Optimal publishing strategies

Kevin J.S. Zollman*
Carnegie Mellon University

March 5, 2009

Abstract

Journals regulate a significant portion of the communication between scientists. This paper devises an agent-based model of scientific practice and uses it to compare various strategies for selecting publications by journals. Surprisingly, it appears that the best selection method for journals is to publish relatively few papers and to select those papers it publishes at random from the available "above threshold" papers it receives. This strategy is most effective at maintaining an appropriate type of diversity which is needed to solve a particular type of scientific problem. This problem and the limitation of the model is discussed in detail.

Agent based modeling techniques are becoming increasingly popular in the social sciences. Among other things, these techniques allow relatively easy comparison across different possible worlds. One can study the effect of structural or individual change on an experimental population *in silico* without engaging in expensive, difficult, and occasionally unethical studies on actual populations. This ability to easily compare different worlds to one another should attract those epistemologists who are interested in comparing the properties of different belief-guiding practices. For philosophers an appealing population to study in this way is the population of scientists. Science represents one of the most successful epistemic enterprises, but it is often plagued with apparently pathological behavior. Understanding which parts of science contribute to its success and which parts might be altered to achieve a more successful enterprise is a project which has interested a variety of scholars, using a variety of techniques.

In this paper, I will bring agent based models to bear on one small aspect of the scientific enterprise, the method by which journals select publications. This model will be used to answer two questions about journal selection strategies. First, ceterus paribus should journals publish more rather than less? This question may seem to have an obvious answer, but earlier work on scientific practice (Zollman, 2007a,b, 2009) may shed doubt on that apparently clear judgment. This question is tackled in section 2. Our second question is how should journals choose from among many publications when they cannot publish them all? This question has been asked and answered in a different fashion by Alvin Goldman (Goldman, 1999, Section 8.11). His model presumes that the editor of the journal can gauge the effect of a particular publication on its readership. This is an

*To contact the author write to Department of Philosophy, Baker Hall 125, Carnegie Mellon University, Pittsburgh, PA, 15213-3890 or email kzollman@andrew.cmu.edu

assumption that we will leave out of this analysis – considering situations were it is true will be left for another day.[1]

In section 1, I will illustrate a particular model, previously used to analyze the effect of unregulated communication in science (Zollman, 2007a,b, 2009). Utilizing this model, we uncover some interesting results. If we hold the publication strategy of the journal fixed, it is occasionally better to publish more rather than less. This result depends on the publication strategy adopted by the journal, however (details of this result are discussed in section 2). If we are allowed to vary both the publication strategy and the amount published we find it is best to publish less rather than more (see section 4). With respect to selection strategies, we find that when constrained to publish relatively few publications the best selection strategy is the one which chooses from among the "above threshold" publications at random. When the percentage of published materials is larger, the best strategy is the one which publishes the papers which represent the greatest successes (see section 4). This coincides more clearly with the strategy we imagine journals use.

This model is limited in a variety of ways. First, it deals only with one type of learning situation which does not encompass every situation that could be faced by scientists. While this lack of generality is unfortunate, I think we are unlikely to get any more general meaningful answers (Zollman, 2007b). Second, it only considers the effect of publication practices for scientists that are working in the same area trying to solve the same problems. That is, it only considers the distribution function of journals, not their service as archivist of scientific work. The importance of this later function should not be understated, but analysis of it will be left to later work.

1 A model of a scientific problem

1.1 Bandit problems

Beginning in the 1950's there was increasing interest in designing statistical methods to more humanely deal with medical trials (see, e.g. Robbins, 1952). A scientist engaging in medical research is often pulled in two different directions by her differing commitments. On the one had she would like to gain as much information as possible, and so would prefer to have two large groups, one control and one experimental group. On the other hand, she would like to treat as many patients as possible, and if it appears that her treatment is significantly better than previous treatments she might opt to reduce the size of the control group or even abandon the experiment alltogether.[2] These two different concerns lead to the development of a class of problems now widely known as bandit problems.

The underlying metaphor is that of a gambler who is confronted with two slot machines which payoff at different rates. The gambler sequentially chooses a slot machine to play and observes the outcome. For simplicity we will assume

[1] As is often true we must exclude interesting cases in order to develop a model which can be understood in some detail. Extending this model to consider the decision procedure recommended by Goldman would not be difficult, but space and time prevent its consideration here.

[2] For example, a recent study on the effect of circumcision on the transmission of HIV was stopped in order to offer circumcision to the control group because the effect was found to be very significant in the early stages of research.

that the slot machines have only two outcomes, "win" or "lose". The gambler has no idea about the probability of securing a "win" from each machine; he must learn by playing. Initially it seems obvious that he ought to try out both machines in order to secure information about the payoffs of the various machines, but when should he stop? The controlled randomized trial model would have him pull both machines equally often in order to determine with the highest accuracy the payoffs of the two machines. While this strategy would result in the most reliable estimates at its conclusion, it will also often not be the most remunerative strategy for the gambler.

The gambler is confronted with a problem that is identical to the clinical researchers problem, he wants to gain information but he would also like to play the better machine. Discovering the optimal strategy in these circumstances is very difficult, and there is significant literature on the subject (see, e.g. Berry and Fristedt, 1985).

While used primarily as a model for individual clinical trials, situations of this structure pervade science. Scientists often must choose between different methodologies in approaching a particular problem and different methods have different intrinsic probabilities of succeeding.[3] For instance, an ecologist may be confronted with a variety of different methodological choices in studying some phenomena of interest, say altruism in animals. The ecologist could persue population genetics models, game theory models, field investigations, or laboratory experiments. She could choose to focus on species at every level of complexity, ranging from bacteria to primates. If her ultimate goal is the deeper understanding of the mechanism underlying altruism, each of these choices may make her more or less likely to succeed.

I (Zollman, 2007a) have also suggested that the bandit problem model fits well with Laudan's (1996) model of paradigm change, since Laudan believes these changes are based on something like expected utility calculations. So not only might this model apply to small choices of methods for individual scientists, but might fit the big scale choices of scientific groups as well.

In this paper the payoff of the bandit is analogous to a successful application of a given theory. One attempts to apply a theory to a particular case and succeeds to different degrees. Individual theories have underlying probabilities of success which govern the likelihood that they can be successfully applied in some specified domain. Since they are interested in successful applications of a theory, scientists would like to work only on those theories which are mostly likely to be successfully applied.

1.2 Journals

When a scientist successfully applies a theory, he usually seeks to advertise this success by publishing the result in a scientific journal. Whether this is a successful empirical study, a new mathematical result, or a new theoretical explanation for a phenomenon, results are often judged by some standard which rates the degree of their success. Theoretical explanations can be better or worse and empirical results can be more or less significant. Journals attempt

[3]Success here can be defined in almost any way you like it. Different methods might have different probabilities of generating true explanations, adequate predictions, usefull policy suggestions, etc. My hope here is to develop a model which would appeal to people with varying commitments about what constitutes scientific success.

to rate this level of success and publish those results which represent the best applications (or at least this is how we imagine they work).

Scientists then observe what the journals publish, as well as their own work, and then reform their belief about which methodology is superior. To model this process we will assume that individuals have beta distributions over the expected effectiveness of a given methodology and they update their beliefs using Bayesian reasoning.[4] The scientist will attempt to apply the methodology they currently think is best, and submit their results to a central "journal."[5] The journal then uses one of a variety of strategies to choose what results to disseminate. After a round of experiments every scientists updates her belief on the basis of her own experiment and on the basis of the other experiments published by the journal. Although what the journals publish contains some information about the results of unpublished attempts, we will assume our scientists do not speculate in this way – they treat what they read as if those are the only results obtained. After modifying their beliefs, scientists will then attempt to apply the methodology they currently think is best and the publication process is repeated.

This will allow us to analyze the effect of different methods of selection for an individual journal. That journal could select both how many and which individual results to publish and distribute to the group. We will vary this strategy in an attempt to answer two major questions about journal publishing strategies. First, all else being equal is it better to publish more results? And, second, given that we have a fixed number of papers to publish how ought the material be selected?

One feature of our model that is worth noting is that there are no differences amongst the individual scientists in terms of the quality of their work. Each individual attempt to apply a methodology is equally informative about the underlying success of that methodology as any other. As a result, one important purpose of editorial review in journals has been assumed away – namely, the need to remove misleading, uninformative, or dishonest papers from circulation. But even assuming this process away leaves us with interesting questions that can be investigated. I will leave it for another time to consider how journals should best deal with situations where papers might not all be "above threshold" in this way.

2 Is more better?

In earlier work, I (Zollman, 2007a,b, 2009) analyzed a model of unmediated communication using the bandit problem model of scientific practice. In this model individual results were communicated between scientists without regard for their quality, either two scientists communicated all their results to one another or they communicated none. In this model I found that reducing the amount of information transmitted could significantly improve the long run reliability of a group of scientists.

[4]Individuals begin with initial parameters drawn from a uniform distribution over [0,4]. Detailed explanation of applying beta distributions to this circumstance are provided in (Zollman, 2007b, 2009).

[5]Each attempt to apply a given methodology is modeled as 1,000 flips of a biased coin that comes up either 0 or 1. The higher the total number of 1's the greater the success. Each methodology has an intrinsic probability of producing a 1. In the simulations reported here these were .5 and .49 respectively.

This is, of course, a rather surprising result. One's initial intuition would be that the communication of more information would make the individual (and thus the group) more reliable. However, in this case it does not. When many results are communicated short term erroneous results cause the entire community to adopt (or abandon) a particular theory. Since, in bandit problems, information is only gained by observing outcomes of a particular project, once one has been abandoned the group never learns of its error.[6]

In this way, diversity in scientific practice is important for at least a short while. This diversity is destroyed too quickly by the widespread communication of results and limiting the spread of information is one way that this diversity can be maintained.

This earlier model is realistic for the communication of result informally, perhaps using the internet or post to distribute results to others that one knows. However, this is not an accurate model for journals. Journals exercise control over what is published, and perhaps this might be able to maintain diversity even if a large number of results are communicated. Considering this model of journal publication, we can ask if this same result holds true, i.e. is less information better?

It will turn out that whether and to what extent information is usefull depends on what strategy is used by the journal to select the information it distributes. We have chosen four potential strategies here, the reason for these choices will become clear in the next section.

As a baseline we will use the strategy *Random*. *Random* simply selections a certain number of results without considering any feature of them and distributes them to everyone. Given the earlier results from (Zollman, 2007a,b, 2009), we should expect that group reliability would decrease as the number of publications increases, and this is exactly what is found. Figure 1 shows the results for a simulation involving a single journal and ten scientists. Here we see clearly that an increase in the number of publications results in worse groups.

This assumes that on each round the journal randomizes which results it chooses to publish. It also might simply choose a random group of "heroes" and always publish the results from these people, regardless of quality. Interestingly, with this strategy more information is superior (as can be seen in Figure 2).

Both of these strategies fail to capture anything like the selectivity exercised by real journals. In keeping with the interpretation offered above, we regard each individual trial as representing an attempt to apply a given scientific methodology to a problem in its domain. Since journals generally only report significant results that are the successful application of a given methodology, we should consider a strategy which is purely meritocratic in this way. The strategy *Best* does just this. It looks at all the results reported and chooses the ones which represent the most successfull application of a particular method without regard from which method that individual result comes.

Figure 3 shows the reliability of the group compared against the number of publications for this strategy. Here we see that the previous results no longer hold. There is now a peak in reliability around where around 60% of the reported results are published. The reader might remark at how large a number this is, certainly higher than the acceptance rate of many of the best journals. However,

[6]It is important to note that information here is totally cost free. The harm comes from the effect that short term biased results have on behavior and by extension on the availability of information.

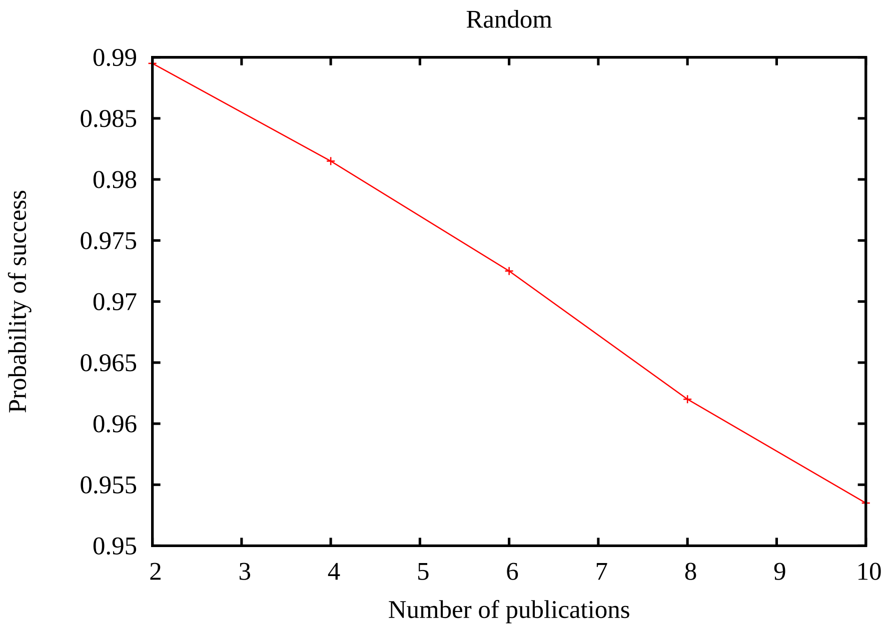

Figure 1: Random on each round

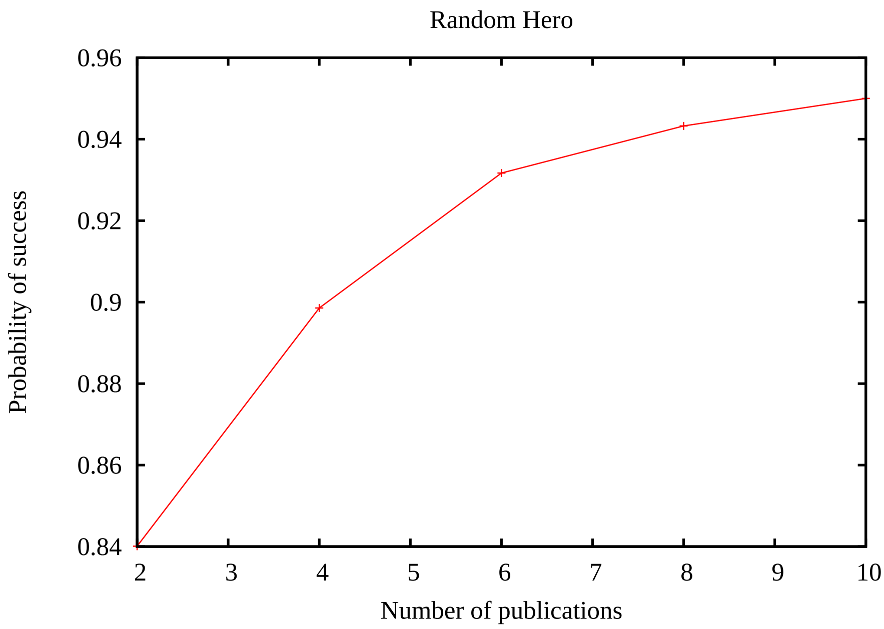

Figure 2: Random Hero

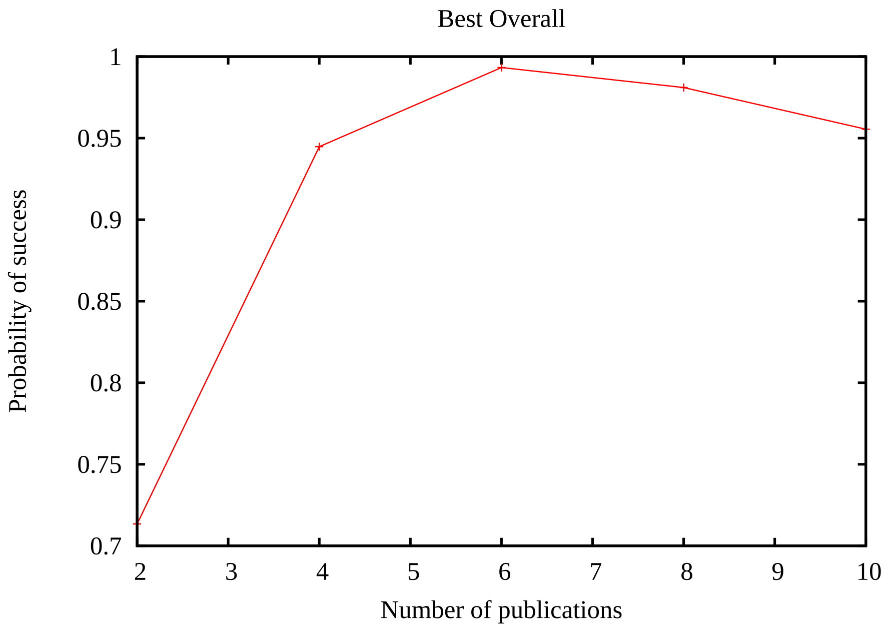

Figure 3: Best publishing strategy

recall that an idealizing assumption of this model is that all results are not erroneous. That is, the only subject of variation is natural statistical variation, incompetence, misconduct, or irrelevance are not captured by this model.

Best makes no attempt to be evenhanded, and thus no attempt to maintain diversity. Nonetheless it appears to succeed in doing so. One might also imagine a publishing strategy which attempts to give "equal time" to competing methodologies. Like *Best*, *Best-from-each* uses successful application to judge methodologies but it chooses the best from each of the competing methodologies to publish. Figure 4 shows the relative success of this publishing strategy. Again we see a peak in the optimal number of publications, although here it is lower – around 40% of the possible publications.[7]

Given that information does not appear to uniformly harmful in these models one might be inclined to dismiss the earlier results from (Zollman, 2007a,b, 2009). However, doing so would miss important distinctions. Our model of publication strategies here assumes that there is central authority controlling the distribution of scientific results, while the earlier models assumed that information transmission was done without regulation. While the central authority model is appropriate for the informational distribution of science in the last few centuries, it is becoming less accurate in recent times. The invention of internet repositories for unvetted articles (like arXiv) is changing the methods of distribution and moving science toward a more unregulated system of information distribution.

[7] The jump at 10 papers should be taken with a grain of salt. Since there are only 10 scientists in these simulations, that represents publishing each and every experimental result which renders the decisions procedures identical. More will be said about the comparison in section 3.

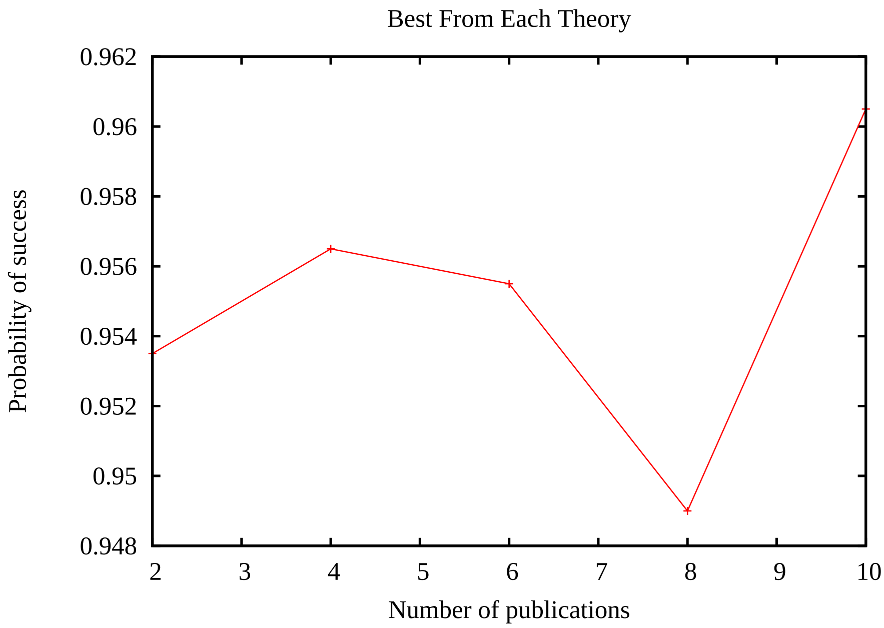

Figure 4: Best from each arm

In addition, our analysis of information holds fixed the publication strategy of the journal. In section 4 we will see there is a very real sense in which information is harmful when we allow both the number of publications and also the publication strategy to vary.

3 The optimal selection strategy

Before now our primary concern has been with whether more or less information is better. This question was primarily motivated by some concerns over the widespread – and unregulated – distribution of results. We found that there were no consistent answers about the benefits of widespread distribution of information. For some publication strategies it is better for journals to publish more rather than less, but on others it is better for them to publish less rather than more.

These are hardly the only questions we might ask about journal publication strategies, however. Here I will consider three questions:

1. Should we publish negative results?
2. Should we give "equal time" to competing theories, methodologies, or paradigms?
3. Is bias toward arbitrarily chosen famous people harmful?

Each of these questions can be answered relatively easily using the simulation methodology presented here.

3.1 Negative results

Experiments may fail in a variety of ways. Individuals may fail to uncover statistically significant correlations between two variables of interest, predictions of a theory may fail to occur, or mathematical theorems may turn out to have counterexamples. Sometimes these negative results produce important positive results, like the celebrated negative results of Michaelson and Morley, but more often such results never see the light of day.

This is unquestionably a feature of the publication strategies of science, but it is not always clear why such a system should be. Single surprising results may be more informative that single instances of negative results, but a publication strategy that ignores negative results entirely may serve to bias a field. Thinking of our model of methodology choice, if a journal only publishes the successes of a given methodology and not its abject failures, it may produce a skewed representation of the effectiveness of that methodology.[8]

Because of concerns like this, there is a growing interest in archiving negative results and distributing them among scientists. For instance, an online journal has been created for the publication of negative results in ecology and evolutionary biology and others have been developed for other fields.

We can use this simulation methodology to evaluate whether one type of negative result should be published. Under our interpretation of bandit problems, a negative result is an attempt to use a methodology to produce an important result which fails to do so. This is somewhat different from another type of negative result, the statistically insignificant study. While studying this is warrented, it is beyond the scope of this paper.

Figure 5 shows the comparison of two publication strategies. *Best* publishes only the most successful applications of any methodology, that is only positive results. *Best and worst* publishes (in equal amounts) the most successful and the most unsuccessful attempts to apply a methodology. For relatively small numbers of publication (publishing less than 50% of available experiments) we find that the publication of negative results is indeed helpful. Knowing that an attempt failed balances against the occasionally misleading grand successes that particular methods could have.

However, if journals are publishing a large proportion of the (informative) studies conducted, then publishing negative results is no longer helpful. The potential danger of being misled appears to vanish, and negative results are no longer informative.

3.2 Equal time

Another vexing question which occasionally surfaces regarding journal publication is whether or not they have an obligation to give competing methodologies equal time. Such disputes can often be very bitter with different methods attempting to control editorial boards and influence decisions at conferences.[9]

[8]In a summary statement, the *Journal of Negative Results* for ecology and evolutionary biology says, "[The selective publication of positive results] may lead to a biased, perhaps untrue, representation of what exists in nature. By counter-balancing such selective reporting, JNR aims to expand the capacity for formulating generalizations."

[9]David Hull (1988) provides a rather detailed account of such a dispute in taxonomy research.

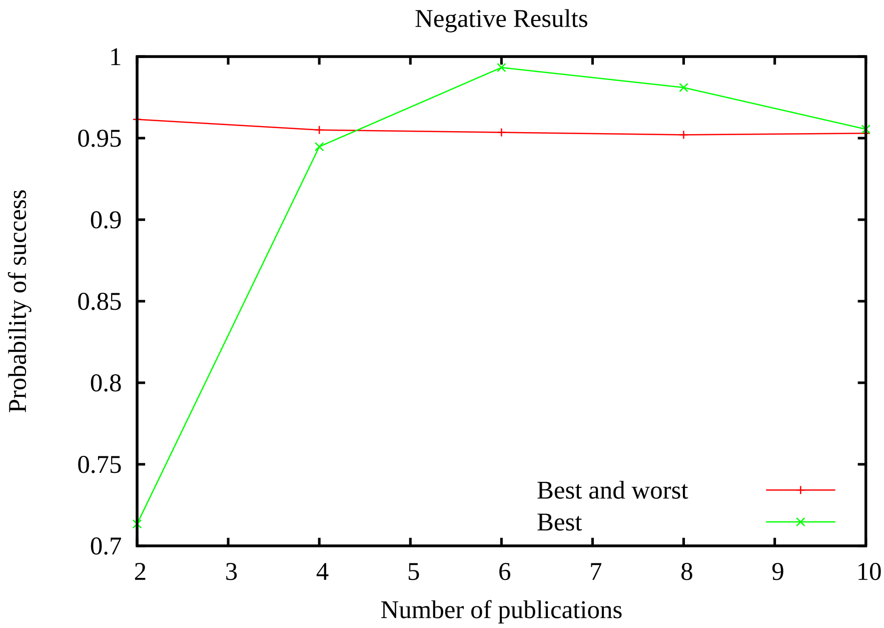

Figure 5: Comparison of two strategies regarding negative results

It should be clear that an arbitrary bias toward one methodology would be harmful in resolving a dispute between two methodologies. But, should journals make a conscious attempt to be even handed? That is, should they publish a paper of lower quality in order to ensure that a particular methodology is not given short shrift? In terms of our model, the question is addressed by comparing the *Best* publication strategy with *Best-from-each*. The former ignores concerns of equal time and only publishes the most successful application of any methodology, while the later chooses the best application from each and publishes an equal number of both.

We find an almost identical result to the earlier one above (see Figure 6). For small numbers specific attention to equality is important, but as the proportion of studies published grows the benefit is lost. Again, this occurs for the same reason, small number of publications can be misleading and some attempt to be evenhanded reduces the chance for early abandonment of a superior theory. However, as the percentage of published studies increases the risk of such a problem decreases and the value of evenhandedness is swamped by the information available under the *Best* strategy.

3.3 Random heroes

It has been noted that certain famous individuals often are more likely to succeed in publishing their results than other less famous ones. This could be the result of a common cause, they are famous and more successful both because they are superior scientists, or it could be that fame results in greater ease in publishing. The former represents no problem for the meritocratic system of science, but the later gives some concern. How worrisome would this be, if it did in fact

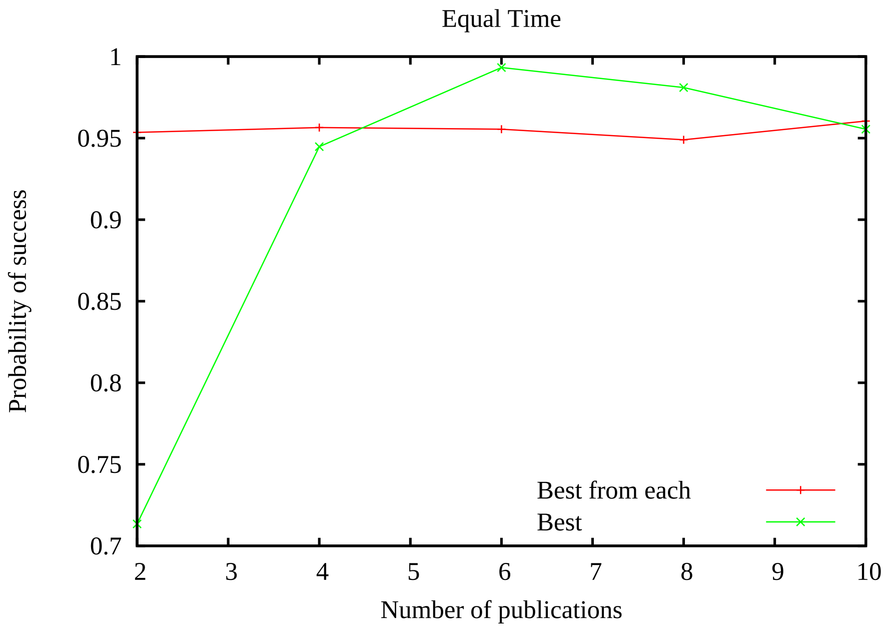

Figure 6: Comparison regarding equal time

occur?

Here we will look at an extreme example of this behavior, represented by the *Random heroes* strategy described above.[10] At the beginning of a simulation, a certain number of heroes are chosen at random, and their results are published regardless of content. How does this strategy fair compared with a meritocratic one?

Figure 7 shows the comparison between *Best* and *Random Hero*. Interestingly for very small publication rates *Random Hero* outperforms *Best*, but very quickly *Best* overtakes *Random Hero*. This underwrites our general intuition that an arbitrary bias toward particular figures in publication is harmful.

4 The "best" publication strategy

Given that we've considered a variety of different publication strategies, it is natural to ask which strategy is superior in given circumstances. Caution should be taken here especially, since our selection is only limited to a few potential strategies and so we may have missed a superior system.

So far we have only considered situations with a single journal, but we can also expand our search to include 2 or more journals publishing any number of publications.[11] Table 1 illustrates which publication strategy is best for

[10] This is not as terrible as one could imagine, however. Recall that in our simulation every experiment is informative and so every publication is at least above a certain threshold of usefullness.

[11] When there is more than one journal, each scientist chooses between 1 and n journals to subscribe to. He submits his results to every journal he subscribes to and reads everything in every journal he subscribes to.

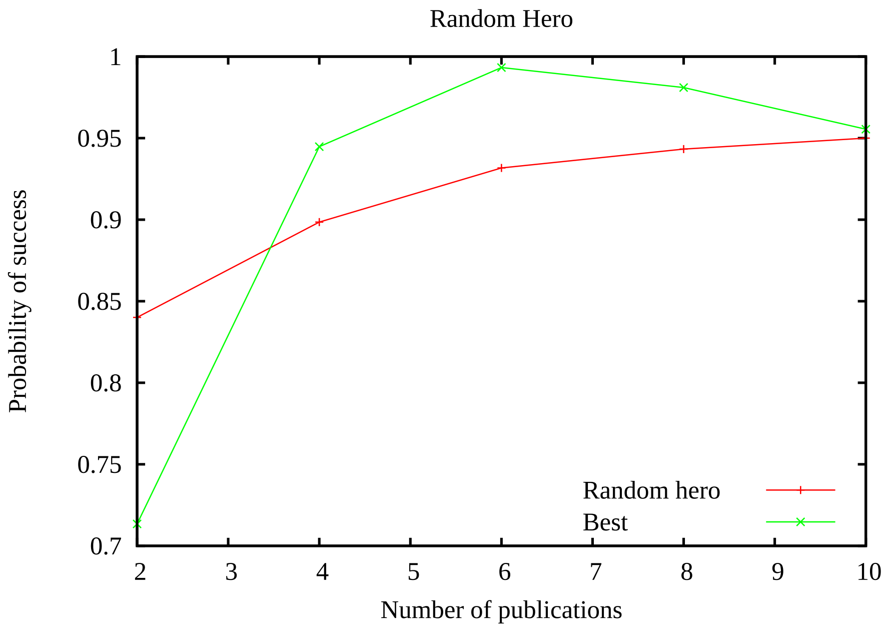

Figure 7: Comparison of two strategies regarding negative results

	Number of publications/journal			
Number of journals	2	4	6	8
1	Random	Random	Best	Best
2	Random	Random	Best	Best
3	Random	Random	Best	Best
4	Random	Random	Best	Best
5	Random	Random	Best	Best

Table 1: The optimal publication strategy for various numbers of journals and publications

each of these cases. For small numbers of publications surprisingly the best publication strategy is *Random*, the strategy which employs no discrimination at all. As the number of publications approaches a larger percentage of the "above threshold" submissions, then the strategy which chooses the best from the available publications becomes the optimal strategy.

The reason for this difference is not difficult to understand. In bandit problems there is some benefit to short term diversity. Groups of learners can converge to a suboptimal methodology, and the probability of this is heightened by giving everyone access to the same, relatively small, and potentially biasing information. When initial evidence suggests that the suboptimal methodology is superior, this evidence can convince everyone to abandon the superior methodology. While this possibility can happen with the *Random* publishing strategy, it is more likely when the publishing strategy explicitly seeks out the best (and thus potentially most misleading) publications.

A yet more surprising result comes when we compare across the all available parameters. If we were to choose the number of journals, number of publications, and publication selection strategy which made the group most likely to settle on the best methodology, we would settle on a single journal publishing only two publications per round using the *Random* publication strategy.

This confirms the earlier results regarding unregulated communication – that less information is superior. Fewer publications are superior for the same reason that less communication is superior in the unregulated case; diversity is maintained better by restricting information flow.

5 Conclusion

This last result is perhaps the most surprising. This suggests that publishing relatively few publications and exercising very little editorial choice represents the best strategy amongst those surveyed. This fits with the earlier results on informal distribution of information, where it was better to limit information than have it widely distributed. But the results depart from the current practice of science where the number of journals is growing and there is significant effort dedicated to making informed choices among different publications.

In addition, we have addressed some other questions regarding journal publication practices. We found that when relatively few papers are published, it is important to be consciously evenhanded and it is also beneficial to publish failures as well as successes. However, as the number of publication grows such biases are no longer helpful. Finally, we found general confirmation that arbitrary bias toward particular researchers is inferior to a publication strategy which does not have that bias.

As noted in the beginning, we have limited ourselves in a few ways. First, these conclusions only hold for scientific problems that fit the bandit problem model. If we expand to include other types of problems, we are likely to see very different results. Also, our model is limited by the particular strategies we choose to consider. For instance, none of our strategies react to the general consensus of the community of researchers. Since we know diversity has some benefit in bandit problems, a strategy which actively maintains this diversity might do better than the more limited one we considered here. However, the level of information required by such a strategy might be so high as to make the strategy infeasible.

Overall this study may give us reason to question many apparently obviously good practices in science. Some practices, it seems, are justified (like avoiding random heroes) but others may not be so helpful.

References

Berry, Donald A. and Bert Fristedt. 1985. *Bandit Problems: Sequential Allocation of Experiments*. London: Chapman and Hall.

Goldman, Alvin. 1999. *Knowledge in a Social World*. Oxford: Clarendon Press.

Hull, David. 1988. *Science as a Process*. Chicago: University of Chicago Press.

Laudan, Larry. 1996. *Beyond Positivism and Relativism: Theory, Method, and Evidence*. Boulder: Westview Press.

Robbins, Herbert. 1952. "Some aspects of the sequential design of experiments." *Bulletin of the American Mathematical Society 58*: 527–535.

Zollman, Kevin J.S. 2007a. "The communication structure of epistemic communities." *Philosophy of Science 74*(5): 574–587.

Zollman, Kevin J.S. 2007b. *Network Epistemology*. Ph. D. thesis, University of California, Irvine.

Zollman, Kevin J.S. 2009. "The epistemic benefit of transient diversity." *Manuscript*.