

Warning Signs of Bogus Progress in Research in an Age of Rich Computation and Information

Yi Ma, ECE, University of Illinois

November 4th, 2007

More and more frequently, some of my students get very confused about papers published in the literature and ask if they should follow the same style in order to get their work published more quickly. My answer to such questions is always: The only thing that is worse than not publishing is to publish wrong or bogus results. As wrong results are easy to tell and correct, the question boils down to what constitutes “bogus results.” Very surprisingly, I could not find any precise, satisfactory description for them in the literature.¹ The reason may be that there has never been a period in the history of science when bogus results could be produced and reproduced with such efficiency and at such volume, due to the rich computational and information resource now available. So I have decided to give it a try myself. Realizing that subtleties of the definition could differ a lot for different people or for different fields, I hereby try to only identify some of the most common troubling signs of bogus results, tailored to the resource-rich scenario, and hope to provide some guidelines for my students.

1. Justify by Successful Instances. With access to tremendous computational resource and data, it has become relatively easy for researchers to produce successful instances to problems that are of high complexity, say NP-hard problems. For instance, using some heuristic or random search techniques², one may have a good chance to find correct solutions to an inherently hard problem, at least to some of its instances in a particular application. Such successful examples, no matter how many are given or how impressive they might seem to be, are no evidence or proof that the proposed solution has truly alleviated in any way the difficulty of the problem in question, let alone solving it. Nevertheless, good heuristic solutions often point to promising directions for a more systematic investigation. That is the true value of a heuristic solution, especially one that can later be rigorously justified.

2. Compound and Conquer. Powerful computers have made it possible to put together solutions to smaller subproblems into a large software system to tackle, in a “holistic” fashion, a comprehensive, challenging task that previously was deemed unlikely. While such a system may demonstrate some success and have good technological values, it does not necessarily help better understand whether the task indeed consists of any new, unsolved problems and how much progress one has truly made

¹The closest I get to is the article by Robert L. Park on “The Seven Warning Signs of Bogus Science” published in *The Chronicle Review*, January 31, 2003. See <http://chronicle.com/free/v49/i21/21b02001.htm>

²They are not to be confused with random techniques that approximate certain NP-hard problems with guaranteed chance of success.

towards solving them. The danger of compounding too many subproblems together is at least two-fold: First, if these problems have different assumptions and their solutions have different conditions, then no logically consistent conclusions can ever be drawn. Second, the complexity of the compounded problem can be too complex to analyze rigorously. This has often been used as a lame excuse to dismiss analysis and instead to look for an alternative, grossly simplified solution. There is a good reason why “divide and conquer” has always been a tenet intimately cherished by modern science.³

3. Results Too Complicated to Reproduce. The cardinal rule of science for accepting any results is that the same results can be obtained independently by others following the proposed methods and experiments. In a resource-rich age, many softwares and experiments can be designed, almost on purpose, to be so complicated that it is daunting for others to reproduce if they do not exert the same amount of resource. This makes review of such work particularly difficult. Many peers can only rely on their experience or common sense to judge the validity of the work, which allows false results or claims to pass temporary scrutiny with high probability. For this reason alone, we should in fact down-play the importance of one-time conference publications. Even for archival publications, we should enforce retraction of work that others cannot obtain the same experimental results or if the authors refuse to make their code and data available to others for scrutiny.

4. Reinvent the Wheel in Mass-Production Scale. Rich computation and information resource make it easier for researchers empirically to screen their guesses, ideas, or hypotheses. Researchers can arrive at results new to themselves with unprecedented frequency. However, many of the results may have been discovered long before or in another field. There is still good value if old results are reinvented in a new context and at a better time, as long as they are properly acknowledged. However, the same rich resource, used for expediting discovery, has not been equally adequately used to verify the novelty of the discovery, very often due to the unwillingness by the researchers⁴ or the lack of communications among different research fields. As result, many rediscovered results are labeled as new, using different jargons or terminologies. A self-claimed independent research field should bear the burden of proving beyond any doubt to other fields its unique contributions to the scientific community, in terms of novel discoveries and methodologies that can truly stand the test of time.

5. Tackle Ill-posed Problems Directly. Ill-posed problems are not unimportant. In fact, most difficult tasks in the real world are ill-posed because they often contain inconsistent or conflicting objectives.⁵ In scientific research, the role of ill-posed problems is to inspire new well-defined problems, either an idealized version or a smaller part of the

³The defendants of the traditional Chinese medicine always claim that Chinese medicine takes a holistic approach to heal people, as opposed to the “Western” medical practice examining people to the level of molecule. While many empirical prescriptions from ancient Chinese medical practice remain mystically effective, the overall methodology simply cannot stand any scientific test.

⁴Honest ignorance about past results are often excused, and in fact are often credited as independent work. The famous example is about the solvability for polynomial equations of degree higher than four – Abel proved it first and Galois did it independently. But deliberately exaggerating one’s originality is doomed to be mocked by history.

⁵For example, the problem of regulating the stock markets contains complicated economical, social, political, technological, and even cultural factors.

original ones, which can then be addressed through scientific means and be provided with consistent solutions. This is the only way we can truly improve the overall solution to the real-world problems, ill-posed or not. Unfortunately, rich resource may give some researchers the illusion that they are able to directly tackle a complex real-world task empirically and no longer need to follow the rigor of scientific and mathematical investigation. To them, there is no fundamental difference between the development of a commercial product and a piece of scientific work.

6. Solicit Popularity and Publicity. Number of citations used to be a good indicator for the importance of a piece of scientific work. That may no longer be the case in a resource-rich age.⁶ With ever more people having access to ever more literature, they are biased to choose and use work that is more accessible, simple, and intuitive. As a result, a watered-down, late reinvention of an important result might attract the most citations. Thus, popularity is not necessarily an indication of the originality, depth, and ultimate importance of the work. Deep scientific investigation should not be degraded as a popularity contest, especially in the new age of universal accessibility. Science is not a part of the entertainment business, where no publicity is bad publicity. Modesty is still a cherished virtue in science. One should not use clever examples, cute demos, or fancy stories to attract attention (from nonexperts) but hide the lack of substance (from experts).

7. Occam's Razor Reversed. These days it is not uncommon to see an algorithm or solution that uses a complex parametric model to represent the data since powerful machines allow people to simulate and search in ever larger parameter space. However, people tend to abuse the computational resource and often use unnecessarily complex models for their problems at will. The redundant parameters are often tuned manually or set heuristically, in order to obtain good testing results.⁷ This makes it almost impossible for others to reproduce the same results with the same model, unless provided with the exact setting of the parameters. Even worse, if the performance of a model on the testing data is not satisfactory, the reflex is to further increase the complexity of the model by introducing additional variables and parameters. However, any performance gained in this way is at the expense of the validity of the approach: The testing data has become part of the training, with a human expert in the feedback loop.

8. Monkey Collects Corn Cobs. Before an old and more basic problem is completely solved, one would accept a partial, suboptimal solution and use it in a solution to a new or supposedly more challenging problem. When a partial solution or result is found, one moves on again. Such a constant shift of focus is very counterproductive.⁸ When others try to revisit some of the problems, one would claim that the problems have already been addressed and there would be no need for any improvement. It is impossible to measure progress or compare different solutions if they are all to some

⁶Of course, truly important work normally does have a large number of citations.

⁷The physicist Fermi once quoted the mathematician John von Neumann: "with four parameters I can fit an elephant, and with five I can make him wiggle his trunk."

⁸An old Chinese saying describes this as "a monkey collects corn cobs." In a corn field, the monkey constantly tries to grab new cobs while at the same time losing the ones in his arm. So the total number of cobs he can collect is always the same.

extent partial. From a pedagogical perspective, from partial results like that, it is impossible to develop a systematic body of knowledge that can be effectively transferred to future students and researchers.

Disclosure (added on December 14): Incidentally, I read a recent article by Donald Geman: “*Ten Reasons Why Conference Papers Should be Abolished*,” which can be found on his website. It reflects some similar thoughts and feelings about publication. I have then decided to follow his style and add this disclosure. Over the past ten years, I myself could have published papers that fall into some of the above categories, mostly under peer pressure or the pressure from research funding and tenure promotion. I wish I could have done better so that more of my publication does not belong to the above categories.