Improving Management Science: Problems and Solutions

J. Scott Armstrong
Wharton School, University of Pennsylvania
Philadelphia, PA

Armstrong GAMMA Award for Brilliance in Research in Marketing
2016 Global Marketing Conference at Hong Kong
July 22, 2016

Guidelines for science-HK-42 (7/21/16)

Benjamin Franklin called for the *discovery* and *dissemination of useful knowledge* *(Franklin 1743)*

Survey time:

1. How many of you are are involved in some way with scientific research in the social and management sciences (e.g., researcher, teacher, administrator, consultant, lawyer, policy maker. . . . .?) [Raise your hand & hold it up to be counted]
2. How many of you also believe that fewer than ten percent of papers published in the journals that you read provide useful scientific knowledge? [Keep hand hand up; otherwise put hand down for our estimators.]

I will discuss this issue today. I have been doing research on science since the 1970s.

Why this talk is relevant to you

We have a problem:

- Policymakers, managers, journalists, journal editors, reviewers, funders of research, teachers and researchers do not understand (or choose to ignore or violate) important scientific principles.
- Most papers in the management and social sciences provide useless and misleading advice, and the small proportion providing useful scientific findings is decreasing.

There is a solution:

- You can identify papers with useful scientific knowledge by using a checklist that requires only a few minutes to apply, if you take two hours to read our "Guidelines for Science" paper.

Scientific principles provide the basis for progress in society

Let me describe how scientific principles can lead to useful knowledge for business and public policy management.

Study 1: Can advertising effectiveness be improved by science?

Assume that a review of experimental findings formulated operational persuasion principles.

Assume further that 195 principles and a checklist scheme—called the Persuasion Principles Index (PPI) to rate ads based on their compliance with the principles resulted, and that the PPI checklist:

- produced predictions of which ads are more effective that were much more accurate than predictions by experts or by copy testing.
- could also be used to identify how to improve the persuasiveness of the ads.
- and a one-hour self-administered training course was available for free on the internet.

How would potential users respond? (Take a moment to speculate.)

Possible reactions to this discovery in advertising

1. Universities incorporated the PPI into their courses.
2. Some advertisers cut ties with their agency and did their own advertising.
3. Advertising agencies used the PPI as one of their tools because it was more effective

This is a true situation. See Predictive Validity of Evidence-based Advertising.

Actual outcomes? No measureable progress to date.

But it is early days… Scientific discoveries spread slowly, as illustrated by the many decades delay before the highly effective scientific procedures for personnel selection were adopted by sports teams (See *Moneyball*).

A checklist should allow clients as well as forecasters to assess credibility by determining whether the forecasting methods complied with scientific principles.

The checklist should allow a trained novice to judge whether a forecasting system complied with a list of guidelines.

The guidelines should be available for free online with a checklist and instructions.

How would such a *Golden Rule of Forecasting* checklist be received?

Reception of the Forecasting Tool

- This case too is true. The Golden Rule of Forecasting was published in 2015.
- It contains a checklist of 28 guidelines.
- The violation of a typical guideline increases forecast error by 40%.
- It is free (available at ForPrin.com, GoldenRuleofForecasting.com, and as an app).
- Can be used by experts and novices.
- And the reception?

To date, the paper has attracted 37 Google Scholar citations, and that number is growing rapidly.

Can be used in practical situations to rate the credibility of the models used to forecast global warming, the effects of Brexit, and other key issues.

Statisticians who do forecasting pay little attention to the Golden Rule. Why? They do advocacy research (as described later in this talk).

No evidence of widespread adoption in practice. Why?

My opinion: Clients for important forecasts typically want forecasts not for accuracy, but to support their policies and procedures. Also, the Golden Rule of Forecasting seems *too simple*.

Assumptions behind these two examples of the development of useful knowledge

- Useful knowledge is *most effectively and efficiently* developed from scientific evidence
- Knowledge must be formulated as operational guidelines to be *useful*
- Guidelines must be presented as checklists to be *used*.
- Checklist use should be independently verified to *ensure implementation*
- Decisions should be based on "compliance with the guidelines" to *deliver improvements*

What is “useful science”?

A process that studies *important problems using evidence on multiple hypotheses from experiments*.

The process involves using

1. cumulative scientific knowledge
2. systematic measurement with valid and reliable data
3. valid and simple methods for analysis
4. logical deduction that does not go beyond the evidence presented

and requires

5. full disclosure of all information needed to replicate the process.

“What’s the problem? The scientific method is well known!”

The scientific approach is routinely violated in the management and social sciences.

The problem has gotten much worse.

One reason why there is a lack of useful scientific papers

Adam Smith asked why Scotland's relatively few academics—who received little funding from the government—were responsible for many scientific advances in the Industrial Revolution, while England's large number of academics—who were well supported by the government—contributed little.

The question contained his answer:

- with government support for academics in England, there was little motivation to do useful research (Kealey 1996, p. 60-89).

The approach used in this study

- Reviewed definitions of science from over the centuries to identify scientific principles
- Developed understandable operational guidelines for implementing the principles
- Revised guidelines using the findings of our review of experimental research.

Objectivity vs. advocacy

- Objectivity is the primary principle of the scientific method
- Advocacy involves searching only for evidence that supports a desired hypothesis, therefore
  - is *not* objective
  - is *not* scientific

Advocacy is common in the management sciences, as was shown by two audits:

- 64% of the 120 empirical papers in *Management Science* from 1955 to 1976 used advocacy (Advocacy and Objectivity in Science, 1979).
- 74% of 1,700 empirical papers in six leading marketing journals from 1984 to 1999 used advocacy (Hypotheses in Marketing Science, 2001).

Advocates use non-experimental data

A preferred hypothesis is easily supported by applying multiple regression analyses to non-experimental data (see Illusions of Regression Analysis).

Illustrated by:

- Analyses of *non*-experimental data suggest that competitor-oriented objectives—such as market share—lead to increased profits for firms, but
- Analyses of *experimental* studies find that market-share objectives harm the profitability and survival of firms (Competitor-oriented objectives).

Complex writing by experts is persuasive

Reviewers rated papers more highly when:

1. abstracts were written in more complex ways (Armstrong 1980b).
2. papers included irrelevant complex mathematics (Eriksson 2012).
3. papers included irrelevant words related to neuroscience (Weisberg et al. 2008).

Mahoney's (1977) experiment obtained reviews of a paper from 75 (unsuspecting) reviewers.

Two versions were used:

- version A *supported* current beliefs,
- version B *challenged* current beliefs.

Those who received

- version A typically *accepted* the paper pointing to its *sound* methodology,
- version B typically *rejected* the paper pointing to its *poor* methodology.

University incentives lead to violations of the scientific method

1. Journals typically insist on tests of statistical significance for empirical studies, even though it is an invalid and harmful procedure.
   (Blinding us to the obvious?)
2. Researchers are rewarded for publishing papers in scientific journals—not for discovering and disseminating useful findings.
3. These incentives have led to an increase in cheating, such as proposing hypotheses after data have been analyzed.

   A survey of management faculty found that 92% knew of researchers who, within the previous year, had developed hypotheses after analyzing the data. (Bedeian, Taylor & Miller 2010)

Complexity can conceal cheating

Cheating is often hidden by complex writing.

For an extreme example, computer software (SCIgen) randomly selects complex words that are commonly used in a topic area. The software then uses grammar rules to produce academic papers.

- At least 120 such papers were published in reputable peer-reviewed scientific journals (Lott, 2014). The title of one such paper was, "Simulating Flip-flop Gates Using Peer-to-peer Methodologies."

Effects of violations of scientific principles

A meta-analysis of 804 replication outcomes in 16 studies in 7 areas of management science found that conflicts arose in 46% of the studies (Hubbard 2016, p.140-141).

Ioannidis (2005), described how incentives, flexibility in research methods in medicine, the use of statistical significance testing, and advocacy of a favored hypothesis, led to incorrect findings.

An empirical analysis of economics journals by Doucouliagos and Stanley (2013) concluded that for topics where established researchers strongly favor a hypothesis, journals avoid publishing unsupportive studies.

Guidelines should be consistent with experimental evidence

- Guidelines lacking evidence can lead people to use invalid guidelines more consistently.
  - For example, the Boston Consulting Group (BCG) matrix for portfolio planning is widely used even though it has been shown to be harmful to decision-making (Effects of Portfolio Planning Methods on Decision Making).

The 26 guidelines cover six aspects of the practice of science *(number of guidelines)*

1. Selecting problems *(6)*
2. Designing a study *(3)*
3. Collecting data *(2)*
4. Analyzing data *(5)*
5. Writing a scientific paper *(6)*
6. Disseminating the findings *(4)*

Guideline 6: *If you need funding, ensure that you will have control over all aspects of your study.*

- Beginning in the 1930s, the US has gradually increased regulations related to the speech of scientists.
- 'Currently, no research using human subjects may be done in institutions receiving federal research funds unless an "Institutional Review Board" (IRB) licenses and monitors it'. (Schneider 2015, first page of "Introduction").
- The US government now controls
  - what can be studied
  - how studies are designed
  - how studies are reported.
- Those restrictions apply if the government supports the institution (e.g., most universities) even for researchers who receive no funding for their own research projects.
- Many of these restrictions likely violate the First Amendment (Free Speech) in the U.S. Attempts to restrict speech as a way to improve welfare in the U.S. have failed all of the many experimental studies to date (Ben-Shahar & Schneider 2014).

Prior knowledge should be based on experimental evidence. For example, use findings from prior experiments to select causal variables.

Increasingly, researchers ignore prior evidence and analyze non-experimental data without having developed evidence-based hypotheses. That is an invalid approach to knowledge development (*Illusions in Regression Analysis*).

> Ziliac and McCloskey (2004) found that 32% of papers published in the *American Economic Review* in the 1980s used statistical significance to select causal variables.
>
> By the 1990s, it increased to 74%.

Guideline 8: Test Hypotheses by Experiments

- Non-experimental research concludes that consumer satisfaction surveys improve consumer satisfaction.
- In contrast, experimental research concludes just the opposite (Ofir at al 2001).
- This also applies to higher education; and learning by students also suffers substantially

Guideline 9: *Compare multiple reasonable hypotheses*

Chamberlin (1890) observed that sciences making the most progress used experiments to test multiple reasonable hypotheses. Doing so encourages objectivity.

Kealey's review of the scientific research behind the agricultural and industrial revolutions provides evidence from natural experiments (1996, p 47-89).

The approach of multiple reasonable hypotheses is rare in research in the management sciences. For example…

1. Of 120 empirical papers in *Management Science* from 1955 to 1976, only 22% used the method of multiple reasonable hypotheses (Armstrong, 1979).
2. Of 1,700 empirical papers in six leading marketing journals from 1984 to 1999, only 13% used multiple reasonable hypotheses. Of those that did, only 11% included conditions. Thus, only about 1% of the papers in leading marketing journals complied with the scientific method.

Guideline 13: *Use simple methods*

Simplicity in science, commonly called Occam's razor, has been revered by scientists, but is it is commonly violated by those who publish in scientific journals.

A search was conducted for published forecasting studies comparing the out-of-sample accuracy of forecasts from simple vs. complex methods. ("Simple vs. complex forecasting")

> Simplicity increased forecast accuracy in all 32 papers (97 comparisons).
> Simplicity decreased forecast errors by 21% for the 25 papers with quantitative comparisons.

Guideline 18: *When presenting evidence, cite only relevant scientific papers that you have read.*

Citations in scientific papers imply evidence. Give the reader an indication about the evidence in the cited work. If the paper only provided an opinion, make that clear to the reader.

Avoid citing advocacy papers *as evidence*.

Researchers commonly cite papers without reading them (Wright & Armstrong 2008). To help avoid the problem, authors should include a statement verifying that at least one author has read each of the articles cited.

Guideline 19: *Ensure summaries of prior findings that you cite are correct*

Try to contact authors of all papers you cite to ask if you have described their findings correctly.

Why? Authors often provide incorrect summaries of other researchers' findings. For example, in Wright & Armstrong (2008), 98% of a sample of 50 studies incorrectly cited the findings of Armstrong and Overton (1977).

None of the thousands of researchers who cited that paper checked with the authors if they were correctly describing the findings.

In addition, ask the authors if you have overlooked any relevant research, with particular emphasis on papers with evidence that might challenge your conclusions.

Guidelines are not sufficient; Checklists are necessary

People cannot solve complex problems using *unaided judgment*.

See the SeerSucker Theory and Expert Political Judgment

Figure 1: Guidelines for Scientists

Figure 1: Guidelines for Scientists

Selecting problems

1. ☐ Seek important problems that you could investigate objectively
2. ☐ Be skeptical about current findings, theories, policies, methods, data, and opinions
3. ☐ Consider problems where conclusive experimental evidence is lacking or ignored
4. ☐ Consider replications and extensions of useful papers that examine experimental evidence
5. ☐ Check the likely usefulness of your proposed study
6. ☐ If you need funding, ensure that you will nevertheless have control over all aspects of your study

Designing a study

7. ☐ Acquire existing knowledge about the problem (*a priori* analysis)
8. ☐ Test hypotheses or procedures by designing experiments with specified conditions
9. ☐ Compare multiple reasonable hypotheses or procedures

Collecting data

10. ☐ Obtain valid data
11. ☐ Obtain reliable data

Analyzing data

12. ☐ Use validated methods
13. ☐ Use simple methods
14. ☐ Use methods that include cumulative knowledge
15. ☐ Estimate effect sizes
16. ☐ Draw logical conclusions from the evidence on the practical implications of findings

Writing a scientific paper

17. ☐ Fully disclose research hypotheses, procedures, and data
18. ☐ Cite only relevant scientific papers when presenting evidence
19. ☐ Ensure summaries of prior findings that you cite are correct
20. ☐ Explain why your findings are important
21. ☐ Write clearly and succinctly for the widest audience for whom the findings might be useful
22. ☐ Obtain extensive peer review *before* submitting

Disseminating the findings

23. ☐ Provide thorough responses to journal reviewers, including reasons for not following suggestions
24. ☐ Challenge rejection, but only if your case is strong
25. ☐ Consider alternative ways to publish your findings
26. ☐ Inform those who can use your findings

Value of Checklists

Evidence-based checklists help even when decision-makers know the proper guidelines.

- Review of 15 experimental studies in healthcare found that checklists led to substantial improvements in outcomes. Hales and Pronovost (2006)

Checklists are especially effective when experts know little about the principles.

- Compliance with an evidence-based checklist enabled novice raters to outperform advertising experts, who lacked knowledge of the evidence-based principles in judging which ads are most effective.
  - Using a sample of 100 pairs of ads, unaided experts were correct on 55%—compared to 75% correct for the novices using the checklist. (Predictive Validity of Advertising Principles).

Evidence-based *Guidelines for Science*

We drew upon the Guidelines for Scientists to develop the "Guidelines for Science" checklist.

The 7-item checklist is designed to allow researchers or novices to evaluate compliance to scientific principles.

These are guidelines that can be rated by reading a paper.

The checklist was developed to help journal editors, funders, users, courts, managers, politicians, journalists, and other stakeholders to evaluate whether a research project is scientific.

Suggestions for scientific journals

- Certify papers for compliance to scientific principles using a checklist, such as the "Guidelines for Science" checklist
- Ratings done by a team of trained raters who would inform the authors within one week whether the paper is:
  1. acceptable for publication as a work of science,
  2. acceptable subject to the resolution of some scientific procedures,
  3. not acceptable due to violations of science guidelines (including incoherent writing).

Suggestions for universities

1. Hire professors who demonstrate that they do scientific work.

2. Certify researchers as knowing how to do scientific research by using online training and assessment, as will be provided on GuidelinesForScientists.com

3. Evaluate researchers' research contributions only on the basis of papers that discovered useful scientific knowledge, and their citations. (Other citations are useless or harmful as they spread misinformation).
4. Avoid government funding so as to allow scientists to freely pursue useful research.

Limitations to our research to date

1. We have not tested the reliability or predictive validity of the checklists.
2. We have only rough estimates of the costs and benefits of this approach to evaluation versus that of other procedures for peer review.
3. We have only rough estimates of the percentages of papers with useful scientific findings currently, or over time.

This is only the end of the beginning

While the scientific principles never change, we will continue to learn more about how to apply the principles

Barriers to implementation

1. Universities have become dependent on government funding.

2. Unelected government officials and the courts (not elected legislators) have been writing regulations, and they view failure of their regulations as arguments for more regulations.

Purpose of a conference paper is to improve it

Our research continues.

Please send me relevant experimental research that we have overlooked, *especially research that challenges the conclusions on the next page*.

The draft working paper that is the basis of this talk is available from ResearchGate as:

Armstrong, J. S. & Green, K. C. (2016). Guidelines for Science: Evidence and Checklists. [Working Paper Version].

Conclusions

1. Fewer papers in leading journals in management sciences comply with scientific principles.
2. The primary causes are:
   a) government funding, especially of advocacy research,
   b) government regulation as to what topics are permissible to study, how the study must be designed, and how it should be reported,
   c) government suppression of free speech by scientists,
   d) the use by universities of invalid criteria for evaluating research, and
   e) mandatory journal peer review.
3. We propose two evidence-based guidelines for "compliance to science", one for scientists, and one for those who fund scientists, hire them, publish their papers, and use their research findings. *The process relies first and foremost on researchers.*

Scott Armstrong's opinion

- My estimate? Fewer than one percent of papers in leading management and social science journals provide useful findings.

- Use the Checklist for Science and prove me wrong. I hope I am wrong!

Observations

1. I am *grateful* to the Wharton School for providing me with resources to work on problems that I believe to be important and to do so in the way that I consider most productive.
2. I am *confident* that the creativity of people around the world will continue to find solutions (such as the Internet and private foundations) to enable scientists the freedom to do useful scientific research.
3. I am *optimistic* that the U.S. will be able to restore free speech for scientists.
4. I am *hopeful* that the U.S. government will end the regulation of science. (See the IronLawofRegulation.com)
5. I *expect* that some of you will respond to Benjamin Franklin's call for useful scientific papers.

To be continued

A link to these slides will be posted in the next few days at GuidelinesForScience.com.

The follow-up panel session on "Disseminating & gaining acceptance for Guidelines for Science" will be held from **10:25 to 11:40 in the Hennessy Room**.

The panel session will be chaired by Arch Woodside, with a panel including Kesten Green, Roger Marshall, and myself.

Audience participation will be welcomed.