July 1996

The Ombudsman: Management Folklore and Management Science: On Portfolio Planning, Escalation Bias and Such

J. Scott Armstrong
University of Pennsylvania, armstrong@wharton.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/marketing_papers

The Ombudsman: Management Folklore and Management Science: On Portfolio Planning, Escalation Bias and Such

Abstract

Management folklore sometimes leads to unprofitable decision making. Thus, studies of the value of such folklore should be of interest to managers, especially when they identify unprofitable procedures. I reviewed empirical research on scientific publishing and concluded that studies supporting management folklore are likely to be favorably reviewed for publication and to be cited. However, researchers who obtain findings that refute folklore are likely to encounter resistance in publication and are less likely to be cited. My experience with papers on portfolio planning methods and escalation bias illustrates the problem. To encourage the publication of papers that challenge management folklore, editors should use results-blind reviews and, in some cases,constrain, reduce, or eliminate peer review.

**Management Folklore and Management Science –
On Portfolio Planning, Escalation Bias, and Such**

J. Scott Armstrong
The Wharton School
University of Pennsylvania

Abstract

Management folklore sometimes leads to unprofitable decision
making. Thus, studies of the value of such folklore should be of
interest to managers, especially when they identify unprofitable
procedures. I reviewed empirical research on scientific publishing
and concluded that studies supporting management folklore are
likely to be favorably reviewed for publication and to be cited.
However, researchers who obtain findings that refute folklore are
likely to encounter resistance in publication and are less likely to
be cited. My experience with papers on portfolio planning
methods and escalation bias illustrates the problem. To encourage
the publication of papers that challenge management folklore,
editors should use results-blind reviews and, in some cases,
constrain, reduce, or eliminate peer review.

Management contains folklore. By folklore, I mean techniques and concepts that

managers adopt without any formal evaluation of their effectiveness simply because others are

using them. Sometimes the folklore proves useful. Often however it is not useful, and sometimes

it is harmful. Folklore seems to have a long life, even when useless or harmful.

Management folklore probably arises as a way of recognizing what seems obvious. Often

this folklore is adopted by academics in their teaching, and it appears in textbooks. This

anointment by academics may contribute to the credibility and popularity of the folklore. Lee

[1980] examined popular management techniques from texts and papers; he concluded that many

are simply based on common beliefs. Miner [1984] examined 32 well- regarded organizational

theories; he concluded that only four had been shown to be valid and useful. One example of folklore that has been used in management is Maslow's hierarchy of needs [Maslow 1954]. This states that people satisfy basic needs, such as physiological and safety needs, before moving on to higher needs, such as affiliation, then achievement, then selfactualization. As this happens, the satisfied needs become less important as motivators. According to a review by Soper, Milford, and Rosenthal [1995], Maslow's hierarchy was adopted based only on the argume nt that it made sense. Conflicting evidence followed. For example, Hall and Nougaim [1968] studied the first five years of a group of managers' careers and concluded that needs became more important, not less, the more they were satisfied.

Another example of management folklore is the experience curve. Because costs decrease as cumulative production volume increases, advocates of the experience curve advise managers to increase production to gain economies of scale. Then they should reduce prices to gain more volume before their competitors can catch up, thus moving faster down the experience curve.

Research on Management Folklore

Papers that test management folklore would seem to be valuable. Consider an analogy to medical science. Folklore (for example, do not sit in a draft, get lots of rest, or eat an apple a day) is tested along with new treatments, and the testing is replicated and extended. Ideally, those who do the replications and extensions strive for objectivity. Such a process helps to determine which treatments are useful. Medical doctors rely on such testing. Without it, one would expect useless treatments to persist. (This is not to say that medical journals are free of the problems discussed here. They are also more likely to publish papers that support folklore, as summarized by Coursol and Wagner [1986].)

Replications and extensions, while useful for testing folklore, are uncommon in management science and in many of the social sciences [Hubbard and Armstrong 1994]. (Some

areas, such as personnel psychology and survey research, do use replications and extensions.) Surveys of reviewers suggest that they are biased against accepting replications [Kerr, Tolliver, and Petree 1977; Neuliep and Crandall 1990]. Typically, the argument against replications has been that scarce space in journals is needed for reports of new research. The analogous argument would sound ludicrous in the field of medicine: "We have so many proposals for new medicines that we cannot devote resources to careful evaluations of the ones we already use."

Sometimes testing shows folklore to be correct. Many researchers have made successful careers by doing research that supports folklore. For example, the 1995 Nobel prize in economics recognizes the claim that people have rational expectations about the future. Similarly, many findings in psychology appear to be based on common sense. Mischel [1981] asked fourth and sixth-grade students to predict the outcomes of 17 classic experiments in psychology; they correctly predicted the outcomes for 12 of them.

Papers that support management folklore seem to be favorably reviewed, they sometimes get a lot of attention, and the folklore becomes even more entrenched. This occurs because reviewers tend to be biased in favor of publishing a paper if they agree with the results. This was shown by Goodstein and Brazis (1970), who asked 282 psychologists to review one of two abstracts that were identical except for the results. They rated those in which the results were in accord with their own beliefs as better designed and said that they were more suitable for publication. Abramowitz, Gomes, and Abramowitz [1975] did a similar study and reached the same conclusion.

Early papers on the "Hawthorne effect" fit this description; their authors interpreted the data as evidence that workers respond positively to any attention from management, a conclusion that fit well with what managers and researchers believed at the time.

If our aim is to improve management practice, papers that reveal folklore to be ineffective should be of special interest because they might lead to improved practices. I believe that such papers are published on occasion, and I could cite examples describing research in marketing, strategic planning, survey techniques, personnel selection, and forecasting that has demonstrated improvements over existing management practices. One might hope that journals would publish many studies with controversial findings. But it appears that they do so only rarely. Editors of 16 psychology journals reported that reviewers dealt harshly with papers that contained controversial findings [Armstrong and Hubbard 1991].

The aversion to disconfirming evidence seems to be widespread, as shown by a stream of research dating back to Festinger, Riecken, and Schacter's [1956] paper about a cult that predicted the end of the world; it did not end on the date that they predicted, and thus led the cult members to have more confidence in their beliefs. In a study by Batson [1975], subjects who believed that Christ was God were given what they believed to be authentic evidence that he was not. As a result, these subjects increased their belief that Christ was God.

The effects of bias against controversial findings by social scientists seem so great that they are obvious to many observers. For example, in Gans and Shepherd's [1994] survey of eminent economists, many reported that they had difficulties in getting papers accepted for publication when their findings departed from accepted beliefs. More important than the anecdotal evidence, though, empirical studies support this view. In an experiment conducted by Mahoney [1977], 75 psychologists thought that they were providing reviews of an actual submission; they were more likely to accept papers when the findings agreed with the reviewers' existing beliefs. When they didn't, the reviewers were much more likely to reject the paper, explaining that the methodology was flawed. The methodology was in fact the same for both versions of Mahoney's fictitious submission.

Publication, while important, will not completely solve the problem. Researchers and practitioners may ignore the findings. Consider the Hawthorne effect. Franke and Kaul [1978] reanalyzed the data and concluded that there was little evidence for a Hawthorne effect. Nevertheless, this belief in a Hawthorne effect persists not only among managers, but also among many academics. Such situations have also been studied in the social sciences, and it turns out that disconfirming evidence is often ignored. The Little Albert study by J. B. Watson [Samelson 1980] provides another example. Watson's study, based on a sample of one baby, supported existing folklore about the conditioning of behavior. Partly because of this study, Watson was one of the most highly regarded psychologists of the early 1900s. Other researchers tried to replicate the study with little success. Some eventually concluded that Watson's sample size was zero rather than one. Meanwhile, descriptions of the Little Albert study kept appearing in textbooks without reference to the failed replications.

Given the potential for bias in the citation of papers [Wanous, Sullivan, and Malinak 1989], researchers should take care to ensure that they include positive and negative findings in a review of the literature. Typically, the recommendation is to use a systematic search procedure and to avoid excluding papers based on judgments of methodology. For example, Greenley [1986], based on a review of eight empirical studies, concluded that formal strategic planning was not useful for manufacturing companies. This belief is consistent with management folklore. I attempted to conduct a comprehensive review of this topic [Armstrong 1991]. I found 28 studies; of these, 20 found better performance with formal planning, five found no difference, and three found planning to be detrimental.

Given these findings, researchers concerned about their careers might conclude that it is unwise to submit papers that conflict with existing folklore. This might help to explain why

editors of psychology journals report that they rarely receive papers with controversial findings [Armstrong and Hubbard 1991].

To illustrate the problems an author can encounter in trying to publish papers that refute folklore, I describe my experiences with two papers; one concerns portfolio planning methods, and the other deals with escalation bias. I am assuming that my experiences are representative of what one might expect. They are representative of the treatment accorded to much of my other work, and people who have published research with controversial finding have told me that their experiences have been similar to mine. I use my own cases because I know them much better than those of others.

Note the direction of the argument. I am not generalizing from my experience. Prior findings have established that folklore in many areas is persistent and difficult to attack. My experiences are consistent with those findings. There is the alternative explanation that my research is worthless. The reader will have to decide whether this is so by reading the original papers. In my own opinion, the two studies are among my best.

Portfolio Planning Methods

Portfolio planning methods typically base product planning on the market share of the firm's product and the growth rate of the product category. For example, the Boston Consulting Group's (BCG) Product Portfolio Matrix implies that managers should eliminate "dogs" (low market share products in markets that are not growing) and invest in "stars" (products with a high market share in high growth areas). In other words, managers should disinvest in products that are stagnant and invest in products positioned to grow.

Descriptive and theoretical papers on portfolio planning matrix methods were originally published over two decades ago. These papers tend to reinforce folklore that says, when deciding

which products to emphasize, managers should stick with their winners. Management and marketing textbooks describe the BCG matrix, generally in an approving manner.

Despite the widespread use of portfolio planning matrices, we were able to find only two published studies describing empirical tests of their value [Capon, Farley, and Hulbert 1987, pp. 316-317; and Slater and Zwirlein 1992]. These used cross-sectional data on firms; they concluded that those firms that used the portfolio matrix methods were less profitable than the firms in their sample that did not use them.

Armstrong and Brodie [1994], which will be called A&B, was the first published experimental test of portfolio planning methods. A&B asked subjects to adopt the role of a marketing vice-president and to make a decision as to which of two investment projects should be selected. The subjects had enough information to make simple calculations of the return on investment. The difference between the two projects was enormous. One project doubled its investment over the 10- year horizon, and the other lost half of its investment. Our purpose was to determine whether those who were informed about the BCG matrix would have trouble selecting the most profitable project—because it was a "dog" that yielded the gain and a "stat" that lost money.

Of those exposed to information about the BCG matrix, 64 percent selected the unprofitable investment. Of those who used the BCG matrix as a decision aid, 87 percent selected the unprofitable decision. Thus, the BCG matrix misled decision makers.

Reviewers' Criticisms of the A&B Paper

A&B was originally submitted to a journal in mid-1989. The reviewers often differed with one another. When two reviewers stated, "I cannot imagine obtaining the results reported here," and the "conclusion . . . seems to lack a certain face validity," an editor summarized these reviews by stating that the results were not controversial because "the BCG portfolio matrix has

been widely criticized in the popular and academic literature over the past 10 years." One reviewer said, "The author shows good knowledge of the relevant literature." Another said, "The author's review of the literature is scant" (but did not cite any missing research studies). As to whether firms use the BCG, one reviewer said, "the author makes no attempt to explain why so many firms appear to be using an approach which has so little to recommend it." Another said, "I do not think that many managers rely on the BCG approach." One reviewer said, "I cannot help but think that the undergrads cannot rise to the occasion and that the task might have been too demanding for them," while another reviewer said that the task was "too simple."

The reviewers seldom showed evidence, nor did they otherwise support their criticisms. For example, one said, "The task was unfair to the BCG matrix," but did not suggest what type of test would have been fair. A reviewer said that the BCG was "the weakest of the portfolio matrices," but did not say how he or she came to this conclusion.

Over its three-year review period, the paper went through seven rounds of reviews with 14 referees at four journals. There were many supporters, but some referees recommended rejection each time. Despite continued improvements in the study, the reviews did not become more favorable. In the end, an editor agreed with our rebuttal and accepted the paper despite negative reviews. The paper was published along with a comment by a reviewer [Wensley 1994].

Recognition of Challenges to Portfolio Matrices

The only published empirical evaluations of the portfolio matrices are those by Armstrong and Brodie [1994], Capon, Farley, and Hulbert [1987], and Slater and Zwirlein [1992]. All found evidence that use of the BCG matrix was harmful. Other than our citations of the other two studies, there were no citations of this evidence about the BCG matrix according to the *Social Science Citation Index.* Overall, this yields an average of less than 0.2 citations per year for the 11 "paper years." As there were no empirical studies that were favorable, it was not

possible to make a comparison. However, one could say that these papers with negative results have been virtually ignored by academics.

Escalation Bias

The original study on escalation bias [Staw 1976] showed that managers tend to reinvest in projects that have gone poorly. This supports management folklore that managers throw good money after bad. Managers believe that other managers suffer from escalatio n bias.

Staw asked subjects (acting as managers) to invest in one of two R&D projects. Half of the subjects were then told that their investment had done well over a subsequent period, and the other half that their investment had done poorly. They were given a chance to invest more, but this time they could split their investment between the same two projects. Subjects who learned that their selected investment was doing poorly tended to invest more money in the same project than did those who were told that their selected project had done well.

Armstrong, Coviello, and Safranek [1993], referred to here as ACS, studied escalation bias using the materials that Staw had given to his subjects. (Stew provided us with the materials.) However, ACS changed the context from an R&D investment to either an advertising investment or a product design investment. ACS also changed the dates to make the case more contemporary. The results did not support the original findings; escalation bias did not generalize to these marketing decisions.

Were the ACS findings controversial? One way to view the study is that it simply helped to define the limits to which escalation bias can be generalized. This does not seem overly controversial. However, another viewpoint is that it showed that, in general, the research findings about escalation bias are not generalizable. Furthermore, ACS, like others before, concluded that the problem in Staw's original study design did not have a correct decision. That is, the subjects did not receive information that would allow them to make a rational decision. To do so they

9

would need information about the future profitability for each alternative. Because they did not have such information, one could not say that escalation bias harmed decision making in these studies. Given these considerations, ACS concluded that 15 years of research that was spawned by Staw's original study produced little of value to managers.

Reviewers' Criticisms of the ACS Paper

Although the reviews were blind, the referees apparently came largely from among those cited in ACS who had done successful replications. Initially, ACS was rejected because this replication itself had not been replicated. The reviewers suggested that the results might have been due to differences in administrative procedures. To address this, ACS did a replication with changes in the administrative details. The results were virtually identical to the first study. However, the reviewers rejected the resubmission because it examined only one situation (advertising), and there might be something unusual about that situation. While Staw's original research was also based on only one situation, this objection to our study had merit. As a result, we conducted an extension involving a product design decision. Again, the extension failed to show an escalation bias.

The initial reviews of ACS seemed more encouraging than later reviews. In the earlier ones, the reviewers discussed potential defects and suggested further research. ACS devoted a lot of time to gathering and analyzing new data for the revisions. However, after ACS incorporated the new studies into the paper, the reviewers became more negative and their rejections more terse. They made few substantive recommendations for change.

One criticism was that the original study by Staw was defective and should not have been published, so this extension should not be published. Instead, a different study should have been done, one that would correct the major defect of the earlier studies, in particular, the lack of adequate information. The reviewer seemed to be unaware of Bateman [1986] who corrected

some of these defects. (Schaubroeck and Davis [1994], in a later paper, also corrected a design defect.) But even if one argues that there were defects in the original study, it had a major impact in the mass media, and it had strong implications for managers. Other researchers also paid attention: Staw's original paper had been cited over 150 times by mid-1995, and numerous replications had been conducted. Thus, it seemed important to assess the generalizability of these findings. ACS did this by conducting experiments using the identical procedure in a new context.

A second criticism was that ACS did not examine all aspects of the original study. In particular, it only studied the case in which the decision maker was highly committed to the earlier decision. However, ACS clearly stated that the study was limited to the situation in which escalation bias was expected to be strongest (that in which the subject was responsible for the first decision). If no effect were found here, examining the situation in which the effect was expected to be weak (in which someone else was responsible for the first decision) would be irrelevant.

A third criticism was that ACS did not explain why the extensions did not produce the same results, thus it does not improve understanding of the phenomena. (ACS tested possible explanations, but found no support for them.) According to this argument, extensions should be published only if the reasons for different results can be identified. Certainly it would be desirable to be able to explain why replication results differ. But what if one examines the leading explanations and finds none satisfactory? This might happen, for example, if the original results were spurious. Having to explain the reasons for differing results would impose an undue burden on those conducting replications. Such a requirement would lead to a bias in the literature, favoring publication of successful replications over unsuccessful ones.

A fourth criticism was that no additional replications of escalation bias are needed. Interestingly, escalation bias is unusual in management science in that many replications have

been published. They typically supported the original findings. Thus, at this stage, a failure to replicate would seem to provide more information than a successful replication.

I believe that ACS addressed an important issue, that the methodology was proper for an extension, and that the results were interesting because they conflicted with expectations and with prior research. We discussed these points in the submission letters. Also, we asked all the editors whether their journals published replications, and whether it would be possible to have reviewers selected from those outside the mainstream research. We also offered to share prior reviews with each editor. In only one case, did we receive a letter that addressed our concerns. The other responses were form letters.

ACS was originally sent for review in October 1987. It was eventually reviewed by eight journals, two of which had two independent rounds of reviewing. Thus, it went through a total of 10 formal reviews. The editor of one journal asked for new experiments of a well-specified nature but then rejected the paper, after these were successfully completed, because the results did not support a possible explanation for the failure to replicate. A total of 28 journal reviewers were used over a five-year period. Including the peer reviews that we ourselves obtained, 37 people reviewed the paper.

ACS was eventually accepted for publication in the J*ournal of the Academy of Marketing Science* in 1993. This journal is interested in controversial work as a matter of policy [Peterson 1992]. The editor, Robert Peterson, responded to the points raised in our submission letter. He also suggested that we make revisions before he sent the paper out to reviewers to increase the likelihood of its acceptance. The reviewers disagreed with our conclusions but proposed no substantive changes. We addressed the concerns in our response and the editor supported our position.

ACS may also have suffered in the reviewing because it did not report statistically significant results. A substantial literature indicates such a bias (for a summary of this evidence, see Hubbard and Armstrong [1992]). By the time the paper was finally accepted, it had grown from one experimental comparison to seven. While the reviews led to many improvements in the paper, especially in the early stages, the additional experiments and analyses consistently supported the original conclusion.

Recognition of Challenges to Escalation Bias

In addition to ACS, we are aware of four prior studies that failed to replicate findings about escalation bias: Barton, Duchon, and Dunegan [1989], Bateman [1986], Schwenk [1988], and Singer and Singer [1985 or 1986]. Researchers have largely ignored these papers. We came to this conclusion after examining the *Social Science Citation Index* through August 1995. ACS had not yet been cited. The Singer and Singer papers were cited nine times, Barton, Duchon, and Dunegan was cited six times, Bateman was cited three times, and Schwenk was cited twice – overall, 20 citations. This works out to 0.46 citations per paper per year over 43 "paper years." There were 10 successful replications, which we cited in ACS [Bazerman, Beekun, and Schoorman 1982; Brockner 1992; Conlon and Parks 1987; Fox and Staw 1979; Garland 1990; Garland and Newport 1991; Garland, Sandefur, and Rogers 1990; McCain 1986; Staw 1981; and Staw and Fox 1977]. They were cited 4.3 times per year over this sample of 96 "paper years." Thus, the successful replications were cited nine times as frequently as those with disconfirming evidence.

This difference in citation rates is statistically significant at less than 0.001 [Wilcoxon-Mann-Whitney test from Siegel and Castellan 1988]. The most frequently cited of the four disconfirming replications was Barton, Duchon, and Dunegan [1989] with an average of 1.0 citations per year, which was a lower rate than all of the 10 confirming replications.

13

Discussion

An argument can be made that papers with findings that refute management folklore should be subject to extra scrutiny because they may have a great impact. Additional peer review should help to reduce defects and improve the writing. On the other hand, one could also argue that papers that refute folklore should get preference. Consider the analogy to medical science. Should not results that challenge the use of an existing procedure be published rapidly to prevent potential harm and to alert other researchers? The BCG matrix, for example, was adopted with no basis for support and has been used for over two decades with seemingly detrimental results. One virtue of academic publishing is that researchers can submit their work to many journals. Despite biases in reviewing, papers with controversial findings can eventually find their way into the literature. However, this may cost the authors much effort and time. My own papers with controversial findings have taken three to seven years from original submission to final publication. Thus, attempting to publish papers with potentially controversial findings may be risky for professors without tenure.

The journal that eventually publishes the paper may not be optimal for reaching relevant researchers. For example, the papers describing failures to replicate escalation were not published in the journals that published the original paper and the successful replications, except for Schwenk's paper, where replication was an incidental issue.

Some management journals welcome controversial papers. These include the *International Journal of Forecasting*, *International Journal of Research in Marketing, Journal of the Academy of Marketing Science, Journal of Business Research, Journal of Management*, and *Marketing Letters*. Undoubtedly there are others. However, despite their good intentions, few editors back them up with formal procedures for reviewing papers with controversial findings. Without such procedures, the chances that they will publish controversial papers are slim. It is

14

difficult after the fact to identify bias against papers with controversial findings, and they are likely to be rejected for "poor methodology."

The *International Journal of Forecasting* has formal procedures (1) to seek papers with controversial findings, and (2) to review them in ways that avoid biases. The *Journal of Management* calls for "interesting papers." A new journal, *Iconoclastic Papers*, has been founded with the intent of publishing papers that challenge management practices and beliefs. New procedures should be of special interest to editors of leading journals. Because they accept only a small portion of submitted papers, they may adopt bureaucratic rules, such as "accept all papers favorably reviewed by all three reviewers," and "reject any paper that receives mostly bad reviews." Kupfersmid and Wonderly [1994] summarized three studies on the relationship of reviewers' recommendations to the final decision. They concluded that the relationship was strong for the *American Sociological Review, Journal of Abnormal Psychology*, and *Thorax*, a medical journal. For *Thorax*, if the review was split, the paper was published only 10 percent of the time. Bureaucratic rules, while fair, would increase the likelihood of rejecting papers with controversial findings about folklore Various procedures might improve the reviewing process for papers with controversial findings and especially those involving replications:

1. Editors could accept proposed papers on controversial topics based on the study design. That is, the referees could review the study design prior to the study being done. As long as the researchers followed the plan, their paper would be published. Such proposals have also been made for journals in medical science [Newcombe 1987] and in the social sciences [Walster and Cleary 1970]. The *International Journal of Forecasting* plans to adopt such a procedure on an experimental basis. One benefit of this procedure is that it ensures that the hypotheses are developed prior to the analyses of the data.

2. Editors can invite authors to provide lists of potential reviewers, from which they will choose at least one. The editors of *Organization Science* follow this procedure. This should help to ensure that the reviewers include researchers who are not biased against the findings.

3. Authors can request "results-blind" reviews. Referees would review the paper without knowing what results were obtained. This procedure guards against reviewers' rejecting a paper because they disagree with the findings [Kupersmid and Wonderly 1994, pp. 99-105]. The editors of the *Journal of Social Behavior and Personality* ask that author start each new section on a separate page so that this process can be followed when needed.

4. Editors should not ask reviewers to say whether the paper should be published or not. Given that papers with controversial findings receive mixed reviews at best [Armstrong and Hubbard 1991] and that papers with mixed reviews have a low probability of being published (Kupfersmid and Wonderly 1994], current policies are unlikely to lead to the publication of papers with controversial findings. Reviewers should focus on the methodology and how it might be improved. Editors should decide which papers should be published. This will improve the likelihood that papers with controversial findings make their way successfully through the reviewing process. Of course, this assumes that editors are more interested than the reviewers in publishing papers about studies that have controversial findings.

An alternative approach would be for some journals to dispense with the reviewing process, at least for a section of the journal. Instead, editors would solicit papers from well-established authors. There are some arguments as to why this should work. First, a good

predictor of whether a paper will be useful seems to be whether previous papers by this author have been useful [Abrams 1991]. In addition, well-established authors are not likely to want to harm their reputations, so they will seek peer review on their own. Finally, editors can employ such safeguards as open peer review or letters to the editors. The *Journal of Economic Perspectives* has used this strategy with apparent success.

The problem I describe may eventually be solved by electronic publishing. Because the costs of dissemination will be low, the need to reject papers will no longer exist. Scientific results can then be reported more rapidly to a wider audience. Presumably, the academic reward system would no longer focus on number of publications and instead be based on measures of readership, citations, open peer review, published comments, and evidence on successful implementation of the findings. In addition, as there would be no reward for the mere act of putting a paper on the electronic network, this procedure might lead to a reduction in the mountain of senseless papers that is currently published.

Researchers should recognize that replications and extensions of studies on folklore might lead to negative findings. In general, replications and extensions tend to conflict with the original findings in about half of the published studies [Hubbard and Vetter 1996].

Authors can take steps to increase the likelihood for acceptance of papers that contradict folklore. One possibility is to frame the paper as an attempt to examine the conditions under which a previous generalization holds, rather than as disconfirming existing folklore. Another step is to persist in submitting such papers.

The failure to cite disconfirming evidence should also be addressed. Inasmuch as those who do research on a topic are the ones who are likely to review challenges to the folklore, one would expect that they would cite this research in their future work. This does not appear to be happening. Perhaps those who submit favorable evidence about folklore should be asked to

conduct searches for contradictory evidence in order that all relevant evidence may be considered. Referees might be asked to explicitly consider whether contradictory evidence has been overlooked by a submitted paper. Systematic computer-aided literature search procedures may help researchers to find relevant papers.

A Proposal

"The Ombudsman" column of *Interfaces* was founded with the intention of provided space for papers with controversial findings about management science [Armstrong 1982]. I renew this offer. it does not extend to papers that merely say controversial things; they must contain empirical findings. While controversial replications are of particular interest, we will also consider replications that support previous findings. To submit such a paper, send it with a cover letter saying that it contains controversial findings and that you would like the paper to be handled by "The Ombudsman." You are invited to submit a list of possible reviewers. If you provide the names and addresses of four qualified researchers or more, we will ask at least one to provide a review. We will follow a results-blind procedure. Thus, you will need to submit a version of your paper in which each section starts on a new page and ensure that the results are not discussed in the design section. We will not ask reviewers whether the paper should be published. An editor will make the decision.

We are also interested in papers for which researchers can show that they obtained prior peer review or that relevant journals have resisted publishing their work. We invite authors to submit copies of previous reviews, suitably disguised so as not to criticize individuals.

Perhaps other journals will adopt procedures that encourage publication of papers that challenge folklore. Such papers are vital to the growth of scientific knowledge and are useful for improving management practice.

The road to publishing controversial replications is long and difficult. It might be useful to know what to expect so that you do not give up in your attempts to publish. Knowing that you have a publication outlet at *Interfaces* may encourage you to study important issues. Of course, in encouraging people to persist in their publication efforts, we would not want to create an escalation bias.

Conclusions

Management folklore offers seemingly sensible solutions. When researchers study folklore, the reviewing process is likely to favor papers that support it. This bias can contribute to managers' belief in the folklore.

Prior research suggested that papers that refute management folklore will meet resistance in the reviewing process. I illustrated this with two studies, one on portfolio planning matrices and the other on escalation bias. They were subjected to an average of eight rounds of reviews … by six journals … using 21 referees … over four years.

Consistent with expectations, the papers that disconfirmed folklore have received little attention. Evidence from 43 "paper-years" indicates that these papers have been largely ignored. Reviewers should insist that unfavorable evidence should also be included in reviews of prior research.

I suggest that provisions be made to allow for reviews based on the design of a study rather than on its results. This could be done on an experimental basis. In addition, it might be useful to allocate sections in prestigious journals for papers that are not subject to the traditional peer review process. Such procedures might help to weed out false folklore. Meanwhile, I have little expectation that these examples of folklore will die easily.

References

Abramowitz, S. I., Gomes, B., and Abramowitz, C. V. (1975), "Publish or politic: Referee bias in manuscript review," *Journal of Applied Social Psychology*, Vol. 5, No. 3, pp. 187-200.

Abrams, Peter A. (1991), "The predictive ability of peer review of grant proposals: The case of ecology and the US National Science Foundation," *Social Studies of Science*, Vol. 21, No. 1, pp. 111-132.

Armstrong, J. S. (1982), "Is review by peers as fair as it appears?" *Interfaces* Vol. 12, No. 5 (September-October), pp. 62- 74.

Armstrong, J. S. (1991), "Strategic planning improves manufacturing performance," *Long Range Planning*, Vol. 24, No. 4, pp. 127-129.

Armstrong, J. S. and Brodie, R. (1994), "Effects of portfolio planning methods on decision making: Experimental results," *International Journal of Research in Marketing*, Vol. 11, No. 1, pp. 73- 84.

Armstrong, J. S., Coviello, N., and Safranek, B. (1993), "Escalation bias: Does it extend to marketing decisions?" *Journal of the Academy of Marketing Science*, Vol. 21, No. 3, pp. 247-253.

Armstrong, J. S. and Hubbard, R. (1991), "Does the need for agreement among reviewers inhibit the publication of controversial findings?" *Behavioral and Brain Sciences*, Vol. 14 No. 11 (March), pp. 136-137.

Barton, S. L., Duchon, D., and Dunegan, K. J. (1989), "An empirical test of Staw and Ross's prescriptions for the manageme nt of escalation of commitment behavior in organizations," *Decision Sciences*, Vol. 20, No. 3, pp. 532-544.

Bateman, T. S. (1986), "The escalation of commitment in sequential decision making: Situational and personal moderators and limiting conditions," *Decision Sciences*, Vol. 17, No. 1, pp. 33-49.

Batson, C. D. (1975), "Rational processing or rationalization? The effect of disconfirming information on a stated religious belief," *Journal of Personality and Social Psychology*, Vol. 32, No. 1, pp. 176-184.

Bazerman, M. H., Beekun, R. I., and Schoorman F. D. (1982), "Performance evaluation in a dynamic context: A laboratory study of the impact of a prior commitment to the ratee," *Journal of Applied Psychology*, Vol. 67, No. 6, pp. 873-876.

Brockner, J. (1992), "The escalation of commitment to a failing course of action: Toward theoretical progress," *Academy of Management Review*, Vol. 17, No. 1, pp. 39-61.

Capon, N., Farley, J. U., and Hulbert, J. M. (1987), *Corporate Strategic Planning*. Columbia University Press, New York.

Conlon, E. J. and Parks, J. M. (1987), "Information requests in the context of escalation," *Journal of Applied Psychology*, Vol. 72, No. 3, pp. 344-350.

Coursol, A. and Wagner, E. E. 1986, "Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias," *Professional Psychology: Research and Practice*, Vol. 17, No. 2, pp. 136-137.

Festinger, L., Riecken, H. W., Jr., and Schacter, S. (1956), *When Prophecy Fails*. University of Minnesota Press, Minneapolis, Minnesota.

Fox, F. V. and Staw, B. M. (1979), "The trapped administrator: Effects of job insecurity and policy resistance upon commitment to a course of action," *Administrative Science Quarterly*, Vol. 24, No. 3, pp. 449-471.

Franke, R. H. and Kaul, J. D. (1978), "The Hawthorne experiments: First statistical interpretation," *American Sociological Review*, Vol. 43, No. 5, pp. 623-643.

Gans, Joshua S. and Shepherd, G. B. (1994), "How are the mighty fallen: Rejected classic articles by leading economists," *Journal of Economic Perspectives*, Vol. 8, No. 1, pp. 165-179.

Garland, H. (1990), "Throwing good money after bad: The effect of sunk costs on the decision to escalate commitment to an ongoing project," *Journal of Applied Psychology*, Vol. 75, No. 6, pp. 728-731.

Garland, H. and Newport, S. S. (1991), "Effects of absolute and relative sunk costs on the decision to persist with a course of action," *Organizational Behavior and Human Decision Processes*, Vol. 48, No. 1, pp. 55-69.

Garland, H., Sandefur, C. A., and Rogers, A. C. (1990), "De-escalation of commitment in oil exploration: When sunk costs and negative feedback coincide," *Journal of Applied Psychology*, Vol. 75, No. 6, pp. 721-727.

Goodstein, L. D. and Brazis, K. L. (1970), "Credibility of psychologists: An empirical study," *Psychological Reports*, Vol. 27, No. 3, pp. 835-838.

Greenley, G. E. (1986), "Does strategic planning improve company performance?" *Long Range Planning*, Vol. 19, No. 2, pp. 101-109.

Hall, D. T. and Nougaim, K. E. (1968), "An examination of Maslow's need hierarchy in an organizational setting," *Organizational Behavior and Human Performance*, Vol. 3, No. 1, pp. 1235.

Hubbard, R. and Armstrong, J. S. (1994), "Replications and extensions in marketing: Rarely published but quite contrary," *International Journal of Research in Marketing*, Vol. 11, No. 3, pp. 233-248.

Hubbard, R. and Armstrong, J. S. (1992), "Are null results becoming an endangered species in marketing?" *Marketing Letters*, Vol. 3, No. 2, pp. 127-136.

Hubbard, R. and Vetter, D. E. (1996), "An empirical comparison of published replication research in accounting, economics, finance, management and marketing," *Journal of Business Research*, Vol. 35, No. 2, pp. 153-164.

Kerr, S., Tolliver, J., and Petree, D. (1977), "Manuscript characteristics which influence acceptance for management and social science journals," *Academy of Management Journal*, Vol. 20, No. 1, pp. 132-141.

Kupfersmid, J. and Wonderly, D. M. (1994), *An Author's Guide to Publishing Better Articles in Better Journals in the Behavioral Sciences*. Clinical Psychology Publishing Co., Brandon, Vermont.

Lee, J. A. (1980), *The Gold and the Garbage in Management Theories and Prescriptions.* Ohio University Press, Athens, Ohio.

Mahoney, M. (1977), "Publication prejudices: An experimental study of confirmatory bias in the peer review system," *Cognitive Therapy and Research*, Vol. 1, No. 2, pp. 161-175.

Maslow, A. H. (1954), *Motivation and Personality.* Harper and Row, New York.

McCain, B. E. (1986), "Continuing investment under conditions of failure: A laboratory study of the limits to escalation," *Journal of Applied Psychology*, Vol. 71, No. 2, pp. 280-284.

Miner, J. B. (1984), "The validity and usefulness of theories in an emerging organizational science," *Academy of Management Review*, Vol. 9, No. 2, pp. 296-306.

Mischel, W. (1981), "Metacognition and rules of delay," in *Social Cognition Development*, eds. John H. Flavell and L. Ross, Cambridge University Press, Cambridge, England.

Neunliep, J. W. and Crandall, R. (1990), "Editorial bias against replication research," *Journal of Social Behavior and Personality*, Vol. 5, No. 4, pp. 85-90.

Newcombe, R. G. (1987), "Towards a reduction in publication bias," *British Medical Journal*, Vol. 295 (September 12), pp. 656-659.

Peterson, R. (1992), "Introduction to the special issue," *Journal of the Academy of Marketing Science*, Vol. 20, No. 4, pp. 295-297.

Samelson, F. (1980), "J. B. Watson's Little Albert, Cyril Burt's twins, and the need for a critical science," *American Psychologist*, Vol. 35, No. 7 (July), pp. 619-625.

Schaubroeck, J. and Davis, E. (1994), "Prospect theory predictions when escalation is not the only chance to recover sunk costs," *Organizational Behavior and Human Decision Processes*, Vol. 57, No. 1, pp. 59-82.

Schwenk, C. R. (1988), "Effects of devil's advocacy on escalating commitment," *Human Relations*, Vol. 41, No. 10, pp. 769-782.

Siegel, S. and Castellan, N. J. (1988), *Nonparametric Statistics for the Behavioral Sciences.* McGraw Hill, New York.

Singer, M. S. and Singer, A. E. (1985), "Is there always escalation of commitment?" *Psychological Reports*, Vol. 56, No. 3 (June), pp. 816-818.

Singer, M. S. and Singer, A. E. (1986), "Individual differences and the escalation of commitment paradigm," *Journal of Social Psychology*, Vol. 126, No. 2 (April), pp. 197-204.

Slater, S. F. and Zwirlein, T. J. (1992), "Shareholder value and investment strategy using the general portfolio model," *Journal of Management*, Vol. 18, No. 4, pp. 717-732.

Soper, B., Milford, G. E., and Rosenthal, G. T. (1995), "Belief when evidence does not support theory," *Psychology and Marketing*, Vol. 12, No. 5, pp. 415-422.

Staw, B. M. (1976), "Knee-deep in the big muddy: A study of escalation commitment to a chosen course of action," *Organizational Behavior and Human Performance*, Vol. 16, No. 1 (June), pp. 2744.

Staw, B. M. (1981), "The escalation of commitment to a course of action," *Academy of Management Review*, Vol. 6, No. 4 (October), pp. 577-587.

Staw, B. M. and Fox, F. V. (1977), "Escalation: The determinants of commitment to a chosen course of action," *Human Relations*, Vol. 30 No. 5 (May), pp. 431-450.

Walster, G. W. and Cleary, T. A. (1970), "A proposal for a new editorial policy in the social sciences," *American Statistician*, Vol. 24, No. 2, pp. 5-10.

Wanous, J. P., Sullivan, S. E., and Malinak, J. (1989), "The role of judgment calls in metaanalysis," *Journal of Applied Psychology*, Vol. 74, No. 2, pp. 259-264.

Wensley, R. (1994), "Making better decisions: The challenge of marketing strategy techniques," *International Journal of Research in Marketing*, Vol. 11, No. 1, pp. 85-90.