3-2022

Note to self: How I can be a better reviewer?

Reddi KOTHA
Singapore Management University, reddikotha@smu.edu.sg

Note to Self: How I Can Be a Better Reviewer

Reddi Kotha
Lee Kong Chian School of Business
Singapore Management University
Email: reddikotha@smu.edu.sg Draft
Date: 14 March 2022

The progress of science depends on reviewers as much as it does upon researchers developing new science. Reviewers provide impartial, anonymous, and expert advice to researchers, and they screen which research gets published in scientific journals. When the review process works well, it acts as an effective filter and enables dissemination of rigorous scientific work. Although much of scientific progress depends upon the review process, reviewers often learn this practice simply by doing. Of course, there are articles that suggest some general norms of behavior, such as being punctual, polite to authors, and respecting rules of confidentiality (Romanelli, 1996). However, if the field of management is to advance, a more comprehensive approach to reviewing and additional training may be warranted. While such an endeavor is indeed worthwhile, in this article I strive to offer an immediate stopgap by reviewing some of the common errors I have made as a reviewer and suggesting ways to avoid them. These errors point to reasons why the same manuscript could receive widely heterogeneous feedback from two or more reviewers. Avoiding these review errors will lead to greater convergence between reviewers on the "true" deservingness of the science as reported in academic papers.

Before I review the errors I have made, let me first suggest a mental model that allows for scientific papers to be ascertained on "true" deservingness. Since we are situated within the domain of the social sciences there may be multiple explanations for the same phenomenon. When and why one explanation should be preferred over others is guided by whether it: i) explains more of a "critical experiment" that contrasts between different approaches and ii) makes fewer assumptions (Lave & March, 1993). To clarify for the remainder of this paper, I define a "critical experiment" as a setting that allows for two or more otherwise competing explanations that make the same predictions, but rely on different processes to make them.

Critical experiment: For an abstract example of a critical experiment, consider a paper from an author that uses theory A to make two predictions on the relationship between input X and outcomes 1 and 2 (see Table 1, column 4). Assume further that this author has tested and found support for these predictions. Reviewer #1 uses a different theory, theory B, but makes the same predictions on the relationship between input X and the outcomes (column 5). Since the causal processes in theory A and B are not the same, how can we determine which theory should be preferred? A critical experiment allows us to break through this deadlock. We could also find another outcome, for example outcome 3, in the same setting where theories A and B make opposite predictions on the relationship between input X and outcome

3. After the critical experiment, only one theory will explain all three outcomes, and that theory should be preferred. For instance, the author's theory A could make a positive prediction on the relationship between input X and outcome 3 (column 4), whereas reviewer #1's theory B could make a negative prediction or no prediction at all (column 5). The converse also allows us to set up a critical experiment. When tested, if the relationship between input X and outcome 3 is positive, then theory A as offered by the author should be preferred (over theory B, as offered by reviewer #1). There are other aspects of Table 1 that will be referenced further ahead, but for now, the key point is that outcome 3's setting enables us to conduct a critical experiment to determine whether the author's or reviewer #1's explanation is preferred.

---INSERT TABLE 1 ABOUT HERE---

With this context of a critical experiment established, let me highlight some common errors that I have made as a rookie reviewer and explain how I have learned to guard against them.

Seven rookie errors I have made as a reviewer and how to correct them:

1) Heroic assumption error: Authors are allowed to make any assumption they desire. In the end, if they explain outcomes using what reviewers see as incredulous assumptions, their explanation merits publication. The only point the reviewers should focus on is this: do they propose an alternative model that can make the same predictions, albeit with fewer assumptions? For instance, in Table 1, let us assume that reviewer #1's explanation, theory B (column 4), makes fewer assumptions than the author' explanation, theory A (column 5), for outcomes 1 and 2. For the purpose of explaining outcomes 1 and 2, reviewer #1's explanations should be preferred. However, there is no reason to select a theory that makes a simpler assumption if the theory at hand cannot be used to make predictions that explain all the facts of interest. For instance, if reviewer #2's theory C makes the least number of assumptions, but does not explain any additional outcome in Table 1, it is an inferior explanation compared to the author's explanation and to reviewer #1's (column 5).

   Developmental review that corrects the heroic assumption error: In the paragraph above, I have taken a strong line on assumptions stating that it is only outcomes that matter, rather than assumptions. Having made this point, let me make a concession. If past work

has repeatedly found evidence for processes and outcomes that are consistent with a different set of assumptions than the ones made by the author, it is worthwhile to highlight this evidence. However, the *absence* of specific work with which a reviewer can disagree is akin to maintaining a dogmatic view that holds back the management field.

2) Not counterintuitive error: Many excellent scholarly statements are available on the subject of what makes for a good theoretical contribution. If a paper meets the conditions necessary for a theoretical contribution, then trivializing the paper for not revealing counterintuitive insights is misplaced.

   Developmental review that corrects the not counter-intuitive error: If the insights of a paper are not counterintuitive to a reviewer, then perhaps the reviewer's I.Q.is extremely high. Though tongue-in-check, my point is simple: what is not counterintuitive to one person may be to someone else. My advice is to not apply this mode of thinking in the review process. Since there is no objective yardstick for what makes a paper counterintuitive, evaluating the appeal of a given paper to a wider audience should be best left to the editors. The caveat is when there is a body of work in a field which shows the same relationship that the authors have replicated. In that case, it is advisable to point to this work and question what additional value the focal paper offers.

3) Many incomplete-alternative mechanisms error: The most common mistake I have made as a rookie reviewer, especially in graduate seminars, is what I call the "many incomplete-alternative mechanisms error." Reviewers are encouraged to articulate an alternative explanation, not to list multiple mechanisms that make some but not all the predictions made by the author. The task is to come up with an alternative mechanism that explains *all the stated predictions and more* within the context of the focal paper. By generating many alternative mechanisms that are different for each of the author's predictions, the noise only increases in a given review. For example in Table 1, reviewer #2 has two theories, C and D, that explain only one outcome of interest for outcome 1 and outcome 2, respectively (columns 6 and 7). These theories C or D should not be preferred to either the author's or reviewer # 1's explanation, as they explain only one outcome each (column 8).

   Developmental review that corrects the incomplete-alternative mechanisms error: The solution is simple. Is there a *single* alternative explanation that can be used to explain *all* of the predictions? Highlighting such an explanation is helpful for the author, since

it allows both the reviewers and the author to focus on a critical experiment that could differentiate the two explanations for the outcomes at hand.

4) Implicit implication that authors fabricated data error: One of the worst offenses I have made as a reviewer is to use a different theory than the author's and to propose predictions contrary to their own. Using Table 1 as an example, let us say that reviewer #2 keeps stating that the author's explanation should not be preferred, since theory D makes a different prediction on the relationship between X and outcome 2 (column 7). Why should anyone, in the context of someone else's paper, care a theory that makes the opposite predictions and is not supported by the data at hand? If a mechanism is internally consistent and is supported by data in the context, why is it the authors' responsibility to explain why another theory that makes opposite predictions does not hold in this context? There are few reasons for levying this critique, unless reviewer #2 believes the authors have fabricated their data.

   Developmental review that corrects the assumed data fabrication error: A caveat to correcting this error may exist if prior work has found the relationship between input X and an outcome to be that of the reviewer's proposed theory rather than the author's. For instance, if prior work has repeatedly found that the relationship between input X and outcome 2 is negative - as suggested by reviewer # 2's theory D (column 7, Table 1) - then cite this body of work and ask the authors why they proposed and found support for the opposite relationship. By referring to such a body of countervailing evidence, good reviewers could question whether or not the author's setting is idiosyncratic enough, or if the measures used by the author are imperfect.

5) Disagreeing with derivations consistent with authors' starting assumptions error: Assume that reviewer #2's theory C suggests that the relationship between input X and outcome 3 is negative (column 6). Then, criticizing the authors for making a positive relationship prediction between input X and outcome 3 is unwarranted. If the authors have been internally consistent with their starting assumptions in the theory they used, theory A in this case, to make a positive relationship prediction from input X and outcome 3, it is inadvisable to criticize the prediction because the authors can make no other prediction from their starting assumption.

   Developmental review that corrects the derivation disagreement error: Disagreement with derivations consistent with authors' starting assumptions is not a valid critique. It is valid

to point out derivations in the rest of the paper that deviate from the authors' proposed mechanism and assumptions. This not only helps bring deviations from the author's assumptions to light, but can also suggest how the author could make the argument consistent with their corresponding mechanism. Once the paper is internally consistent, then both the reviewers and authors can turn their attention to a critical experiment that can determine whether the author's explanation should prevail over another.

6) "I was not creative enough to think of a critical experiment" error: Frequently, my single biggest failing as a referee was my inability to devise a critical experiment that the authors could conduct within their dataset in order to rule out alternative explanations. For instance in Table 1, evidence from the test of outcome 3 would be sufficient to separate out the author's explanation from reviewer #1's. I would often suggest new data collection that would serve as the basis for a critical test. For instance, outcome 4 would serve as one such critical test as well as a means to separate the author's explanation from reviewer #1's. However, outcome 4 would ultimately be a poor critical test, as it would require the author to collect new data. It is not a good use of resources for the authors to collect new data when already-collected data is sufficient to determine whose explanations should be preferred.

   Developmental review that corrects the lack of a creative critical experiment error: The field depends upon the imagination of contributors. Without the creation of a critical test, the field cannot advance. An ideal critical test would be outcome 3 in Table 1, something that the author can readily test with data at hand. When reviewers recommend a critical test, they must also commit to the authors if the evidence is consistent with their mechanism. Then, the paper merits being published.
   Accordingly, reviewers are welcome to suggest that the authors change their theory to the proposed mechanism if the results of the critical test support their point of view.
   Both authors and reviewers should be open-minded enough to revise their explanations based on the evidence from a critical experiment.

7) Emotional filibuster: I have rarely written reviews that are very long; however, I have given four pages of critical points, without including advice on how to improve the paper, putting the editors in an untenable position. Writing four pages on the reasons to reject a paper is not an act of service to the field. Rather, it simply raises the emotional bar that the editor has to cross in order to go against a reviewer's recommendation. As such, I am

inclined to call this approach "emotional filibustering, an approach that attempts to convince the editor to reject the paper.

Developmental review that corrects the emotional filibuster error: If reviewers must write four pages of single-spaced comments, they can serve the field by including two sections in the review: A first section that states the critical experiment(s) (for instance testing outcomes 3 or 4 in Table 1) that the author can conduct, in addition to explaining how they can make their arguments more internally consistent. Then, reviewers can include a separate section with additional commentary more oriented toward their own individual tastes and preferences, which, of course, authors and editors are free to ignore.

Appeal to Editors

Before I wrap up this note, I would like to address an appeal to my senior colleagues, the editors. I've overheard editors say that they value a referee for their diligence even if they have not received a single referee report recommending publication during the course of a given term. As such, I offer two conclusions here based on what I have observed. Somehow, it seems that poor quality papers are often sent to the same referee. In that case, the referees deserve our sympathies for being thus abused. Alternatively, the reviewer may not have a strong mental model of which papers can be accepted. Any referee must have a null model of what they will recommend for publication. The reviewer must reveal this through their comments on what specific changes will make a paper worthy of publication in the journal. Without evidence of this, the editors must help the field by educating the reviewer. If the reviewer does not change their habits, editors must consider removing such reviewers from the reviewer pool.

Conclusion

We in the management field face a conundrum when a paper given to two or more experts results in little convergence of opinion on its deservingness for publication. Accordingly, either we do not understand what makes a paper deserving, or the experts have subjective tastes that do not converge. Kuhn (1962) suggests that strong paradigms solve the taste issue by getting scholars to agree on what questions are interesting and what explanations are worthy of pursuit. Being in domains like management, which contains relatively weak paradigms due to heterogeneity in assumptions and explanations used, does not mean that we should let subjective tastes run amok

in the review process.

While nothing that I have mentioned is new, this article serves as a template for identifying and responding to common errors that reviewers make during the review process. I have started with the assumption that the deservingness of a paper can be objectively measured, following Lave & March (1993) with the factors that make a good paper. If one agrees with Lave & March's model that allows for objective classification of papers, the only remaining problem to solve is to educate ourselves on how to be good reviewers - and that involves being willing to set aside our individual tastes.

A good reviewer follows a widely accepted model which can objectively classify papers as worthy of publication or not, regardless of their personal tastes. As such, I encourage us to sign up to be good reviewers by following a given model, articulating which model we are following, and sticking to that model throughout the review. If we commit to this as a field, I anticipate more convergence in our evaluations of papers. Let me personally confess that, previously, my reviews did not follow the guidelines I listed above, and that I am guilty of committing mistakes as a reviewer. However, I am learning to be a good reviewer, and I hope you'll join me in this journey toward stronger reviews.

References

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* University of Chicago Press.

Lave, C. A., & March, J. G. (1993). *An introduction to models in the social sciences*.

University Press of America.

Romanelli, E. (1996). Becoming a reviewer: Lessons somewhat painfully learned. *Rhythms of academic life: Personal accounts of careers in academia*, 263-268.

Table 1: A Stylized Representation of an Author's and Reviewers' Explanations

	Column 1	*Column 2*	*Column 3*	*Column 4*	*Column 5*	*Column 6*	*Column 7*	*Column 8*
	Outcomes	**Observed in the data**	**Tested by the author**	**Author's theory A makes a prediction between input X and outcomes**	**Reviewer #1's theory B makes a prediction between input X and outcomes**	**Reviewer #2**		**Remarks**
						Theory C makes a prediction between input X and outcome	**Theory D makes a prediction between input X and outcome**	
Focal Paper Submitted to Review	1	Yes	Yes	(+)	(+)	(+)	No prediction	We cannot separate out reviewer #1's explanation (column 5, theory B) and the author's explanation (column 4, theory A) when we examine the first two outcomes on the relationship between input X and outcomes 1 & 2. Reviewer #2's theory C (column 6) and theory D (column 7) explain only one outcome each 1 and 2 respectively. Hence are inferior and should henceforth be ignored for the purpose of this paper.
	2	Yes	Yes	(+)	(+)	No prediction	(-)	
Post-Review Options	3	Yes	No	(+)	(-)	No prediction	No prediction	Outcome 3 is an ideal critical experiment to which any reviewer should aspire. Reviewer #1's theory B (column 5) predicts a negative relationship between input X and outcome 3, whereas the author's theory A (column 4) makes an opposite, positive prediction. Examining outcome 3 will conclusively support either the author's explanation or reviewer #1's explanation. Hence this forms an ideal critical experiment, an outcome of which will support either the author's theory A or the reviewer # 1's theory B.
	4	No	No	(+)	(-)	No prediction	No prediction	A reviewer can only ask for additional data collection if he/she fails to identify a critical experiment that can be tested within already collected data. For instance, when examining outcome 4, reviewer #1's theory B predicts that the relationship between input X and outcome 4 would be negative, whereas the author's theory A predicts a positive relationship. Thus, outcome 4 is a setting that qualifies as a critical experiment. However, this is inferior to testing outcome 3 if this outcome already exists in the collected data.