**REVIEW ARTICLE**

# The Paradox of Reviewer Accuracy

**Wolfgang Lederer**

*Department of Anesthesiology and Critical Care Medicine, Medical University of Innsbruck, Austria.*

Corresponding Author: Lederer W. Medical University of Innsbruck, Department of Anesthesiology and Critical Care Medicine, Anichstr. 35, 6020 Innsbruck, Austria.

## Abstract

The presented paradox is based on the assumption that an obviously well-designed and well- conducted hypothetical study reveals experts being unable to accurately interpret the scientific value of research findings outside of their areas of expertise. In case of two or more experts reviewing research off the mainstream there is a certain chance of getting a positive assessment from one reviewer while getting a negative assessment from another reviewer. Amazingly, regardless of opposing assessments from different reviewers, both judgements may support the conclusion of the hypothetical study. A primarily negative assessment leads to rejection of the study but indirectly may confirm the accuracy of the conclusion e.g. by authority bias. A primarily positive assessment may directly confirm accuracy of the study conclusion e.g. by confirmation bias. At the same time, it questions the accuracy of the evaluation taking into account that the reviewer is inexperienced in this issue. This imagined paradox raises serious concern about the assumed objectivity of an independent and unbiased peer review of studies outside mainstream research.

**Keywords:** Empirical Science, Medical Philosophy, Peer Review, Probability Theory, Scientific Evidence.

## 1. Introduction

Scientific integrity and research ethics are integral properties of an all-encompassing explanatory scientific theory (Chatfield and Law, 2024). However, science creates divergent perspectives of research findings and allows different interpretations of them. This demands an open-minded and heterogeneously arranged scientific community. Contrary to debates between the rigid limits of hard lined confessions of faith and doctrines, in science the individual point of view matters, when attempting to increase insight into a certain topic. The Roman philosopher Horatius (31 BC) emphasized in his first letter that the individual should dare to think in his own way outside the box (literally: "*Dimidium facti qui coepit habet: sapere aude*," meaning that: "he who has begun is already halfway finished: dare to be wise"). The critical assessment of manuscripts submitted to peer-reviewed journals is classified as a high-quality review of the scientific nature of a report (ICMJE, 2024). Reviewers who are considered experts in their research area enable the publication of high-quality investigations according to current standards. This process relies on an independent and unbiased assessment. The evidential strength of different research designs can be rated with a multistory pyramid (Guyatt *et al*, 2013). The higher up, the more rigorous methodology is anticipated. In the pyramid of evidence expert opinions reside at the bottom of the pyramid. Amazingly, at the top of the pyramid there are peer-reviewed systematic reviews and meta-analyses of previously peer-reviewed studies. The fact that expert opinion makes the base of the pyramid while expert review makes the top of the pyramid is strange but not yet paradoxical. In general, assessments in the scientific community are very heterogeneous. Scientific investigation will always produce results that meet one person's expectations

while concomitantly disagreeing with another person's point of view (Lederer, 2023). Accordingly, the assessment of two experts can be contradictory from different perspectives (Deroover *et al*, 2023). The quality of assessment is additionally affected by the imperfectness of language that influences our understanding (Wittgenstein, 1960). The role of the editor is crucial in recognizing novelties. His choice of suitable reviewers strongly influences the chance of the paper being published. Despite the thematic limitations of the individual journals, it is important to think and interpret outside the box. Possibly, the law of triviality, as previously described by Parkinson on high finance also largely applies to reviewer decisions indicating that most attention and time is spend on findings with mean dimensions (Parkinson, 1957). Critical reviews are intended to improve research but it would be unethical if they prevented research. On the one hand, specialist guidance, e.g. advisory service from a statistician on data analysis can increase the explanatory power of a study and avoid low-quality research. On the other hand, a review of excellent research needs extraordinary experts. In this critical analysis I want to question the quality of peer review outside mainstream research and to point out a paradoxical condition in reviewer accuracy.

## 2. Methods

According to the Stanford Encyclopedia of Philosophy (2021) a paradox can be defined as a statement claiming something which goes against what is usually believed or held. It is an understandable but self-contradictory statement that can be true and false at the same time. In logical thinking a paradox is an invalid argument which is created, for example, through the use of different definitions (Stanford, 2021). In this essay the paradox describes a condition that excludes itself. This paradox does not take into account that an expert`s decision could also be influenced by non-technical and ignoble aspects.

It is not my intention to provoke a never ending discussion between philosophical dogmatism and skepticism that finally leads to absurdity. This analysis rather aims to determine a weakness in the epistemological pursuit to find the truth. I do not generally question the expertise of individual reviewers and I trust them having high sensitivity and specificity in their fields of experience. However, I dare to doubt the predictive value of expert assessments outside main stream research. I propose there must be conditional probabilities of the accuracy of reviewer comments (Fig. 1).
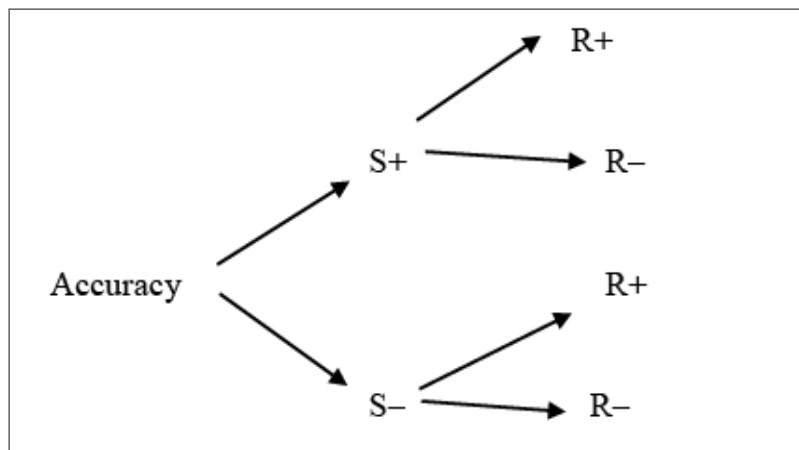


**Figure 1.** *The assumed distribution of conditional probabilities of accurate reviewer assessment (R+) and inaccurate reviewer assessment (R–) on accurate study conclusions (S+) and on inaccurate study conclusions (S–) outside main stream research using a tree diagram*

Conditional probabilities depend on predetermined total probabilities. In this essay the considerations of an accurate reviewer assessment follow a strict binomial distribution of independent events in a dichotomous, homogenous, constant and rigid system. When applying Bayes' theorem (Bayes, 1763; Papoulis, 1984) the probability (P) of accurate reviewer assessment (R+) given that the study conclusion from research off the mainstream is accurate (S+) appears to be quite low (1).

$$P(S+) = P(S+|R+) \times P(R+) + P(S+|R-) \times P(R-)$$

$$P(R+|S+) = \frac{P(S+|R+) \times P(R+)}{P(S+)}$$

$$P(R+|S+) = \frac{P(S+|R+) \times P(R+)}{P(S+|R+) \times P(R+) + P(S+|R-) \times P(R-)}$$

$$P(R+) = \frac{P(S+) - P(S+|R-) \times P(R-)}{P(S+|R+)} \tag{1}$$

P (S+) probability that the study conclusion is accurate

P (S–) probability that the study conclusion is inaccurate P(R+) probability that the reviewer assessment is accurate

P(R–) probability that the reviewer assessment is inaccurate

P(R+|S+) probability of accurate reviewer assessment given that the study conclusion is accurate

## 3. The Reviewer Paradox

The paradox of reviewer accuracy deals with peer review and applies to both, the observed expert and the observing expert. Consequently, the paradox of reviewer accuracy arises from the fact that the person who analyses and the person who gets analyzed are identical. Generally, following peer review there are two possibilities of opposite assessment, acceptance (usually demanding minor or major revisions) and rejection.

This paradox is based on the assumption, that an obviously well-designed and well- conducted hypothetical study reveals that experts cannot accurately interpret the scientific value of research findings outside their areas of expertise. Taking into account the conditional probabilities mentioned above, the chances of such a submitted work being rejected are high. However, applying dichotomous thinking, then both acceptance and rejection would confirm the accuracy of the study results.

Condition A: the reviewer accepts the conclusion and thereby admits that experts cannot really interpret the scientific value of research findings outside their areas of expertise. Confirming the accuracy of the conclusion the reviewer also confirms the potential inaccuracy of his comment. This would question the reviewer`s previous assessment.

Condition B: the reviewer does not accept the conclusion regarding limited assessment abilities of experts. When the reviewer does not confirm the accuracy of the conclusion then he involuntarily admits the correctness of the hypothesis and therefore, the inaccuracy of his statement. Amazingly, regardless of the contrary opinion both reviewer judgements support the conclusion of the hypothetical study. While a positive assessment directly confirms accuracy of the study conclusion (e.g. by confirmation bias) a negative assessment indirectly confirms the accuracy of the conclusion (e.g. by authority bias).

## 4. Implications

The findings of this hypothetical consideration suggest that publication of results from studies outside mainstream research has a low chance. In the case of just guessing, the number of one correct answer divided by two total possibilities results in 0.5 indicating a fifty-fifty chance similar to a coin toss. It implies that the total of all acceptances (a) matches the total of all rejections (r), corresponding to a harmonious distribution. However, the accuracy of expert assessment can potentially even be lower than the probability of guessing. Colson and Cooke reported in their validation of 33 studies, that less than one-third of the individual experts were statistically accurate when scoring experts performance against empirical data (Colson and Cooke, 2018).

When the above-mentioned study is reviewed by two or more experts, the final decision based on differing assessments will most likely result in rejection. Comparably, a to the power n ($a^n$) minus r to the power n ($r^n$) results in zero ($|a^{n}| - |b^{n}| = |0|$). This also concerns the ratio of acceptance to equal rejection as the logarithm of total a to the base 10 related to the logarithm of total r to the base 10 results in zero (2).

$$\log_{10}(a) - \log_{10}(r) = \log_{10}(a/r) = \log_{10}(1) = 0 \quad (2)$$

This underlines the demand that the scientific value of research with little background knowledge should not be judged solely on the basis of comments made by a single point of view from an expert. Instead, a study's validity should be assigned by the scientific community together with the individual observer, reviewer and editor. Expert elicitation allows for quantification of uncertainty in order to achieve scientific consensus. When expert assessments are weighted according to scores based on expert performance, this combination seems to be superior to alternative approaches, such as likelihood-based and social network weights, scientific community weights, peer weights, averaging expert's quantiles, and harmonic weighting. The performance-based combination of experts is regarded more statistically accurate and more informative than an equal weighting of experts (Colson and Cooke, 2018). However, there is a blurred boundary between scientific knowledge and scientific belief. Elicited probabilities of expert statements about unknown parameters demands training the experts to express their beliefs as probabilities (O'Hagan et al, 2006). In addition, transparency demands that a reviewer's comments on a published study should always be

accessible to the reader. Ideally, editorial decisions should be based on the validity of the work and its importance to the scientific community independent from ideological and commercial goals of the publisher (ICMJE, 2024). Expert judgement plays an important role in uncertainty analysis but should not provide the final word on an issue (Colson and Cooke, 2018). It is the contrastive discussion and the exchange of ideas in a scientific forum that make individual observations meaningful as science progresses by dialogue.

## 5. Conclusion

Experts in research and ethic committees decide the scientific value of research. In most scientific journals the editorial decision based on peer review determines whether a submitted study will be published or not. While high quality peer-review can be assumed for most research there is a certain risk of confirmation bias and authority bias especially when studies were conducted outside mainstream research.

**Competing Interests**

The author has no conflicts of interests to declare including financial, consultant, institutional and other relationships that might lead to bias or a conflict of interests.

## 6. References

1. Bayes, T. (1763). LII. *An Essay towards solving a Problem in the Doctrine of Chances.* By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S. https://bayes.wustl.edu/an.essay.pdf

2. Chatfield, K., Law, E. (2024). 'I should do what?' Addressing research misconduct through values alignment. *Research Ethics*. https://doi.org/10.1177/17470161231224481

3. Colson, A. R., Cooke, R. M. (2018). Expert Elicitation: Using the Classical Model to Validate Experts' Judgments. *Review of Environmental Economics and Policy*. 12:1 https://www.journals.uchicago.edu/doi/epdf/10.1093/reep/rex022

4. Deroover, K., Knight, S., Burke, P. F., Bucher, T. (2023). Why do experts disagree? The development of a taxonomy. *Public Understanding of Science*, 32(2), 224–246. https://doi.org/10.1177/09636625221110029

5. Guyatt, G. H., Thorlund, K., Oxman, A. D., Walter, S. D., Patrick, D., Furukawa, T. A., Johnston, B. C., Karanicolas, P., Akl, E. A., Vist, G., Kunz, R., Brozek, J., Kupper, L. L., Martin, S. L., Meerpohl, J. J., Alonso-Coello, P., Christensen, R., Schunemann, H. J. (2013). GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. *Journal of Clinical Epidemiology*, 66(2), 173–183. https://doi: 10.1016/j.jclinepi.2012.08.001.

6. ICMJE. (2024). Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. (2024). Retrieved May 24, 2024, from https://www.icmje.org/icmje-recommendations.pdf

7. Lederer, W. (2023). The Importance of "Nothingness" in Empirical Science – a Hypothesis". *Journal of Philosophy and Ethics*; 5(2): 2–4https://sryahwapublications.com/journals/journal-of-philosophy-and-ethics/volume-5/issue-2

8. O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R.; Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., Rakow, T. (2006). Chapter 4: *The Elicitation of Probabilities*, in: Uncertain Judgements: Eliciting Experts' Probabilities. West Sussex, John Wiley & Sons, UK.

9. Papoulis, A. (1984). Chapter 2: *Total Probability and Bayes' Theorem*, in: Probability, Random Variables, and Stochastic Processes (4th ed.), McGraw-Hill, New York: p. 32.

10. Parkinson, N. C. (1957). *High finance, or the point of vanishing interest*, in: Parkinson's law and other studies in administration. Cambridge Massachusetts, The Riverside Press, USA.

11. Stanford. (2021). *Paradoxes and Contemporary Logic*, in: Stanford Encyclopedia of Philosophy, *First published Tue Oct 16, 2007; substantive revision Tue Apr 20, 2021*, https://plat.stanford.edu/entries/paradoxes-contemporary-logic/

12. Wittgenstein, L. (1960). Philosophische Untersuchungen [Philosophical Investigations] ed.: J. Schulte, Suhrkamp, Frankfurt am Main, Germany