
10-12-2020

The Best Predictions in Experimental Biology are Critical and Persuasive

Douglas S. Fudge

Chapman University, fudge@chapman.edu

Andy J. Turko

McMaster University

Follow this and additional works at: https://digitalcommons.chapman.edu/sees_articles



Part of the [Animal Experimentation and Research Commons](#), [Biology Commons](#), [Laboratory and Basic Science Research Commons](#), and the [Other Life Sciences Commons](#)

Recommended Citation

Douglas S. Fudge, Andy J. Turko; The best predictions in experimental biology are critical and persuasive. *J Exp Biol* 1 October 2020; 223 (19): jeb231894. <https://doi.org/10.1242/jeb.231894>

This Article is brought to you for free and open access by the Science and Technology Faculty Articles and Research at Chapman University Digital Commons. It has been accepted for inclusion in Biology, Chemistry, and Environmental Sciences Faculty Articles and Research by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.

The Best Predictions in Experimental Biology are Critical and Persuasive

Comments

This article was originally published in *Journal of Experimental Biology*, volume 223, issue 19, in 2020.
<https://doi.org/10.1242/jeb.231894>

Copyright

The Company of Biologists

COMMENTARY

The best predictions in experimental biology are critical and persuasive

Douglas S. Fudge^{1,*} and Andy J. Turko²

ABSTRACT

A powerful way to evaluate scientific explanations (hypotheses) is to test the predictions that they make. In this way, predictions serve as an important bridge between abstract hypotheses and concrete experiments. Experimental biologists, however, generally receive little guidance on how to generate quality predictions. Here, we identify two important components of good predictions – criticality and persuasiveness – which relate to the ability of a prediction (and the experiment it implies) to disprove a hypothesis or to convince a skeptic that the hypothesis has merit. Using a detailed example, we demonstrate how striving for predictions that are both critical and persuasive can speed scientific progress by leading us to more powerful experiments. Finally, we provide a quality control checklist to assist students and researchers as they navigate the hypothetico-deductive method from puzzling observations to experimental tests.

KEY WORDS: Scientific method, Hypothesis testing, Experimental design

Introduction

The scientific method (i.e. the hypothetico-deductive method) is a powerful means of discovery because it provides a mechanism for evaluating explanations of how the world works, allowing us to reject bad explanations and increase our confidence in good ones (Deutsch, 2011). The difference between good and bad explanations has everything to do with the predictions they make; good explanations make accurate predictions, and bad explanations fail this test (Feynman et al., 1965). In this Commentary, we discuss how to answer open-ended questions by testing competing explanations, focusing on the predictions the explanations make and where those predictions diverge. We argue that the best predictions have the potential to do two things – disprove incorrect explanations and increase confidence in correct ones. We also provide a checklist that we hope will be useful for experimental biologists and students as they make the most of limited time and resources to tackle open-ended questions (Box 1).

First, let us clarify what we mean by open-ended questions. These are questions for which the range of possible answers is not constrained and it is not obvious how they should be answered. Often, these questions inquire about mechanistic explanations of natural phenomena, usually taking the form ‘How does this work?’ or ‘Why does this occur?’. Many ‘big’ questions in biology are open-ended – for example, ‘How do animals maintain homeostasis?’ or ‘Why does individual variation persist?’ – and

answers to open-ended questions therefore often form the basis for general theory. However, many specific and fascinating natural history questions are also open-ended. Consider an observation that our undergraduate marine biology students recently made while exploring the rocky intertidal zone in Maine, USA. They noticed aggregations of the marine springtail *Anurida maritima* floating on the surface of several tidepools (Fig. 1), which led them to ask ‘Why do these animals form rafts?’. Although this is a simple question, it is also a difficult one, because it is not obvious how to go about answering it. Possible answers to this question (i.e. explanations) range widely, from physical mechanisms (e.g. they stick to each other like Cheerios in a bowl of milk) to biological ones (e.g. their food occurs in patches and they go where their food is). In contrast to open-ended questions, constrained questions have fewer possible answers. For example, we might ask the question ‘What is the effect of wind speed on the size of *A. maritima* rafts?’. The form that the answer to this question will take is obvious (increased wind speed might increase raft size, decrease it or have no effect), as is the experiment you could do to answer it (expose springtails to varying wind speeds and measure raft size). We do not wish to minimize the complexities of answering constrained questions; indeed, they often require elaborate experimental designs, impressive technical skill and sophisticated use of statistics. Our point is that open-ended questions are particularly difficult because the number of potential answers is unlimited, and it is these kinds of questions that we will focus on in this Commentary.

What makes a good prediction?

The hypothetico-deductive method is a particularly powerful tool for answering open-ended questions (Deutsch, 2011). In his seminal paper on ‘strong inference’, Platt (1964) sums up the hypothetico-deductive method in three steps: (1) devise alternative explanations (in science, these are called hypotheses); (2) devise a crucial experiment (i.e. one that can disprove one or more of the hypotheses under consideration); and (3) carry out the experiment so as to get a clean result. The vast majority of the training that we give and receive as experimental biologists focuses on step 3 of this process, i.e. experimental design and data analysis. Step 1, hypothesis generation, is a fascinating and mysterious creative process (Fudge, 2014), but it is not our focus here. Step 2, which involves moving from a list of alternative hypotheses to a ‘crucial experiment’ is not as simple as it sounds in Platt’s instructions. In the remainder of this Commentary, we will demonstrate how predictions can be a powerful tool for navigating this transition from step 1 (hypothesis generation) to step 2 (crucial experiment).

Before we discuss how to come up with crucial experiments, we would like to address a common misconception about the function of predictions when using the hypothetico-deductive method. Our use of the word ‘prediction’ does not simply mean a guess about how an experiment will turn out (Hutto, 2012). Instead, predictions are the logical outcomes of hypotheses under a given set of

¹Schmid College of Science and Technology, Chapman University, 1 University Dr., Orange, CA 92866, USA. ²Department of Biology, McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada.

*Author for correspondence (fudge@chapman.edu)

© D.S.F., 0000-0001-9569-1374; A.J.T., 0000-0002-6330-5798

conditions, and therefore they must be considered before the experiments are designed. In fact, it is the act of finding these predictions that leads us to our experiments. Although it is not obvious from the way that the hypothetico-deductive method is often taught, making predictions for constrained questions has little to no utility (Hutto, 2012). For example, if we want to know the effect of wind speed on springtail rafts, we can just conduct our experiment and find the answer. In this case, generating an *a priori* guess about the results is nothing more than a distraction and might even bias our data collection.

As mentioned above, when Platt refers to a ‘crucial experiment,’ he means one that has the potential to disprove one or more of the hypotheses under consideration, and the best way to disprove a hypothesis is to test the predictions that it makes. But how exactly do we figure out what a particular hypothesis predicts? We have learned from teaching this method to our students (and from our own struggles) that it is all too easy to make flawed predictions. Flawed predictions are ones that don’t align with the hypothesis, that aren’t clear about the experimental test or that won’t yield useful information if they are tested. We have therefore devised some simple guidelines for evaluating predictions before proceeding to the experimental testing phase. To evaluate the quality of a prediction, we always ask ourselves the following two important questions. (1) What would you learn if the prediction were found to be false? Would such a result force you to reject the hypothesis that made the prediction? If the answer is yes, we call this a ‘critical prediction’. The importance of criticality has been recognized for a long time, as falsification of a hypothesis necessarily relies on testing critical predictions (Popper, 1959; Lakatos, 1970). However, we have found that emphasizing only criticality can sometimes lead to experiments that fail to provide much insight into the phenomenon at hand. To avoid this pitfall, which we describe in more detail below, we recommend asking a second question. (2) What would you learn if the prediction were found to be true? Would such a result increase your confidence in the hypothesis that made the prediction? If the answer is yes, we call this a ‘persuasive prediction’. The importance of persuasiveness in hypothesis testing has not been widely recognized, but in our experience, explicit consideration of persuasion is helpful for evaluating the power of various experimental options and finding the experiments that will be most illuminating.

It is important to stress that criticality and persuasiveness are not mutually exclusive; in fact the best predictions to test are both critical and persuasive.

Generating critical and persuasive predictions – a case study

In order to illustrate the concepts of criticality and persuasiveness, let us consider some hypotheses and predictions that our students generated when thinking about the springtail example we introduced earlier. Their observation that intertidal springtails form rafts on the surface of tidepools led them to ask the open-ended question ‘Why do intertidal springtails form rafts?’. The first hypothesis they came up with posits that this phenomenon has something to do with reproduction.

Hypothesis: The rafts are mating aggregations. We will refer to this as the ‘mating aggregation’ hypothesis.

Based on this hypothesis, they came up with the following prediction.

Prediction: *Anurida maritima* is a sexually reproducing species of springtail.

First, we ask whether the prediction is critical. In other words, if we found it to be false, i.e. if this species of springtail is an

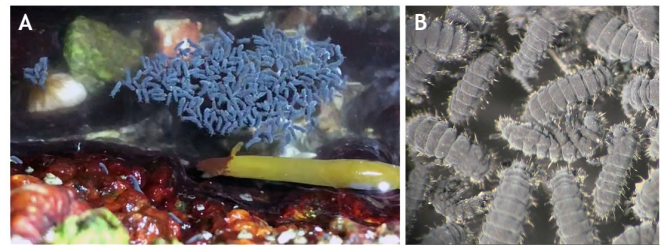


Fig. 1. The springtail *Anurida maritima* forms rafts of many individuals on the surface of tidepools. (A) Large and small rafts of springtails on the water surface of a tidepool. (B) Close-up of a raft of *A. maritima*.

obligately asexual species, would that force us to reject the mating aggregation hypothesis? The answer is yes: the mating aggregation hypothesis could not survive if we found that *A. maritima* were not a sexually reproducing species. Next, we ask whether the prediction is persuasive. In other words, if we found it to be true, i.e. if *A. maritima* is indeed a sexual species, would that increase our confidence in the mating aggregation hypothesis? The answer to that question is no, because the vast majority of animal species reproduce sexually, and so demonstrating sexual reproduction does little to explain why these rafts form. The lesson here is that it is possible to come up with predictions that are critical, but not persuasive. Such a prediction has the potential to disprove a hypothesis, but little power to convince a skeptic that the hypothesis is true. More persuasive predictions would contain more detail about the connection between reproduction and rafting. For example, the mating aggregation hypothesis also predicts that only sexually mature adult springtails should be found in rafts, and that evidence of reproduction (e.g. release of spermatophores by males) should be observable (Hopkin, 1997). Both of these predictions are more persuasive than the one about *A. maritima* being a sexual species.

Let us consider a completely different hypothesis that we could put forth for the same question about why springtails form rafts.

Hypothesis: Because of their small mass and the surface tension of water, springtails on the surface of a tidepool stick to each other via the ‘Cheerios effect’ (Vella and Mahadevan, 2005). We will refer to this as the ‘Cheerios’ hypothesis.

Whereas our first hypothesis was biological in nature, this one poses an entirely physical explanation, i.e. springtails aggregate because of attractive forces that arise from surface tension effects. What are some predictions made by this hypothesis?

Prediction: Because the Cheerios effect relies on surface tension, lowering the surface tension of water with a surfactant will stop springtails from aggregating.

Firstly, is this prediction critical? That is, if we found that adding a surfactant (like soap) had no effect on the tendency of the springtails to form aggregations, must we reject the Cheerios hypothesis? Not necessarily, because we do not know how much the surface tension will be lowered by the addition of soap. If it is not lowered enough to abolish the Cheerios effect, then it is possible that the aggregations will persist in the presence of soap even if the Cheerios hypothesis is true. Therefore, this is not a critical prediction. Secondly, how persuasive would it be if we found that soap strongly inhibited the formation of springtail rafts? The prediction is somewhat persuasive, because it demonstrates a possible link between surface tension and raft formation, just as the hypothesis proposes. However, a skeptic might say there are other plausible explanations for how soap might affect raft formation that have nothing to do with the Cheerios effect. For example, if you

Box 1. Checklist for answering open-ended questions in experimental biology

1. Identify observations or patterns that are unexplained and 'puzzling'.
2. Ask an open-ended question about the knowledge gap identified in step 1. These often begin with 'How' or 'Why'.
3. Generate a list of plausible, intellectually satisfying and logically consistent answers (i.e. hypotheses) to the question posed in step 2. Hypotheses should be written in the present tense and should read like explanations. Hypotheses written in the future tense are easily confused with predictions.
4. Give each hypothesis a short name (e.g. the Cheerios hypothesis). This will make it easier to think through the logic of the predictions that it makes (e.g. 'If the Cheerios hypothesis is true, then when we reduce temperature...').
5. Reflect on each hypothesis and make sure it represents a satisfying answer to the question posed in step 2. One way to do this is to ask whether the hypothesis predicts the original puzzling observations identified in step 1.
6. Generate a list of predictions for each hypothesis. Predictions should be written in the future tense, as they should describe what a hypothesis predicts will happen under a given set of experimental conditions. In this way, predictions are a bridge between hypotheses, which are abstract ideas, and experiments, which are concrete scenarios in the real world. If you are having trouble coming up with predictions, ask yourself what the essential differences are amongst the hypotheses you are considering. Under what conditions will your competing hypotheses make divergent predictions?
7. Ask whether your predictions are critical and/or persuasive. Ideally, you will generate some predictions that are both critical and persuasive. If a prediction is lacking in criticality or persuasiveness, try revising it to address this.
8. If you have multiple experimental options, start with the experiments that test the most critical and persuasive predictions.
9. Run the experiment and collect the data.
10. Analyze the data and decide (using statistical methods if necessary) whether the various predictions made by your hypotheses are true or false.
11. For critical predictions found to be false, reject the hypotheses that made them. For persuasive predictions found to be true, increase your confidence in those hypotheses.
12. Experimental results often generate new and puzzling observations. To find answers to new questions raised by these observations, return to step 1.

suspected that rafts form because springtails actively paddle towards their nearest neighbor, then one can imagine that adding soap to the water might poison the springtails and stop them from paddling. Thus, one component of a prediction's persuasiveness is its potential to exclude competing hypotheses.

In the above example, asking whether a prediction is critical and/or persuasive forced us to think hard about what we might learn from testing it, and we concluded that the prediction as written is not critical. Is it possible to change it so that it becomes more critical? Doing a bit of reading about the physics of surface tension leads us to the fact that surfactants reduce surface tension in a concentration-dependent manner. Perhaps making our prediction more specific would make it more critical. Consider the following revision.

Prediction: Inhibition of raft formation in springtails will depend on surfactant concentration.

Is this prediction now more critical than the original? What would we learn if it were found to be false, i.e. if raft formation were completely unaffected by surfactant concentration? By doing the experiment over a wide range of concentrations, we are much more likely to lower the surface tension to a point where it will interfere

with the Cheerios effect. If we find that raft formation is unaffected by the addition of surfactant over the entire range of concentrations, this would deal a more serious blow to the Cheerios hypothesis than the simpler experiment of just adding some soap and seeing whether it affects raft formation. Thus, our new prediction is more critical than its predecessor. Is it entirely critical though? The answer is no, because it is possible that the chosen surfactant does not lower the surface tension enough to interfere with the Cheerios effect in a way that would inhibit rafting. In this case, the lack of criticality is the result of our lack of knowledge of two things: (1) the amount of surface tension required for the Cheerios effect, and (2) the degree to which a given concentration of our surfactant will lower the surface tension. Consideration of these issues before doing an experiment will likely lead us to a deeper understanding of the mechanisms we are proposing and would push us to make our hypothesis and predictions even more explicit. It is also possible that springtails might compensate for a surfactant-induced loss of surface tension by paddling to stay close to their neighbors. Indeed, the ability of organisms to change their behavior, physiology or morphology can often confound our attempts to probe mechanistic hypotheses by reducing the criticality of our predictions.

How persuasive is the new prediction? What would we learn if we found it to be true, i.e. that raft formation was inhibited more and more as we increased surfactant concentration? We decided that the original prediction was somewhat persuasive, but that it could be more persuasive, given that other hypotheses could account for an inhibition of raft formation with the addition of surfactant. Is the new prediction more persuasive? The answer is yes, because finding it to be true would not only establish that rafting is affected by a surfactant but also demonstrate a more detailed quantitative relationship between these two variables that is consistent with the hypothesis. Of course, finding this prediction to be true would not confirm the Cheerios hypothesis, but it would reduce the number of hypotheses that can explain both the original puzzling observation (springtails form rafts) and our new observations (raft formation and surfactant concentration are negatively correlated). Earlier, we raised the alternative possibility that rafts form when springtails paddle toward their nearest neighbor. If we found that adding soap inhibits raft formation, the 'nearest neighbor' hypothesis could account for this result, but it would have a much harder time explaining why the effect should be concentration dependent. Although it is possible that the nearest neighbor hypothesis might predict a surfactant concentration-dependent effect, it is unlikely that the shape of the response curve would be the same as that predicted by a mechanism involving a loss of surface tension. To summarize, the more specific prediction is more critical and more persuasive than the original; therefore, the experiment it leads to is better because it has greater potential to either falsify or bolster the hypothesis.

Thinking about the persuasiveness of a prediction has the added benefit of forcing us to think about alternative hypotheses, which in turn can lead to new lines of inquiry. Let us explicitly consider one of these hypotheses – the nearest neighbor hypothesis – and a prediction that it makes.

Hypothesis: Springtails tend to paddle toward their nearest neighbor, which over time leads to rafts.

Prediction: Immobilizing the springtails will abolish raft formation.

Is this a critical prediction? In other words, if we immobilized springtails and they still formed rafts, would it force us to reject the nearest neighbor hypothesis? The answer seems to be yes; it would be difficult for the hypothesis to survive such a result. What about

the persuasiveness of this prediction? What would we potentially learn if immobilizing the springtails inhibited raft formation? Thinking deeply about this question makes us realize that the persuasiveness might depend on exactly how we immobilize them, because our method will determine whether we can simultaneously evaluate the Cheerios hypothesis. For example, if we immobilize the springtails by killing them with soap, a negative effect on raft formation would not be very persuasive for the Cheerios hypothesis because it would not be clear whether rafting had been disrupted by the reduction in surface tension or the reduction in paddling. Ideally, we would find an experiment for which the Cheerios and nearest neighbor hypotheses make divergent predictions. Such an experiment is what Platt would refer to as a ‘crucial experiment’.

What if we used cold temperature to immobilize the springtails? Because springtails are ectotherms, their activity should decrease at low temperatures, and thus the nearest neighbor hypothesis predicts that raft formation should be inhibited as we reduce the temperature and should be maximally inhibited at a temperature when they stop moving completely. Conveniently, lowering temperature increases surface tension; this is ideal, because when temperature is reduced, the nearest neighbor hypothesis predicts that rafting should be inhibited, and the Cheerios hypothesis predicts that it should be strengthened. In this case, thinking hard about the persuasiveness of a prediction has led us to a ‘crucial’ experiment.

What makes predictions persuasive?

Above, we have provided several examples where considering persuasiveness has led to more informative experiments, but what exactly makes a prediction persuasive or unpersuasive? From the above examples, we can see that persuasiveness comes from the ability to convince a skeptic that a hypothesis has merit. A skeptic can be defined as ‘one who is willing to question any claim to truth, asking for clarity in definition, consistency in logic, and adequacy of evidence’ (Kurtz, 1992). We realize that our definition of persuasion is inherently fuzzy, because convincing someone that an idea has merit is not an all-or-none endeavor, but rather relies on the subjective judgment of a skeptical person. Unlike the process of falsification, where a false critical prediction can lead to the wholesale rejection of a hypothesis, there is no clear moment when a skeptic declares that they are persuaded. This situation arises from the fundamental hypothetico-deductive principle that hypotheses can never be definitively proven, only supported or disproven.

Although the persuasiveness of a prediction can be somewhat fuzzy, we have identified some common themes of effective scientific persuasion. At a minimum, a skeptic will want to see honest, good faith attempts to falsify a hypothesis, which underscores the importance of critical predictions. However, as we have shown, not all critical predictions are persuasive, so clearly there is more to the story. We have found that the most persuasive predictions are those that push the limits of what should be observable if a given hypothesis is true. These predictions are usually highly specific, quantitative and contain details that align closely with the hypothesis. Unpersuasive predictions tend to be ‘safe’ and focus on observations that are vague and likely to be true. For example, our prediction about the relationship between surfactant dose and rafting response is more persuasive than a prediction about simply adding a single dose of surfactant. An added benefit of highly specific predictions is that they have greater potential to falsify competing hypotheses because the more elaborate and specific a prediction is, the less likely it is that other hypotheses will make the same prediction. In short, persuasive

predictions help lead us to Platt’s ‘crucial experiments’ by focusing on experimental conditions for which competing hypotheses are more likely to make divergent predictions. Of course, because no single persuasive prediction can ‘prove’ a hypothesis, finding multiple persuasive predictions to be true builds a body of evidence that is often more convincing than the results of a single experiment.

Conclusions

We hope we have persuaded you that both the criticality and persuasiveness of predictions should be considered when deciding which lines of experimentation to pursue. If you realize that a given prediction is neither critical nor persuasive, think about how it (and the experiment it implies) might be revised to increase the chances that you will learn something regardless of the experimental outcome (Box 1). We should add that the whole point of doing things this way is to speed progress. If trying to generate predictions that are both critical and persuasive leads to paralysis, then it makes good sense to forge ahead by testing predictions that might still lack one of these elements. As most scientists know, generating a fresh set of observations can sometimes break a logjam, even if the exact implications of those observations are not clear before the experiment is done.

Thinking hard about whether predictions are critical and persuasive has the added benefit of making it easier to write manuscripts and grant proposals and respond to reviewer critiques. Laying out a study in terms of puzzling observations, hypotheses, predictions and experiments aligns well with the logical structure of scientific papers and is especially helpful for writing Introduction and Discussion sections in which the narrative is obvious and the stakes of each experiment are clear. Furthermore, testing the strongest possible critical predictions signals to a reviewer that you have taken your responsibility to falsify your hypothesis seriously, and striving for persuasive predictions means that you have pushed your hypothesis to its logical limits and have considered a number of reasonable alternative hypotheses.

Working through the scientific process in the manner we describe here is clearly a lot of work. Why bother to ask open-ended questions, develop competing hypotheses and evaluate the criticality and persuasiveness of predictions when it is often easier to focus on constrained questions? Our view is that open-ended questions are a simple and powerful tool for developing broad, mechanistic explanations of how the world works. Answering constrained questions is undoubtedly an important part of the scientific process and can provide detailed observations about how variables interact under specific conditions. However, there are drawbacks to starting with constrained questions and considering their implications only after the data are in. One risk is that thinking hard about the implications of a dataset often reveals that it would have been better to do the experiment in a different way; having this realization earlier in the process is almost always beneficial. Another consideration is that writing a thoughtful Discussion section of a manuscript involves thinking deeply about what the data say about the merit of various ideas in the scientific literature, and doing this rigorously involves asking what each of these ideas predicts for the experiments that were carried out. If a researcher needs to engage in this process while writing a manuscript, why not do it before the experiments are planned and executed? The difference in the timing of this process can sometimes be the difference between the rigorous testing of hypotheses and hand waving.

There are other important advantages to this approach. Asking an open-ended question is a deliberate act of open-mindedness that

creates space for multiple competing hypotheses (Chamberlin, 1890). In contrast, constrained questions often have an explanatory hypothesis built into them, which can result in the experimenter becoming ‘attached’ to a particular explanation before the data are collected (Betini et al., 2017). We feel that the approach we are advocating can help remedy some possible causes of the reproducibility crisis that threatens to undermine the credibility of science and scientists (Ioannidis, 2005). By focusing on open-ended questions, entertaining multiple working hypotheses and testing multiple predictions for a given hypothesis, we are less likely to fall into the traps of p-hacking, only reporting ‘significant’ data or hypothesizing after results are known (‘HARKing’; Kerr, 1998). Carrying out an experiment for which hypotheses make divergent predictions (i.e. a ‘crucial’ experiment) means that its outcome will illuminate the question at hand, regardless of whether the results are ‘significant’ or not.

Acknowledgements

We are grateful for thoughtful feedback from the following mentors and readers: E. Don Stevens, Patricia Wright, Sigal Balshine and members of the Balshine Lab, Gary Burness and members of the Burness Lab, Trevor Pitcher and members of the Pitcher lab, William Wright, Charlene McCord, Dennis Taylor, members of the Fudge Lab, and our students and colleagues at the Shoals Marine Lab, who helped us refine the ideas in this essay over many stimulating summers. Steve Crawford taught us the trick of giving hypotheses short nicknames.

Competing interests

The authors declare no competing or financial interests.

Funding

A.J.T was supported by an E. B. Eastburn Fellowship from the Hamilton Community Foundation.

References

- Betini, G. S., Avgar, T. and Fryxell, J. M.** (2017). Why are we not evaluating multiple competing hypotheses in ecology and evolution? *R. Soc. Open Sci.* **4**, 160756. doi:10.1098/rsos.160756
- Chamberlin, T. C.** (1890). The method of multiple working hypotheses. *Science* **15**, 92-96. doi:10.1126/science.ns-15.366.92
- Deutsch, D.** (2011). *The Beginning of Infinity: Explanations that Transform the World*, p. 496. UK: Penguin.
- Feynman, R. P., Leighton, R. B. and Sands, M.** (1965). The Feynman lectures on physics; vol. i. *Am. J. Phys.* **33**, 750-752. doi:10.1119/1.1972241
- Fudge, D. S.** (2014). Fifty years of JR Platt’s strong inference. *J. Exp. Biol.* **217**, 1202-1204. doi:10.1242/jeb.104976
- Hopkin, S. P.** (1997). *Biology of the Springtails*. Oxford: Oxford University Press.
- Hutto, R. L.** (2012). Distorting the Process of scientific inquiry. *Bioscience* **62**, 707-708. doi:10.1525/bio.2012.62.8.3
- Ioannidis, J. P.** (2005). Why most published research findings are false. *PLoS Med.* **2**, e124. doi:10.1371/journal.pmed.0020124
- Kerr, N. L.** (1998). HARKing: Hypothesizing after the results are known. *Pers. Soc. Psychol. Rev.* **2**, 196-217. doi:10.1207/s15327957pspr0203_4
- Kurtz, P.** (1992). *The New Skepticism. Inquiry and Reliable Knowledge*. Buffalo: Prometheus Books.
- Lakatos, I.** (1970). Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of Knowledge* (ed. I. Lakatos and A. Musgrave), pp. 170-196. doi:10.1017/CBO9781139171434.009
- Platt, J. R.** (1964). Strong inference. *Science* **146**, 347-353. doi:10.1126/science.146.3642.347
- Popper, K. R.** (1959). *The Logic of Scientific Discovery*. Basic Books.
- Vella, D. and Mahadevan, L.** (2005). The ‘Cheerios effect’. *Am. J. Phys.* **73**, 817-825. doi: 10.1119/1.1898523