

WHAT MAKES FOR A GOOD SCIENTIFIC THEORY?

The standard “scientific method” is roughly: identify a question, formulate a hypothesis, design an experiment, run the experiment, and confirm/disconfirm the hypothesis based on the result.

This is perhaps how science operates in many cases. But in other cases, it conflates two different parts of scientific procedure: *discovery* versus *justification*.

Discovery can be rather chaotic! We all know stories where an important discovery occurred by accident (x-rays, penicillin, etc.).

Justification is needed to verify that we have not simply made a mistake. This is often where experiments are performed, in order to *double-check* our theory of *X* rather than make an initial discovery.

So often, assessing a scientific theory is not simply a matter of asking whether the scientific method was followed. But how should we assess a scientific theory? There’s been lots of discussion on this; below is my attempt to summarize the insights of those discussions.

Assessing a Scientific Theory

A typical **scientific theory of *X*** has (roughly) three parts:

- (A) Background Theories—theories taken for granted when making observations, predictions, and explanations about *X*. (E.g., chemistry often takes for granted some physics, plus the reliability of certain tools and methods, etc.)
- (B) Observation Record or Data Set—recorded observations about *X*.
- (C) One or more Predictions/Explanations about *X*—inductive and abductive conclusions meant to be supported by (A) and (B).¹

A scientific theory of *X* is **a good theory to the degree that:**²

- (1) The theories in (A) are **widely accepted** in the field.
- (2) (B) includes only **verifiable observations** and yet is **comprehensive** (i.e., it includes *all* verifiable observations that are relevant to *X*). Any experiments used in generating the data set are **well-designed experiments**.
- (3) Assuming (A) and (B): The predictions/explanations in (C) are **more likely than any competitor** predictions/explanations.
- (4) The theory is **internally coherent** and **exactly formulated** (e.g., mathematical precision is used when possible).
- (5) The theory exhibits **pragmatic virtues** (see below).

¹ To avoid complications, I will also assume that a “scientific theory” is deductively closed: It is assumed to include all the deductive consequences of (A), (B), and (C).

² Each of (1)-(5) are not equally important to the worth of a theory. E.g., well-designed experiments are more important than fruitfulness (a pragmatic virtue). Yet it is unclear how much weight each of (1)-(5) should have.

Judging (1)-(5) is aided by **further guidelines**:

Re: (2). There are many types of well-designed experiment, but in medicine and in social science, a prominent example is a **randomized controlled trial** (RCT) with an **appropriate sample size**, subjected to much **peer review**. Where applicable, an RCT should also be **double blind** and **placebo controlled** (see Forseman et al., ch. 6).³

Further, data should be **unambiguous** (no fuzzy pictures of bigfoot, etc.). Data-collection and experimental procedures should be **well documented**. Also, **flawed data should be eliminated** (e.g., from faulty data collection). This is to ensure that the experiments are **replicable**, both in execution and in results.

Insofar as possible, **anomalies should be accounted for**. Plus, **limitations should be recognized**, including the limits on any retrospective studies, in vitro studies, case studies & anecdotal evidence, or non-human animal studies (see Forseman et al., ch. 9).

Re: (3). Correlations that justify predictions/explanations should be **statistically significant** ($p\text{-value} \leq .05$).⁴ (See the handout on significance testing.) Causal hypotheses can be evaluated using the **Bradford Hill criteria** (see separate handout).

Re: (5). The theory should be **falsifiable** and should exhibit, as much as possible, the following virtues (cf. Forseman et al, ch. 9):⁵

- a. **Predictive power**: The theory reliably predicts phenomena concerning *X*.
- b. **Explanatory scope**: The theory comprehensively explains phenomena concerning *X*.
- c. **Coherence with other accepted theories**: The theory is consistent with other theories we accept, and some of its elements are predicted/explained by those other theories (or vice-versa).
- d. **Simplicity**: The theory does not posit more entities than necessary to support its predictions/explanations. The theory is not needlessly conceptually complicated.
- e. **Fruitfulness**: Where applicable, the theory generates new, compelling hypotheses for investigation.

Sometimes there must be “trade offs” between these virtues. For example, an increase in explanatory power often requires a decrease in simplicity. But the trade may be worthwhile, depending on our goals and interests.

Note well: An unjustified theory according to (1) – (5) is not necessarily *false*. But it is a theory which is lacking in scientific justification.

³ There are additional criteria for a well-designed experiment. For social and health sciences, criteria for “internal validity” are described well at <https://psych.athabasca.ca/open/validity/>. (The website also has a useful self-test to ensure you understand the criteria.)

⁴ The *p*-value is the probability that an effect is due simply to random chance. Standardly, a *p*-value of less than or equal to .05 is necessary for ruling out random chance, but it is not sufficient. The so-called “replication crisis” is partly due to assuming otherwise. On the insufficiency of a low *p*-value, I highly recommend the following interactive website: <https://projects.fivethirtyeight.com/p-hacking/>.

⁵ Forseman et al. include “replicability” on their list of virtues, but I included it earlier as a virtue of an experiment rather than a theory. Also, in a – e I am only concerned with the “pragmatic” virtues of a theory.

Signs of a Hack

You can identify unjustified theories according to the above guidelines, but **advocates** of unjustified theories can often be identified by the following vices:

1. They advocate their views **not in scientific journals** but rather through the media and popular culture.
2. They **resist independent testing** of their views.
3. They are prone to **fallacies**, especially the appeal to authority and/or tradition, and do not adhere to the pragmatic virtues:
 - i. Their authorities are often believed to have “special” abilities for knowing the truth; others must simply defer to their judgment.
 - ii. Unfalsifiable or overly complex conspiracy theories are offered to explain why their views aren’t more widely accepted.
4. They exhibit **biases**, especially apophenia, the overconfidence effect, and confirmation bias (cherry-picking favorable evidence, ignoring unfavorable evidence).
5. They emphasize **low-quality evidence** among their favorable evidence, e.g., ambiguous and anecdotal evidence, poorly designed and/or non-repeatable experiments.

A way to sum up 2 – 5 is: A hack is **dogmatic**.

Unfortunately, some legitimate scientists can exhibit dogmatism as well... So the above signs are not infallible indicators of a hack, but they should raise suspicion. (Though again, an unjustified theory is not necessarily false, even if peddled by a dogmatist.)