



Methodological Function of Hypotheses in Science: Old Ideas in New Cloth

Krzysztof Lastowski and Wojciech Makalowski

Genome Res. 2000 10: 273-274

Access the most recent version at doi:[10.1101/gr.10.3.273](https://doi.org/10.1101/gr.10.3.273)

References This article cites 1 articles, 1 of which can be accessed free at:
<http://genome.cshlp.org/content/10/3/273.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

A horizontal banner advertisement with a teal background. On the left, the text "CRISPR and RNAi Genetic Screening. Your new superpower." is written in white. In the center, there is a white rectangular button with the text "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero cape and mask, with a green molecular structure logo and the word "CELLECTA" below it.

CRISPR and RNAi Genetic Screening.
Your new superpower.

LEARN MORE

CELLECTA

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Methodological Function of Hypotheses in Science: Old Ideas in New Cloth

Krzysztof Łastowski¹ and Wojciech Makalowski^{2,3}

¹*Institute of Philosophy, Adam Mickiewicz University, Poznań, Poland;* ²*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894 USA*

The August 1999 issue of *Genome Research* presented a very interesting editorial, "Hypothesis-Limited Research," on the status and discovery function of scientific hypotheses in disciplines generating large amounts of data (Goodman 1999). Dr. Goodman points to two types of limitations generated by hypotheses—one theoretical and one pragmatic. She argues that generating at least some type of data, for example, large-scale sequencing, does not require hypotheses. From a practical point of view, a hypothesis often complicates the scientific discovery flow, generating a hypothesis is time consuming and biases data interpretation. Consequently, Goodman urges giving up on proposing theory first and collecting data afterward. She supports her suggestion with historical examples, presenting great scientific discoveries that were not based on a hypothesis, although contrary interpretation has been presented by others (Łastowski 1996). Let us have a look at how heretical Goodman's proposal is, if at all.

A Critique of Goodman's Standpoint

From Goodman's argumentation the conclusion can be drawn that molecular biology is the first experimental scientific discipline in which accumulation of data is so dynamic and fast that researchers cannot follow it by formulating an adequate number of hypotheses. This statement does not decrease the value of hypotheses in science but, as we will show here, it increases it.

From a methodological point of

view, Goodman argues against a hypothetical program of research (Popper 1959). This program suggests the following procedure of scientific discovery:

[H] A hypothesis—Experimental data—Falsification
tests → *Refutation of the hypothesis*

where—reads *follows to* and → reads *implies*.

L. Goodman replaces this procedure with another one. It seems that (according to her) the scientific discovery in molecular biology should follow the scheme:

[G] Experimental data—Hypotheses—Pattern's discovery—Tests supporting hypotheses (verification) → Pattern's acceptance

The main argument for the scheme [G] is the fact that the pace of data accumulation is much greater than the generation of the hypothesis. Another important fact is that data are collected with minimal cost. Following procedure [G] we avoid procedure [H], which significantly limits, according to Goodman, the efficiency of scientific discovery in molecular biology. In exchange we gain a greater chance of the pattern's discovery.

Let us note that this is not a new idea; it resembles another scheme that was suggested quite long ago. It belongs to the positivist tradition of science and was developed within *logical empiricism* (see Hempel 1966). It can be presented as follows:

[P] Experimental data (facts)—Empirical hypotheses—Empirical pattern's

discovery—Tests verifying patterns → Pattern's acceptance

[G] is very similar to [P]. In both, the scientific procedure begins with facts (experimental data), and both accept the necessity of verification of discovered patterns by confirmation of data. The degree of hypotheses' confirmation increases with the amount of accumulated data.

In light of the juxtaposition of formula [G] and [P], the key methodological problem in Goodman's argument is the meaning of hypotheses and their role in scientific discovery. The question is how to reconcile the heuristic function of hypothesis with the demonstration of the hypothesis' correctness by identifying tests that falsify the hypothesis. If the test verifies the hypothesis, we assume that we can accept it as empirically justified.

Goodman suggests that the experimental data are a starting point in molecular biology research. This is true in many fields of science. Only in the next step do scientists propose hypotheses that then are checked by falsification tests. In this procedure hypotheses are based on "solid" facts and then in the empirical test undergo a falsification, not a verification. This procedure is not described by scheme [H], in which a hypothesis is a starting point, nor by scheme [G] in which data are a starting point but facts verifying a hypothesis are a final result, nor by scheme [P] in which facts are both the alpha and omega point.

Consequently, a new problem appears, which we call the dilemma of scientific hypotheses' discovery validity [D]. The dilemma is formulated as fol-

³**Corresponding author.**
E-MAIL makalowski@ncbi.nlm.nih.gov; FAX (301) 480-9241.

lows: What is the role of a hypothesis' explanatory status in relation to the usage of observations (i.e., facts) in falsification of the hypothesis?

Proposed Solution to Dilemma [D]

This dilemma [D] cannot be solved within the philosophy of science presented by Goodman. However, it can be addressed within the Idealization Theory of Science (ITS) (Nowak 1977, 1981). According to ITS, research in the experimental sciences proceeds according to the following scheme:

[ITS] Experimental data—Assumptions hypotheses understood as data selection filters → Scientific discovery—Tests falsifying hypotheses; (a) If test falsifies a hypothesis → Rejection of a hypothesis; (b) If test does not falsify a hypothesis → Temporary acceptance of a hypothesis

In the scheme above, the Assumptions hypotheses step is understood as theoretical filtering for data selection. This step assesses the facts, pointing out those that are explainable and those that are not. The latter are called anomalies. According to [ITS], a hypothesis is a description of relations simplified by theoretical assumptions. The function of these simplifications is the identification of *nonessential* features associated with the investigated phenomena or objects, so that they may be disregarded. As a result, a hypothetical description of the *most important* relations within the investigated phenomena is created. The role of facts is to show which of the proposed simplified descriptions are valid and which are not. From this point on, [ITS] is identical to Popper's (1959) scheme, that is, it states that researchers are focused on facts that falsify a hypothesis; therefore, they are focused on facts that are parts of negative tests.

In contradiction to this, Goodman interprets scientific hypothesis quite narrowly, that is; as a conjectural description of empirical regularities in data. In this meaning of hypothesis, regardless of how numerous the accumulated facts, they always are interpreted as

components of a hypothesis' positive test.

In light of [ITS] it is clear that (1) researchers do not have a single but, rather, many different descriptions of relations; (2) researchers, depending on their theoretical knowledge, can provide different descriptions (hypotheses); and (3) facts (experimental data) alone do not imply anything about their explanatory status. On the basis of their hypotheses, researchers decide how to relate them to the facts. Therefore, according to statements 1, 2, and 3, those theorized descriptions should be treated as theoretical filters for experimental data selection.

Consequently, we can conclude that in science, from a certain set of possible theoretical descriptions (hypotheses) one is chosen that is the most adequate according to objective criteria. How can one choose the best hypothesis? Falsification criteria should be applied to all competing hypotheses, and one that fulfills the following two conditions should be selected: (1) The hypothesis is not falsified by the test(s) used, and (2) it offers the largest domain of explanation—has the highest explanatory power. That is, the hypothesis' basic function is to offer the best understanding of facts. Facts, but not all of them, decide about the rightfulness of hypotheses. A "good" hypothesis is one that clearly proposes a new interpretation of facts and consequently can be easily falsified.

In contrast to Goodman's view, we conclude that hypotheses are not limitations to science. Quite the contrary: Facts decide which hypotheses survive and which die; ergo, facts are limitations to hypotheses not the other way around. Survival of a hypothesis that does not conform to the facts is simply an "adjustment" to the specific set of facts that are explained by the hypothesis. A hypothesis' natural death is the lack of such an adjustment or, in the extreme case, empty explanatory space. Therefore, contrary to Goodman's assertion, research is not limited by hypotheses but, rather, experimental data (facts) decide if our hypotheses possess

the required explanatory power. The more numerous and heterogeneous the sets of facts are, the more interesting the hypotheses, and the better the chance for an interesting discovery.

Editor's Response

The Editorial published in the August issue of Genome Research (Goodman 1999) was meant only to focus on the current perception in laboratories and especially funding agencies that all scientific studies should begin first with a hypothesis. (It is agreed that this is not a new idea; however, current policies and teaching have tended to cement hypothesis-first into current practice.) In this regard, in focusing on another point, the Editorial did not appropriately indicate how hypotheses should be handled after they are made and erroneously spoke of "proving" a hypothesis. Thus, the authors above make an extremely important point; one that is wholly agreed with here. Unlike mathematical theorems, hypotheses in other scientific fields, especially molecular biology, cannot be proven but can only be accepted as true until proven false. In this regard, the best hypotheses (no matter what stage in the scientific process they are generated) are those that can be best tested for falseness. One additional note, the method for scientific procedure proposed in the Editorial was meant to be clearly indicated in the following order: [G] experimental data—Pattern's discovery—THEN Hypotheses—after which hypotheses undergo appropriate testing. For hypothesis-free data collection, testable hypotheses should not need to be formed until general patterns are noted from the data.

Laurie Goodman

REFERENCES

- Goodman, L. 1999. *Genome Res.* **9**: 673–674.
 Hempel, C. 1966. *Philosophy of natural sciences*. Prentice Hall, Englewood Cliffs, NJ.
 Łastowski K. 1996. In *New trends in molecular biology, genetic engineering, and medicine*. (English translation), pp. 9–26. Sorus, Poznań, Poland.
 Nowak, L. 1977. *Introduction to idealizational theory of science*. (English translation). PWN, Warszawa, Poland.
 Nowak, L. 1981. *The structure of idealization*. Reidel, Boston, MA.
 Popper, K.R. 1959. *The logic of scientific discovery*. Basic Books, New York, NY.