Hypotheses should Better be Well-Founded and Not Just Testable

Hartmut Traunmüller
Department of Linguistics, Stockholm University

Abstract

This is a contribution about the scientific method – about falsificationism and its non-applicability in phonetics and the life sciences in general – about the advantage of distinguishing between well-founded, provisional and fictitious hypotheses and the a priori confidence we can have in these types, marginally also about the principle of parsimony – and about path-dependence and the lock-in effect of "normal science".

Introduction

I have, for several years, been teaching the scientific method to small groups of students of phonetics. This was mainly about how to test hypotheses, about statistical methods, and about various interfering factors that are often met in experiments, especially with humans. Knowledge of these methods and factors is certainly essential. However, it does not offer the students any guidance in choosing their hypotheses.

During the past century, the basic question of the testability of hypotheses or theories by observations has been much talked about by philosophers of science. Popper's (1935) teaching, inspired by Einstein, that theories and 'laws of nature' cannot be verified (by testing them even repeatedly) but only falsified (even by a single test) has come to be widely respected, although talk about 'proving' theories is still common. Meanwhile, it has become clear that even falsifications may not be definitive, and it appears as a weakness that they provide no guidance for how to proceed. It is not enough to require of all assumptions to be testable and simple. They should also be as well-founded as possible, and, in a given theory, they must be compatible with each other.

My present contribution is inspired by my criticism of standard cosmology (Traunmüller, 2018), in which a lack of well-foundedness and compatibility problems stand out prominently. In this field, and in the basic sciences like physics and chemistry, falsificationism is widely applicable, while this is not so in phonetics and in the life sciences in general, in which variation between and within subjects must be allowed for. In these cases, less categorical statistical methods have to be used (Dolby, 1982), but the impossibility to 'prove' theories remains and also the advantage of well-founded hypotheses over questionable and fictitious ones.

Empirical science

Empirical science involves acquiring knowledge with an aim to organize, explain and understand phenomena. Among this knowledge, the 'existential' (about what there is) and the 'universal' (such as so called 'laws of nature') stand out. Universal knowledge, which is most prominent in Popper's (1935) philosophy of science, can be thought of as *a set of empirically testable universal statements that have not yet been convincingly falsified* and so remain *tenable* (at our present state of knowledge). Science makes also *existential statements*. These can only be *verified* rather than falsified empirically. In any case, alleged laws of nature can only be claimed to be *tenable* rather than "*true*". This holds even if Popper's falsificationism does not apply.

In this conception of science, it is not necessary for postulates and hypotheses, which give rise to the statements, to be understood. It suffices for them to be *tenable* given the empirical evidence. However, it can be argued that it is the ultimate aim of the basic sciences to extend the body of empirical knowledge that can be rationally explained without reliance on any assumption that is not understood. Questionable assumptions still have a function in science, but they remain tentative and provisional until they are shown to be either untenable or redundant, i.e. predictable within a wider frame. There is no better fate for a questionable assumption.

In order to really *understand* phenomena and the relations between these, we need theories that rest on a foundation of solid knowledge. This may involve other well-founded, more fundamental theories. The most well-founded theories are based solely on *definitions* and *first principles* of the kind that cannot be easily rejected using the Cartesian method of doubt. These are principles that are accepted even outside the frame of the particular theory.

An 'axiom' or 'postulate' does not necessarily qualify as a first principle in the sense adopted here, but indispensable axioms whose validity is independent of nature lie at the foundation of the formal sciences. These give us the rules of logic, algebra and geometry, which then can be taken as first principles in all sciences. A theory that builds on a postulate that is nowhere necessary outside the theory it serves is not well-founded in our sense, but speculative, conditional and provisional. This remains so even if its predictions are compatible with all empirical evidence, no matter how accurately. It will remain 'just a theory' even if 'corroborated' by evidence. While many theories are of this kind, there are also well-founded *ab initio* approaches at least in some fields.

One might, perhaps, expect *ab initio* approaches to be most common in theoretical physics, but it appears as if physicists did not always find enough to build on for proper *ab initio* approaches. Such approaches are actually most common in chemistry (e.g. ab initio quantum chemistry, ab initio molecular dynamics).

In particular in the life sciences, the scope of deductive approaches is limited to sub-problems unless a statistical distribution of observables is operated with instead of single values. In phonetics, the acoustic theory of speech production describes how the formant frequencies depend on the shape of the vocal tract. This is a physical-geometrical problem, for which an ab initio approach can be attempted. Models of glottal flow, e.g., (Fant, Liljencrants & Lin, 1985), appear already more remote from ab initio approaches. Here, the aim is primarily to make it possible to model observed flow patterns at all, with the minimal number of necessary parameters.

Research in speech perception can make use of well-founded and testable assumptions rooted in the wider field of psychoacoustics, but strict falsificationism is no longer applicable there.

In addition to well-founded assumptions, one persistently retains in many fields of science also traditional assumptions and standard paradigms that may fall short of satisfying the criteria of well-foundedness. Outside science, inferior paradigms and standards can also persist because of the legacy they have built up, like the QWERTY layout in typewriters (David, 1985). This exemplifies "path dependence", and there are cases in the history, teaching and practice of science in which a traditional path showed itself to be an impasse. This had, in effect, already been noticed by Kuhn (1962) in his study of scientific practice, but the undesirable "lock-in effects" of path dependence have still not found the attention they require (Jolink & Vromen, 2001, Peakock, 2009) in science.

The history of science shows us that traditional assumptions tend to be retained not only as long as they remain compatible with the empirical evidence but as long as they can be made compatible with it by ad hoc means. Falsifications are thus often fudged away by excuses in the form of ad hoc assumptions and constructs, also purely imaginary ones. In standard cosmology, dark matter is an example and dark energy is a purely imaginary one (Merritt, 2017, Traunmüller, 2018). Such adherence to and protection of traditional paradigms is in fact characteristic of what Kuhn (1962) called "normal science" as opposed to "revolutionary science" and of what Lakatos (1976) called a "research programme" with its "irrefutable hard core".

Reliability check

Among scientific approaches to natural phenomena one can distinguish between inductive, phenomenological ones, which are founded on empirical observations, and deductive ones, founded on theoretical premises. There is often interplay between these. Definitions are essential in both cases.

In purely inductive approaches, regularities among observations are searched and described without offering an explanation. They yield organized particular knowledge, empirical relationships, and superficial or probabilistic understanding. This is always feasible, whether or not falsificationism would apply in a corresponding deductive approach.

Studies in phonetics, as in the life sciences in general, are often inductive. Notable philosophers of science have denied such studies the status of being "scientific". They consider deductive reasoning as essential, but most scientists disagree. Systematic descriptions of experiences undoubtedly fall within the common definition of science as "the intellectual and practical activity encompassing the systematic study of the structure and behaviour of the physical and natural world through observation and experiment" (Oxford English dictionary).

All inductive modelling (curve fitting) requires an assumption about the kind of relation that might be present. The linear relations that are often tested by default are of limited use. When I worked with my thesis, around 1980, I made much use of the psychoacoustic critical band rate scale, which was specified in form of a table. There were also several equations published for it, but none of them was good enough. Finally, I found a better one myself by considering critical band rate z to be related to log(f) by a logistic function (a sigmoid curve),

$$z = [26.81/(1+1960/f)] - 0.53, \quad (1)$$

with f in Hz (Traunmüller, 1990). In the curve-fitting procedure, an exponent was present, but this turned out not to be significantly different from 1. This led to a striking simplification of the equation, which was meant for practical purposes only, but is there any well-founded hypothesis from which this would follow?

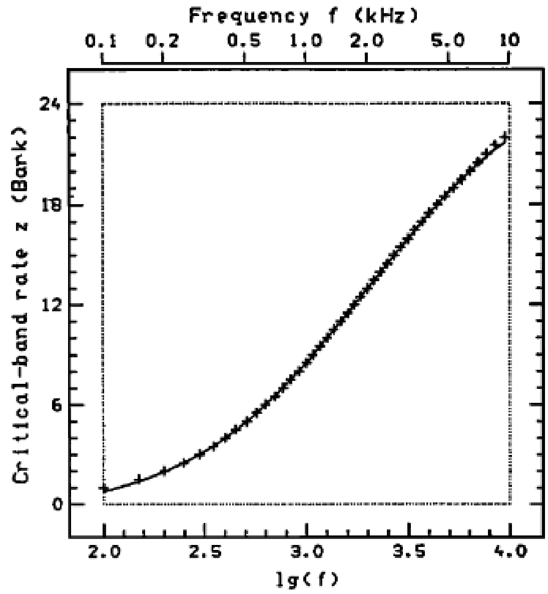

Figure 1. Critical-band rate z as a function of frequency f. Plus signs represent data from Zwicker (1961) and the curve represents equation (1).

Deductive approaches offer explanations of observations. They provide support for universal statements to the extent to which we can be confident in their premises. Here, three types can be distinguished (1, 2 and 3 in Table 1), which differ grossly in the confidence they lend.

1) First principles. If no other type is involved, these lead to well-founded theories and reliable predictions and to explanations that can be understood. Approaches that are founded on definitions and first principles alone embody the deepest understanding of phenomena.

2) Tentative assumptions or "postulates", which in some way appear reasonable but remain subject to doubt since they are not rooted outside the theory in question. These can never be proven within it. They lead to provisional (conditional) theories and to explanations that hold to the extent to which the assumptions hold. This is characteristic of the "hypothetico-deductive method", which fails to distinguish between the types 1, 2 and 3. Philosophers of science typically imagine all assumptions as fallible without distinction.

3) Assumptions that, in addition to not being rooted outside the theory in question, also lack independent empirical support. Any reasoning based on these remains within the domain of imagination. Such assumptions are fictitious and lead to epistemically unsupported beliefs. Modern theoretical physics offers a range of "fairy tale physics" (Baggott, 2013) in which fictitious assumptions

are either primary, as in string theory, or supplementary, as in the “dark sector” of big bang cosmology.

Table 1. Epistemologically different types of premises in deductive models, the confidence C these impart (a multiplicative variable), their type of adequacy and their epistemic yield.

Premise	Confidence	Adequacy	Epistemic yield
1 Well-founded hypothesis	$C = 1$	Descriptive and explanatory	Deep understanding
2 Provisional hypothesis	$0<C<1$	Descriptive and tentatively explanatory	Superficial understanding and uncertain deeper one
3 Fictitious hypothesis	$C = 0$	Formal	Epistemically unsupported belief

The values listed in Table 1 under “Confidence” express the a priori confidence we can have in the hypotheses and explanations these suggest. They depend on how well the hypotheses are rooted in what is already understood.

If we trust our prior knowledge, we can be fully confident if the hypotheses are well-founded (type 1). This can remain so ($C = 1$) even when we are confronted with discrepant empirical data. However, in such cases, there must be a factor that remained unaccounted for – which can often be expected in the life sciences.

If, on the other extreme, an entity or process is fictitious within the frame of our knowledge (type 3), the confidence it deserves, its explanatory power and its epistemic value cannot be asserted to be larger than zero ($C = 0$). This holds even if the approach leads to predictions that are compatible with the evidence, no matter how well.

In the provisional approaches, type 2, we have $0 < C < 1$. In these cases, a numerical rating of confidence that would be generally valid is not obvious, except at the level of rank order. It is, e.g., justified to attach more confidence to a reasoning based on a simple general assumption that has not been falsified than to a less general alternative that can be said to involve the same assumption under a restrictive condition that needs to be specified. The latter is equivalent to having an additional assumption, and the higher confidence in an approach that needs fewer assumptions reflects the principle of parsimony (Ockham’s razor), which applies here.

Sufficiently, even fully reliable predictions of entities that have never been observed are not precluded in this scheme. In order for us to be confident at $C > 0$ into their real existence, it is only required that $C > 0$ for each of the hypotheses on which the prediction is based.

While falsificationism is largely inapplicable within the life sciences, and proposed laws of nature cannot be verified but only falsified empirically, even falsifications are not firmly conclusive. A statement that stood falsified may become tenable again in the light of new knowledge. Universal statements can only be claimed “to be tenable” or “to stand falsified” (given our knowledge), unless it follows from definitions and logic alone that they are “true” or “false”.

The classification of a hypothesis or a presumed entity as fictitious ($C = 0$) might also change in the light of new knowledge, but as long as we lack this knowledge, our confidence in it must remain at zero if we wish to remain within empirical science. This remains so whether or not falsificationism applies.

When confronted with discrepant evidence, the descriptive adequacy of a theory can often be saved by introducing an ad hoc parameter or “fudge factor”. However, such an approach has no explanatory power. Worse yet, it invites circular reasoning, and if it represents a fictitious entity, the approach turns into one of type 3, which just yields a mere belief, e.g. in dark energy. It promotes ‘credence’ – not ‘science’.

Theories usually involve several assumptions, and they do not gain in confidence if the number of first principles they invoke (all with $C = 1$) is reduced. They gain in confidence if the number of tentative assumptions (all with $C < 1$) is reduced, provided that no fictitious assumption (with $C = 0$) is involved.

The types of assumptions in theories can be listed in the following order:

- definitions
- well-founded first principles
- generalizing assumptions
- more specific testable assumptions
- ass. involving fictitious entities or processes.

In order to obtain a clearly more well-founded theory, this list needs to be shortened from its end.

Consistency check

As compared with judging the reliability of hypotheses and the simplicity of the set of assumptions, it is of course an even more fundamental concern to check theories for self-consistency between all their premises – also between premises and conclusions from other premises. The reasoning should be free from conceptual, logical and mathematical errors and from crucial lacunae.

The normal science problem

The *reliability check*, inspired by Descartes, may suggest that certain assumptions should better be skipped. This, as well as even a *consistency check* can go against “normal science” or the “hard core” of a research program, in which scientists accept the conventional tenets of established theories without proper reflection.

Scientific journals often publish articles on speculative modifications of mainstream doctrines, but articles that discredit the “irrefutable hard core” in the respective research program run a very high risk of being rejected by referees and editors. These can easily identify deviations from the established doctrine and practice, while it requires a higher effort and self-conquest to follow and evaluate a path of reasoning that is not one’s customary one. Together with the similar disposition by teachers and grant providers, this leads to the perseverance of aberrations from the path to reliable knowledge in what Kuhn (1962) called “normal science” and Lakatos (1976) “research programmes”. The activities so labeled can widen our knowledge, but they cannot lead to a fundamental improvement in our *understanding* of nature. This would require a more critical attitude, but only adherence to traditional paradigms is safe for those who aim for or depend on positive judgments by teachers, editors, referees, or grant providers. Nevertheless, approaches that require approval of traditional fallacies and 'credence' in ad hoc assumptions can hardly be claimed to remain within the bounds of 'science'.

References

Baggott, J. (2013) *Farewell to Reality: How Modern Physics Has Betrayed the Search for Scientific Truth*, Pegasus Books, New York.

David, P. A. (1985) Clio and the economics of QWERTY, *Am. Econ. Rev. 75*, 332.

Dolby, G. R. (1982) The role of statistics in the methodology of the life sciences, *Biometrics 38*, 1069-1083.

Fant, G., Liljencrants, J. & Lin, Q. (1985) A four-parameter model of glottal flow, *STL-QPSR 26* (4), 1-13.

Jolink, A. & Vromen, J. (2001) Path dependence in scientific evolution, in: *Evolution and Path Dependence in Economic Ideas* (Eds. P. Garrouste & S. Ioannides), Elgar, Cheltenham, p. 205.

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

Lakatos, I. (1976) Falsification and the methodology of scientific research programmes, in: *Can Theories be Refuted?* (Ed. S. G. Harding), Springer, Dordrecht, pp. 205-259.

Merritt, D. (2017) Cosmology and convention, Studies in *History and Philosophy of Modern Physics 57*, 41.

Peacock, M. S. (2009) Path dependence in the production of scientific knowledge, *Social Epistemology 23*, 105.

Popper, K. (1935) *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft*, Springer, Wien. [English version *The Logic of Scientific Discovery*].

Traunmüller, H. (1990) Analytical expressions for the tonotopic sensory scale, *Journal of the Acoustical Society of America 88*, 97–100.

Traunmüller, H. (2018) Towards a more well-founded cosmology, *Zeitschrift für Naturforschung A 73*, 1005-1023.

Zwicker, E. (1961) Subdivision of the audible frequency range into critical bands, *Journal of the Acoustical Society of America 33*, 248.