Is Bayesian epistemology a better alternative to Popperian falsifiability?

By Victor Christianto, email: victorchristianto@gmail.com

Abstract

According to a site (see http://psychology.wikia.com/wiki/falsifiability), there are many people who reject Popperian falsifiability, such as Paul Feyerabend and Alan Sokal. There is another two-fold critique on falsificationism by Shockley, see http://pk.b5z.net/i/u/2167316/i/A_Critique_of_Falsificationism_by_Karl_Popper.pdf

Introduction

According to a site (see http://psychology.wikia.com/wiki/falsifiability), there are many people who reject Popperian falsifiability, such as Paul Feyerabend and Alan Sokal. There is another two-fold critique on falsificationism by Shockley, see http://pk.b5z.net/i/u/2167316/i/A_Critique_of_Falsificationism_by_Karl_Popper.pdf

From reading this criticism, perhaps we can propose a new idea, that is: Bayesian acceptance of a theory. The idea stems from Bayesian epistemology (see http://plato.stanford.edu/entries/epistemology-bayesian/). The basic idea here is that each time an observation/data supports a theory then it has greater acceptance or probability nearer to 1.

For example, if each morning we observe that the sun always rises in the morning, and this phenomenon has been observed since thousand years ago, then we have probability of almost 1 that the sun will also rise tomorrow morning.

On the contrary, if an observation disproves a theory, then the acceptance chance of that theory will diminish or approaching zero, although there are still supporters of it.

The proposed Bayesian acceptance of a theory may be applied to astrophysics, cosmology, biology, chemistry, and experimental physics as well. See also John Harsanyi's paper discussing acceptance of empirical hypothesis: http://www.jstor.org/discover/10.2307/20115945?uid=3738224&uid=2460338175&uid=2460337935&uid=2&uid=4&uid=83&uid=63&sid=21103914255951

Answers:

[1] **Karen Gold**

Can you give an example of how one would conduct a placebo-controlled drug trial under a strictly Bayesian rubric? I cannot. While there are difficulties in a strict positivist framing of inquiry, the the quantification framework of Popper's falsificationism is remarkably elegant and clean. I have seen hybrid frameworks where a Bayesian prior has been added to account for knowledge aquired prior to an experiment (in the Popper falsification framework of rejecting a null in favor or an alternative hypothesis) but I have not seen one which altogether disgards this sort of test and enables drawing an applicable conclusion. Theories are great, but without an appropiate armory of machinery, they have no teeth.

[2] **Victor Christianto**

Dear Karen, thank you for your question. I am no specialist in drug trial, i just try to argue on how to improve falsifiability without recourse to anarchism view of science (Feyerabend). For drug clinical trial, perhaps you will find this article in Nature,2006, useful as guide. It was written by Donald A. Berry (see http://www.nature.com/nrd/journal/v5/n1/abs/nrd1927.html). His email is dberry@mdanderson.org.

Other than Bayesian acceptance of a theory, another alternative is perhaps SOC model of scientific discovery (Soc means self organized criticality). But so far, i do not find any reference on this proposal. Best wishes

[3] **Jochen Wilhelm**

Karen, you state that Popper's falsificationism is remarkably elegant and clean. In terms of Boolean logic it is. However, I do not see how it is applicable in scenarios with incomplete knowledge. Data can be extremely unlikely under H0, but this does not mean that H0 is falsified by the data. Even more severe, although the data may be so unlikely under H0 that H0 will be rejected, it might be even less likely under a reasonable alternative. I do not see how Popperian philosophy of falsification and probabalistic methods (not sure if this is the correct term, but I think you'll know what I mean) are compatible. Looking forward reading your comments on this :)

[4] **Andrew Messing**

Dear Victor Christianto:

You raise some interesting issues, but alas they touch upon a number of some of the most debated matters in the philosophy of science, the philosophy of probability, epistemology, and scientific methods. I would find it helpful if you could clarify certain comparisons and distinctions you seem to have made or not made that I find rather central.

For example, you set up something of a dichotomy between Popperian epistemology and Bayesian epistemology. But subjective interpretations of probability aren't limited to Bayesian, and the opposing approach to Bayesian analysis isn't really any philosophy of science per se but rather the frequentist philosophical interpretation of probability. This is not to say that Bayesian epistemology is compatible with Popperian philosophy of science. It is simply to point out that objective interpretations of probability and subjective interpretations are perhaps better contrasting epistemological approaches than comparing an extension of conditional probability from statistical inference to a formalized scientific epistemology over and against a rather specific, detailed, and singular philosophy of science.

Likewise, I find it a bit confusing to refer to Feyerabend's views as critical of Popper's, which of course it is, but seemingly gloss over how the same is just as incompatible (or nearly so) with Bayesian epistemology. On a related note, the primary opposition to Popper's philosophy of science comes from not from Bayesian epistemology but the Quine-Duhem thesis (although for a critical review on the appropriateness of this nomenclature, see Zammito's A Nice Derangement of Epistemes, pp. 17-25). For a brief but relatively comprehensive treatment on the changes to the philosophy of sciences originating around the beginning of the 20th century, see e.g., Hrheinberger's Historische Epistemologie (also, Zammito's text cited above is a good source, albeit ~200 pages longer). In general, it seems as if your presentation of Popper on the one hand and Bayesian epistemology on the other seems to miss the rather important differences between what each were intended for as well as the other, more directly comparable approaches to the philosophy of science and/or epistemology (including the philosophy of probability).

As for falsification, perhaps THE method for hypothesis testing is the ritualistic null hypothesis significance testing (NHST). This is the standard "reject the null, accept the alternative" (given some alpha level) used in the medical, psychological, social, and other sciences. I hate to rehash material I've already repeated here, but as I took some time to find a short but comprehensive set of resources, I will anyway:

"One wishes to know the probability that a biological or medical hypothesis, H, is true in view of the sadly incomplete facts of the world...But the statistical tests used in many sciences (though not much in chemistry or physics) do nothing to aid such judgments. The tests that were regularized or invented in the 1920s by the great statistician and geneticist Ronald A. Fisher (1890-1962) measure the probability that the facts you are examining will occur assuming that the hypothesis is true....The mistake here is known in statistical logic as "the fallacy of the transposed conditional." If cholera is caused not by polluted drinking water but by bad air, then economically poor areas with rotting garbage and open sewers will have large amounts of cholera. They do. So cholera is caused by bad air. If cholera is caused by person-to-person contagion, then cholera cases will often be neighbors. They are. So cholera is caused by person-to-person contact. Thus Fisherian science." McCloskey, D. N., & Ziliak, S. T. (2009). The Unreasonable Ineffectiveness of Fisherian" Tests" in Biology, and Especially in Medicine. Biological Theory, 4(1), 44. ([http://stephentziliak.com/doc/Transposed%20conditionals%20in%20Biology%20and%20Medicine.pdf)](http://stephentziliak.com/doc/Transposed%20conditionals%20in%20Biology%20and%20Medicine.pdf)

"In a recent article, Armstrong (2007) points out that, contrary to popular belief, "there is no empirical evidence supporting the use of statistical significance tests. Despite repeated calls for evidence, no one has shown that the applications of tests of statistical significance improve decision making or advance scientific knowledge" (p. 335). He is by no means alone in arguing this. Many prominent researchers have now for decades protested NHST, arguing that it often results in the publication of peer-reviewed and journal endorsed pseudo-science. Indeed, this history of criticism now extends back more than 90 years (e.g., Armstrong, 2007; Bakan, 1966, 1974; Berkson, 1938; Boring, 1919; Campbell, 1982; Carver, 1978, 1993; Cohen, 1990, 1994; Edwards, 1965; Falk, 1998; Fidler, Thomason, Cumming, Finch, & Leeman, 2004; Fisher, 1955, 1956; Gigerenzer, 1987, 1993, 2004; Gigerenzer et al., 1989; Gill, 1999; Granaas, 2002; Greenwald, 1975; Hubbard & Armstrong, 2006; Hubbard & Lindsay, 2008; Hubbard & Ryan, 2000; Jones & Tukey, 2000; Kirk, 1996, 2003; Lindsay, 1995; Lykken, 1968, 1991; Meehl, 1967, 1978, 1990; Nunnally, 1960; Rosnow & Rosenthal, 1989; Rozeboom, 1960; Schmidt, 1992, 1996; Schmidt & Hunter, 1997; Sedlmeier & Gigerenzer, 1989; Skinner, 1972; Thompson, 1996, 1999, 2002, 2006, 2007; Tukey, 1991)…It is certainly a "significant" problem for the social sciences that significance tests do not actually tell researchers what the overwhelming majority of them think they do (Bakan, 1966). Bakan thought that "everybody knows this" and that to say it out loud is to be like the child who pointed out that the emperor is wearing no clothes. He argues that if we pull out the strand of NHST much of the tapestry of psychology would fall apart. Indeed, NHST, Gerrig and Zimbardo (2002) state, is the "backbone of psychological research" (p. 46). So, instead of abandoning it, which could be very embarrassing, we make the adjustment of simply misinterpreting what it actually tells us, which is…not much." Lambdin, C. (2012). Significance tests as sorcery: Science is empirical—significance tests are not. Theory & Psychology, 22(1), 67-90. ([http://psychology.okstate.edu/faculty/jgrice/psyc5314/SignificanceSorceryLambdin2012.pdf)](http://psychology.okstate.edu/faculty/jgrice/psyc5314/SignificanceSorceryLambdin2012.pdf)

"402 Citations Questioning the Indiscriminate Use of Null Hypothesis Significance Tests in Observational Studies"
([http://warnercnr.colostate.edu/~anderson/thompson1.html)](http://warnercnr.colostate.edu/~anderson/thompson1.html)

"The Null Ritual: What You Always Wanted to Know About Significance Testing but Were Afraid to Ask"
([http://www.sozialpsychologie.uni-frankfurt.de/wp-content/uploads/2010/09/GG_Null_20042.pdf)](http://www.sozialpsychologie.uni-frankfurt.de/wp-content/uploads/2010/09/GG_Null_20042.pdf)

Gigerenzer, G. (2004). Mindless statistics. The Journal of Socio-Economics, 33(5), 587-606.
([http://people.umass.edu/~bioep740/yr2009/topics/Gigerenzer-jSoc-Econ-1994.pdf)](http://people.umass.edu/~bioep740/yr2009/topics/Gigerenzer-jSoc-Econ-1994.pdf)

At any rate, thanks for broaching the topic. I'm looking forward to others' input (particularly given the two experts who have already contributed and will, I hope, continue to).

[5] **Jochen Wilhelm**

From Lambdin (p. 69): "Social scientists, he argues, all too commonly employ a jargonized statistical analysis to smokescreen the seriously flawed reasoning underpinning their conclusions. He calls this practice "quantification as camouflage." This had gotten so bad, Andreski thought, that the "quantophrenics" in psychology should take a break from learning statistical technique to study some elementary logic and analytic philosophy."

I strongly support this!

However, Lambdin misses to correctly define or distinguish "chance" and "probability" (p. 75): "A pvalue does not and cannot tell you the odds your results are due to chance, because it is calculated based on the assumption that this probability is already 100%. In other words, a pvalue tells you the odds of your results given that you assume they are in fact due to chance. This brings back to mind the Andreskian admonition that less statistics classes and more courses in simple analytic philosophy are in order."

What is "due to chance"? What is "probability"? What is a "probability of something being due to chance"? This is a completely circular argumentation. If "chance" should not be used to open Pandorra's box on determinism, it can only be related to our own state of knowledge or certainty about events we may observe (or data we may measure). The very same, but more quantitatively elaborated, is meant by "probability". To my opinion, claiming that "chance" is some physical thing being out there in the world is not a penny better than claiming that a p-value is a measure for reliability, evidence, or support for a hypothesis. So I'd reformulate his sentence "Assuming my results are due to chance, my obtained mean difference is very unlikely. Therefore chance may not be the culprit." to "Under my assumptions I would not expect such (or more 'extreme') data. Therefore, my assumptions might not be appropriate"

Unfortunately, this problem is not even recognized as such from the APA (from p. 80): "Caution: Do not infer trends from data that fail by a small margin to meet the usual levels of significance. Such results are best interpreted as caused by chance and are best reported as such" (APA, 1974, p. 19)"

Am I wrong here? Is there a rationale to see "chance" as something physically existing? I tried a lot to find out, but I am always coming back to the conviction that "change" is only expressing a lack of precice knowledge. Maybe we might theoretically even have the knowledge but it would not be related to the factors we are interesetd in and/or would not be helpful for predictions.

Great, and so true (p. 85): "As Roseanne Conner's father says in the sitcom Roseanne, "You can lead a horse to water, but you can't make him think." Paraphrasing Fidler et al., (2004), "You can lead a social scientist to a confidence interval, but you can't make him think about statistics." " - I have seen scientists using confidence intervals simply to accept and reject H0s based on the fact whether or not the intervals included H0. Give a latest optical quantum-effects computer to a monkey, and he will likely just try to eat it :)

Great paper, fun to read.I only missed a more detailed discussion of points pro NHST.

[6] **Andrew Messing**
Dear Dr. Wilhem:

Regarding your (in my opinion correct) criticism that Lambdin does not sufficiently define "due to chance" (or chance), and in particular does not define it so that it may be properly distinguished from probability: I had overlooked this, as I am so used to thinking of "due to chance" and "probability" in terms of NHST where "due to chance" is defined in terms of a failure to meet some alpha level (or p-value, the term Lambdin does). The logic of NHST is to assume that, given some probability distribution (usually approximately normal), there is some outcome that is sufficiently far from the expected value that we can assert it isn't "due to chance". I think Lambdin is expecting the reader to be familiar with the ways in which standard hypothesis testing methods are defined in terms of "chance" as outcomes that are less probable than some alpha level and are thus not distinct from probability but are defined by some interval within a probability distribution.

That said, I agree with your remarks, and thanks to your remarks I feel it necessary to add something to what I said. I was thoroughly indoctrinated to accept NHST, and have been required to teach it for years. I sometimes unconsciously assume, unfortunately, a stance against the kind of understanding of NHST I had years ago and my students have had- i.e., a limited understanding both of its underlying logic, its criticisms and defenses against these, and in general the kind of nuanced understanding of both the philosophy of probability and the logic of statistical inference. While this is in general appropriate when I deal with students, it isn't here. So I would welcome links and/or references to comprehensive defenses of NHST, as I have spent far too long trying to break the death-grip of its ritualistic use to concentrate on the counters to those arguments I use to challenge complacency.

[7] **Jochen Wilhelm**

It is in fact difficult to get a reference pro NHST. Most papers about that are logically flawed and do not provide a grounds for discussion. Good papers advocating significance tests or hypothesis tests are absed on a correct mathematical framework that is not neccesarily relevant for applied research.

I have some references that I find relatively ok (but none of them positively "solving" the problem that p-values and hypothesis tests are not sensible for standard applied research (industrial reasearch and screening technologies may be exceptional cases) :

Statistical Methods and Scientific Induction. Ronald Fisher. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 17, No. 1. (1955), 69-78.

Neyman J, Pearson E S 1928 On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika20A: 175–240, 263–94

Lehmann E L 1993 The Fisher, Neymann–Pearson theories of testing hypotheses: one theory or two?Journal of the American Statistical Association 88: 1242–9

International Encyclopedia of the Social & Behavioral Sciences, ISBN: 0-08-043076-7 (maybe especially the article "Hypothesis Testing: Methodology and Limitations" of T. A. B. Snijders, p 7121ff)

Hurlbert & Lombardi: Final collapse of the Neyman-Pearson decision theoretic framework and rise of the neoFisherian. Ann Zool. Fennici 46: 311-349.

Clarice R. Weinberg: It’s Time to Rehabilitate theP-Value. Epidemiology May 2001, Vol. 12 No. 3

No 2 Practice - What Is to Be Done with P Values? Epidemiology 2013, Vol. 24, and Theory David A. Savitz: Reconciling

[8] **Victor Christianto**

Dear Karen, Jochen, and Andrew. Thank you so much for all your answers. I need some time to digest all your deep comments, before making further remarks.
Allow me to say that i just got that idea of Bayesian acceptance of a theory yesterday, and then a quick search at bing.com reveals that only a few people brought this topic before, therefore i think it would be interesting to compare it vis a vis Popperian falsifiability, which some scientists have objection to. One of my reasoning to propose Bayesian epistemology is because Falsifiability is merely a general concept, while bayesian probability has strict definition.
Now i got another idea: provided that Bayesian epistemology is worth to consider, then perhaps we can extend it further to become Dempster-Shafer epistemology. That is because Dempster-Shafer theory is a generalization of Bayesian probability, see this article http://www.glennshafer.com/assets/downloads/articles/article48.pdf. So what do you think? Thanks and best wishes

[9] **Jochen Wilhelm**

Counter-question (without having read the book): Why should a theory be "fully accepted" at all? I do not see the point. A theory is set of models, and whatever applies to model does so for theories. A theory can be more or less supported by data, and some theories can even be falsified (in the strict Popperian sense).

[10] **Daniel Courgeau**

Dear Victor,

I entirely agree with Jochen and Andrew criticisms saying that Popperian philosophy and probabilistic theory are not directly compatible. In order to go further, I will show that the opposition between the more general *Popperian deductivism*, for which falsificationism is only a part, and the *Baconian inductivism*, is in fact the important dichotomy.
For Popper, who considers induction as inconsistent, the only way to scientific methodology is not to prove a hypothesis but only to show that we have the possibility to falsify it. This way of reasoning from a hypothesis to the facts implied by it, will lead to as much theories as there are hypotheses to verify, and will restrict scientific research to a task of validation. However, as Hume, he defines induction in an incorrect way, as a generalization from specific observations. In this case it is true that such a generalization may always be refuted by a new observation.
But for Bacon, induction has another meaning more thorough. For the previous meaning, he clearly said: “For the induction which proceeds by simple enumeration is childish; its conclusions are precarious and exposed to peril from a contradictory instance; and it generally decides on too small a number of facts, and on those only which are at hand.” (The new Organon, 1620, Book 1, CV). The second meaning, which had been used by Galileo, Descartes and Newton, is defined by Bacon as: “The other derives from the senses and particulars, rising by a gradual and unbroken ascent, so that it arrives at the most general axioms last of all. This is the true way, but as yet untried.” (The new Organon, 1620, Book 1, XIX). For him the axioms are not “self-evident and necessary truths”, as the

Webster's Dictionary said, but are drawn from experience. They are not everlasting as new experiences may lead to new axioms. They are principles inferred by experience. The way Newton found his laws, not from hypotheses but by induction in the Baconian meaning, is a perfect example of this way of reasoning. Nowadays this Baconian induction leads to the mechanistic theory applied as well in natural sciences , as in biological sciences or in social sciences (see for example: Franck ed., 2002, The explanatory power of models, or Craver and Bechtel, 2006, Mechanisms, in The philosophy of science: an Encyclopedia, p. 469-479). Looking forward for your comments on this point.

[11] **Victor Christianto**

Dear Daniel, thank you for your answer. I think that both deduction and induction are both required for a proper scientific method. The question is how. Best wishes

[12] **Daniel Courgeau**

Dear Victor,

I don't think that deduction and induction are both required for a proper scientific method. As Sir Harold Jeffreys said in his *Theory of probability* (1939): "Running through the whole is the tendency to claim that scientific method can be reduced in some way to deductive logic, which is he most fundamental fallacy of all: it can be done only by rejecting its chief nature, induction." This induction is not, as I previously said, a generalisation from particular instances, but "consists in discovering a system's principle from a study of its properties, by way of experiment and observation", as Franck said in his 2002 book. So that it appears to be clearly conceived as the inverse procedure which is followed in deduction, and this last procedure has to be abandoned as a method of scientific explanation, if we use induction.

[13] **Victor Christianto**

Daniel, you are right that some people rejected induction as a form of logic, for instance Peter Medawar and Karl Popper.
See http://www.ssr.org/sites/ssr.org/files/uploads/attachments/node/16/rothchild_scimethod.pdf

But i still think that both approaches are required. For instance, when you study some literatures in order to form a hypothesis, you use deduction. But when you use statistics, and infer conclusion from a set of given data, then you use induction. At least that is what i learned from a lecture of research methodology. Of course, one can argue which form of logic is better. Best wishes

[14] **Jochen Wilhelm**

Victor, I do not understand how a hypothesis is derived by deduction. AFAIK, the context where a hypothesis enters reasoning is the *modus ponens*, where a conditional statement and a hypothesis are given, in order to deduce a conclution. The hypothesis is not deduced. I think that the creation of a hypothesis rather is a deeply inductive process.

After that, we may use this hypothesis and data and -deductively!- reason by *modus tollens* that the hypothesis is wrong.

[15] **Victor Christianto**

Jochen, you may be right that even in forming hypothesis, a scientist can use induction too. What i meant with deduction is logical conclusion derived from past findings/theories. For instance, if we believe a theory which says that serotonin has relation to juvenile delinquency (youth crimes), then one can hypothesize that certain food types can trigger serotonin activity therefore they may affect anger behavior (related to juvenile delinquency). Then one can begin designing certain experiment to validate or invalidate that hypothesis, that is by inductive reasoning. But perhaps i can make mistakes here.

[16] **Jochen Wilhelm**

Victor, sure this is absolutely correct what you say, that hypothese are derived from past experiences. But this process is in fact an induction. The example you gave is also an induction. The (in-)validation later on is a deductive process. So it seems that one of us is just confusing induction with deduction (It might be me, but from my point of view I believe that I am right...).

[17] **Victor Christianto**

Dear Jochen, yes perhaps there is misinterpretation here. If you want simple definitions of deduction vs. induction, please check: http://www.ssr.org/sites/ssr.org/files/uploads/attachments/node/16/rothchild_scimethod.pdf.

For example, induction is generalization from particular observation/data using statistical inference. For example, if you conduct polling or survey, then you always take certain sample, but then you try to make conclusion for the entire population. This is called statistical inference, and it is determined by method of and size of sampling. But if you use logical reasoning from a general theory to derive a particular hypothesis, that is called deductive reasoning. From what i learned several years ago, the textbooks also explain like that. Best wishes

[18] **Daniel Courgeau**

Dear Jochen, dear Victor,

I think that the main difference between deduction and induction may be the following: deduction consists in deducing consequences from principles taken as given; induction consists of the deduction of principles from the study of their consequences. So that induction can be seen as the exact opposite of deduction. We can say that the modern sciences, at the time of Bacon, were born from the abandonment of deduction as a method of explanation, in favor of induction. Unfortunately later, with the empiricist tradition put forward by Hume or Mill and developed by Popper, deduction became again the main method for explanation, as they considered induction only as a generalisation from particular instances. Only recently a renewal of interest in classical induction appears. Once the functional architecture is known (by induction) then it can guide (by deduction) the empirical investigation of the mechanism which generate the observed properties. See for more details Franck, 2002, The explanatory power of models.

Concluding remarks

It seems possible to propose Bayesian epistemology as the basis of acceptance for a theory, although it has not been widely used yet.

References:

[1] Stephen Hartmann (2009) Bayesian epistemology, http://www.stephanhartmann.org/HartmannSprenger_BayesEpis.pdf