## Title

A Note on Statistical Hypothesis Testing: Probabilifying Modus Tollens Invalidates Its Force? Not True!



## Journal

Educational and Psychological Measurement, 83(6)

## Author

Widaman, Keith

## Publication Date

2023-12-01

## DOI

10.1177/00131644221145132

Peer reviewed

*Original Research Article*

Educational and Psychological Measurement
2023, Vol. 83(6) 1160–1172

DOI: 10.1177/00131644221145132
journals.sagepub.com/home/epm
SAGE

# A Note on Statistical Hypothesis Testing: Probabilifying *Modus Tollens* Invalidates Its Force? Not True!

**Keith F. Widaman**[1]

**Abstract**

The import or force of the result of a statistical test has long been portrayed as consistent with deductive reasoning. The simplest form of deductive argument has a first premise with conditional form, such as $p \rightarrow q$, which means that "if $p$ is true, then $q$ must be true." Given the first premise, one can either affirm or deny the antecedent clause ($p$) or affirm or deny the consequent claim ($q$). This leads to four forms of deductive argument, two of which are valid forms of reasoning and two of which are invalid. The typical conclusion is that only a single form of argument—denying the consequent, also known as *modus tollens*—is a reasonable analog of decisions based on statistical hypothesis testing. Now, statistical evidence is never certain, but is associated with a probability (i.e., a $p$-level). Some have argued that *modus tollens*, when probabilified, loses its force and leads to ridiculous, nonsensical conclusions. Their argument is based on specious problem setup. This note is intended to correct this error and restore the position of *modus tollens* as a valid form of deductive inference in statistical matters, even when it is probabilified.

**Keywords**

statistical hypothesis testing, *modus tollens*

In scientific endeavors, confirmation of predictions and disconfirmation or falsification of predictions have long been contrasted. In articles and textbooks discussing

[1]University of California at Riverside, USA

**Corresponding Author:**
Keith F. Widamam, School of Education, University of California, 900 University Avenue, Riverside, CA 92521, USA.
Email: keith.widaman@ucr.edu

hypothesis testing, confirmation of predictions is often touted. Researchers are encouraged to develop theories, which lead to conjectures regarding expected patterns in data. If the predicted pattern in data is found, an investigator typically promotes this confirmation as evidence that the theory driving the empirical work was supported. This impulse to confirm hypotheses is consistent with the most basic contentions of the Vienna Circle and the associated Berlin Circle of philosophers and philosophically oriented scientists and mathematicians in the 1920s and 1930s. The Vienna and Berlin Circles propounded the approach known as logical positivism. One fundamental proposition of logical positivism was that a statement had meaning only to the extent that it could be verified or confirmed. Indeed, the meaning of a statement was, essentially, its verification. If a statement could not be verified with certainty, it was meaningless.

The opposing position—emphasizing the importance of disconfirmation or falsification of predictions—was promoted by Popper (1934/1959/1992) in his influential tome *The Logic of Scientific Discovery*. Popper argued that disconfirmation of predictions was the road toward scientific progress, because it revealed problems for the theory under investigation. Prior to conducting a study, a scientist should state clearly the empirical outcome from a study that would disconfirm or falsify the theory-derived hypotheses driving the study. If the empirical outcome did falsify predictions, the scientist should discard the theory and develop a theory that could account for the results. To emphasize the importance of falsification, Popper argued that one could separate scientific endeavors from nonscientific ones based on whether their conjectures could be falsified. If a theory led to conjectures that could be falsified, it was scientific; if the conjectures derived from a theory could not be falsified, the theory was not scientific.

The force of disconfirmation has been questioned for over three decades primarily on the basis of the lack of certainty regarding empirical findings. Admittedly, scientific findings are never certain, but have associated $p$ values. Consider a nil null hypothesis, or $H_0$, that the means for the treatment and control conditions in an experiment are equal in the population. If the nil null hypothesis is tested and leads to a test statistic with a small $p$ value (e.g., $p < .05$ or $p < .01$), researchers are encouraged to reject the null hypothesis of no mean difference in favor of the alternative hypothesis that the group means differ in the population. This decision to reject $H_0$ could be in error. Indeed, $H_0$ could be true, and rejection of $H_0$ when it is indeed true is termed a Type I error, with the likelihood of a Type I error embodied in the $\alpha$ level chosen. The $\alpha$ level is a conditional statistic, representing the long run probability of incorrectly rejecting a true $H_0$.

Pollard and Richardson (1987) argued that the probabilistic nature of the decision based on a test statistic rendered the falsificationist position to be untenable. Indeed, they argued that a probabilistic representation of falsificationist reasoning leads to comically nonsensical conclusions. In a highly cited paper,[1] Cohen (1994) doubled down on the reasoning of Pollard and Richardson, highlighting the nonsensical conclusion supported by a probabilistic form of deductive reasoning. More recently, Miller (2017) supplied the problem setup proposed by Pollard and Richardson and repeated by Cohen, but countered their position by pointing out certain problematic issues with that setup.

The goal of the current article is to extend the critique by Miller (2017) by exposing fundamental errors in the problem setup by Pollard and Richardson (1987) and espoused by Cohen (1994). To do so, I first review simple forms of syllogistic reasoning. I then discuss how probability enters into that reasoning, based on statistical evidence. The upshot is that the parody of probabilistic reasoning regarding the import of statistical evidence promulgated by Pollard and Richardson and echoed by Cohen is based on fallacious reasoning. Furthermore, the logic of drawing conclusions from empirical data is confirmed, even acknowledging the probabilistic nature of the data, the statistical test values, and associated conclusions.

## Four Forms of Deductive Inference

To understand the basis of Popper's argument in favor of falsificationist logic, it is useful to describe four simple forms of deductive argument or syllogistic reasoning. Here, I use the following symbols: $p$ stands for an antecedent claim, $q$ for the consequent claim, $\rightarrow$ is a conditional implication, ~ stands for negation, $\wedge$ represents conjunction (i.e., the logical ''and'' operation), $\vee$ stands for disjunction (i.e., the logical ''or'' operation), and $\therefore$ for conclusion (i.e., ''therefore''). We call a premise, such as $p \rightarrow q$, a conditional premise involving two assertions or propositions, $p$ and $q$. In this conditional premise, $p \rightarrow q$ means, symbolically, that ''if $p$, then $q$,'' or ''if $p$ is true, then $q$ is true.'' Given the conditional premise, we can then either affirm or deny the antecedent assertion ($p$) or affirm or deny the consequent assertion ($q$) to arrive at a conclusion, leading to four forms of deductive argument. Two of these forms of argument are valid and lead to correct inferences or conclusions, and two are deductively invalid. The validity of the forms of argument can be illustrated with simple propositions.

### *Form 1: Affirming the Antecedent, or* Modus Ponens

To allow clearer understanding of the *modus ponens* form of reasoning, let $p$ stand for the antecedent proposition that ''it is raining'' and $q$ stand for the consequent proposition that ''there are clouds in the sky.'' The major or conditional premise $p \rightarrow q$ then becomes ''If it is raining, then there are clouds in the sky.'' This appears to be an unequivocal claim because, notwithstanding the query by Fogerty,[2] rain only occurs if clouds are in the sky. Indeed, if no clouds are in the sky, any purported precipitation would not be rain. If we affirm the truth of the antecedent $p$, we then conclude that the consequent $q$ is true, as:

| | |
|---|---|
| $p \rightarrow q$ | If it is raining, then there are clouds in the sky |
| $\underline{p}$ | <u>It is raining</u> |
| $\therefore q$ | $\therefore$ There are clouds in the sky |

So, in reasoning by *modus ponens* or affirming the antecedent, if one can affirm the truth of the antecedent premise, it is valid to conclude that the consequent is true.

## *Form 2: Denying the Consequent, or* Modus Tollens

In *modus tollens* reasoning, one denies the consequent, and then concludes by denying the antecedent, as:

| | |
|---|---|
| $p \rightarrow q$ | If it is raining, then there are clouds in the sky |
| $\underline{\sim q}$ | <u>No clouds are in the sky</u> |
| $\therefore \sim p$ | ∴ It is not raining |

Under *modus tollens*, if one can deny the consequent assertion with certainty, it is valid to conclude that the antecedent must not be true. In the present instance, if it is obvious that the consequent assertion—the presence of clouds in the sky—is false, it is safe, and imminently logical, to conclude that any apparent precipitation is not rain (e.g., it might instead derive from a lawn sprinkler or a bird).

## *Form 3: Affirming the Consequent*

Under this third form of argument, here we affirm the consequent (rather than denying it) and would like to conclude that the antecedent is true, which can be written as:

| | |
|---|---|
| $p \rightarrow q$ | If it is raining, then there are clouds in the sky |
| $\underline{q}$ | <u>There are clouds in the sky</u> |
| $\therefore p$ | ∴ It is raining |

It is certainly true that, if it were raining, there must be clouds in the sky. But, the simple observation of clouds in the sky does not allow one to conclude with certainty that it is raining; clouds often fail to produce rain. Hence, affirming the consequent is an invalid form of deductive argument for concluding that the antecedent premise is true.

## *Form 4: Denying the Antecedent*

The final form of argument involves denying the antecedent assertion, presuming we could then conclude by denying the consequent, as:

| | |
|---|---|
| $p \rightarrow q$ | If it is raining, then there are clouds in the sky |
| $\underline{\sim p}$ | <u>It is not raining</u> |
| $\therefore \sim q$ | ∴ There are no clouds in the sky |

As is clear from the example, denying the antecedent does not allow one to conclude that the consequent is false. In this case, if we accept the claim that it is not raining, this offers no basis at all for any conclusion regarding whether there are clouds in the sky.

## Implications for Scientific Research

At this point, one might wonder what implications these four forms of deductive inference have for scientific practice. To illuminate this issue, do the following: Replace $p$ with $T$ (which stands for *theory*), and replace $q$ with $D$ (which stands for *data* or *a predicted outcome in data*). The conditional premise then becomes $T \rightarrow D$, which represents the claim that ''if my theory is correct, then a predicted outcome in data will occur.'' This reasonable assertion is consistent with Introduction sections of most empirical articles, in which scientists develop the theoretical background for their study and then offer predictions regarding hypothesized patterns in data based on their theorizing. As one example, *modus tollens* reasoning then becomes

| | |
|---|---|
| $T \rightarrow D$ | If my *T*heory is correct, then a predicted outcome in *D*ata will occur |
| $\underline{\sim D}$ | <u>The predicted outcome in *D*ata did not occur</u> |
| $\therefore \sim T$ | ∴ My *T*heory is incorrect |

At this point, what do the four forms of argument hold for scientific argument? The first form of argument, *modus ponens*, cannot be used. We can never affirm with absolute certainty that our theory is true or correct. If we could do so, the deductive argument of *modus ponens* would allow us to conclude that we must ever and always observe certain patterns or outcomes in data. However, just as we can never affirm absolutely the correctness of our theory, we can never conclude that predicted outcomes in data must invariably hold—because predicted outcomes in empirical data often are not found. Hence, *modus ponens*, which is a valid form of deductive argument, cannot be used as the basis for scientific practice.

The form of argument used most often in empirical research conforms to Form 3, affirming the consequent, which is an invalid form of deductive argument. A scientist might argue that a theoretical conjecture is reasonable and that, if it were true, certain predicted patterns in data should be found. Suppose that the predicted pattern in data is affirmed in a study. A researcher often argues that this affirms the truth of the theory or at least the theoretical conjecture driving the study. However, affirming the consequent (e.g., finding the predicted pattern in data) does not allow one to conclude validly that the antecedent (e.g., the theory) has been affirmed. At most, the researcher can claim inductive support for the theory, but must always be vigilant because the predicted result may have come about for reasons other than those hypothesized, such as an alternative theory that supports the same empirical conjecture.

The fourth form of argument, denying the antecedent, is a poor choice for scientific endeavor. It is difficult to conceive how research would be conducted if a researcher began the research enterprise by denying the reasonableness of theory. In fact, if theory is embedded as the antecedent phrase of the conditional premise, neither affirming nor denying the antecedent is open as a gambit as the second premise.

This leaves Form 2, *modus tollens*, as the only valid form of deductive argument open to the practicing scientist. In science, we become inculcated in one theory or another, and we then formulate conjectures, which lead to predicted patterns or outcomes in data. If we find the predicted patterns in data, this leads to affirming the consequent, which (as noted above) may indeed provide additional inductive support for our theory. But, if we fail to find the predicted pattern in data, we have thereby denied the consequent. When the consequent is denied, the ‘‘arrow of *modus tollens*’’ is directed at the antecedent, and it is valid to deny the antecedent and conclude that the theory, or at least the theoretical conjecture, has been falsified.

Note that Meehl (1990) supplemented the basic form of *modus tollens* by considering additional features of scientific investigations to yield a more informed version of scientific reasoning based on empirical outcomes. Meehl allowed $T$ to stand for core theory postulates, but added $A_t$ (for auxiliary theoretical conjectures), $A_i$ (for auxiliary instrument or measurement assumptions), $C_n$ (for conditions of the experiment or study), and $C_p$ (for the ceteris paribus clause, or ‘‘all other things being equal’’). If we continue to use D to stand for a predicted outcome in data, *modus tollens* then becomes:

$$(T \wedge A_t \wedge A_i \wedge C_n \wedge C_p) \rightarrow D$$

$$\underline{\sim D \quad\quad\quad}$$

$$\therefore \sim (T \wedge A_t \wedge A_i \wedge C_n \wedge C_p) = \sim T \vee \sim A_t \vee \sim A_i \vee \sim C_n \vee \sim C_p$$

The reasoning depicted above means that if the conjunction of ‘‘my core theory *and* auxiliary theoretical conjectures *and* auxiliary instrument assumptions *and* conditions of the experiment *and* all other things are equal’’ is reasonable and justified, then a predicted pattern in data will be observed. If the pattern in data is not observed (i.e., $\sim D$), the proper conclusion is that the antecedent conjunction is falsified. When the parenthesis in the antecedent is removed, this means that one or more of the five components of the conjunction may be at fault: core theory may be faulty *or* auxiliary theoretical conjectures may be faulty *or* auxiliary instrument assumptions may be problematic *or* conditions of the experiment may be faulted *or* perhaps all other things were not equal. This allows the ‘‘arrow of *modus tollens*’’ to be directed at any of the components of the conjunction, allowing the core theory $T$ to remain immune from rejection if problems can be identified elsewhere.

## Probabilifying *Modus Tollens*

There is general agreement that *modus tollens* is a valid form of deductive inference when premises in an argument can be verified categorically as being either true or false, but some, including Pollard and Richardson (1987), Cohen (1994), and Falk and Greenbaum (1995), have argued that *modus tollens* loses its force when truth or falsity of premises can be determined only probabilistically. If probabilifying a premise weakens or destroys the force of *modus tollens*, this might be a serious problem, disrupting the falsificationist stance.

## Setup 1

I argue here that the claim that probabilifying *modus tollens* renders it moot or untrustworthy is unsound and due to improper problem setup. Recall our earlier presentation of *modus tollens* in the section on deductive arguments, which I here call Setup 1:

| | |
|---|---|
| $T \rightarrow D$ | If my *T*heory is correct, then a predicted outcome in *D*ata will occur |
| $\sim D$ | The predicted pattern in *D*ata did not occur |
| $\therefore \sim T$ | ∴ My *T*heory is incorrect |

This rendition of *modus tollens* reasoning is valid, albeit of questionable application to empirical analyses due to the probabilistic nature of conclusions about patterns in data.

## Setup 2

In developing an argument that *modus tollens* reasoning with respect to statistical testing is faulty, Pollard and Richardson (1987) focused on the null hypothesis, $H_0$, rather than theoretical predictions, reorienting Setup 1 acceptably as the version I call Setup 2:

| | |
|---|---|
| $H_0 \rightarrow \sim D$ | If $H_0$ is true, then a predicted outcome in *D*ata will *not* occur |
| $\sim(\sim D)$ | The predicted *nil* pattern in *D*ata did *not* occur |
| $\therefore \sim H_0$ | ∴ $H_0$ is false (or rejected) |

A standard nil null hypothesis $H_0$ represents the conjecture that a nil difference is true in the population, such as that treatment and control group means are equal in the population or that a given regression weight is zero in the population. Hence, the prediction is made that, if $H_0$ is true, the empirical finding will be nil or will not depart significantly from nil. When data are collected, assume that a statistical test is consistent with the conclusion that a significant departure from zero has been found (e.g., group means differ significantly). The conclusion in this case to reject the antecedent (i.e., $\sim H_0$) is justified, because the second premise, $\sim(\sim D)$, is the denial of the consequent, as the denial of ''not-*D*'' is *D* itself, or $\sim(\sim D) = D$. Thus, this reframing of the *modus tollens* form of argument is valid, but still of questionable utility to scientists due to the failure to consider the probabilistic nature of statistical decisions.

## Setup 3

Pollard and Richardson (1987) then went further and offered the following probabilified version, which I term Setup 3, as:

| | |
|---|---|
| $H_0 \rightarrow p\text{~}D$ | If $H_0$ is true, then a predicted outcome in Data is very unlikely |
| $\underline{D}$ | The predicted pattern in Data did occur |
| $\therefore p\text{~} H_0$ | $\therefore$ $H_0$ is very unlikely |

where I took the liberty to create the operator ''*p*~'' to mean ''probably not'' or ''very unlikely,'' there being no commonly accepted operator in symbolic logic of which I was aware to designate ''probably not'' or ''very unlikely.'' Pollard and Richardson opined that the problematic nature of this probabilified version of *modus tollens* would become obvious if one substituted ''This person is American'' in place of $H_0$ and ''This person is a member of Congress'' in place of *p*~*D*, but they left the reader to demonstrate this.

In a section headed ''The Permanent Illusion,''Cohen (1994) elaborated on several forms of deductive reasoning, concluding that probabilistic forms of modus tollens led to specious conclusions. In particular, Cohen contrasted arguments conforming to Setups 2 and 3 to demonstrate that the probabilified Setup 3 is invalid. Cohen followed the suggestion by Pollard and Richardson (1987) to replace $H_0$ by ''person is an American,'' ''*D*'' with ''person is a member of Congress'' (so ~*D* = ''person is *not* a member of Congress''). Now, imagine a person you do not know walks through the door and presents his or her passport and ID to you, Setup 2 becomes:

If the person is an American, then the person is not a member of Congress [FALSE!]

The person is a member of Congress

∴ The person is not an American

The preceding is an acceptable instance of *modus tollens* reasoning, because the second premise constitutes a denial of the consequent in the first premise, but it is clear that the conclusion is not justified. Why? Because a conclusion using *modus tollens* is true or valid or correctly drawn only if both the antecedent and consequent claims are verified as true, and the first premise is clearly false, as every member of Congress must be an American citizen. If a person is an American, it is improper to conclude that they are not a member of Congress, so the first premise is false.

Cohen (1994), following Pollard and Richardson (1987), argued that the probabilistic form provided by Setup 3 seems, on the face of it, to be more acceptable, but it leads to the nonsensical conclusion as:

If the person is an American, then the person is *probably not* a member of Congress

The person is a member of Congress

∴ The person is *probably not* an American

Because members of Congress are few in number (535 voting members, consisting of 435 representatives and 100 senators), a random person walking through the door is unlikely to be a member of Congress. However, if the person were verified to be a member of Congress (second premise), the conclusion that the person is probably not an American is clearly illogical and nonsensical, implying that the probabilified version of *modus tollens*, Setup 3, is faulty.

## Setups 4 and 5

The problem with this entire line of reasoning by Pollard and Richardson (1987) and Cohen (1994) is that Setup 3 fails to represent accurately the reasoning embodied in statistical testing. To convey statistical testing properly, the first premise must remain categorical (i.e., unqualified) and the second premise is the one that is probabilified. Researchers never argue that, if their theory is correct, they will *probably* observe a predicted pattern in data. Instead, they argue that a predicted outcome *will* occur if their theory is correct. Stated conversely, if the null hypothesis is true, then the theory-predicted outcome will not occur.

The probabilifying occurs in the evaluation of the empirical results, where one obtains the probability—not certainty—of the data, given the null hypothesis, or $P(D|H_0)$. To be clear, the probabilifying occurs with regard to the second premise. Thus, one can never assert categorically (i.e., without doubt) that a particular pattern in data has occurred, as required by Setups 1, 2, and 3. Instead, one can assert only that a predicted result occurred accompanied by a small probability value under the nil null hypothesis, because the result might be a statistical fluke, a chance outcome that might be unlikely, but acknowledged by the choice of an alpha level when evaluating statistical test results. Properly leaving the first premise categorical and allowing the second premise to be probabilified leads to Setup 4:

| | |
|---|---|
| $H_0 \rightarrow \sim D$ | If $H_0$ is true, then a predicted outcome in *D*ata will not occur |
| $p\sim(\sim D)$ | The predicted pattern in *D*ata probably did occur |
| $\therefore p\sim H_0$ | ∴ $H_0$ is very unlikely |

or, when reoriented equivalently toward evaluating the theory, Setup 4 becomes Setup 5:

| | |
|---|---|
| $T \rightarrow D$ | If my *T*heory is correct, then a predicted outcome in *D*ata will occur |
| $p\sim D$ | The predicted pattern in *D*ata probably did *not* occur |
| $\therefore p\sim T$ | ∴ My *T*heory is probably incorrect |

Note that the ''$p\sim$'' in the second premise of Setups 4 and 5 does not lead to a categorical or absolute denial of the consequent in the first premise, but in the ''likely denial'' of the consequent. This then leads to the ''likely denial'' of the antecedent.

The key issue with Setups 4 and 5 is that the first premise is categorical or unqualified, and the probabilifying is applied to the second premise. Because of this, Setups 4 and 5 can handle the situation of the nationality and occupation of the random person identified above. The first premise in categorical form can be verified as true if membership in Congress requires that a person be an American citizen. Setup 5 becomes:

If the person is member of Congress, then the person is an American

The person is probably not an American

∴ The person is probably not a member of Congress

Every member of Congress must be an American citizen, so the first premise is true. After inspecting the person's passport, suppose you notice several indications that the passport might be a fake, leading you to suspect strongly that the person is not an American citizen, rendering the consequent claim—that the person is an American—unlikely. This probabilistic denial of the consequent leads to the reasonable conclusion that the person is very likely not a member of Congress, even if the person is wearing a ''member of Congress'' lapel pin.

The upshot of the argument in this section is that probabilifying *modus tollens* does not lead to a ridiculous or illogical argument form. Instead, Pollard and Richardson (1987), Cohen (1994), and others who have followed this line of attack provided a flawed rendition of how probability enters *modus tollens*. Properly reoriented, the probabilified form of *modus tollens* exemplifies the tenuous nature of inferences we make based on statistical tests of parameter estimates and models.

## A Final Caveat or Two

As a final note, it is important to understand just what has been argued here and the proper interpretation of my claims. The central claim in the current article is that previous arguments that a probabilized form of *modus tollens* reasoning leads to problematic or fallacious conclusions are incorrect, having been based on a faulty characterization of the manner in which researchers formulate hypotheses, conduct studies, and then draw conclusions from their data. Theorizing is done in categorical fashion, developing ideas about the psychological processes that generate data. Based on theory regarding these processes, certain patterns in empirical data should occur, and these predicted patterns constitute the researcher's conjectures motivating a study. After data are collected and analyses are performed, probability statements are based on results of statistical tests conducted on the data. These probabilistic conclusions are rendered with regard to the data, based on the hypothesis tested.

To reiterate, a statistical evaluation of results of a study provides only the probability of the data, given the hypothesis. More specifically, we obtain the probability of obtaining results this extreme or more extreme under the hypothesis being tested. If the hypothesis is a typical nil null hypothesis, we get $P(D|H_0)$. That is, the probability statement based on a test statistic is a statement about the probability of the data given the hypothesis under test, not a probability statement about the hypothesis. If the probability of the data is very low (e.g., $p < .01$) given the hypothesis tested, it is reasonable to reject the tested hypothesis in favor of an acceptable alternative. It is unfortunate that, in most psychology studies, the hypothesis tested is a nil null hypothesis that ''nothing is going on,'' and the alternative is a relatively uninformative one that ''something is going on.'' If, on the basis of a statistical test, one rejects the hypothesis that ''nothing is going on,'' it is in fact reasonable and logical to conclude that ''something is going on'' (or, at least, that ''there is not nothing going on'') even if the magnitude of the ''something that is going on'' is wholly unspecified and therefore relatively uninformative.

Setup 4 is an acceptable probabilified version of *modus tollens* reasoning when evaluating $H_0$ and provides a conclusion that appears to conform to the probability of $H_0$ given the data, or $P(H_0|D)$. That should *not* be the takeaway message. The null hypothesis $H_0$ is either true or it is false, so the probability of $H_0$ is either 1 (i.e., it is true) or 0 (i.e., it is false), and no empirical investigation can provide indisputable evidence with regard to the truth value of $H_0$ or even a justified probability of its truth value.[3] In Setup 4, a probabilistic decision is made based on an empirical outcome. If the probability of the data is very low given the hypothesis tested, then the hypothesis tested is not a likely account of the data—because the data are very unlikely under that hypothesis. As a result, it is reasonable to reject the hypothesis as an account of the data even if a probability cannot be assigned to the hypothesis tested.

We never get the probability of the null hypothesis given the data, or $P(H_0|D)$, nor do we get the more desirable probability of the alternative hypothesis given the data, or $P(H_A|D)$. Meehl (e.g., 1990) often argued that, at least in his areas of research, the point nil hypothesis $H_0$ was always false, because everything was correlated at least a little bit with everything else. Even Tukey (1969) observed that the point nil hypothesis $H_0$ was unlikely ever to be true even in true experiments with random assignment of participants to conditions—if one took results out to a sufficient number of decimal places. Based on reasoning by eminent scholars such as Meehl and Tukey, it is tempting to propose that the probability of the point nil hypothesis is zero, so $P(H_0) = 0.0$, and the probability of the alternative hypothesis is 1.0, or $P(H_A) = 1.0$. Note that neither of these probabilities has any connection with data, consistent with the very general claims by Meehl and Tukey.

The evaluation of empirical data will always be done in a probabilistic fashion, evaluating data in relation to a hypothesis that is tested. If the data obtained or data even more extreme are very unlikely given the hypothesis tested, then it is logical to reject the hypothesis as an account of the data in favor of an alternative hypothesis. Contrary to the claims by detractors, the probabilified version of *modus tollens*—

properly portrayed—provides a valid and justifiable model for drawing conclusions about the acceptability of hypotheses as accounts of data, even if the probabilities involved are in reference to the data, and not the hypotheses under test, and even if the conclusions regarding acceptance or rejection of hypotheses might be drawn in error.

## Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Keith F. Widaman https://orcid.org/0000-0002-6424-3998

## Notes

1. A Google Scholar search on November 8, 2022 indicated that the Cohen (1994) paper is a citation classic, having been cited 5,857 times.
2. Lyrics by John Fogerty for ''Have you ever seen the rain?'' Available from https://www.azlyrics.com/lyrics/johnfogerty/haveyoueverseentherain305650.html
3. When attempting to predict a very unlikely outcome, even a screening test with excellent sensitivity and specificity can lead to incorrect decisions. Cohen (1994) discussed a screening test for schizophrenia, and Falk and Greenbaum (1995) presented a problem in diagnosing whether a fetus has Down syndrome using results from amniocentesis. Both of these examples concern outcomes with extremely low base rates, where screening test results may provide very misleading information. That is, in such situations, the probability of $H_0$ (i.e., non-schizophrenic or non-Down status) given the data (i.e., a positive sign of abnormal status from the screening test), or $P(H_0|D)$, can be very high even if the $P(D|H_0)$ is extremely low and supports a decision to reject $H_0$. Falk and Greenbaum argued that ''The very existence of a counter-example [i.e., an example in which an incorrect decision to reject $H_0$ is made when $P(D|H_0)$ is very low] discredits the logic of tests of significance'' (p. 79). This claim would be supportable only if one assumed that decisions based on statistical tests were categorical and therefore allow one to reject a tested hypothesis, such as $H_0$, with certainty. If, however, one assumes, as is common practice, that decisions based on statistical tests to accept or reject $H_0$ are probabilistic in nature and can be made in error, the counter-example does not discredit the logic of tests of significance. Instead, the Cohen and Falk and Greenbaum counter-examples are extremely useful illustrations of how wrong one can be in assuming that $H_0$ is very unlikely in certain situations in which the $P(D|H_0)$ is so low that rejection of $H_0$ seems justified.

## References

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory & Psychology*, *5*(1), 75–98. https://doi.org/10.1177/0959354395051004

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, *1*(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1

Miller, J. (2017). Hypothesis testing in the real world. *Educational and Psychological Measurement*, *77*(4), 663–672. https://doi.org/10.1177/0013164416667984

Pollard, P., & Richardson, J. T. E. (1987). On the probability of making Type I errors. *Psychological Bulletin*, *102*(1), 159–163. https://doi.org/10.1037/0033-2909.102.1.159

Popper, K. R. (1934/1959/1992). *The logic of scientific discovery*. Routledge. https://doi.org/10.4324/9780203994627

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, *24*(2), 83–91. https://doi.org/10.1037/h0027108