# The Duhem-Quine Thesis: A Non-Standard Approach [1]

Abstract:

There are two versions of the Duhem-Quine thesis: (i) confirmation holism and (ii) evidence holism. After disentangling the notions of confirmation and belief from that of evidence, we propose two Bayesian accounts to address these two versions of the Duhem-Quine thesis. We further distinguish confirmation holism and evidence holism into two respective sub-varieties. We argue that neither of the versions of the Duhem -Quine thesis is tenable. (Word Count 66)

1. Overview
2. Two Distinctive Accounts: Confirmation and Evidence
3. Varieties of Confirmation Holism and Evidence Holism
4. The Paradigm Example and the Problems of Confirmation Holism
5. Problems of Evidence Holism
6. Summing Up

WORD COUNT for the paper is 7,740 that include the word count for six tables, footnotes and logical/mathematical notations.

# The Duhem-Quine Thesis: A Non-Standard Approach

## 1. Overview

W. V. Quine was perhaps the first philosopher to articulate a coherent picture of how theories are verified as a whole along with their auxiliaries in a broader context although the physicist/philosopher Pierre Duhem proposed theory confirmation on a holist line much earlier. [2] Later on, their views on holism are often lumped together as the Duhem-Quine thesis.[3] However, there is a great deal of controversy concerning what *the* Duhem-Quine thesis is. But, generally considered, it seems to have two variants. The first involves the claim that hypotheses are never confirmed or disconfirmed in isolation, but only in conjunction with other auxiliaries, ultimately with all of the hypotheses and their auxiliaries that constitute science itself at any given time. This claim might be called confirmation holism (CH). The second version involves the claim that, given the evidence, investigators are unable to discriminate between the conjoined hypotheses and auxiliaries after a hypothesis has been confirmed/disconfirmed. Hence, one might call this second claim evidence holism (EH). When the hypotheses along with their auxiliaries are confronted with recalcitrant data, EH implies that evidence does not single out any one of them in particular for rejection. The only requirement for EH is that the *whole* system of hypotheses and auxiliaries be "kept squared with experience"[4] in which case the decision to reject one or another of its components can only be made on the basis of extra-evidential, pragmatic considerations. It should be clear that these two versions of holism rest on a distinction between confirmation and evidence. Consequently, it is necessary to develop this distinction at the outset.

## 2.      Two Distinctive Accounts: Confirmation and Evidence

We propose distinct accounts of confirmation and of evidence, within a Bayesian framework, to respond to two types of questions, here called the belief question and the evidence question.[5] The belief question consists in asking, "given the data what should we *believe* about a hypothesis and to what degree?" In contrast, the evidence question asks, "what do the data say about *evidence* for the hypothesis against its alternative?" Since the belief question has to do with what an agent should believe, its answer lies in an account of confirmation which measures an agent's degree of belief. An account of confirmation provides a confirmation relation, C(H, D, A,B) among the hypothesis, H, data, D, auxiliaries for the hypothesis, A, and the agents' background information, B. Because a confirmation relation is a belief relation, it must satisfy the probability calculus, including the rule of conditional probability, together with some reasonable constraints on one's *a priori* degree of belief in an empirical proposition. The satisfaction of the rules of the probability calculus, including the rule of conditional probability by an agent's degree of belief, is known as the *coherence condition*. According to this account, one should believe a hypothesis to a degree if its posterior probability (P(H|D)) is greater than its prior probability (P(H)). The posterior and prior probabilities are essential features of the Bayes' theorem, which states that the posterior probability of the hypothesis equals its prior probability multiplied by its likelihood function (P(D|H)) and then divided by the marginal probability of the data (P(D)).

$$P(H \mid D) = \frac{P(H) \times P(D \mid H)}{P(D)} \qquad\qquad \text{(EQ 1)}$$

The prior probability of a hypothesis depends on the agent's degree of belief in the hypothesis before data for it have been gathered. The likelihood function provides an answer to the question, "How likely are the data, given the hypothesis?" The marginal probability represents the probability that D would obtain, averaged over the hypothesis's being true and false. Confirmation for a hypothesis becomes weak or strong depending on the strength of the difference between its prior and posterior probabilities.

In contrast, the account of evidence takes the likelihood ratio (LR) of the two likelihood functions as basic.[6] While the account of confirmation is confined to a single hypothesis embodied in EQ 1, an account of evidence *compares* the merits of two hypotheses, H1 and H2 (or ~ H1) relative to the data D, and background information B.

$$LR = \left[ \frac{P(D \,|\, H1 \,\&\, B)}{P(D \,|\, H2 \,\&\, B)} \right] \qquad\qquad \text{(EQ 2).}$$

Note that in (EQ2) if $1 < LR \leq 8$, then D is often said to provide weak evidence for H1 against H2, while when $LR > 8$, D provides strong evidence. This cut-off number follows a statistical practice common among investigators.

Here, hypotheses are taken to be *simple* statistical hypotheses. A hypothesis is simple if the distribution of the data under of the hypothesis is fully specified. The evidence relation that our account captures rests on the likelihood ratio which does not satisfy the probability calculus. Like the account of confirmation, the account of evidence is also Bayesian. However, unlike the account of confirmation which is understood in terms of an agent's degree of belief in a hypothesis, the account of evidence is agent-independent. In fact, the likelihood ratio which captures the evidence relation is objective

in the sense that the ratio is independent of an agent's prior probability (Berger, 1985, p.146).[7]

As we have noticed since both accounts, an account of confirmation, and an account of evidence, are devised to respond to two types of questions, they are conceptually distinct. However, there is a close relationship between these two accounts shown by this theorem: [P(H|D) > P(H)] if and only if LR > 1].[8]  The theorem states that data (D) confirm the hypothesis to some degree when the posterior probability of the hypothesis is greater than its prior probability just in case the likelihood ratio is greater than 1.


## 3.        Varieties of Confirmation Holism and Evidence Holism

Earlier we distinguished two versions of the Duhem-Quine thesis, (i) confirmation holism (CH) and (ii) evidence holism (EH).[9]  We also distinguish CH into two sub varieties: (a) confirmation holism$_1$ (CH1) and confirmation holism$_2$ (CH2) and evidence holism (EH) into (c) evidence-ignoring holism (EIH) and (d) evidence-revisionary holism (ERH): for clarity

Confirmation Holism$_1$ (CH1):

(a): If a hypothesis, H, as a whole along with auxiliaries, A, has been confirmed together, then one should never be able to confirm H and A separately.

Confirmation Holism$_2$ (CH2):

(b): If H as a whole along with A has been disconfirmed together, then one should never be able to disconfirm H and A separately.

Evidence holism provides a response to the question, "what epistemic attitude should an investigator adopt toward a hypothesis and its auxiliaries when they have jointly been confirmed or disconfirmed?" As already noted, there are two variants of EH: c)

evidence-ignoring holism (EIH) and (d) evidence-revisionary holism (ERH). The motivating force behind EH is Quine's slogan: "[a]ny statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system…[c]onversely, by the same token, no statement is immune to revision."[10] EIH captures this attitude of disregarding evidence when a hypothesis as a whole has been confirmed or disconfirmed. In contrast, ERH captures this revisionary aspect of theory confirmation in *rejecting* a hypothesis which is held to be *correct*. However, according to EH, the investigator's choice (which statement to hold and which one to reject) is based on, for example, simplicity and usefulness, non-evidential/pragmatic considerations. Thus:

 Evidence-Ignoring Holism (EIH):

> (c): If a sentence would otherwise be rejected due to recalcitrant data, one could nevertheless retain that sentence come what may, provided one makes enough adjustments inside the entire system of sentences to square them with the data.

Evidence-Revisionary Holism (ERH):[11]

> (d): If a sentence is otherwise held to be true, then one could reject that sentence if one would like to do so, provided one makes enough adjustments inside the entire system of sentences to square them with the data.

Naturally, whatever policy one adopts toward adhering or jettisoning a sentence in the system, one should remain vigilant that the system as a whole should fit the empirical data.

## 4. The Paradigm Example and Problems of Confirmation Holism

Our paradigm example concerns two hospitals, B (for Billings) and the other in SLC (for Salt Lake City).[12] They participated in a study designed to measure the risk of lung cancer (LC) for smokers and non-smokers. In all, 2000 smokers and 4000 non-smokers were observed during the 15 years of the study. None of the subjects had experienced symptoms of LC prior to commencement of the study. Both among smokers and non-smokers there were

an equal percentage of LC cases, about 10 percent, implying that there seems to be "no association" between smoking and LC. The overall results of the two hospitals are displayed in Table I.

| | First episode of LC within 15 years? | | |
|---|---|---|---|
| | Yes | No | Total |
| Smokers | 200 (10%) | 1800 (90%) | 2000 |
| Non-smokers | 400 (10%) | 3600 (90%) | 4000 |

**Table I**

To evaluate the "no association" claim between two variables, smoking and LC, we need to consider a third variable to examine whether it acts as a confounder or confounder variable. If the effect of the smoking variable on the LC variable is mixed up with the effect of a third variable (i.e., the location) on the LC variable, then the third variable is called a confounder. In this case, to investigate the no association claim between smoking and LC, we take the third variable to be the location of the study, either Billings or SLC, to find out whether the latter's effect masks the real association between smoking and LC. If the effect of the location of study is mixed up with the effect of smoking on LC, then the location will be called a confounder. The data are separated by hospitals, in Tables II and III.

| **Billings Hospital** | First episode of LC within 15 years? | | |
|---|---|---|---|
| | Yes | No | Total |
| Smokers | 90 | 1410 | 1500 |
| Non-smokers | 20 | 980 | 1000 |

**Table II**

| SLC Hospital | First episode of LC within 15 years? | | |
|---|---|---|---|
| | Yes | No | Total |
| Smokers | 110 | 390 | 500 |
| Non-smokers | 380 | 2620 | 3000 |

**Table III**

In both Tables II and III, when the data are separated by hospitals, Billings and SLC, each location shows an association between smoking and LC. For example, in Table II, in Billings, a greater percentage of smokers have LC, that is, 6%, than non-smokers, that is, 2%. Similarly, in Table III, in SLC, a greater of percentage of smokers has the disease, that is, 22%, than non-smokers, that is, 12%. Table IV summarizes results of the last three tables succinctly.

| | First Episode of LC within 15 years | | | | |
|---|---|---|---|---|---|
| | Billings and LC Total Smokers =1500 Total Non-smokers=1000 | | SLC and LC Total Smokers = 500 Total Non-smokers=3000 | | |
| | LC | Percentage | LC | Percentage | Difference |
| Smokers | 90 | 6% | 110 | 22% | 16% |
| Non-Smokers | 20 | 2% | 380 | 12% | 10% |
| Difference | | 4% | | 10% | 6% |

**Table IV**

To show that that the paradigm example serves as an example of confirmation, one is required to show that it satisfies the coherence requirement.[13] From Table I, P(LC|S) = 0.1 and P(LC|~S) = 0.1 where S stands for smoking data and ~S the non-smoking data. To prove coherence for P(LC|S), one must show that

$P(LC \mid S) = [P(LC \mid B \& S) \times P(B \mid S)] + [P(LC \mid \sim SLC \& S) \times P(SLC \mid S)]$. It is easy to

show that the RHS is equal to the LHS = 0.1. $\left( \dfrac{90}{1500} \times \dfrac{1500}{2000} \right) + \left( \dfrac{110}{500} + \left( \dfrac{500}{2000} \right) \right) = 0.1$, so

the example satisfies the coherence requirement. Similarly, one could show that P(LC|~S) = 0.1

For both versions of confirmation holism (CH), we assume that in the paradigm example H1 stands for the hypothesis that smoking is a causal factor for LC, and H2 (i.e., ~H1) the denial of H1. A is the auxiliary that there is no a confounding variable that masks the real effect of smoking on LC. A here is a function of the data. Notice A need not be probabilistically independent of H1 in the dataset under consideration.[14] By this, we mean that the truth or falsity of A could affect the probability of H1, and therefore its confirmation. As we will see that in the dataset in question A, the auxiliary is likely to be false affecting the probability of the hypothesis, H1, that smoking is a causal factor for LC. We will argue that in the dataset there is confounding going on regarding the auxiliary (i.e., ~A) contributing to the probability of the hypothesis being false (i.e., ~H1). If our argument is correct then it follows from the rule of the probability theory that H1 and A1 are probabilistically dependent as are ~H1 (i.e., H2) and ~A.[15] Although in the dataset there is a probabilistic dependence between the hypothesis and its auxiliary, we leave open the possibility that in some other dataset we might find a probabilistic independence relationship between the hypothesis and its auxiliary as most discussions on resolving the Duhem-Quine thesis have assumed.[16] The presence of the probabilistic dependence between hypotheses and their auxiliaries in the dataset does not, however,

preclude the possibility of confirming/disconfirming hypotheses or auxiliaries without falling back either on their own specific auxiliaries or hypotheses in question.

Consider how this paradigm example shows that the two sub varieties of CH are false. CH1 says: if H1 along with A has been confirmed together, then one should never be able to confirm H1 and A separately. This is how a supporter of this brand of confirmation holism might argue.

1. $(H1 \, \& \, A) \rightarrow D$.
2. D. Therefore,
   $(H1 \, \& \, A)$.

Here "D" stands for the association between smoking and LC. Both Tables II and III support D, because the latter shows that there is association between smoking and LC. Since confirmation relation is the inverse of the deductive consequence relation, a confirmation holist$_1$ contends, the data provide confirmation for (H1 & A) together.

We argue, however, that although in the antecedent, H1 along with A could be confirmed together, in the consequent, H1 or A could be confirmed separately. Let's assume the antecedent. Some past and present celebrated observational studies on smoking and lung cancer have confirmed H1 strongly independent of A. Observational studies involve data that are not gathered by controlled experiments. The latter situation arises if the investigator is able to impose treatments to observe their effects on subjects or to assign subjects at random to different procedures. The purpose of observational studies is usually to understand the relationships between two variables in which it is not feasible to use controlled experiments. After reviewing a vast literature, the 1964 Surgeon General's Advisory Committee Report concluded:[17]

"Cigarette smoking is causally related to lung cancer in men; the magnitude of the effect of cigarette smoking *far outweighs all other factors*. The data for women, though less extensive, point in the same direction."

Based on those observational studies and the like, even the Tobacco industries have recently conceded that there is a weak link between smoking and lung cancer, and settled billions of dollars of law-suits with their consumers. Therefore, to confirm H1 one does not need to fall back on its auxiliary because there is independent research data (D) to support H1, i.e., $P(H1|D) > P(H1)$, thus exposing the pitfall of CH1. One point needs to be made clear. Although it is evident that for confirming H1 we exploit some theory/background information as correct, unlike confirmation holism it does not follow that for confirming H1 we must fall back on all of the hypotheses and their auxiliaries that ultimately constitute science at a given point or auxiliaries used along with H1 to derive observable consequences from the latter. For similar reasons, CH2 fairs as well.

According to CH2, if a hypothesis is disconfirmed along with its auxiliaries as a whole, then one should never be able to disconfirm a hypothesis or its auxiliaries in isolation. This is how a confirmation holist$_2$ might defend her position.

1. $(H2 \ \& \ A) \rightarrow \sim D$.
2. D. Therefore,
   $\sim (H2 \ \& \ A)$.

The data "D" in 2 are true because Table II shows that there is association between smoking and LC. Hence, it seems that H2 (i.e., ~H1) along with A has been disconfirmed together by the data. For the sake of the argument, let's assume the antecedent of CH2. We will argue, however, that the consequent of CH2 does not follow. That is, we will disconfirm either H2 or A independent of one another. In fact, we have already cited observational studies showing that H2 (i.e, ~H1), that is, there is no causal relation between smoking and LC, is very likely to be false independent of A. We argue that A is very likely to be false given Tables III and IV, because in SLC the rate of LC is higher

both among smokers and non-smokers compared to the rate of LC in B. In SLC among smokers, the rate is 16% higher and among non-smokers, the rate is 10% higher. The falsity of A has been shown independent of assuming the truth/falsity of H1 or H2. Thus, both versions of confirmation holism, confirmation holism$_1$ and confirmation holism$_2$, have turned out to be untenable.

One possible objection to our resolution of confirmation holism is that we have in fact defended confirmation holism instead of rejecting it by our own admission. The objection continues if our argument for objecting to confirmation holism rests on another auxiliaries or theories different from the theory or auxiliary in question then we have in fact presupposed some existing theories or auxiliaries to be correct providing a clear defense for confirmation holism. As a result, confirmation holism$_1$ at the hand of the objector, now turns out to be the theory according to which if a hypothesis, H, as a whole along with auxiliaries, A, has been confirmed together, then one should never be able to confirm H and A separately without falling back on some other theory or auxiliaries. On this *new* construal of confirmation holism, it becomes obvious that there is no way confirmation holism$_1$ would be falsifiable. For confirmation/disconfirmation of a theory or an auxiliary, we have to fall back on our background knowledge containing some theories or auxiliaries and there is no escape from exploiting background knowledge for confirmation/disconfirmation of a theory or its auxiliary when we reject any of the versions of confirmation holism. This position of confirmation holism being unfalsifiable is similar to the position of some form of psychological egoism which has been criticized as being irrefutable. According to psychological egoism, human beings are so constituted by nature that they can't but pursue their own well-being. According to this version of

psychological egoism, whatever we do, good or bad in some sense, we always support this theory by our action, and hence, it does not seem that we are ever be able to reject the theory by any means. However, it is well-known by now that this form of psychological egoism is irrefutable, and that it is irrefutable is not counted as a merit for psychological egoism.[18] In the same vein, under this new construal of confirmation holism$_1$, the latter becomes irrefutable, which should not, however, be taken as a merit for the theory. Therefore, the objection to our resolution of confirmation holism that we make use some background information other than theories or auxiliaries in question is correct, but the objection, in turn, makes confirmation holism irrefutable, which is not a desirable feature of any theory.

One might wonder that we are presumably uncharitable toward confirmation holism. Either we have proposed a version which is liable to an easy refutation or formulated a version which is irrefutable. The expectation of this worry is that perhaps there is a version of confirmation holism which is just right not too weak to be refuted easily and not too strong which is irrefutable. It is not logically impossible to have such a version. However, the onus is on the confirmation holist to come up with a "Goldilocks" variety of confirmation which will fill the bill. Unless we see such a version we conclude the section that confirmation holism is untenable.


## 5.    Problems of Evidence Holism

For the sake of simplicity, we take two hypotheses (H3, H4) to be mutually exclusive and jointly exhaustive. Then, rejecting a true hypothesis H3 implies accepting a false hypothesis ~H3 (i.e., H4) and its converse is also true. In this section, we will

evaluate both forms of evidence holism. Consider first the Evidence-Ignoring Holism (EIH): If a sentence would otherwise be rejected due to recalcitrant data, one could nevertheless retain that sentence come what may, provided one makes enough adjustments inside the entire system of sentences to square them with the data.

According to Quine, one could hold on to H4 even when there is recalcitrant evidence against it by pointing to that evidence as, say, erroneous or pleading hallucinations. Although Quine casts his holism in deductive terms, since it is a general thesis about theory confirmation/disconfirmation, if it were correct, it should hold across the board both in deductive and in probabilistic reasoning. In probabilistic terms, Quine's recommendation to save H4 from refutation could be taken to mean that recalcitrant evidence is misleading implying that it would be fine to overlook it. By "misleading evidence" epistemologists mean "evidence against a truth."[19]

On a probabilistic account of evidence, evidence could contain some errors; and in this sense, could be misleading. Consider an example involving tuberculosis in which H3 represents the hypothesis that an individual is suffering from tuberculosis and ~H3 (i.e., H4) the hypothesis that she is not.[20] Assume D represents a positive X-ray test result. Let's assume that the following table provides a summary of test results for 100,000 people with the probability for having a positive X-ray for members with or without tuberculosis.

|  | Positive | Negative |
|---|---|---|
| Disease present | 0.7333 | 0.2667 |
| Disease is absent | 0.0285 | 0.9715 |

**Table V: X-RAY RESULTS**

Take "A" for the auxiliary assumption that the instrument measuring the positive and negative results is reliable. "A" could vary from one context to another. In this connection, "A" stands for the reliability of the X-ray instrument. The observed result is an example of evidence. The LR = 0.733/0.0285 (i.e., P(D & A|H3)/P(D& A|H4)) ≈ 26 constitutes strong evidence that the subject has tuberculosis. This inference based on the data is correct regardless of whether the subject really has the disease. However, this observation of the positive test result could itself be misleading even though there is strong evidence for the hypothesis. If the disease is absent, Table V shows that the probability of a misleading positive test is only 0.0285. The latter number represents the probability of observing misleading evidence for the correctness of the hypothesis (even when the likelihood ratio is that strong, i.e., 26 times), because the number implies that disease is present, although the disease is in fact absent. However, when we have a sample of that size (100,000 people), as the evidence becomes strong, fortunately the probability of misleading evidence for the LR that strong becomes low. The last point could be explained by a comparison between two examples, the tuberculosis case and what we call the PAP case. The PAP smear test is used widely to check whether a woman of a varying age group suffers from cervical cancer. H5 represents that the subject in question actually has the disease, and ~H5 (i.e., H6) represents she does not actually have the disease. Let's assume that the following provides a summary of test results for 100,000 people where D represents a positive test result. [21]

|  | Positive | Negative |
| --- | --- | --- |
| Disease present | 0.8375 | 0.1625 |
| Disease is absent | 0.1864 | 0.8136 |

**Table VI: PAP Test Results**

Take "A" for the auxiliary that the Pap smear instrument is reliable. Given that the test comes out positive, LR = P(D& A|H5)/P(D&A|H6) = 0.8375/0.1864 = 4.49. This shows that evidence is weak for the presence of the disease, while the probability of a misleading positive test is 0.1864 (when the LR is not strong, just 5 times) if the disease is not present. When the last two tables, the TB and PAP results, are compared, one finds that if the evidence is strong, i.e., the LR is that strong for the presence of the disease as in the case of TB case, the probability of a misleading positive test is lower i.e., 0.0285, compared to when the evidence is weak, i.e., the LR is weak for the presence of the disease as in the case of PAP case. These examples suggest that when we have a sample of that size (100,000 people), as evidence becomes strong, the probability of misleading evidence for the LR that strong in favor of the hypothesis, decreases.

What is the bearing of this discussion on two forms of holism, (i) evidence-ignoring holism (EIH), and (ii) evidence revisionary holism (ERH)? EIH says that one could ignore data and it is perfectly justified if one does so, albeit they are in favor of the correct hypothesis, whereas ERH says that one could reject a correct theory if one makes enough adjustments within one's belief-system to fit the data. Both forms of the holism are like two sides of the same coin. Consequently, rejection/acceptance of one implies the rejection/acceptance of the other and the converse is also true. In light of the tuberculosis example, consider a scientist who adopts the policy held by the EIH. The latter could contend that there is a justification for ignoring strong evidence in the tuberculosis example because if tuberculosis were really absent (i.e., ~H3) then the data should not

have showed a probability of observing the presence of tuberculosis.[22] However, it does show up which is 0.0285 even for the LR that strong in favor of the correct tuberculosis hypothesis. In the PAP case, an adherent of the EIH could offer the same sort of justification. She could ignore weak evidence for the presence of cervical cancer by pointing out that there is a probability of observing misleading evidence if the disease is absent even when the LR is that weak for the correct hypothesis. As we know, that probability of misleading evidence in the PAP case is 0.1864. However, on a closer inspection, this enthusiasm for EIH disappears because of a theorem (Royall, 2000).

The theorem states that when the sample size is sufficiently large so that the probability of strong evidence is reasonably large for the correct hypothesis, the probability of getting strong evidence that is misleading for the correct hypothesis will become very small. Therefore, although the claim made by EIH, that is, one could ignore evidence, weak or strong, for a hypothesis over its rival by pleading it to be misleading, holds in a small sample size, it does not hold as the sample size increases adequately. As the sample becomes sufficiently large, the probability of strong evidence for the correct hypothesis approaches one whereas, the probability for observing misleading evidence approaches zero. Therefore, there is no sound probabilistic justification for holding onto EIH.

Consider now Evidence-Revisionary Holism (ERH): If a sentence is otherwise held to be true, then one could reject that sentence if one would like to do so, provided one makes enough adjustments inside the entire system of sentences to square them with the data. The motivation for ERH draws its strength from Quine's *revisionary* claim that

"no statement is immune to revision." If one is careful to appreciate the drift of his pragmatism, then one realizes that perhaps Quine has in mind something similar to ERH.

One could find pitfalls in ERH if one looks at it in its dynamic scenario. In this scenario, it is logically possible to think that an investigator has been continuously bombarded by the data against her pet hypothesis and she has been constantly revising her entire system in accordance with the ripple effects of new contrary evidence. As an evidence-revisionary holist, she could refuse to reject her pet false hypothesis, H4 (i.e., the person in question does not have tuberculosis), when its alternative H3 (i.e., the person in question does have tuberculosis) is true. She could adopt two protocols to save her theory from refutation by recalcitrant experiments. The first is that she calculates the likelihood ratio (LR) after each new observation. If the LR favors H4 (i.e., the false theory) by a factor of k, then she stops the experiment. If not, then she adopts the second protocol. Under this second protocol, she keeps on doing more and more experiments to support H4, although the alternative, H3, is true. There are two dire mathematical results that cast doubt on the legitimacy of ERH. The first one is that if H3 is true, then the probability is at least $1 - 1/k$ that the *experiment* will *never end,* implying that there is some probability that the scientist will never be able to show that her false theory is well-supported. [23] The second result which follows from the first result is that if H3 is true, then the probability of the accumulated evidence to provide strong evidence for H4 over H3 will never exceed the universal bound for the probability of observing misleading evidence, 1/k, where k indicates the strength of evidence.[24] Since ours is a non-standard approach to the Duhem-Quine thesis, we will take extra care to spell out the consequence of the second mathematical result intuitively. An investigator inspired by ERH could try

to defend the falsity of the tuberculosis hypothesis (i.e., the person in question does have tuberculosis), albeit it is the correct hypothesis, by carrying out experiments indefinitely to provide strong evidence for her false hypothesis (i.e., the person does *not* have tuberculosis) However, her experiments to be able to produce the probability of misleading evidence for the correct hypothesis that the person in question has tuberculosis would never exceed the bound 1/k. In the tuberculosis case, it is $1/26 \approx 0.038$ which is slightly higher than 0.0285, which, recall, is the probability of misleading evidence for the tuberculosis hypothesis for LR that strong when the sample size is 100,000.

So there are problems for both versions of evidence holism. If one is an evidence-ignoring holist, then, for the time being in the case of small sample size, she could enjoy the luxury of ignoring evidence supporting one hypothesis over its rival by contending that evidence is misleading because there is obviously a probability of observing misleading evidence even when the LR is that strong or weak in favor of the so-called correct hypothesis. However, as the sample size becomes sufficiently large, the probability of strong evidence for the correct hypothesis tends to one. These two factors, (i) the sample size becoming sufficiently large and (ii) the probability of the strong evidence for the correct hypothesis going to one, will force the probability of observing misleading evidence for the correct hypothesis to tend to zero. The situation for an evidence-revisionary holist is also worse. Please be reminded that this theorem-based argument works against both incorrect hypotheses and auxiliaries when they are being held against an increasing amount of evidence against either of them. We contend that although an investigator could be dogmatic regarding holding onto a false hypothesis or

auxiliaries while making the entire system fit the data, the price the investigator will pay for her dogmatism is purely epistemological, (since the probability of misleading evidence is epistemological) and there is nothing pragmatic about the error she commits in being an evidence-ignoring holist or an evidence-revisionary holist. Hence, EH along with CH are both false in both their versions purely based on epistemological considerations.

## 6. Summing Up

The Duhem-Quine thesis construes the practice of science in an initially plausible but ultimately defective way. If it were correct, when an investigator needs to revise a part of her hypothesis due to recalcitrant experimental data, she would be unable to do so without affecting the hypothesis globally along with its auxiliaries. However, in science we often do those local surgeries inside the hypothesis to incorporate impacts of significant experimental data without affecting its auxiliaries. Is it possible to propose an approach that would be epistemologically viable, yet compatible with the practice of science?

In response to this, we have proposed two Bayesian accounts; an account of confirmation and an account of evidence, to address two forms of holism; confirmation holism and evidence holism.[25] We further established that two varieties of holism-confirmation and evidence- are often run together in statements of "the Duhem-Quine thesis". Although our underlying theory is Bayesian, its application does not depend on an agent's subjective prior probability to influence evaluation of confirmation holism.[26] We have not assumed that the subjective probability of the hypothesis that smoking is a causal factor of lung cancer to be high, although it is evident that background information

20

in the form of observational studies has been brought to bear on adjudicating the hypothesis insofar as confirmation holism is concerned. However, this background information could be anybody's background information, not up for grabs like an agent's subjective prior probability. In addition, unlike most solutions to the Duhem-Quine problem, we don't assume that the hypothesis and its auxiliaries have to be probabilistically independent. As a result, it makes the Duhem-Quine problem more realistic than it has usually been. We also challenged two versions of evidence holism unnoticed in the literature. Based on purely epistemological considerations, we conclude that the Duhem-Quine holism should be rejected en bloc including confirmation holism and evidence holism together, thus narrowing the gap between philosophy of science and the actual practice of science in which hypotheses and auxiliaries could be selectively tested.

## Select Bibliography

P. Achinstein (1983): *The Concept of Evidence*. (ed.) Oxford University Press, Oxford.

J. Berger. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York. (Second edition.).

J. Bernardo, and A. Smith (2000): *Bayesian Theory*. Wiley, New York.

L. Bovens, L., and S. Hartmann (2003): *Bayesian Epistemology*. Oxford. Oxford University Press.

W. Cohran (1964): "The Planning of Observational Studies of Human Populations." (With Discussion). In *Journal of the Royal Statistical Society*, Series A, 128, pp.134-155

E. Conee (2004): "Heeding Misleading Evidence." In *Evidentialism*. (With Richard Feldman.) Oxford University Press, Oxford.

D. Gillies (1993): *Philosophy of Science in the Twentieth Century*. Blackwell, Oxford.

J. Dorling, (1979): "Bayesian Personalism, the Methodology of Scientific Research Programs, and the Duhem's Problem." In *Studies in the History and Philosophy of Science*, x, pp.177-187.

P. Duhem (1954): *The Aim and Structure of Physical Theory*. Princeton University Press, Princeton.

J. Earman, (2002): *Bayes or Bust?* MIT Press, Cambridge.

J. Feinberg, "Psychological Egoism." In *Reason and Responsibility*: Readings in Some Basic Problems of Philosophy, Ninth edition, ed. J. Feinberg, Wadsworth Publishing Company, 1996, pp.497-508.

C. Glymour, (1980): *Theory and Evidence*. Princeton University Press. Princeton.

C. Howson, and P. Urbach (2006): *Scientific Reasoning*: *The Bayesian Approach*. Open Court, La Salle, Illinois.

P. Kitcher (2001): *Science, Truth, and Democracy*. Oxford University Press. Oxford.

L. Laudan (1990): "Demystifying Underdetermiantion." In C. W. Savage, ed., *Scientific Theories*, vol. 14, *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press.), pp.267-297.

D. Mayo (1996): *Error and the Growth of Experimental Knowledge*. University of Chicago Press. Chicago.

M. Pagano, and K. Gauvreau (2000): *Principles of Biostatistics*. Duxbury. Australia.

Poirier, D. (1995). *Intermediate Statistics and Econometrics*. MIT Press. Cambridge. Mass.

W. Quine (1953): "Two Dogmas of Empiricism." In *From a Logical Point of View*. Harvard University Press, Cambridge.

W. Quine (1970): *The Philosophy of Logic*. Prentice-Hall, Englewood Cliffs.

H. Robbins (1970): "Statistical Methods Related to the law of the Iterated Logarithm." In *The Annals of Mathematical Statistics*, vol. 41, no. 5 (October), pp.1397-1409.

P. Rosenbaum (2002): *Observational Studies*. Springer. New York

R. Royall (2004): "The Likelihood Paradigm for Statistical Evidence." In *The Nature of Scientific Evidence*. M.L. Taper and S. Lele (eds.) University of Chicago Press, Chicago.

R. Royall (2000): 'On the Probability of Misleading Statistical Evidence." In *JASA*, Vol. 95, no. 451, pp.760-768.

W. Salmon (1983): "Confirmation and Relevance." In *The Concept of Evidence*. P. Achinstein (ed.) Oxford University Press, Oxford.

B. Skyrms, and K. Lambert (1995): 'The Middle Ground: Resiliency and Laws in the Web of Belief." In *Laws of Nature: Essays on the Philosophical, Scientific, and Historical Dimensions*. F. Weinert, (ed.) Walter de Gruyter, Berlin, pp. 139-156.

Sober, E (2004): "Likelihood, Model-Selection, and the Duhem-Quine Problem." In *Journal of Philosophy*. Vol. CI, no. 5, May, pp 221-241.

A. Wald (1947): *Sequential Analysis*. John Wiley. New York.

[2] W. V. Quine, "Two Dogmas of Empiricism," in *From a Logical Point of View* (Cambridge: Harvard University Press, 1953), pp.20-46, and *Philosophy of Logic* (Englewood Cliffs: NJ: Prentice-Hall, 1970). P. Duhem, *The Aim and Structure of Physical Theory* (Princeton: Princeton University Press, 1954). Duhem's book was originally published in French in 1905.

[3] It is not perhaps hard to state the source of this identification. Quine's classic paper has generated this identification followed by several papers including a well-cited recent paper by Laudan (See Laudan, 1990). Gillies, however, thinks that Duhem's holism needs to be distinguished from Quine's holism, because, he thinks, the former has a narrow scope than the latter (See Gillies, 1993).

[4] Quine (1953).

[5] This distinction between the belief and evidence questions is due to Richard Royall. This paper owes a great deal to his work. See "The Likelihood Paradigm for Statistical Evidence" in M.L. Taper, and S. Lele (eds.) *The Nature of Scientific Evidence*. (Chicago: University of Chicago Press, 2004), pp.119-152.

[6] Kitcher is one philosopher who is critical about the prospect of such an account of evidence. See, Kitcher, "The Idea of Objectivity," in *Science, Truth, and Democracy*, Oxford, 2001.

[7] A Bayesian of our kind should very well be able to construe an evidence relation to be independent of what the agent believes about those two competing hypothesis. The present day subjective Bayesian-dominated-philosophy of science is, however, slow to appreciate this point. For a subjective Bayesian, since the evidence relation is defined in terms of a ratio of two probabilities, those probabilities have to be understood in terms of an agent's degree of belief. Therefore, according to her, the evidence relation must be subjective. However the situation need not be like that. A subjective Bayesian favored text-book on econometrics explains this point well and discusses how an objective Bayesian could approach the likelihood function which is at the core of our evidential account. The author writes, "[t]he likelihood function was simply the *data generating process* (DGP) viewed as a function of the unknown parameters, and it was thought of having an "objective existence" in the sense of a meaningful repetitive experiment. The parameters themselves, according to this view, also had an objective existence independent of the DGP."… It is possible to adopt the so-called *objective Bayesian* viewpoint in which the parameters and the likelihood maintain their objective existence…" (Poirier, 1995, pp.288-9)

[8] We owe this point to Robert Boik. $\text{Prob}(H|D) > \text{Prob}(H) \leftrightarrow LR > 1$.
Proof:
$\text{Prob}(H|D) > \text{Prob}(H) \leftrightarrow [\text{Prob}(D|H) \text{Prob}(H)]/ \text{Prob}(D)] > \text{Prob}(H) \leftrightarrow [\text{Prob}(D|H) > \text{Prob}(D)] \leftrightarrow \text{Prob}(D|H) > \{[\text{Prob}(D|H) \text{Prob}(H)] +[\text{Prob}(D| \sim H) \text{Prob}(\sim H)]\} \leftrightarrow 1 >[ \text{Prob}(H) + \text{Prob}(\sim H)/LR] \leftrightarrow (1-\text{Prob}(H)) > ( 1-\text{Prob}(H))/LR \leftrightarrow LR > 1$.

[9] We owe some of our ideas to W. Salmon, "Confirmation and Relevance," in *The Concept of Evidence*. Achinstein, P. ed. (Oxford: Oxford University Press, 1983), pp. 95-123. For selective confirmation/bootstrapping, see C. Glymour, *Theory and Evidence*. (Princeton University Press, Princeton). For a different take on the relationships among networks and sub-networks of belief in connection with Quine's "Two Dogmas" see Skyrms and Lambert's subjective probability approach. (Skyrms and Lambert, "The Middle Ground: Resiliency and Laws in the Web of Belief." In *Laws of Nature*: *Essays on the Philosophical, Scientific and Historical Dimensions*. (ed.) F. Weinert, Walter de Gruyter: Berlin, 1995, pp. 139-156.)

[10] W.V. Quine (1953).

[11] This name has been suggested to us by Elliott Sober in an email communication.

[12] This example has been suggested to us by the statistician Martin Hamilton. The following data are made-up. However, they are realistic enough to expect to recur in various places including any two hospitals like the above. This example is known as an instance of Simpson's paradox in the literature. There are several version of this paradox. The instance of the paradox we are using in this section arises when the association we observe between two variables, "smoking" and "lung cancer" in the subpopulations ceases to exist when the data are pooled together to form the entire population.

[13] See J. Bernardo and A. Smith *Bayesian Theory*, Wiley, New York, 2000 for more on the coherence requirement for Bayesians.

[14] *Almost* all discussions on the Duhem-Quine thesis including both Bayesian and non-Bayesians' assumed auxiliaries and hypotheses to be independent. Here are some samplings from Bayesians. See J. Dorling, "Bayesian Personalism, the Methodology of Scientific Research Programs, and the Duhem's Problem," *Studies in the History and Philosophy of Science,* x (1979), pp. 177-187. Both C. Howson and P. Urbach also began with Dorling's article about the assumption and added their spin on the issue. See. C. Howson, and P. Urbach: *Scientific Reasoning*: *The Bayesian Approach*, (La Salle, IL:

Open Court, 2006, 3ʳᵈ edition). A section of R. Jeffrey's latest book was also devoted to a discussion of Dorling's work with the same assumption (R. Jeffrey, *Subjective Probability*, Cambridge: Cambridge University Press, 2004). Like other Bayesians, J. Earman began with Dorling's solution while addressing the Duhem-Quine thesis. However, he was careful to observe that the assumption, i.e., the hypothesis and auxiliaries need to be independent for its solution, is "unrealistic." J. Earman, *Bayes or Bust?* (Cambridge: MIT, 1992). Here are some samplings from non-Bayesians. D. Mayo, *Error and the Growth of Experimental Knowledge* (Chicago, Chicago University Press, 1996). See also E. Sober, "Likelihood, Model Selection, and the Duhem-Quine Problem," in *Journal of Philosophy*. Vol. no. 5, May, 2004. , pp. 221-241. The only exception to this general stance toward it seems to be Bovens and Hartman's recent work. (See, L. Bovens and S. Hartmann, *Bayesian Epistemology*, Oxford: Oxford University Press, 2003). Howson and Urbach made a comment about Bovens and Hartmann's work for relaxing this standard assumption in their book referred to above.

[15] The rule of the probability calculus in point is [P(H1|A) ≠ P(H1)] just in case [P(~H1|~A) ≠ P(~H)].

[16] See footnote 13 for some recent treatments of the thesis.

[17] W. G. Cochran, "The Planning of Observational Studies of Human Populations (with Discussion). *Journal of the Royal Statistical Society*, Series A, 128, pp.134-155. (1964). Emphasis is ours. This passage has also been cited by P. Rosenbaum, *Observational Studies* (Springer, New York: 2002, 2ⁿᵈ edition.).

[18] J. Feinberg, "Psychological Egoism." In *Reason and Responsibility*: Readings in Some Basic Problems of Philosophy, Ninth edition, ed. J. Feinberg, Wadsworth Publishing Company, 1996, pp.497-508.

[19] This is the meaning E. Conee has attached to the term. E. Conee, "Heeding Misleading Evidence," in *Evidentialism*. (E. Conee and R. Feldman, Oxford University Press. 2004, New York, pp.259-276.)

[20] Recall that one assumption of the paper is that all hypotheses are simple statistical hypotheses including the tuberculosis hypothesis including what we call the "cervical cencer hypotheses". Deborah Mayo worries that this tuberculosis example can't be a simple statistical hypothesis because, she thinks, a possession of a specific feature does not necessarily make the tuberculosis hypothesis true (in a private conversation.). If H were not a simple statistical hypothesis then it would have been a complex hypothesis. It is well-known that a complex hypothesis is a disjunction of mutually exclusive and jointly exhaustive simple statistical hypotheses with prior weights on each of them. It would further follow from this is that those weights are an agent's subjective prior probabilities infecting the belief/evidence distinction. This was the motivation behind Mayo's worry for calling our tuberculosis hypothesis as not a simple hypothesis. In consultation with an expert on the subject, we gladly report that it is perfectly legitimate to treat H to signify a person to suffer from tuberculosis and thus to treat H as a simple statistical hypothesis. Tuberculosis is caused by a bacterium called mycobacterium tuberculosis. If one is affected by mycobacterium tuberculosis, then one has tuberculosis and converse is also true. If one does not have that specific bacterium then one does not have that disease. However, the case would be different if we would have treated H as the person suffering from cancer. Since no single cause can be found for any kind of cancer it won't be right to treat H as a simple statistical hypothesis in the case of cancer. We owe this point to Mujib Rahman. We thank both Mayo and Rahman for helping us to appreciate the significance for taking the tuberculosis hypothesis as a simple hypothesis.

[21] M. Pagano and K. Gauvreau. *Principles of Biostatistics*. Duxbury: Australia. (00)

[22] Obviously, the motivation of this argument derives its strength from deductive inference in which the argument is an all or nothing affair, and not from probabilistic reasoning in which although one might have very strong evidence for a theory, yet the theory could very well be false.

[23] See Royall (2000)'s paper that has already been cited. Here he is referring to Robbins' work referred to above.

[24] Royall, R. (2000): "On the Probability of Misleading Statistical Evidence." In *JASA*. Vol. 95, no. 451.pp.760-768. In fact, for the mathematical result to hold, a sequence of observations need not be independent. Here Royall has referred to H. Robbins' "Statistical Methods Related to the Law of the Iterated Logarithm," in Robbins, *The Annals of Mathematical Statistics*, vol. 41, no. 5. (Oct, 1970), pp. 1397-1409.

[25] Commenting on an allied paper of ours, Elliott Sober observes that our account is more like a "strengthened likelihood" position (in an email communication). It is hard to know what he really means by the strengthened likelihoodism. On one sense, our position could be regarded as a form of likelihood postion if we would endorse priors which are only frequency-based, in addition to accepting the likelihood framework for addressing the evidence question. However, as our two accouts, an account of confirmation and an account of evidence, make it explicit that our account of confirmation allows both frequency-based as well as an agent's degree of belief-based priors as rational. As a result, our Bayesian position should not be called a strengthened likelihood postion. Our postion will be reduced to such a strengthened likelihood position if our priors are only frequency-based together with the likelihood framework for addressing the evidence question. We would like to thank Sober for forcing us to make our Bayesian stance clear.

[26] Both a Bayesian like Earman (Earman, 1992) and a non-Bayesian like Sober (Sober, 2004) worry that a Bayesian solution to the Duhem-Quine problem must ultimately fall back on an agent's subjective prior probability for its solution. However, this paper shows why this is not necessarily the case although it has proposed a Bayesian solution.