Synthese (2020) 197:337–354
https://doi.org/10.1007/s11229-018-1845-1

Duhem's problem revisited: logical versus epistemic formulations and solutions

Michael Dietrich[1] · **Phillip Honenberger**[1]

Received: 2 October 2017 / Accepted: 5 June 2018 / Published online: 4 July 2018

Abstract When the results of an experiment appears to disconfirm a hypothesis, how does one know whether it's the hypothesis, or rather some auxiliary hypothesis or assumption, that is at fault? Philosophers' answers to this question, now known as "Duhem's problem," have differed widely. Despite these differences, we affirm Duhem's original position that the logical structure of this problem alone does not allow a solution. A survey of philosophical approaches to Duhem's problem indicates that what allows any philosopher, or scientists for that matter, to solve this problem is the addition of epistemic information that guides their assignment of praise and blame after a negative test. We therefore advocate a distinction between the logical and epistemic formulations of Duhem's problem, the latter relying upon additional relevant information about the system being tested. Recognition of the role of this additional information suggests that some proposed solutions to the epistemic form of Duhem's problem are preferable over others.

Keywords Duhem's problem · Bayesian confirmation theory · Error statistics · Darden, Lindley · Sober, Elliott · Inference to the best explanation

1 Introduction

The canonical philosophy-of-science puzzle known as "Duhem's problem" (Duhem 1914/1974) might be described at its most general level as follows: how do H&A $\models$ ~ E,

Michael Dietrich and Phillip Honenberger have contributed equally to this paper.

✉ Phillip Honenberger
philliphonenberger@gmail.com

[1] Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, USA

and E, taken together, affect the respective confirmation values of H and of A, where H is the hypothesis under test, A is an auxiliary hypothesis, and E is an experimental result? As Duhem famously described the problem in *La Théorie Physique*: "[T]he physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one ought to be changed" (Duhem 1914/1954, p. 187). The question, in short, is which hypothesis or assumption, among those that might be taken to be disconfirmed by a negative test result, ought to be taken as disconfirmed. In other words, where should we place the blame for a failed test? What is now called "Duhem's problem" is just the difficulty of satisfactorily answering that question.

The general formulation given above implemented a few simplifications for the sake of readability. For instance, the formulation is deductive rather than probabilistic, but one might weaken the reading of '$\models$' from 'entails' to 'supports,' and assign probabilities between 0 and 1 to H, A, and E. Likewise, it assumes a contest between just two propositions, H and A, whereas real scientific theories are more complex and thus better represented by larger conjunctions (e.g. H&A1&A2&A3…), any one of which (or some combination of which) might be taken to be disconfirmed by E, if H&A1&A2… $\models$ ~E. Indeed, in light of work on philosophy of experiment, it is important to recognize that generating experimental results also involves many assumptions, which may be singled out as sources of error in a scientific test. These expansions of the elements relevant to a scientific test add complexity, but do not alter the logical form of Duhem's problem. After a test in which the prediction and empirical evidence do not agree, the scientist must choose where to place blame from the hypothesis and all of the relevant assumptions involved in the test.

Duhem's assertion that the logical structure of the test does not indicate where to place blame remains regardless of how many additional assumptions we recognize. Without additional information about the epistemic value of elements in the theoretical system (the hypothesis and assumptions), there is no way to squeeze unequal degrees of confirmation or disconfirmation for H and A out of the disagreement between prediction and empirical result. We refer to this as the logical formulation of Duhem's problem.

Duhem's problem and its possible solutions change dramatically, however, if one admits additional information that bears on the epistemic value of any of the assumptions or hypothesis in question. Additional epistemic information about any of the elements involved in the test add structure to the case described, changing it from a case where all of the assumptions and the hypothesis have equal epistemic status (epistemic equivalence) to a case where some of the assumptions and/or the hypothesis have differential epistemic weights. This additional epistemically relevant information can make it possible to justify greater suspicion of a particular assumption, for instance. These more epistemically informed cases should be distinguished from the logical formulation of Duhem's problem: they constitute a different kind of problem, what we call the epistemic form of Duhem's problem, and they allow different kinds of solution, including those offered by scientists daily.

Duhem's problem has been discussed by many philosophers since Duhem introduced it (for instance: Dorling 1979; Darden 1992; Mayo 1997; Strevens 2001; Sober 2004; and Weber 2009). Their proposed solutions have differed widely. Here we survey these efforts and show a recurrent pattern wherein most philosophers import additional information in their proposed solutions. These solutions do not address the logical form of the problem and should rather be understood and evaluated as solutions to the epistemic form of the problem. A principle criterion for evaluating solutions to the epistemic form of Duhem's problem, we argue, should be how well that solution accommodates a range of relevant epistemic information.

2 Review of previous solutions

2.1 Duhem's "good sense" solution and recent revivals

Duhem's own solution to the problem that bears his name centered on what he called "good sense" (*bon sens*), which he described as "motives which do not proceed from logic and yet direct our choices"; as "reasons which reason does not know" (quoting Pascal); and as coming from the "supple mind" (*l'ésprit de finesse*) rather than the geometrical mind (*l'ésprit geometrique*), borrowing one of Pascal's key distinctions (1914/1954, p. 217).

In *La théorie physique*, Duhem reminds us that logic alone will not decide between the alternative available responses to a disconfirmatory experiment. This means some scientists may respond one way, while others respond differently, and "logic" alone won't settle which view is correct (1914/1954, pp. 216–217). This doesn't mean "we [*l'on*] cannot very properly prefer the work of one of the two to that of the other" (217). What enables us to have this preference is "good sense." Duhem considers an historical example involving competition between rival hypotheses. Eventually, Duhem claims, good sense will favor one side or the other in such disputes. Recognizing such moments is itself a capacity of good sense. Scientists can thus try to hasten the resolution, and thus the progress of science itself, by developing good sense within themselves (218).

The main problem with "good sense" as a solution to Duhem's problem is that it fails to offer any criteria for choice that is justified on epistemic grounds (that is, grounds we have reason to believe favor epistemic values such as truthfulness or correspondence with present and future evidence). This problem is highlighted by asking of Duhem's texts: "What kind of faculty is good sense? What is its content, and how does it work?" In *La théorie physique*, Duhem tells us that "impartiality and faithfulness" are among the characteristics of good sense (1914/1954, p. 218). In his discussion of historical research in *German Science*, he writes that "pursuit of truth not only requires intellectual abilities, but also calls for moral qualities: rectitude, probity, detachment of all interests and all passions" (1915/1991, p. 43). And in "Physics of a Believer" (1905) he writes (without using the phrase "good sense") of the resolution of Duhem's-problem situations as following "considerations of elegance, simplicity, and convenience, and grounds of suitability which are essentially subjective, contingent, and variable with time, with schools, and with persons" (1914/1954, p. 288).

However, such virtues do not supply univocal and epistemically justified solutions to Duhem's-problem-type situations. Disinterestedness, for instance, does not solve the problem: wouldn't a disinterested reasoner, faced with the standoff expressed in the general formulation of the problem, suspend judgment rather than choose? Likewise, the variable "grounds of suitability" appealed to in the "Physics of a Believer" passage (which are "subjective, contingent, and variable") could as well lead to opposed judgments as convergent ones. "Elegance" and "simplicity," on the other hand, demand further epistemic justification before they should be accepted as guidelines for solving Duhem's problem in all cases, in addition to ambiguities in the precise content of these criteria and how to apply them. And "probity" is virtually a synonym for "capacity to get things right," which begs the question. Suppose on the other hand that, rather than appeal to substantive guidelines for how scientists' exhibiting "good sense" approach their problems, we rest (as Duhem sometimes appears to do) on the vaguer description of good sense as simply a kind of intuition—a cognitive "direct access" to solid interpretations of the evidence. Then we are faced with the difficult question of how to tell when we ourselves, or others, are in possession of this coveted faculty.

Recent commentators have aimed to revive the "good sense" notion as a solution to Duhem's problem (Stump 2007, Ivanova 2010, Fairweather 2012, Ivanova and Paternotte 2013). While these discussions have usefully illuminated various relevancies of agency, social context, and the history of science to resolutions of Duhem's problem, the revivals ultimately suffer the same weakness as Duhem's original proposal: lack of epistemically justified concrete guidance in the face of Duhem's-problem-type situations.

Some recent revivals of the good sense notion have sought to interpret the concept as an instance of virtue epistemology (Stump 2007, Fairweather 2012). This approach involves enriching the account of the cognitive ability of good sense itself and then arguing that this ability grounds the epistemic preferability of the research decisions that scientists exercising good sense (under this description) make. To this end, Stump (2007) and others have collected those passages wherein Duhem describes what having good sense is like. However, as we have seen in our own collection of many of those passages above, the descriptions do not provide an effective procedure for revolving Duhem's-problem-type situations. Similarly, Ivanova (2010) proposes that scientists use the history of science to construct the notion of an "ideal scientist," which they may then use as a guide to choices in Duhem's-problem-type situations. It is an open question, however, how the attributes of this ideal scientist are to be selected, as well as how they will guide resolutions of the problem. Ivanova admits that the "ideal scientist" notion alone doesn't provide an epistemically justified effective procedure for solving Duhem's problem. The justification of one or another resolution of Duhem's problem, she proposes, comes only from the future accumulation of evidence (which thereby confirms or disconfirms the choices made at the time that the problem was logically insoluble).

Ivanova and Paternotte (2013) look for a social solution to Duhem's problem. Good sense has the function of smoothing the social decision-making process in science because it "makes individual opinions more similar" (2013, pp. 1125–1126). This by itself contributes to resolutions of Duhem's problem and thus benefits scientific progress. (Fairweather 2012 makes a similar claim at the individual level.) Duhem's association between good sense and impartiality is thereby explained: impartiality just recommends against the "individual biases" that stand in the way of consensus (that is, "partiality" is to be read not as a barrier to truth, but as a barrier to agreement). However, this solution leaves the content towards which good sense will drive empty (or, dependent on a shifting social and historical context), and it seems impossible to deny that consensus can as well lead away from the realization of epistemic values such as truth, as towards them. Thus, "good sense" remains epistemically unjustified on the social account, as Ivanova and Paternotte admit (2013, p. 1127).

In sum, the fundamental problem with Duhem's "good sense" solution, in all its guises, is that either it fails to be action guiding, or it fails to be normatively justified. Commentators disagree about the precise content of Duhem's "good sense" notion: Is it a kind of direct, intuitive access to the truth; or some definite set of intellectual and moral virtues; or an arbitrary mechanism to produce scientific consensus? But for each of these answers, good sense fails on one or the other of these grounds. The solutions to epistemic formulations of the problem considered below do not share this shortcoming.[1]

2.2 Darden's diagnostic solution

Another solution only suggested metaphorically by Duhem is to consider the problem of localizing error as a doctor would, as a problem of diagnosis (1914/1954, pp. 187–188). Lindley Darden has developed this approach into one of the most promising solutions to Duhem's problem (1990; 1991, note 17). The key to Darden's approach to Duhem's problem is her decomposition of the system being tested and her emphasis on the normal function of the theories involved. It is this emphasis on the normal functioning of theories and their components that allows her to diagnose the theory's faults. Other problems in the test are identified in a step-wise anomaly-resolution process that proceeds as follows:

(1) "Reproduce anomalous data": that is, ensure that the disconfirmatory result E' can be achieved repeatedly and regularly
(2) "Localize potentially problematic components": that is, identify those parts of the theoretical system that are most plausibly at fault for the incorrect prediction
(3) "Generate alternative hypotheses to account for the anomaly."
(4) "Assess among the alternative hypotheses."
(5) "Evaluate the nature of the hypothesis with evidence in its favor": for instance, does accepting the new hypothesis require fundamental changes to the original theory, or only the recognition of special cases to which the original theory does not apply? (text in quotations from Darden 1991, p. 113; cf. 269)

[1] An anonymous reviewer usefully notes the possibility of "pragmatic" resolutions that are not themselves epistemic: for instance, the choice to favor a hypothesis that would lend theoretical support to more easily or cheaply conducted experiments over a rival hypothesis. We agree that such pragmatic considerations play a role in scientific decision-making and are important topics of study, but we restrict our inquiry here to the prospects of epistemically justified resolutions. Our selection and evaluation of proposed solutions in the following sections thus rests on appeals to epistemic values such as fit with the evidence, independent support for assumptions, or explanatory power, rather than pragmatic ones.

Darden offers no quantitative model of the steps above. However, her view employs a richer account of several relevant factors, such as theory structure, the relation between theory and the domains it putatively describes or explains, and the practical conditions of experimental inquiry, when compared to some quantitative approaches, such as Bayesian accounts, which will be discussed below. Due in part to this enrichment, Darden offers detailed and practicable "strategies" (her term) for localizing blame for anomalous results and revising one's theory and practice.

Darden distinguishes three kinds of "strategy" for localizing the source of an anomaly:

a. "Direct" strategy: find components that directly account for the anomalous domain item (or its normal, nonanomalous counterpart)
b. "Explanation" strategy: trace all the components involved in the explanation of the anomalous item (or its normal, nonanomalous counterpart)
c. "Implicit" strategy: uncover implicit assumptions in the explanation of the anomalous item (or its normal, nonanomalous counterpart) (Darden 1991, p. 113)

Darden illustrates her approach with a case study from classical genetics. In 1905 Lucien Cuénot announced that his breeding experiments on mice had produced results that did not agree with the expected 3:1 ratios of classical Mendelian theory. Cuénot was interested in coat color and knew that yellow was dominant, but when he crossed yellow hybrids (heterozygotes), the ratio of yellow to non-yellow in the next generation was about 2.55:1, not 3:1. Moreover, when he bred the yellows from this generation, he could not obtain any homozygous yellow mice (Darden 1991, p. 99). Shortly after this anomaly appeared, three different hypotheses were offered to explain it. These hypotheses shared the anomaly-resolution strategy of changing the theory to accommodate a change in the domain. Furthermore, each agreed that it was the theory's account of segregation that needed to be changed, but the changes proposed were different. The components concerning "segregation" were singled out for closer study because it was these components that predicted the 3:1 ratios. If different ratios were observed experimentally, it was reasoned, then something in these components must be wrong. This is an instance of Step (2) of the anomaly resolution process.

Cuénot's hypothesis was that the germ cells did not combine randomly as the theory had formerly supposed they had. Germ cells containing the factor for yellow preferred to combine with germ cells containing a non-yellow factor. This selective fertilization resulted in no yellow-yellow combinations (Darden 1991, p. 100). Thomas Hunt Morgan, by contrast, explained the 2:1 ratio by denying the (formerly assumed) purity of the gametes, i.e, that the germ cells are of one parental type or another but not both. Morgan instead proposed that there was an active factor and a latent factor within each germ cell, even those that appeared to be pure dominants. Pure breeding dominants didn't exist, according to this scheme, and Morgan explained Cuénot's observations by claiming that he had found a case wherein the latent recessive factors had expressed themselves earlier than usual (Darden 1990, pp. 100–101). Thirdly, W. E. Castle and C. C. Little proposed an explanation based on the claim that the embryos formed from the combination of two germ cells with the yellow factor were not viable; the yellow homozygote combination was lethal. This explanation was further supported by the observation of reduced numbers of young in Cuenot's crosses (Darden 1991, pp. 101–102). The explanation didn't involve a direct rejection of an explicit component of the Mendelian theory, as had Cuénot's and Morgan's explanations, but rather the identification of an implicit and not previously considered premise of the theory–namely, that all offspring of crosses were equally viable.

Of the three possible explanations considered above, Cuénot's selective fertilization explanation and Morgan's impure gametes explanation are examples of Darden's direct strategy. Castle and Little's explanation is an example of the implicit strategy, since until that time the assumption of equal viability of all the combinations had not been considered. The only strategy of Darden's not used in the 2:1 ratio case was the explanation strategy. In order for it to have been appealed to someone would have had to propose fault with one of the theoretical components that the account of segregation depended on, for instance, the thesis that inheritance occurs through the transmission of unit-characters. Thinking through these possibilities helps to show how Darden's approach suggests specific strategies for resolving a case of Duhem's problem, that is, Step (3) of the anomaly resolution process as she describes it.

The main strength of Darden's approach is its sophisticated treatment of the modularity of theory, domain, and relation between the two, and the diversity of localization and theory-revision strategies that it pulls from an analysis of these relations. Further strengths include the useful reminders that we ought to be sure we can replicate the anomalous results before even trying explain them, and that anomalies themselves are often "information rich" and closer study of them can provide resources to guide choices among competing resolutions.

However, Darden's approach also carries some limitations. Theoretical systems may sometimes be more densely interconnected than (say) classical Mendelian genetics. At a certain point, such complexity might make a modular approach untenable. Also, Darden treats localization as a response to a stable anomaly. Duhem's problem is thus construed as one of localizing blame in the theoretical system responsible for producing the prediction. But Duhem's problem extends to theoretical presuppositions of the experiment as well as those involved in prediction generation. Step 1, confirmation of the replicability of the anomaly in question, is itself subject to Duhem's problem. Effectively Darden has left out a large part of the theoretical system by making the claim that when a negative result is found, the theory is always looked to first as a source of the error. The search for the source of fault need not, and should not, overlook the possibility of experimental error. Greater attention to this, including the statistical signals of such error, may therefore be recommended (see sections on Mayo and Sober, below).

One further note on what Darden calls the *diagrammatic* strategy of localization is in order (1990; 1991, pp. 191–204, 271–272). In a diagrammatic localization strategy, one first diagrammatically represents the theory itself. The theory diagram is a description of the "normal" or "typical" functioning of the systems the theory putatively describes (1991, p. 202). Placing the anomalous data at the appropriate step of the diagram then allows one to identify the steps prior to this step, within the "normal functioning" of the system, that could plausibly have led to the appearance of the anomalous data. Finally, one seeks to determine which of that set of potentially problematic steps is actually at fault for the mismatch between theory and anomalous evidence.

Darden’s diagrammatic strategy allows for the same threefold distinction between “strategies of localization”: direct, explanatory, and implicit. But by diagrammatically representing theory and domain as following a “normal” functioning pattern, the approach employs additional information (some of it merely hypothetical or implicitly assumed) concerning the temporal and causal order of steps in such processes. Darden’s diagrammatic strategy thus takes us some way towards her later views on mechanistic explanations (Machamer et al. 2000; see also discussion of Weber’s IBE approach below). However, such a strategy may be limited in application, or inherently idealizing of some systems, as some theories and domains may not fit mechanistic assumptions (for instance: computation-heavy inferences such as “searching the space of all possible phylogenetic trees”). There are also important and difficult questions about how to determine the scope, frequency, and boundary conditions of the “normal form” of the system.

The normal function of a theory is more than a proposed set of logical relationships among the parts of a theory. Because it also encompasses relationships to phenomena that it is understood to explain, and because parts of the theory may be understood to have independent sources of epistemic support, the “normal function” carries with it some level of epistemic support. In other words, the theory that is understood to have a “normal function” does so in virtue of additional information that lends parts of that theory epistemic support. This differential support contributes to the apportionment of praise and blame after a test.

2.3 Bayesian solutions

While different Bayesian approaches have been defended by different writers (e.g. “personalist” Bayesians like Dorling 1979 versus “objectivist” Bayesians like Strevens 2001, both discussed below), the core of any Bayesian solution to Duhem’s problem can be summarized as an effort to use a comparison between (i) the level of conviction an epistemic agent has in various propositions *before* knowing the new experimental result, and (ii) the change in the level of conviction the agent would or should have in various propositions *after* learning the experimental result, to calculate—using Bayes’ rule and Bayes’ theorem, recounted below, as well as other theorems of probability theory and statistics—(iii) the degree to which belief in various specific propositions (such as H and A) should change in such a transition (cf. Dorling 1979, Howson and Urbach 1989, Earman 1992, Strevens 2001). The most recognizable trademark of Bayesian approaches is their reliance on Bayes’ rule and Bayes’ theorem within this general procedure. Strevens (2001, p. 517) states these elements as follows:

Bayes’ Rule	$\mathrm{Pr}^{+}(\mathrm{H}) = \mathrm{Pr}(\mathrm{H}|\mathrm{E})$
Bayes’ Theorem	$\mathrm{Pr}(\mathrm{H}|\mathrm{E}) = [\mathrm{Pr}(\mathrm{E}|\mathrm{H})/\mathrm{Pr}(\mathrm{E})]\mathrm{Pr}(\mathrm{H})$

where H = hypothesis, E = the new evidence, Pr(H|E) = probability of H given E, Pr = prior probability (that is, probability assigned before the awareness of the new evidence E), and Pr^{+}=posterior probability (that is, the probability assigned after the new evidence E comes in). The main terms of these two equations may be understood, intuitively, as follows: $Pr^+(H)$ = new confidence in the hypothesis; Pr(H|E) = *confirmation* of the hypothesis by the evidence, i.e. how strongly does the evidence *confirm* the hypothesis?; Pr(E|H) = *explanatoriness* of the hypothesis vis-à-vis the evidence, i.e. how strongly does the hypothesis *predict* the evidence?; Pr(E) = prior probability (expectation) of seeing the evidence, i.e. how *surprising* is the evidence?; Pr(H) = prior probability of the hypothesis, i.e. How already confirmed or disconfirmed is the hypothesis? Notice that the rule and the theorem can be connected; knowing how well the evidence supports the hypothesis (i.e. Pr(H|E)) tells us how confident we should be in the hypothesis now that the new evidence has rolled in (i.e. $Pr^+(H)$).

Bayesianism's main advantages are (1) the authority it claims by virtue of its reliance on well-established mathematical procedures such as Bayes' rule and theorem, and mathematical probability and statistics more generally, and (2) the promise of comprehensiveness in its approach. Regarding (1): Bayesianism rests its claims about Duhem's problem on the highly reliable or at least widely accepted canons of probability theory and statistics (e.g. Bayes' theorem). Regarding (2), Bayesianism seems to enable inclusion and comparison, within a formal calculus, of evidence of *any* kind that would (intuitively, arguably) increase or decrease the probability of some proposition's truth. This evidence can include a single observation, a string of previous observations, raw or processed data sets, testimony of other scientists, consistency with other well-confirmed beliefs, relative strength *vis-à-vis* alternatives, hunches, and plenty of other things besides. Of course, the importation of information to weight the prior probabilities renders any Bayesian solution to Duhem's problem a solution to the epistemic form of the problem.

One of the main weakness of Bayesian approaches is the difficulty of assigning quantitative (or even just relative) values to the terms in their equations in non-arbitrary ways (as noted, for instance, by Mayo 1997 and Sober 2004). For instance: What is the probability that the tides would change as they do if Newton's theory of gravity were false? (Mayo 1997, pp. 226–229). There doesn't seem to be any non-arbitrary way of assigning a value to this term (Pr[E| ~ H] in the formalism used above).[2] Since a Bayesian's conclusion about the extent to which H or A is disconfirmed by E varies based on the value assigned to such terms, however, it is worried that a Bayesian's conclusions on this matter will be inescapably arbitrary.

Dorling's solution Dorling (1979) presents a well-known attempt to develop a Bayesian solution to Duhem's problem. Dorling describes the problem as follows: a number of hypotheses, say Ha and Hb, entail some expected result, E, which is in disagreement with some experimental result, E'. If we were testing the hypothesis Ha, then we want to know Pr(Ha|E'), but all the test tells us is that Pr(E'|Ha&Hb) = 0. The test tells us nothing about the likelihood of the individual hypotheses. Dorling's proposed solution uses subjective probability assignments and Bayes theorem to compare the damaging effects of the refuting evidence, E', on Ha and on Hb.

[2] For one thing, Pr(E| ~ H) plausibly depends on how many alternative theories (e.g. H_1, $H_{2,}$… H_n) would predict E, and with what strength. Mayo (1997) points out that since the number of alternative theories that could predict E can plausibly be expanded or contracted at will with the help of some theoretical ingenuity and historical selectivity, the strength of Pr(E| ~ H) can likewise be varied at will.

More specifically, subjective probability values must be attributed to p(Ha), Pr(Hb), Pr(E'|Ha& ~ Hb), Pr(E'| ~ Ha&Hb), and Pr(E'| ~ Ha& ~ Hb). Pr(E|Ha&Hb) is assumed to be 1, p(E'|Ha&Hb) is assumed to be 0, and Ha and Hb are assumed to be probabilistically independent. Once subjective probability values are assigned, Dorling can compute the consequences for p(Ha|E') and p(Hb|E') using the rules of probability theory and use these values to decide between blaming Ha or Hb.

Dorling uses the Bayesian approach outlined above to argue that potentially refuting evidence can have a decidedly asymmetric effect, which cannot be resolved in the direction of blaming Ha or Hb arbitrarily or by mere fiat. This indeed constitutes a "solution" to Duhem's problem, but a solution to particular epistemic formulations of the problem, rather than a solution to the logical formulation. As Dorling puts it, to obtain a preference for rejecting one component over another—say, Hb over Ha—"it is only necessary that [Ha] start off more probable than not and substantially more probable than [Hb] and that E' be no more readily explainable on any plausible rival theory to [Ha]" (Dorling 1979, p. 184).[3] The Bayesian scheme, thus, shows that in some conditions one is warranted in blaming auxiliary hypotheses, while in others one is warranted in blaming the main theory.[4] Importing additional epistemically relevant information is necessary for creating this asymmetry.

Dorling's approach to Duhem's problem has its own limitations as a general solution. For one thing, Dorling assumes that the components of the partitioned theoretical system are probabilistically independent: that Pr(Ha|Hb) = Pr(Ha|~ Hb) (Dorling 1979, p. 181). As Dorling admits, this will not be true in all cases. Strevens (2001, p. 526) notes that when it comes to relations between theories under test and auxiliary assumptions pertaining to the equipment used to test them (a classic sort of situation in which Duhem-type problems arise), the relations will most often *not* be independent. This need not be a reason to abandon Dorling's approach entirely, however; the Dorling calculation could still apply in cases where we have reason to believe (or are willing to assume) that these components are theoretically independent.

Strevens' solution. In addition to the limitations noted above, Dorling's Bayesian solution depends on there being an asymmetry in how the refuting evidence affects the partioned theoretical system, i.e. H and A (or, Ha and Hb). This asymmetry is produced only under limited conditions. Strevens (2001) exploits this attribute of Dorling's solution to suggest that Dorling hasn't actually solved Duhem's problem, but rather avoided it, insofar as Dorling's solution relies on information other than E and E ⊨ ~ (H&A) to differentially assign confirmation effects to H and to A. Strevens (2001) notes there can be motives for change in probability assignment to H and to A, on the basis of E, that are not directly due to E's confirmatory or disconfirmatory effects on the conjunction of H&A. This means previous "solutions" to Duhem's problem such as Dorling's have actually avoided the problem rather than solved it, since they may unwittingly appeal to or rely on features that are strictly speaking not

[3] In this quotation, we've substituted "Ha" and "Hb" for Dorling's notation ("T" and "H" respectively), to maintain consistency with our notation.

[4] Note that by treating theory and auxiliary statements differently in his model, Dorling imports another kind of asymmetry into his description of the problem. For instance, he counsels us to consider explanations by alternative theories, but doesn't describe the evaluation of auxiliaries in the same way.

part of the E $\models$ ~(H&A) entailment from which alone any fair solution to Duhem's problem ought to be wrestled, according to Strevens.

Strevens' proposed response is to develop a way of calculating the confirmation effects of E on A and on H that are due to the premises E and E $\models$ ~(H&A) alone. He admits that not every situation in which Duhem's problem arises can be covered by this calculation; thus, his general solution is conditional on there not being significant confirmatory effects of E on A or on H, other than what follows from the confirmatory effects of E on the conjunction of A and H. Given these conditions, Strevens unsurprisingly finds that the relative confirmation effect of E and E $\models$ ~(H&A) on A and H depends entirely (or nearly entirely) on the relative prior probability assignments to A and to H. On this basis, Strevens recommends that tests of H be performed using only A's of the highest possible prior probability – otherwise, a disconfirmatory result E will not lower the probability of H very much, but *will* lower the probability of the weaker A's significantly.

Strevens' concrete suggestion to use A's of high probability, within the kinds of cases he considers, is a useful result. But we think Strevens' objection to Dorling assumes too high a bar for acceptable solutions to Duhem's problem, driving the criteria for any acceptable "solution" too far in the direction of the formal version of the problem. Strevens' procedure restricts the content that may be legitimately considered relevant to solutions to Duhem's problem to an unhelpful degree. Rather, we think addressing Duhem's problem as it appears within the complex, information-rich environments of science in practice—that is, *epistemic* rather than *logical* instances of the problem—is likely to benefit from recognizing as broad a set of potential sources of asymmetry between E's disconfirmatory effects on A and on H as possible.

Relatedly, Fitelson and Waterman (2005) argue that any situation meeting Strevens' condition for a true solution is a situation in which E is inferentially *irrelevant* to the conclusions drawn: Strevens' model allows us to infer the resulting $P^+(A)$ and $P^+(H)$ on the basis of P(A) and P(H) alone! They also provide a simple case in which $P(A) = P(H)$, E $\models$ ~(A&H), and E, and yet $P^+(A) > P^+(H)$ – that is, in which Strevens' conditions on an acceptable solution are violated:

> Let H be the claim that either the winning ticket will be among tickets #1-#5 or it will be ticket #11. Let A be the claim that the winning ticket will be among tickets #4-#9, and let E be the claim that the winning ticket is among tickets #1-#3 or it is among tickets #6-#9. (Fitelson and Waterman 2005, p. 296).

Obviously $P(A) = P(H)$, yet $P^+(A) > P^+(H)$ in this case. Strevens doesn't deny the existence of cases of this sort, but he seems obliged to admit his proposed solution wouldn't apply to them. The simplicity and intuitiveness of this case suggest that Strevens' limiting assumptions may restrict the applicability of his solution rather narrowly. As a general solution to Duhem's problem, it is thus arguably less attractive and useful than the woolier version of Dorling and his predecessors. At least the woolier version can capture the asymmetry between A and H in the example above. Fitelson and Waterman are quite explicit in preferring Dorling's version of Bayesianism for this reason.

All Bayesian solutions to Duhem's problem explicitly depend on epistemic information not included in the formal problem, information that must support the precise kind of asymmetry (per the Bayesian formulas) required to place blame. The difference between Dorling's and Strevens' approaches is in how much, and what kind of, information is taken as allowable within a Bayesian analysis of a case of Duhem's problem.

2.4 Mayo's error-statistics solution

Deborah Mayo's proposed solution to Duhem's problem appeals to well-known procedures for estimating the probability that some experimental result is an error, in order to determine the probability that auxiliary hypotheses (A) or the main hypothesis (H) are the source of the disconfirmatory test result. Mayo's solution is part of a larger philosophical program, which she calls "error statistics," and which promises to use the notion of "learning from error," and standard statistical estimates of the probability that errors of various kinds and ranges have been made, to solve a number of canonical problems in philosophy of science.

Mayo faults Bayesian approaches such as Dorling's for construing all probability assignments as part of a single "probability pie." Mayo's approach, by contrast, recognizes a variety of different parts to empirical scientific reasoning: in particular, she follows Patrick Suppes in distinguishing data models, experimental models, and hypothesis models (Mayo, 1996, pp. 230–232; 1997). Because each of these can be split off and evaluated independently, she believes that solutions to the epistemic form of Duhem's problem are possible. Mayo describes her approach as a "piecemeal approach." In her words, solutions to Duhem's problem proceed in two steps: "the first task is to determine if the data itself are reliable, to determine if there is a real effect (a real anomaly) that needs explaining; the second is to determine if the assumptions of an experiment are met sufficiently, that is, to the problem of checking if alternative auxiliary factors are intervening or if the experiment is adequately controlled" (Mayo, 1996, p. 230). Like Duhem and Darden, Mayo invokes the diagnostic metaphor, and believes that the division of a theoretical system into different kinds of models and their independent assessment, provide the foundation for a differential diagnosis.

Without claiming to provide a complete solution (Mayo 1997, pp. 229, 242), Mayo presents some strategies that will apply to many cases. These strategies are: (i) adding a (specific) auxiliary hypothesis (that is, an A_{n+1}) from which E would follow despite H and A; (ii) retaining the auxiliary hypothesis under question (A), whatever other changes are made; and (iii) rejecting the auxiliary hypothesis under question (that is, concluding ~ A), whatever other changes are made. Mayo doesn't claim that such attempts to preserve H should always be accepted or rejected; rather, she presents a procedure for deciding whether some proposed attempt should be accepted. In particular, she claims that only those new auxiliary hypotheses, or rejections of old ones, that can pass a *severe test* should be accepted, where a severe test is defined as: "H's passing test T (with result e) is a severe test of H just to the extent that there is a very low probability that test procedure T would yield such a passing result, if hypothesis H is false" (Mayo 1997, p. 232). These tests become parts of a process designed to reduce error.

Compared to Bayesian accounts, Mayo is right to claim that her view appreciates the piecemeal nature of the system of models or theories in a scientific test as well as the meaningfulness of anomolies and negative results themselves. We agree with her that in order to present solutions to the epistemic form of Duhem's problem what is needed are "separate tools to detect whether specific auxiliaries are responsible for observed anomalies, tools for discriminating signals from noise, ruling out artifacts, distinguishing backgrounds, and so on. And these tools should be applicable with the kind of information scientists actually tend to have or can obtain" (Mayo 1997, p. 242). This naturalistic approach shares a number of features with Darden's approach.

2.5 Sober's likelihood analysis approach

Elliott Sober (2004) aims to demonstrate that Duhem's problem is at least sometimes soluble – that is, that when "(H&A) $\models$ ~ E, and E," then "evidence bearing on (H&A) can have an impact on H that differs from the impact it has on A" (Sober 2004, p. 221). The type of solution Sober describes employs likelihood analysis. Consider an experiment wherein the hypothesis is that a certain cell drawn from tissue will be stained. The researcher knows she is using one of two microscopes, with different probabilities of being faulty, but doesn't know which microscope she is using. When the observation is made, the cell in question appears to be unstained. In this case, the researcher wants to decide whether (a) her hypothesis is incorrect (i.e. the cell is unstained) or (b) the appearance is due to failure of the microscope. Suppose further that the experimenter believes that all cells drawn from this tissue, as this one was, will be stained. A likelihood analysis of this case would assign likelihoods to four different scenarios, represented in the chart below. This answers the question for each scenario—that is, for each combination of factors—"What is the likelihood that the cell appears unstained, given this combination of factors?" (cf. Sober 2004, pp. 225–226):

	Microscope P	Microscope R
Cell is stained (H)	*A*	*C*
Cell is not stained (~ H)	*B*	*D*

Where do the numbers in these four quadrants come from—that is, how do we assign quantitative values to *A*, *B*, *C*, and *D*? Sober's answer is that the right kind of testing can provide objective reasons to set these numbers in one way rather than another. In a case like this one, the proposal would amount to conducting a large enough number of trials of the same kind. Specifically, it would involve peering into microscope P at a sufficiently large number of cells one already knows to be stained or unstained (say, 100 each), and peering into microscope R at a sufficiently large number of cells one, again, knows to be stained or unstained (again, say 100 each). For this example, such a data set might look like this:

MICROSCOPE P	100 stained cells	100 unstained cells
+ test result (appears stained)	94	2
− test result (appears unstained)	6	98

MICROSCOPE R	100 stained cells	100 unstained cells
+ test result (appears stained)	86	1
− test result (appears unstained)	14	99

On the basis of this data, an estimate can be made of the likelihood, in any of the four scenarios, that a specific observation would be made. For the observation that the cell appears *unstained*, for instance, this estimate would be:

Cell appears not stained

	Microscope P	Microscope R
Cell is stained (H)	6/100	14/100
Cell is not stained (~ H)	98/100	99/100

The numbers in each quadrant represent the likelihood that the test result would come out as it did (E = the cell appears not stained) if that scenario held.

On the basis of this chart, we can say that observation E gives some substantial support to rejecting the hypothesis that the cell is stained. This is because the likelihoods in the "Cell is not stained" row are much greater than the likelihoods in the "Cell is stained" row, regardless of which microscope is used: that is, 98/100 > 6/100 and 99/100 > 14/100. In other words: If the cell *were* actually stained, there would be only a 6–14% chance that it would appear unstained, regardless of which microscope was used. At the same time, observation E gives some small support to the conclusion that microscope R rather than microscope P is being used, insofar as the likelihoods in the rightmost column are greater than the likelihoods in the leftmost column: 14/100 > 6/100, and 99/100 > 98/100. Note also that the difference in likelihoods is much greater in the row-to-row than in the column-to-column comparison, and thus the observation offers *more information* relevant to deciding the stained/unstained question than the microscope question. This shows very clearly that some evidence E may bear differently on H than it does on A, or vice versa.

One limitation to Sober's approach, which Sober himself recognizes, is that it only applies to cases wherein we compare exhaustive alternatives – for instance, where we choose between concluding H or ~ H, or between concluding A or ~ A. That means Sober's approach won't directly help us decide between two rival hypotheses (e.g. between H_1&A and H_2&A). The exhaustivity condition further means that specific likelihood analyses may ignore factors that in reality are more important in generating the patterns we see (or preventing us from seeing patterns that are indeed in operation) than the ones highlighted in that particular likelihood analysis. What factors we choose to study makes a big difference to the information on which our choice of H or ~ H, A or ~ A will be determined. These limitations are partially overcome in approaches such as Darden's that take a more comprehensive view of the experimental and theoretical situations in which Duhem-type problems arise and offer heuristic guidance in the generation and comparison of rival hypotheses.

For Sober's solution to a case of Duhem's problem to apply, information not contained in the formal statement of Duhem's problem is required. In the cases above, the additional information includes the set of prior observations that establishes the likelihoods. Thus Sober's solution, like the other solutions considered here, is a solution to the epistemic form of the problem rather than the logical form.

2.6 Weber's IBE approach

Marcel Weber's "inference to the best explanation" (hereafter IBE) approach to Duhem's problem supplements a Darden-style modular approach with a strong epistemological argument in favor of specific resolutions (Weber 2009). The IBE approach, in a nutshell, consists in ranking possible explanations of an experimental result in terms of their explanatoriness and adopting the explanation that is most explanatory. By "explanatoriness," Weber means sufficiency (and perhaps breadth) of explanation of the data by Woodward's interventionist criterion of explanation (Weber 2009, p. 36). While it is always possible that one or more auxiliary hypotheses may be wrong, the superior explanatoriness of the preferred hypothesis, in comparison with competitors, actually gives support to the auxiliary hypotheses that are part of that explanation. The auxiliaries become "inferential hitchhikers" on the selected (preferred) hypothesis.

Consider a molecular biologist *circa* 1958, curious about the mechanism of DNA synthesis (Weber 2009, pp. 23–29). Three different but incompatible hypotheses are live options: Watson and Crick's "semi-conservative" mechanism (whereby the two strands of the DNA molecule are split, and two new strands are made to fit each of the separated strands); Stent's "conservative" mechanism (whereby the entire double-helix is copied without splitting); and Delbrück's "dispersive" mechanism (whereby parts of the old strands and newly formed strands are interspersed). A set of researchers (Meselson and Stahl 1958) conduct an experiment that seems to disconfirm the conservative hypothesis. They induce *E. coli* to incorporate heavy nitrogen (nitrogen-15) into their DNA, then transfer the *E. coli* to an environment containing only light nitrogen (nitrogen-14). By running samples of the bacterial DNA in an ultracentrifuge, they claim to distinguish different stands of DNA by weight. After one generation, DNA of intermediate weight (between heavy and light) appear. This appears to many interpreters of the experiment to definitively refute the conservative hypothesis.

An epistemic case of Duhem's problem arises here: Does the result actually refute the conservative hypothesis, or is it due to an error in the researchers' experiment—perhaps the assumption that the DNA placed in the centriguge is not relevantly altered by the extraction process (Weber 2009, p. 29)? In Weber's reading of this case, Meselson and Stahl's result, combined with what else was known at the time, made the "dispersive" and "conservative" hypotheses clearly less explanatory than the "semi-conservative hypothesis." This made the semi-conservative hypothesis the best available explanation, and it ought to have subsequently been favored for that reason.

Weber's approach has the advantage of capturing a form of reasoning that is surely common in many experimental contexts. The greater relative explanatory power of a theoretical position vis-à-vis imaginable rivals can indeed make the former legitimately preferable to the latter. There are, however, a number of limitations to Weber's approach.

First: IBE functions much like Mill's eliminative induction, which recommends that we exhaustively list possible explanations, experimentally rule out as many as we can until only one remains, and select that one. IBE does have an appropriately tempered sense of the extent to which we can really "eliminate" any such explanation—we ought to think of the elimination as comparative rather than absolute. However, IBE shares a major weakness of eliminative induction in that its strongest form depends on giving an exhaustive tally of possible explanations, and such an exhaustive tally cannot be had in many contexts. Weber tries to save IBE from this objection in the context of answering Van Fraasen's "bad lot" objection to IBE (Weber 2009, pp. 43–5), but this reply depends on the additional assumption of a mechanistic criterion of explanation. In short, Weber argues that the set of hypotheses considered is likely to be exhaustive due to the constraints on possible explanations imposed by mechanistic assumptions.

This strikes us as the wrong direction in which to walk from the bad lot argument. We ought rather to recognize the limitations on any particular application of IBE. And mechanistic assumptions may not always be applicable, as we noted in connection with Darden's "diagrammatic" approach. It's possible that IBE itself can be freed from such mechanistic commitments, but any non-ambiguous ranking of hypotheses in terms of "explanatoriness" will still require a single criterion of explanatoriness, mechanistic or otherwise. It's unclear whether this can reasonably be expected or required.

Weber faces another problem in the "inferential hitchhiking" of auxiliary hypotheses. An anonymous critic asks why alternative explanations can't *also* bring their auxiliary assumptions along for the ride; Weber responds that these explanations (at least in the Meselson-Stahl case) didn't actually explain their auxiliaries, so the contest would be settled in favor of the semi-conservative hypothesis. In some situations, however, the choice of which competing hypothesis is "more explanatory" of its auxiliaries may not be so easily settled, not least because of competing standards of explanation.

Weber's solution to Duhem's problem, like the other solutions considered here, depends on consideration of information not contained in the formal statement of the problem – in this case, the relative explanatoriness of different conceivable hypotheses (or, sets of hypotheses including auxiliaries). The adopted criteria of explanatoriness, as well as the range of hypotheses considered, can justify preference for one solution over another where such information generates sufficient asymmetry.

3 Conclusion

Our review of the literature on Duhem's problem affirms Duhem's original insight that logic alone is not sufficient to offer a solution to Duhem's problem. Indeed, the epistemic form of the problem is and should be the focus of philosophical analysis.

The value of recognizing the epistemically informed version of Duhem's problem goes beyond philosophy. Scientists find ways to solve Duhem's problem every day, both epistemically and pragmatically. Their solutions invariably rely on their own particular knowledge about the theoretical and experimental systems with which they are engaged. They are making use of their knowledge to give different epistemic weights to assumptions that will then allow them to make an informed plan for revision and future research. The more epistemic form of Duhem's problem, thus, bears on actual scientific practice in a way that the logical formulation does not.

Not all of the proposed philosophical solutions to the epistemic form of Duhem's problem are equally strong. Duhem's own proposed solution in terms of good sense was too vague to indicate how the epistemic problem is solved. More contentful strategies for solving Duhem's problem (such as those proposed by Mayo, Sober, or Weber) seem to be well-matched to particular kinds of cases, but not to others. Bayesian strategies express some general and valid logical conditions on acceptable solutions but require concrete input values to generate decisions about which of the competing conclusions is probabilistically better supported. These values are difficult to assign in a non-arbitrary (that is, epistemically justified) way, and must meet specific asymmetry conditions in order to work.

Nevertheless, our survey of philosophical approaches to Duhem's problem does offer some insight into solution strategies and areas for further philosophical inquiry. The most promising philosophical approaches are those that incorporate justifiable criteria for evaluating additional epistemically relevant information, such as those proposed by Darden, Mayo, Sober, and Weber. In general, the Bayesian account is correct that a solution to the epistemic form of the problem requires an asymmetry or differential epistemic status ascribed to the elements used in a test. Of all the approaches considered, Darden's and Mayo's are perhaps the most promising avenues of further inquiry in that they appreciate that (1) the elements of a scientific test are numerous, varied, and related to each other in specific structures (models and modules), (2) the different elements (assumptions) used in a scientific test carry different epistemic weights that can be evaluated independently from the test at hand, and (3) failures are informative in so far as a negative result is not simply the negation of the expected result but often an informative deviation. Each of these features, we believe, holds for more naturalistic descriptions of scientific testing. Accordingly, exploring how the conjunction of these features of scientific tests inform solutions to instances of Duhem's problem holds the most promise for explicating Duhem's diagnostic metaphor and for analyzing specific scientific cases.

References

Burnham, K., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304.

Darden, L. (1990). Diagnosing and fixing faults in theories. In J. Shrager & P. Langley (Eds.), *Computational models of scientific discovery and theory formation* (pp. 319–346). San Mateo: Morgan Kaufmann Publishers.

Darden, L. (1991). *Theory change in science: Strategies from Mendelian genetics*. Oxford: Oxford University Press.

Dorling, J. (1979). Bayesian personalism, the methodology of scientific research programmes, and Duhem's problem. *Studies in History and Philosophy of Science, 10*(3), 177–187.

Duhem, P. (1914/1954). *The aim and structure of physical Theory*, 2nd (Edn.) (P. Weiner, Trans.). New York: Atheneum.

Fairweather, A. (2012). The epistemic value of good sense. *Studies in History and Philosophy of Science, 43,* 139–146.

Fitelson, B., & Waterman, A. (2005). Bayesian confirmation and auxiliary hypotheses revisited: A reply to Strevens. *British Journal for the Philosophy of Science, 56*(2), 293–302.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach*. La Salle: Open Court.

Ivanona, M., & Paternotte, C. (2013). Theory choice, good sense and social consensus. *Erkenntnis, 78,* 1109–1132.

Ivanova, M. (2010). Pierre Duhem's good sense as a guide to theory choice. *Studies in History and Philosophy of Science, 41,* 58–64.

Mayo, D. G. (1997). Duhem's problem, the Bayesian way, and error statistics, or 'What's belief got to do with it?'. *Philosophy of Science, 64*(2), 222–244.

Mayo, D. G. (1996). *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.

Sober, E. (2004). Likelihood, model selection, and the Duhem-Quine problem. *The Journal of Philosophy, 101*(5), 221–241.

Strevens, M. (2001). The Bayesian treatment of auxiliary hypotheses. *British Journal for the Philosophy of Science, 52,* 515–537.

Stump, D. (2007). Pierre Duhem's virtue epistemology. *Studies in History and Philosophy of Science, 38,* 149–159.

Weber, M. (2009). The Crux of crucial experiments: Duhem's problems and inference to the best explanation. *British Journal for the Philosophy of Science, 60*(2009), 19–49.