Data, Evidence, and Explanatory Power

Pascal Ströing*

Influential classical and recent approaches to explicate confirmation, explanation, or explanatory power define these relations or degrees between hypotheses and evidence. This holds for both deductive and Bayesian approaches. However, this neglects the role of data, which for many everyday and scientific examples cannot simply be classified as evidence. I present arguments to sharply distinguish data from evidence in Bayesian approaches. Taking into account this distinction, we can rewrite Schupbach and Sprenger's measure of explanatory power and show the strengths of this adaptation by applying it to the example of the detection of gravitational waves in physics.

1. Introduction. Explications of confirmation, explanation, and explanatory power are widely discussed subjects in the philosophy of science, historically and recently. Some approaches are based on deductive logic, and some are Bayesian. This article is motivated by a critical view of the fact that in all of these discussions—at least the most cited ones—these three important concepts are described as relations between *hypotheses* and *evidence* (formally often denoted as h and e) plus, in some accounts, further auxiliary assumptions.

I criticize how in these discussions the notion of evidence seems to be commonly understood to cover rather simple everyday propositions like "she saw a black raven" from toy examples in the same way it covers large sets of data that occur in more realistic examples from actual science. In fact, an analysis of scientist's jargon, as well as the epistemic inferences involved in their reasoning, reveals a principle difference between evidence and data. In the most simple terms, the difference is that evidence can be epistemically modeled or explicated by a *proposition* in a deductive logical system, whereas data are often vastly complex (e.g., millions of numbers, photographs) and show uninteresting features like noise.

Received August 2016; revised April 2017.

*To contact the author, please write to: Geschwister-Scholl-Platz 1, München, Bayern D-80807, Germany; e-mail: pascal.stroing@gmail.com.

Philosophy of Science, 85 (July 2018) pp. 422–441. 0031-8248/2018/8503-0005$10.00
Copyright 2018 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

The introduction of conceptual interlayers between the two ends of scientific inference, the—depending on the specific terminology of the account—hypothesis/theory on the one side and the observation/data on the other side, is not new. Suppes (1960) introduces *models* and more specifically (1962) *models of data*. Bogen and Woodward (1988) introduce a distinction between data and *phenomena*. In section 7, I elaborate on the reasons why the notion of evidence is of particular importance for my approach. Roughly speaking, I prefer the term "evidence" over "model" for pragmatic reasons, and phenomena are described as features of the real world about which scientists make claims with their theories, whereas evidence is detected in data.

I justify the distinction between data, evidence, and hypotheses further by examples. The observation of gravitational waves (h) by a LIGO (Laser Interferometer Gravitational-Wave Observatory) measurement (d) via the pattern that a binary black hole merger (e) would produce is one of these examples.

To provide a more concrete and applicable account of a distinction between data and evidence, I develop a measure of explanatory power of a hypothesis over a set of data. The substantiation of this measure is not trivial and takes up the majority of this article. The starting point is Schupbach and Sprenger's (2011) measure for explanatory power of a hypothesis h over some evidence e:

$$\mathcal{E}(e, h) \coloneqq \frac{\mathbb{P}(h|e) - \mathbb{P}(h|\neg e)}{\mathbb{P}(h|e) + \mathbb{P}(h|\neg e)} \in [-1, 1],$$

for which they provide a careful epistemological substantiation. My newly suggested measure $\hat{\mathcal{E}}(d, e, h) \in [-1, 1]$ is the trivalent measure for explanatory power of a hypothesis h over some set of data d, whereas the inference from data to a hypothesis is restricted by the detection of a certain stipulated evidence e.

This article is organized as follows. In section 2 the differences between hypotheses, evidence, and data regarding the application of a probability measure $\mathbb{P}$ are scrutinized. I introduce three brief examples from everyday reasoning and science. In section 3 I introduce measures of explanatory power from the literature. In sections 4 and 5 I construct $\hat{\mathcal{E}}$ based on formalized constraints for it. In section 6 I present the case study of the history of finding experimental confirmation for the existence of gravitational waves. The adapted measure is applied to different experimental approaches from this history. Section 7 provides a comparison of the distinction between data and evidence to the influential accounts from Suppes and from Bogen and Woodward. It also provides some further notes regarding other examples from science and $\hat{\mathcal{E}}$.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

2. Investigating Probability Assignments to Hypotheses, Evidence, and Data. We begin this section with a brief discussion about the roots of logical confirmation theory and the Bayesian motivation for measuring explanatory power. This introduction prepares the ground for a cornerstone of our argumentation: an adequate and applicable explication of explanatory power in science should include inferences from data to evidence.

Historically, the formal investigation of the relation between a hypothesis and some evidence is associated with the expression of the hypothesis and the evidence as sentences in a formal logical language. Classical confirmation theory is built on the assumption that such deductive logical explications or models for reasoning are helpful to normatively investigate confirmation.[1] In these approaches to understanding confirmation the deductive entailment relation for two logical sentences plays a crucial role.

In the Bayesian framework, the explanandum $e \in \mathcal{A}$ and an explanans $h \in \mathcal{A}$ are modeled as propositions or events in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.[2] In applicatory contexts we need to specify a probability measure $\mathbb{P}$ and what it models, denotes, or represents (e.g., a community's or a single agent's degree of belief, statistical degrees of certainty).

The role that entailment ($h \vDash e$) plays in deductive confirmation theory conceptually translates into conditional probability $\mathbb{P}(e|h) \in [0, 1]$ in Bayesian approaches. The assignment of probabilities to evidence and hypotheses alone without any conditionalization, that is, $\mathbb{P}(e)$ and $\mathbb{P}(h)$, is crucial and more intricate than it may seem prima facie.

2.1. A Principle Difference between $\mathbb{P}(h)$ and $\mathbb{P}(e)$. The explanans $h \in \mathcal{A}$ *models* a hypothesis, which is a conclusion or scientific theory about some empirical object or system. The following list shows some very basic and different examples:

$\langle h_a \rangle$ the patient N has measles,

$\langle h_b \rangle$ all ravens are black, and

$\langle h_c \rangle$ there are gravitational waves as predicted by Einstein's general theory of relativity.

In our notation, from now on, with $\langle x \rangle$ we refer to the actual epistemic content of reasoning (e.g., a proposition), and with x, to its representation

1. I refer to Hempel's (1943, 1945) work.

2. Referring to events $a, a_1, a_2 \in \mathcal{A}$ in the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we use the following notions to denote operations on sets: for negation ($\neg a \coloneqq \mathcal{A} \setminus a$), conjunction ($a_1, a_2 \coloneqq a_1 \wedge a_2 \coloneqq a_1 \cap a_2$), and disjunction ($a_1 \vee a_2 \coloneqq a_1 \cup a_2$).

https://doi.org/10.1086/697733 Published online by Cambridge University Press

in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$; that is, $x \in \mathcal{A}$. The reason for this syntactical distinction is that my further argumentation is based on a precise understanding of what counts as a proposition and how we should interpret a probability assignment to it.

Probability $\mathbb{P}(h_\cdot)$ denotes the degree to which $\langle h_\cdot \rangle$ is believed in or true or empirically adequate.[3] The point that I want to stress is that the explanans in $\mathcal{E}$ is a proposition and a *human-made* expression (as opposed to propositions from, e.g., artificial intelligences or animals) that cannot be determined to be true by mere observation because of practical (e.g., untestability) or theoretical problems (e.g., induction problem). Those are characteristics of hypotheses, and that is why, realistically, $\mathbb{P}(h_\cdot) < 1$.

The explanandum $\langle e \rangle$, the evidence, is more intricate concerning the interpretation and construction of an assignment of degrees of truth or belief by $\mathbb{P}$. The reason for this is that it refers to observational reports about the empirical world. Examples corresponding to $\langle h_\cdot \rangle$ include

$\langle e_a \rangle$ patient N has a red rash,

$\langle e_b \rangle$ the raven I saw today was black, and

$\langle e_c \rangle$ on September 14, 2015, LIGO recorded gravitational waves of a black hole merger.

This evidence is of human-made propositions, too (and not data), but the values $\mathbb{P}(e_\cdot)$ have a somewhat different meaning than $\mathbb{P}(h_\cdot)$ in the Bayesian interpretation. In the cases of the explanans $\langle h_\cdot \rangle$, $\mathbb{P}(h_\cdot) < 1$ describes that there is a positive probability that the hypothesis is not believed in or inadequate or false. In example $\langle h_a \rangle$ the patient may not have a measles infection, despite symptom $\langle e_a \rangle$; thus, $\mathbb{P}(h_a | e_a) < 1$.

For the examples $\langle e_\cdot \rangle$, the values $\mathbb{P}(e_\cdot) \in [0, 1]$ prima facie do not express very interesting information, if we interpret the application of $\mathbb{P}(e_\cdot)$ as providing the same information about $\langle e_\cdot \rangle$ as $\mathbb{P}(h_\cdot)$ provides about $\langle h_\cdot \rangle$. We built up our scenario in a way that, trivially, $\mathbb{P}(e_\cdot) = 1$ or $(=0)$ because the $\langle e_\cdot \rangle$ are just reports of—more or less—*directly observable* actual occurrences, whereas the $\langle h_\cdot \rangle$ are not directly observable. If $\mathbb{P}$ are the degrees of belief of an agent who knows the past observation, then $\mathbb{P}(e_\cdot) \in \{0, 1\}$, but $\mathbb{P}(h_\cdot) \in (0, 1)$, that is, an open interval.

In Bayesian frameworks of confirmation $\mathbb{P}(e)$ is often interpreted as how surprising the future occurrence or observation of $\langle e \rangle$ would be. McGrew (2003, 557–58) stresses this point. The difference here is that $\langle e \rangle$ can practically be known, whereas $\langle h \rangle$ cannot. Why is this important to us? Actual

3. The term "$h_\cdot$" denotes one among the following possibilities: h_a or h_b or h_c.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

scientific inferences do not refer to human-made and possibly true propositions as explananda. Scientists look at data and try to interpret them. They do not refer to the data in any propositional way. Data are rather a large series of numbers or images or the like.

2.2. Evidence Is Information in Sets of Data. As opposed to the examples for an explanans $\langle h \rangle$, the occurrence of some evidence $\langle e \rangle$ is detected via a *pattern* in data. Observations must be interpreted by some observer to become evidence. For example $\langle e_b \rangle$ the observer has to make the decision that the dark object he saw in the sky was in fact a black raven. But as we discussed above, $\mathbb{P}(e)$ does not express anything regarding the inference from a set of data to the proposition, which is the evidence. The reason for my understanding is that in the discussions of measurements of explanatory power the evidence is assumed to be an *epistemically accessible* proposition such as our examples $\langle e. \rangle$ above and not sets of data. Only for these epistemically accessible expressions of the sort of $\langle e. \rangle$, which are propositions that are expressed in a human language and relatively brief, can we decide how surprising they are to us.

An objection to my point would be to claim that data are as epistemically accessible as the evidence $\langle e. \rangle$. Figure 1 shows examples of sets of data that correspond to our examples $\langle h. \rangle$ and $\langle e. \rangle$. In this case, $\langle d. \rangle$ does not denote a human-made proposition, it denotes the set of data itself that is relevant for scientific reasoning; $d. \in \mathcal{A}$ denotes its representations in the probability space. Confronted with the complexity of the data—millions of perceptible pixels, thousands of numbers—there is no meaningful way in interpreting $\mathbb{P}(d.)$ with $\langle d. \rangle$ being the actual set of data. How surprising is it to get exactly this picture of a raven or this series of numbers? That is why, in many realistic examples from science, evidence $\langle e \rangle$ in Bayesian frameworks, which make use of $\mathbb{P}(e)$, cannot be observed data. Our brain is not designed to specifically focus on raw data, but in science we often do so. The inference from data to evidence is a crucial epistemic step in contexts of scientific explanations. The measures of explanatory power that Schupbach re-

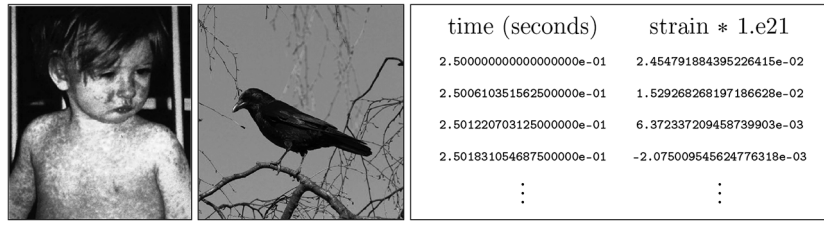


Figure 1. Examples of data. (Sources: Wikimedia, http://pixabay.com, http://losc.ligo.org.)

https://doi.org/10.1086/697733 Published online by Cambridge University Press

TABLE 1. Candidate Measures of Explanatory Power

Measure	Values in Interval
$E_D(e, h) \coloneqq \mathbb{P}(e\|h) - \mathbb{P}(e)$	$[-1, 1)$
$E_C(e, h) \coloneqq \mathbb{P}(e\|h) - \mathbb{P}(e\|\neg h)$	$[-1, 1]$
$E_P(e, h) \coloneqq \frac{\mathbb{P}(e\|h)-\mathbb{P}(e)}{\mathbb{P}(e\|h)+\mathbb{P}(e)}$	$[-1, 1)$
$E_M(e, h) \coloneqq \ln\left(\frac{\mathbb{P}(e\|h)}{\mathbb{P}(e)}\right)$	$(-\infty, \infty)$
$\mathcal{E}(e, h) \coloneqq \frac{\mathbb{P}(h\|e)-\mathbb{P}(h\|\neg e)}{\mathbb{P}(h\|e)+\mathbb{P}(h\|\neg e)}$	$[-1, 1]$

Note.—Measures as listed by Schupbach (2011). We refer to his text for further historical references regarding these measures.

fers to (see table 1) take as explananda evidence and neglect the inference from data to evidence.

The representations of data $\langle d_\cdot \rangle$ in figure 1 are to a certain degree still approximations. Digital images are restricted by their size in pixels, the data may more accurately be represented by a video, and so on. However, these approximations describe $\langle d_\cdot \rangle$ much better than just identifying it with the propositional evidence $\langle e_\cdot \rangle$.

Example $\langle d_c \rangle$ is of particular interest to us. In cases $\langle d_a \rangle$ and $\langle d_b \rangle$ the inferences from data to evidence can be explained by common sensitive-cognitive abilities and proper background knowledge about the discussed phenomena. These abilities cannot be described easily and raise scientific questions by themselves, but we can agree on their being natural capabilities of human beings and therefore reasonably addressed by $\mathbb{P}(e_\cdot|d_\cdot)$, which gives the degree to which the evidence can be inferred from the given set of data.

This description does not apply to $\langle d_c \rangle$. Here, the inference from data to evidence is based on an algorithm to detect the patterns that we can infer from Einstein's theory of general relativity and our knowledge about cosmic bodies. Also relevant to the inference are controlling mechanisms to delete known disturbing influences from the data. The physicists present their results with quantifications of the degree of statistical certainty, which is common in experimental discoveries.[4] Expression $\mathcal{E}(e_c, h_c)$ gives the explanatory power excluding this statistical uncertainty for the inference from $\langle d_c \rangle$ to $\langle e_c \rangle$. Probability $\mathbb{P}(e_c|d_c)$ is the degree to which $\langle e_c \rangle$ can be inferred from $\langle d_c \rangle$, and this probability is completely different from some degree of surprise that $\langle e_c \rangle$ occurs.

4. Abbott et al. (2016) present the detection of the relevant pattern with "a false alarm rate estimated to be less than 1 event per 203 000 years." The "median values with 90% credible intervals" are given for the parameters to describe the detected black holes. Note that this first information provides an answer to the probability $\mathbb{P}(d_c|\neg e_c)$, which can be used to infer the demanded probability $\mathbb{P}(e_c|d_c)$; we should carefully aim for the avoidance of a base rate fallacy.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

In many related discussions the difference between the evidence $\langle e \rangle$ and the set of data $\langle d \rangle$ is neglected. A common understanding seems to be that this inference can be neglected as a minor influence to scientific and everyday reasoning that becomes an insignificant victim in the process of approximation. Example $\langle \cdot_c \rangle$ is associated with a history of false inferences from data to evidence,[5] or to say it more accurately: it is a history of overestimating the degree to which the evidence is given by the observed data or overestimating the degree to which $\langle h_c \rangle$ is true if some evidence is given. We elaborate on this history in section 6.

In Schupbach (2011) participants to an experiment have to provide personal degrees of belief after investigating a series of 10 balls that are drawn from an urn *A* and not from urn *B*. From both urns, they know the composition of black and white balls, and they are asked to decide from which urn the balls were drawn. An outcome of the draw is the finite information fully described by, for example, $\langle e_s \rangle$ bwwbwbbbbw. Here, $\langle e_s \rangle = \langle d_s \rangle$; in this example there is no difference between the evidence and the data, and no inference from data to evidence is necessary (or it is trivial). I take this choice of an example as an indication for the common negligence of the inference from data to evidence. Overall, this example is a very artificial simplification of actual scientific or everyday examples to which a measure of explanatory power should be applicable to.

2.3. $\mathbb{P}(d)$—the Bayesian Degree of a Set of Data. Can and do we need to apply $\mathbb{P}$ to a set of data $\langle d \rangle$? If we claim that $\mathbb{P}(d)$ gives the degree of how surprising the occurrence of (exactly) $\langle d \rangle$ is, then we face the problem of $\mathbb{P}(d)$ being in almost all realistic cases extremely low. Every slight deviation in the data from $\langle d \rangle$ is, strictly speaking, an alternative set of data $\langle \hat{d} \rangle$.

The example $\langle d_c \rangle$ indicates how the distortion, additive noise, or vagueness of evidence-revealing patterns in data is tackled with the body of statistical methods. These statistical methods are a common mathematical theory that in principle does not deviate from mathematics that is applied to express and reason about empirical theories.

The important point is that we cannot make use of a value $\mathbb{P}(d)$—except for the claim $\mathbb{P}(d) > 0$; that is, the observed data should be a possible observation—in our adapted measure of explanatory power that includes the inference from data to evidence. Probability $\mathbb{P}(d)$ is a very low value and provides no meaningful information for our purpose.

3. Measures of Explanatory Power. Schupbach and Sprenger (2011) discuss measures of explanatory power. These measures provide a quantifica-

5. Collins (2004) provides a survey of this history including social aspects.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

tion (e.g., a value in the interval $[-1,1]$) of explanatory power that a hypothesis has over some evidence. Formally,

$$E_{\mathbb{P}}(e, h) : \mathcal{A}^2 \times \mathcal{P} \rightarrow \mathbb{R},$$

with $\mathcal{P}$ being the set of all probability measures over the measurable space $(\Omega, \mathcal{A})$. They provide a Bayesian account to answer questions like 'how strongly does a measles infection explain a red rash?'. The measures explicate explanatory power in contexts of everyday and of scientific reasoning. Measures of explanatory power are already available in the literature, but Schupbach and Sprenger argue in favor of a new version, which they denote with "$\mathcal{E}(e, h)$." I use their result as a recently discussed example to further describe problems and suggest an adaptation that incorporates the distinction between evidence and data. A historical overview of some measures is given in table 1.

Schupbach and Sprenger claim that "our grasp on the notion of *explanation* is at least sufficiently strong to ground a precise formal analysis of *explanatory power*, even if it is not strong enough to determine a general account of the nature of explanation" (2011, 107). Their technique to define the measure $\mathcal{E}(e, h)$ for a proposition e and a hypothesis h is to introduce a list of reasonable constraints that $\mathcal{E}$ has to fulfill. These constraints express criteria of normalization and symmetry, as well as a proper interpretation of independencies between the relevant propositions.

To infer or explicate the different measures of explanatory power, the authors make use of several conditionalized and unconditionalized assignments of probabilities. They assign probabilities to the explanandum $\mathbb{P}(e)$, to the explanandum constrained on the explanans $\mathbb{P}(e|h)$ and its negation $\mathbb{P}(e|\neg h)$, and to the explanans constrained on the explanandum $\mathbb{P}(h|e)$ and its negation $\mathbb{P}(h|\neg e)$. We have to consider these assignments, when we introduce d to the measure.

This new version[6]

$$\hat{\mathcal{E}}_{\mathbb{P}}(d, e, h) : \mathcal{A}^3 \times \mathcal{P} \rightarrow \mathbb{R}$$

aims to explicate the explanatory power of a hypothesis over some set of observed data, which, in many cases, cannot be interpreted as a special form of evidence. From a formal perspective, the cornerstone of $\hat{\mathcal{E}}$ is to somehow incorporate the inference from data to evidence (e.g., by reference to $\mathbb{P}(e|d)$ and $\mathbb{P}(\neg e|d)$).

6. At other positions in this article, the subscripted $\mathbb{P}$ is not printed because of the reference to Schupbach and Sprenger's original notation.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

4. Explanatory Power and Data: Separating Data from Evidence. How can we adapt $\mathcal{E}$ to measure the explanatory power that a hypothesis has over a set of observed data, instead of some propositional evidence? The intention for the adaptation is that we want to explain what we actually observe (or at least what we recorded from the observation) and provide a justified explication of this trivalent measure $\hat{\mathcal{E}}(d, e, h)$.

The probabilistic expressions $\mathbb{P}(h|e)$ and $\mathbb{P}(h|\neg e)$ are the parameters of $\mathcal{E}(e, h)$. Our approach for $\hat{\mathcal{E}}$ is to adapt these building blocks for the measure of explanatory power over observed data accordingly. The first step is to adapt $\mathbb{P}(h|e)$ by formalizing the preformal intuition

$$\mathbb{P}(h[e]d) \text{ for the probability of } \langle h \rangle, \text{ given } \langle d \rangle \text{ and inference only via } \langle e \rangle$$

to replace

$$\mathbb{P}(h|e) \text{ for the probability of } \langle h \rangle, \text{ given } \langle e \rangle$$

from $\mathcal{E}$. The notation "$\mathbb{P}(\cdot[\cdot]\cdot)$" is more or less arbitrary and is chosen to express the described intentions by syntactically replacing the conditionalization "$|$" with a conditionalization "$[e]$" in which we are implicitly restricted to a propositional evidence $\langle e \rangle$ at the inference from $\langle d \rangle$ to $\langle h \rangle$. The term "$\mathbb{P}(\cdot[\cdot]\cdot)$" is only a notation and, as will become clearer below, I do not intend to extend the standard probability calculus in any way.

The evidence still plays an important role for $\hat{\mathcal{E}}$. We aim to measure the explanatory power for a set of observed data via some evidence. In science, as well as in everyday explanations, data are always interpreted somehow; we have to know what we look for in the data. In many realistic scientific examples, the evidence (e.g., patterns from gravitational waves) for which we look in the data (e.g., time series from LIGO measurement) is already well described before the data are gathered. This may even be seen as a indicator of good scientific method. Furthermore, if we imagine a case of Popperian falsification (see sec. 7.3), evidence may also be chosen with a particularly negative relevance to the hypothesis in question. Evidence can, roughly speaking, be explicated in the language of hypothesis confirmation or falsification, whereas data cannot (at least not in a direct way); our cases $\langle h. \rangle$, $\langle e. \rangle$, and $\langle d. \rangle$ provide examples for this. A set of data cannot be the explanandum of an explanation without the stipulation of some propositional evidence that should be revealed in the data by some pattern.

My technique to explicate $\mathbb{P}(h[e]d)$ is to, first, list constraints that are necessary to really express the intention as preformally described above. We establish the following constraints for

https://doi.org/10.1086/697733 Published online by Cambridge University Press

$$\mathbb{P}(h[e]d) : \mathcal{A} \times \mathcal{A} \times \mathcal{A} \rightarrow [0, 1]$$

and all h, h_1, h_2, e, d, d_1, $d_2 \in \mathcal{A}$ and $\mathcal{D}, \mathcal{H} \subset \mathcal{A}$:

(i)	**Informativeness**:	$\mathbb{P}(h)$, $\mathbb{P}(e)$, $\mathbb{P}(d) \in (0, 1)$, that is, open intervals
(ii)	**Evidential Data**:	If $e = d$, then $\mathbb{P}(h[e]d) = \mathbb{P}(h\|e)$
(iii)	**Weak Data**:	$\mathbb{P}(h[e]d) \leq \mathbb{P}(h\|e)$
(iv)	**Independence**:	If $\mathbb{P}(e\|\bar{d})$ is constant for all $\bar{d} \in \mathcal{D}$, then $\mathbb{P}(h[e]\bar{d}_1) = \mathbb{P}(h[e]\bar{d}_2)$ for all $\bar{d}_1, \bar{d}_2 \in \mathcal{D}$, and if $\mathbb{P}(\bar{h}\|e)$ is constant for all $\bar{h} \in \mathcal{H}$, then $\mathbb{P}(\bar{h}_1[e]d) = \mathbb{P}(\bar{h}_2[e]d)$ for all $\bar{h}_1, \bar{h}_2 \in \mathcal{H}$
(v)	**Monotonicity**:	If $\mathbb{P}(e\|d_1) \leq \mathbb{P}(e\|d_2)$, then $\mathbb{P}(h[e]d_1) \leq \mathbb{P}(h[e]d_2)$, and if $\mathbb{P}(h_1\|e) \leq \mathbb{P}(h_2\|e)$, then $\mathbb{P}(h_1[e]d) \leq \mathbb{P}(h_2[e]d)$

The intentions concerning the constraints are mostly self-explanatory. We introduce i to exclude tautologies and contradictions and to avoid division-by-zero problems. Constraint ii aims to cover the—often very artificial—examples in which the data are already designed to be evidence. Constraint iii expresses that the assumption of $\langle e \rangle$ (or $\langle e \rangle = \langle d \rangle$) always nondecreases the $\mathbb{P}(h[e]d)$, since the inference from $\langle d \rangle$ to $\langle h \rangle$ is restricted to only infer $\langle e \rangle$ (and no other possibly evidence) from $\langle d \rangle$. Constraint iv states that the two steps, inference from data to evidence and inference from evidence to hypothesis, are statistically independent. Constraint v expresses that $\mathbb{P}(h[e]d)$ increases, if one of the two relevant degrees of inference increase with the other one being constant.

> **Remark** A solution consistent with constraints ii–v for all probability spaces $(\Omega, \mathcal{A}, \mathbb{P})$ and all h, e, $d \in \mathcal{A}$ for which i holds is
>
> $$\mathbb{P}(h[e]d) \coloneqq \mathbb{P}(h|e) \cdot \mathbb{P}(e|d) .$$

This solution for $\mathbb{P}(h[e]d)$ is the one we use in the following course of this article. Note that this remark does not imply uniqueness of the solution. I intentionally introduced the intuitive notation $\mathbb{P}(h[e]d)$ despite the simple alternative given by this definition. However, when it comes to estimating $\mathbb{P}(h[e]d)$ in empirical contexts (i.e., degrees of belief or truth or appropriateness), cases may occur in which we estimate $\mathbb{P}(h[e]d)$ directly without a further analysis of its mathematical structure.

The probabilistic models for $\mathcal{E}$ and $\hat{\mathcal{E}}$ imply specific probabilistic dependencies between the set of observed data, the evidence, and the hypothesis. The Bayesian networks shown by figure 2 provide an accessible approach

https://doi.org/10.1086/697733 Published online by Cambridge University Press

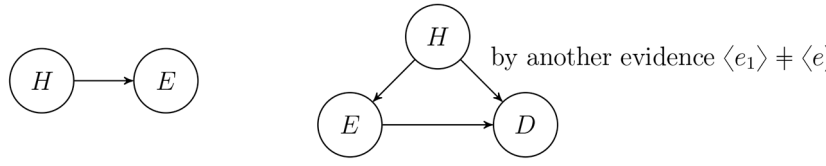


Figure 2. Bayesian networks that describe the underlying dependencies. *Left*, Schupbach and Sprenger's binary relation between hypothesis and evidence, as implied by $\mathcal{E}$. *Right*, my trivalent approach with data, as regarded by $\hat{\mathcal{E}}$.

to express these dependencies.[7] The notation with capitalized letters is common in the literature on Bayesian networks.[8]

Figure 2, right side, indicates how the inference from hypothesis to data depends on a specific evidence. Note that for $\hat{\mathcal{E}}(d, e, h)$ the inference from $\langle d \rangle$ to $\langle h \rangle$ (which is represented by the arrow from H to D) is irrelevant. This evidence in many cases in science is described before the observation or gathering of the data. Figure 3 shows a simple example to highlight these aspects and how our invention $\hat{\mathcal{E}}$ can capture it properly, whereas $\mathcal{E}$ cannot. In this example, if we measure very precisely, $\hat{\mathcal{E}}(d, e_2, h)$ is high, but $\hat{\mathcal{E}}(d, e_1, h)$ is not, and the reason is, of course, the different appropriateness of the choice of the evidence. Note that the amount of alternative evidence is high or even infinite.

5. Explanatory Power and Data: Constructing the Adapted Measure. To replace the expressions $\mathbb{P}(h|e)$ and $\mathbb{P}(h|\neg e)$ from $\mathcal{E}$ to a three-placed probabilistic expression over d, $(\neg)e$, and h and to express the measure of explanatory power for a set of data, we need to answer the following question. How do we translate the conditional probability $\mathbb{P}(h|\neg e)$ from $\mathcal{E}$ into our trivalent version, if $\mathbb{P}(h|e)$ translates to the $\mathbb{P}(h[e]d)$ as discussed above?

By constraints iv and v, $\mathbb{P}(h[e]d)$ can only be a high value if $\mathbb{P}(h|e)$ and $\mathbb{P}(e|d)$ are high values. In contrast, the trivalent adaptation of $\mathbb{P}(h|\neg e)$ should be high, if either $\mathbb{P}(h|\neg e)$ or $\mathbb{P}(\neg e|d)$ is high. The reason for this is that to achieve high explanatory power the negated evidence should neither lead to a high support of the hypothesis nor be strongly supported by the data. A simple solution to formalize this intuition is to translate $\mathbb{P}(h|\neg e)$ for the trivalent adaptation into $\max(\mathbb{P}(h|\neg e), \mathbb{P}(\neg e|d))$. Note that the calcu-

7. For an introduction into Bayesian networks, see Pearl (1988), Bovens and Hartmann (2003), and Neapolitan (2003).

8. A proposition $\langle x \rangle$ can also be modeled by a two-valued random variable $X : \Omega \rightarrow \{\mathrm{X}, \neg\mathrm{X}\}$ in a probability space, instead of $x \in \mathcal{A}$. For our purposes the differences are merely syntactical.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

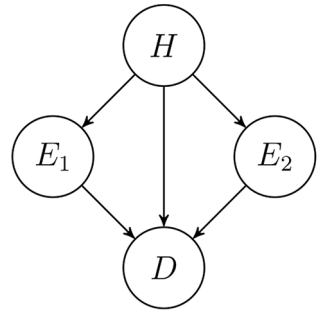


Figure 3. Bayesian network for $\langle h \rangle$, the earth rotates around the sun; $\langle e_1 \rangle$, pattern of a round orbit; $\langle e_2 \rangle$, pattern of a nonround elliptic orbit; and $\langle d \rangle$, records of the sun's position. The actual orbit is nonround elliptic.

lations made in the case study about gravitational waves in section 6 further illuminate these intuitions and formalization.

Note that $\langle \neg d \rangle$ expresses, strictly speaking, every possible outcome of the measurement except the very specific set of data $\langle d \rangle$. Since $\langle d \rangle$ includes one specific realization of noise and $\mathbb{P}(d)$ is a very low value, $\langle \neg d \rangle$ does not have any useful meaning for our approach. In an empirical setup of explanatory power, $\langle d \rangle$ solely plays the role to confirm or disconfirm some specified evidence $\langle e \rangle$ to a certain degree (i.e., $\mathbb{P}(e|d)$ or $\mathbb{P}(\neg e|d)$).

We can now explicate the measure of explanatory power of a hypothesis $\langle h \rangle$ over a set of observed data $\langle d \rangle$ via some evidence $\langle e \rangle$:

$$\hat{\mathcal{E}}(d, e, h) = \frac{\mathbb{P}(h[e]d) - \max(\mathbb{P}(h|\neg e), \mathbb{P}(\neg e|d))}{\mathbb{P}(h[e]d) + \max(\mathbb{P}(h|\neg e), \mathbb{P}(\neg e|d))}.$$

This adaptation makes use of the advantages of $\mathcal{E}$ over the alternative measures of explanatory power (see table 1) by the adaptation of its basic structure. However, from a semantic perspective and concerning the derivation of the measure analogous to Schupbach and Sprenger's, we do something substantially different: the explanandum does not need to be some proposition $\langle e \rangle$. In our account the explanandum can also be a set of observed data $\langle d \rangle$.

The occurrence of the maximum operator at $\hat{\mathcal{E}}$ may seem unpleasant from an analytical mathematical perspective, since it makes $\hat{\mathcal{E}}$ nondifferentiable. The intuition behind it is that, again, first, $\mathcal{E}$ gets significantly negative or close to zero, if $\mathbb{P}(h|\neg e)$ is high. Second, at $\hat{\mathcal{E}}$ this can occur in two different ways, which are a high value for $\mathbb{P}(h|\neg e)$ or a high value for $\mathbb{P}(\neg e|d)$. This aims to ensure that $\hat{\mathcal{E}}$ cannot be high, if one of these two scenarios occurs. Note that one of these two values being high should lead to a significantly low explanatory power—the examples "weak evidence in strong data" and

https://doi.org/10.1086/697733 Published online by Cambridge University Press

"strong evidence in weak data" in section 6 exemplify this point. And this desired feature is exactly what the maximum operator expresses, whereas there is no philosophical need for $\hat{\mathcal{E}}$ to be differentiable.

Schupbach and Sprenger's $\mathcal{E}$ is taken as a template for my adaptation because of the normative arguments they give in addition to Schupbach's (2011) empirical arguments. These are that $\mathcal{E}$ performs best among the candidate measures according to an empirical investigation of decisions from 26 participants. However, if you take a more critical perspective on $\mathcal{E}$, my preformal discussion and some aspects of the formalization of $\hat{\mathcal{E}}$ may also be picked up to define adaptations of the other measures that are listed in table 1 to incorporate the distinction between evidence and data. However, the notation "$\hat{\mathcal{E}}$" is justified by the fact that for $e = d$ we have $\hat{\mathcal{E}}|_{e=d}(d, e, h) \equiv \mathcal{E}(e, h)$.

6. A Comparison of $\mathcal{E}$ and $\hat{\mathcal{E}}$: The Example of Gravitational Waves. We briefly discuss the historical example of the detection of gravitational waves (GW) to show how our approach for $\hat{\mathcal{E}}$ is applicable and very appropriate for explicating explanatory power. GW are a theoretical result from general relativity that also allows for the possibility of detection by various ways. I mention several observations or measurements to detect GW and how the explanatory power of the theory of general relativity over the measurement results is reflected by $\hat{\mathcal{E}}$.[9]

I suggest some very rough probability assignments for the application of $\mathcal{E}$ and $\hat{\mathcal{E}}$. These assignments aim to highlight the implications of $\hat{\mathcal{E}}$ rather than being perfectly accurate under all possible considerations, which demands a more lengthy discussion. Furthermore, the probability assignments are very conservative in the sense that, for instance, "very strong" translates to 0.95 to make the calculations more accessible, whereas in actual experimental physics "very strong" would rather translate to $0.99\bar{9}$. In what follows, $\langle h \rangle$ denotes the hypothesis that GW exist, and $\mathbb{P}$ are the degrees of belief of the mainstream gravitational physicists community.

6.1. Weak Evidence in Strong Data: Rotating Cylinders. The physicist Joseph Weber detected expansions and contractions of a heavy solid cylinder when another heavy solid cylinder rotated in an orthogonal position to it ($\langle e_W \rangle$). He first conducted experiments of this kind in the late 1960s, and other physicists repeated modernized versions of the experiment until the year 2000. The data showed clear effects, and Weber claimed that this evidence confirms GW. His critics claim that this effect can be explained by gravitation alone without the assumption of GW. They claim that the de-

9. Our sources for this history are Bartusiak (2000) and Collins (2004). I refer to their texts for further historical investigations and references.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

tected effect is similar to tides on earth that are caused by the moon's gravitation. The majority of physicists who work on GW criticize Weber's conclusion. Let $\langle d_W \rangle$ denote his measurement results:

$$\mathbb{P}(h|e_W) = P(h|\neg e_W) = \mathbb{P}(h) = .5, \quad \mathbb{P}(e_W|d_W) = .95, \quad \mathbb{P}(\neg e_W|d_W) = .05.$$

The value for $\mathbb{P}(h)$ depends a lot on the point in time (e.g., before or after certain other discoveries), and .5 is more or less arbitrarily chosen, not close to 0 or 1. This gives

$$\hat{\mathcal{E}}(d_W, e_W, h) = \frac{.5 \cdot .95 - \max(.5, .05)}{.5 \cdot .95 + \max(.5, .05)} \approx -.026 \approx 0.$$

It gives $\mathcal{E}(e_W, h) = 0$, as well, but our approach makes clear that this low explanatory power is due to the interpretation of the experimental results and not due to unsatisfying measurement results.

6.2. Medium-Strong Evidence in Strong Data: Orbit Degradations. Hulse and Taylor (1975) measured orbits of a binary star system and detected degradations, which are consistent with the energy loss explained by GW. This evidence is considered strongly confirmed by the data but not very reliable because of strong additional background assumptions for the interpretation of the measurements.[10] Their set of data is $\langle d_{\mathrm{HT}} \rangle$, and their evidence is $\langle e_{\mathrm{HT}} \rangle$.

$$\mathbb{P}(h|e_{\mathrm{HT}}) = .8, \quad P(h|\neg e_{\mathrm{HT}}) = .4, \quad \mathbb{P}(e_{\mathrm{HT}}|d_{\mathrm{HT}}) = .95, \quad \mathbb{P}(\neg e_{\mathrm{HT}}|d_{\mathrm{HT}}) = .05$$

$$\hat{\mathcal{E}}(d_{\mathrm{HT}}, e_{\mathrm{HT}}, h) = \frac{.8 \cdot .95 - \max(.4, .05)}{.8 \cdot .95 + \max(.4, .05)} \approx .31$$

is, again, close to $\mathcal{E}(e_{\mathrm{HT}}, h) = 1/3$. This is always the case for the strong data examples but not for weak data.

6.3. Strong Evidence in Weak Data: Early Interferometers. The two LIGO interferometers are active in the search for GW since 2002. Interferometers provide measurement data for the detection of GW and are widely accepted as actually doing so in the physicists mainstream community. After finding that the data ($\langle d_{L02} \rangle$) are too noisy to detect possible signs of GW ($\langle e_L \rangle$), the more sensitive Advanced LIGO was put into operation in 2015. For the early LIGO, we have

10. Others, e.g., Weisberg, Taylor, and Nice (2010), see a much stronger confirmation of GW by the evidence than my probability assignments indicate. However, I chose these assignments according to the mentioned historical sources, which help to highlight the philosophical aspects of $\hat{\mathcal{E}}$.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

$$\mathbb{P}(h|e_L) = .99,\ \ P(h|\neg e_L) = .02,\ \ \mathbb{P}(e_L|d_{L02}) = .5,\ \ \mathbb{P}(\neg e_L|d_{L02}) = .5, \text{ and}$$

$$\hat{\mathcal{E}}(d_{L02}, e_L, h) = \frac{.99 \cdot .5 - \max(.02, .5)}{.99 \cdot .5 + \max(.02, .5)} \approx .005 \approx 0, \text{ but}$$

$\mathcal{E}(e_L, h) \approx .96$, and $\mathcal{E}(d_{L02}, h)$ is not determinable by the problems of interpreting $\mathbb{P}(h|d)$, which we discussed in section 2.2.

6.4. Strong Evidence in Strong Data: LIGO 2015 Interferometers. Recently, with a stated degree of $\mathbb{P}(e_L|d_{L15}) \in (0.99, 1)$, reliable confirmation of GW was published by Abbott et al. (2016). With the assignments that we know from the case before this one, we get

$$\hat{\mathcal{E}}(d_{L15}, e_L, h) = \frac{.99 \cdot .995 - \max(.02, .005)}{.99 \cdot .995 + \max(.02, .005)} \approx .96.$$

Without the separation of evidence and data, we are not able to distinguish between some profoundly different scenarios. Strongly explained evidence may be given only weakly in data, and weakly explained evidence may strongly be inferred from the data. For $\mathcal{E}$ these different effects have to be semantically incorporated into the evidence $\langle e \rangle$, which leads to unreasonable and indeterminable probability assignments.

7. Further Comments on History and Applicability. In the introduction and other parts of the article, I make several remarks about the history of the distinction between data and evidence. In this section, I offer a more thorough discussion on this matter and on the actual applicability of constraint v.

7.1. Comparison to Bogen and Woodward's Phenomena. Bogen and Woodward (1988) emphasize a distinction between data and phenomena, which was the subject of discussion by articles from Woodward (1989, 2000) and others.[11] In the inaugurating article, phenomena are introduced as features of the real world about which our theories make claims. This pictures a contrast to a logical empiricist's view that scientists aim to describe the observation with their theories without any conceptual intermediate layer. Scientists often use the term "phenomenon" in the same way or at least with a very similar meaning.

An example from the discussion are bubble chamber photographs to detect weak neutral currents. For this detection, roughly 100 of 290,000 photographs were preselected to exclude all the data that did not show any evidence or, in other words, did not imply any information about the sought-after phenomenon.

11. Most notably McAllister (1997), Glymour (2000), and entries from the *Synthese* special issue (Machamer 2011) on the subject.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

What is the difference between a phenomenon and evidence? Much evidence can correspond to a phenomenon but not vice versa. Take the biological phenomenon of albinism for example. Here, we have a *phenotype*, white hair and reddish eyes, and a *genotype*, a certain mutation in the 11th chromosome. Both can be regarded as different evidence, which is evidence for the phenomenon of albinism, whereas we know from a vast amount of biological background knowledge that each type of evidence corresponds to albinism. I think that this interpretation of the terminology is most accurate in a descriptive account of phenomena and evidence.

7.2. Comparison to Suppes's Models. In Suppes's (1962) classical account, he introduces the notion of *models of data* to explain the inference from raw data to the information in it that actually stands in an assumed semantic relation to hypotheses. The model of some data is a modified version of the raw data to strip away noise and 'uninteresting' information in the data. This idea is an approach to introduce a conceptual intermediate step to the inference from data to a hypothesis.

Earlier, Suppes (1960) introduces models as nonlinguistic and set-theoretically definable entities that satisfy a theory, which is always linguistic. Models of data are therefore special cases of models and have to correspond to theories of data. A model of a hypothesis is a further conceptual intermediate step at the inference from data to a hypothesis.

What is new with my approach and how is it related to Suppes's models? As figure 3 indicates, I focus particularly on the possibility of introducing alternative evidence for the inference from one set of raw data to some hypothesis. Or to say it in Suppes's terminology, not just one model of data plays a role in the inference from data to a hypothesis, but many of them can with different degrees of explanatory success. I see no reason why the evidence in my approach should not be identified with models of data, if the theory of the data is sufficiently powerful.

A model of the hypothesis is not regarded in my approach, since it does not play any paramount role for a Bayesian explication of the inference from data to a hypothesis but rather for a characterization or identification of the hypothesis itself. My approach aims to underline how we epistemically separate data, evidence, and hypotheses with the aim to explicate explanatory power. Suppes aims to lay down a framework that defines the borders between theories, models, data, experiment, and also the so-called prima facie conditions. My deviation from Suppes's classical account of models is pragmatic rather than a rejection of his views.

7.3. Popperian Falsification and $\hat{\mathcal{E}}$. A further strength of my suggested measure of explanatory power $\hat{\mathcal{E}}(d, e, h) \in [-1, 1]$ is the result that it provides for cases of *falsification*, as discussed by Popper (1935). If a set of

https://doi.org/10.1086/697733 Published online by Cambridge University Press

data provides strong support for falsifying evidence, then the explanatory power of the falsified hypothesis should be close to -1.

This is a classical example:

$\langle h \rangle$ Light propagates linearly through space without being influenced by gravitation.

$\langle e \rangle$ Stars at daytime with an eclipsed sun appear in the same arrangement as in earth's night sky.

After waiting for a proper sun eclipse to gain the desired image $\langle d \rangle$, $\langle h \rangle$ was considered to be falsified because of strong support for $\langle \neg e \rangle$ by $\langle d \rangle$. In Bayesian terms, this leads to something very roughly like

$$\mathbb{P}(e|d) = .01, \quad \mathbb{P}(\neg e|d) = .99, \quad \mathbb{P}(h|\neg e) = .01, \quad \mathbb{P}(h|e) = .9,$$

and

$$\hat{\mathcal{E}}(d, e, h) = \frac{.01 \cdot .9 - \max(.01, .99)}{.01 \cdot .9 + \max(.01, .99)} = \frac{.009 - .99}{.009 + .99} = \frac{-.981}{.999} \approx -.98.$$

This result corresponds perfectly with a possible notion of a degree of falsification (modulo a change of sign).

7.4. $\hat{\mathcal{E}}$ and the Sciences. At $\hat{\mathcal{E}}$ all aspects about the inference from data to evidence are formally incorporated via $\mathbb{P}(e|d)$ and $\mathbb{P}(\neg e|d) = 1 - \mathbb{P}(e|d)$. In many mathematically worked-out scientific contexts, these probabilities straightforwardly refer to a degree of statistical support that some data provide for some evidence. In addition to the GW example from section 6, many other examples from physics and other fields can easily be found, as figure 1 indicates.

Another question is whether the evidence and the data can clearly be identified in examples from science. As the GW case exemplifies, the case is often unambiguous for mathematically explicated physical investigations. In examples from social sciences, psychology, or medicine, the notions of evidence and data are often sufficiently supported by the scientist's use of the terminology, too. A scientific claim like "people underestimate the danger from extreme catastrophes" (hypothesis) is usually investigated by the definition of some measurable pattern like "in the survey more than 50% marked 'I do not care'" (evidence) in the rawest form at which the records are available, like stapled papers or a computer file (data). An even more illuminating example is the long-standing discussion about *stylized facts* in

https://doi.org/10.1086/697733 Published online by Cambridge University Press

economics,[12] which are certain statistical properties of daily financial asset price data (e.g., extreme returns appear in clusters, returns are not correlated) that inspire two threads of discussion: How strongly is a certain stylized fact confirmed by the data (inference from data to evidence)? What does a certain stylized fact imply regarding the investor's behavior (inference from evidence to hypothesis)? A discussion about stylized facts is therefore a discussion about the evidence that needs to be explicated for the inference from data to a hypothesis. Value $\hat{\mathcal{E}}$ captures this situation adequately.

8. Implications for Measures of Confirmation. Expression "$\mathbb{P}(h[e]d)$" denotes the intuition that any positive inference from data to a hypothesis is only possible by the stipulation of some evidence. This fact influences not only the discussion about measures of explanatory power but also the discussion about measures of confirmation of a hypothesis by some scientific observational result. As opposed to the measures of explanatory power that are listed in table 1, measures of confirmation from the recent literature often take into account that confirmation could be a three-placed relation between evidence $\langle e \rangle$, hypothesis $\langle h \rangle$, and some auxiliary hypotheses and assumptions $\langle k \rangle$.[13]

In Crupi's (2015) notation with variable names as given above, a degree of confirmation is

$$C_{\mathbb{P}}(h, e|k) : \mathcal{A}^3 \times \mathcal{P} \to \mathbb{R},$$

with $\mathcal{P}$ being the set of all probability measures over the measurable space $(\Omega, \mathcal{A})$. This notation also implies that $\langle k \rangle$ with $k \in \mathcal{A}$ and $\mathbb{P} \in \mathcal{P}$ are assumed to be given when the degree of confirmation is measured.

How would a $\hat{C}_{\mathbb{P}}(h, e, d|k)$ that makes use of the idea $\mathbb{P}(h[e]d)$ look? A full answer to this question is part of this article and demands an elaborated discussion of its own. But at this point we can speculate how relevant our discussion for $\hat{\mathcal{E}}$ is for $\hat{C}_{\mathbb{P}}(h, e, d|k)$. The original intentions for the introduction of the auxiliaries $\langle k \rangle$ include the problem that what is often referred to as "evidence" is hardly logically formulated or expressible and fits more the notion of data as introduced in this article. However, the issue that one set of data can confirm a hypothesis to different degrees depending on the choice of the evidence is still not addressed with the introduction of auxiliaries at $C_{\mathbb{P}}(h, e|k)$. The Bayesian network from figure 3 exemplifies this point for confirmation, too. That is why $\hat{C}_{\mathbb{P}}$ has to deviate from $C_{\mathbb{P}}$ by philosophical intentions similar to those underpinning how $\hat{\mathcal{E}}$ deviates from $\mathcal{E}$.

12. For a textbook introduction on stylized facts, see chap. 2 of Cristelli (2014).

13. This point was raised by Glymour (1980).

https://doi.org/10.1086/697733 Published online by Cambridge University Press

9. Conclusion. Approaches to model explanatory power stipulate propositional evidence as explananda. I claim that—in opposition to the vast majority of the tradition of this discussion—observed data are the more adequate explanandum for many scientific applications.

We inferred that for a measure of explanatory power a distinction between evidence and data must be represented in a mathematical explication, which is $\hat{\mathcal{E}}$. Measure $\hat{\mathcal{E}}$ is a relatively simple adaptation of Schupbach and Sprenger's suggested measure $\mathcal{E}$. The adaptation $\hat{\mathcal{E}}$ solves the problem of the separation of evidence from data.

The construction of $\hat{\mathcal{E}}$ is based on a list of constraints. These constraints include, first, that, even if we are interested in the explanatory power of a hypothesis for a set of observed data, it can—by epistemic reasons—only be inferred with the stipulation of some specified evidence in it. Second, we refer to $\mathbb{P}(e|d)$ in $\hat{\mathcal{E}}$ to explicate how strongly evidence can be inferred from the set of data under investigation.

We discussed the example of gravitational waves to highlight the importance of the separation of evidence and data for the application of measures of explanatory power. The argument applies to everyday examples, too. The arguments that were raised in favor of a distinction between evidence and data for the explication of explanatory power similarly apply to explications of measures of confirmation.

REFERENCES

Abbott, B. P., et al. 2016. "Observation of Gravitational Waves from a Binary Black Hole Merger." *Physical Review Letters* 116:061102.

Bartusiak, Marcia. 2000. *Einstein's Unfinished Symphony: Listening to the Sounds of Space-Time*. Washington, DC: Henry.

Bogen, James, and James Woodward. 1988. "Saving the Phenomena." *Philosophical Review* 97 (3): 303–52.

Bovens, Luc, and Stephan Hartmann. 2003. *Bayesian Epistemology*. Oxford: Oxford University Press.

Collins, Harry. 2004. *Gravity's Shadow: The Search for Gravitational Waves*. Chicago: University of Chicago Press.

Cristelli, Matthieu. 2014. *Complexity in Financial Markets: Modeling Psychological Behavior in Agent-Based Models and Order Book Models*. Springer Theses: Recognizing Outstanding Ph.D. Research. Cham: Springer.

Crupi, Vincenzo. 2015. "Confirmation." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University. http://plato.stanford.edu/archives/fall2016/entries/confirmation/.

Glymour, Bruce. 2000. "Data and Phenomena: A Distinction Reconsidered." *Erkenntnis* 52 (1): 29–37.

Glymour, Clark. 1980. *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Hempel, Carl Gustav. 1943. "A Purely Syntactical Definition of Confirmation." *Journal of Symbolic Logic* 8 (4): 122–43.

———. 1945. "Studies in the Logic of Confirmation." *Mind* 54:1–26 and 97–121.

Hulse, Russel, and Joseph Taylor. 1975. "Discovery of a Pulsar in a Binary System." *Astrophysical Journal* 195:L51–L53.

https://doi.org/10.1086/697733 Published online by Cambridge University Press

Machamer, Peter. 2011. "Phenomena, Data and Theories: A Special Issue of *Synthese*." *Synthese* 182, no. 1.
McAllister, James W. 1997. "Phenomena and Patterns in Data Sets." *Erkenntnis* 47 (2): 217–28.
McGrew, Timothy. 2003. "Confirmation, Heuristics, and Explanatory Reasoning." *British Journal for the Philosophy of Science* 54 (4): 553–67.
Neapolitan, Richard E. 2003. *Learning Bayesian Networks*. Upper Saddle River, NJ: Prentice Hall.
Pearl, Judea. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann.
Popper, Karl. 1935. *Logik der Forschung*. Vienna: Springer.
Schupbach, Jonah N. 2011. "Comparing Probabilistic Measures of Explanatory Power." *Philosophy of Science* 78 (5): 813–29.
Schupbach, Jonah N., and Jan Sprenger. 2011. "The Logic of Explanatory Power." *Philosophy of Science* 78 (1): 105–27.
Suppes, Patrick. 1960. "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences." *Synthese* 12 (3): 287–301.
———. 1962. "Models of Data." In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, ed. E. Nagel, P. Suppes, and A. Tarski, 252–61. Stanford, CA: Stanford University Press.
Weisberg, Joel M., Joseph H. Taylor, and David J. Nice. 2010. "Timing Measurements of the Relativistic Binary Pulsar PSR B1913+16." *Astrophysical Journal* 722 (2): 1030.
Woodward, James. 1989. "Data and Phenomena." *Synthese* 79 (3): 393–472.
———. 2000. "Data, Phenomena, and Reliability." *Philosophy of Science* 67 (Proceedings): S163–S179.

https://doi.org/10.1086/697733 Published online by Cambridge University Press