Canadian Journal of Philosophy (2020), 50: 2, 174–194
doi:10.1017/can.2019.18

ARTICLE

Confirmation Based on Analogical Inference: Bayes Meets Jeffrey

Christian J. Feldbacher-Escamilla[1]* and Alexander Gebharter[2]

[1]Duesseldorf Center for Logic and Philosophy of Science (DCLPS), University of Duesseldorf, and [2]Department of Theoretical Philosophy, University of Groningen. Email: alexander.gebharter@gmail.com
*Corresponding author. Email: cj.feldbacher.escamilla@gmail.com

Abstract
Certain hypotheses cannot be directly confirmed for theoretical, practical, or moral reasons. For some of these hypotheses, however, there might be a workaround: confirmation based on analogical reasoning. In this paper we take up Dardashti, Hartmann, Thébault, and Winsberg's (2019) idea of analyzing confirmation based on analogical inference Bayesian style. We identify three types of confirmation by analogy and show that Dardashti et al.'s approach can cover two of them. We then highlight possible problems with their model as a general approach to analogical inference and argue that these problems can be avoided by supplementing Bayesian update with Jeffrey conditionalization.

Keywords: Analogies; indirect evidence; confirmation; Jeffrey conditionalization; Bayesian networks

1. Introduction

Sometimes evidence for a hypothesis cannot be directly observed. This might be the case if the evidence is inaccessible for theoretical reasons. An example would be evidence for certain hypotheses about the dynamics of black holes (Winsberg, 2009; Dardashti, Thébault, and Winsberg, 2015). But even if observing the evidence for a hypothesis is theoretically possible, we still might not possess the know-how or the right tools to measure it, or the costs to produce it, or building the tools required to measure it might be too high. In such cases, evidence cannot be accessed for different practical reasons. The existence of widely recognized moral reservations might also make it impossible to observe evidence. Producing evidence to directly confirm a certain psychological hypothesis might, for example, require surgical interventions on the brains of test subjects.

Cases in which a hypothesis H cannot be directly confirmed by observing evidence E obviously cause trouble for scientists. Though such a hypothesis cannot be directly confirmed, it might make perfectly reasonable true or false claims about the world. Is there really no way to confirm (or disconfirm) such a hypothesis at all? One possible option consists in trying to find systems s' that are similar (or analogous) enough to the systems s about which H claims this and that. One could then formulate a corresponding hypothesis H' for these similar enough systems s'. Contrary to the systems s, these systems s' might produce evidence E' that can be observed directly. Now, the hope is that our original hypothesis H can somehow be confirmed on the basis of observing E'. After all, H' makes a claim about systems s' that is analogous to what H claims about systems s. If E' can somehow be used to confirm H, then it seems that there is a possibility to empirically assess hypotheses whose corresponding evidence cannot be observed (for whatever reasons).

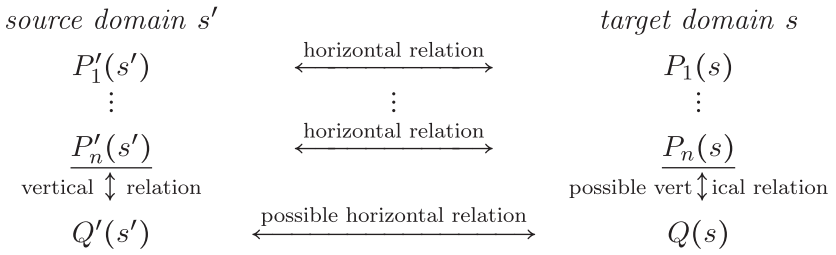

Figure 1. Horizontal and vertical relations in analogical reasoning.

Some kind of confirmation on the basis of analogical reasoning is clearly applied in sciences such as biology, climate science, economics, medicine and pharmacology, etc. However, whether evidence E' that directly confirms a hypothesis H' can be used to confirm an analogous hypothesis H is controversial (see, e.g., the critique of Duhem 1991, 97ff.; Bartha 2010, sec. 1.9). A recent approach put forward by Dardashti et al. (2019) seems to support the view that confirmation based on analogical inference is quite reasonable. They propose a Bayesian analysis of confirmation on the basis of analogical reasoning (for confirmation within a Bayesian framework, see, e.g., Bovens and Hartmann 2003; Hartmann and Sprenger 2011). In particular, they argue that if the systems described by H and H' (at least partially) share the same structural features, there might be a connection between H and H' that establishes probability flow between evidence E' and hypothesis H. This seems to be everything required for E' to (indirectly) confirm H Bayesian style and, thus, confirmation based on analogical inference would turn out to be a kind of Bayesian confirmation.

In this paper, we take up Dardashti et al.'s (2019) idea to make sense of confirmation based on analogical inference in a Bayesian framework. We first identify three more or less classical types of analogical inference in section 2. In section 3, we then introduce and illustrate Dardashti et al.'s approach by means of a simple toy example. We argue that their approach—in its original version—covers only one of the types of analogical inference and show that it can be expanded in such a way that it also covers a second type. We then generalize their approach to scenarios in which common causes play the same role as shared structures (or analogies) play in their account in section 4. This move will turn out to be quite straightforward since, from a formal point of view, common causes work exactly like shared structures. We also discuss general problems (subsections 4.a and 4.b) and more specific problems with the view that evidence E' for a hypothesis H' can confirm another hypothesis H making a claim about a different system (subsection 4.c). In section 5, we discuss approaches not plagued by the possible problems due to cross-system confirmation. In particular, we make a new proposal for the missing type of inference by analogy and develop a model which supplements Bayesian update by Jeffrey conditionalization for cases in which direct evidence for the hypothesis of interest is inaccessible. We conclude in section 6.

2. Three types of analogical inference

In this section, we briefly discuss the traditional characterization and distinguish three types of analogical inference. This will provide the basis for the discussion of analogical inference in the Bayesian framework in subsequent sections. A standard form of an argument by analogy goes as follows (*cf.* Walton 2005, 96):

Similarity Premise: Generally, case s' is similar to case s.
Base Premise: Q' holds in s'.
Conclusion: Q, which is similar to Q', holds in s.

Similarly, Bartha (2010, 1 and 13) states that "an analogical argument is an explicit representation of analogical reasoning that cites accepted similarities between two systems in support of the conclusion that some further similarity exists," which amounts to the following schema:

(1) s' is similar to s in certain respects P'_i and P_i (where P'_i and P_i with $1 \leq i \leq n$ are similar).
(2) s' has some additional feature Q'.
(3) Therefore, s has feature Q (where Q and Q' are similar).

In her seminal monograph on analogical reasoning, Hesse (1966, 59f) suggests to represent analogies by help of a tabular representing features of a source and target system:[1]

(source) properties of *sound* [s']	[s] properties of *light* (target)
echoes [$P'_1(s')$]	[$P_1(s)$] reflection
loudness [$P'_2(s')$]	[$P_2(s)$] brightness
pitch [$P'_3(s')$]	[$P_3(s)$] color
detected by ear [$P'_4(s')$]	[$P_4(s)$] detected by eye
propagated in air [$Q'(s')$]	[Hence: $Q(s)$] propagated in ether

Note that Hesse (1966, 59f; see also Hesse, 1974) already distinguished between horizontal and vertical relations between the target and the source system's features. The horizontal relations between echoing and reflection, loudness and brightness, etc., consist in similarity or identity, while the vertical relations between echoing, loudness, etc., and propagation in air consist in causal dependence. Explicitly describing these relations constitutes an argument by analogy and brings the tabular notion in line with the schemata above. In accordance with Bartha (2010, 24), one can further distinguish between empirically established relations and merely possible (or conjectured) relations. The similarity of s and s' with respect to P_i and P'_i (with $1 \leq i \leq 4$), for example, was a de facto established relation back at the time of ether theories. Hence, the similarities between P_i and P'_i were considered to be horizontal relations (simpliciter). However, the similarity between Q and Q' was not an established relation, but also not excluded back then. It was, hence, considered to be a possible horizontal relation. It is interesting to note that Bartha (2010) also applies this more subtle distinction to vertical relations. The relation between echoing and propagation in air, for example, was an already established causal relation and, hence, considered to be a vertical relation (simpliciter). The similarity between reflection and propagation in ether, on the other hand, was not established on empirical grounds; it was considered to be a possible vertical relation. By extending the approach of Bartha (2010, 24), who considers possible horizontal and vertical relations not independent of each other, but as mutually constrained, we can distinguish two types of analogical inference, namely the type of inferring a (possible) horizontal relation and the type of inferring a (possible) vertical relation. The schemata of these two types of inference are provided in Figure 1.

For an analogical inference of a (possible) horizontal relation we have already provided an example above: the ether example. A prime example of an analogical inference of a (possible) vertical relation would be the famous violinist case presented by Thomson (1971). Here it is argued that similarly to the case where a violinist's right to live does not establish a claim on someone else's

[1]There are more subtle schemata of inferences by analogy—for example, some schemata including dissimilarities, or so-called *negative* analogies (Keynes 1921). To keep things simple, we stick to this general schema of *positive* analogies.

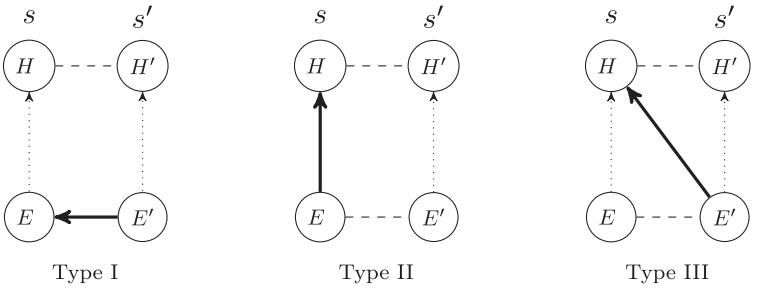

Figure 2. Three types of analogical inference. Thick arrows represent inferred possible relations, dotted arrows established relations, and dashed lines similarity relations.

body (whose kidney is used to keep the violinist alive), a baby's right to live does not establish a claim on the mother's body. Put into the argumentation schema above, this amounts to:

(1) Generally, the violinist case s' is similar to the case of unwanted pregnancy s, i.e., P'_1, P_1 and Q', Q are similar.
(2) A violinist's right to live $P'_1(s')$ does not establish a right to use someone else's body $Q'(s)$.
(3) Therefore, also a baby's right to live $P_1(s)$ does not establish a right to use the mother's body $Q(s)$.

Let us simplify this structure by considering only single hypotheses (conclusions) about and evidence (premises) for features of the source and target systems. Let H stand for such a hypothesis about and E for such evidence for features of the target system s. Likewise, let H' and E' stand for such a hypothesis about and such evidence for features of the source system s'. Furthermore, let us, in accordance with Hesse (1964) and Bartha (2010), assume that the possible horizontal and vertical relations are confirmational relations. Then characterizing the first kind of analogical inference (type I)—which is based on a possible horizontal relation between $Q(s)$ and $Q'(s')$—amounts to the task of evaluating the confirmational impact of E' on E. Note that, speaking in terms of the traditional terminology, in this type of confirmation the possible horizontal relation is established on the basis of established positive analogies (similarities); what role these positive analogies play will become more clear when we present Dardasthi et al.'s (2019) Bayesian approach in the next section. Characterizing the second kind of analogical inference (type II)—establishing a possible vertical relation between $Q(s)$ and $P_1(s), \ldots, P_n(s)$—amounts to the task of evaluating the confirmational impact of E on H on the basis of E''s confirmational impact on H'.

Note that these two types are not the only possibilities of how evidence about features of the source systems could be used for analogical reasoning. It is easy to see that there is also a third kind of analogical inference (type III) which consists in evaluating the confirmational impact of evidence E' (source system) on hypothesis H (target system). This third type is not explicitly discussed in the traditional literature on analogical reasoning. However, as we will see in the next section, it is the central kind of inference discussed in the literature on so-called *analogue simulation*. In the traditional framework, this type of analogical reasoning could be described as an inference along the lines of a possible *diagonal* relation. The three different types of inference by analogy are depicted in Figure 2.

3. Confirmation by analogy Bayesian style

In this section, we present Dardasthi et al.'s (2019) Bayesian approach to confirmation based on analogical reasoning and show that it can cover type I and III analogical inferences. We will use a toy example for illustration throughout the paper: assume we are interested in the efficacy of a new antiviral compound on humans who suffer from a certain viral disease. Experts think that this new

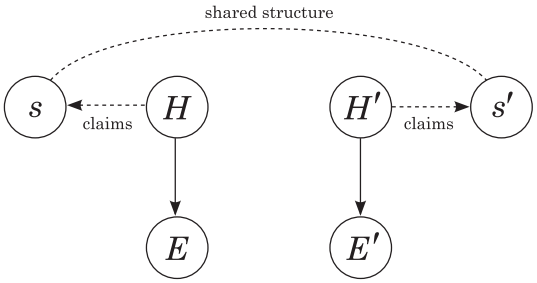

Figure 3. Schema of analogical reasoning.

antiviral compound is incredibly promising. To avoid possible negative side effects, however, physicians decide to first test the antiviral compound on nonhuman model organisms such as rats. Only after several successful trials with rats would the physicians consider cautiously starting to investigate the efficacy of the antiviral compound with human test subjects. This procedure seems to suggest that they make some kind of inference from the success of the treatment on rats to the future success of the treatment on humans. In other words: they seem to assume that a successful treatment of the model organism can be employed to confirm a corresponding hypothesis about humans. Physicians might justify this by pointing to the fact that the immune system of rats works analogously to the immune system of humans in relevant respects.

Let us now describe the toy example introduced above in a more precise way. (For an illustration, see Figure 3.) Let s stand for human test persons and s' for rats. Let H be a hypothesis about the human immune system and H' a hypothesis about the immune system of rats. If H is true, then humans suffering from the viral disease who have been treated with the antiviral compound will recover without severe negative side effects. Likewise, if H' is true, then rats treated with the antiviral compound will recover without noteworthy negative side effects. Finally, let E describe whether infected humans treated with the antiviral compound recover without severe negative side effects and E' for whether infected rats do.

What goes on in the above toy example seems to be the following: at the beginning, the physicians are quite confident that H (claims about the human immune system) holds, but they still consider it as too risky to perform human trials in order to directly confirm H. So they start to test the antiviral compound on rats instead. They also think that a hypothesis H' similar to H holds for the immune system of rats. So the experts are quite confident that the immune systems of humans and rats (s and s', respectively) are adequately modeled by H and H', respectively, and that the immune systems of humans (s) and rats (s') share important structural features (i.e., are in many relevant respects analogous). They then collect evidence E' (treatment success with rats), which directly confirms H'. Because the immune systems of humans and rats are assumed to share relevant structural features, physicians conclude that the results of the rat study can somehow be used to at least weakly confirm H.

So far the rough schema. The problem is that it is still not well understood how the confirmation involved here works. Here is a rational reconstruction of how the confirmation would work in Bayesian terms according to Dardashti et al. (2019). They would stress the fact that the inference seems to crucially employ assumptions about the structure shared by the immune systems of humans and rats.[2] Consequently, they would propose to model this shared structure with a

[2]Note that this is our reconstruction of how Dardashti et al. (2019) would apply their approach to this specific case. In this work, they focus on the dynamics of black holes instead, and the structural similarities they rely on are based upon universality arguments or, more generally, upon connections between background assumptions of the two models (see also Dardashti et al. 2015). Since we are not specifically interested in physics but rather in the more general picture, we choose the much easier to comprehend rat study case as our main example throughout this paper.

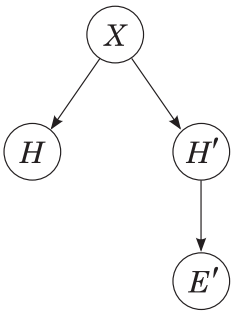

Figure 4. Bayesian network of analogical reasoning type III.

variable X that is a common ancestor of H and H' in a Bayesian network. Bayesian networks are especially well suited for Dardashti et al.'s endeavor because they allow for modeling and graphically representing the paths over which probabilistic information spreads between variables. A Bayesian network consists of a set $\mathbf{V}$ of random variables $X_1, \ldots, X_n$, a set $\mathbf{E}$ of directed edges ($\rightarrow$) connecting some of these variables, and a probability distribution P over $\mathbf{V}$. A triple $\langle \mathbf{V}, \mathbf{E}, P \rangle$ is a Bayesian network if and only if it conforms to the Markov factorization (Pearl 2000, 16):

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathbf{Par}(X_i)), \tag{1}$$

where $\mathbf{Par}(X_i)$ is the set of X_i's parents in the Bayesian network, i.e., all $X_j \in \mathbf{V}$ with $X_j \rightarrow X_i$. If P factors according to equation 1, then one can read off certain independencies from the graph $\langle \mathbf{V}, \mathbf{E} \rangle$. In particular, every $X_i \in \mathbf{V}$ has to be independent of every X_j (where $j \neq i$) that is not connected to X_i over a path $X_i \rightarrow \ldots \rightarrow X_j$ conditional on $\mathbf{Par}(X_i)$.

Now, Dardashti et al. (2019) would analyze the kind of inference used in our toy example above by means of the Bayesian network in Figure 4: this Bayesian network allows for probability flow from E' to H over the path $H \leftarrow X \rightarrow H' \rightarrow E'$. In particular, observing E' confirms H'. This increases the probability of X which, in turn, increases the probability of H. Thus, E' increases the probability of H which, according to Dardashti et al., shows that E' indirectly confirms H. So, according to their understanding, confirmation based on analogical reasoning coincides with Bayesian confirmation of H given E'. They claim that their approach provides a general schema for indirect confirmation by analogy, where by "indirect confirmation" we mean confirmation citing evidence E' from the source system, but not evidence E from the target system.

Before we go on to generalize Dardashti et al.'s (2019) approach in such a way that it can also cover causal scenarios (which play a central role in the traditional literature on analogical inference; see, e.g., Hesse 1964; Bartha 2010) and discuss several possible problems, some remarks are in order. First, note that the role of X (the shared structure) is to mediate probabilistic influence between the source system s' described by H' and the target system s modeled by H. How exactly X must be specified is not entirely clear yet. We will come back to this issue and provide a general theoretical recipe in subsection 4.b. Second, the interpretation of the directed edges in the underlying Bayesian network is more or less flexible. The arrow $H' \rightarrow E'$ stands for some kind of direct confirmation relation, while the arrows $X \rightarrow H$ and $X \rightarrow H'$ express that X describes a common feature of both systems relevant for H and H'. We will argue in section 4 that X could also be understood as a common cause of H and H'. Third, note that Dardashti et al. do not speak of hypotheses H and H', but rather of models that are assumed to adequately represent certain systems. This is due to their interest of explicating what is called an *analogue simulation* in the modeling literature (Winsberg 2009). We have decided to speak of hypotheses instead because this is much more in line with the traditional philosophical debate on analogies and confirmation (see, e.g., Carnap 1962,

appendix D; Hesse, 1964). However, it does not make that much of a difference whether one speaks of hypotheses or of models, because it seems quite straightforward to think of models as complex hypotheses about certain systems. Fourth, note that E' can increase the probability of H only under certain circumstances. Dardashti et al. identify the following four conditions that have to be assumed in addition to the structure of the Bayesian network depicted in Figure 4:

(i) $0 < P(X) < 1$
(ii) $P(H|X) > P(H|\neg X)$
(iii) $P(H'|X) > P(H'|\neg X)$
(iv) $P(E'|H') > P(E'|\neg H')$

Condition (i) states that the probability distribution over X is not extreme. Conditions (ii) and (iii) say that H and H' must become more plausible in the light of X. Finally, condition (iv) is a necessary condition for E' to be considered evidence for H'. It can be proven that $P(H|E') > P(H)$ holds if conditions (i) through (iv) are satisfied and the probability distribution P over $\mathbf{V} = \{X, H, H', E'\}$ factors according to equation 1 (Dardashti et al. 2019, Theorem 1). Dardashti et al. see this as support for the view that evidence E' of the source system can actually confirm hypothesis H about the target system. As a measure for indirect confirmation by analogy, one could, according to this approach, simply use the ordinary Bayesian difference measure

$$Bconf(H|E') = P(H|E') - P(H). \tag{2}$$

In this paper, we use this particular measure as a proxy for all kinds of Bayesian confirmation measures. (For an overview, see Fitelson 1999.)

Note that the Bayesian approach proposed by Dardashti et al. (2019) amounts to a model of what we labeled type III analogical inference in section 2. It aims at confirming H on the basis of E'—where the confirmational impact can be expressed by $Bconf(H|E')$—which corresponds to the confirmatory relation indicated by the thick diagonal arrow in Figure 2. Their approach can be easily expanded to a model that can also cover type I analogical inference by adding a variable E standing for (unknown) direct evidence for H to one's model and by assuming the additional condition (v) $P(E|H) > P(E|\neg H)$. It follows that $P(E|E') > P(E)$ holds and, hence, that E can be Bayes confirmed by E', which corresponds to the confirmatory relation expressed by the thick horizontal arrow in the type I pattern in Figure 2. The confirmational impact can, again, be expressed by $Bconf(E|E')$.[3]

4. Shared structures, common causes, and problems with indirect confirmation

In this section, we will do two things: First, we will generalize Dardashti et al.'s (2019) Bayesian approach to indirect confirmation by analogy in such a way that it can also cover causal scenarios. Since analogies involving causal dependencies are particularly appreciated in the traditional literature on analogies (Bartha, 2010), supplementing their approach with such an interpretation seems promising. To this end, we will start this section by introducing the basics required for the causal interpretation of Bayesian networks. Second, we will highlight possible problems this

[3]Note that confirmation is typically understood relative to some background knowledge K. To account for the role this background knowledge plays, the confirmatory impact of E' on H would better be measured by $P(H|E', K) - P(H|K)$ than by $P(H|E') - P(H)$. In order to keep things simple, we bracket background knowledge K in our discussion. It is, however, important to mention that Bartha (2019, sect.5.1) discusses the worry that analogical arguments might not be able to provide confirmation because it is inappropriate to treat E' as evidence. Rather, E' should be seen as part of the background knowledge K. Observing E' would, thus, not increase the probability of H given K. (Note that this is a version of the *problem of old evidence*). We agree with Dardasthi et al.'s (2019, sec. 2) argumentation for why E' should be considered as *new* evidence anyway.

generalized approach has to face. These problems do not depend on causal setups and, hence, might pose a threat also to Dardashti et al.'s original non-causal approach. However, they are not intended to deny the rationale of confirmation by analogy type I and III in general. By putting them forward we rather want to stress that inferences by analogy type II (which are not exposed to these problems) and other strategies for using evidence of the source system for confirming a hypothesis about the target system are significant as well.

The causal interpretation of Bayesian networks was developed by Spirtes, Glymour, and Scheines (1993) and later by Pearl (2000). In this interpretation, the arrows of a Bayesian network are interpreted as direct causal dependencies: If $X_i \to X_j$, then X_i is a direct cause of X_j. If there is a path of the form $X_i \to \ldots \to X_j$, then X_i is called a (direct or indirect) cause of X_j. If a variable X_k lies on a path of the form $X_i \to \ldots \to X_k \to \ldots \to X_j$, then X_k is called an intermediate cause. A variable X_k lying on a path $X_i \leftarrow \ldots \leftarrow X_k \to \ldots \to X_j$ is called a common cause of X_i and X_j, and a variable X_k lying on a path $X_i \to \ldots \to X_k \leftarrow \ldots \leftarrow X_j$ is called a common effect of X_i and X_j.

Under the causal interpretation of Bayesian networks, the Markov factorization (equation 1) implies (among other things) that conditionalizing on intermediate causes or common causes X_k on a path $X_i \to \ldots \to X_k \to \ldots \to X_j$ or $X_i \leftarrow \ldots \leftarrow X_k \to \ldots \to X_j$, respectively, screens X_i and X_j off each other (provided X_i and X_j are not connected via other paths), and that variables X_i and X_j, connected only via a common effect path $X_i \to \ldots \to X_k \leftarrow \ldots \leftarrow X_j$, are independent but might become dependent after conditionalizing on their common effect X_k or on one of its effects.

Note that from a formal point of view, all that a Bayesian network does is provide a graphical representation from which certain independencies in an associated probability distribution can be read off. A common direct ancestor structure such as $H \leftarrow X \to H'$, for example, produces the same screening off effect regardless of whether the directed edges are causally interpreted or not. Whatever the true nature of the relation between X and the hypotheses H and H' might be, as long as it has the right formal properties, it can be represented by the structure $H \leftarrow X \to H'$.[4] If dependence due to shared structure is adequately represented by $H \leftarrow X \to H'$, then it seems to be straightforward to assume that if Dardashti et al.'s (2019) story about indirect confirmation via a path $H \leftarrow X \to H'$ through a variable X representing relevant shared structural properties is correct, then E' should also confirm H if X represents a common cause of H and H' instead of a shared structural feature as long as the conditions (i) through (iv) from section 3 are satisfied. Indirect confirmation via common cause paths would work exactly like indirect confirmation by analogy in Dardashti et al.'s approach does. Thus, Dardashti et al.'s Bayesian approach to indirect confirmation by analogy type III (and also type I) can be straightforwardly generalized in such a way that it also covers causal structures.

Let us now highlight several possible problems with this generalized Bayesian approach to indirect confirmation. The first two problems (subsections 4.a and 4.b) can be solved. We will argue that the other problems (subsection 4.c) should raise doubt with respect to the kind of cross-system confirmation employed by type I and III analogical inferences. Our diagnosis will be that sometimes confirmation should be restricted to system-internal confirmation and, hence, to analogies of type II (to be discussed in section 5). We will also suggest an alternative procedure for how one can confirm H on the basis of collecting evidence E' that can avoid these problems.

4.a Paths between hypotheses

The first possible objection one might raise is that it is not clear how to draw the arrows between the variables X, H, H', and E'. It is quite uncontroversial that hypothesis H' and evidence E' are correctly represented by $H' \to E'$. (The reason for this is simply that further evidence for H' is typically

[4]It seems, for example, that supervenience, constitutive relevance, and the grounding relation also produce the Markov factorization (see, e.g., Gebharter 2017a, 2017b; Schaffer, 2016).

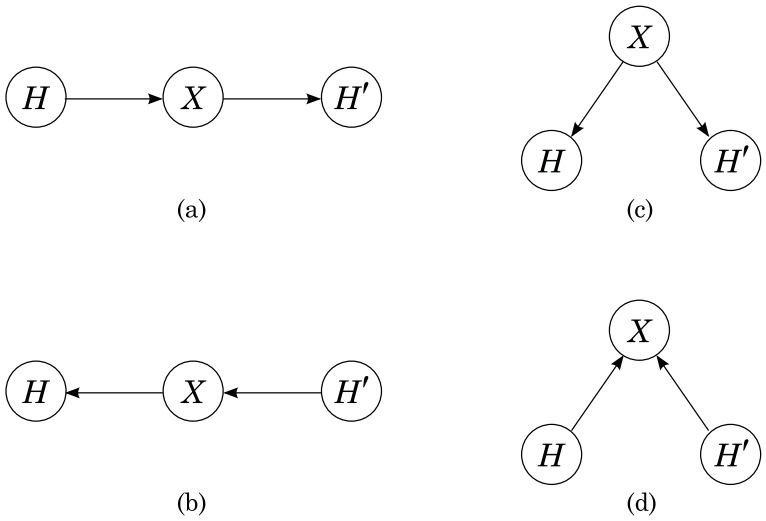

Figure 5. Four possibilities of connecting H and H' via X.

assumed to be independent of E' conditional on H'.) But what about the connection of X to H and H'? Recall Dardashti et al. (2019) originally introduced X to mediate probabilistic information between H and H'. There are basically four possible structures that would allow X to play this role (see Figure 5). Since X represents something the systems described by H and H' have in common, it seems plausible to eliminate possibilities (a) and (b). But why should (c) and not (d) be the correct structure? In the common cause scenario it is quite trivial that (c) is the right representation. But what about the shared structure interpretation?

We think that we can come up with the following argumentation for why (c) is also the correct representation in the shared structure case: (c) seems, contrary to (d), to be able to capture just the right dependencies and independencies. Assume that the probability distribution over X is not extreme (i.e., satisfies condition (i) from section 3). Now H and H' can be expected to be dependent. When we learn, for example, H', then this might increase the probability of H simply because H and H' partially share the same structure (described by X). But once we conditionalize on X, H and H' will become independent. When fixing X's value, the general structural features (represented by X) of the systems described by H and H' cannot change anymore. So all changes in H or H' must be independent of these structural features. But since probability flow between H and H' is only established via the shared structure described by X, learning about other features of the system described by H' would give us no additional probabilistic information for H and vice versa. The dependence and independence features we have described are exactly the ones that come with structure (c); they are, on the other hand, incompatible with structure (d). This will become even clearer in the next subsection.

4.b Choosing X and the Markov factorization

Here is another possible problem that is somehow related to the first one. To establish probability flow between the target and the source system, not only conditions 1 through 4 are required, but one must also be justified in assuming that $H \leftarrow X \rightarrow H' \rightarrow E'$ is a Bayesian network—one must be justified in assuming that one's probability distribution factors according to equation 1. In particular, X does not only have to mediate between the source and the target system, but also to screen the two systems off each other if conditioned on. A necessary condition for the latter is that

$$P(H|\neg X, H') = P(H|\neg X) \tag{3}$$

holds. But why exactly should X satisfy this screening off condition? The problem is that X cannot be chosen just in any way sufficient for establishing probability flow between the target and the source system. Assume, for example, X would have the two possible values $H \wedge H'$ and $\neg H \vee \neg H'$.[5] We would then get $P(H|\neg X, H') = P(H|\neg H \vee \neg H', H') = 0$ and $P(H|\neg X) = P(H|\neg H \vee \neg H') > 0$ and, thus, equation 3 would be violated. So, again, are there general rules how X must be specified for analogical inference, and how can introducing an X that satisfies these rules be justified?

First of all, note that finding and specifying X has two aspects: a theoretical and an empirical aspect. The theoretical task consists in two subtasks, in identifying general constraints for X that guarantee that X will have the right formal properties—it must mediate probabilistic influence between the target and the source system and screen both systems off each other if conditioned on—and in finding a general method for inferring the existence of such an X. Let us say a few words on the latter first. One possibility to infer the existence of a suitable X is creative abduction (Schurz 2008). The basic idea is that a stable strict correlation between two systems requires an explanation. According to Reichenbach's (1956) principle of the common cause, there must be a common cause if these systems are not directly causally connected. One key feature of common causes is that they screen off their effects, which guarantees that the inferred X has the right formal properties. This basic idea can be generalized to nonstrict correlations and noncausal structures as well (Feldbacher-Escamilla and Gebharter 2019). In that case, one infers dispositional or shared structural features of the correlated systems instead of a common cause (see also Glymour 2019).

Though creative abduction can be used for inferring the existence of a suitable X, this does not yet tell us more about X than that it must have the right formal properties. So, how could X generally be specified? With this question, we have arrived at the other part of the theoretical aspect. A possible answer to it comes again from the formal similarity between common causes and shared structural features. Common causes do screen off if they are maximal, i.e., if all common causes are subsumed under X. The same holds for shared components or constitutents (see, e.g., Gebharter 2017b) and should also hold for shared structural features. If probability flow between the target and the source system is due to shared structural features, then X must, to fully explain this correlation and to guarantee the screening off property, cover all the features shared by both systems. One can, for example, think in terms of INUS conditions (Insufficient but Necessary parts of a condition which is itself Unnecessary but Sufficient; *cf.* Mackie 1980) about these matters. Assume C is an effect to be explained and A and B are causally relevant factors for C. Their logical connection to C can then be described via

$$AU \vee BV \vee W \leftrightarrow C, \tag{4}$$

where U, V, and W stand for further possibly unknown explanatory relevant factors. Note that each causally relevant factor (such as A and B) is an INUS condition for C, i.e., an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition for C. Similar representations have been used for constitutively relevant parts (Couch 2011; Harbecke 2010) and might, in the same way, be used for characterizing structural relevance as well. Note that for establishing probability flow between H and H', two descriptions such as equation 4 having at least one INUS condition in common are required. These descriptions could, for example, be

$$AU \vee BV \vee W \leftrightarrow H \tag{5}$$

and

$$AU' \vee BV' \vee W' \leftrightarrow H' \tag{6}$$

[5]We are indebted to an anonymous referee for pointing us to this problem.

which share the structural features A and B. Now, for X to screen H and H' off each other, it is required that all INUS conditions appearing in both descriptions are featured in X. In particular, X's values must be the possible conjunctions of INUS conditions (and their negations) appearing in both descriptions; it is in this sense that X must be maximal. According to this idea, X would, in our example, have the four values $A \wedge B$ (both structural features are present), $A \wedge \neg B$ (only A is present), $\neg A \wedge B$ (only B is present), and $\neg A \wedge \neg B$ (neither A nor B are present).

We think that there is not much more that can be said about X from a theoretical perspective. We pointed to a method for inferring the existence of a suitable X and tried to provide a general logical pattern for how X could be specified in order to have the right formal properties. Inferring and specifying X in concrete cases (such as the dynamics of black holes studied by Dardashti et al. 2019, Dardashti et al. 2015, or our rat case example) cannot be done a priori. Filling in these details is the empirical task mentioned before that might require different methods depending on the specific domain of application and the specific science (or sciences) concerned.

4.c Possible problems with cross system confirmation

In what follows, we put forward another type of problems whose brief discussion aims at a further clarification of the model(s) on confirmation based on analogical reasoning. As we will argue, confirmation across systems (type I and type III) might undermine intuitions about *confirmation* by licensing inflation, scattering, employment of distrusted evidence, and outweighing of direct evidence by indirect evidence.

Confirmation inflationism. This problem arises for the shared structure as well as for the common cause interpretation. Let us have a look at the common cause interpretation first. Physicists are quite confident (but not absolutely certain) about the existence of the Big Bang. Let us say that X describes the Big Bang. Since the Big Bang is a common cause of everything going on in the universe at any later point in time, we can formulate hypotheses H and H' about any two systems s and s' in the world. Given the generalization of Dardashti et al.'s (2019) story about indirect confirmation—given evidence E' can actually confirm a hypothesis H if there is some mediator X between H and H'—then evidence E' for any hypothesis H' indirectly confirms hypothesis H about (almost) any other system simply because all current systems have the Big Bang as a common cause. Everything required for indirect confirmation in a Bayesian framework is that conditions (i) through (iv) from section 3 are satisfied, which seems to be very plausible in the case of the Big Bang.

A similar point can be made for the shared structure interpretation. It seems that for almost all real world systems there are some common structural features that could be described by X. Note that conditions (i) through (iv) from section 3 may already be satisfied if there are very weak dependencies of hypotheses H and H' on X. This means that if all relations of confirmation by analogy would be modelled by analogical confirmation type I and III only, then almost everything would at least very weakly confirm almost everything else. We would have to face some kind of full blown confirmation inflationism.

Confirmation of scattered hypotheses. Here is another possible problem. It is related to the last one. If the generalized Bayesian approach to indirect confirmation is correct, then confirmation just consists in Bayesian update, meaning that everything needed for a hypothesis H to be confirmed by some E' is that $P(H|E') > P(H)$ holds. But then variables that are quite scattered across a Bayesian network adequately representing a scenario of interest can, in principle, be used for confirmation. $P(H|E') > P(H)$ might already hold in some circumstances if there is some (maybe complex) path connecting E' with H in such a way that it can propagate probabilistic influence from E' to H. This leads to cases of confirmation that seem, at least at first glance, quite odd. Let us illustrate this by means of the exemplary Bayesian network whose structure is depicted in Figure 6: the ALARM network which was originally developed as a tool for monitoring medical systems of a certain kind (Beinlich, Suermondt, Chavez, and Cooper 1989).

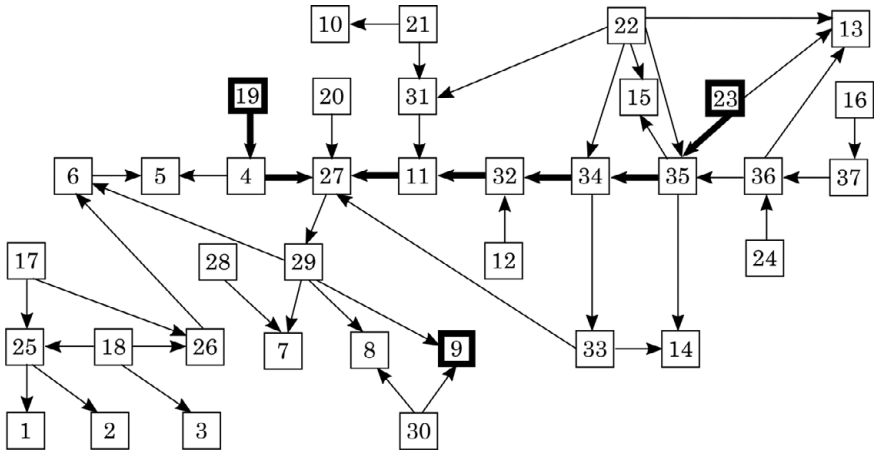

Figure 6. The ALARM belief network (reconstruction based on Spirtes et al., 1993, p. 12).

The details of the ALARM network and the medical system modeled are not that important for the point we want to make. Let us rather focus on the following two variables of the ALARM network: 19 and 23. 19 could represent the hypothesis that an anaphylaxis occurs, and 23 the hypothesis that the ventilation tube is kinked. Now 19 and 23 are not directly connected to each other in the ALARM network; they are rather scattered. They are, however, connected by several quite complex paths. The path consisting of the bold arrows in Figure 6, for example, can transport probabilistic dependence between 19 and 23 conditional on 9 (which stands for heart rate obtained from oximeter). Now assume that we observe some evidence E' for 23. Since conditionalizing on 23 increases the probability of 19 conditional on 9, it might well be the case that $P(19|E',9) > P(19|9)$ also holds. But in the generalized Bayesian approach to indirect confirmation, this just means that E' confirms 19 in the context of 9. However, one might hesitate to speak of confirmation in this situation simply because the hypotheses 19 and 23 are too scattered. It is not clear why E' should count as evidence for 19.

Distrust in indirect evidence. The third possible problem is that the generalized Bayesian approach to indirect confirmation by analogy type I and III does not seem to be in line with how many scientists understand and use the concept of confirmation as well as with the common sense understanding of that concept. Let us illustrate this by means of the rat study example introduced in section 3. Assume that Jones was affected with a certain virus. She goes to a physician who offers her two possible treatments: medical compound A or B. A has been tested on humans and it turned out that 80% of humans affected with the virus recovered after taking A. B, on the other hand, has only been tested on rats so far. Let us assume that 95% of rats affected with the virus recovered after taking B. Let us further assume that physicians have used Dardashti et al.'s (2019) method of indirect confirmation. They chose the priors of X and everything else required for indirect confirmation to the best of their knowledge and are quite certain that no mistakes occurred. They might then come to the conclusion that 80% of humans affected with the virus should recover after taking compound B. So which compound should Jones choose? Our guess is that almost every patient (and physician) would prefer compound A over B, simply because A has been tested directly on humans. The rat study investigating B clearly gives us probabilistic information for whether B will work when given to humans and makes it equally plausible that B will lead to recovery when given to affected humans. Still, we distrust B to some extent. Maybe the best explanation for this is that we hesitate to accept results from the rat study as evidence for hypotheses about B's effect on human recovery. If this diagnosis is correct, then at least sometimes confirmation has to be constrained by structural rules (system internal confirmation versus confirmation across systems).

Irrelevance of direct evidence. Another possible problem for considering the generalized Bayesian approach to confirmation of a hypothesis H by analogy type III is that direct evidence E for the hypothesis H about the target system does not play any role for confirming H in that approach. All the confirmatory work is actually done by evidence E' for the hypothesis H' about the source system and the connection between H and H' established by the common cause or shared structure X. However, E often plays an important role in confirming H in scientific practice, even when scientists reason on the basis of analogies. If the generalized Bayesian approach to confirmation by analogy type III were the only mode of indirect confirmation of a hypothesis H about the target system, then it is not clear why, in cases such as the rat study example, scientists are looking for direct evidence E though they have already confirmed hypothesis H about the target system indirectly. Even more problematic, it is easy to find scenarios in which indirect evidence E' has more confirmatory impact on H than direct evidence E has. Here is an exemplary probability distribution:

$P(X) = .65$	$P(H\|X) = .75$	$P(E\|H) = .55$	$P(H'\|X) = .75$	$P(E'\|H') = .75$
	$P(H\|\neg X) = .25$	$P(E\|\neg H) = .45$	$P(H'\|\neg X) = .25$	$P(E'\|\neg H') = .25$

From this distribution the impact of indirect evidence E' on H can be computed as $Bconf(H|E') = .053$, which overshoots the impact of direct evidence E on H which can be computed as $Bconf(H|E) = .048$.

Here is our diagnosis of the possible problems highlighted above. We think that the culprit is an unconstrained application of Bayesian confirmation. In particular, it seems to be problematic to only rely on probabilistic information $P(H|E') > P(H)$ when doing confirmation, especially if there is also structural information available. The situation seems to be similar to what was historically going on in several other philosophical debates such as debates about explanation, causation, decision theory, etc. According to Hempel's (1965, 338) classical inductive-statistical model of explanation, for example, an explanation of a phenomenon described by E consists of a statistical law of the form $P(E|C_1,\ldots,C_n) \approx 1$ and statements $C_1,\ldots,C_n$ about the occurrence of certain events. If $P(E|C_1,\ldots,C_n) \approx 1$ and $C_1,\ldots,C_n$ together make E very likely, then $P(E|C_1,\ldots,C_n) \approx 1$ and $C_1,\ldots,C_n$ explain E. After striking counterexamples (see, e.g., Salmon 1984) it turned out that structural information was missing in Hempel's approach: for a successful explanation it is also required that the events described by $C_1,\ldots,C_n$ are causally relevant for the event described by E. Naïve probabilistic theories of causation also share this problem. They claim that C is a cause of E if and only if $P(E|C) > P(E)$ (this is, admittedly, an oversimplified version of a probabilistic theory of causation and more sophisticated probabilistic theories typically make additional assumptions; they might, for example, require that the cause always occurs before the effect, etc.). It turned out that a purely probabilistic criterion for causation is not enough. Here, one also needs additional information about structure. Modern accounts (e.g., Cartwright 1979) rather characterize causation by referring to additional causal (i.e., structural) information. Finally, a similar move can also be found in the philosophical literature on decision theory. It is well known that the purely probabilistic account (i.e., evidential decision theory) has to face several problems. Some of them can be solved by endorsing causal decision theory, which already comes with some structural constraints. The approaches proposed by Meek and Glymour (1994) and Hitchcock (2016) consider richer causal structural information, which allows them to handle even more problems. We think that structural restrictions might also help to overcome problems that a generalized Bayesian approach to analogical inference has to face.

In this section, we discussed three types of possible problems for analogical inference Bayesian style. We think that the first two (subsection 4.a and 4.b) can be solved. The third type of problems (subsection 4.c) concerns confirmation across systems and it seems that these problems force one

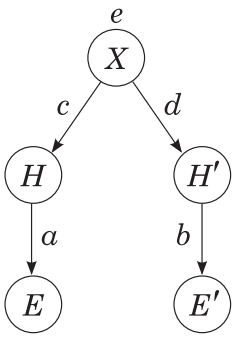

Figure 7. Parameters in the Bayesian network for analogical reasoning type II.

to make a choice depending on whether one finds them alarming. If one does, one might want to subscribe, as indicated above, to a structural constraint for confirmation:

(⋆). Only evidence E accessible via paths *within* a particular system s can be used for confirmation of a hypothesis H about s.

We do not want to commit ourselves to one of the two options mentioned. Rather, we think that whatever path one takes, it still stresses the relevance of confirmation by analogy type II. (For details, see section 5.) However, it is clear that if one subscribes to (⋆), then cross-system confirmation of type I and III is excluded and type II is the only classical inference pattern (of the three introduced in section 2) left. But even if one considers indirect confirmation of type I and III to be perfectly fine, it seems that type I and III only work within their proper bounds: when considering confirmation on the quantitative scale, for example via $Bconf$, then the problems of confirmation inflationism and confirmation of scattered hypotheses might vanish simply because the degree of the general background confirmation or that of confirmation of scattered hypotheses might be so low that it can be considered negligible. However, these strategies will not work for a purely qualitative model of confirmation.

In the next section, we will discuss alternatives that conform to (⋆) and, thus, are not plagued by the cross-system problems discussed in subsection 4.c. If evidence E of the target system is accessible, one can go for the last remaining classical inference pattern: type II analogical inference. For the case that E is inaccessible, we will propose an alternative to cross-system confirmation of type I and III that adheres to the structural constraint (⋆).

5. Confirmation by analogy: Bayes meets Jeffrey

Let us start this section with type II analogical inference. The basic idea is to use information about confirmation within the source system s' for making confirmational claims about the target system s. Schematically, the problem is to define $Bconf(H|E)$ on the basis of $Bconf(H'|E')$. This corresponds to establishing the confirmational relation indicated by the vertical thick arrow in Figure 2. To this end, we first need to add a variable for evidence about the target system (E) to our model (see Figure 7). Next, we have to think about how to establish the vertical confirmatory relation within the target system on the basis of what we can learn about the source system. Note that we cannot simply equate $Bconf(H|E)$ with $Bconf(H'|E')$. According to equation 2, $Bconf(H|E)$ is defined on the basis of $P(H|E)$ and $P(H)$ and, hence, $Bconf(H|E)$ cannot be determined independently of the Bayesian network's parameters a, c, and e. (Likewise, $Bconf(H'|E')$ is defined on the basis of $P(H'|E')$ and $P(H')$ and $Bconf(H'|E')$ cannot be specified independently of the parameters b, d, and e.) The problem is that the nonparameter probabilities within a Bayesian network are fully determined by and cannot be varied independently of the Bayesian network's parameters a through e, where $a = \langle P(E|H), P(E|\neg H)\rangle$, $b = \langle P(E'|H'), P(E'|\neg H')\rangle$, $c = \langle P(H|X), P(H|\neg X)\rangle$,

$d = \langle P(H'|X), P(H'|\neg X)\rangle$ and $e = P(X)$. Since parameters b through e are already fixed, specifying a is the only way to influence the probabilities $P(H|E)$ and $P(E)$ relevant for determining $Bconf(H|E)$. Because of this, we suggest implementing type II analogical inference via equating parameter a of the target system with the respective parameter b of the source system: $a = b$. This move amounts to inferring a similar vertical relation for the target system as for the source system on the basis of the similarity between these systems expressed via $H \leftarrow X \rightarrow H'$.

Note that one could say a little bit more about the similarity relations between features of the source and the target system in terms of the parameters c and d. That the first element of $c = \langle P(H|X), P(H|\neg X)\rangle$ is significantly greater than the first element of $d = \langle P(H'|X), P(H'|\neg X)\rangle$ could, for example, represent the fact that the structural features modeled by X have more impact on the target system than on the source system. Equal parameters ($c = d$), on the other hand, could model perfect similarity between the impact of these features on s and s'. In case of such a perfect similarity, making a type II analogical inference, i.e., identifying a with b, implies that E's confirmational impact on H equals E''s confirmational impact on H':

Observation: *If* $c = d$ *and* $a = b$, *then* $Bconf(H|E) = Bconf(H'|E')$.

Note that type II analogical inference adheres to the structural constraint (*) from section 4: confirmation does not happen over cross-system pathways. Thus, type II analogical inference does not have to face the problems of cross-system confirmation discussed in subsection 4.c. However, type II analogical inference clearly requires that evidence E of the target system is accessible. This requirement will often not be satisfied. This was what made type I and III analogical inference so tempting in the first place. So, what if E is inaccessible? Can we still somehow use E for confirming H given the similarity between the source and the target system while, at the same time, respecting the structural constraint (*)? This question will be our focus in the remainder of this section. It is especially relevant if we find the problems discussed in subsection 4.c alarming and are somewhat skeptical regarding type I and III analogical inference.

Here comes our alternative to cross-system confirmation of type I and III: we propose a two-step approach: in the first step, one updates the probability distribution over the target system by conditionalizing on evidence E' of the source system. So the first step consists in Bayesian updating; no confirmation is involved here and, thus, (*) is respected. Confirmation happens system-internal in the second step: here we use the updated probability of evidence E to confirm H. Of course, E is not observed (in the strict sense); in the first step we have just reduced our uncertainty about E to some degree. For this reason, we have to use an update method in step 2 that can handle uncertain evidence. A standard method for this is Jeffrey conditionalization (Jeffrey 1983).

If one is ready to subscribe to (*), then one still needs to make sense of scientific practice as, for example, illustrated in the rat study case. But if one finds cross-system confirmation type I and III problematic, then one will not find the reconstruction of the rat study case provided in section 3 very tempting. In the following, we will use this example again for two purposes: First, to show how it can be rationally reconstructed in terms of our two-step approach if one endorses (*), and second to illustrate the two-step approach. In the rat case example physicians were, for moral reasons, not able to collect evidence E about how humans react to the antiviral compound under investigation. So they were not able to directly confirm the hypothesis H about the human immune system. For this reason, they studied the effects of the antiviral compound on a model organism such as rats. Let us assume that the physicians were able to gather evidence E' for a similar hypothesis H' about the immune system of rats. Since it has already been empirically established that the source system and the target system share relevant structural features, they infer, from observing E', that if they were to start testing the antiviral compound on humans, they would very likely find evidence E. This is the first step in our two-step approach based on ordinary Bayesian updating; no confirmation is involved here. They then use this inferred decreased uncertainty about E as evidence for H, meaning

that they infer that it is sufficiently likely that infected humans will recover without severe side effects if treated with the antiviral compound. This is the second step in our two-step approach; it is based on Jeffrey conditionalization and marks the point where (system-internal) confirmation happens. Having confirmed H that way will be sufficient for physicians to start tests with humans to further confirm H (directly).

Here is our story again, this time with all the technical details: we assume that the rat study example can be adequately represented by a Bayesian network with the structure depicted in Figure 7. To confirm H, one needs access to direct evidence E. But, in our example, E cannot be observed directly. However, what the experts have available is their (prior) estimation of the probability that direct evidence E would occur if the antiviral compound were administered to affected humans. What they can do now is employ the information they gathered from their investigation of the model organism to update the probability $P(E)$ Bayesian style to $P^*(E)$ via

$$P^*(E) = P(E|E'). \tag{7}$$

Since E' is ordinarily gathered evidence—the physicians just observe E'—this move seems to be straightforward. In this way, evidence E' about the source system is used to achieve an empirically informed estimation of the probability of the direct evidence E. Now this updated probability $P^*(E)$ can be used to confirm H directly. Because E is still uncertain—not observed in the strict sense—we use Jeffrey conditionalization for this last step. We say that uncertain evidence E confirms H by analogy if

$$P(H|E) \cdot P^*(E) + P(H|\neg E) \cdot P^*(\neg E) > P(H) \tag{8}$$

holds. The left part of this inequality expresses the impact of direct but uncertain evidence E on hypothesis H Jeffrey style (i.e., the impact on H given the updated probability of E). We have to use the prior conditional probabilities $P(H|E)$ and $P(H|\neg E)$ from the original distribution P here. Using the posterior conditional probabilities $P^*(H|E)$ and $P^*(H|\neg E)$ would amount to computing the impact of E' on H by simple Bayesian conditionalization and, thus, we would face at least the problems of distrust in indirect evidence and irrelevance of direct evidence which we highlighted in subsection 4.c. The structural constraint **(*)** which only allows for confirmation within a system impedes to directly use external evidence for confirmation. However, such evidence is not completely irrelevant: even if E' cannot be used to confirm H, observing E' can still (cross system-wise) decrease uncertainty about our expectations of direct evidence E. We can then use this decreased uncertainty to directly confirm (system-internal) H.

Before we go on to define a measure for this kind of confirmation, note that Jeffrey conditionalization (but also Bayesian conditionalization) requires a rigidity condition to hold: one has to use the *initial* conditional probabilities in calculating the *unconditional* ones. Since the conditional probabilities remain unchanged, our two-step approach adheres to this rigidity condition. Next to the uncertainty of evidence E of the target system, this is another important respect in which the two-step approach differs from type II inference by analogy in form of parameter mapping: parameter mapping changes the conditional probabilities and, hence, is a form of nonrigid updating.

Based on the two-step approach to confirmation outlined above, we can define the simple incremental confirmation measure

$$BJconf_{E'}(H|E) = [P(H|E) \cdot P^*(E) + P(H|\neg E) \cdot P^*(\neg E)] - P(H). \tag{9}$$

BJ stands for the two steps: *B* for the first step using Bayesian update of the probability of E in the light of E' and *J* for the second step using Jeffrey conditionalization for confirmation.

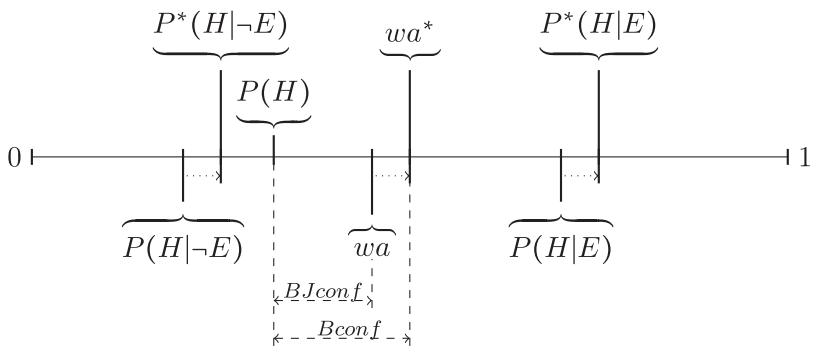

Figure 8. Illustration of confirmational impact of Bconf and BJconf.

A similar result as the one Dardashti et al. (2019) proved for $Bconf$ can be proven for $BJconf$ if

$(v)\ P(E|H) > P(E|\neg H)$

is added as a fifth condition to (i) through (iv) from section 3. Condition (v) is analogue to (iv). It guarantees that E can serve as evidence for H. If all five conditions are satisfied and the probability distribution P over $\mathbf{V} = \{X, H, H', E, E'\}$ factors according to equation 1, where the parents are determined by the structure of the Bayesian network in Figure 7, then E confirms H on the basis of observing E', even if E cannot be accessed directly.

Theorem 1. $BJconf_{E'}(H|E) > 0$, *if (i) through (v) are satisfied.*

Before we go on, let us briefly compare the measures $Bconf$ and $BJconf$. It can be shown that under conditions (i) through (v) $BJconf$ is always weaker than or equal to $Bconf$.

Theorem 2. $BJconf_{E'}(H|E) \leq Bconf(H|E')$, *if (i) through (v) are satisfied.*

The confirmational impact according to the two measures can be illustrated by means of Figure 8: the formula in the squared brackets of equation 9 is a weighted average wa with maximum $P(H|E)$ and minimum $P(H|\neg E)$ and weights $P^*(E)$ and $P^*(\neg E)$, respectively. According to Theorem 1, wa lies above $P(H)$ and the distance between wa and $P(H)$ corresponds to the degree of confirmation according to $BJconf$. When replacing $P(H|E)$ with $P^*(H|E)$ and $P(H|\neg E)$ with $P^*(H|\neg E)$ in equation 9, one gets Dardashti et al.'s (2019) confirmation measure $Bconf$. According to Theorem 2, wa^* of the posterior probabilities is equal to or lies above wa used in our measure $BJconf$ simply because the former maximum and minimum are equal or lie above the latter. Again, the distance between wa^* and $P(H)$ represents the degree of confirmation according to $Bconf$.

Recall from section 4 that, according to Dardashti et al.'s (2019) approach, indirect evidence E' sometimes provides more confirmation of a hypothesis H than direct evidence E for H, which contradicts in relevant cases our intuitions and scientific practice. The next theorem shows that the two-step approach does not share this problem with the generalized Bayesian approach to confirmation by analogy type III.

Theorem 3. $BJconf_{E'}(H|E) \leq BJconf_{E}(H|E)$, *if (i) through (v) are satisfied.*

This result fits nicely with our intuitions and scientific practice. It shows that our approach allows for an explanation of why, for example, the scientists in the rat study scenario still collect direct evidence for the efficacy of the antiviral compound on humans though they have already

tested it on rats: in this case, direct evidence has an impact at least as strong as evidence from different but structurally analogous systems; even more, direct evidence will almost always have more impact. So, we think that in the case of the rat study, for example, the Bayes-Jeffrey measure for confirmation provides a more adequate model of the confirmational relation in question than the Bayes measure for type III confirmation.

6. Conclusion

This paper started with the observation that in many cases it is not possible to directly observe evidence E for a hypothesis H. In such cases scientists often investigate different but to some degree analogous systems, models, or simulations instead. They consider H to be in some sense confirmed if they succeed in collecting evidence E' for a corresponding hypothesis about such a source system. In section 2, we have, based on the traditional literature on analogies, identified three types of analogical inference: type I, which establishes a horizontal relation between the hypotheses or pieces of evidence and statements of the source and target system (E',E); type II, which establishes a vertical relation between the evidence and hypothesis of the target system (E,H) on the basis of such a relation within the source system (E',H'); and type III, which establishes a diagonal relation between evidence of the source system and a hypothesis of the target system (E',H). In section 3, we introduced Dardasthi et al.'s (2019) approach, according to which confirmation by analogy simply consists in standard Bayesian update. We argued that it can cover type III analogical inference and can be expanded to cover type I. In section 4, we generalized their approach to scenarios in which common causes play the role of analogies. We then discussed several possible problems for this generalized approach. Our diagnosis was that the more serious problems (subsection 4.c) might arise due to missing structural constraints. In section 5, we proposed a model of confirmation by analogy type II which does not fall victim to these problems. It requires, however, that evidence E of the target system is accessible. If one takes the problems discussed in subsection 4.c serious and E is inaccessible, then one can replace cross-level confirmation type I and III by a two-step approach adhering to the structural constraint (*) which requires that only paths within a system can be used for confirmation. In a first step, one decreases uncertainty about direct evidence E of the target system by Bayesian update on evidence E' of the source system. Confirmation happens in the second step. Here we use Jeffrey conditionalization on the decreased uncertainty about E to confirm H. We finally showed that our two-step approach allows for confirmation of H by E on the basis of E' (if conditions (i) through (v) are satisfied; Theorem 1). We also demonstrated that our measure of confirmation $BJconf$ is always weaker than or equal to $Bconf$ (Theorem 2) and that confirmation of H by E based on E' can never overshoot direct confirmation of H by E (Theorem 3). Finally, it seems appropriate to once more emphasize that this paper was about the theoretical underpinning of analogical inference to the background of a Bayesian setting. In the end, case studies from actual science more realistic than our simple rat study example are required to further test its adequacy and to explore specific empirical methods for establishing structural similarities X satisfying all the requirements specified throughout the paper.

Acknowledgements. This work was supported by Deutsche Forschungsgemeinschaft (DFG), research unit Inductive Metaphysics (FOR 2495). We would like to thank Radin Dardashti, Stephan Hartmann, Daniel Koch, Winfried Löffler, Gerhard Schurz, Karim Thébault, and Marieke Williams for important discussions and two anonymous referees for helpful comments.

Christian J. Feldbacher-Escamilla is a research fellow and lecturer at the Duesseldorf Center for Logic and Philosophy of Science (DCLPS). His area of expertise is within social epistemology and philosophy of science. For more information, see http://cjf.escamilla.one.

Alexander Gebharter is a postdoc at the Department of Theoretical Philosophy at the University of Groningen. He is mainly interested in philosophy of science and its intersection with metaphysics and philosophy of mind. For more information, see www.alexandergebharter.com.

References

Bartha, P. (2010). *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. New York: Oxford University Press.

Bartha, P. (2019). "Analogy and Analogical Reasoning." In *The Stanford Encyclopedia of Philosophy* (Spring 2019 ed.), edited by E. N. Zalta. Stanford, CA: Metaphysics Research Lab, Stanford University.

Beinlich, I., H. Suermondt, R. Chavez, and G. Cooper. (1989). "The ALARM Monitoring System: A Case Study with Two Probabilistic Inference Techniques for Belief Networks." In *Proceedings of the Second European Conference on Artificial Intelligence in Medicine*, London, August 29–31, 1989 (247–56). Berlin: Springer.

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

Carnap, R. (1962). *Logical Foundations of Probability*. London: Routledge and Kegan Paul.

Cartwright, N. (1979). "Causal Laws and Effective Strategies." *Noûs*, 13 (4): 419–37.

Couch, M. B. (2011). "Mechanisms and Constitutive Relevance." *Synthese*, 183 (3): 375–88.

Dardashti, R., S. Hartmann, K. Thébault, and E. Winsberg. (2019). "Hawking Radiation and Analogue Experiments: A Bayesian Analysis." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics.* doi: 10.1016/j.shpsb.2019.04.004

Dardashti, R., K. Thébault, and E. Winsberg. (2015). "Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity." *British Journal for the Philosophy of Science*, 68 (1): 55–89.

Duhem, P. M. M. (1991). *The Aim and Structure of Physical Theory*. Princeton, NJ: Princeton University Press.

Feldbacher-Escamilla, C. J., and A. Gebharter. (2019). "Modeling Creative Abduction Bayesian Style." *European Journal for Philosophy of Science*, 9 (1): 9.

Fitelson, B. (1999). "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity. *Philosophy of Science*, 66 (3): S362–78.

Gebharter, A. (2017a). "Causal Exclusion and Causal Bayes Nets." *Philosophy and Phenomenological Research*, 95 (2): 353–75.

Gebharter, A. (2017b). "Uncovering Constitutive Relevance Relations in Mechanisms." *Philosophical Studies*, 174 (11): 2645–66.

Glymour, C. (2019). "Creative Abduction, Factor Analysis, and the Causes of Liberal Democracy." *Kriterion: Journal of Philosophy*, 33 (1): 1–22.

Harbecke, J. (2010). "Mechanistic Constitution in Neurobiological Explanations." *International Studies in the Philosophy of Science*, 24 (3): 267–85.

Hartmann, S., and J. Sprenger. (2011). "Bayesian Epistemology." In *The Routledge Companion to Epistemology*, edited by S. Bernecker and D. Pritchard, 609–20. London: Routledge.

Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hesse, M. B. (1964). "Analogy and Confirmation Theory." *Philosophy of Science*, 31 (4): 319–27.

Hesse, M. B. (1966). *Models and Analogies in Science*. Notre Dame, ID: University of Notre Dame Press.

Hesse, M. B. (1974). *The Structure of Scientific Inference*. Berkley, CA: University of California Press.

Hitchcock, C. (2016). "Conditioning, Intervening, and Decision." *Synthese*, 193 (4): 1157–76.

Jeffrey, R. C. (1983). *The Logic of Decision*. 2nd ed. Chicago: The University of Chicago Press.

Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.

Mackie, J. L. (1980). *The Cement of the Universe*. Oxford University Press on demand.

Meek, C., and C. Glymour. (1994). "Conditioning and Intervening." *British Journal for the Philosophy of Science*, 45 (4): 1001–21.

Pearl, J. (2000). *Causality*. 1st ed. Cambridge: Cambridge University Press.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Schaffer, J. (2016). "Grounding in the Image of Causation." *Philosophical Studies*, 173 (1): 49–100.

Schurz, G. (2008). "Patterns of Abduction." *Synthese*, 164 (2): 201–34.

Spirtes, P., C. Glymour, and R. Scheines. (1993). *Causation, Prediction, and Search*. 1st ed. Dordrecht: Springer.

Thomson, J. J. (1971). "A Defense of Abortion." *Philosophy and Public Affairs*, 1 (1): 47–66. http://www.jstor.org/stable/2265091.

Walton, D. (2005). *Fundamentals of Critical Argumentation*. Cambridge: Cambridge University Press.

Winsberg, E. (2009). "A Tale of Two Methods." *Synthese*, 169 (3): 575–92.

Appendix

For the proofs of the theorems below, we assume that conditions (i) through (v) hold and that the target and the source system are adequately represented by the Bayesian network in Figure 7.

Proof of Theorem 1. To show:

$$BJconf_{E'}(H|E) > 0$$

Proof. We first show that $P^*(E) > P(E)$. Since $P^*(\cdot) = P(\cdot|E')$ and the conditional probabilities $P(X_i|\mathbf{Par}(X_i))$ do not change after conditionalizing on nondescendants of X_i in a Bayesian network (equation 1), we can state $P^*(E)$ as

$$P^*(E) = \underbrace{P(E|H) \cdot P^*(H) + P(E|\neg H) \cdot P^*(\neg H)}_{:=t_2}.$$

It is a probabilistic fact that

$$P(E) = \underbrace{P(E|H) \cdot P(H) + P(E|\neg H) \cdot P(\neg H)}_{:=t_1}.$$

The terms t_2 and t_1 express weighted averages. From (Dardashti et al. 2019, Theorem 1) we know that $P^*(H) > P(H)$, and from (v) that $P(E|H) > P(E|\neg H)$. It follows that $P^*(E) > P(E)$.

Due to equation 9,

$$\underbrace{P(H|E) \cdot P^*(E) + P(H|\neg E) \cdot P^*(\neg E)}_{:=t_4} - P(H) \;=\; BJconf_{E'}(H|E) \tag{10}$$

holds. By probability theory, also

$$\underbrace{P(H|E) \cdot P(E) + P(H|\neg E) \cdot P(\neg E)}_{:=t_3} \;=\; P(H) \tag{11}$$

holds. The formulæ t_4 and t_3, again, express weighted averages. From (i), (ii), and (v) it follows that $P(H|E) > P(H|\neg E)$. Since $P^*(E) > P(E)$, it follows that $t_4 > t_3$. Then, by equation 10 and equation 11, we get $BJconf_{E'}(H|E) = t_4 - t_3 > 0$. □

Proof of Theorem 2. To show:

$$BJconf_{E'}(H|E) \leq Bconf(H|E')$$

Proof. According to Bayes' theorem,

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E|H) \cdot P(H) + P(E|\neg H) \cdot (1 - P(H))} \tag{12}$$

holds. Since $P^*(\cdot) = P(\cdot|E')$ and probabilities $P(X_i|\mathbf{Par}(X_i))$ do not change after conditionalizing on nondescendants of X_i in a Bayesian network (equation 1), also

$$P^*(H|E) = \frac{P(E|H) \cdot P^*(H)}{P(E|H) \cdot P^*(H) + P(E|\neg H) \cdot (1 - P^*(H))} \tag{13}$$

holds. If we define $a := P(E|H)$, $b := P(H)$, $b^* := P^*(H)$, $c := P(E|\neg H)$, $d := P(H|E)$, and $d^* := P^*(H|E)$, we can write equation 12 and equation 13 down as

$$d = \frac{a \cdot b}{a \cdot b + c \cdot (1 - b)} \quad \text{and} \quad d^* = \frac{a \cdot b^*}{a \cdot b^* + c \cdot (1 - b^*)}.$$

From (Dardashti et al. 2019, Theorem 1) we know that $b^* > b$, from (v) that $a > c$, and from (i) and (ii) that $0 < b < 1$. It follows that $d = d^*$ if $c = 0$ or $b^* = 1$, and that $d < d^*$ if $c > 0$ and $b^* < 1$. Thus, $P(H|E) \leq P^*(H|E)$ holds. Similarly, it can be shown that also $P(H|\neg E) \leq P^*(H|\neg E)$ holds.

Recall from section 5 that $Bconf$ can be formulated as

$$Bconf(H|E) = [P^*(H|E) \cdot P^*(E) + P^*(H|\neg E) \cdot P^*(\neg E)] - P(H), \tag{14}$$

and that $BJconf$ can be formulated as

$$BJconf_{E'}(H|E) = [P(H|E) \cdot P^*(E) + P(H|\neg E) \cdot P^*(\neg E)] - P(H). \tag{15}$$

Since the formulæ in the squared brackets are weighted averages and $P^*(H|E)$ and $P^*(H|\neg E)$ are greater than $P(H|E)$ and $P(H|\neg E)$ respectively, it follows that $BJconf_{E'}(H|E) \leq Bconf(H|E')$ holds. (For a graphical illustration, see Figure 8.) □

Proof of Theorem 3. To show:

$$BJconf_{E'}(H|E) \leq BJconf_{E}(H|E)$$

Proof. The left term of this inequality is defined as in equation 15. The right term is defined as

$$BJconf_{E}(H|E) = [P(H|E) \cdot 1 + P(H|\neg E) \cdot 0] - P(H). \tag{16}$$

If the weight $P^*(E)$ in equation 15 equals 1, then $BJconf_{E'}(H|E)$ equals $BJconf_{E}(H|E)$, and if $P^*(E)$ is smaller than 1, then $BJconf_{E'}(H|E) < BJconf_{E}(H|E)$. Thus, $BJconf_{E'}(H|E) \leq BJconf_{E}(H|E)$. □

Cite this article: Feldbacher-Escamilla, C. J. and Gebharter, A. 2020. Confirmation Based on Analogical Inference: Bayes Meets Jeffrey. *Canadian Journal of Philosophy* 50: 174–194, doi:10.1017/can.2019.18