Bayesian Confirmation from Analog Models

Cameron Beebe[*] & Roland Poellinger[†]

Abstract

Analog models can be used to investigate aspects of a target system we might not have easy empirical access to. Evidence from an analog model has, under certain strict conditions, been used to argue for the *confirmation* of a target theory (Unruh (2008), Dardashti *et al.* (2017)). We investigate what a Bayesian account of such confirmation might require, and illustrate the details by discussing a water-wave analog system of the quantum Casimir effect. We argue that the analogical reasoning involved in this case cannot be sufficiently expressed by traditional Bayesian networks, and therefore employ an extension of (causal) Bayes nets to more capably handle the case study. Our formalization of the concept of analogy provides a novel reconstruction of Bayesian confirmation from analog models, which crucially preserves the essential symmetry involved in analogies. Finally, we take our formal analysis of the Casimir effect case to shed new light onto theoretical pre-unification via analogical reasoning.

Contents

[*]Research Center for Neurophilosophy and Ethics of Neurosciences, GSN-LMU Munich
[†]Munich Center for Mathematical Philosophy (MCMP), LMU Munich

1 Introduction

A recent argument by Dardashti *et al.* (2017) introduces the notion of analog *simulation* under certain syntactic isomorphisms to justify the claim that analogous phenomena in fluid 'dumb hole' simulations can confirm the hypothesis of black hole Hawking radiation. The authors distinguish the notion of analog reasoning in general from analog simulation and assert that a good theory of confirmation will be able to incorporate confirmation from analog simulation:

> It should be noted that here we are using confirmation in the most general and intuitive sense of the term, and not making a specific claim regarding the possibility of characterizing certain cases of analogue simulation in terms of a particular philosophical model of confirmation. Rather, we would insist that certain cases of analogue simulation must be counted amongst the explananda for which the models of confirmation are intended to provide the explanans. Thus, from our perspective, if a philosophical model of confirmation proves not to be able to accommodate analogue simulation, then so much worse for the model. (Dardashti *et al.* 2017, p. 14)

The aim of this article is to discuss whether (and how) *Bayesian Confirmation Theory* may be able to meet this challenge. One of the problems we will see for a Bayesian account is that analogical relations are genuinely symmetric. If one wishes to incorporate confirmation from an analog model, it will be crucial to utilize a framework that handles this symmetry. A Bayesian framework should also allow for the intuition that, although there is symmetry in analog relations, scientists may *confirm* asymmetrically. That is, at any moment, confirmation might only flow in one direction. Confirmation through a bi-directional relation may be one-directional in practice.

In the following, we will first (Sec. 4) provide a background on analogical reasoning (Sec. 2) and on Bayesian confirmation theory (Sec. 3). We then look in more detail at the concept of analog simulation as developed in Dardashti *et al.* (2017). We apply the idea of analog simulation to a case study of an analog Casimir effect system, and check whether existing frameworks can model the confirmatory probabilities from analogy seen in this case (Sec. 5). We argue that standard Bayesian networks cannot, and we suggest criteria we think a formal representation of analogy should cover in order to be used for Bayesian confirmation. These criteria are then formalized in an extension of the standard Bayes net framework by introducing a symmetric relation to represent relevant sub-systems isomorphism (Sec. 6.1). In Sec. 6.2, we reassess the case study in light of this extension and supplement this discussion with technical details on the the relevant sub-systems isomorphism (Sec. 6.3). We comment on the relationship between analogical reasoning in science and theory pre-unification in Sec. 7 and conclude with a summary in Sec. 8.

2 Analogical reasoning

The word analogy comes from the greek ἀναλογία, meaning proportion. Following the treatment found in Hesse (1966), an analogical argument can be understood as very similar to solving proportion problems in mathematics. We are given the relationships between two terms, and one more term from a second relationship. From these three terms we can determine the fourth term. Using fractional notation such problems are of the form

$$\frac{A}{B} :: \frac{C}{x} \tag{2.1}$$

For simple mathematical ratios, we could have for example

$$\frac{2}{3} :: \frac{x}{9} \tag{2.2}$$

where we read out "two is to three, as x is to nine". Solving for $x = 6$ involves interpreting the :: as an equality, and we simply multiply the diagonals getting $3x = 18$. However, without some further clarification three possible interpretations might seem equally likely, that $x = 4, 6$ or 8. The latter option interprets the relationship between 3 and 9 as the addition of 6, and applies the same to the numerator. The first option interprets the relationship as squaring 3, and thus squaring 2 is 4. By convention, and use in analogical reasoning, we say that in this case $x = 6$ since it is the *proportion* that must remain the same—that is, it is not the relationship between 3 and 9 but the structural relationship between 2 and 3 that should be preserved on the right hand side for an analogical relation.

This relational structure mapping is the key to analog relations according to Gentner (1983). When we use the form of (2.1) to talk about scientific reasoning from an analog model, we might read out "A is related to B, which is similar to C related to D". That is, there is some relation $Rel_1(A, B)$ that is similar to a relation $Rel_2(C, D)$. This relational structure is what is mapped between representations in an analogy, on Gentner's account, and is denoted here by the double colon notation '::'.

3 Analogy and confirmation

The notion of theory confirmation in science is represented in the Bayesian framework by confirmatory probabilities—that is, a theory obtains higher probability given confirming evidence compared to the theory without this evidence. Hypothesis and evidence are considered to be random variables, visualized as nodes in a network representation of these variables and their epistemic connections, so called DAGs—Directed Acyclic Graphs. For nodes H and E, we say that E confirms H when $P(H \mid E) > P(H)$. If H is a hypothesis and E some evidence for the hypothesis, we would expect our subjective degree of belief[1] in H to be positively influenced by an observation of relevant evidence. In a causal Bayesian network representation of the situation, we intuitively expect E to be a descendant of H—that is, there is some causal or other directed connection between H and E. This aspect of a Bayesian Network will be slightly revised later in order to accommodate our view of analog confirmation, since *analog* evidence E' is not, under minimal assumptions, a descendent of H nor is it causally connected.

In a recent book, Paul Bartha makes explicit statements concerning analogical arguments and Bayesian epistemology (Bartha 2010, p. 31):

> For Bayesians, it may seem quite clear that an analogical argument cannot provide confirmation. In the first place, it is not obvious how to represent an analogical argument as an evidential proposition E. Second, even if we can find a proposition E that expresses the information about source and target domains used in the argument, that information is not new. It is "old evidence," and therefore part of the background K. This implies that $E \wedge K$ is equivalent to K, and hence that
>
> $$Pr(H \mid E \wedge K) = Pr(H \mid K) \tag{3.1}$$
>
> According to the definition, we don't have confirmation. Instead, we have an instance of the familiar "problem of old evidence" (Glymour 1980). Third, and perhaps most important, analogical arguments are often applied to novel hypotheses H for which the "prior" probability $Pr(H \mid K)$ is not even defined. Again, the definition is inapplicable.

Bartha goes on to suggest that the role of analogy in Bayesian epistemology is to raise prior probabilities of a considered hypothesis. While we agree that analogical considerations may impact prior probabilities, we think that Bartha has only discussed one way in which analogical reasoning may be used by

[1]The interpretation of probability being used is that they are subjective degrees of belief.

scientists—and thus only one way in which analogies may shape Bayesian models. There is an epistemic difference between a situation in which a specific analogical relation is *granted* by a scientist, and a situation in which the discovery or establishment of an analogy is taken as evidence. The latter is arguably what Bartha is referring to, whereas we will mainly be concerned with the former.

We argue that Bartha's idea is captured in an epistemic network in which the analogy is modeled as a node—a random variable representing the possible existence or non-existence of an analogical relation. Confirmation from an analog model, on the other hand, *grants* the existence of an analogical relation. We think this is better modeled as an informational link (a *contour*) between theoretical domains. We will see how analogy *does* play a role in evidential statements which do not succumb to the old evidence problem. Therefore, traditional Bayesian confirmation is also accounted for in our approach, showing that analogical considerations can also impact posterior probabilities by taking into account evidence from a model.

4 Analog simulation

In a recent paper on analogical inference in physics, Dardashti *et al.* (2017) discuss analogies between experimentally accessible test setups and potentially less accessible target system we want to gain insights about. The authors introduce the formal concept of *analog simulation* for this purpose which shall be introduced briefly in the following before it is applied for our case study.

Analog simulation bridges two basic frames: The source system is prepared, manipulated, and observed to make inferences about the target system. Let us introduce some terminology first to relate all concepts in a formal way:[2]

1. The target system T (a class of situations of interest) is to be modeled as $\mathcal{M}_T$ in a suitably chosen modeling framework $\mathcal{L}_T$;

2. $\mathcal{M}_T$ is constrained by certain background assumptions $\mathcal{A}_T$, summarizing theoretical and empirical knowledge as well as the domain of conditions $\mathcal{D}_T$ to which the model is intended to apply;

3. $\mathcal{M}_T$ can be used to predict phenomena E_T and will in turn be validated by evidence in accordance with E_T;

4. The accessible source system S is to be modeled as $\mathcal{M}_S$ in a suitably chosen modeling framework $\mathcal{L}_S$;

5. $\mathcal{M}_S$ is constrained by background assumptions $\mathcal{A}_S$, containing the domain of conditions $\mathcal{D}_S$ to which the model is intended to apply;

[2]In our interpretation, we deviate from Dardashti *et al.* (2017) in notational details.

6. Just as on the T side, $\mathcal{M}_S$ can be used to predict phenomena E_S and will in turn be validated by evidence in accordance with E_S.

The source system S will now allow *analog simulation* of target T's behavior if (i) there exist exploitable structural similarities between $\mathcal{M}_S$ and $\mathcal{M}_T$ sufficient to define a syntactic isomorphism robust within the domains $\mathcal{D}_S$ and $\mathcal{D}_T$, respectively, and if (ii) this isomorphism is prompted by and based on a set of *model-external empirically grounded arguments*, abbreviated as MEEGA.

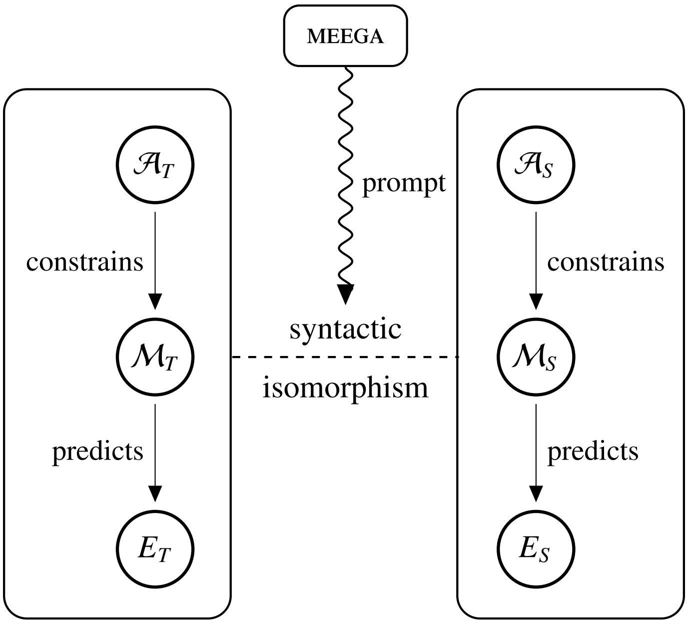

Figure 1: The analog simulation scheme: Framework $\mathcal{L}_T$ (left box) is used to model target system T in model $\mathcal{M}_T$; source system S is accordingly treated in framework $\mathcal{L}_S$ (right box).

Figure 1 relates these elements in a conceptual graph: The rounded box on the left side contains all elements of the target frame, while the right box contains all elements of the source system. MEEGA prompt the establishment of a bridge between theoretical networks in the form of a syntactic isomorphism as translation between the systems' components.

For Dardashti *et al.* (2017), the terminology is illustrated with an example from physics, where observations of phenomena E_S in table-top fluid systems boost confidence in theoretical assumptions $\mathcal{A}_T$ about gravitational phenomena described in framework $\mathcal{L}_T$. The syntactic isomorphism (motivated by additional knowledge about the underlying physics of both frames) allows for the transfer of knowledge about acoustic Hawking radiation in the fluid system to Hawking radation in black holes.

5 Water wave analog of the Casimir effect

The example of an analog model we will consider here is the table-top fluid model of the Casimir Effect investigated by Denardo *et al.* (2009), a physical or material analogy: The Casimir effect is produced between two very small (and very thin) uncharged parallel metal plates that are placed close together in a vacuum. At certain distances d the plates are pushed together, at others they are pushed apart. In quantum theory there is a non-zero energy associated with the ground state of each mode in the quantum vacuum $hf/2$ (where f is the frequency of the harmonic oscillator associated with the mode and h is Planck's constant). We can account for such behavior in quantum electrodynamics by calculating the relative difference in pressure between the force of electromagnetic radiation outside the plates and inside the plates, since the closeness of the plates excludes certain wavelengths in the background spectrum. The spectrum of zero point frequencies is infinite both on the outside and inside of the plates, but after renormalization Denardo *et al.* (2009, p. 1095) note that the result of the calculation gives a force of $\pi^2\hbar c/240d^4$ per unit area.

The water wave analog model that the authors construct consists of a table-top bath which is vertically vibrated according to a range of frequencies (10-20 Hz) which excites surface waves. Two acrylic or PVC plates are hung in parallel above the bath and dipped into the vibrating fluid bath (VFB). The surface waves and VFB are analogous to the zero point fluctuations, providing an explanation in terms of difference in pressure due to the exclusion of certain waves between the plates.[3]

We can model the relationship between these two systems in the following way according to the similarity relation discussed above. Say the phenomena of two parallel plates being pushed together when dipped in the VFB is E_F. To differentiate, let us say that E_Q is the *similar* effect on Casimir plates. The VFB for the analog model is B_F, and it has a causal relationship to E_F in that it causes the relative pressure due to wave motion to be greater on the surface area outside the plates than on the surface of the interior. Thus far we can say—in the form of a fractional notation:

$$\frac{E_F}{B_F} :: \frac{E_Q}{X_Q}. \tag{5.1}$$

Our choice for X_Q, if we had no other knowledge of the target system, would arguably be an X_Q that fulfills similar conditions as B_F (that is, X_Q should be some causal explanation like B_F, i.e., $X_Q = B_F{}'$). It seems that our expectation for an $X_Q = B_F{}'$ is *greater* than other options—given that we have established

[3]Also mentioned in both Denardo *et al.* (2009) and Barrow (2002, §7) is a larger instance of macroscopic 'Casimir' effects where parallel ships rolling on a swell will result in the destructive interference of waves between the ships, thus allowing the relative pressure difference from the waves on the outer surfaces to push the ships closer together.

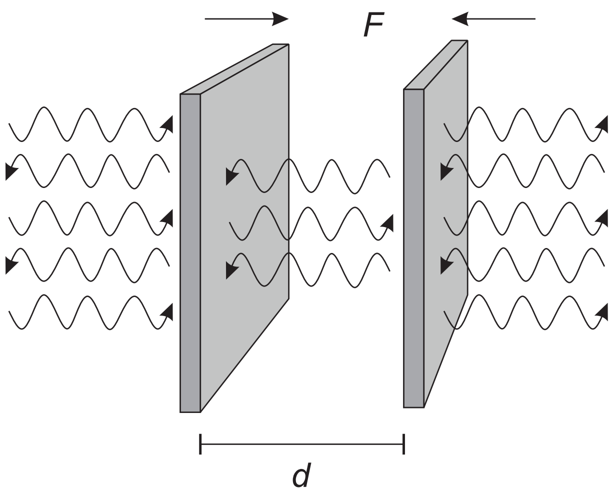

Figure 2: The abstract scheme underlying both the quantum Casimir and the fluid Casimir effect; F indicates the attractive force, d the distance between the two plates.

some justification for applying the analog model in the first place (i.e., we have $E_F :: E_Q$).

Importantly, this set up is different from that considered in Dardashti *et al.* (2017) since it is not the existence of a particular phenomenon that is the unknown variable in the analogical argument. In their discussion, the phenomenon of Hawking radiation is what is inaccessible—black holes are presumed to exist. Here, what is inaccessible is not the phenomenon (we already have observed Casimir plates coming closer together). Rather, what is inaccessible is the background medium or field which is supposed to be our explanation of how the observed phenomenon is produced—*but which cannot be directly observed*.[4] What is confirmed in their example is the existence of Hawking radiation. What is potentially confirmed here by the argument $X_Q = B_F{}'$ is any quantum theory of the vacuum which gives an ontology of non-zero energy density using a mechanism or term that is similar to the vibrating fluid bath (i.e., a $B_F{}'$). We will now refer to the quantum mechanism as B_Q.

However, we know that the analogical relationship at least breaks down with respect to the *dimensions* of the analog model and the target system. The authors consciously note other deficiencies in the analogy (Denardo *et al.* 2009, p. 1095):

> The analogy of our water wave system to the Casimir effect is not exact. Because the water waves are driven, the energy density of the spectrum is not infinite, so a regularization procedure is not needed.

[4]"Although [zero point energy] cannot be directly observed, the presence of the plates discretizes the spectrum between and transverse to the plates, which causes the imbalance of the radiation force."(Denardo *et al.* 2009, p. 1095)

> Furthermore, we are primarily concerned with the case of closely spaced plates, which yields a force that is independent of the separation distance d. This behavior is in contrast to the Casimir force, which has a $1/d^4$ dependence due to the divergence of the ω^3 spectrum at high frequencies.

Furthermore, there are terms in our formal representation of the fluid such as viscosity and surface tension which have unclear analogical relationships with the quantum world. However, in our view these are not particularly troublesome. The analogy concerns the relative difference in pressure between the exterior and interior of two parallel surfaces in an oscillating medium composed of a range of frequencies.

Before introducing a Bayesian network in the next section which can appropriately model confirmation from B_F, the analog model, we should first like to know the kinds of probabilities that should hold in such a network in order to ensure confirmation. As mentioned earlier, we can consider quantum electrodynamics T_Q to be the theory of electromagnetic radiation in a quantum vacuum—the relevant theory for explaining the Casimir effect. Unfortunately, measuring the zero point fluctuations directly is not possible, and thus—obviously—the existence of such an ontology is independent from the theory. In other words, since we have not observed B_Q, then $P(T_Q \,|\, B_Q) = P(T_Q)$. This is a problem since B_Q is supposed to give us the causal explanation of E_Q, and surely we should have that $P(T_Q \,|\, E_Q) > P(T_Q)$—i.e., that observing the Casimir effect confirms a quantum theory of the vacuum.

Considering the analog model, however, it seems that as a subjective scientist the explanation offered by the fluid system confirms (increases the probability of) an analogous explanation in the target system. It seems that $P(B_Q \,|\, B_F) > P(B_Q)$. It is also evident that $P(B_F \,|\, E_F) > P(B_F)$—i.e., observing evidence predicted by a model of the bath system should confirm the model. In the end we will see also that $P(T_Q \,|\, E_F) > P(T_Q)$, that the analog phenomenon confirms the target theory. This stems from the final positively correlated module $P(T_Q \,|\, B_Q) > P(T_Q)$.

So the Bayesian network should allow the following assumptions to hold:

$$P(T_Q \,|\, B_Q) > P(T_Q) \tag{5.2}$$

$$P(B_Q \,|\, B_F) > P(B_Q) \tag{5.3}$$

$$P(B_F \,|\, E_F) > P(B_F) \tag{5.4}$$

In the example of the fluid analog model of the Casimir Effect, the theories of fluid mechanics T_F and quantum electrodynamics T_Q have evidence domains of macroscopic fluids E_F and electromagnetic radiation in a vacuum E_Q respectively. The difference in orders of magnitude justifies the assumption that the overlap between evidence of these theories is, under normal conditions, nonexistent. This is in contrast with the overlapping evidence in the networks discussed in Dizadji-Bahmani *et al.* (2011). However, it is seen that *after* the analogical argument is admitted into the network, both B_F and E_F are common descendants of T_F and T_Q.

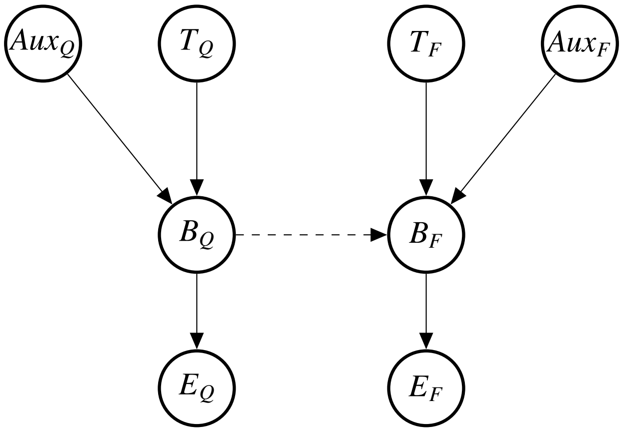

Figure 3: A Bayesian network representing the relationship between the analog Casimir effect E_F and the Casimir effect E_Q. The dashed edge marks analogy as asymmetric link in order to capture the subjective sense in which $P(T_Q \,|\, E_F) > P(T_Q)$, by making E_F and B_F descendants. Also $P(T_Q \,|\, B_F) > P(T_Q)$.

An analogy is made with the relative pressures of waves on parallel partitions in the (practically) random 'bath'—either surface waves in a macroscopic fluid or electromagnetic radiation in a quantum vacuum. These waves are assumed to be linear, since non-linear mechanics begin to appear upon sufficiently high amplitude vertical oscillations of the fluid bath when droplets are ejected. See Denardo *et al.* (2009) and Terwagne and Bush (2011), for example. Thus, the limitation of the system to approximately sinusoidal waves is a relevant aspect of the analogy that we can call an *auxiliary* condition. These and other auxiliary conditions from the experimental set up that produces both the Casimir effect and the analog effect, plus the respective relevant theories, gives us descriptions of the fluid and quantum systems B_F and B_Q respectively, which are related through the analogical argument given earlier. However, it still isn't exactly straightforward how to represent the analogical relationship between B_F and B_Q in the network. A possible choice for directing an edge between these nodes in the Bayesian network would be from B_Q to B_F: In this way, B_F and E_F are descendants of T_Q, but the arrow also introduces asymmetry in the graph.

6 Bridging models

The above network fails to capture the symmetry required of an analog relation. It represents, if you will, only *half* of a superposition of *two* Bayesian networks, each yielding answers to queries for different inferential directions—if B_F and B_Q behave analogously, evidence for either side should inform the respective other side in a symmetric way. In this section we provide an alternative account of Bayesian confirmation, utilizing undirected edges to bridge theoretical frameworks and preserve the symmetry of analog relations.

6.1 Directed and undirected relations

Our choice to direct the edge from B_Q to B_F in Fig. 3 is rooted in the epistemic status and goals of a scientist (or philosopher) in a particular situation. Simply put, we want to confirm T_Q, not T_F. If we wanted to confirm T_F we would switch the arrow the other way around. [5] So, E_F and B_F must be descendants. If the arrow pointed in the other direction and they weren't descendants, the collider structure at B_Q would d-separate T_Q from B_F and E_F, giving us $T_Q \perp\!\!\!\perp B_F, E_F$. In other words, if the collider were present then $P(T_Q \,|\, B_F) = P(T_Q)$, where instead we want that $P(T_Q \,|\, B_F) > P(T_Q)$.[6]

For a specific inquiry, confirmation from an analog model can be represented in a DAG. Yet, we may wish to have confirmation flow the other way (e.g., to T_F) and our formal system should be prepared with the extra information on hand to do this. If we were to use *only* the above DAG in our representation of the problem, then we would lose the information relevant to the symmetries of an analogy and the potential to confirm in the opposite direction (i.e., $P(T_F \,|\, E_Q) > P(T_F)$). Since we are not looking for a case-by-case account of analog confirmation, we want to preserve this information in a more general framework that ties in with Bayesian confirmation theory. As shown, a standard Bayesian network is insufficient to adequately handle representing confirmation from analog relations as we have construed them. The question remains: How can intertheoretical, symmetric relations be integrated in a formal model from which genuine (Bayesian) confirmation claims can be derived?

We suggest a new type of edge—a non-directed, non-causal, informational link, capable of propagating information instantaneously. Furthermore, it is not to be deactivated by (causal) interventions in the model. It should work like

[5] However, the unobservability of B_Q may present some issues. For example, it seems like it should be the case that $P(T_F \,|\, B_Q) = P(T_F)$. If we can't observe B_Q directly (this is the reason the analogical argument was made in the first place) then we can't condition on it like it is observed evidence.

[6] This argument similarly goes for an edge between E_Q and E_F, were we to choose to express the analogical relationship between the two frames at the evidential level. Also, our previous discussion used analogical reasoning to map the structure between B_F and E_F to suggest B_Q. If the fluid bath was not granted as analogical, a mere similarity of evidence would arguably not justify the strong intuition that we might want to confirm T_Q. One might be getting similar evidence from *dissimilar* systems—e.g., mere numbers from point measurements, similar but unstructured. What is important is precisely the structural mapping between the descriptions of systems.

synonyms, mathematical inter-definitions, or logical relations (which certainly all belong to the pool of knowledge we use for decision making). In standard statistical modeling, extensionally equivalent variables would certainly be collapsed into one single variable (node, respectively). For our purposes, though, we would like to disambiguate in the model the intensional distinction between connected variables. Consequently, the final model ought to contain two separate nodes and mark these nodes as tightly, functionally dependent.

In briefly discussing the possibility of embedding such non-causal links into causal Bayes net structures, Verma and Pearl acknowledge the usefulness of such hybrid models:

> The ability to represent functional dependencies would be a powerful extension from the point of view of the designer. These dependencies may easily be represented by the introduction of deterministic nodes which would correspond to the deterministic variables. Graphs which contain deterministic nodes represent more information than d-separation is able to extract; but a simple extension of d-separation, called D-separation, is both sound and complete with respect to the input list under both probabilistic inference and graphoid inference. (Verma and Pearl 1988, p. 75)

We propose to utilize this idea and build analogy on a relation between strictly correlated variables. We understand analogy as a non-causal and non-directional relation and construct it on top of a syntactic isomorphism (formalized as a 1-1 function) in extensions of Bayes net causal models. Such hybrid structures have been discussed in philosophy as *causal knowledge patterns* (*CKP*) in Poellinger (2012), as well as statistics (e.g., as chain graphs in Lauritzen and Richardson (2001)). Poellinger's *CKP* extend a standard causal model $M = \langle U, V, F\rangle$, where U is a set of exogenous variables, V a set of endogenous variables, and F a set of functional causal mechanisms—cf. e.g. (Pearl 2000, def. 7.1.1, p. 203). Such extensions can be defined as quadruples $\mathcal{K} = \langle U, V, F, C\rangle$, where C is a set of *epistemic contours*, i.e., a set of 1-1 functions $i_{j,k}$ that take the value of some variable V_j and assign the value $i_{j,k}(V_j)$ to some other variable V_k in the pattern. Importantly, intervening on one of these entangled variables will not break the contour (and will thus not determine any directionality).

Contours possess exactly the properties we want for our analog relations. Yet, embedding entangled variables of this kind in Bayesian networks precisely renders them non-Markovian.[7] In the general case, the inferential framework must be tweaked to retain soundness,[8] but in our special case with a single intertheoretical bridge, we extract from the larger project only the idea of utilizing an

[7]When Pearl claims that "[t]he Markovian assumption [...] is a matter of convention, to distinguish complete from incomplete models"(Cf. (Pearl 2000, p. 61) he naturally has Bayes net causal models (with distinct variables) in mind, which we just dismissed.

[8]For consistent reasoning and efficient computation of causal knowledge patterns to remain

undirected functional link to join two probabilistic chains (i.e., our two frames). So, how can we spell out analogical inference across this newly introduced bridge?

6.2 Analogical inference across symmetric links

In our proposal, the model-external postulate (or assumption, or also perception) "B_Q is similar to B_F (in certain known respects)" prompts the inclusion of a translation relation rather than the insertion of a new node. Analogical reasoning begins with a consciously initiated domain comparison which we characterize as the insertion of an inter-theoretical bridge.[9] Figure 4 is a rendition of the Casimir effect example discussed above with the contour i marking the analogical relationship between the frames at the level of models B_Q and B_F.

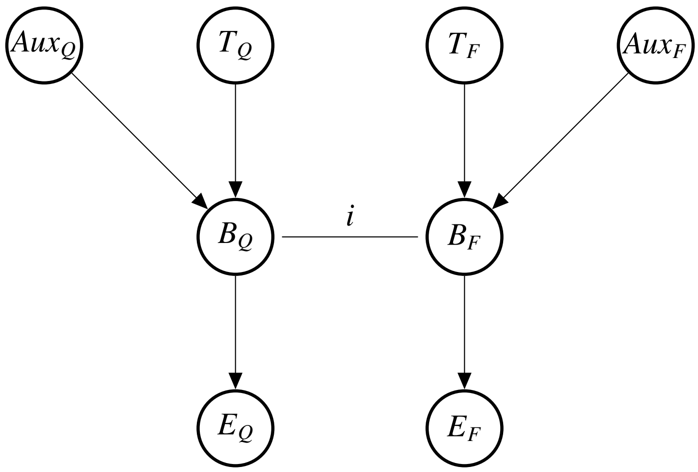

Figure 4: Model-level analogy as epistemic contour i with the intertheoretical bridge i between B_Q and B_F.

In this graph, the undirected edge between B_Q and B_F, along with the formal explication we have introduced, provides a means for implementing analog confirmation as we have construed it. A scientist or an artificial system can obtain $P(T_Q \,|\, E_F) > P(T_Q)$ while retaining the information of the more general undirected edge as well as the ability at some later time to provide confirmation for T_F.[10]

possible at all, acyclicity, independence (as expressed in the graphical d-separation criterion), and the *identifiability of causal effects* receive new explications. Poellinger (2012) introduces a further graphical criterion, the *principle of explanatory dominance*, to define under which conditions the Markov requirement can be reclaimed and *CKPs* utilized for causal inference.

[9] This insertion can formally be understood as a structural transformation by which two frames are joined.

[10] While in this case we might not need or want to confirm T_F, a general account *should* provide for this.

We might also wish to represent the analogical contour between T_Q and T_F (Figure 5). However, this would be a much stronger claim since B_Q and B_F are determined to an additional extent by auxiliaries. An analogy at the theory level is, in some sense, an analogy that could be a step further towards unification than one at the model level. We will return to (pre-)unification in Sec. 7.

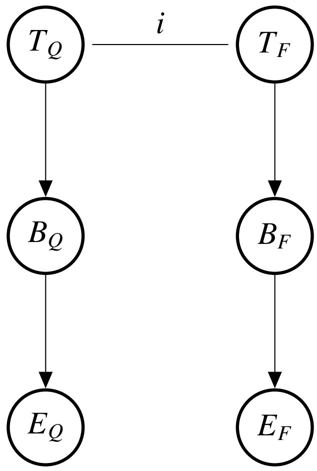

Figure 5: Theory-level analogy as epistemic contour i with the intertheoretical bridge i between T_Q and T_F.

A potential option would also be to insert a collider structure between the model levels representing an analogy. However, as we have argued, granted analogies should not be represented as a node. It is unclear what the content of the node would be, and the values it could take would arguably depend upon a meta-level analysis of the network (i.e., it would be a self-referential node). We think our approach of modeling the analogy as a functional relation is more consistent with the case study, as well as mathematically useful for future applications of the method.

6.3 Translation via relevant sub-isomorphisms

We take analog contours to be an expression of a modeling relationship between frameworks. It can be thought of formally as a translation relation based on a *relevant sub-isomorphism*, which has been anticipated in the literature on models and representations in science, cf. Frigg and Hartmann (2012):[11]

> One version of the semantic view, one that builds on a mathematical notion of models (see Sec. 2), posits that a model and its target have

[11] We have included in our bibliography the complete references cited in this quote for completeness.

to be isomorphic (van Fraassen 1980; Suppes 2002) or partially isomorphic (Da Costa and French 2003) to each other. Formal requirements weaker than these have been discussed by Mundy (1986) and Swoyer (1991). Another version of the semantic view drops formal requirements in favor of similarity (Giere 1988 and 2004, Teller 2001). This approach enjoys the advantage over the isomorphism view that it is less restrictive and also can account for cases of inexact and simplifying models. However, as Giere points out, this account remains empty as long as no relevant respects and degrees of similarity are specified. The specification of such respects and degrees depends on the problem at hand and the larger scientific context and cannot be made on the basis of purely philosophical considerations (Teller 2001).

We follow this line of reasoning and formulate a relevance filter in order to capture the purpose-driven selection of theoretical entities to be translated. Of course, basing analogy on a purpose-driven relevance concept makes the concept of analog models context-specific. We embrace this fact and call B_Q and B_F *analog models relative to*

1. a relevance filter *Rlv*;
2. a bijection between the relevant properties of B_Q and B_F (an isomorphism between sub-structures of B_Q and B_F).

The filter function *Rlv*, an indicator function over the descriptive elements of both frameworks, selects for each semantic category (for individual objects and each set of n-ary relations between such objects) subsets of equal magnitude; i.e., for each category:

$$\| Rlv(B_Q) \| = \| Rlv(B_F) \| . \tag{6.1}$$

If B_Q and B_F behave alike with respect to relevant parts (i.e., parts selected by *Rlv*) that are described by $P_Q(x)_Q$ and $P_F(x)_F$ (with properties P of objects x in the respective models), then the following formula explicates the analog relation between frameworks via translation i:

$$\forall P_F, \mathbf{x}_F (P_F(\mathbf{x}_F) \leftrightarrow P_Q(\mathbf{x}_Q)) . \tag{6.2}$$

Note that this isomorphism might be the result of iteratively fine-tuning non-bijective translations between the frameworks.[12]

[12]We are thankful to Mark Colyvan for valuable discussions about the nature of this morphism.

Having defined the propagation of information across the epistemic contour in this way, tracing confirmatory support in Fig. 4 yields the following:

$$P(B_F \mid E_F) > P(B_F) \tag{6.3}$$
$$P(T_Q \mid B_Q) > P(T_Q) \tag{6.4}$$
$$P(i(B_F) \mid B_F) > P(i(B_F)) \tag{6.5}$$

where $i(B_F)$ represents specific information about the properties of B_Q relevant for the analogical inference (i.e., as chosen by the filter function).[13] Eq. 6.5 exploits the characterization of contour i as 1-1 function: Learning B_F tells us more about the *Rlv*-selected properties and objects at the core of B_Q, thereby raising our degree of belief in those B_Q that are compatible with $i(B_F)$. Now, by transitivity, 6.3, 6.4, and 6.5 together entail

$$P(T_Q \mid E_F) > P(T_Q), \tag{6.6}$$

which was implied by our list of desiderata above—Eq. 5.2, Eq. 5.3, Eq. 5.4, chained together. Formula 6.6 is an instance of Bayesian confirmation—this time across theoretical frameworks, though, and it encodes what we set out to achieve: Bayesian confirmation from an analog model.

7 Analogy and (pre-)unification

In her structure-mapping account of analogical reasoning, Dedre Gentner makes an important distinction between mere similarity, analogy, and abstract generalities or law-like statements. These represent different stages of learning about the relationship between two domains, moving from early similarity comparisons, to analogies, to generalizations:

> This sequence can be understood in terms of the kinds of differences in predicate overlap discussed in this paper. In the structure-mapping framework, we can suggest reasons that the accessibility and the explanatory usefulness of a match may be negatively related. Literal similarity matches are highly accessible, since they can be indexed by object descriptions, by relational structures, or by both. But they are not very useful in deriving causal principles precisely because there is too much overlap to know what is crucial. Potential analogies are less likely to be noticed, since they require accessing the data base via relational matches; object matches are of no use. However, once found, an analogy should be more useful in deriving the key principles, since the shared data structure

[13] As soon as one learns of a specific instantiation of $i(B_F)$, i.e., the relevant core of B_Q, those theoretical entities not in the *Rlv* mapping must be updated in line with consistency requirements.

> is sparse enough to permit analysis. [...] To state a general law requires another step beyond creating a temporary correspondence between unlike domains: The person must create a new relational structure whose objects are so lacking in specific attributes that the structure can be applied across widely different domains. (Gentner 1983, p. 167-168)

This contextualizes our approach and the way in which analogy can be seen as the pre-unification of two theoretical frames: The way we have modeled the link between these frames may be understood as a "temporary correspondence between two unlike domains". The Casimir example discussed is arguably in the middle of Gentner's spectrum between bare similarity and abstract generality. There are literal similarity matches—but some features in representations of the respective domains must also be thrown out as not similar. There is relational structure being mapped—but the objects still have enough specific attributes that it seems unwarranted at this stage to make any conclusive generalization about an entire class of domains.

That said, we can imagine that such a class could be built up in a case-by-case manner, and eventually justify a unified claim regarding the structure of domains in the class. Indeed, there are cases in science where strong or systematic analogies can be thought of as almost unificatory (see Bartha (2013)). We think there is strong motivation for interpreting contours at the level we have utilized them (i.e., between model representations) as pre-unificatory analogies. This might be contrasted with, for example, an approach that models the inter-theoretical relationship as a sort of common cause (i.e., a parent node of both structures at the uppermost theory level). We think that these two approaches can coexist, representing different stages of epistemic modeling. An *analog knowledge pattern* can precisely represent a scientist's nuanced view of an inter-theoretic relation *before* she might wish to consider that the theories under question should somehow be unified into one theory.

8 Conclusion

The extended subjective Bayesian network presented here is able to account for confirmation from analog models and analog simulation. Thus, slightly modified, Bayesian confirmation theory is able to meet the challenge offered in Dardashti *et al.* (2017). Importantly, our account preserves the informational symmetries involved in analogical reasoning, as demonstrated in an application to a case study from philosophy of physics, the Casimir effect. It should be reemphasized that, in this case, what is inaccessible about the target system is not the phenomenon—both the target and analog systems have shown the plates moving closer together—but rather the theoretical and ontological explanation of why the target system produces the phenomenon. Thus, we are confirming a theory of a phenomenon by offering an explanation of an analogous phenomenon. We should consider confirmation of the general class of quantum theories such that with regard to the quantum vacuum they contain a B_Q—which is analogous to the vibrating bath in the analog model.

In this sense, a particular interpretation of the Casimir effect seems to be implied by the conceptual aspect of the explanation: the argument $X_Q = B_F{}'$ supports a field-theoretic explanation in terms of a spectrum of modes rather than that the effect is due to Van der Waals force. This is perhaps a problem for explaining the Casimir effect in terms of the Van der Waals force.[14] Furthermore, there are many other systems which exhibit similar phenomena (Denardo *et al.* 2009, p. 1100):

> Casimir-type effects occur, in general, for two bodies in a homogeneous and isotropic spectrum of any kind of random waves that carry momentum. A net attractive force occurs between two parallel plates in the typical case where the radiation force is reduced between them.

The various analog models of Casimir-type effects seem to provide some sort of unifying explanation, whereas an alternative explanation of the Casimir effect in terms of the van der Waals force is in contrast to such models.[15] However, the systems are not all described by a single unifying theory—and thus the weaker analogies between models might be taken to supply a form of *pre*-unificatory explanations in the cases where well-defined structural similarities might eventually lead to theoretical unification. Grounded in an analytic approach towards the concept of analogy, our constructive proposal provides a novel template for making such epistemological advances explicit.

Cameron Beebe · Roland Poellinger
Ludwig-Maximilians-Universität München
cameron.beebe@campus.lmu.de · r.poellinger@lmu.de

References

Barrow, J. D. [2002]: *The Book of Nothing: Vacuums, Voids, and the Latest Ideas about the Origins of the Universe*, Vintage.

Bartha, P. [2013]: 'Analogy and Analogical Reasoning', in E. N. Zalta (*ed.*), *The Stanford Encyclopedia of Philosophy*, fall 2013 edition.

[14] Also, an explanation in terms of 'virtual particles' flitting in and out of existence seems inferior, given these results, to an explanation in terms of crests and troughs in a fluctuating medium.

[15] Unless van der Waals and Casimir forces were shown to be equivalent, but as non-experts we remain skeptical yet open to further discussion on this point.

Bartha, P. F. A. [2010]: *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*, Oxford University Press.

Costa, N. D. and French, S. [2003]: *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*, Oxford University Press.

Dardashti, R., Thébault, K. and Winsberg, E. [2017]: 'Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity', *The British Journal for the Philosophy of Science*, **68**(1), pp. 55–89.

Denardo, B. C., Puda, J. J. and Larraza, A. [2009]: 'A water wave analog of the Casimir effect', *American Journal of Physics*, **Vol. 77**(No. 12), pp. p. 1095–1101.

Dizadji-Bahmani, F., Frigg, R. and Hartmann, S. [2011]: 'Confirmation and reduction: a Bayesian account', *Synthese*, **179**, pp. p. 321 – 338.

Frigg, R. and Hartmann, S. [2012]: 'Models in Science', *Stanford Encyclopedia of Philosophy*.

Gentner, D. [1983]: 'Structure-Mapping: A Theoretical Framework for Analogy', *Cognitive Science*, **7**, pp. 155–170.

Giere, R. [1988]: *Explaining Science: A Cognitive Approach*, University of Chicago Press.

Giere, R. [2004]: 'How Models Are Used to Represent Reality', *Philosophy of Science*, **71**(5), pp. 742–752.

Glymour, C. [1980]: *Theory and Evidence*, Princeton University Press.

Hesse, M. B. [1966]: *Models and Analogies in Science*, University of Notre Dame Press.

Lauritzen, S. and Richardson, T. S. [2001]: 'Chain Graph Models and their Causal Interpretations', *B*, **64**, pp. 321–361.

Mundy, B. [1986]: 'On the General Theory of Meaningful Representation', *Synthese*, **67**(3), pp. 391–437.

Pearl, J. [2000]: *Causality: Models, Reasoning, and Inference*,.

Poellinger, R. [2012]: *Concrete Causation: About the Structures of Causal Knowledge*, Ludwig-Maximilians-Universität München (doctoral thesis).

Suppes, P. [2002]: *Representation and Invariance of Scientific Structures*, CSLI Publications.

Swoyer, C. [1991]: 'Structural representation and surrogative reasoning', *Synthese*, **87**(3), pp. 449–508.

Teller, P. [2001]: 'Twilight of the Perfect Model Model', *Erkenntnis*, **55**(3), pp. 393–415.

Terwagne, D. and Bush, J. W. M. [2011]: 'Tibetan singing bowls', *Nonlinearity*, **Vol. 24**, pp. p. R51–R66.

Unruh, W. G. [2008]: 'Dumb holes: analogues for black holes', *Philosophical Transactions of The Royal Society A*, **366**, pp. p. 2905–2913.

van Fraassen, B. C. [1980]: *The Scientific Image*, Oxford University Press.

Verma, T. and Pearl, J. [1988]: 'Causal Networks: Semantics and Expressiveness', in *Proceedings of the 4th Annual Conference on Uncertainty in Artificial Intelligence (UAI-88)*, New York: Elsevier Science.