Humanitas Hodie

2024 | Vol. 7, n°. 1 | Recibido: 26 de abril de 2023 | Aprobado: 30 de agosto de 2023
DOI: 10.28970/ hh.2024.1.a3

When should we draw arrows? Assessing the use of Bayesian networks in philosophy of science

¿Cuándo deberíamos dibujar flechas? Evaluando el uso de redes Bayesianas en filosofía de la ciencia

Pablo Rivas-Robledo*

Abstract

In this paper I present the criteria under which it is admissible to draw arrows in Bayesian networks. I argue that arrows can be used to represent asymmetric, non-transitive, difference-making relationships in Bayesian networks. Moreover, since these graphs are not heterogeneous, there should be uniformity in what the arrows are representing. That is, a single Bayesian network cannot represent a variety of difference-making relationships. I use this asses two models found in the literature for transferring evidence in intertheoretic reduction. I conclude that these models are either unable to represent the fact that evidence from one theory is evidence for the other theory, or that they operate with assumptions that contradict the central tenets of intertheoretic reduction. I offer a possible way out of this predicament by proposing heterogeneous graphs that can be condensed into Directed Acyclic Graphs (DAGs).

Keywords: confirmation, intertheoretic reduction, bayesian models, bayesian epistemology, graph theory, heterogeneous graph.

Resumen

En este artículo presento los criterios bajo los cuales es admisible dibujar flechas en redes Bayesianas. Argumento que las flechas pueden ser utilizadas para representar relaciones asimétricas, no transitivas y que hacen la diferencia en las redes Bayesianas. Además, dado que estos gráficos no son heterogéneos, debería existir uniformidad en lo que las flechas están representando. Es decir, una única red Bayesiana no puede representar una variedad de relaciones que hacen la diferencia. Utilizo esto para evaluar dos modelos encontrados en la literatura para transferir evidencia en la reducción interteórica. Concluyo que estos modelos no pueden representar el hecho de que la evidencia de una teoría es evidencia para la otra teoría, o que operan con supuestos que contradicen los principios centrales de la reducción interteórica. Ofrezco una posible salida de

* Institute for Logic, Language and Computation. Contact: p.rivasrobledo@uva.nl. ORCID: 0000-0003-2716- 2715

este dilema proponiendo gráficos heterogéneos que pueden ser condensados en Grafos Acíclicos Dirigidos (GAD).

Palabras clave: confirmación, reducción interteórica, modelos bayesianos, epistemología bayesiana, teoría de grafos, grafo heterogéneo.

Introduction

Bayesianism is a theory of uncertainty (Hájek & Hartmann, 2010, p. 93). It can be used to model the reasoning of agents whose epistemic attitudes are taken as degrees of belief and follow the laws of probability to develop a theory of subjective Bayesian inference (Sprenger & Hartmann, 2019, p. 1). Although initially developed as a mathematical tool, Bayesian inference has found its way into philosophy in recent years, shedding light on various problems in different areas, more prominently in epistemology (see Bovens and Hartmann, 2004; Talbott, 2016). More recently, subjective Bayesian inference has been used to explain scientific concepts or to capture arguments that are part of scientific reasoning, giving rise to Bayesian philosophy of science (see Sprenger and Hartmann, 2019, p. xxv).

A central tool of subjective Bayesian inference is the use of Bayesian networks (BNs). These are a form of representation used to organise one's knowledge about a given situation into a coherent whole, facilitating an economical representation of joint probability distributions and efficient inferences from observations (Pearl, 1988, pp. 77, 116; Pearl, 2009; Wiegerinck *et al.*, 2013, p. 402, see also (Darwiche, 2009)). These have a number of important properties and limitations, which I will discuss in detail in section 2. For present purposes, I will note that BNs are a modelling tool that plays an important role in Bayesian philosophy of science. They are essentially directed acyclic graphs (DAG) consisting of a pair G = (V, E), where V is a finite set of nodes or vertices and E is a finite set of edges (also known as arcs or arrows) between nodes (Pearl, 1988, Chapter 3; Pearl, 2009, Sec. 1.2.2)[1].

Arrows are the main theme of this paper. Originally, arrows were used in BNs to show the correlation between probabilistic variables (Wright, 1921, pp. 559-560). A few decades later, they were also used to encode the existence of direct causal influences between any two connected nodes (Pearl, 1985, p. 331; Pearl, 1988, p. 77) and temporal relations (S. Lauritzen, 1979/1982; Wermuth & Lauritzen, 1982; Kiiveri *et al.*, 1984).

1 I will use the term arrow because it conveys a sense of direction that the other two terms do not.

The study of arrows per se and their role is a relatively unexplored but worthwhile topic, as many simply take it for granted. Here, I will argue that a careful analysis of what arrows really are provides a very important condition for the construction of Bayesian models in the philosophy of science, since they can describe only one kind of relation per graph. That is, one cannot draw an arrow between two nodes in the same graph that represents more than one kind of influence on the nodes.

Traditional mathematical and probabilistic Bayesian models today are usually data-driven, so this particular property of BNs is of little importance for these models. However, I think that the thesis defended here is of great importance for Bayesian philosophy of science, as I believe there have been models in recent years that have abused the modelling tool. An example of this is the models of intertheoretic reduction offered by Dizadji-Bahmani, Frigg, and Hartmann (2010; 2011, see also Sprenger and Hartmann, 2019, Variation 8) and the derivative model presented by Tešić almost a decade after (2019). I will call the former GNS and the latter GNS*.

Based on my analysis of arrows, I will argue that these models fails to describe intertheoretic reduction because they represent multiple relations with just one arrow, even relations that are symmetrical. I argue that these models are faced with the problem of either accepting that all arrows represent confirmation, and therefore they do not adequately describe intertheoretic reduction, or they must present a new model in which two BNs are used to describe intertheoretic reduction. To this end, in section 2, I will first give an overview of BNs and explain the representational constraints imposed by arrows when modelling probability distributions. In section 3, I will explain how Bayesian philosophy uses these tools to answer philosophical questions. In addition, I will present GNS and GNS* critique them in light of what was defended in section 2. Some conclusions will be presented at the end.

Bayesian networks: a primer

Today, BNs enjoy a widespread positive reputation, as they are often used in real-world applications, which include forecasting (Abramson, 1994; Gu *et al.*, 1994), automated vision (Levitt *et al.*, 1990; Rehg *et al.*, 1999), manufacturing control (Nadi *et al.*, 1991), integration of biological data (Beaumont & Rannala, 2004; Needham *et al.*, 2007), diagnostics (Andreassen *et al.*, 1987; Breese *et al.*, 1992), and healthcare (Kyrimi *et al.*, 2021). On top of that, they are pretty much ubiquitous in artificial intelligence. In this section, I will introduce some fundamental concepts relevant for the later discussion, as well as some criteria under which it is possible to draw arrows in BNs.

As mentioned earlier, BNs are a modelling tool for the economic specification of joint probability distributions, meaning that they are mainly used to simplify calculations. To see this, let us first imagine a probability distribution P defined on n discrete variables ordered arbitrarily as $X_1, X_2,...,X_3$. By the chain rule, we know that we can decompose P into a product of n conditional distributions:

$$P\,(x_1,..., x_n) = \prod_j P\,(x_j \mid x_1,..., x_{j-1}) \tag{1}$$

Yet it may be that the conditional probability of a variable X_j is not sensitive to all its predecessors. So, to calculate the value of X_j we choose the set of predecessors that do influence it's value (I will call such set *PAj*), and we calculate the value as.

$$P\,(x_j \mid x_1,..., x_{j-1}) = P\,(x_j \mid pa_j) \tag{2}$$

Which considerably simplifies the required input data. Thus, instead of specifying all possible values of all predecessors, we deal only with the values of PA_j, the *Markovian parents* or, simply put, *parents* of X_j (Pearl, 2009, p. 14). We define the parents of X_j (PA_j) as any subset $\{X_1,...,X_{j-1}\}$ that satisfies the equation (2) such that no other suitable subset of PA_j also satisfies it. In this case, X_j is dubbed the *child* node.

This is when Directed Acyclic Graphs (DAGS) come in handy, as we can graph the assignment of each variable X_j to a set PA_j. They are directed, which means that that each edge is oriented from one edge to another, and they are acyclic, which means following the direction of the edges will never create a closed loop.

The construction of a DAG, which models the distribution, can follow a recursive method (Pearl, 2009, p. 15): we start with the pair (X_1,X_2), we draw an arrow from the first to the second only if they are dependent according to equation (2); we continue with X_3 by drawing no arrows if it is independent of $\{X_1,X_2\}$, otherwise we draw an arrow from X_2 to X_3, if X_2 screens off X_3 from X_1 ($P(x_3 \mid x_2, x_1) = P(x_3 \mid x_2)$) and we draw an arrow X_1 to X_3 for the case that X_1 screens off X_3 from X_2; if no screening is found, then we draw arrows from X_1 and X_2 to X_3. And so on for every other variable in the distribution. In general, at each *j-stage* of drawing, take the minimal set of X_j's predecessors that screen off X_j from other predecessors (see equation (2)), this is the set PA_j, and draw an arrow from each member of PA_j to X_j.

For example, consider the typical distribution P defined by the binary variables X_1, X_2, X_3, X_4, X_5, where these variables represent (respectively) the season of the year, the amount of precipitation, whether the sprinkler is on, the wetness of the sidewalk, and whether the sidewalk is slippery. We can draw a DAG G to model this using the procedure just presented (Fig. (1)):

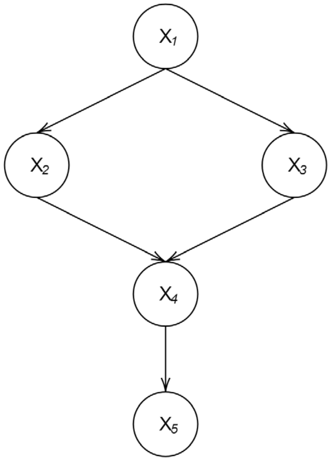

Figure 1: A DAG for a probability distribution P defined on binary variables X_1, X_2, X_3, X_4, X_5
Based on: (Pearl, 2009, p. 15).

G then induces a decomposition for the calculation of the joint probability distribution, namely $P(x_1, x_2, x_3, x_4, x_5) = P(x_1)P(x_2 \mid x_1) P(x_3 \mid x_1)P(x_4 \mid x_2, x_3)P(x_5 \mid x_4)$. If this decomposition is compatible with the distribution P, we say that G is Markov compatible with P.

A normal mathematical presentation of BNs would dive much deeper into the *d*-separation criterion, the parental Markov condition, minimality, and faithfulness. Space constraints don't allow me to explain these concepts in detail. However, I think that with the tools presented I can make my point about the meaning of the arrows, to which I now turn[2].

What do arrows mean?

To answer this question simply with "probabilistic dependence" is simply begging the question, as BNs are graphical models of pro-

2 A clear exposition of the above concepts can be found in various texts on the subject. The original treatment is, of course, contained in (Pearl, 1988, Chapter 3) and can also be found in (Pearl, 2009, Section 1.2-1.3). A more mathematically minded presentation is the one found in (Neapolitan, 2004, Part I). For a more philosophically informed exposition, see (Hitchcock, 2020, Section 4).

babilistic dependence. The real question is: what relationships can be considered probabilistic dependence? On the surface, one could argue that there is a relative consensus about what arrows represent in BNs. There is a sense in which an arrow between two nodes X_j and X_k indisputably means a direct influence from the variable in X_j to the variable in X_k. And it is precisely this formulation ("direct influence") that is found in much of the literature on the subject (S. L. Lauritzen and Spiegelhalter, 1988, p. 160; Spiegelhalter et al., 1993, p. 221; Bovens and Hartmann, 2004, p. 68; Wiegerinck *et al.*, 2013, p. 405), but it is also spoken of in more abstract terms, such as "class-property relationship" (Pearl, 1988, p. 77). And sometimes even the term "causal influence" is used casually (Pearl, 1988, pp. 77, 117; Sprenger & Hartmann, 2019, p. 33), a notoriously dicult term to pin down, but in this case, it simply means that there is a causal difference for the value of the child node, so that the way the parent influences the child is not mediated by any other variable (Hitchcock, 2020, Sec. 2.3).

The most basic way to understand this (especially for philosophical purposes), I think, is to understand an arrow as *difference-making*: an arrow between two nodes says that the variable in the parent node makes a difference in the value that the variable in the child node takes. The deliberately vague wording of the definition corresponds to the fact that in most cases the particular interpretation or type of difference-making will vary depending on the application. One might well use it to represent physical causality, but that may not be the case, since arrows work perfectly fine to represent other types of difference-making relationships (S. L. Lauritzen & Spiegelhalter, 1988, p. 160).

When should we draw arrows?

However, I think that not all difference making-relationships are t to be represented as an arrow in a BN. We basically already know that we should draw arrows between variables encoded by nodes that have certain relationships, but we need to know what kinds of relationships are representable. Some criteria can be established if we recall the original motivation for using BNs and some of the principles of graph theory.

Let us first address the original motivation for BNs. Pearl explains that one of the main reasons for using DAGs as graphs for BNs was that (i) they allowed him to represent real dependencies between variables and (ii) they encoded non-transitive relations (Pearl, 1988, p. 116; Pearl, 2009, p. 21).

The type of edges in DAGs are also important. They go only in one direction, which puts the D in DAG (Gutin, 2018, p. 163). This implies that they are not suitable for properly representing symmetric relationships: Their arrows run in only one direction (Bondy & Murty, 2008, pp. 31-32), using biderectional edges would immediately create a cycle, removing the A from the DAG. This makes them particularly good for representing asymmetric relationships.

Last but not least, we must remember that DAGs are not heterogeneous graphs, which means that the type of information encoded by their nodes and arrows is singular (Wang *et al.*, 2019, pp. 2022, 2024; Gan *et al.*, 2021; DGL, 2022). This means that the nodes in DAGs, and therefore in BNs, represent one type of object and the arrows represent one type of relation. There cannot be an arrow describing more than one type of difference-making relationship.

To summarise what has been said so far: BNs are a type of DAG, compatible with a probability distribution. When constructed from data, BNs pose almost no problems in constructing a DAG compatible with the distribution. However, as a modelling tool, the arrows in BNs can only represent one type of difference-making relationship, and it must be non-transitive and asymmetric.

From Bayesian networks to Bayesian philosophy of science

The remarks made in the previous section are crucial to my assessment of the use of BNs in Bayesian philosophy of science. In this section, I will set out how BNs are used in this branch of philosophy, with a particular focus on the difference-making relation of confirmation.

Bayesian philosophy of science seeks to explain scientific concepts and capture arguments that are part of scientific reasoning. As such, it rarely uses data to construct its BNs; rather, the construction can occur in two ways: Bayesians either (i) specify a probability distribution for a probability space and construct a BN that is compatible with the distribution (which in turn helps them provide a graphical representation of the conditional independencies between variables) (Bovens and Hartmann, 2004, see p.68 specially and Chapter 4 for an example); or (ii) they relate a set of terms using a DAG and derive a set of probability distributions compatible with the DAG such that said graph is considered a BN (Sprenger & Hartmann, 2019, Variations 2, 3, 8)[3].

So, in the first case, it looks like the BN is little more than a visual aid, since the distribution is already given. In the second case, the DAG is crucial for determining the distribution. This methodology (known as parametrization) is not completely unheard of (see Darwiche, 2009, Sec. 4.3; Wiegerinck *et al*., 2013, Sec. 2.2).

Probably the most important example in this context is the encoding of the relationship between a hypothesis (H) and a body of evidence (E). Bayesians want to represent the relationship between these two as E as depending probabilistically on H, since E might be more likely to be the case if H is the case than if it is not. The DAG is fairly straightforward in this case since we know the direction of influence, see Fig. (2).

Figure 2: A DAG to represent the Bayesian understanding of the relationship between a hypothesis (H) and a body of evidence (E).
Source: own elaboration based on (Dizadji-Bahmani et al., 2011; Tesic, 2019).

The parametrization of the said network is simple, as we only need to specify the value of H and the conditional probabilities of the node E given the node H, i.e. P(H), and P(E | H), and P(E | ¬ H), which lie in the interval (0, 1). According to the Bayesian framework, some evidence E confirms a hypothesis H if P(H | E) > P(H) and disconfirms it if P(H | E) < P(H).

In this case, there is good reason to believe that the difference-making relationship is direct causal influence. For example, let H be the proposition "S has COVID" and E be the proposition "The test for COVID is positive". In this case, P(H) is our prior degree of belief that S has COVID, that is how probable is for her to have the virus without us seeing the results, e.g., the proportion of people in her community who have COVID. P(E | H) would be the truepositive rate and P(E | ¬ H) the falsepositive rate[4].

3 Credit where credit is due: the ideas presented in these chapters of (Sprenger & Hartmann, 2019) first appeared in (Dawid *et al*., 2015), (Sprenger, 2016), (Dawid *et al*., 2015), and (Dizadji-Bahmani *et al*., 2010) and (Dizadji-Bahmani *et al*., 2011).

4 Example is borrowed from (Tesic, 2019, p. 1104).

So obviously having COVID has an influence on whether the test is positive or not. In this case, it seems straightforward to acknowledge that there is a causal flow from one proposition to another: having COVID symptoms leads to a positive test, provided that the test is good (which it is, see (Jarrom *et al.*, 2022)). Now it is not correct to say in general that the arrow between H and E represents (physical) causation, because sometimes the same BN is used in a more abstract way[5].

Two Bayesian models of intertheoretic reduction and their problems

Sometimes the relationship between hypothesis and evidence is more abstract, think on the relationship that exists between a scientific theory and a body of evidence. For example, the relationship that thermodynamics has to the Joule-Thomson process (Dizadji-Bahmani *et al.*, 2011, pp. 323-324): the expected amount of physical temperature change of a real gas when forced through a valve or porous plug (E) can be calculated using the principles of thermodynamics (H), so we can have a confirmation of thermodynamics ($P(\mathrm{H} \mid \mathrm{E}) > P(\mathrm{H})$), but we cannot say that the principles of thermodynamics physically causes the temperature change. What we can say is that the principles of thermodynamics (H) affects the measurement (the calculations themselves) of the temperature change, in the sense that a change in the principles leads to a change in the measurement.

In this case, of course, we don't have physical causation as in the COVID example, but we still can ascertain that the principles make a difference in measurement. Nevertheless, the relationship is clearly non-transitive (changing the principles doesn't change anything the evidence influences, since the evidence has no child node) and asymmetric (changing the way we measure doesn't necessarily change the principles of thermodynamics). We can then safely conclude that there is naturally a probabilistic dependence between these nodes precisely because $P(\mathrm{E}) \neq P(\mathrm{H} \mid \mathrm{E})$[6]. Therefore, we are safe if we draw the arrow between the nodes.

The Generalized Nagel-Schaffer model attempts to exploit this kind of relationship between hypothesis and evidence to prove how evidence for one theory can actually turn out to be evidence for another one. This model asserts that when the former theory is reduced to the later, then evidence for the first is evidence for the second (Dizadji-Bahmani *et al.*, 2010, p. 406). For example, some evidence

5 For starters, note that in this framework one can represent a logical consequence for the case $P(A \mid B) = 1$.

6 Alternatively, $P(H \cap E) \neq P(H)P(E)$.

in favour of statistical mechanics is evidence for thermodynamics because statistical mechanics is reducible to thermodynamics.

There are two Bayesian models for Generalized Nagel-Schaffner reduction, GNS and GNS*. The GNS model was developed in (Dizadji-Bahmani *et al.*, 2010; Dizadji-Bahmani *et al.*, 2011). A unified version of it can be found in (Sprenger & Hartmann, 2019, Variation 8). For this very reason, I will quote only from Sprenger and Hartmann rather than jump back and forth between the two original texts. The GNS* model is a modified version of GNS, proposed a few years later and found in (Tesic, 2019). Before explaining both and pointing out some problems, I will briefly show how intertheoretic reduction works according to Schaffner (1967, pp. 139140) and Sprenger & Hartmann (2019, pp. 209-210).

Consider a phenomenological theory T_{p} and a fundamental theory T_{f}, each of them identified with a finite set of empirical propositions, as follows: $\mathrm{T}_{\mathrm{p}} = \{P^{1}_{\mathrm{p}},..., P^{n}_{\mathrm{p}}\}$ and $\mathrm{T}^{\mathrm{f}}_{\mathrm{p}} = \{P^{1}_{\mathrm{f}},..., P^{m}_{\mathrm{f}}\}$, where *n* and *m* are not necessarily the same. The reduction consists in performing the following steps: (1) introduce boundary conditions and auxiliary assumptions to form an extended version of T_{f}, $\mathrm{T}^{*}_{\mathrm{f}}$; (2) connect different terms in both theories using bridging laws, then substitute terms in $\mathrm{T}^{*}_{\mathrm{f}}$ with the terms of T_{p} according to the laws to obtain $\mathrm{T}^{*}_{\mathrm{f}}$; (3) show that each element of $\mathrm{T}^{*}_{\mathrm{f}}$ is strongly analogous to T_{p}.

Two models

In synthesis, the general idea behind intertheoretic reduction is to show that the concepts and laws of a phenomenological theory can be reduced to some other concepts and laws of a fundamental theory in an overlapping domain (Sprenger & Hartmann, 2019, p. 207). If we establish reducibility between two theories, this means not only that there is consistency between them, but also that we can use evidence in favour of one theory as evidence for the other (Nagel, 1961, Sec. III.1). This last point is a well-accepted consequence of intertheoretic reduction (Wimsatt, 1974, p. 678; Van Riel, 2014, Sec. 8.7.2; Sarkar, 2015, p. 47), but the biggest challenge for Bayesians has been to be able to prove it using the modeling tools they value so much: probability distributions, BNs, and the relationship between hypothesis and evidence that presented above.

One way is to follow GNS and represent the intertheoretical reduction as follows: first draw a DAG with evidence for the phenomenological theory (E_{p}), evidence for both theories (E), and evidence for the fundamental theory (E_f). This gives an overview of the flow of confirmation before a reduction is reached (see Fig. (3)). Then draw a new one containing the phenomenological theory (T_p), the fundamental theory (T_f), the phenomenological theory with terms replaced (T^*_p), and the fundamental theory with assumptions (T^*_f). This gives us an insight into the flow of confirmation after reduction (see Fig. (4))[7].

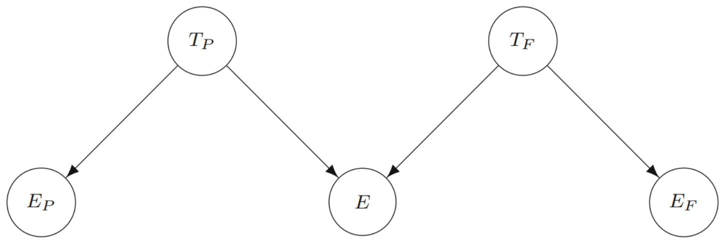

Figure 3: A DAG for GNS before the reduction. Source: (Tesic, 2019, p. 1105).

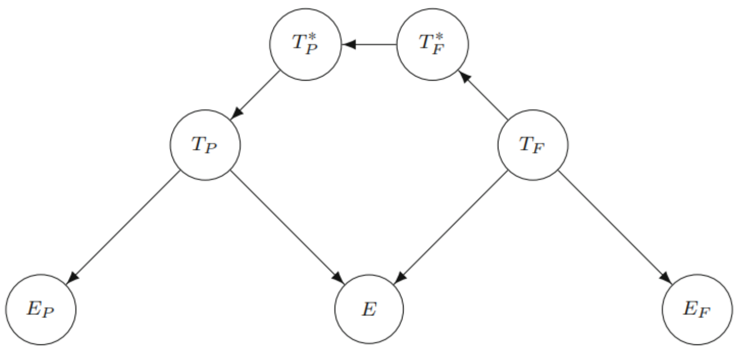

Figure 4: A DAG for GNS after the reduction. Source: (Tesic, 2019, p. 1106).

From there, GNS parameterizes a probability distribution compatible with said DAG, and we get our BN. For brevity, I give only the most important values, the entire distribution can be found in (Sprenger & Hartmann, 2019, pp. 213-215). So, the relevant probabilities are:

7 For clarity, the example refers to a theory with only one empirical proposition.

$$P(\mathrm{T_f}) = \mathrm{t_f}$$

$$P(\mathrm{E_f} \mid \mathrm{T_f}) = \mathrm{r_f},\ P(\mathrm{E_f} \mid \mathrm{T_f}) = \mathrm{q_f}$$

$$\mathrm{P(E_p \mid T_p) = r_p,\ P(E_f \mid T_p) = q_p}$$

$$\mathrm{P(T_p \mid T_p^*) = r_p^*,\ P(T_p \mid \neg T_p^*) = q_p^*}$$

$$\mathrm{P(T_f^* \mid T_f) = r_f^*,\ P(T_f^* \mid \neg T_f = q_f^*}$$

$$\mathrm{P(T_p \mid T_p^*) = r_p^*,\ P(T_p \mid \neg T_p^*) = q_p^*}$$

$$\mathrm{P(T_p^* \mid T_f^*) = 1,\ P(T_p^* \mid \neg T_f) = 0}$$

The beauty of this BN is that we can prove (i) that $\mathrm{E_f}$ confirms $\mathrm{T_p}$ if $(\mathrm{r_f} - \mathrm{q_f})(\mathrm{r_f^*} - \mathrm{q_f^*})(\mathrm{r_p^*} - \mathrm{q_p^*}) > 0$; and (ii) that $\mathrm{E_p}$ confirms $\mathrm{T_f}$ if $(\mathrm{r_p} - \mathrm{q_p})(\mathrm{r_f^*} - \mathrm{q_f^*})(\mathrm{r_p^*} - \mathrm{q_p^*}) > 0$ (see Theorems 8.1 and 8.2 of (Sprenger & Hartmann, 2019)). Hence the body of evidence that was exclusive to one theory becomes evidence for both.

Teŝić (2019) disagrees with this model because (i) the conditional probabilities between the bodies of evidence may not remain the same after reduction, (ii) it does not allow for partial reductions, and (iii) it allows for an equality that is not always the case, namely $P(\mathrm{T_p^*}) = P(\mathrm{T_f^*})$, since in GNS $P(\mathrm{T_p^*} \mid \mathrm{T_f^*}) = 1$. He then presents a different Bayesian treatment of the matter (GNS*), which gives a more prominent role to the bridge laws: whereas they were previously supposed to be implicit in the process of obtaining $\mathrm{T_p^*}$ from $\mathrm{T_f^*}$, they now appear as a root node with an edge dominating $\mathrm{T_p^*}$ (Fig. (5)):

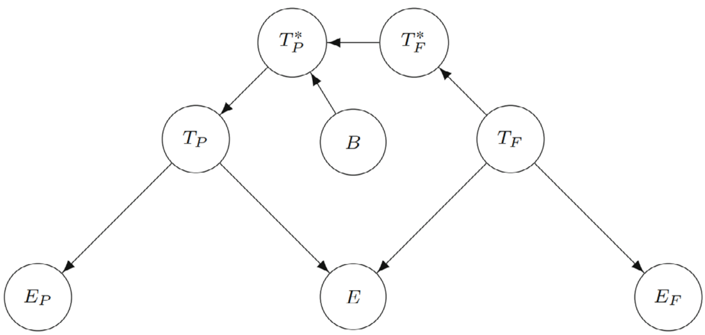

Figure 5: A DAG for GNS*.
Source: (Tesic, 2019,p. 1112).

The new distribution changes slightly because values for P(B) must now be given, the value of $P(T^*_p \mid T^*_f)$ is not set to 1, and there is a conditional distribution of the new conditional probability (T^*_p at B). This allows GNS* not only to prove the same thing as GNS (evidence for the reduced theory is evidence for the reducing one and vice versa), but also to determine when this transfer is actually relevant, i.e., when E_p actually contributes to E_f confirmation of T_f and vice versa (see Theorems 5-8).

Some problems for these models

Recall that arrows in DAGs mean an asymmetric and non-transitive difference-making relationship and that you can only represent one kind of relation by arrows, since DAGs are not heterogeneous graphs. It is also worth remembering that constructing BNs from DAGs is about parameterizing DAG to find a Markov compatible probability distribution.

Although I agree with Teŝić assessment of GNS, I generally disagree with the way the intertheoretical reduction is modeled. The first problem I see with both models is the large number of relations modeled in both BNs. Before the reduction, we have some nice arrows connecting theories to evidence, as explained in the section (3) and at the beginning of the section (4). This relation is a very particular one that relates an empirical set of propositions to some evidence for them.

After the reduction, however, we witness how a number of new relations are added to the model. The problems are the following:

1. The relation between T_f and T^*_f is that of derivation. While this can be represented probabilistically, it is not the same relationship as that between T_f and the evidence supporting it. The derivation in the case of T_f and T^*_f is transitive, while there is by definition no transitive relation between the theory and its evidence. This creates two problems for this arrow: it is not coherent with the DAG and is transitive.
2. In GNS the relation between T^*_f and T^*_p is a relation between the vocabularies of the terms present in both theories, it is a conventional but semantic relation obtained thanks to the bridge laws (Nagel, 1961, p. 354). The problem with this relation is that it is symmetric and is not captured by restricting $P(T^*_p \mid T^*_f)$ to 1, because that tells us nothing about how the vocabulary in T^*_p is related to the vocabulary in T^*_f. To repre-

sent this, we would have to accept that $P(T^*_p \mid T^*_f) = P(T^*_f \mid T^*_p)$, which cannot be represented in a DAG, since this would create a loop.

3. As for the same arrow in GNS*, the new arrow no longer represents the logical consequence, but since the bridge laws are now exogenously defined: we are forced to read the relation between the restricted version of the fundamental theory and T^*_p as T^*_f depending not only on T^*_f, but also on B. The problem is that the relation between T^*_f and T^*_p is different from that between B and T^*_p. So even if we could prove that one of the two relations is the same as the one represented before the reduction in BN, we would still have an alien relation in both cases.
4. Lastly in both GNS and GNS* we have found an arrow between T^*_p and T_p that attempts to capture strong analogy in both cases. However, analogy is usually a symmetric relationship, so if analogy exists between objects/propositions X and Y, then it also exists between Y and X. Even if one could argue that it is non-transitive (which is also not always the case, especially in scientific reasoning. see (Achinstein, 1964, p. 342)), it is still not representable with an arrow.

In view of the above, I believe that both GNS and GNS* are not adequate BNs, since they violate the principles of modeling just listed. I will now move on to the conclusions, where I will point out some consequences of this and outline some possible ways out.

Conclusion: a way out?

In this paper I defended that arrows in DAGs and consequently in BNs can only be drawn to represent asymmetric, non-transitive, difference-making relationships. Moreover, arrows should represent only one such relation per DAG. I then showed how this is relevant to Bayesian philosophy of science by arguing that two models for inter-theoretic reduction do not reflect these principles and are therefore not suitable for modeling purposes.

There are three possible ways out of this predicament, each with its own difficulties. The first is to accept the difficulties of the model after the reduction and return to the original BN, which satisfies the criteria presented in this paper. One could establish the main results of GNS and GNS* with non-Bayesian arguments, but this perhaps defeats the purpose of constructing a Bayesian model of intertheoretic reduction.

The second option would be to reduce the relationships between all nodes in the DAG to be relations of confirmation, as shown in Fig. (2). This may be promising, but would change the basic assumptions about how variables influence each other, completely altering the framework originally offered by Nagel and Schaffner. For example, one would have to accept that strong analogy is nothing more than an asymmetric relationship.

There is a third possible way out, which I personally consider more promising, but which unfortunately requires major changes in our modeling apparatus. That is, to diversify the formal apparatus of Bayesian philosophy and introduce condensable heterogeneous graphs to better model concepts that require different kinds of difference-making relationships (and perhaps even different kinds of variables). A graph is heterogeneous if it has different types of nodes and arrows, and it is condensable if each strongly connected component can be contracted into a new graph (i.e., each subset of nodes that can reach each other becomes a single node (Bondy & Murty, 2008, p. 63)). We could apply this by constructing a heterogeneous cyclic graph that might contain loops, and then condensing it into a DAG. For example, consider this new graph, which disregards the limitations imposed by the theory of BN to favor a graphical representation (Fig. (6)).

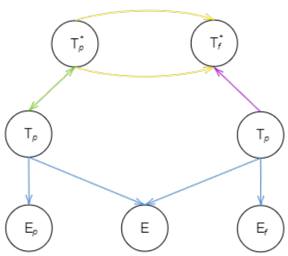

Figure 6: A heterogeneous graph for intertheoretic reduction that can represent the different relations observed. Theory-evidence in blue, strong analogy in green, inter-definition via bridge laws in yellow and derivation in purple. Source: own elaboration

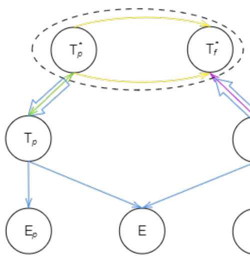

Figure 7: A DAG for intertheoretic reduction where the original graph has been condensed and thus can be read off as probabilistic dependencies. Source: own elaboration

Since any compactification of a directed graph directed graph is a DAG (Harary et al., 1965, Theorem 3.6), we can find a dag that meets our needs (Fig. (7)). This may prove to be very effective, since we can still represent all the different relationships and at the same time mask them as probability dependencies. So, it may work as a better graphical representation of the concept.

References

Abramson, B. (1994). The design of belief network-based systems for price forecasting. *Computers & electrical engineering*, *20*(2), 163-180.

Achinstein, P. (1964). Models, analogies, and theories. *Philosophy of Science*, *31*(4), 328-350.

Andreassen, S., Woldbye, M., Falck, B., & Andersen, S. K. (1987). Munin: A causal probabilisticnetwork for interpretation of electromyographic findings. *Proceedings of the 10th international joint conference on Artificial intelligence-Volume 1*, 366-372.

Beaumont, M. A., & Rannala, B. (2004). The bayesian revolution in genetics. *Nature Reviews Genetics*, *5*(4), 251-261.

Bondy, J., & Murty, U. (2008). *Graph theory* (Vol. 244). Springer New York.

Bovens, L., & Hartmann, S. (2004). *Bayesian epistemology*. Oxford University Press.

Breese, J. S., Horvitz, E. J., Peot, M. A., Gay, R., & Quentin, G. H. (1992). Automated decisionanalytic diagnosis of thermal performance in gas turbines. *Turbo Expo: Power for Land, Sea, and Air*, 78972, V005T15A015.

Darwiche, A. (2009). *Modeling and reasoning with bayesian networks*. Cambridge University Press.

Dawid, R., Hartmann, S., & Sprenger, J. (2015). The no alternatives argument. *The British Journal for the Philosophy of Science*, *66*(1), 213-234.

DGL. (2022). Heterogeneous graphs. In *Deep graph library (DGL) tutorials and documentation*. https://docs.dgl.ai/guide/graph-heterogeneous.html?highlight=heterogenous

Dizadji-Bahmani, F., Frigg, R., & Hartmann, S. (2010). Who's afraid of nagelian reduction? *Erkenntnis*, *73*(3), 393-412.

Dizadji-Bahmani, F., Frigg, R., & Hartmann, S. (2011). Confirmation and reduction: A bayesian account. *Synthese*, *179*(2), 321-338.

Gan, Q.,Wang, M., Li, M., Karypis, G., & Zhang, Z. (2021).Working with heterogeneous graphs. In *Deep graph library (dgl) tutorials and documentation*. https://docs.dgl.ai/en/0.6.x/ tutorials/basics/5_hetero.html

Gu, Y., Peiris, D., Crawford, J.W., NcNicol, J., Marshall, B., & Jefferies, R. A. (1994). An application of belief networks to future crop production. *Proceedings of the Tenth Conference on Artificial Intelligence for Applications*, 305-309.

Gutin, G. (2018). Acyclic digraphs. In J. Bang-Jensen & G. Gutin (Eds.), *Classes of directed graphs* (pp. 125-172). Springer International Publishing. https://doi.org/10.1007/978-3-319-71840-8_3

Hájek, A., & Hartmann, S. (2010). Bayesian epistemology. In J. Dancy (Ed.), *A companion to epistemology* (pp. 93-105). Blackwell.

Harary, F., Norman, R. Z., & Cartwright, D. (1965). *Structural models: Introduction to the theory of directed graphs*. John Wiley & Sons Inc.

Hitchcock, C. (2020). Causal Models. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Summer 2020)*. Metaphysics Research Lab, Stanford University.

Jarrom, D., Elston, L., Washington, J., Prettyjohns, M., Cann, K., Myles, S., & Groves, P. (2022). Effectiveness of tests to detect the presence of sars-cov-2 virus, and antibodies to sars-cov-2, to inform covid-19 diagnosis: A rapid systematic review. *BMJ evidence-based medicine*, *27*(1), 33-45.

Kiiveri, H., Speed, T. P., & Carlin, J. B. (1984). Recursive causal models. *Journal of the australian Mathematical Society*, *36*(1), 30-52.

Kyrimi, E., McLachlan, S., Dube, K., Neves, M. R., Fahmi, A., & Fenton, N. (2021). A comprehensive scoping review of bayesian networks in healthcare: Past, present and future. *Artificial Intelligence in Medicine*, 117, 102108. https://doi.org/10.1016/j.artmed.2021.102108

Lauritzen, S. (1979/1982). *Lectures on contingency tables* (2nd ed.). University of Aalborg Press (Originally published in 1979).

Lauritzen, S. L., & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50 (2), 157-194.

Levitt, T. S., Agosta, J. M., & Binford, T. O. (1990). Model-based influence diagrams for machine vision. In *Machine intelligence and pattern recognition* (pp. 371-388, Vol. 10). Elsevier.

Nadi, F., Agogino, A. M., & Hodges, D. A. (1991). Use of influence diagrams and neural networks in modeling semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, 4(1), 52-58.

Nagel, E. (1961). The structure of science, problems in the logic of scientific explanation.

Neapolitan, R. E. (2004). *Learning bayesian networks*. Pearson Prentice Hall.

Needham, C. J., Bradford, J. R., Bulpitt, A. J., & Westhead, D. R. (2007). A primer on learning in bayesian networks for computational biology. *PLoS computational biology*, 3(8), e129.

Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. *Proceedings of the 7th conference of the Cognitive Science Society*, University of California, Irvine, CA, USA, 329-334.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*.Morgan Kaufmann.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Rehg, J. M., Murphy, K. P., & Fieguth, P. W. (1999). Vision-based speaker detection usingbayesian networks. *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, 2, 110-116.

Sarkar, S. (2015). Nagel on reduction. *Studies in History and Philosophy of Science Part A*, 53, 43-56.

Schaffner, K. F. (1967). Approaches to reduction. *Philosophy of science*, *34*(2), 137-147.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., & Cowell, R. G. (1993). Bayesian analysisin expert systems. *Statistical science*, 219-247.

Sprenger, J. (2016). The probabilistic no miracles argument. *European Journal for Philosophy of Science*, *6*(2), 173-189.

Sprenger, J., & Hartmann, S. (2019). *Bayesian philosophy of science*. Oxford University Press.

Talbott, W. (2016). Bayesian Epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Winter 2016)*. Metaphysics Research Lab, Stanford University.

Tesic, M. (2019). Confirmation and the generalized Nagel-Schaffner model of reduction: A Bayesian analysis. *Synthese*, *196*(3), 1097-1129.

Van Riel, R. (2014). *The concept of reduction*. Springer.

Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., & Yu, P. S. (2019). Heterogeneous graph attention network. *The World Wide Web Conferenc*e, 2022-2032. https://doi.org/10.1145/3308558.3313562

Wermuth, N., & Lauritzen, S. L. (1982). *Graphical and recursive models for contigency tables*. Institut for Elektroniske Systemer, Aalborg Universitetscenter.

Wiegerinck, W., Burgers, W., & Kappen, B. (2013). Bayesian networks, introduction and practical applications. In M. Bianchini, M. Maggini, & L. C. Jain (Eds.), *Handbook on neural information processing* (pp. 401-431). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36657-4_12

Wimsatt, W. C. (1974). Reductive explanation: A functional account. PSA: Proceedings of the biennial meeting of the Philosophy of Science Association, 1974, 671-710.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, (20), 557-585.

Cómo citar en APA: Rivas-Robledo, P. (2024). When should we draw arrows? Assessing the use of Bayesian networks in philosophy of science. *Humanitas Hodie*, *7*(1), H71a3. https://doi.org/10.28970/ hh.2024.1.a3