Confirmation and reduction: a Bayesian account

F. Dizadji-Bahmani · R. Frigg · S. Hartmann

Abstract   Various scientific theories stand in a reductive relation to each other. In a recent article, we have argued that a generalized version of the Nagel-Schaffner model (GNS) is the right account of this relation. In this article, we present a Bayesian analysis of how GNS impacts on confirmation. We formalize the relation between the reducing and the reduced theory before and after the reduction using Bayesian networks, and thereby show that, post-reduction, the two theories are confirmatory of each other. We then ask when a purported reduction should be accepted on epistemic grounds. To do so, we compare the prior and posterior probabilities of the conjunction of both theories before and after the reduction and ask how well each is confirmed by the available evidence.

1 Introduction

A number of contemporary scientific research programs are reductionist in the sense that their aim is to account for the behaviour of a system at a certain level of organization

F. Dizadji-Bahmani · R. Frigg
Department of Philosophy, Logic and Scientific Method, London School of Economics
and Political Science, London, WC2A 2AE, UK
e-mail: F.Dizadji-Bahmani@lse.ac.uk

R. Frigg
e-mail: R.P.Frigg@lse.ac.uk

S. Hartmann (✉)
Tilburg Center for Logic and Philosophy of Science, Tilburg University,
PO Box 90153, 5000 LE, Tilburg, The Netherlands
e-mail: S.Hartmann@uvt.nl

in terms of the behavior of its constituents. Statistical mechanics (SM) is a case in point: it aims to account for the laws of thermodynamics (TD) in terms of the laws of mechanics and probabilistic assumptions. Other examples of (putative) intertheoretical reductions are chemistry to atomic physics, rigid body mechanics to particle mechanics, psychology to neuroscience, and agent-based modeling in the social sciences.

These programs have attracted the attention of philosophers and have led to a renewed interest in Nagelian reduction, which, for many years, has been considered a dead end.[1] In an earlier article (Dizadji-Bahmani et al. 2010), we have argued that the points leveled against Nagel's original model (1961, Ch. 11) can either be overcome in a more sophisticated version of the approach, which we call the Generalized Nagel-Schaffner Model of reduction (GNS), or turn out to be red herrings on closer inspection. Part of our defence of Nagelian reduction was that it has desirable consequences for the confirmation of theories: given two theories whose target domains are identical (or have significant overlap), evidence confirming one theory should also confirm the other theory, but this can only happen if an intertheoretic link is established. GNS provides this link. In this paper, we show in detail how the establishment of an intertheoretical reduction (in the sense of GNS) boosts the confirmation of both theories. More specifically, we use the framework of Bayesian confirmation theory and show that if there is a reductive relation between two theories, then confirmation flows both from the reducing to the reduced theory and from the reduced to the reducing theory. For instance, evidence that exclusively confirms SM before the reduction also confirms (though perhaps to a lower degree) TD after the reduction, and *vice versa*.

The plan is as follows. Section 2 sets the scene by briefly outlining GNS and introducing the case of SM, which will serve as our example throughout the paper. In Sect. 3, we present a short summary of Bayesian confirmation theory. Section 4 contains the main argument: we consider confirmation in a scenario in which there is no intertheoretic reduction and then compare it with one in which there is. The result is that reduction boosts confirmation. Section 5 discusses various implications of this result. In Sect. 6, we sum up our results and outline a number of open problems.

2 The generalized Nagel-Schaffner model

TD describes systems like gases and solids in terms of macroscopic properties such as volume, pressure, temperature and entropy, and gives a correct description of the behaviour of such systems. The aim of statistical mechanics is to account for the laws of TD in terms of the dynamical laws governing the microscopic constituents of macroscopic systems and probabilistic assumptions.[2] In particular, SM aims to show that the Second Law of TD is a consequence of the mechanical motion of the molecules of the gas. For example, consider a container divided in two by a partition wall. The left half is filled with a gas, while the right half is empty. If we now remove the partition, the gas will spread and soon be evenly distributed throughout the entire container; the

---

[1] See, for example, Klein (2009); Needham (2009) and Winther (2009).

[2] For a detailed discussion of SM see Frigg (2008) and references therein.

gas's entropy increases as it spreads. This is an instance of a process obeying the Second Law of TD. Roughly speaking, the Second Law says that the entropy of a closed system cannot decrease, and usually increases when the system is left on its own in a non-equilibrium state. The aim of SM is to account for the Second Law in terms of the equations governing the motion of the molecules of the gas and some probabilistic assumptions; that is, it aims to show that the Second Law is a consequence of its basic postulates. Or almost. In fact, it is impossible to derive the *exact* Second Law (which is an exceptionless law) from a probabilistic theory. So what we aim to do instead is to derive a probabilistic law that is strongly analogous to the TD Second Law, in this case the proposition that entropy is highly likely to increase (which is known as Boltzmann's Law).

That analogous versions of the laws of the phenomenological theory (here TD) should follow from the laws of the fundamental theory (here SM) is the basic idea of GNS.[3] Consider a phenomenological theory $\mathbf{T_P}$ and a fundamental theory $\mathbf{T_F}$. Each theory has a number empirical propositions. Let $\mathcal{T}_P := \{T_P^{(1)}, \ldots, T_P^{(n_P)}\}$ and $\mathcal{T}_F := \{T_F^{(1)}, \ldots, T_F^{(n_F)}\}$ be the set of empirical propositions of $\mathbf{T_P}$ and $\mathbf{T_F}$ respectively.[4] The reduction of $\mathbf{T_P}$ to $\mathbf{T_F}$ consists of the following three steps (for details and discussion, see Dizadji-Bahmani et al. (2010) and Schaffner (1967)).

1. Adopt assumptions (so-called 'auxiliary assumptions') describing the particular setup under investigation. Here, these are assumptions about the mechanical properties of the gas molecules. Then derive from these and $\mathcal{T}_F$ a restricted version of each element $T_F^{(i)}$ of $\mathcal{T}_F$. Denote these by $T_F^{*(i)}$ and the corresponding set by $\mathcal{T}_F^* := \{T_F^{*(1)}, \ldots, T_F^{*(n_F)}\}$.

2. $\mathcal{T}_F$ and $\mathcal{T}_P$ are formulated in different vocabularies. In our example, SM talks about trajectories in phase space and probability measures while TD talks about macroscopic properties such as pressure and temperature. In order to connect the two theories, we adopt bridge laws. These connect terms of one theory with terms of the other, for instance mean kinetic energy in SM with temperature in TD. Substituting the terms in $\mathcal{T}_F^*$ with terms from the macro theory as per the bridge laws yields $\mathcal{T}_P^*$, i.e. the set $\{T_P^{*(1)}, \ldots, T_P^{*(n_P)}\}$.

3. Show that each element of $\mathcal{T}_P^*$ is strongly analogous to the corresponding element in $\mathcal{T}_P$.

If these conditions obtain, $\mathbf{T_P}$ is reduced to $\mathbf{T_F}$. Following standard terminology, we say that $\mathbf{T_P}$ is the reduced theory and that $\mathbf{T_F}$ is the reducing theory.

Let us now consider how theories are supported by evidence. With regard to our two theories, there are three kinds of evidence: evidence that only confirms, to some degree, the phenomenological theory; evidence that only confirms, to some degree, the fundamental theory; and evidence that confirms, to some degree, both. We make

---

[3] For a detailed exposition of GNS as well as a defence see our (2010).

[4] The empirical propositions of a theory are its various laws. GNS has it that two theories stand in a reductive relation to one another in virtue of certain relations obtaining between their empirical propositions, as is set out in the main text. Notice, however, that we are not committed to the view that a theory *just is* a set of laws. i.e. $\mathbf{T_A}$ is not identified with $\mathcal{T}_A$.

this clear with examples from TD and SM. For the first case, consider what is known as the Joule-Thomson process: there are two chambers of different dimensions connected to each other by a permeable membrane, filled with a gas. At the end of each chamber, there is a piston which allows the pressure and volume for the gas in each chamber to be varied by applying a force. The pressure in the first chamber is higher than the pressure in the second. Now push the gas from the first chamber into the second, but so slowly that the pressure remains constant in both chambers and no heat is exchanged with the environment. Then, the gas in the second chamber cools down. The amount of cooling can be calculated using the principles of TD, and is found to coincide with experimental values. So we have a confirmation of TD, but not of SM because no SM assumptions have been used in the argument. For the second, consider the dependence of a metal's electrical conductivity on temperature. From SM, one can derive an equation relating the change in the electrical conductivity of certain metals given a change in temperature which is what one finds in experiment. TD, in contrast, is entirely silent about this phenomenon. Third, consider again the gas confined to the left half of the box which spreads evenly when the dividing wall is removed. It follows from TD that the thermodynamic entropy of the gas increases; at the same time, it is a consequence of SM that the Boltzmann entropy increases in that process. So the spreading of the gas confirms both SM and TD.

3 A primer on Bayesianism

Bayesianism is, arguably, our best account of the logic of science.[5] It is a quantitative framework in which what it means for a piece of evidence to confirm a hypothesis is made precise and various puzzles pertaining to confirmation are resolved. The plan for this section is to briefly outline the Bayesian model of confirmation and introduce Bayesian Networks, which are a useful representational tool. This will allow us to explore the relation between reduction and confirmation in a precise way in the next section.

Consider a hypothesis, H and a piece of evidence, E. E confirms H if $P(\text{H}|\text{E}) > P(\text{H})$; E disconfirms H if $P(\text{H}|\text{E}) < P(\text{H})$; E is irrelevant for H if $P(\text{H}|\text{E}) = P(\text{H})$. $P(\text{H}|\text{E})$ is related to $P(\text{E}|\text{H})$, $P(\text{H})$ and $P(\text{E})$ via Bayes' Theorem:

$$P(\text{H}|\text{E}) = \frac{P(\text{E}|\text{H})P(\text{H})}{P(\text{E})} \tag{1}$$

$$= \frac{P(\text{E}|\text{H})P(\text{H})}{P(\text{E}|\text{H})P(\text{H}) + P(\text{E}|\neg\text{H})P(\neg\text{H})} \tag{2}$$

By fixing the relevant probabilities, Bayes theorem gives us the value of $P(\text{H}|\text{E})$, which, under the above criterion, determines the confirmatory relation between the evidence and the hypothesis.

---

[5] Bayesianism is presented and critically discussed in Earman (1992) and Howson and Urbach (2005). Recent surveys of Bayesian epistemology are Hájek and Hartmann (2010) and Hartmann and Sprenger (2010).

Fig. 1 The Bayesian Network
representing the relation
between *E* and *H*
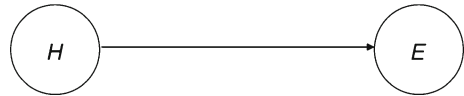
For our purposes, it will be useful to explore the relation between reduction and confirmation in terms of Bayesian networks.[6] Bayesian networks considerably help to simplify (or make feasible) calculations in complicated multi-variate situations. More specifically, they help in representing the relations between (sets of) variables; graphical structure allows us to read these off directly.

A *directed graph* is a graph consisting of nodes and arrows. *Parent nodes* are nodes with outgoing arrows; *child nodes* are nodes with incoming arrows; and *root nodes* are nodes with only outgoing arrows. A *directed acyclical graph* (DAG) is a graph consisting of nodes and arrows such that one cannot run in a cycle; that is, a directed graph such that there is no path from a parent node back to itself. A Bayesian network is a DAG whose nodes represent propositional variables, and arrows encode the relationship between these variables.

Let us now frame the confirmatory relation between a hypothesis and a piece of evidence in terms of a Bayesian network. We introduce two binary propositional variables *E* and *H*. *H* can take two values: H, the hypothesis is true, and ¬ H, the hypothesis is false. *E* has two values as well, viz. E, the evidence obtains, and ¬E, the evidence does not obtain.[7] The relation between *E* and *H* can be represented in the Bayesian network depicted in Fig. 1.

The arrow from the propositional variable *H* to *E* means that there is a direct influence of *H* on *E*.[8] The arrow goes from *H* to *E* because the truth or falsity of the hypothesis affects the probability of the evidence obtaining. For example, it may be more likely that *E* obtains if *H* is true than if *H* is false.

In general, to fully specify the Bayesian network, we have to fix two sets of parameters. First, the prior probabilities of all root nodes. In our case, this means that we have to specify $P(H)$. Second, the conditional probabilities of all other nodes, given their parents. In the present case, this means that we have to fix the likelihoods $P(E|H)$ and $P(E|\neg H)$. Using Bayes' theorem, we can then fully specify all the other probabilities.

To simplify discussion of the relation between reduction and confirmation in the next section, we introduce the following notional conventions: $h := P(H)$, $h^* := P(H|E)$, $p := P(E|H)$ and $q := P(E|\neg H)$. Then Eq. 2 becomes:

---

[6] Bayesian networks are introduced in Neapolitan (2003) and Pearl (1988). Bovens and Hartmann (2003) give a non-technical introduction and discuss applications from epistemology and philosophy of science.

[7] Propositional variables are denoted by italicized letters; their particular values are denoted by non-italicized letters. Throughout, we are only concerned with two-valued propositional variables, *A*, which can take values A or ¬ A.

[8] Unlike some users of Bayesian networks, for instance Spirtes et al. (2001), we are only interested in probabilistic dependencies between variables, and not in causal relations.
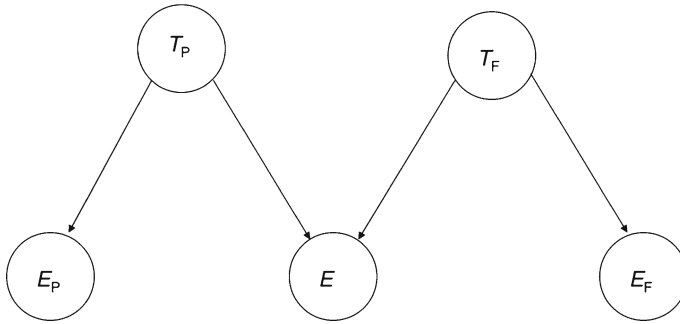
Fig. 2 The Bayesian network representing the situation *before* the reduction

$$h^* = \frac{h\,p}{h\,p + (1-h)\,q} = \frac{1}{1 + \frac{1-h}{h} \cdot \frac{q}{p}} \tag{3}$$

The expression $x := q/p$ is called the *likelihood ratio*. If E confirms H, then $x \in (0, 1)$. If E disconfirms H, then $x > 1$. If E and H are independent, then $x = 1$. Finally we set $\bar{z} := 1 - z$ for all parameters z and use this abbreviation in the remainder.

To avoid technical complications, we assume that all probabilities are non-extreme throughout, i.e. that they lie in the open interval $(0, 1)$. This with the exception of conditional probabilities $P(A|B)$ where A is a logical consequence of B; in this case we have $P(A|B) = 1$.

4 Reduction and confirmation

4.1 Before the reduction

We examine the situation before a reduction is attempted. To simplify things, we assume that $\mathcal{T}_F$ and $\mathcal{T}_P$ have only one element, viz. T_F and T_P respectively. The generalization to more than one element is conceptually straightforward. Furthermore, E confirms T_F and T_P, E_F only confirms T_F and E_P only confirms T_P. Introducing corresponding propositional variables T_F, T_P, E, E_F and E_P, we can represent the situation before the attempted reduction in the Bayesian network depicted in Fig. 2.

Following our methodology, we have to specify the prior probabilities of T_F and T_P (i.e. of all root nodes) and the conditional probabilities of E, E_F and E_P (i.e. of all child nodes), given their parents. Let the corresponding probability measure be P_1. We denote:

$$
\begin{aligned}
P_1(T_F) &= t_F, & P_1(T_P) &= t_P \\
P_1(E_F|T_F) &= p_F, & P_1(E_F|\neg T_F) &= q_F \\
P_1(E_P|T_P) &= p_P, & P_1(E_P|\neg T_P) &= q_P \\
P_1(E|T_F, T_P) &= \alpha, & P_1(E|T_F, \neg T_P) &= \beta \\
P_1(E|\neg T_F, T_P) &= \gamma, & P_1(E|\neg T_F, \neg T_P) &= \delta
\end{aligned}
\tag{4}
$$

These parameters cannot be freely chosen as we assume that the following conditions hold: First, E_F confirms T_F, hence $p_F > q_F$. Second, E_P confirms T_P, hence, $p_P > q_P$. Third, E confirms T_F and fourth E confirms T_P. The last two conditions entail the following constraints on α, β, γ and δ (all proofs are in the Appendix)[9]:

$$(\alpha - \beta)\, t_F + (\gamma - \delta)\, \bar{t}_F > 0 \tag{5}$$
$$(\alpha - \gamma)\, t_P + (\beta - \delta)\, \bar{t}_P > 0 \tag{6}$$

These inequalities hold, for example, if $\alpha > \beta, \gamma > \delta$, which seems to be a natural condition.

Given this network structure and the conditional independences encoded in it, it is easy to see, for example, that the variable E_F is independent of T_P given T_F and that E_P is independent of T_F given T_P. In symbols:

$$E_F \perp\!\!\!\perp T_P | T_F, \quad E_P \perp\!\!\!\perp T_F | T_P \tag{7}$$

Hence, E_F does not confirm (or disconfirm) T_P and E_P does not confirm (or disconfirm) T_F:

$$P_1(T_P | E_F) = P_1(T_P), \quad P_1(T_F | E_P) = P_1(T_F) \tag{8}$$

We conclude that there is no flow of confirmation from one theory to the other. The intuitive reason for this is that there is no chain of arrows from E_F to T_P. Note also that the variables T_F and T_P are probabilistically independent before the reduction:

$$P_1(T_F, T_P) = P_1(T_F)\, P_1(T_P) = t_F\, t_P \tag{9}$$

All this may, however, not be right in practice. Scientists may feel, for example, that the two theories are much more intimately connected. An indication for this may be that there is, as we assume, evidence E that supports both theories. Another reason may be that there are formal (or other) relations between the two theories. In this case, scientists will attempt to reduce one theory to the other. Let us now model this situation.

4.2 After the reduction

Recall the three steps involved in reducing one theory to another set out in Sect. 2: First, derive T_F^* from the auxiliary assumptions and T_F. Second, introduce bridge laws and obtain T_P^* from T_F^*. Third, show that T_P^* is strongly analogous to T_P.

The situation after the reduction can then be represented in the Bayesian network depicted in Fig. 3. To complete the network, we specify the following conditional probabilities:

---

[9] One may also want to require that $P_1(T_F | E, E_F) > P_1(T_F)$ and $P_1(T_P | E, E_P) > P_1(T_P)$. Note, however, that both inequalities follow from the above four conditions (proof omitted).
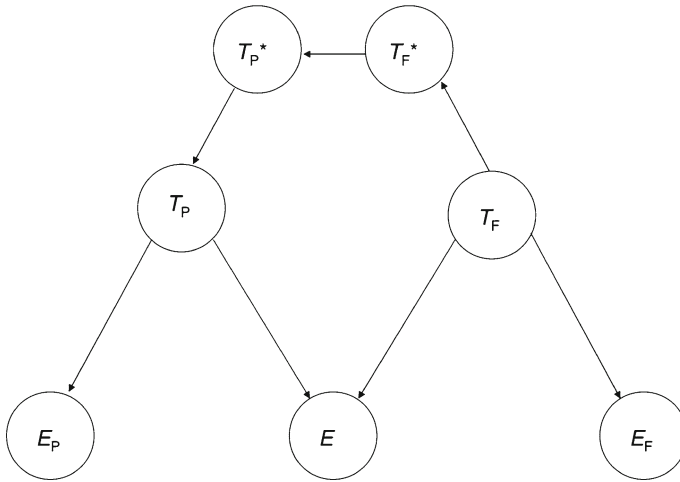
Fig. 3 The Bayesian Network representing the situation *after* the reduction

$$P_2(T_P|T_P^*) = p_P^*, \quad P_2(T_P|\neg T_P^*) = q_P^* \qquad (10)$$

$$P_2(T_F^*|T_F) = p_F^*, \quad P_2(T_F|\neg T_F) = q_F^* \qquad (11)$$

Note that Eq. 10 replaces the second equation in the first line of Eq. 4. We also have to represent the bridge law in probabilistic terms. Naturally, we require:

$$P_2(T_P^*|T_F^*) = 1, \quad P_2(T_P^*|\neg T_F^*) = 0 \qquad (12)$$

All other probability assignments hold as in the case of P_1. Requiring this condition makes sure that we can compare the two scenarios later, i.e. the situations before and after the reduction.

Three remarks about the three steps in the reduction are in order. First, T_F^* may be more or less good. How good it is depends on the context (i.e. the application in question and the auxiliary assumptions made) and on the judgement of the scientists involved. In line with our Bayesian approach, we assume that the judgement of the scientists can be expressed in probabilistic terms. Second, the move from T_F^* to T_P^* in virtue of the bridge laws may well be controversial amongst scientists. Whilst bridge-laws are non-conventional factual claims, different scientists may assign different credences to them. Third, what counts as strongly analogous will also depend on the specific context and on the judgement of the scientists. For example, whether entropy fluctuations can be neglected or not cannot be decided independently of the specific problem at hand, see Callender (2001).[10] All this fits our Bayesian account well.

Note that, in the Bayesian network in Fig. 3, there is now a direct sequence of arrows from T_F to T_P, i.e. the path through T_F^* to T_P^*. And hence, we expect that E_F is now probabilistically relevant for T_P and that E_P is now probabilistically relevant for

---

[10] Again, for a full defence of the GNS see Dizadji-Bahmani et al. (2010).

T_F. And this is indeed what we find: the independencies formulated in Eq. 7 do not hold any more. We state our results in the following two theorems:

Theorem 1 *E_F confirms T_P iff* $(p_F - q_F)(p_F^* - q_F^*)(p_P^* - q_P^*) > 0.$

This theorem entails that E_F confirms T_P if the following three conditions hold: (i) E_F confirms T_F (i.e. $p_F > q_F$), (ii) T_F confirms[11] T_F^* (i.e. $p_F^* > q_F^*$), and (iii) T_P^* confirms T_P (i.e. $p_P^* > q_P^*$). These conditions are immediately plausible. Condition (i) was assumed from the beginning, and conditions (ii) and (iii) make sure that there is a positive flow of confirmation from T_F to $T_F^* \equiv T_P^*$ (*qua* bridge law) and from T_P^* to T_P.[12]

Theorem 2 *E_P confirms T_F iff* $(p_P - q_P)(p_F^* - q_F^*)(p_P^* - q_P^*) > 0.$

This theorem is analogous to the previous theorem. It entails that E_P confirms T_F if the following three conditions hold: (i) E_P confirms T_P (i.e. $p_P > q_P$), (ii) T_F confirms T_F^* (i.e. $p_F^* > q_F^*$), and (iii) T_P^* confirms T_P (i.e. $p_P^* > q_P^*$).

Note that, in our representation, the bridge law states a *perfect correlation* between T_F^* and T_P^*. A bridge law is posited by scientists working in a particular field, and it may happen that not everybody in that community is convinced of it. Thus, different scientists may assign different credences to a particular bridge law. In a case where a lower likelihood is assigned to a bridge law, the reduction may still be epistemically valuable—the flow of confirmation will just be less. How much confirmation will flow depends, of course, on the values of the likelihoods.

For future reference, let us calculate the prior probability of the conjunction of both theories. We obtain:

$$P_2(T_F, T_P) = t_F \left(p_F^* \, p_P^* + \overline{p}_F^* \, q_P^* \right) \tag{13}$$

In a similar way, the posterior probability of both theories given the total evidence, i.e. the expression $P_2(T_F, T_P | E, E_F, E_P)$, can be calculated (see Appendix).

Finally, let us remark on the specific representation we have chosen in the Bayesian Network in Fig. 3. Clearly, having a sequence of arrows from T_F to T_P ensures that confirmation can flow from one theory to the other. This sequence of arrows from the reducing theory to the reduced theory, however, makes sense. It is not just driven by our wish to establish a flow of confirmation from the reducing theory to the reduced theory. First, T_F^* is an approximation of T_F. It follows from it and (or so we argue) depends on it in a probabilistically sense, which justifies the direction of the arrow. Second, we have drawn an arrow from T_F^* to T_P^* although the propositional variables in question are, *qua* the bridge law, intersubstitutable with each other. This is modeled by assigning appropriate conditional probabilities. The arrow could have also been drawn

---

[11] Some authors use the word 'confirm' only to refer to the relation between between a hypothesis and a piece of evidence and the word 'support' to (also) include the relation between two theories. We do not follow this usage and use the word 'confirm' in both cases.

[12] Interestingly, however, E_F also confirms T_P if condition (i) holds and if both $p_F^* < q_F^*$ and $p_P^* < q_P^*$. This result, which shows up in related confirmation-theoretical contexts, should not worry us as all that is is important is that E_P confirms T_F if $p_F^* > q_F^*$ and if $p_P^* > q_P^*$, as is, indeed, the case.

from T_P^* to T_F^*. In this case we had to require $P(T_F^*|T_P^*) = 1$ and $P(T_F^*|\neg T_P^*) = 0$. These conditions are, however, equivalent to Eq. 12 for non-extreme priors. Third, it may look strange that we have drawn an arrow from T_P^* to T_P to model the relation of strong analogy as a symmetrical relation. We would like to reply to this objection that, then, 'analogy' is perhaps not the right word as T_P^* is indeed stronger than T_P, and so it makes sense to draw an arrow from T_P^* to T_P. We conclude that the chain of arrows from T_F to T_P is indeed plausible.

5 Why accept a purported reduction?

Under what conditions should we accept a proposed reduction? More specifically, given everything we *know* about the domains of the two theories, when should we accept a proposed reduction and when should we reject it? In the Bayesian framework the aim is to raise relevant probabilities, and we accept a reduction if it achieves this goal. The crucial question then is which probabilities are relevant. Is it the prior probability of the conjunction of T_F and T_P? Or the posterior probability of the conjunction of T_F and T_P, i.e. the probability of T_F and T_P given the total evidence (i.e. E, E_F and E_P)? Or should we rather accept a proposed reduction if the conjunction of T_F and T_P is better *confirmed* by the evidence after the reduction (compared to the situation before the reduction)? If one decides to follow the last proposal, then one will also have to choose one of the various confirmation measures (Fitelson 1999). We examine these three proposals in turn.

Let us first compare the prior probabilities of the conjunction of T_F and T_P before and after the reduction. Before the reduction, the two theories are independent, as expressed in Eq. 9. For convenience, let us restate the condition:

$$P_1(T_F, T_P) = t_F\, t_P \tag{14}$$

We also restate the prior probability of the conjunction of T_F and T_P after the reduction and obtain

$$P_2(T_F, T_P) = t_F \left(p_F^* \, p_P^* + \overline{p}_F^* \, q_P^* \right). \tag{15}$$

While the expression in Eq. 14 is an explicit function of t_P, the expression in Eq. 15 is not. This is because, after the reduction, T_P is no longer a root node, and so it is not assigned a prior probability. In order to meaningfully compare the situation before and after the reduction, we not only have to assume that $P_2(E_P|T_P) = P_1(E_P|T_P)$ etc., but also that $P_2(T_P) = P_1(T_P)$. Let us therefore calculate:

$$\tilde{t}_P := P_2(T_P) = t_F^* \, p_P^* + \overline{t}_F^* \, q_P^* \tag{16}$$

with

$$t_F^* = P_2(T_F^*) = P_2(T_P^*) = p_F^* \, t_F + q_F^* \, \overline{t}_F. \tag{17}$$

Alternatively, we have:

$$\tilde{t}_P := (p_F^* p_P^* + \overline{p}_F^* q_P^*) \, t_F + (q_F^* p_P^* + \overline{q}_F^* q_P^*) \, \overline{t}_F \tag{18}$$

This equation follows if we insert Eq. 17 into Eq. 16 or by direct calculation from the Bayesian network depicted in Fig. 3. We now require $P_2(T_P) = P_1(T_P)$, i.e.

$$t_P = \tilde{t}_P \tag{19}$$

and replace t_P in Eq. 14 by the expression for $\tilde{t}_P$ given in Eq. 18. With this we calculate the difference,

$$\Delta_0 := P_2(T_F, T_P) - P_1(T_F, T_P) \tag{20}$$

and obtain:

$$\Delta_0 = (p_F^* - q_F^*) \, (p_P^* - q_P^*) \, t_F \, \overline{t}_F \tag{21}$$

Hence,

Theorem 3 $\Delta_0 = 0$ iff $(p_F^* = q_F^*)$ or $(p_P^* = q_P^*)$. And $\Delta_0 > 0$ if $(p_F^* > q_F^*)$ and if $(p_P^* > q_P^*)$.

The first part of the theorem says that if either T_F and T_F^* are independent or if T_P^* and T_P are independent, then T_F and T_P remain independent after the 'reduction'. The second part of the theorem says that the conjunction of T_F and T_P is more likely after the reduction if T_F confirms T_F^* and if T_P^* confirms T_P.

Next, let us compare the posterior probabilities of the conjunction of T_F and T_P before and after the reduction. To do so, we calculate the difference,

$$\Delta_1 := P_2(T_F, T_P | E, E_F, E_P) - P_1(T_F, T_P | E, E_F, E_P) \tag{22}$$

and obtain:

$$\Delta_1 = (p_F^* - q_F^*) \, (p_P^* - q_P^*) \, t_F \, \overline{t}_F \cdot \alpha \, \tilde{\Delta}_1, \tag{23}$$

The explicit expression for $\tilde{\Delta}_1$ is given in the Appendix. Eq. 23 then entails the following theorem:

Theorem 4 $\Delta_1 = 0$ if $(p_F^* = q_F^*)$ or $(p_P^* = q_P^*)$.

This result has an intuitive interpretation: If either T_F and T_F^* or T_P^* and T_P are independent, then the flow of confirmation from T_F to T_P (and vice versa) is stopped and the epistemic situation before and after the 'reduction' are the same.

Using the expression for $\tilde{\Delta}_1$, we obtain:

Theorem 5 $\Delta_1 > 0$ if the following three conditions hold: (i) $\beta, \gamma > \delta$, (ii) $0 < x_F, x_P < 1$, and (iii) $(p_F^* - q_F^*) \, (p_P^* - q_P^*) > 0$.

Condition (i) seems natural in the light of inequalities (5) and (6). In fact, it is a rather weak condition which also holds, for example, for Set 2, below. Condition (ii) makes sure that E_F confirms T_F and E_P confirms T_P; we have assumed this throughout. Condition (iii) is our usual condition. Hence, none of these conditions is in any way problematic. Given this, we conclude that the posterior probability of the conjunction of T_F and T_P indeed increases after a reductive relationship is established between the two theories.

Finally, let us compare the degree of confirmation of the conjunction of T_F and T_P before and after the reduction. To do so, we use the *difference measure*.[13] Adapted to our case, the difference measure is defined as follows:

$$d_i := P_i(T_F, T_P | E, E_F, E_P) - P_i(T_F, T_P) \quad \text{for } i = 1, 2 \tag{24}$$

We can now calculate

$$\Delta_2 := d_2 - d_1 \tag{25}$$

and obtain

$$\begin{aligned} \Delta_2 &= \Delta_1 - \Delta_0 \\ &= (p_F^* - q_F^*)(p_P^* - q_P^*) \, t_F \, \bar{t}_F \cdot (\alpha \, \tilde{\Delta}_1 - 1). \end{aligned} \tag{26}$$

Hence,

Theorem 6 $\Delta_2 = 0$ *if* $(p_F^* = q_F^*)$ *or* $(p_P^* = q_P^*)$.

That is, if either T_F and T_F^* or T_P^* and T_P are independent, then the evidence confirms the conjunction of T_F and T_P equally well before and after the reduction.

Comparing Eq. 26 with Eq. 23, we see that, for given values of p_F^*, $q_F^* < p_F^*$, p_P^*, $q_P^* < p_P^*$, and t_F, $\Delta_2 > 0$ entails $\tilde{\Delta}_1 > 1/\alpha$. Notice, however, that the converse does not hold. Hence, it is 'easier' to boost the posterior probability of both theories with a reduction than it is to increase the confirmation of the conjunction of both theories. Especially if $1/\alpha$ is large, the confirmation may not be greater after the reduction.

In order to better understand the properties of Δ_2, we now consider two concrete numerical examples. In both cases, we set $p_F^* = p_P^* = .8$ and $q_F^* = q_P^* = .3$, i.e. we assume that there is a flow of confirmation from the reducing theory to the reduced theory. The first case is then defined as follows:

Set 1: $t_F = .8, \alpha = .8, \beta = .6, \gamma = .4, \delta = .2$.

These assignments are consistent with Eqs. 5, 6 and 19. Here a high value is given to t_F. In general, the assignments of this set exhibit much confidence in the fundamental theory. The *phase diagram* on the left-hand side of Fig. 4 plots the likelihood of the phenomenological theory (x_P) against the likelihood of the fundamental theory (x_F). All points below the phase curve correspond to $\Delta_2 < 0$. All points above the phase curve correspond to $\Delta_2 > 0$.

---

[13] Results may depend on the chosen confirmation measure, see Fitelson (1999). Whether our results do is a question that should be addressed in future research.
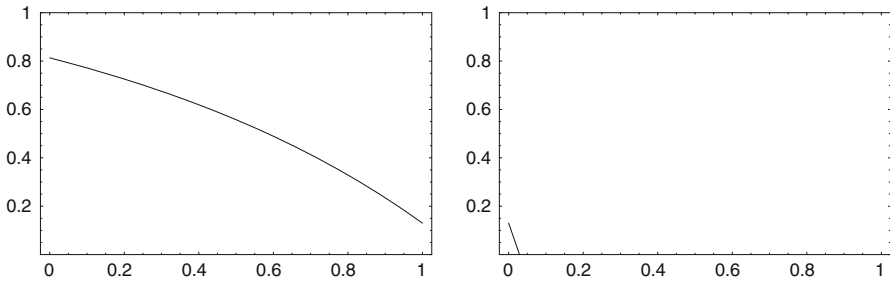
Fig. 4 The phase diagram for Δ_2. Here x_P is plotted as a function of x_F for $p_F^* = p_P^* = .8$ and $q_F^* = q_P^* = .3$ and the parameters specified in Set 1 (*left* figure) and in Set 2 (*right* figure). All points below the phase curve correspond to $\Delta_2 < 0$. All points above the phase curve correspond to $\Delta_2 > 0$

We see that the conjunction of both theories gains from a reduction in three typical cases: (i) If T_F is strongly supported by E_F (i.e. small x_F) and T_P is weakly supported by E_P (i.e. $x_P \gtrsim .8$). (ii) If T_F is weakly supported by E_F (i.e. large x_F) and T_P is strongly supported by E_P (i.e. $x_P \lesssim .2$). (iii) If both theories are not particularly well confirmed by E_F and E_P respectively (i.e. $x_F, x_P \gtrsim .5$). If $x_F + x_P \lesssim 1$, then the conjunction of both theories is better confirmed before the reduction.

The second case consists in the following assignments:

Set 2: $t_F = .2, \alpha = .5, \beta = .3, \gamma = .6, \delta = .1$.

Again, these assignments are consistent with Eqs. 5, 6 and 19. The phase diagram on the right-hand side of Fig. 4 plots the likelihood of the phenomenological theory (x_P) against the likelihood of the fundamental theory (x_F). All points below the phase curve correspond to $\Delta_2 < 0$. All points above the phase curve correspond to $\Delta_2 > 0$.

In contrast to Set 1, we now assign a low value to t_F. In general, the assignments of this set exhibit low confidence in the fundamental theory. It turns out that $\Delta_2 > 0$ for almost all values of x_F and x_P. Only if $x_F, x_P \approx 0$, i.e. if T_F is strongly supported by E_F and if T_P is strongly supported by E_P, then the confirmation is higher before the reduction than after the reduction.

To summarize, a GNS reduction between two theories, such as TD and SM, is epistemically advantageous in virtue of the three results above. We have shown that a reduction makes sure that evidence which, prior to reduction, only supported one of the theories, due to the reduction comes to support the other theory as well. Moreover, a successful reduction increases both the prior and the posterior probabilities of the conjunction of both theories. And in many cases, a reduction results in the conjunction of both theories being better confirmed.

6 Conclusion

We have discussed how the Generalized Nagel-Schaffner model of reduction impacts on the confirmation of theories by evidence. We formulated criteria that help us assess proposed reductions epistemically, and we have shown how a reduction facilitates the flow of confirmation from the reducing theory to the reduced theory and back. Our

Bayesian account also shows to what extent the various judgements depend on the probabilistic judgements of the scientists, connecting—or so we argue—our account to the practice of science. Disagreement amongst scientists can be traced back to disagreement about the assignment of prior probabilities and likelihoods.

We would like to end this paper with proposals for several follow-up projects: First, it would be interesting to compare the coherence before and after the reduction.[14] Here, one might want to focus on the two theories in question, or on the conjunction of the theories and all available evidence. It might be reasonable to focus on the latter, as the evidence is also uncertain and one might, in the end, be interested in the 'package' as a whole, comprising all available theories and all available evidence. Should coherence considerations play a role when it comes to decide whether a theory should be accepted?

Second, as already mentioned, several other confirmation measures have to be checked and the stability of our results have to be explored.

Third, one may want to examine the situation where evidence for, say, the fundamental theory disconfirms the phenomenological theory.[15] How shall one deal with these situations?

Fourth and finally, other types of intertheoretic relation should be studied from a Bayesian point of view. Here, we are thinking of Batterman's singular limits (Batterman 2002) and Hartmann's 'stories' (1999). But there will surely be other examples. This project requires the collaboration between philosophers of science, who conduct case studies, and formal philosophers, who provide the corresponding Bayesian analysis. It may also be asked which picture about the structure of science as a whole emerges from all this. It seems plausible to find something like a network structure, with more or less connected theories and models, and it might be interesting to discuss the implications of this for the debate about the (dis-)unity of science.

Appendix

Let us start with the situation before the reduction and the Bayesian network represented in Fig. 2. The joint distribution $P_1(T_F, T_P, E, E_F, E_P)$ is given by the expression

---

[14] For a discussion of various coherence measures and of the relation between coherence and truth, see Bovens and Hartmann (2003).

[15] Thanks to Jan Sprenger for bringing up this point.

$$P_1(T_F) \, P_1(T_P) \, P_1(E|T_F, T_P) \, P_1(E_F|T_F) \, P_1(E_P|E_P).$$

Using the methodology described in Bovens and Hartmann (2003, Ch.3), we obtain:

$$P_1(\mathrm{T_F}, \mathrm{E}) = \sum_{T_P, E_F, E_P} P_1(T_F, T_P, E, E_F, E_P)$$
$$= t_F \, (t_P \, \alpha + \bar{t}_P \, \beta) \tag{27}$$

Similarly, we calculate

$$P_1(\mathrm{T_P}, \mathrm{E}) = t_P \, (t_F \, \alpha + \bar{t}_F \, \gamma) \tag{28}$$
$$P_1(\mathrm{E}) = t_F \, (t_P \, \alpha + \bar{t}_P \, \beta) + \bar{t}_F \, (t_P \, \gamma + \bar{t}_P \, \delta) \tag{29}$$
$$= t_P \, (t_F \, \alpha + \bar{t}_F \, \gamma) + \bar{t}_P \, (t_F \, \beta + \bar{t}_F \, \delta) \tag{30}$$

To prove Eq. 5 we note, using the definition of conditional probability, that $P_1(\mathrm{T_P}|\mathrm{E}) > P_1(\mathrm{T_P})$ iff $P_1(\mathrm{T_P}, \mathrm{E}) - P_1(\mathrm{T_P}) \, P_1(\mathrm{E}) > 0$ and obtain using Eqs. 28 and 30

$$P_1(\mathrm{T_P}, \mathrm{E}) - P_1(\mathrm{T_P}) \, P_1(\mathrm{E}) = t_P \, \bar{t}_P \, \left[(\alpha - \beta) \, t_F + (\gamma - \delta) \, \bar{t}_F \right], \tag{31}$$

from which Eq. 5 immediately follows. The proof of Eq. 6 proceeds accordingly using Eqs. 28 and 29.

Next, we calculate the prior probability of the two theories.

$$P_1(\mathrm{T_F}, \mathrm{T_P}) = \sum_{E, E_F, E_P} P_1(T_F, T_P, E, E_F, E_P)$$
$$= P_1(\mathrm{T_F}) \, P_1(\mathrm{T_P}) = t_F \, t_P$$

Similarly, we obtain for the posterior probability $P_1^* := P_1(\mathrm{T_F}, \mathrm{T_P}|\mathrm{E}, \mathrm{E_F}, \mathrm{E_P})$:

$$P_1^* = \frac{P_1(\mathrm{T_F}, \mathrm{T_P}, \mathrm{E}, \mathrm{E_F}, \mathrm{E_P})}{P_1(\mathrm{E}, \mathrm{E_F}, \mathrm{E_P})}$$
$$= \frac{t_F \, t_P \, p_F \, p_P \, \alpha}{t_F \, t_P \, p_F \, p_P \, \alpha + t_F \, \bar{t}_P \, p_F \, q_P \, \beta + \bar{t}_F \, t_P \, q_F \, p_P \, \gamma + \bar{t}_F \, \bar{t}_P \, q_F \, q_P \, \delta}$$
$$= \frac{t_F \, t_P \, \alpha}{t_F \, (t_P \, \alpha + \bar{t}_P \, x_P \, \beta) + \bar{t}_F \, x_F \, (t_P \, \gamma + \bar{t}_P \, x_P \, \delta)}, \tag{32}$$

with the likelihood ratios $x_F := q_F/p_F$ and $x_P := q_P/p_P$.

Let us now turn to the situation after the reduction and the Bayesian network represented in Fig. 3. The joint distribution $P_2(T_F, T_P, T_F^*, T_P^*, E, E_F, E_P)$ is given by

$$P_2(T_F) \, P_2(E|T_F, T_P) \, P_2(E_F|T_F) \, P_2(E_P|E_P) \, P_2(T_P|T_P^*) \, P_2(T_P^*|T_F^*) \, P_2(T_F^*|T_F).$$

To simplify our notation, we introduce the following abbreviations:

$$\varphi_\alpha := p_F^* \, p_P^* + \overline{p}_F^* \, q_P^*, \quad \varphi_\beta := p_F^* \, \overline{p}_P^* + \overline{p}_F^* \, \overline{q}_P^*$$
$$\varphi_\gamma := q_F^* \, p_P^* + \overline{q}_F^* \, q_P^*, \quad \varphi_\delta := q_F^* \, \overline{p}_P^* + \overline{q}_F^* \, \overline{q}_P^*$$

For later use, we note that $0 < \varphi_\alpha, \varphi_\beta, \varphi_\gamma, \varphi_\delta < 1$ and

$$\varphi_\alpha - \varphi_\gamma = \varphi_\delta - \varphi_\beta = (p_F^* - q_F^*) \, (p_P^* - q_P^*) \tag{33}$$

$$\varphi_\alpha + \varphi_\beta = \varphi_\gamma + \varphi_\delta = 1. \tag{34}$$

We then obtain for the prior probability of the conjunction of both theories after the reduction

$$P_2(\text{T}_\text{F}, \text{T}_\text{P}) = t_F \, \varphi_\alpha . \tag{35}$$

For the posterior $P_2^* := P_2(\text{T}_\text{F}, \text{T}_\text{P}|\text{E}, \text{E}_\text{F}, \text{E}_\text{P})$, we obtain:

$$P_2^* = \frac{t_F \, \alpha \, \varphi_\alpha}{t_F \left(\alpha \, \varphi_\alpha + x_P \, \beta \, \varphi_\beta \right) + \overline{t}_F \, x_F \left(\gamma \, \varphi_\gamma + x_P \, \delta \, \varphi_\delta \right)} \tag{36}$$

Similarly, we calculate

$$P_2(\text{T}_\text{P}) = t_F \, \varphi_\alpha + \overline{t}_F \, \varphi_\gamma \tag{37}$$

$$P_2(\text{T}_\text{P}|\text{E}_\text{F}) = \frac{t_F \, \varphi_\alpha + \overline{t}_F \, x_F \, \varphi_\gamma}{t_F + \overline{t}_F \, x_F} \tag{38}$$

$$P_2(\text{T}_\text{F}|\text{E}_\text{P}) = \frac{t_F \left(\varphi_\alpha + x_P \, \varphi_\beta \right)}{t_F \left(\varphi_\alpha + x_P \, \varphi_\beta \right) + \overline{t}_F \left(\varphi_\gamma + x_P \, \varphi_\delta \right)}. \tag{39}$$

We now calculate

$$\begin{aligned}
P_2(\text{T}_\text{P}|\text{E}_\text{F}) - P_2(\text{T}_\text{P}) &= \frac{t_F \, \overline{t}_F \, (\varphi_\alpha - \varphi_\gamma) \, (1 - x_F)}{t_F + \overline{t}_F \, x_F} \\
&= \frac{t_F \, \overline{t}_F \, (p_F - q_F) \, (p_F^* - q_F^*) \, (p_P^* - q_P^*)}{p_F \left(t_F + \overline{t}_F \, x_F \right)}.
\end{aligned}$$

This proves theorem 1. Similarly, we calculate

$$\begin{aligned}
P_2(\text{T}_\text{F}|\text{E}_\text{P}) - P_2(\text{T}_\text{F}) &= \frac{t_F \, \overline{t}_F \, (\varphi_\alpha - \varphi_\gamma) \, (1 - x_P)}{t_F \left(\varphi_\alpha + x_P \, \varphi_\beta \right) + \overline{t}_F \left(\varphi_\gamma + x_P \, \varphi_\delta \right)} \\
&= \frac{t_F \, \overline{t}_F \, (p_P - q_P) \, (p_F^* - q_F^*) \, (p_P^* - q_P^*)}{p_P \left[t_F \left(\varphi_\alpha + x_P \, \varphi_\beta \right) + \overline{t}_F \left(\varphi_\gamma + x_P \, \varphi_\delta \right) \right]},
\end{aligned}$$

which proves theorem 2.

To proof Eq. 21, we note that, using Eq. 35

$$\Delta_0 = (\varphi_\alpha - t_P) \, t_F.$$

We now use Eqs. 19 and 37 and obtain

$$\Delta_0 = (\varphi_\alpha - t_F \, \varphi_\alpha - \bar{t}_F \, \varphi_\gamma) \, t_F$$
$$= (\varphi_\alpha - \varphi_\gamma) \, t_F \, \bar{t}_F.$$

Eq. 21 then follows using Eq. 33.

Let us finally calculate Δ_1 using Eqs. 32 and 36. We obtain

$$\Delta_1 = (\varphi_\alpha - \varphi_\gamma) \, t_F \, \bar{t}_F \cdot \alpha \, \tilde{\Delta}_1, \tag{40}$$

with

$$\tilde{\Delta}_1 = N_1^{-1} \, N_2^{-1} \cdot \tilde{\Delta}'_1 \tag{41}$$

and

$$N_1 = t_F \, (t_P \, \alpha + \bar{t}_P \, x_P \, \beta) + \bar{t}_F \, x_F \, (t_P \, \gamma + \bar{t}_P \, x_P \, \delta)$$
$$N_2 = t_F \, \big(\alpha \, \varphi_\alpha + x_P \, \beta \, \varphi_\beta\big) + \bar{t}_F \, x_F \, \big(\gamma \, \varphi_\gamma + x_P \, \delta \, \varphi_\delta\big).$$

Note that $N_1, N_2 > 0$. We are therefore most interested in $\tilde{\Delta}'_1$, which is given by

$$\tilde{\Delta}'_1 = t_F \, x_F \, (\varphi_\alpha - \varphi_\gamma) \, (\gamma - \delta \, x_P) + t_F \, x_P \, (\beta - \delta \, x_F)$$
$$+ \gamma \, \varphi_\gamma \, x_F + \delta \, \overline{\varphi}_\gamma \, x_F \, x_P.$$

From conditions (i) and (ii) of theorem 5, we conclude that $\gamma > \delta \, x_P$ and $\beta > \delta \, x_F$. Hence $\tilde{\Delta}'_1 > 0$, which proves the theorem.

References

Batterman, R. W. (2002). *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence.* Oxford: Oxford University Press.

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology.* Oxford: Oxford University Press.

Callender, C. (2001). Taking thermodynamics too seriously. *Studies in the History and Philosophy of Modern Physics, 32,* 539–553.

Dizadji-Bahmani, F., Frigg, R., & Hartmann, S. (2010). Who is afraid of Nagelian reduction? *Erkenntnis* (forthcoming).

Earman, J. (1992). Bayes or bust? Cambridge, MA: The MIT Press.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science, 66,* S362–S378.

Frigg, R. (2008). A field guide to recent work on the foundations of statistical mechanics. In D. Rickles (Ed.), *The Ashgate companion to contemporary philosophy of physics* (pp. 99–196). London: Ashgate.

Hájek, A., & Hartmann, S. (2010). Bayesian epistemology. In J. Dancy (Ed.), *A companion to epistemology* (pp. 93–106). Oxford: Blackwell.

Hartmann, S. (1999). Models and stories in Hadron physics. In M. Morgan & M. Morrison (Eds.), *Models as mediators* (pp. 326–346). Cambridge: Cambridge University Press.

Hartmann, S., & Sprenger, J. (2010). Bayesian epistemology, to appear. In S. Bernecker & D. Pritchard (Eds.), *Routledge companion to epistemology*. London: Routledge.

Howson, C., & Urbach, P. (2005). *Scientific reasoning: The Bayesian approach*. La Salle: Open Court.

Klein, C. (2009). Reduction without reductionism: A defence of Nagel on connectability. *Philosophical Quarterly, 59*, 39–53.

Nagel, E. (1961). *The structure of science*. London: Routledge and Keagan Paul.

Neapolitan, R. (2003). *Learning Bayesian networks*. Upper Saddle River, NJ: Prentice Hall.

Needham, P. (2009). Reduction and emergence: A critique of Kim. *Philosophical Studies, 146*, 93–116.

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

Schaffner, K. (1967). Approaches to reduction. *Philosophy of Science, 34*, 137–147.

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction and search*. Cambridge, MA: MIT Press.

Winther, R. (2009). Schaffner's model of theory reduction: Critique and reconstruction. *Philosophy of Science, 76*, 119–142.