Unification and Confirmation*

Ilkka Niiniluoto

Received: 19/10/2014
Final Version: 24/01/2015

BIBLID 0495-4548(2016)31:1p.107-123
DOI: 10.1387/theoria.13084

ABSTRACT: According to the traditional requirement, formulated by William Whewell in his account of the "consilience of inductions" in 1840, a scientific hypothesis should have unifying power in the sense that it explains and predicts several mutually independent phenomena. Variants of this notion of consilience or unification include deductive, inductive, and approximate systematization. Inference from surprising phenomena to their theoretical explanations was called abduction by Charles Peirce. As a unifying theory is independently testable by new kinds of phenomena, it should also receive confirmation from its empirical success. The study of the prospects of probabilistic Bayesianism to motivate this kind of criterion for abductive confirmation is shown to lead to two quite distinct conceptions of unification, linking up and screening off, and in both cases the unifying theory can be seen to receive probabilistic support from empirical phenomena.

Keywords: abduction, confirmation, consilience, explanatory power, systematization, testability, unification.

RESUMEN: De acuerdo con un requisito tradicional, formulado por William Whewell en su explicación de la «consiliencia de las inducciones» en 1840, una hipótesis científica debería tener poder unificador, en el sentido de que explique y prediga varios fenómenos mutuamente independientes. Las variantes de esta noción de consiliencia o unificación incluyen la sistematización deductiva, inductiva y aproximada. Charles Peirce llamó abducción a la inferencia que va de fenómenos sorprendentes hasta sus explicaciones teóricas. Puesto que una teoría unificadora puede contrastarse independientemente a partir de nuevas clases de fenómenos, también debería recibir confirmación a partir de su éxito empírico. Se muestra que el estudio de las perspectivas del bayesianismo probabilístico para motivar este tipo de criterio para la confirmación abductiva conduce a dos concepciones distintas de la unificación, vinculación (*linking up*) y anulación (*screening off*), y en ambos casos puede observarse que la teoría unificadora recibe apoyo probabilístico a partir de fenómenos empíricos.

Palabras clave: abducción, confirmación, consiliencia, poder explicativo, sistematización, contrastación, unificación.

1. *Whewell on Consilience*

It is often required that a scientifically interesting hypothesis has to explain several mutually independent phenomena, thereby showing that these phenomena are effects of a hypothetical common causal principle or a postulated unobservable entity (see Schurz 2008a).This idea is historically related to the classical exposition of the method of hypothesis by William Whewell in *The Philosophy of the Inductive Sciences* in 1840. According to

* I am grateful to Gerhard Schurz for illuminating discussions on unification and to anonymous referees of this journal for useful comments.

Whewell, explanatory scientific hypotheses are discovered by induction to account for phenomena, but they should be testable by new kinds of predictions:

> The hypothesis which we accept ought to explain phenomena which we have observed. But they ought to do more than this: our hypothesis ought to *fortel* phenomena which have not yet been observed; - at least of the same kind as those which the hypothesis was invented to explain. ...But the evidence in favour of our induction is of a much higher and more forcible character when it enables us to explain and determine cases of *a kind different* from those which were contemplated in the formation of our hypothesis. (Whewell 1847, 62-65)

Charles S. Peirce agreed with Whewell that scientific hypotheses, which are first inferred by abduction as explanations of surprising phenomena (*CP* 5.189), should be subjected to the test of experiment by deducing from them "most unlikely experiental predictions" (*CP* 7.182). Also Karl Popper repeated Whewell's view in his requirement that a new theory should be independently testable: "it must have new and testable consequences (preferably consequences of a new kind)" (Popper 1963, 241) (cf. Niiniluoto 1984, 37-38).

Whewell argued further that the strongest sign of scientific progress occurs with "the consilience of inductions", where "an induction, obtained from one class of facts, coincides with an induction, obtained from another class". In consilience, inductions "jump together": two separate generalizations are found to be consequences of the same comprehensive theory. His paradigmatic example was Newton's mechanics which was able to explain Galileo's and Kepler's empirical laws and many other apparently unconnected phenomena like the perturbations of the moon and planets by the sun and the precession of the equinoxes. Whewell illustrated the historical development of sciences by "inductive tables" which look like inverted genealogical trees, with the most general and powerful theory as the trunk – the theory of universal gravitation in the case of astronomy, and the undulatory theory of light in the case of optics. These inductive tables show how the truths of these sciences "tend to simplicity" and "converge to unity".

Whewell regarded consilience as "a test of the theory in which it occurs", or "a criterion of its reality, which has never yet been produced in favour of falsehood". Whewell seems to propose here an optimistic meta-induction from the history of science, claiming that consilience has always brought about true theories. Even though Whewell admitted that incomplete and even erroneous hypotheses "may often be of service to science", his claim about certainty due to consilience is clearly exaggerated. Any fallibilist, who follows Peirce in thinking that all scientific conclusions are uncertain and corrigible (*CP* 1.149), has to admit that theoretical unification in science has led to errors as well. The discovery of relativity theory and quantum theory in the early twentieth century shows that even Newton's long admired theory is not strictly speaking true but at best approximately true or truthlike (cf. Niiniluoto 1999a).

Whewell's notions of explanation and prediction are deductive, so that his inductions are converse deductions: an inductively inferred theory "subordinates" or "colligates" a class of facts as its instances. Peirce's account of abduction, as inference to a theory, allows the theory to give probabilistic explanations as well (see Niiniluoto 1999b). Another relevant extension is the notion of approximate explanation and prediction (cf. Section 5).

Some sceptical anti-realists have suggested against Whewell and Peirce that the success of scientific theories in explanation and prediction does not in any way indicate that these theories are *true* (in the realist sense of correspondence with reality) or that the postulated theoretical entities exists in reality. For example, Bas van Fraassen (1989) proposes to conclude only that successful theories are *empirically adequate* (i.e., all of their observable consequences are true). Larry Laudan (1990), who thinks that truth is a utopian aim of science, is still willing to speak about the empirical support or confirmation of theories, but for him such support of a theory concerns only its *reliability* (i.e., true expectations about its next observable predictions).

It is therefore important to ask four questions: What adequacy conditions does the notion of confirmation satisfy? (see Section 2). How can one measure the strength of the empirical support or confirmation that a theory receives from its successful explanations and predictions? (see Section 3). How does this confirmation depend on the unifying power of the theory? (see Section 4).

2. Inductive and Abductive Confirmation

Carl G. Hempel's important contribution in the 1940s was the idea of analysing the concept of empirical confirmation by qualitative or structural principles (see Hempel 1965). One intuitive notion is to regard confirmation as a weakening of the relation of logical consequence or deductive entailment: according to the *Entailment* condition, if an observation report entails a hypothesis, it also confirms the hypothesis.

(E) If evidence E logically entails hypothesis H, then E confirms H.

As entailment itself is transitive, one may also suggest that confirmation is transmitted to logical consequences. This is the *Special Consequence* principle:

(SC) If evidence E confirms hypothesis H, and K is entailed by H, then E confirms K.

Another idea is related to confirmation by hypothetico-deductive testing, where we check whether the observable consequences of a hypothesis are true or not. In the negative case, the hypothesis is refuted, and in the positive case, the hypothesis is confirmed. Thus, according to the *Converse Entailment* condition,

(CE) If hypothesis H logically entails non-tautological evidence E, then E confirms H.

For example, if theory H achieves *deductive systematization* between empirical statements E and E', i.e., $H \& E \vdash E'$ and not $E \vdash E'$, then $H \vdash (E \rightarrow E')$, so that $E \rightarrow E'$ confirms H by CE (see Hempel, 1965). By the *Converse Consequence* condition, evidence confirming a hypothesis also confirms all logically stronger hypotheses:

(CC) If K logically entails H and E confirms H, then E confirms K.

Hempel further observed that CC and SC are incompatible, since a notion of confirmation satisfying both of them would be trivial in the sense that any statement confirms any other statement, and concluded that CC is not "a sound general condition of adequacy" (*ibid.*, p. 33). The same holds for the pair CE and SC.

Howard Smokler (1968) noted that E and SC are typical in enumerative and eliminative induction, where a generalization receives empirical support from its positive instances. On the other hand, he proposed that CE and CC are satisfied by "abductive inference", where a hypothetical theory is supported by its power to explain surprising phenomena. This is in line with Peirce's famous formulation of *abduction* or the "operation of adopting an explanatory hypothesis":

(PA) The surprising fact E is observed;
But if H were true, E would be a matter of course.
Hence, there is reason to suspect that H is true.

(*CP* 5.189). This schema shows how a hypothetical theory can be "abductively conjectured" if it accounts "for the facts or some of them".

The straightforward definition of "deductive confirmation" (E confirms H iff H entails E) would satisfy Smokler's conditions CE and CC.[1] But these adequacy conditions have been criticized for allowing confirmation too easily, since entailment is monotonic and thus admits arbitrary strengthening of the premise (i.e., if $H \vdash E$, then $H\&K \vdash E$ for any K; hence, if E deductively confirms H, it also deductively confirms H&K for any K). More plausible versions of these principles —call them CE* and CC*— are obtained by replacing deductive entailment $\vdash$ by the stronger condition of deductive explanation. Define *abductive confirmation* by

(AC) E abductively confirms H iff H deductively explains E,

where H is consistent and E is non-tautologous. Then confirmation defined by (AC) satisfies the central principle CE*, but not generally CC*. (See Niiniluoto and Tuomela, 1973, 227).[2] While (AC) is in line with Peirce's account of abduction (PA), it should be generalized to cover cases of inductive-probabilistic explanation as well.

The incompatibility of conditions of CE and SC has an important consequence to the once popular transitivity interpretation of Whewell's consilience of inductions (see Hesse 1974, 141-146; Niiniluoto and Tuomela 1973, 228-230).[3] Assuming that a theory H entails empirical statements or regularities E and E', some authors suggested that direct evidence for E (or E itself) counts via H as indirect evidence for E' (see Kneale 1949, 108; Nagel 1961, 64-65). However, such an argument would apply first CE in concluding that E confirms H and then SC in concluding that E confirms E'. As CE and SC cannot be satisfied at the same time, this kind of argument presupposes wrong kinds of transitivity properties for the notion of confirmation.

---

[1] The concept of conditional deductive confirmation of H by E can be defined by the requirement that there is an observational statement C such that H achieves deductive systematization between C and E, i.e., $H\&C \vdash E$. To avoid the trivial choice of C as $H \rightarrow E$, one should require that C is logically independent of H (see Kuipers, 2000, 36).

[2] The concept of deductive explanation is much stronger than merely deductive relations between statements, as one should rule out self-explanations and demand that general premises are lawlike. Further, the explanans should be ontologically or causally prior to the explanandum. Similarly, inductive-probabilistic explanations presuppose laws with objective probabilities or propensities (see Fetzer, 1981).

[3] We have seen in Section 1 that this was not Whewell's own interpretation, as consilience for him provided an argument in favour of the unifying theory H. See also Section 4.

3. *Bayesian Confirmation*

The *Bayesian* approach analyses inferences in terms of epistemic probabilities which express coherent degrees of belief of rational agents. Bayes's Theorem

$$(B)\quad P(H/E) = \frac{P(H)P(E/H)}{P(E)}$$

tells how prior probabilities P(H) of hypotheses are transformed to posterior probabilities P(H/E) via the likelihood P(E/H). While for subjectivists such probabilities are personal degrees of belief, systems of inductive logic include more objective ways of allocating probabilities to statements expressible within a language (for an overview, see Gabbay *et al.* 2011). The Bayesians can also employ objective physical probabilities especially in their treatment of likelihoods P(E/H) and probabilistic laws.

If confirmation is explicated by the *High Probability* criterion (E confirms H iff P(H/E) is sufficiently high, i.e. larger than some threshold q above .5), then conditions of Entailment E and Special Consequence SC are satisfied. This follows from the facts that $E \vdash H$ implies $P(H/E) = 1$ and $H \vdash K$ implies $P(H/E) \leq P(K/E)$. On the other hand, Converse Entailment CE and Converse Consequence CC are not satisfied. This notion of HP-confirmation thus fits the idea of enumerative inductive inference.

Another definition of incremental confirmation is the *Positive Relevance* criterion (E confirms H iff $P(H/E) > P(H)$; E disconfirms H iff $P(H/E) < P(H)$). Here PR-confirmation means that evidence E increases the rational degree of belief in the truth of hypothesis H. Now the principle CE receives immediately a justification by Bayes's Theorem (B):

(1) If H logically entails E, where $P(H) > 0$ and $P(E) < 1$, then $P(H/E) > P(H)$.

This result is completely general in the sense that it is valid for all epistemic probability measures P, which satisfy the probability axioms, and for all non-zero prior probabilities P(H). On the other hand, Positive Relevance does not generally satisfy the controversial principle CC, since P(K/E) may be small and even zero in the case where $K \vdash H$ and $P(H/E) > P(H)$.Still, one might defend the modified principle CC* for cases where explanation is transitive (see Niiniluoto, 2007).

As positive relevance is a symmetric relation, $P(H/E) > P(H)$ holds if and only if $P(E/H) > P(E)$, and this holds if and only if $P(E/H) > P(E/\neg H)$. Therefore, it is sufficient for the PR-confirmation of H by E that H is positively relevant to E. For example, if an infection H increases the probability of fever E, then the fever supports the hypothesis of infection. This allows us to generalize (1) from universal theories H with deductive consequences to cases of probabilistic theories with inductive consequences.

If *inductive explanation* is defined by the positive relevance condition, i.e., by requiring that the explanatory theory H increases the probability of data E (see Niiniluoto and Tuomela 1973), then we have the general result:

(2) If hypothesis H deductively or inductively explains evidence E, then E PR-confirms H.

This is a generalized form of the condition CE* for the relation of explanation and confirmation. Indeed, if the second premise of Peirce's schema (PA) is understood to require that theory H deductively or inductively explains E, then (PA) can be interpreted as stating the principle (2) of abductive confirmation.The same result about indirect confirmation holds of course for successful predictions E from a hypothetical theory H. The strength of the quantitative Bayesian treatment is thus its ability to account by (1) and (2) for the most central qualitative principles CE and CE* of abductive confirmation and their extension to cover probabilistic settings.

The basic result (2) can be refined by introducing quantitative degrees of explanatory power and degrees of confirmation, with the aim of showing comparatively that *better* explanations receive *stronger* confirmation (see Niiniluoto 1999b). The first measure of *explanatory power* of theory H with respect to evidence E was defined by Hempel in 1948 as

$$expl_1(H,E) = P(\neg H/\neg E)$$

(see Hempel 1965). Here E may be a conjunction of several empirical statements. Hempel derived this measure as the ratio between the common information content of H and E (i.e., 1-P(HvE)) and the content of E (i.e., 1-P(E)).Another information-theoretic definition, due to Jaakko Hintikka (1968), is

$$expl_2(H,E) = \frac{P(E/H) - P(E)}{1 - P(E)}.$$

Measure $expl_2(H,E)$ is an improvement of the simple ratio measure P(E/H)/P(E) (see McGrew 2003) or its logarithm (see Good 1960), which remain constant when an irrelevant proposition E' is added to the explanandum E (see Schupbach and Sprenger 2011). This criticism does not hit $expl_2$, since $expl_2(H,E\&E') < expl_2(H,E)$ in case $P(E'/E\&H) = P(E'/E) \neq 1$. The proposal of Schupbach and Sprenger (2011) is structurally similar to the Kemeny – Oppenheim measure of "factual support":

$$expl_3(H,E) = \frac{P(H/E) - P(H/\neg E)}{P(H/E) + P(H/\neg E)}.$$

All of these three measures receive the maximal value one if H deductively explains (the whole of) E, so that they can primarily be used for the comparison of non-deductive explanations. Such non-deductive explanations may include partial deductions, where H entails most but not all conjuncts of E, or inductive inferences from a probabilistic theory H. Following Hempel, these formulas are called measures of *systematic power*, when the capacities of a theory to give successful explanations and predictions are combined.

The simplest definition of *degrees of confirmation*, related to the High Probability criterion, is posterior probability:

$$conf_1(H/E) = P(H/E).$$

The difference and ratio measures are related to the Positive Relevance criterion:

$$\text{conf}_2(H/E) = P(H/E) - P(H)$$

$$\text{conf}_3(H/E) = \frac{P(H/E)}{P(H)} = \frac{P(E/H)}{P(E)}$$

$$\text{conf}_4(H/E) = \log P(H/E) - \log P(H).$$

(See Kuipers 2000). Variants of these definitions include I.J. Good's 1950 measure of "weight of evidence" log P(E/H) – log P(E/¬H)(cf. Good, 1960) and J. Kemeny's and P. Oppenheim's 1952 measure for "the degree of factual support" (cf. Foster and Martin 1966). All of these measures satisfy the requirement that confirmation increases when the evidence is more surprising or improbable. Fitelson (1999), who himself favors Good's weight of evidence, notes that an important virtue of the difference measure conf_2(H/E), in comparison to the ratio measure conf_3(H/E), is that evidence gives strongest support to *a minimal explanation*, i.e., only to the part of an explanatory hypothesis H that is relevant to the explanation of the evidence E.

Relating these measures gives results like the following. If H explains E better than K in the sense of expl_2 or expl_3, then conf_3(H/E) > conf_3(H/E) and conf_4(H/E) > conf_4(H/E). If H explains E better than K in the sense of the expected value of expl_1, then conf_2(H/E) > conf_2(K/E) (see Niiniluoto 1999a, 187).

Peircean abduction (PA) is often treated as a rule of inference rather than as a principle of confirmation. In *selective* abductive reasoning, the potential explanations H in (PA) are compared to each other, and the best one is chosen. The measures of explanatory power allow us to define the *best* explanation of evidence E as that hypothesis H which maximizes expl(H,E). Thereby selective abduction can be expressed by an acceptance rule which tentatively recommends *inference to the best explanation*:

(IBE) A hypothesis H may be inferred from evidence E if H is the best explanation of E among all the rival hypotheses.

As a qualification to the rule IBE, one may require that the best explanation should be sufficiently good (see Lipton 2001, 104). Otherwise, suspension of judgment is the most rational option before new evidence is gathered by observation and experimentation.

The results (1) and (2) are compelling also to anti-realists, who like van Fraassen and Laudan admit that theoretical hypotheses have truth values, as soon as they accept the probabilistic Bayesian framework. However, one way of denying that empirical success is truth-conducive is to assume that all scientific theories have zero probability: if P(H) = 0, then P(H/E) = 0 for any evidence E (see van Fraassen, 1989). But this choice of prior probability is dogmatic in the sense that no evidence whatsoever can change it, so that the general recommendation to choose P(H) = 0 for all theories H is an expression of dogmatic theoretical anti-realism.

The concept of *probable approximate truth* allows treatments of cases where the prior probability of a hypothesis H is zero.[4] For example, H may be a sharp hypothesis with measure zero, but still its probable approximate truth may be larger than zero.

[4] For this concept, see Niiniluoto (1987), p. 280. The probability of a hypothetical theory H relative to our background knowledge may be zero also in cases, where H is an idealization with counterfactual presuppositions or has known counterexamples. Such cases can be handled with the concept of truthlikeness.

Peirce was not a Bayesian, as he wanted to analyze induction and abduction by means of probabilities as truth-frequencies instead of degrees of belief. Some philosophers of science, like Stathis Psillos (2009), take abduction and IBE as valid principles which need no probabilistic justification (cf. discussion by Iranzo 2007). Gerhard Schurz (2008a) defends abduction as a mode of inference with some value in justification, but rejects the Bayesian approach as being unable to demarcate scientifically worthwhile hypotheses from pure speculations.

To show that Bayesian incremental confirmation is too easy, Schurz (2008b) considers the God's will hypothesis

(G-E) God wants E, and whatever God wants, happens.

where E is any empirical phenomenon. As G-E deductively entails E, by (1) E PR-confirms G-E, even though G-E is purely speculative and should not receive any scientific confirmation. This argument raises many intricate issues that belong to the philosophy of religion. Some religious thinkers, most notably Richard Swinburne (2004), have applied Bayesian confirmation theory to the hypothesis that God exists. If H is the hypothesis of theism, and E states the existence of a complex physical universe (or the orderliness of the universe, the existence of consciousness, etc.), he argues that $P(E/H) > P(E/\neg H)$, which implies that E PR-confirms H. Further, according to Swinburne, the hypothesis that God has all the supernatural powers (like omniscience, omnipotence, and moral perfection) is simpler and hence initially more probable than any rival hypothesis, which implies that $P(H/E) > ½$, i.e., H receives HP-confirmation from the total evidence E. So the Bayesian approach is indeed flexible enough to reconstruct the thinking of a religious person who sees divine providence everywhere and thereby finds confirmation for his or her faith. Here one may recall that confirmation is a weak epistemic concept which does not yet guarantee that a hypothesis is acceptable as true – especially when a natural explanation as a rival to G-E is available (Niiniluoto 2008). But I think there are good reasons to agree with Schurz that G-E should not receive any *scientific* confirmation. However, this does not require that Bayesianism is rejected, as Schurz pleads, since a Bayesian can block the confirmation of such speculative religious hypothesis by giving them zero prior probability. This move need not be based on dogmatic atheism, but rather with inherent difficulties with a hypothesis like G-E. All theologians would not accept that evil events and miseries happen by God's wishes, so that they would exclude the explanation of such events by G-E. There are famous arguments to show that an unrestricted notion of omnipotence is inconsistent - for example, could God create a stone so heavy that even He could not lift it? (See Grim 2006.) Already medieval scholastics who accepted divine foreknowledge attempted to avoid fatalism by denying that God's knowledge *causes* future events. Further, a Bayesian, who takes seriously the idea of abductive confirmation, might accept that G-E entails E, but deny that G-E *explains* E, since it includes an *ad hoc* assumption that God wants E – and hence there is no genuine confirmation of G-E in the sense of (AC).

4. *The Virtue of Unification*

We are now ready to consider whether the Bayesian approach is able to justify the criterion of theoretical or causal unification. In other words, is it possible to prove that a hypothesis which explains many different kinds of phenomena has stronger confirmation than a hypothesis which explains only one kind of phenomenon? Can the Whewellian idea of "consilience of inductions" be justified in Bayesian terms? Or are there other interesting notions of unification which as theoretical virtues are rewarded by empirical confirmation? The answers to all of these questions turn out to be affirmative.

One natural proposal is to regard a theoretical explanation the better, the more empirical phenomena it explains, and the less new entities or principles it postulates. There is a trade-off between these desiderata, as new consequences may be achieved by adding assumptions and thereby making the theory more complex.This idea was formulated already by Eino Kaila in his Finnish monograph *Inhimillinen tieto* in 1939 with his notion of "relative simplicity".[5] According to Kaila, a theory has scientific value only to the extent that the multiplicity of its explanatory principles is smaller than the multiplicity of facts of experience that can be derived from it. Thus, the *relative simplicity* of theory H is the ratio between the multitude of empirical data E derivable from H and the number of logically independent basic assumptions of H. In modern terms, this definition can be stated as the ratio between the explanatory power of H and the complexity of H. Kaila, who had written in 1926 a monograph on *Wahrscheinlichkeitslogik* but had thereafter become sceptical about the possibility of quantitative inductive logic, further suggested that this ratio *would* be equal or at least proportional to the inductive probability of H given the data E, *if* it were measurable. His guess was not quite correct: when Hempel defined in 1948 the first formal measure of explanatory or systematic power of a theory H relative to evidence E, its value turned out to be $P(\neg H/\neg E)$ rather than $P(H/E)$.[6] But Kaila's important point was that relative simplicity in some way indicates the truth or epistemic acceptability of a theory, and is not only a pragmatic or conventional virtue in theory preference. Thus, his notion of relative simplicity can be viewed as a generalization of Whewell's consilience of inductions.

For example, Newton's mechanics has high relative simplicity in Kaila's sense, since the law of gravitation explains many different kinds of phenomena (such as orbits of planets, free fall near the Earth, pendulum, etc.). On the other hand, the religious hypothesis G-E would have minimal relative simplicity in Kaila's sense, since for each empirical statement E it requires the *ad hoc* premise that God wants E. Kitcher (1989) says that G-E achieves "spurious unification", but in fact it would be better to say that it does not yield any kind of unification.

One problem with measures of explanatory power like $expl_1(H,E)$ and $expl_2(H,E)$ is that they receive their maximal value one when H is a contradiction – recall that a contradiction entails all statements.[7] This a reason why Laudan (1977) proposes that the problem-solving ability of a theory H should be measured by the weighted number of solved empirical problems (which corresponds to $expl_1(H,E)$) subtracted with the number of "conceptual problems" (such as inconsistency). Laudan's proposal is clearly a variant of Kaila's notion of relative simplicity (see Niiniluoto 1999a, 167, 182).

[5] Kaila's early treatment of unification was mentioned by Niiniluoto (1999a, 182), but it has remained largely unknown among philosophers of science, since its English translation had to be waited for 75 years (see Kaila 2014).

[6] See measure $expl_1(H,E)$ in Section 3, cf. Hempel (1965).

[7] Note that $expl_3(H,E)$ is undefined, when H is a contradiction.

Kaila applied his ratio measure n/m to the case of curve-fitting, where a curve runs through n observation points and can be determined by m points (e.g. m = 2 for a straight line, m = 3 for a circle, etc.). But in the general case Kaila did not have the logical tools to measure the complexity of a theory. As the difficulties in the explication of the notion of complexity show (cf. Foster and Martin 1966), it is not easy to give a precise account of counting the number of the "logically independent basic assumptions" of a theory.[8] An attempt in this direction was made by Michael Friedman (1974) in his notion of a partition of a sentence or a set of sentences into its "atomic" parts, but his proposal was refuted by Philip Kitcher (1977). Kitcher suggested that instead of counting the number of independent laws one should consider "the repeated use of a small number of *types* of laws which relate a large class of apparently diverse phenomena". This idea was worked out in Kitcher (1981): a theory achieves "unifying power" by "generating a large number of accepted sentences as the conclusions of acceptable arguments which instantiate a few stringent patterns".

A more sophisticated way of analysing a belief system into a conjunction of its "relevant elements" by using the machinery of relevance logic is given by Gerhard Schurz and Karel Lambert (1994), who allow deductive, approximate and inductive unification in their account of scientific understanding. Their treatment avoids Friedman's (1974) problem that the conjunction of two separate theories (e.g. Kepler's laws and Boyle's law) should not count as unification. Success of unification, which "goes hand in hand" with success in realistic truth approximation, increases with the weighted number of "actually and potentially assimilated phenomena" and decreases with the number of "basic and dissimilated phenomena" (see also Schurz 1999). Again we have a variant of Kaila's ratio measure of relative simplicity.[9]

Another way of analysing unified theories in terms of structures rather than statements is given by the structuralist school (see Balzer, Moulines, and Sneed 1987). Theories are represented as nets of theory-elements <M,I>, where M is a class of structures and I is a set of intended applications. Such theory-nets linked with a specialization relation look very much like Whewell's inductive tables. But the difference to Whewell's consilience and other realist accounts of unification is the instrumentalist spirit of this approach, as it denies that theories have truth values and thus could be confirmed by their empirical success.

Let us return to the question of independent testability. With the formal machinery of Bayesian epistemic probabilities, one can give a simple and straightforward argument about the Whewellian situation where theory H deductively explains two independent phenomena E and E' (see Niiniluoto 2008).

Let E_1, E_2, ..., E_n be repeated occurrences of the phenomenon E, and let E(n) be their conjunction. Suppose that H logically entails E(n), so that P(E(n))/H) = 1. Then by (1) evidence E(n) PR-confirms H. But repetitions of the same kind of evidence give diminishing returns, since by inductive learning the (n+1)st occurrence of E is more probable than its earlier occurrences: $P(E_{n+1}/E(n)) > P(E_{n+1}) = P(E_1)$ (see Howson and Urbach 1989, 82).

8 This is also illustrated by the debate of Myrwold (2003) and Schurz (2008b) on the comparison of the Ptolemaic and Copernican theories in astronomy.

9 To give one more variant, Dean Peters (2014), who defends "selective scientific realism", formulates his "unification criterion of confirmation" by the requirement that a theoretical posit should be regarded as (approximately) true if it entails more accepted empirical propositions than are required to construct it.

Now suppose further that H logically entails another phenomenon E' which is probabilistically independent of E, i.e., P(E'/H) = 1 and P(E'/E) = P(E'). It follows from these conditions that $P(H/E(n)\&E_{n+1}) < P(H/E(n)\&E')$ iff $P(E') < P(E_{n+1}/E(n))$. The last condition holds under quite broad conditions, e.g., if the initial probabilities P(E) and P(E') are about the same magnitude. Hence, relative to E(n), new kind of evidence E' gives more confirmation to H than old kind of evidence E_{n+1}. This argument, which is valid for all confirmation measures from $conf_1$ to $conf_4$, thus proves that successful explanation of the new kind of phenomenon E' gives more confirmation to theory H than the repetition of the old kind of phenomenon E.

Another important result about unification is proved by Wayne C. Myrwold (2003). Suppose that theory H achieves *inductive systematization* between empirical propositions E and E' (cf. Niiniluoto and Tuomela 1973). This means that E and E' are probabilistically independent, but they are informationally relevant to each other given H:

(3) P(E'/E) = P(E') and 1 > P(E'/E&H) > P(E'/H).[10]

Myrwold proposes to measure the *degree of unification* of E and E' achieved by H by means of the difference

$$(4)\quad U(E,E';H) = \log\frac{P(E'/E\&H)}{P(E'/H)} - \log\frac{P(E'/E)}{P(E')}.$$

It follows that U(E,E';H) > 0 if H achieves inductive systematization between E and E'. Similarly, U(E,E';H) > 0 if H achieves deductive systematization between independent E and E' and H alone does not entail E', since then we have P(E'/E) = P(E') and 1 = P(E'/E&H) > P(E'/H) (cp. (3)).Then Myrwold argues, by applying the logarithmic ratio measure of confirmation $conf_4$, that the degree of confirmation of H by E&E' can be divided into three additive parts: confirmation of H by E, confirmation of H by E', and the degree of unification of E and E' achieved by H:

$$(5)\quad conf_4(H/E\&E') = conf_4(H/E) + conf_4(H/E') + U(E,E';H).$$

Myrworld notes that a similar result about the "virtue of unification" can be proved if $conf_4$ is replaced with Good's weight of evidence.

[10] As an alternative to (3), one may require that H together with E is positively relevant to E'. i.e., 1 > P(E'/H&E) > P(E'). These notions of inductive systematization avoid the transitivity paradox mentioned at the end of Section 2. They are motivated by Hempel's suggestion in 1958 that theoretical concepts could be logically indispensable for inductive systematization. Hempel had noted that deductive systematization by a theory H (i.e., H&E'⊢E and not E'⊢E) can always be achieved by an observational subtheory of H. As a way out of this "theoretician's dilemma", he proposed that inductive systematization might behave differently. Niiniluoto and Tuomela (1973) prove that Hempel's guess was right: theoretical concepts may be logically indispensable for inductive systematization. Niiniluoto (1973) argues that a theory, which does not have any non-tautological deductive observational consequences and therefore has a Ramsey-sentence which is logically true in second-order logic, may still achieve inductive systematization among observational statements. Raatikainen (2012) repeats this argument, and uses it against the thesis of some structural realists that Ramsey sentences exhaust the cognitive content of theories.

Essentially the same result is proved by Timothy McGrew (2003), p. 562. Assume that H_1 and H_2 are equally good explanations of E and E', i.e., $P(H_1) = P(H_2)$, $P(E/H_1) = P(E/H_2)$, and $P(E'/H_1) = P(E'/H_2)$. Assume further that H_1 achieves inductive systematization between E and E' in the sense of (3), but E and E' are independent conditional on H_2, i.e., $P(E/E'\&H_2) = P(E/H_2)$. Then a direct calculation by Bayes's Theorem shows that $P(H_1/E\&E') > P(H_2/E\&E')$. Hence, by all of our measures of confirmation, the theory H_1 which unifies E and E' gains more confirmation by evidence E&E' than the theory H_2 without such unifying power. McGrew illustrates this analysis by the example, where two quasar images are observed to have identical spectrums (E and E'), and this coincidence is explained by the lensing hypothesis (H) that the gravitational field of a massive object has bent the radiation stemming from a single quasar; in this case, E and E' become dependent relative to H, i.e. P(E&E'/H) is much larger than P(E/H) P(E'/H).

These results are related to the issue whether "coherence" is truth conducive. The intuition that a coherent set of beliefs has a high probability of being true may arise from the controversial coherence theory of truth. Some Bayesians have given impossibility proofs against the proposal that coherence of evidence would increase the support for the hypothesis (see Bovens and Hartmann 2003; Olsson 2005). Tomoji Shogenji (2013) argues convincingly that some of the expectations of Bayesian coherentists have been wrong-headed, since it is the *diversity* of evidence that strengthens the support for the hypothesis. Shogenji's (1999) own measure of coherence for statements $E_1, \ldots, E_n$ is defined by

$$Coh(E_1, \ldots, E_n) = \frac{P(E_1 \,\&\, \ldots \,\&\, E_n)}{P(E_1) \ldots P(E_n)},$$

and the conditional coherence of $E_1, \ldots, E_n$ given H is obtained by relativizing these probabilities to H:

$$Coh(E_1, \ldots, E_n/H) = \frac{P(E_1 \,\&\, \ldots \,\&\, E_n/H)}{P(E_1/H) \ldots P(E_n/H)}.$$

Now, applying the ratio measure of confirmation $conf_3$, we have

$$conf_3(H/E_1\&\ldots\&E_n) = \frac{conf_3(E_1/H) \ldots conf_3(E_n/H) Coh(E_1, \ldots, E_n/H)}{Coh(E_1, \ldots, E_n)}.$$

But this equation can be rewritten as

$$(6)\quad conf_3(H/E_1\&\ldots\&E_n) = \frac{conf_3(H/E_1) \ldots conf_3(H/E_n) Coh(E_1, \ldots, E_n/H)}{Coh(E_1, \ldots, E_n)},$$

which shows that the confirmation H by $E_1, \ldots, E_n$ can be expressed as the product of the confirmation of H by each E_i, $i = 1, \ldots n$, and the increase of the coherence of $E_1, \ldots, E_n$ by H. By taking logarithm of both sides of (6), this equation reduces to Myrwold's result (5). In fact, the result (6) can be found already in Myrwold's earlier article (see Myrwold, 1999, 663).

Marc Lange (2004) has presented objections to Myrwold's treatment of unification (see also Schurz, 2008b). With examples from physics and cosmology, Lange argues that Myrwold's unification is in some cases "too easy" to achieve. It is correct that formulas like (4) and (5) cannot reflect the question whether the unifying theory H is lawlike and explanatory or ontologically and causally prior to statements E and E', and one could develop a more restricted notion of unification with these additional conditions (cf. (1) and (2)). But this does not yet discredit the Bayesian analysis of the virtue of unification.

Lange also points out that, for any independent E and E', the simple conditional $E \rightarrow E'$ unifies E and E' in Myrwold's sense: $U(E,E';E \rightarrow E') > 0$, since $P(E'/E\&(E \rightarrow E')) = 1$ and $P(E'/(E \rightarrow E'))$ is smaller than 1. To avoid this technical objection, one can require that the unifying theory H should be logically independent of E' and E. This is in fact required in the definition (3) of inductive systematization.[11]

Myrwold (2003), p. 410, notes that on his account the conjunction E & E' of two independent statements E and E' does not unify them: $U(E,E';E\&E') = 0$, since $P(E/E\&E') = P(E'/E\&E') = 1$. It is appropriate that this kind of trivial unification is excluded. But it may be somewhat surprising that, for similar reasons, Whewell's consilience of inductions is excluded as well (see *ibid*., p. 418): if theory H entails both E and E', which are independent of each other, then by (4) the degree of unification $U(E,E';H)$ is zero, since $P(E'/E\&H) = P(E'/H) = 1$.[12]

Deductive consilience by Newton's theory does not show that e.g. Kepler's and Galileo's empirical laws are rendered informationally relevant to each other. On the contrary, these laws K and G are probabilistically irrelevant to each other given the unifying theory N, since $P(G/N) = P(G/K\&N) = P(K/N) = 1$.

A different probabilistic example with a similar conclusion is given by Lange (2004). Let H be the hypothesis that Jones has a disease called systemic lupus erythematosus, and let E state that Jones has pleuritis and E' that Jones has a malar rash. Lange specifies the probabilities so that pleuritis E and malar rush E' are positively relevant to each other, but given lupus H, these symptoms E and E' are independent of each other. Still, E and E' are separately positively relevant to H and also jointly positively relevant to H, so that E&E' strongly PR-confirms H. Indeed, using terminology from Hans Reichenbach's account of probabilistic causality, lupus is a *common cause* which "screens off" its probabilistic effects from each other. More generally, variable Z *screens off* variable X from variable Y iff X and Y are probabilistically dependent but become independent when conditionalized on Z (see Schurz 2008a, 344). Myrwold's $U(X,Y;Z)$ has a negative value in such cases. But, as we just saw, even in these cases a Bayesian treatment of the PR-confirmation of the common cause is available.

The outcome of our discussion is that there are two quite different concepts of unification, both of them legitimate with important applications in science. One of them is *link-*

[11] See also footnote 1 on conditional deductive confirmation.

[12] Myrwold (2003) proposes to account for consilience by the idea that, in a law with free parameters, the values determined from one class of phenomena agree with those determined from another class. This is systematically illustrated in William Harper's (2011) treatise on Newton's method.

ing up[13]: two empirical phenomena are independent from each other, but given a theory they become positively relevant to each other. This idea covers the notions of deductive and inductive systematization and Myrwold's measure of unification.

The second is *screening off*: two empirical phenomena or variables are positively relevant or even indifferent to each other, but given a theory they become independent of each other.[14] This idea covers Whewell's deductive account of consilience, Kaila's relative simplicity, and both deterministic and probabilistic common causes.

In both cases, the unifying theory or common cause receives probabilistic confirmation from empirical phenomena. This shows that Bayesianism can give a successful account of abductive confirmation both for linking up and screening off.

5. *Concluding remark on approximate unification*

The empirical success of scientific theories in explanation and prediction is often approximate: theory H entails a statement which is only "close" to the observed evidence E.[15] In such cases, H and E are strictly speaking incompatible, so that P(E/H) = 0. For the same reason, probabilistic measures of conditional coherence like $Coh(E_1, ..., E_n/H)$ and degree of unification like U(E,E';H) are not well-defined in the case of *approximate unification*. New conceptual tools are therefore needed, if we wish to extend the Bayesian account of abductive confirmation to cases of approximate unification.

But do theories receive abductive confirmation from their approximate empirical success? Ordinary Bayesian probabilities cannot handle this question. Theo Kuipers (2000) gives a comparative methodological treatment, based on his "Success Theorem": if theory H is at least as close to complete truth as theory H', then H is at least as empirically successful as H' relative to correct data. This justifies a "Rule of Success": if H has so far been empirically more successful than H', then eliminate H' in favor of H.

Another approach is to replace the posterior probability P(H/E) with the notion of *expected truthlikeness* ver (H/E) of a theory H given empirical evidence E (see Niiniluoto, 1987). This notion differs from epistemic probability P(H/E), since ver(H/E) can be high even in cases where H and E are in conflict with each other. Then the notion of confirmation as increase of probability can be replaced with the notion of *ver-confirmation* or in-

---

13 Schurz (2015) uses this term in his treatment of unification by means of causal nets. (The paper was presented in the symposium on coherence and unification in Düsseldorf in January 2014.) If X, Y, and Z are random variables, then linking up means that X and Y are statistically independent but dependent relative to Z. For example, if X is the angle of the sun, Y is the length of a tower, and Z is the length of its shadow, then INDEP(X,Y) but for a fixed value of Z the values of X and Y determine each other.

14 In the notation of Schurz (2015), X and Y are screened off by Z iff DEP(X,Y) and INDEP(X,Y/Z). Here Z may be a common cause of X and Y (e.g. X is barometer reading, Y is coming storm, and Z is atmospheric pressure) or Z may an intermediate cause of X and Y (e.g. X is a light switch, Y is a light bulb, and Z is electric current). Schurz notes that deterministic common causes "trivially" screen off their effects.

15 For illustrations, see Popper (1963), p. 62; Hempel (1965), p. 344; Hintikka (1992); and Niiniluoto (2014).

crease of the expected truthlikeness of H by E. At the same time, abduction (PA) or IBE should be generalized to a rule which recommends us to regard the best available explanation as *truthlike* rather than true:

(IBT) If hypothesis H is the best explanation of evidence E, conclude for the time being that H is truthlike.

(see Kuipers, 2000; Niiniluoto, 2005). IBT allows approximate explanations, and the strength of instances of this pattern of argument can be measured by the value ver (H/E).

To generalize the results of Section 4, one should then show that the degree of approximate unification of theory H with respect to evidence increases our expectation that H is truthlike. This task is left to another paper.

REFERENCES

Balzer, W.; Moulines, C.U. & Sneed, J.D. 1987. *An Architechtonic for Science: The Structuralist Program*. Dordrecht: D. Reidel.

Bovens, L. and Hartmann, S. 2003. *Bayesian Epistemology*. Oxford: Oxford University Press.

Fetzer, J.K. 1981. *Scientific Knowledge: Causation, Explanation, and Corroboration*. Dordrecht: D. Reidel.

Fitelson, B. 1999. The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity. *Philosophy of Science (Proceedings)* 66, S362-S378.

Friedman, M. 1974. Explanation and Scientific Understanding. *The Journal of Philosophy* 71, 5-19.

Foster, M.H. & Martin, M.L. (eds). 1966. *Probability, Confirmation, and Simplicity: Readings in the Philosophy of Inductive Logic*. New York: The Odyssey Press.

Gabbay, D.M.; Hartmann, S. & Woods, J. (eds.). *Inductive Logic: Handbook of the History of Logic*, vol. 10, Amsterdam: North-Holland.

Good, I.J. 1960. Weight of Evidence, Corroboration, Explanatory Power, Information, and the Utility of Experiments. *Journal of the Royal Statistical Society* B22, 319-331.

Grim, P. 2006. Impossibility Arguments. In M. Martin (ed.). *The Cambridge Companion to Atheism*. Cambridge: Cambridge University Press.

Harper, W.L. 2011. *Isaac Newton's Scientific Method*.Oxford: Oxford University Press.

Hempel, C.G. 1965. *Aspects of Scientific Explanation*. New York: The Free Press.

Hesse, M. 1974. *The Structure of Scientific Inference*. London: Macmillan.

Hintikka, J. 1968. The Varieties of Information and Scientific Explanation. In B. van Rootselar and J.F. Staal (eds.), *Logic, Methodology and Philosophy of Science III*. Amsterdam: North-Holland, pp. 151-171.

Hintikka, J. 1992. The Concept of Induction in the Light of the Interrogative Approach to Inquiry. In J. Earman (ed.), *Inference, Explanation, and Other Frustrations*. Berkeley and Los Angeles: University of California Press, pp. 23-43.

Howson, C. & Urbach, P. 1989. *Scientific Reasoning: The Bayesian Approach*. La Salle, IL.: Open Court.

Iranzo, V. 2007. Abduction and Inference to the Best Explanation. *Theoria* 22 (3), 339-346.

Kaila, E. 2014. *Human Knowledge: A Classic Statement of Logical Empiricism*. La Salle, IL.: Open Court.

Kitcher, P. 1977. Explanation, Conjunction, and Unification. *Journal of Philosophy* 73, 207-212.

Kitcher, P. 1981. Explanatory Unification. *Philosophy of Science* 48, 507-531.

Kuipers, T. 2000. *From Instrumentalism to Constructive Realism*. Dordrecht: Kluwer.

Lange, M. 2004. Bayesianism and Unification. *Philosophy of Science* 71, 205-215.

Laudan, L. 1977. *Progress and Its Problems: Toward a Theory of Scientific Growth*. London: Routledge and Kegan Paul.

Laudan.L. 1990. *Science and Relativism: Some Key Controversies in the Philosophy of Science*. Chicago: The University of Chicago Press.

Lipton, P. 2001. Is Explanation a Guide to Inference? A Reply to Wesley Salmon. In G. Hon and S.S. Rakover (eds.). *Explanation: Theoretical Approaches and Applications*. Dordrecht: Kluwer, pp. 93-120.

McGrew, T. 2003. Confirmation, Heuristics, and Explanatory Reasoning. *The British Journal for the Philosophy of Science* 54, 553-567.

Myrwold, W.C. 1996. Bayesianism and Diverse Evidence: A Reply to Andrew Wayne." *Philosophy of Science* 63, 661-665.

Myrwold, W.C. 2003. A Bayesian Account of the Virtue of Unification. *Philosophy of Science* 70, 399-423.

Niiniluoto, I. 1973. Empirically Trivial Theories and Inductive Systematization. In R. J. Bogdan and I. Niiniluoto (eds.). *Logic, Language, and Probability*. Dordrecht: D. Reidel, pp. 108-114.

Niiniluoto, I. 1984. *Is Science Progressive?*. Dordrecht: D. Reidel.

Niiniluoto, I. 1987. *Truthlikeness*. Dordrecht: D. Reidel.

Niiniluoto, I. 1999a. *Critical Scientific Realism*. Oxford: Oxford University Press.

Niiniluoto, I. 1999b. Defending Abduction. *Philosophy of Science (Proceedings)* 66, S436-S451.

Niiniluoto, I. 2005. Abduction and Truthlikeness. In R. Festa, A. Aliseda, and J. Peijnenburg (eds.), *Confirmation, Empirical Progress, and Truth Approximation*. Amsterdam: Rodopi, 255-275.

Niiniluoto, I. 2007. Structural Rules for Abduction. *Theoria* 60, 325-329.

Niiniluoto, I. 2008. Abduction, Unification, and Bayesian Confirmation: Comment on Schurz. In C. Dégremont, L. Keiff & H. Rückert (eds.). *Dialogues, Logics and Other Strange Things*. London: College Publications, pp. 365-370.

Niiniluoto, I. 2014. Approximation of Laws. In G. Guo and C. Liu (eds.), *Scientific Explanation and Methodology of Science*. Singapore: World Scientific, pp. 148-155.

Niiniluoto, I. & Tuomela, R. 1973. *Theoretical Concepts and Hypothetico-Inductive Inference.* Dordrecht: D. Reidel.

Olsson, E. 2005. *Against Coherence: Truth, Probability, and Justification*. Oxford: Oxford University Press.

Peirce, C.S. 1931-35, 1958-60. *Collected Papers* 1-6, ed. by C. Hartshorne and P. Weiss, 7-8, ed. by A. Burks. Cambridge, MA: Harvard University Press.

Peters, D. 2014. What Elements of Successful Scientific Theories are the Correct Targets for 'Selective' Scientific Realism. *Philosophy of Science* 81, 377-397.

Popper, K. 1963. *Conjectures and Refutations*. London: Hutchinson.

Psillos, S. 2009. *Knowing the Structure of Nature: Essays on Realism and Explanation*. Houndmills: Palgrave Macmillan.

Raatikainen, P. 2102. Ramsification and Inductive Systematization. *Synthese* 187, 569-577.

Schupbach, J.N. and Sprenger, J. 2011. The Logic of Explanatory Power. *Philosophy of Science* 78, 105-127.

Schurz, G. 1999. Explanation and Unification. *Synthese* 120, 95-114.

Schurz, G. 2008a. Common Cause Abduction and the Formation of Theoretical Concepts in Science. In C. Dégremont, L. Keiff, and H. Rückert (eds.), *Dialogues, Logics and Other Strange Things*. London; College Publications, pp. 337-364.

Schurz, G. 2008b. Reply to IlkkaNiiniluoto. In C. Dégremont, L. Keiff & H. Rückert (eds.), *Dialogues, Logics and Other Strange Things*. London; College Publications, pp. 371-374.

Schurz, G. 2015. Causality and Unification: How Causality Unifies Statistical Regularities. *Theoria* 30 (1), 73-95.

Schurz, G. and Lambert, K. 1994. Outline of a Theory of Scientific Understanding. *Synthese* 101, 65-120.

Shogenji, T. 1999. Is Coherence Truth Conducive? *Analysis* 59, 338-345.

Shogenji, T. 2013. Coherence of the Contents and the Transmission of Probabilistic Support. *Synthese* 190, 2525-2545.

Smokler, H. 1968. Conflicting Conceptions of Confirmation. *Journal of Philosophy* 65, 300-312.

Swinburne, R. 2004. *The Existence of God*, 2nd ed. Oxford: Oxford University Press.

van Fraassen, B. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.

Whewell, W. 1847. *The Philosophy of Inductive Sciences, Founded upon their History*, 2nd ed. London: John W. Parker and Sons.

Ilkka Niiniluoto has made his academic career at the University of Helsinki: Master of Science in applied mathematics (1968), Doctor of Philosophy (1974), Associate Professor of Mathematics (Logic and Foundations) (1973-1977), and Professor of Theoretical Philosophy (1977-2014). His main works in the philosophy of science include *Is Science Progressive?* (D. Reidel, 1984), *Truthlikeness* (D. Reidel, 1987), and *Critical Scientific Realism* (Oxford University Press, 1999, paperback 2002).

Address: Department of Philosophy, History, Culture, and Art Studies, PO Box 24, 00014 University of Helsinki, Finland. E-mail: ilkka.niiniluoto@helsinki.fi