

# Unification and Explanation: Explanation as a Prototype Concept. A Reply to Weber and van Dyck, Gijsbers, and de Regt \*

Gerhard SCHURZ

Received: 06.04.2013

Final version: 06.08.2013

BIBLID [0495-4548 (2014) 29: 79; pp. 57-70]

DOI: 10.1387/theoria.8729

**ABSTRACT:** In this paper I investigate unification as a virtue of explanation. In the first part of the paper (sec. 1-2) I give a brief exposition of the unification account of Schurz and Lambert (1994) and Schurz (1999). I illustrate the advantages of this account in comparison to the older unification accounts of Friedman (1974) and Kitcher (1981). In the second part (sec. 3) I discuss several comments and objections to the Schurz-Lambert account that were raised by Weber and van Dyck (2002), Gijsbers (2007) and de Regt (2005). In the third and final part (sec. 4), I argue that explanation should be understood as a prototype concept which contains nomic expectability, causality and unification as prototypical virtues of explanations, although none of these virtues provides a sufficient and necessary "defining condition" of explanation.

**Keywords:** unification; explanation; conjunction problem; nomic expectation; causality; proto-type concept.

**RESUMEN:** El artículo analiza la unificación como virtud explicativa. En su primera parte hago una breve exposición del enfoque de la unificación de Schurz y Lambert (1994) y Schurz (1999). Ilustro las ventajas de este planteamiento sobre otros anteriores, como Friedman (1974) y Kitcher (1981). En la segunda parte (secc. 3) discuto varios comentarios y objeciones al enfoque Schurz-Lambert planteados por Weber y van Dyck (2002), Gijsbers (2007) y de Regt (2005). En la tercera y última parte (secc. 4) argumento que la explicación debería entenderse como un concepto prototipo que contiene la esperabilidad nómica, la causalidad y la unificación como virtudes prototípicas de las explicaciones, aunque ninguna de estas virtudes ofrece una condición "definitoria", suficiente y necesaria, de la explicación.

**Palabras clave:** unificación; explicación; problema de la conjunción; esperabilidad nómica; causalidad; concepto prototipo.

## 1. *Unification as a Virtue of Theories and Explanations*

A characteristic property of good scientific theories is their *unification power*, i.e. their ability to predict and explain a variety of empirical phenomena in terms of a small number of basic principles. The idea that *unification* is the main goal of scientific theories has been articulated by prominent philosophers such as Mach (1883, 586f), Whewell (1847) and Feigl (1970, 12).<sup>1</sup>

The ideal of unification is related to, but not identical with the ideal of *coherence* (Lehrer 1974, ch. 7; Shogenji 2005, Bovens and Hartmann 2003, ch. 2). The coherence

---

\* This work was supported by Deutsche Forschungsgemeinschaft, research unit *Causation / Laws / Dispositions / Explanation* (FOR 1063). For valuable discussions on explanation, understanding and unification I am indebted to Michael Friedman, Philip Kitcher, Victor Gijsbers, Erik Weber, Henk de Regt, Margaret Morrison, Matti Sintonen, Manfred Stöckler, and two anonymous referees.

<sup>1</sup> Mach's name for the unification idea was "the principle of cognitive economy", and Whewell's name for it was "the consilience of induction".



of a belief system requires that the believed propositions mutually *support* each other, without discriminating non-circular from circular support relations. This is problematic, insofar purely speculative beliefs can be in mutual support, too. In contrast, unification consists in the coherence of a few hypothetical beliefs with a variety of *observational* beliefs. According to Schurz and Lambert (1994, 72), this means that circular coherence is discounted in unification, and empirical coherence is more rewarded than theoretical coherence.

More recently, unification was suggested to be the major feature of scientific explanations by Friedman (1974), Kitcher (1981) and Schurz and Lambert (1994) and Schurz (1999). These accounts see the main goal of explanations as to provide a deeper *understanding* of the empirical facts by means of *unifying* them, i.e. by reducing a variety of facts (viz. the statements expressing them) to a few basic principles and initial conditions. Schurz (1999) argues that neither Hempel's expectability of explanation nor Salmon's causal approach of explanation can account for the fact that an explanation is satisfying only if its premises are *not less in need of explanation* than the explanandum,<sup>2</sup> where the premises are less in need than the explanandum exactly iff the addition of the explanatory premises to the given knowledge system *increases* its unification. Schurz (1999, §2) gives the following example:

- (1) Peter is flying past the window in the third floor, because one second ago he was flying past the window in the fifth floor (plus laws of nature).

(1) is a causally and predictively perfectly adequate explanatory argument, but not a *satisfying* explanation, because the cause is here just as much in need of explanation as the effect. Similarly, predictively successful scientific laws are only explanatory if they are in conformity with the accepted background theory. An example of non-explanatory laws were Bohr's stability postulates for the electrons in an atom in the year 1913: These postulates could successfully predict the hydrogen spectrum, but they were incoherent with classical mechanics and thus were itself strongly in need of explanation. Therefore, most scientists at that time did not view them as a satisfying explanation.

Schurz and Lambert (1994) suggest explicating this idea in terms of the following *condition of unification*, which has to be *added* to a standard accounts of explanation such as Hempel's accounts of a deductive-nomological (DN) and inductive-statistical (IS) explanation (Hempel 1965), or Salmon's account of causal-probabilistic (CP) explanation (Salmon 1984):

- (U) *Unification condition*: An explanatory (quasi-) argument which satisfies the conditions of an adequate DN-, IS- or CP-explanation is a cognitively *satisfying* explanation of the explanandum E in the sense that it provides *understanding* of E, in a given belief system S, *iff* the addition of this argument to S leads to a successor belief state S\* which is more unified than S.

---

<sup>2</sup> Cf. Scriven (1959, sec. 3), Kutschera (1972, 331).

## 2. Definitions of Unification

It is difficult to *define* the degree of unification of a belief system in a satisfying way. So far, only a few philosophers of science have tried to do that. The problem is that one needs a *representation* of the given (scientific) belief system  $S$  which tells us what counts as “one” (elementary) belief or statement, and what as a conjunction of two or more beliefs. A solution to this problem is needed to solve Hempel’s *conjunction problem* (1965, 273, fn. 36) Hempel asked: Why don’t we simply “explain” and “unify” a set of accepted laws  $\{L_1, \dots, L_n\}$  by their conjunction  $L_1 \wedge \dots \wedge L_n$ ? If we would do this, this would undermine any reasonable unification account. So the explication of “unification” requires criteria by which we can distinguish between elementary statements and conjunctions of them. This is logically non-trivial since the degree of unification of a system of beliefs (propositions) should not depend on their linguistic formulation, so it should be preserved under logically (L-) equivalent transformations. But conjunctions can be L-equivalently transformed into negated disjunctions:  $p \wedge q \leftrightarrow \neg(\neg p \vee \neg q)$ . Moreover, atomic statements can be conjunctively expanded by tautologies:  $p \leftrightarrow p \wedge (q \vee \neg q) \leftrightarrow (p \wedge q) \vee (p \wedge \neg q)$ .

Friedman (1974, 16f) suggested solving the conjunction problem by assuming a primitive concept of “independent acceptability”. Intuitively, a sentence (or belief)  $P$  is acceptable independently of a sentence set  $D$  if in the given epistemic background system  $S$ ,  $P$  can be justified without assuming  $D$ . Friedman defines a sentence  $P$  as “S-atomic” iff it is not equivalent with a conjunction of sentences that are acceptable independently from  $P$ . The S-cardinality of a sentence set  $D$ , abbreviated as S-card( $D$ ), is defined as the cardinality of the smallest set of S-atomic statements that is L-equivalent with  $D$  (Friedman speaks of a “partition”; but this is unusual.) Finally, Friedman defines that a theory (or sentence set)  $T$  is explanatorily unifying iff  $T$  reduces the S-cardinality of its independently acceptable consequences, i.e.  $S\text{-card}(C_i(T) \cup \{T\}) < S\text{-card}(C_i(T))$ , where  $C_i(T)$  is the set of  $T$ ’s consequences that are independently acceptable of  $T$ . Unfortunately, Kitcher (1976) has shown that Friedman’s account doesn’t work well, whenever the given theory consists of two or more independently acceptable axioms. For if  $T = T_1 \wedge T_2$ , then  $C_i(T)$  contains  $\{T_1, T_2\}$ , so  $S\text{-card}(C_i(T)) = S\text{-card}(C_i(T) \cup \{T\}) = \text{card}(\{T_1, T_2\})$ ; so  $T$  does not unify.

Later on, Kitcher (1981, 520f) suggested an account of unification that is not based on the reduction of elementary accepted statements to a set of basic principles, but on the reduction of a set of accepted explanatory arguments to a set of basic *argument patterns*. An argument pattern is an argument in which certain non-logical expressions are replaced by variables, e.g.: “force(x,time) =  $\alpha$ , mass(x) =  $\beta$  / acceleration(x,time) =  $\alpha:\beta$ ” (where the stroke ‘/’ separates the explanans from the explanandum). An argument is reduced to an argument pattern if it is subsumed by it via a replacement of variables by corresponding (constant) terms, according to associated “filling instructions”. Kitcher suggests that a scientific belief system  $S$  with a set of accepted explanatory arguments  $E(S)$  is the more unified, (a) the more conclusions are derivable by arguments in  $E(S)$ , (b) the less basic argument patterns are needed to reduce all arguments in  $E(S)$ , and (c) the more stringent these basic patterns are, where more strin-

gent patterns have less variables and more specific filling instructions than less stringent patterns (*ibid.*, 520-2).<sup>3</sup> Kitcher then asks whether his account can eliminate explanations by spurious unifications—such as

- (2) Whatever God wants, happens; God wants P / (therefore) P.

Kitcher argues that the argument pattern (2) is not stringent enough (*ibid.*, 527f). But this answer is doubtful: we can make the argument pattern (2) more stringent by inflating its premises with some details such as:

- (3) Whatever God wants, happens, provided He has white hairs; God wants P and has white hairs / (therefore) P.

This inflated argument counts as unificatory in Kitcher’s account.

Apart from this problem, Kitcher does not explain what is intrinsically so bad about non-stringent argument patterns. For example, if theoretical physicists would find a “complete” theory, all physical facts could be explained by one general but “non-stringent” argument pattern. Then Kitcher would have to exclude this argument pattern, too. In conclusion, Kitcher’s account fails to solve the problem of spurious unification.

The account of unification of Schurz and Lambert (1994) and Schurz (1999) can be regarded as an elaboration of the basic idea of Friedman (1974). Like Friedman and Kitcher, Schurz and Lambert develop their unification account as part of a theory of scientific understanding and explanation. Like Friedman, their account is based on the conjunctive splitting of the given *cognitive* (or belief) system S into its smallest conjunctive elements. For this purpose, Schurz and Lambert use the method of *relevant* (consequence or content) *elements* which has been developed elsewhere as a means of solving problems of verisimilitude and hypothetico-deductive confirmation (Schurz and Weingartner 1987; Schurz 1991, 1994). A more recent presentation of this method is found in Schurz/Weingartner (2010, sec. 4).<sup>4</sup> We dispense with presenting the logical definitions here; see the quoted literature. We just give some examples. We write “E(S)” for the set of relevant elements of a sentence or set of sentences S.

*Examples of relevant elements:*

$E(\{p \wedge \neg q\}) = \{p, \neg q\}$ ;  $E(\{p \vee \neg q, p \vee q\}) = E(\{p\}) = \{p\}$ ;  $E(\{p \rightarrow q, q \rightarrow \neg r\}) = \{\neg p \vee q, \neg q \vee \neg r, \neg p \vee \neg r\}$ ;  $E(\{\forall x(Fx \rightarrow Gx), Fa\}) = \{\forall x(Fx \rightarrow Gx), Fa, \exists xFx, Ga, \exists xGx\} \cup \{Fa \rightarrow Ga: a \in Ic\}$  (where “Ic” = “set of individual constants of the language”).

Three important facts about relevant elements are:

- (i) *Invariance under L-equivalence*: Every two L-equivalent sentence (sets)  $S_1$  and  $S_2$  have the same content elements, i.e.  $E(S_1) = E(S_2)$ . Therefore, every property of a sentence (set) S explicated in terms of S’s relevant elements is invariant under L-equivalent transformations of S.

<sup>3</sup> Kitcher assumes that E(S) is determined as that set of explanatory arguments with conclusions and premises in S (among a set of competing explanation sets) which produce the highest unification of S according to the criteria (a), (b) and (c).

<sup>4</sup> A congenial method of representation by content parts has been developed by Ken Gemes (1993).

- (ii) *Content-preservation*: For every sentence (set)  $S$ ,  $E(S)$  is L-equivalent with  $S$ ; so no information gets lost by the representation of sentence (sets) by content elements (cf. Schurz and Weingartner 2010, lemma 7.2).
- (iii) *Equivalence with clauses*: Content elements coincide with clauses, which are disjunctions of atomic formulas or their negations (cf. Schurz and Weingartner 2010, lemma 5.1). This provides strong support for the method of relevant elements, insofar clauses are an extremely well corroborated representation method in computer logic.

In the Schurz-Lambert account of unification, a theory  $T$  does not reduce the set of *all* of its relevant elements, as in Friedman's account, but only the set of those relevant elements that are nomologically explained or predicted by  $T$ . This step avoids Kitcher's objection against Friedman. Moreover, not only deductive but also weaker probabilistic forms of nomological arguments are admitted. A *nomological (quasi-) argument* is understood as an argument "SL, SA // E" where SL and SA are sets of accepted general laws and facts, respectively, and E is a fact. Thereby E and the premises in SLUSA must be relevant elements of the given belief system  $S$  (i.e. the given set of accepted statements), and (b) SLUSA entails E or makes E at least "sufficiently probable".<sup>5</sup> If the argument is not deductive but probabilistic, it is more correct to call it a "quasi-argument": for then, the premises must be *maximally specific* to the conclusion (cf. Hempel 1965, sec. 12.3.4), so the argument is non-monotonic and the conclusion is not detachable (cf. Schurz and Lambert 1994, §2.3).

Schurz and Lambert avoid spurious unifications—the major problem of Kitcher's account—by assigning different costs and gains to observations versus hypotheses. In their basic account (1994, sec. 2.2) they define a *unification classification* of the system  $S$  of accepted beliefs (formally a set of sentences) as a partition  $\{B, U\}$  of  $E(S)$ , the set of  $S$ 's relevant elements, such that all elements of  $U$ , the subset of "unified elements", are conclusions of known (deductive or probabilistic) inferences which have their premises in  $B$ , the subset of "basic elements". *The* unification classification is defined as that unification classification which *maximizes* the unification of  $S$  according to the following criteria (Ci):<sup>6</sup>

- (C1) Every content element  $P$  is associated with a certain *weight*  $w(P)$ , which reflects its cognitive *complexity*. The complexity of general or theoretical propositions is greater than that of singular and empirical propositions.
- (C2) If a hypothesis  $H$  is newly added to  $B$ , its weight is paid as a unificatory *cost*  $-w(H)$ , while if a new observation  $O$  is added to  $B$ , this costs nothing.
- (C3) If a hypothesis  $H$  is inferred by a "perfect" explanation and hence added to  $U$ , this costs nothing, while if an observation  $O$  is inferred by a "perfect"

<sup>5</sup> See Schurz and Lambert (1994, 88-92). Schurz and Lambert represent a "cognitive corpus" more precisely as a pair  $(E, I)$  where  $E$  is the set of accepted elementary beliefs and  $I$  a set of known nomological arguments.

<sup>6</sup> If there are ties, i.e. if there exist several maximally unified classifications, the unification classification is the alphabetically first of them.

explanation and added to  $U$ , then its weight is added as a unificatory *gain*  $+w(O)$ . A “perfect” explanation is understood as a causal and deductive-nomological explanation. Schurz and Lambert give two standard arguments for this preference (1994, 74): first, deductive arguments make their conclusion predictable with *certainty* and not merely with probability (as probabilistic quasi-arguments), and second, an argument whose factlike premises cites *causes* goes beyond mere predictability as it enables one to exert practical *control* over the inferred fact. If the explanation is merely probabilistic, or if it is merely predictive but not causal, then the cost of the explanandum is only *partially saved* (Schurz and Lambert 1994, §1.4, §3.2). So if an imperfectly explained hypothesis is added to  $U$ , this costs a certain fraction of  $w(H)$ , and if an observation is imperfectly explained, this brings a certain fraction of  $w(O)$  as gain.

These rules lead to a comparative notion of unification by the method of “shift diagrams”:  $P_{+X}$ ,  $P_{-X}$  and  $P_{X \rightarrow Y}$  stand for adding an element  $P$  to  $X$ , removing it from  $X$ , or shifting it from  $X$  to  $Y$ , where “ $P_{X \rightarrow Y}$ ” is understood as the addition-subtraction-sequence  $(P_{-X}, P_{+Y})$ , and  $X, Y \in \{B, U\}$ . The six basic unification shifts are analyzed as follows (“H” for “hypothesis”, “O” for observation, “u” for “unification”, the explanation is assumed as “perfect”):

- (i)  $O_{+B}$  (adding a new observation  $O$  to  $B$ ) is u-preserving, since costs and gains cancel out.
- (ii)  $O_{+U}$  (adding a new observation as explained) is u-increasing with gain  $+w(O)$ .
- (iii)  $H_{+B}$  (adding a hypothesis to  $B$ ) is u-decreasing with cost  $-w(H)$ .
- (iv)  $H_{-U}$  (subtracting a hypothesis from  $U$ ) is u-preserving.
- (v)  $L_{+B}$ ,  $O_{B \rightarrow U}$  (explaining one datum  $O$  on the cost of adding one new law) is u-decreasing, since the cost  $-w(L)$  dominates the gain  $+w(O)$ .
- (vi)  $L_{+B}$ ,  $\{O_1, \dots, O_n\}_{B \rightarrow U}$  (explaining  $n$  observations on the cost of one new law) is u-increasing for sufficiently large  $n$ , since the gain of sufficiently many derived singular elements is greater than the cost of one new basic law.

Schurz and Lambert (1994, sec. 3.4) extend their basic account to a refined one, which distinguishes between four kinds of elements: basic, actually explained or assimilated, heuristically assimilated (how-possible explanations), and dissimilated (Kuhnian anomalies). They apply their account to a variety of cognitive belief changes including examples from science. The problem of spurious unification is solved as follows: Consider again the spurious unifications (2) and (3) above. In these examples,  $n$  facts ( $F_i$ ) are explained on the cost of  $n$  new ex-post speculations ( $C_i$ ) about God’s wishes (“God wanted fact  $F_i$ ”) plus one new law  $L$  (Whatever God wants, happens). Since  $w(C_i) > w(F_i)$  and  $w(L) > 0$ , the unification decreases (in example (3) the situation is even worse, since more premises are needed). In contrast, typical scientific explanations of empirical laws by higher order theories produce an enormous unification increase (see the examples in Schurz and Lambert 1994, §2.3).

3. *Comments on the Schurz-Lambert Account and Replies*

Weber and van Dyck (2002) raise three objections against the Schurz-Lambert account (thereby referring to Schurz 1999):

*Weber and van Dyck—Objection 1:*

Weber and van Dyck (2002, 150f) write that all explanations in the Schurz-Lambert account are “inferential”, either deductive-nomological arguments or inductive-statistical quasi-arguments. This account would ignore causal *contrast* explanations, as for example the following:

(4) *Weber and van Dyck’s example:*

*Explanandum:* Truckers exceed the speed limit more often than car drivers. Formally:  $P(S|T) > P(S|\neg T)$  (“T” for “trucker” and “ $\neg T$ ” for “non-truckers”, i.e. car-drivers)

*Explanans:* (i) Truckers have more often radar detectors than car drivers (and listen more often to the CB radio which warns about radars), and (ii) having a radar detector is a positive cause of exceeding the speed limit. Formally: (i)  $P(R|T) > P(R|\neg T)$  and (ii)  $P(S|R) > P(S|\neg R)$ .

I think that Weber and van Dyck’s objection misses its target, for two reasons:

First, is incorrect that all explanations in the Schurz-Lambert account are inferential. In sec. 3.2 Schurz and Lambert (1994) explicitly admit: (i) strict or approximative DN-explanations, (ii) probability raising explanations, provided the probability of the explanandum is raised to a value of greater than 1/2, and (iii) causally complete explanations, even if the probability value is not increased above 1/2 (cf. the “random connections” mentioned on p. 92f).

Second, Weber and van Dyck’s explanation example rests on the following probabilistic derivation which employs an implicit causal assumption of screening off (4d):

*Derivation of explanation (4) from probability axioms:*

(4a)  $P(R|T) > P(R|\neg T)$  *Premise.*

(4b)  $P(S|R) > P(S|\neg R)$  *Premise.*

(4c)  $P(S|T) = P(S|T \wedge R) \cdot P(R|T) + P(S|T \wedge \neg R) \cdot P(\neg R|T)$  *Probability theorem.*

(4d)  $P(S|T \wedge R) = P(S|R)$  *Implicit assumption, because the direct cause R screens off the indirect cause T from S.*

(4e) Therefore:  $P(S|T) = P(S|R) \cdot P(R|T) + P(S|\neg R) \cdot P(\neg R|T)$  *From 4c and 4d.*

Writing  $r$  for  $P(R|T)$  and  $q$  for  $P(R|\neg T)$  we get:

(4f) Therefore:  $P(S|T) = P(S|R) \cdot r + P(S|\neg R) \cdot (1-r)$   
 $> P(S|\neg T) = P(S|R) \cdot q + P(S|\neg R) \cdot (1-q).$

- (4f) follows by easy arithmetics from 4e, 4a and 4b: both  $P(S|T)$  and  $P(S|\neg T)$  are weighted averages of  $P(S|R)$  and  $P(S|\neg R)$  with weights  $r$ ,  $(1-r)$  and  $q$ ,  $(1-q)$ , respectively, where by (4a)  $r > q$  and by (4b)  $P(S|R) > P(S|\neg R)$ .

This derivation shows that Weber and van Dyck's explanation example is itself an inferential explanation (based on a deduction in probability calculus). So their example seems to be unsuited to serve their own purpose. Nevertheless, Weber and van Dyck may be right that the notion of "explanatory argument" in Schurz and Lambert (1994) is construed too narrow. Investigation of this question exceeds the limitation of this discussion note.

*Weber and van Dyck—Objection 2:*

Next, Weber and van Dyck (*ibid.*, 151f) object that the premise of their explanation (4), namely that truck drivers have more often radar detectors (and listen more often to radar warnings in the CB radio) is itself equally in need of explanation as the explanandum, i.e. that truck drivers more often exceed the speed limit. Therefore their explanation, so they argue, violates Schurz and Lambert's unification condition (U) (sec. 1). Since their explanation (4) is okay in their eyes, condition (U) would be inappropriate. — However, it seems to me that their argument demonstrates the point of Schurz and Lambert's theory: without further background knowledge, it is indeed *not less* surprising *why* truckers more often have radar detectors than car drivers than why they exceed the speed limit more often. So Weber and van Dyck's explanation (4) is in fact *not a satisfying* explanation in the sense of Schurz and Lambert. Only the third explanation which Weber and van Dyck assume in their background is explanatorily satisfying, namely:

- (5) The income of truckers depends more often and much more on the number of miles they drive than the average income of car-drivers, and this income-dependence on the number of driven miles is a positive cause of exceeding the speed limit.

Weber and van Dyck seem to agree with Schurz that explanation (5) is preferable to explanation (4). Condition (U) is nothing but an explication of this preference. Nevertheless Weber and van Dyck criticize Schurz by arguing that explanation (4) is an acceptable explanation (though not as good as (5)), insofar the explanans of (4) provides causal information which enables us to manipulate the effect variable (exceeding the speed limit)—so the unification condition (U) would be inadequate. However, this objection does not really hit Schurz's condition (U). At this point of the argumentation, my thesis that explanation is a *prototype concept* (which will be elaborated in section 4) becomes crucial. This thesis implies that both causal information and unification (surprise-reduction) are prototypical properties of explanations which *normally* go together, but not always. Condition (U) merely says that an explanatory argument which satisfies all *other* adequacy conditions for explanatory arguments (or quasi-arguments) can only be cognitively satisfying explanation if it provides unification. This is compatible with Weber and van Dyck's claim that the argument (4) has at least some explanatory value.



*Weber and van Dyck—Objection 3:*

Weber and van Dyck (2002) write on p. 153 that in Schurz's account there is no reason to prefer the explanation of several explananda  $Ea_1, \dots, Ea_n$  of the same type (e.g. Peter has blood group A, Mary has blood group A,...) with corresponding antecedents and only *one* law  $L$ , as opposed to explaining every explanandum  $Ea_i$  by a corresponding antecedent using a *different* law  $L_i$ , which needs  $n$  different laws in total. But given our sketchy presentation in sec. 2 above, this assertion is obviously wrong and must be based on a misunderstanding: Every explanatory premise which is not an observation and has to be added to the unification basis increases the cost (cf. also the examples in Schurz and Lambert 1994, sec. 2.3). So the Schurz-Lambert account obviously prefers explanations with only one law as opposed to explanations with  $n$  laws.

Gijsbers (2007) gives an accurate and well-organized reconstruction of Schurz-Lambert's account. He acknowledges (*ibid.*, 488, 500) that the Schurz-Lambert definition of unification works as intended and solves the problems of spurious unification that beset Kitcher's account. His critical remarks concern the relation of unification to explanation, causality and lawhood.

*Gijsbers—Objection 1:*

For Schurz and Lambert, the following argument is a satisfying explanation of Peter's flying past the window in the 3rd floor (see Schurz 1999, §5):

- (6) Peter is flying past the window in the third floor, because the fire brigade is testing a new jumping sheet at our building (and Peter is participating in the event).

In distinction to (1), the answer in (6) is itself not as surprising as the explanandum, since there is nothing puzzling about fire brigades testing jumping sheets (Schurz 1999, 108). As Gijsbers correctly observes, Schurz understands the being-in-need-of-explanation not in a probabilistic sense (this would cause problems), but in terms of *how-possible explanations*: a (description of a) fact  $E$  is in need of explanation in a given cognitive system  $S$  iff there exists no how-possible explanation of  $E$  in  $S$ . Gijsbers (2007, 497) objects that it is impossible that a surprising fact  $E$  (in example (6): "Peter is flying past the window in the third floor") *does not* have a how-possible explanation in a cognitive state  $S$ , although the explanatory answer  $A$  for  $E$  *does* have a how-possible explanation in the expanded cognitive state  $S^*$  which results from  $S$  by adding the answer  $A$  to it. For assume that  $A$  (in example (6): "the fire brigade is testing a new jumping sheet in *our* building") has a how-possible explanation in  $S^*$ . This means that a certain proposition  $A^*$  is compatible with  $S^*$  (though it is not believed in  $S^*$ ) and  $A^*$  makes  $A$  inferable by a (quasi-) argument in  $S^*$  (cf. Schurz and Lambert 1994, 97). In example (6),  $A^*$  could be: "the fire brigade has selected *our* building for testing their new jumping sheet". But then, so argues Gijsbers,  $A^*$  is also compatible with  $S$ , and  $A^*$  offers also an argument for the fact  $E$  which was in need of explanation in the cognitive state  $S$ . This argument is obtained by concatenating the argument going

from  $A^*$  (plus additional premises  $P^*$ ) to  $A$  with the argument going from  $A$  (plus additional premises  $P$ ) to  $E$ .

Gijsbers makes an interesting point. His point would indeed be true if cognitive systems would be reconstructed as sets of believed propositions (represented by relevant elements). But Schurz and Lambert construe cognitive systems as pairs  $(E, I)$  (recall fn. 5), where  $E$  is the set of accepted relevant elements and  $I$  a set of nomological (quasi-) arguments that are known to be correct. What Schurz and Lambert assume is that the explanatory answer in (6) implicitly introduces also a new argument into the cognitive system: that the possibility that the fire brigade is testing a new jumping sheet at our building constitutes a possible argument for the fact  $E$ . Formally, in state  $S = (E, I)$  the argument going from  $A$  to  $E$  is not in  $I$  (so also the argument going from  $A^*$  to  $E$  is not in  $I$ , too), while in state  $S^* = (E^*, I^*)$  the argument going from  $A$  to  $E$  is in  $I^*$ . So by adding the explanatory answer to  $S$ ,  $I$  gets expanded by a new argument which makes the explanatory premise unsurprising. In this way, Gijsbers's objection is resolved.

*Gijsbers—Objection 2:*

Gijsbers (*ibid.*, 499) observes correctly that condition (U) goes beyond the local connectedness of the explanandum with the explanatory premise and requires some kind of global connectedness (or “fitting”) of  $P$  with the cognitive system  $S$ . Gijsbers argues that this global connectedness (i.e. condition U) should not be construed as a *necessary* condition for an explanation. Here I would completely agree: as explained before I do not regard (U) as a necessary condition of explanation. I rather assume that explanation is a prototype concept and regard (U) as a condition for a *satisfying*, i.e. an *epistemically ideal* explanation. However Gijsbers's second objection goes further. He argues that the *only* reason why unification is important at all for explanation is that the degree of unification produced by the explanatory premises correlates with their degree of confirmation, i.e. their epistemic probability. I would agree that this overall increase of confirmation is *one* reason why unification is a virtue; in this sense, Schurz and Lambert (1994, §1.3) write that unification yields confirmation as a byproduct.

But is Gijsbers right that this confirmation-increase is the *only* virtue of unification for explanations? If this were true, the unification account to explanation could be replaced by something simpler. I don't think that Gijsbers is right on this point. Let me ask: Why is the possibility of generating how-possible explanation epistemically important at all? This is not so because of confirmational reasons, since how-possible explanations do not produce a (significant) probability-increase of anything. How-possible explanations give us a *more complete* causal representation of the world, if only as a “plausible possibility”. In the same sense, the virtue of unification for explanations does not only consist in confirmation increase, but also in increasing the *completeness* and *depth* of our cognitive representation of the world. This is a virtue in itself.

*De Regt, and Weber and van Dyck—Joint Objection:*

De Regt (2006, 138) as well as Weber and van Dyck (2002, 152) doubt that the core meaning of causality can be saved when causality is understood—as proposed in Schurz and Lambert (1994) and Schurz (1999, 101)—as relative to an accepted general causality theory CT, where the acceptance of CT is in turn motivated by the search for unification at the theoretical level. However, that one of the most influential contemporary theory of causality—the theory of causal Bayes nets—can indeed be justified exactly in this way is shown in detail in Schurz and Gebharter (2014). Explaining this account here would go far beyond the scope of the present paper. Therefore I confine myself to drawing the attention of the critics to this paper.

*4. Explanation as a Prototype Concept: Three Paradigms of Explanation*

For de Regt (2006), causal and unificatory explanations are *complementary* explanation paradigms. I agree with him, provided “complementary” is not construed as “opposition”. I think, however, that the explanation debate has not produced only two but actually three explanation paradigms:

- (1) *The expectability paradigm:* The key property of the explanation of an explanandum E consists in making E expectable, in one of the following four ways (or ‘versions’ of this paradigm):
  - (i) either by deductively inferring E from the explanatory premises (Hempel 1965, ch. 10; Halonen/Hintikka 2005);
  - (ii) or by making E highly probable (Hempel 1965, ch. 12.3), or at least more probable than non-E, conditional on the explanatory premises (Tuomela 1981, 276; Stegmüller 1983, 972),
  - (iii) or by making E more probable than all of E’s competitors in a given partition of alternative outcomes (van Fraassen 1980, 141ff).

The expectation paradigm leads to a very *broad* notion of explanation which applies to the explanation of singular and general facts, but which does not capture the narrow aspect of explanation as delivering not mere reasons for believing but causes.

- (2) *The causality paradigm:* The crucial property of an explanation consists in delivering a conjunction of causes  $C_E = C_1 \wedge \dots \wedge C_n$  of E which is as complete as possible, together with causal laws connecting  $C_E$  with E.  $C_E$  has to change E’s probability, but it is irrelevant whether  $C_E$  increases or decreases E’s probability (Jeffrey 1971, Salmon 1984; Hempel 1977, 99f).

The causality paradigm works fine for the explanation of singular events, but is hardly applicable to the explanation of laws from higher-order theories. The purpose of these explanations in science is usually expressed in terms of unification:

- (3) *The unification paradigm:* The main goal of explanations consists in providing a deeper understanding of the phenomena by unifying them in the way as explained above.

The fact that the explanation debate has produced three mutually incompatible paradigms of explanation does not seem to be a very satisfying result. However, a little reflection may turn this seemingly negative result into a *positive* insight about the nature of the concept of explanation. For note that in *most* instances of scientific explanations, all three paradigms coincide, i.e. all three corresponding properties of explanations are simultaneously realized: probability-increase, causal information and unification.

Firstly, causal explanations are normally also predictively most efficient and robust, and vice versa. For example, it follows from probability theory that if two events A and B have a common non-deterministic cause C, then the correlation between A and B will be lower than the correlations between C and A and C and B, respectively. Moreover, if a correlation between two events A and B is robust under conditionalization on other events, this constitutes a sign of its causal character (cf. Cartwright 1989, 145f; Schurz 2002, sec. 2). Of course there exist cases of quasi-arguments which provide causal information without increasing the probabilistic expectability of the explanandum. A famous example is Salmon's causal explanation of the white colour pea plants (cf. Salmon 1984, 109). But even low-probability laws convey a high probability value to most observed sample frequencies. Moreover, in science probabilistic laws are in most cases not applied to the explanation of individual events, but to explanations of sample frequencies, and for these applications causality and high probability are normally going together. In conclusion, situations in which causality and predictive efficiency depart are rare.

Secondly, explanatory premises which are predictively most powerful are normally exactly those explanations which offer the highest degree of unification, and vice versa. This holds for our characterization of unification because the percentage of the cognitive cost of a fact E, which is saved by explaining it, is proportional to the degree of the conditional probability of E given the explanans (recall section 2). Of course, there exist kinds of unifications—for example the “random connections” of Schurz and Lambert (1994, 92f)—which do not go in parallel with predictability. But again, these cases are rare.

So the three features which each of the three paradigms takes to be the ‘essence’ of scientific explanations *go normally together*, and depart from each other only in *exceptional* cases.

This situation does not imply that the notion of “explanation” is not a reasonable one, but merely, that it has a semantic property which it shares with many other natural language concepts, namely that it is a *prototype* concept (cf. Margolis and Laurence 1999). The meaning of a prototype concept (e.g. “bird”) cannot be sharply defined by a list of individually necessary and jointly sufficient conditions, but is more or less vaguely given, either by way of prototypical examples (“this is a typical bird”) or by a list of prototypical properties (“normal birds have wings, can fly,...”). The concept of explanation is a prototype concept which is characterized by the three prototypical properties of expectability, causation and unification. Normally these properties go together, but as there exist birds without wings and birds with wings that cannot fly, there exist predictive explanations that are neither causal nor unificatory, causal expla-

nations that are not predictive nor unificatory, and unificatory explanations that are neither causal nor predictive.

The relevance of this insight for the explanation debate is this: Most of its authors attempted to explicate “explanation” by a set of individually necessary and jointly sufficient conditions. But if explanation is a prototype concept, then the fact that the participants of the explanation debate were unable to come up with such a list of conditions should not be conceived as a disappointment, but rather as a positive insight. If “nomic expectation”, “causality” and “unification” are the three major prototypical properties of explanation, then the endless philosophical dispute about which of them constitutes the “essence” of explanations is pointless. One should understand the different models developed in the explanation debate as models explicating *one* of these prototypical characteristics of explanations. In this perspective, many counterarguments to these models which were brought up in the debate—showing that there are *some* cases of explanation which do not fit with the model—lose their force as “refutations” of these models.

A similar situation seems to underlie some arguments discussed in the previous section. For example, Gijsbers (2007) attempts to show that unification is neither necessary nor sufficient for explanation. However, I don’t make any such claim (refuted by Gijsbers) because I regard explanation as a prototype concept. Similarly, Weber and Dyck (2002, abstract) wrongly suppose I would claim that explanation always consists in unification. Again I don’t make this claim. The general conclusion which follows from the prototypical nature of the explanation concept seems to be this: One should be *pluralistic* about intuitions of what “good” and “bad” explanations are. No list of individually necessary and jointly sufficient conditions can ever cover all examples of (quasi-)arguments which according to some well-established intuition count as an explanation. All what one can do is to explicate the major prototypical properties of good explanations, to analyze their consequences and to find out how they are related to each other.

## REFERENCES

- Bovens, Luc and Stephan Hartmann. 2003. *Bayesian Epistemology*. Oxford: Oxford University Press.
- Cartwright, Nancy. 1989. *Nature's Capacities and their Measurement*. Oxford: Clarendon Press.
- De Regt, Henk. 2006. Wesley Salmon's Complementary Thesis: Causalism and Unificationism Reconciled? *International Studies in the Philosophy of Science* 20: 129-147.
- Feigl, Herbert. 1970. The Orthodox View of Theories: Remarks in Defense as well as Critique. In *Minnesota Studies in the Philosophy of Science*, vol. IV. Minneapolis: University of Minnesota Press.
- Friedman, Michael. 1974. Explanation and Scientific Understanding. *Journal of Philosophy* 71: 5-19.
- Gemes, Ken. 1993. Hypothetico-Deductivism, Content, and the Natural Axiomatization of Theories. *Philosophy of Science* 54: 477-487.
- Gijsbers, Victor. 2007. Why Unification is Neither Necessary nor Sufficient for Explanation. *Philosophy of Science* 74: 481-500.
- Halonen, Ilpo and Jaakko Hintikka. 2005. Towards a Theory of the Process of Explanation. *Synthese* 143: 5-61.
- Hempel, Carl Gustav. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

- . 2001. Postscript: More Recent Ideas on Statistical Explanation. In *The Philosophy of Carl G. Hempel*, edited by James Fetzer. Oxford: Oxford University Press.
- Jeffrey, Richard Carl. 1971. Statistical Explanation vs. Statistical Relevance. In *Statistical Explanation and Statistical Relevance*, ed. Wesley C. Salmon, 19-28. London: University of Pittsburgh Press.
- Kitcher, Philip. 1976. Comments and Criticism: Explanation, Conjunction, and Unification. *Journal of Philosophy* 73: 207-212.
- . 1981. Explanatory Unification. *Philosophy of Science* 48: 507-531.
- Kutschera, Franz von. 1972. *Wissenschaftstheorie*. Vol. I and II. Munich: Fink.
- Lehrer, Keith. 1974. *Knowledge*. Oxford: Clarendon Press.
- Mach, Ernst. 2009. *The Science of Mechanics*. Charleston: BiblioBazaar.
- Margolis, Eric and Stephen Laurence. 1999. Concepts and Cognitive Science. In *Concepts*, eds. Eric Margolis and Stephen Laurence, 3-82. Cambridge: MIT Press.
- Salmon, Wesley. 1971. *Statistical Explanation and Statistical Relevance*. London: University of Pittsburgh Press.
- . 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Schurz, Gerhard. 1991. Relevant Deduction. *Erkenntnis* 35: 391-437.
- . 1994. Relevant Deduction and Hypothetico-Deductivism. *Erkenntnis* 41: 183-188.
- . 1999. Explanation as Unification. *Synthese* 120 (1): 95-114.
- . 2002. Ceteris Paribus Laws: Classification and Deconstruction. *Erkenntnis* 57: 351-372.
- Schurz, Gerhard and Karel Lambert. 1994. Outline of a Theory of Scientific Understanding. *Synthese* 101: 65-120.
- Schurz, Gerhard and Paul Weingartner. 1987. Verisimilitude Defined by Relevant Consequence-Elements. In *What is Closer-To-The-Truth?*, ed. Theo A. F. Kuipers, 47-78. Amsterdam: Rodopi.
- . 2010. Zwart and Franssen's Impossibility Theorem Holds for Possible-World-Accounts but not for Consequence-Accounts to Verisimilitude. *Synthese* 172: 415-436.
- Schurz, Gerhard and Alexander Gebharter. 2014. Causality as a Theoretical Concept with Empirical Significance. (submitted paper)
- Scriven, Michael. 1959. Truism as Grounds for Historical Explanations. In *Theories of History*, ed. Patrick Gardiner. New York: The Free Press.
- Shogenji, Tomoji. 2005. Justification by Coherence from Scratch. *Philosophical Studies* 125: 305-325.
- Stegmüller, Wolfgang. 1983. *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*. Vol. I (2<sup>nd</sup> edition). Berlin: Springer.
- Tuomela, Raimo. 1981. Inductive Explanation. *Synthese* 48: 257-294.
- Van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Clarendon Press.
- Weber, Erik, and Maarten van Dyck. 2002. Unification and Explanation. *Synthese* 131: 145-154.
- Whewell, William. 1837. *History of the Inductive Sciences*. London: John W. Parker.

**GERHARD SCHURZ** is full professor of philosophy at the University of Düsseldorf and director of the Düsseldorf Center for Logic and Philosophy of Science (DCLPS). He was associated professor at the University of Salzburg and visiting professor at the University of California at Irvine and at Yale University. His research areas cover philosophy of science, logic, epistemology and cognitive science, and meta-ethics. He is author of more than 100 publications in international Journals. Book publications among others: *The Is-Ought Problem* (Kluwer 1997), *Einführung in die Wissenschaftstheorie* (Wissenschaftliche Buchgesellschaft, 3rd edition 2011), *Philosophy of Science: A Unified Approach* (Routledge 2014).

**ADDRESS:** Department of Philosophy, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany. E-mail: gerhard.schurz@phil.hhu.de