Titel / Title:

Common Cause Abduction and the Formation of Theoretical Concepts

Autor / Author:

Gerhard Schurz

TPD PREPRINTS

Annual 2008 No.2

Edited by Gerhard Schurz and Markus Werning

Vorveröffentlichungsreihe des Lehrstuhls für

Theoretische Philosophie an der Universität Düsseldorf

Prepublication Series of the Chair of

Theoretical Philosophy at the University of Düsseldorf

Common Cause Abduction and the Formation of Theoretical Concepts in Science

Gerhard Schurz (University of Duesseldorf, Germany)

Abstract: Abductions are conceived as special patterns of inference to the best explanation whose structure *determines* a particularly *promising* abductive conjecture (conclusion) and thus serves as an *abductive search strategy* (§1).An important distinction is that between *selective abductions*, which choose an optimal candidate from given multitude of possible explanations, and *creative abductions*, which introduce new theoretical models or concepts (§2). The paper focuses on creative abductions, which are essential for scientific progress, although they are rarely discussed in the literature. It is suggested to demarcate scientifically fruitful abductions from purely speculative abductions by the *criterion of causal unification* (§3). Based on various historical examples it is demonstrated that *common cause abduction from correlated dispositions* is the fundamental abductive operation by which new theoretical concepts are scientifically generated (§4). *Statistical factor analysis* can be regarded as a statistical generalization of common cause abduction (§5). When scientists start to develop theoretical models of their conjectured common (unobservable) causes, common cause abductions turns into what is called *theoretical model abduction* (§6).

1. On the Relation between Inference to the Best Explanation and Abduction

Harman (1965) understood inference to the best explanation (IBE) and abduction as more or less equivalent. Both inferences serve the goal of inferring something about the *unobserved causes* or *explanatory reasons* of the observed events. This was also the understanding of abduction in the mind of the inventor of 'abduction', C.S. Peirce (cf. 1903, CP 5.189). Nevertheless, I suggest to make a difference here. By an 'inference' I mean a certain (logically explicable) *schematic pattern* which *specifies* the conclusion as a (syntactical) *function* of the premises. In the case of an abductive inference, the premises describe the phenomenon which is in need of explanation, possibly together with background knowledge of a certain form, and the conclusion is an abductive conjecture which is set out to further test operations. However, in the case of IBE no such pattern exists. Rather, the space of possible explanatory hypotheses and their goodness-evaluation must already be given in the premises, in order to apply the rule of IBE. In other words, IBE is more an instance of the rule of *rational choice* than an inference operation.

All inferences have an *justificational* (or 'inferential') and a *strategical* (or 'discovery') function, but to a different degree (see also Gabbay/Woods 2005, §1.1). The justificational function consists in the justification of the conclusion, *conditional* to the justification of the premises. The strategical function consists in finding a most promising conjecture (conclusion) which is set out to further test operations, or in Hintikka's words, which stimulates new *questions* (Hintikka 1998, 528). In abductive inferences the strategical function becomes *crucial*. Different from the situation of induction, in abduction problems we are often confronted with thousands of possible explanatory conjectures (or conclusions) – everyone in the village might be the murderer. The essential function of abductions is their role as *search* strategies which tell us which explanatory conjecture we should set out *first* to further inquiry (cf. Hintikka 1998, 528) – ore more generally, which suggest us a *short* and *most promising* (though not necessarily successful) path through the exponentially explosive *search space* of possible explanatory reasons.

Therefore I suggest in (Schurz 2006) to understand abductions as special patterns of inference to the best explanation (IBS) whose structure *determines* a particularly *promising* abductive conjecture (conclusion) and thus serves as an *abductive search strategy* in the space of possible explanatory hypotheses. It is essential for a good search strategy that it leads us to an optimal conjecture not only in a finite but in a *reasonable* time. In this respect, the rule of IBE fails completely. If you ask which explanatory conjecture you should choose for further investigation among thousands of possible conjectures, the rule IBE just tells us: "find out which is the best (available) conjecture and then choose it". To see the joke behind, think about someone in a hurry who asks an IBE-philosopher for the right way to the railway station and receives the following answer: "Find out which is the shortest way among all ways between here and the train station which are accessible to you – this is the way you should choose".

In contrast to their strategical function, the justificational function of abductions is minor. Peirce has pointed out that abductive hypotheses are prima facie not even probable, like inductive hypotheses, but merely possible (1903, CP 5.171). Only upon being confirmed in further tests, an abductive hypothesis may become probable. However, I cannot completely agree with Peirce or other authors (e.g. Hanson 1961, Hintikka 1998) who think that abductions are merely a discovery procedure and their justification value is zero. Niiniluoto has pointed out that "abduction as a motive for pursuit cannot always be sharply distinguished from considerations of justification" (1999, S442). This paper will confirm Niiniluoto's point that for the considered patterns of abduction their strategical function goes hand in hand with a (weak) justificational value.

2. Selective versus Creative Abductions

In Schurz (2006) I provide a classification of different patterns of abduction, reaching from fact-abductions, 1st order existential abductions, and law-abductions to theoretical model abductions and 2nd order existential abductions which may introduce new concepts. One result of the analysis in Schurz (2006) is this: the epistemological function and the evaluation criteria of abduction are rather different for different kinds of abduction patterns. An important distinction in this respect is that between selective and creative abductions (cf. also Magnani 2001, 20). *Selective* abductions choose the best candidate among a multitude of possible explanations which is determined by the background knowledge, according to some selection strategy. The most frequently discussed pattern of selective abduction is backward reasoning through given causal laws:

Known Law: If it rains, the street is wet $\forall x(Cx \rightarrow Ex)$
Known Evidence: The street is wet (Ea)
==========================
Abduced Conjecture: (Probably) It has rained (Ca).

The young Peirce (1878) has explicated abduction in this narrow way. Since the background knowledge contains usually many laws of the form "$C_ix \rightarrow Ex$", and iteratively "$D_jx \rightarrow C_ix$", etc., this kind of abduction leads into backward chaining through known causal from a given explanandum to a set of possible explanatory reason. Among the possible explanatory reasons one has to select the most plausible one. In this form, abductive inference has been studied in detail in AI research (cf. Paul 1993, Josephson/Josephson 1994, Flach/Kakas 2000).

Selective abductions remain within the search space determined by the known background laws. They can never create new laws or even new concepts. Exactly this is the task of *creative* abductions. The later Peirce considered scientific abduction as an essentially creative operation by which new laws, theories or concepts are discovered (1903, CP 5.170)). But how and by which kinds of rules should that be possible? For example, can there ever be rules which achieve the following creative abduction?

Known phenomena: Observable properties of substances (gold, iron, stone etc.)
=========== **???** $\downarrow$ **???** ====================================
Abductive conjecture: Molecular models of these substances

Peirce kept silent about this question. Until today, most philosophers of science are skeptical whether the scientific discovery new theories or theoretical concepts follows any abductive rules. Has *Popper* here the last word who has repeatedly argued that that only the question of justifying an already *given* theory has a 'logic', while the question of discovering a *new* theory is a matter of sheer guessing?

Peirce once remarked there are sheer myriads of possible hypotheses which would explain the experimental phenomena, and yet scientists have usually managed to find the true hypothesis after only a small number of guesses (cf. CP 6.5000). Peirce explained these miraculous ability of human minds by their *abductive instincts* (CP 5.47, fn. 12; 5.172; 5.212). But I suggest we should not put too much trust in the abductive instincts of humans – for these abductive instincts have produced too many speculative or even irrational pseudo-explanations. The same problem arises for Harman's IBE: if IBE is understood as an inference to the best *available* explanation, then it unacceptable, because the best available explanation is not always *good enough* to be rationally acceptable (cf. Lipton (1991, 58). If a phenomenon is novel and poorly understood, then one's best available explanation is usually a *pure speculation*. For example, in the early animistic word-views of human mankind the best available explanations of natural phenomena such as the sun's path over the sky was that the involved entities (here: the sun) are intentional agents. Such speculative explanations are not acceptable in science, because they do not meet important methodological criteria, which are discussed in §§ 3-4.

Therefore the Peircean question *returns*: how was it possible for three centuries of scientists, after millennia of idling speculations, to find out so many true theoretical explanations in the astronomically large space of possible explanatory stories? This questions demands for an answer. The answer depends on whether there can exist anything like a 'logic' of discovery. The true observation of Popper and the logical positivists that the justification of a hypothesis is independent from the way it was discovered does not imply that it would not be *desirable* to have in addition *good heuristic rules for discovering new explanatory hypotheses and concepts* – if there only *were* such rules (cf. also Hanson 1961). In the following sections I try to show that there *are* such rules.

Creative abductions are rarely discussed in the literature. The only mechanism which according to my knowledge has been suggested in the abduction literature is *abduction from analogy*. To take Thagard's example (1988, 67), the new concept of a sound wave is achieved by analogical transfer of the already possessed concepts of *water wave* to the domain of sound. Analogical abductions are certainly very important in science, but I doubt that they can explain the discovery of all new theoretical models and concepts. How should the concepts of gravitational force, electrical force, atom and molecule, polar and non-polar bonding, acid and base, (etc.) be obtained from analogy – analogy with what? Moreover, analogical abductions have led humans more often into error than to the truth – the model of natural phenomena by intentional agents is a nice analogy, but is clearly false from the scientific viewpoint. So there is a need of a more fundamental operation of abduction by which science conjectures new concepts of theoretical models which are not obtained from analogy. I call this kind of abduction *hypothetical (common) cause abduction*, and I will elaborate it in the next sections.

3 Hypothetical (Common) Cause Abduction

Hypothetical (common) cause abduction is the essential operation of advanced explanatory reasoning. It starts at the point where the work of empirical induction has already been done. The explanandum of hypothetical (common) cause abduction consists either (a) in one phenomenon or (b) in several mutually *intercorrelated* phenomena (properties or regularities). One abductively conjectures in case (a), that the phenomenon is the effect of a hypothetical (unobservable) cause, and in case (b) that the phenomena are effects of a hypothetical *common* cause. In both cases, the abductive conjecture postulates a *new unobservable entity* (property or kind) together with *new laws* connecting it with the observable properties, without drawing on analogies to concepts with which one is already familiar. In this section I will argue that only case (b) constitutes a scientifically worthwhile abduction, while (a) is a case of pure speculation. Also Salmon (1984, 213ff) has emphasized the importance of finding common cause explanations for the justification of scientific realism. But Salmon did not inform us about the crucial difference between scientific common cause abduction and speculative (cause) abduction. I will argue that the major criterion for this distinction is *causal unification*.

Ockham's razor is a broadly accepted maxim among IBE-theorists: an explanation of observed phenomena should postulate as few unobservable or new entities or properties as possible (cf. Moser 1989, 97-100, who calls them "gratuitous entities"). After closer inspection this maxim turns into a gradual optimization criterion. For an explanation is the better, the *less* new entities it postulates, and the *more* phenomena it explains (cf. Moser's definition of "decisively better explanations" 1989, 99). But by introducing sufficiently many 'hidden entities' one can 'explain' anything one wants. Where is the borderline between 'reasonably many' and 'too many' entities postulated for the purpose of explanation? I suggest the following

(CU) *Causal Unification Criterion for Conceptually Creative Abduction*: The introduction of *one* new entity or property merely for the purpose of explaining *one* phenomenon is always speculative and ad hoc. Only if the postulated entity or property explains *many intercorrelated* but *analytically independent* phenomena, and in this sense yields a *causal* or *explanatory unification*, it is a legitimate scientific abduction which is worthwhile to be put under further investigation (cf. also Schurz/Lambert 1994, §2.3).

I first illustrate the criterion by way of examples. The simplest kind of a speculative abduction 'explains' every particular phenomenon by a special 'power' who (or which) has caused this phenomenon as follows (for '$\psi_E x$' read 'a power of kind ψ wanted E happen to x'):

Speculative Fact-Abduction :	*Example:*
Explanandum E*:* Ea	John got ill.
======================	======================
Conjecture H*:* $\forall x(\psi_E x \rightarrow Ex) \wedge \psi_E a$	Some power wanted that John gets ill, and whatever this power wants, happens.

This speculative fact-abduction schema has been applied by our human ancestors since the earliest times: all sorts of unexpected events can be explained by assuming one or several God-like power(s). Such pseudo-explanations clearly violate Ockham's razor: they do *not* offer *proper* unification, because for every event (Ea) a special hypothetical 'wish' of God ($\psi_E a$) has to be postulated (cf. Schurz/Lambert 1994, 86). On the same reason, such pseudo-explanations are entirely *post-hoc* and have *no predictive power* at all, because God's unforeseeable decisions can be known only *after* the event has already happened. In §4 it will be shown that there is a systematic connection between causal unification and increase of predictive power. Observe how my analysis differs from Kitcher's analysis (1981, 528f) who refutes the speculative fact-abduction pattern as 'spurious' unification because it is not *stringent* enough, in the sense that one may insert any sentence whatsoever for the statement Ea. But according to my suggested criterion (CU), this schema does not provide merely 'non-stringent' or otherwise defective unification – it does not provide unification *at all*.

A *Bayesian* would probably object to the criterion (CU) that there is no *real need* for it – *all* what we need is a good theory of confirmation, and this is *Bayesian confirmation theory*. To this objection I would counter that it is more based on *wishful thinking* than on truth: Bayesian confirmation theory is much *too weak* for demarcating scientifically productive from speculative abductions. Central to Bayesians is the incremental criterion of confirmation, according to which an evidence E confirms a hypothesis H iff H's posterior probability P(H|E) is greater than H's prior probability P(H). It follows from the well-known Bayes-equation $P(H|E) = P(E|H)\cdot P(H)/P(E)$ that E confirms H as long as H's prior probability P(H) is greater zero, and H increases E's probability ($P(E|H) > P(E)$), which is in particular the case if H entails E and $P(E) < 1$. This implies that (almost) *every* speculative abduction would count as confirmed. For example, that God wanted X and whatever God wants, occurs, would be confirmed by the occurrence of the event X. No wonder that philosophers of religion such as Swinburne (1979, ch. 6) suggest to confirm religious speculations using this Bayesian criterion. Although these facts are well-known by Bayesians and sometimes even regarded as a success (cf. Earman 1992, 54; Howson/Urbach 1996, 119ff; Kuipers 2000, §2.1.2), I am inclined to conclude that they imply a *breakdown* of Bayesian incremental confirmation. A Bayesian might reply that (s)he can nevertheless *gradually* distinguish between speculative and scientific explanatory hypotheses by the fact that the *prior* probability of the 'scientific' hypothesis is much higher than that of the 'speculative' one. But prior probabilities are a subjective matter, relative to one's background system of beliefs, and so this Bayesian reply ends up in the unsatisfying position that the difference between science and speculation depends merely on the *subjective prejudices* which are reflected in one's prior probabilities. In contrast, according to criterion (CU) a speculative explanation of an evidence X by a postulated 'X-wish' of God can *never* be regarded as scientifically confirmed by X alone.

A more refined but still speculative abduction schema is the following:

Speculative Law-Abduction:	*Example:*
Explanandum E*:* $\forall x(Fx \rightarrow Dx)$	Opium makes people sleepy (after consuming it).
========================	========================
Conjecture H*:* $\forall x(Fx \rightarrow \psi_D x)$ $\wedge\ \forall x(\psi_D x \rightarrow Gx)$	Opium has a special power (a 'virtus dormitiva') which causes its capacity to make one sleepy.

Speculative law-abductions of this sort have been common in the explanations of the middle ages: every special effect of a natural agens (such as the healing capacity of a certain plant, etc.) was attributed to a special power which God has implanted into nature for human's benefit. The given example of the "virtus dormitiva" had been ironically commentated by Molieré, and many philosophers have used this example as a typical instance of a vacuous pseudo-explanation (cf. Mill 1865, Book 5, ch. 7, §2; Ducasse 1974, ch. 6, §2). This abduction schema violates Ockham's principle insofar we have already a sufficient cause for the disposition to make one sleepy, namely the natural kind "opium", so that the postulated power amounts simply to a redundant multiplication of causes. More formally, the schema does not offer unification because for every elementary empirical law one has to introduce two elementary hypothetical laws to explain it (cf. Schurz/Lambert 1994, 87). Moreover, the abductive conjecture has no predictive power which goes beyond the predictive power of the explained law.

My explication of causal unification – many 'effects' explained by one or just a few 'causes' – requires formal ways to 'count' elementary phenomena, expressed by elementary statements. To be sure, there are some technical difficulties involved in this. Solutions to this problem have been proposed in Schurz (1991) and Gemes (1993). The following definition is sufficient for our purpose: a statement S is *elementary* (represents an elementary phenomenon) iff S is not logically equivalent to a non-redundant conjunction of statements $S_1 \wedge \ldots \wedge S_n$ each of which is shorter than S. Thereby, the belief system K is represented by those elementary phenomena S which are *relevant* deductive consequences of K in the sense that no (n-placed) predicate in S is replaceable by an arbitrary other (n-placed) predicate, salva validitate of the entailment K $\Vdash$ S. However, the following analysis of common cause abduction does *not depend* on this particular proposal; it merely depends on the assumption that a natural method of decomposing the classical consequence class of a belief system into nonredundant sets of *elementary* statements exists.

I do not want to diminish the value of cognitive speculation by my analysis. In fact, cognitive speculations are the predecessor of scientific inquiry. Humans have an inborn instinct to search for causes (cf. Sperber et al. 1995, ch. 3), or in Lipton's words, they are 'obsessed' with the search for explanations (1991, 130). But as it was pointed out in section 1, the best available 'explanations' are often not good enough to count as rationally acceptable. The above speculative abduction patterns can be regarded as the *idling* of human's inborn explanatory search activities when applied to events for which a proper explanation is out of reach. In contrast to these empty causal speculations, scientific common cause abductions have usually led to genuine theoretical progress. The leading principle of *causal unification* can be explicated in terms of the following principle (cf. Glymour et al. 1991, 151):

(R) *Reichenbach principle*: if two properties or kinds of events are probabilistically dependent, then they are *causally connected* in the sense that either one is a cause of the other (or vice versa), or both are effects of a common cause (where X is a cause of Y iff there leads a path of *causal arrows* from X to Y).[1]

Reichenbach's principle does not entail that every phenomenon must have a sufficient cause and, hence, avoids an empty regress of causal speculations – it merely says that all correlations result from causal connections. This principle seems to be the *rationale* which underlies humans' causal instincts. Together with *constraints* on the *causal mechanisms* underlying causal arrows, Reichenbach's principle becomes *empirically non-empty*. The way how Reichenbach's principle leads to common cause abduction is as follows: whenever we encounter several *intercorrelated phenomena*, and – on some reason or other – we can *exclude* that one causes the other(s), then Reichenbach's principle requires that these phenomena must have some (unobservable) common cause which simultaneously explains all of them. In the next section I will show that the most important scientific example of this sort is common cause

[1] A generalization of (R) is the following principle (M), applying to all triples of variables X, Y, Z: if X and Y are probabilistically dependent given Z, then it holds for the assumed underlying directed acyclic causal graph that either Z is a common effect of X and Y or there exists a causal connection between X and Y which does not go through Z". Thereby, two variables X, Y are called probabilistically dependent (relativ to a probability distribution p) iff $p(X=x_i \wedge Y=y_i) \neq p(X=x_i) \cdot p(Y=y_j)$ holds for at least some values x_i and y_j of X and Y, respectively. (M) is equivalent withof Glymour's Markov-condition (Glymour et al. 1991, 156) and with Pearl's Markov-compatibility.This follows from the fact that a probability distribution over a directed acyclic graph satisfies Pearl's Markov-compatibility iff for all (sets of) nodes X, Y, Z in the graph it holds: X is probabilistically dependent on Y given Z iff there exists an (undirected) path in the graph going from X to Y which is not d-separated by Z (cf. Pearl 2000, 16-18). For the equivalence with Glymour's Markov-condition cf. theorem 1.2.7 in Pearl (2000, 19). (M) implies Reichenbach's principle (R), and it implies Reichenbach's *screening-off criterion* (Reichenbach 1956, 159), which says that direct causes screen of indirect causes from their effects, and common causes screen off their effects from each other – where Z screens off X from Y iff X and Y are probabilistically dependent, but become independent when conditionalized to Z.

abduction from *correlated dispositions*: since dispositions cannot cause other dispositions, their correlations must have a common intrinsic cause.

The foremost way of justifying Reichenbach's principle is a kind of *no-miracle-argument*: it would be as *unplausible* as a miracle that several properties or kinds of events are persistently *correlated* without that their correlations are the result of a certain causal connection. Reichenbach's principle has been empirically corroborated in *almost every* area of science, in the sense that conjectured common causes have been identified in later stages of inquiry. Only quantum mechanics is the well-known exception. Therefore we treat the Reichenbach-principle not as a *dogma*, but as a *meta-theoretical* principle which *guides* our causal abductions.

In scientific common cause abduction, causality and unification go perfectly hand-in-hand. This is worth emphasizing insofar in the recent philosophy of science literature, causality and unification are frequently set into mutual opposition (cf. de Regt 2006). For example, Barnes (1995, 265) has put forward the following 'causal' objection against unification: it may well happen that three (kinds of) events E_i (i=1,2,3) are caused by three independent causes C_i (i=1,2,3), and although the corresponding independent explanations do not produce unification, they are certainly not *inferior* as compared to the case when all three events are explainable by one common cause C. What Barnes' example correctly shows is that because not all events have a common cause, the request for unifying explanations cannot always be satisfied. However, Reichenbach's principle allows are very simple analysis of Barnes' example: *either* (1) the three (kinds of) events are probabilistically independent; then they *cannot* have a common cause, or (2) they are probabilistically dependent; then (2.1) either they are related to each other in form of a causal chain, or (2.2) they are effects of a common cause. It is this *latter* case in which an explanation of the three E_i by three distinct C_i is clearly inferior, because, in contrast to the common cause explanation, it cannot explain the correlations between the E_i – it rather shifts this problem into unexplained correlations between the C_i.

4. Strict Common Cause Abduction from Correlated Dispositions and the Discovery of New Theoretical Kind Concepts

In this section I analyze common cause abduction in a simple *deductivistic* setting, which is appropriate when the domain is ruled by *strict* or almost-strict causal laws. Probabilistic generalizations are treated afterwards. Recall the schema of speculative law-abduction, where *one* capacity or disposition D occurring in one (natural) kind F, was pseudo-explained by a causal 'power' ψ_D. In this case of a *single* disposition, the postulate of a causal power ψ_D which mediates between F and D is an unnecessary multiplication of causes. But in the *typical* case of a scientifically productive common cause abduction, we have several (natural) kinds $F_1,\ldots,F_n$ all of which all have a set of characteristic dispositions $D_1,\ldots,D_m$ in common – with the result that all these dispositions are correlated. Given that it is excluded that one disposition can cause another one, then these correlated dispositions must be the common effects of a certain intrinsic structure which is present in all of the kinds $F_1,\ldots,F_n$ as their common cause. For example, the following dispositional properties are common to certain substances such as *iron, copper, tin, …* (cf. fig. 1): a characteristic glossing, smooth surface, characteristic hardness, elasticity, ductility, high conductivity of heat and of electricity. Already before the era of modern chemistry craftsman have abduced that their exists a characteristic intrinsic property of substances which is the common cause of all these correlated dispositions, and they have called it *metallic character* Mx.

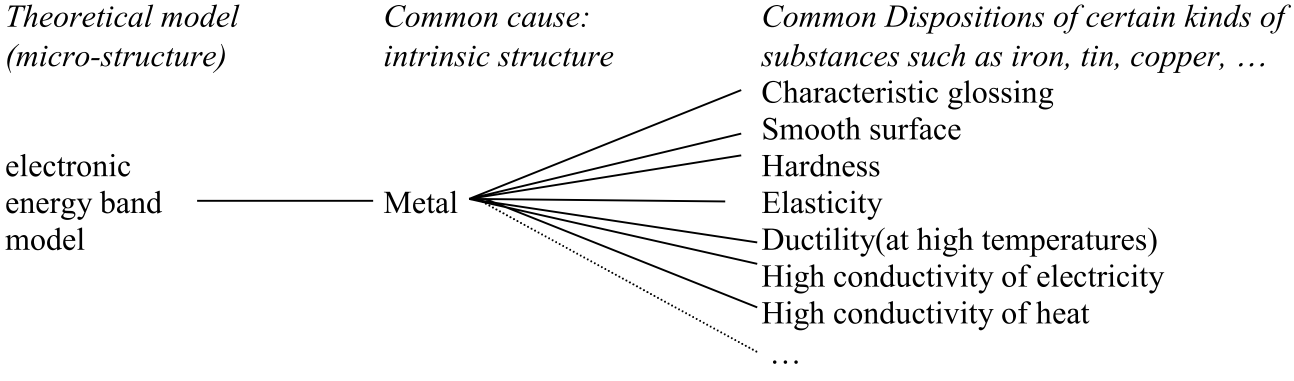

Fig. 1: Common cause abduction of the chemical kind term 'metal'

To be sure, the natural kind term *metal* of pre-modern chemistry was theoretically hardly understood. But the introduction of a new (theoretical) natural kind term is the first step in the development of a *new research programme* in the sense of Lakatos (1970). For, the next step then is to construct a *theoretical model* of the postulated kind *metal*, by which one can give an explanation of *how* the structure of a *metal* can cause all these correlated dispositions at once. Especially in *combination* with *atomic (and molecular) hypotheses* the abduced natural kind terms of chemistry became enormously fruitful. In modern chemistry, the molecular microstructure of metals is modeled as a *band* of densely layered electronic energy levels belonging to different nuclei among which the electron can shift easily around, which offers a unifying explanation of all the common dispositions of metals (cf. Octoby et al. 1999, 708ff).

In the *history of chemistry*, common cause abduction from correlated dispositions was of central importance in the discovery of new (theoretical) kinds of substances. As a second example, consider the 'paradigm' disposition of philosophers: *solubility* in water. Also this disposition does not come in isolation, but is correlated with several further dispositions, such as solubility in ammonium, non-solubility in oil or benzene, electrolytic conductivity, etc (see fig. 2). Abduction conjectures an intrinsic property as a common cause, which in early chemistry was called the *hydrophylic* ('water-friendly') character'. The corresponding theoretical model of modern chemistry are substances having electrically *polarized* chemical bonds, by which they are solvable in all fluids which have themselves polarized bonds, thereby forming weak electrostatic bondings.

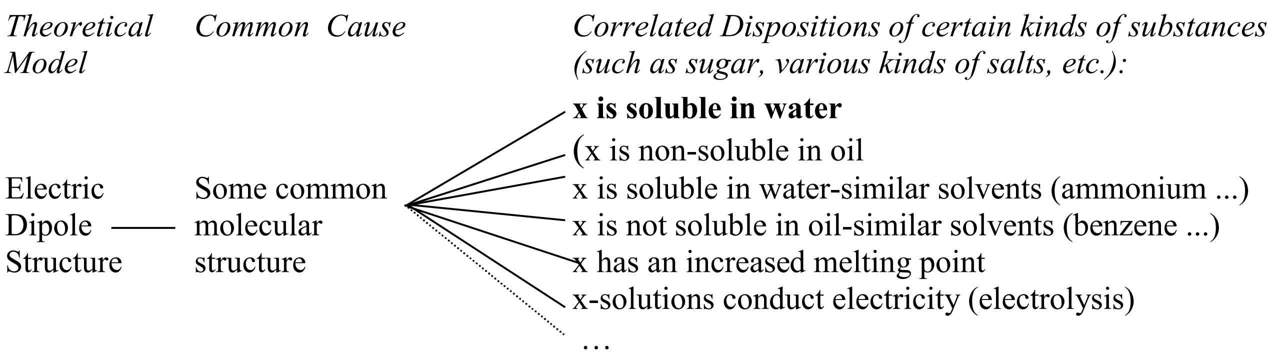

Fig. 2: Common cause abduction of the theoretical term "hydrophylic/polar" molecular structure.

The notion of disposition is discussed rather controversially in the recent literature. According to my understanding of this notion, dispositions are *conditional* (or *functional*) properties. More precisely, that an object x has a (strict) disposition D means that whenever certain initial conditions (or 'stimuli') C are (or would be) satisfied for x, then a certain reaction (or 'response') R of x will (or would) take place, or formally:

(1) $D(x) :\leftrightarrow \forall t\in\Delta(Cxt \rightarrow_n Rxt)$.

Here, $\rightarrow_n$ stands for nomological (or 'counterfactual') implication, and Δ is a more-or-less long temporal interval: if $\Delta = (-\infty,+\infty)$, the disposition is *permanent*, else it is only *temporary*. While (1) expresses a strict disposition, a merely probabilistic disposition is explicated by something like

(2): $D(x) : \leftrightarrow$ (i) $p_{t\in\Delta}(Rxt|Cxt) = \text{high}$ & (ii) $p_{t\in\Delta}(Rxt|Cxt) \geq p_{t\varepsilon\Delta}(Rxt)$

My conditional understanding of dispositions is in according with the 'received view' (cf. Carnap 1956, §IX-X; Pap 1978, 44), which has been defended by Prior, Pargetter and Jackson (1982). Dispositional properties are contrasted with *categorial* properties, which are not defined in terms of conditional effects, but in terms of 'occurrent' intrinsic structures or states (in the sense of Earman 1986, 94). Dispositional properties in this understanding have categorial properties such as molecular structures as their *causal basis*, but they are *not identical* with them. In particular, since dispositions are '2nd order properties', they can only be the effects of certain (categorial) causes, but cannot themselves act as causes (cf. Prior et al. 1982, 255; the same point has been emphasized by Ducasse 1974, ch. 6, §2).

In contrast to this view, philosophers such as Quine (1974, §3-4), Armstrong (1997, 70f) and Mumford (1998, 205) have argued that dispositions should be identified with categorial and causally effective properties, e.g. with molecular structures etc. There are two main counterarguments against the categorial view of dispositions. The first one is the *multiple realization* argument (cf. Prior et al. 1982, 253): the same disposition can be realized by *different* intrinsic structures. For example, a piece of metal and a rubber-band have both the disposition of being *elastic*, although this dispositions are caused by very different molecular properties. The second counterargument to the categorial view of dispositions is the situation of correlated dispositions just explained: if several *different* dispositions all have the same molecular structure as their common cause, then they *cannot* be identical with this molecular structure because then all of them would be mutually identical, which is counterintuitive.[2]

In conclusion, the categorial view of dispositions is not in accord the role dispositions play in science: the chemist understands dispositions such as solubility in water clearly in a *conditional* way and separates them from molecular structures which causally explain them. Only in the following *special* situation, the categorial view of dispositions has a rationale behind it: if one has one *isolated* disposition being a conditional property of an (epistemically or ontologically) *primitive* kind, then one may well identify the categorial nature of this kind with this disposition, instead of performing a *speculative* abduction and multiplying causes beyond necessity. As Molnar (1999) has pointed out, exactly this situation seems to hold in the case of *elementary particles* (electrons etc.) which are characterized by fundamental dispositions (electric charge etc.) without any further causal explanations for them. So at the fundamental levels of physics there may well be causally ungrounded dispositions. But in all higher levels of science one finds mutually correlated dispositions having a common causal basis – and I argue that this situation gives us a *clear reason* to distinguish between conditionally understood dispositions on the one side and their common causal basis on the other side.

A final remark: when I speak of a molecular structure as being the *cause* of a disposition, I understand notion of "cause" in a more general sense that the narrow notion of causation between temporally separated events. My usage of "cause" fits well with ordinary and scientific usage. For the more scrupulous philosopher of causation,

[2] This second counterargument is also a problem for Mumford's "token-identity" view, which would us force to say that this instance of electric conductivity is identical with this instance of elasticity, because both instances are identical with this instance of gold (cf. Mumford 1998, 163).

let me add that may extended usage of "cause" is *reducible* to the notion of event-causation as follows: a disposition Dx, being defined as the conditional property $\forall t \in \Delta(Cxt \rightarrow_n Rxt)$, is *caused* by a categorial property Sx iff each *manifestation* of the disposition's reaction, Rxt, is caused by Sx together with the initial conditions Cxt, or formally, iff $\forall x \forall t \in \Delta(Sx \wedge Cxt \rightarrow_n Rxt)$.[3]

The structural pattern of the two examples (fig. 1+2) can be formalized as follows:

Common cause abduction (abduced theoretical concept: ψ):

Explanandum E: All kinds $F_1,\ldots,F_n$ have the dispositions $D_1,\ldots,D_m$ in common.
$\forall i \in \{1,\ldots,n\} \forall\, j \in \{1,\ldots,m\}: \forall x(F_i x \rightarrow D_j x)$.
==
Abductive conjecture H: All F_1s,…,F_ns have a common intrinsic and structural property ψ which is a sufficient [and necessary] cause of all the dispositions $D_1,\ldots,D_m$.
$\forall i \in \{1,\ldots,n\}: \forall x(F_i x \rightarrow \psi x) \ \wedge \forall j \in \{1,\ldots,m\}: \forall x(\psi x \rightarrow [\leftrightarrow]\ D_i x)$.

The abductive conjecture H logically implies E and it yields a *unification* of n·m empirical (elementary) laws to n+m theoretical (elementary) laws, which is a *polynomial reduction* of elementary laws. H postulates the theoretical property ψx as a merely *sufficient* cause of *all* of the dispositions. If we assume that the dispositions are strictly correlated, then the abductive conjecture even postulates that ψx is both a necessary and sufficient cause of the dispositions (see the version in brackets "[$\leftrightarrow$]"). Note that the given explanandum E would also allow for the possibility that the correlated dispositions $D_1,\ldots,D_m$ have in each kind F_i a *different* common cause ψ_i – but of course, the much *more probable* hypothesis is to assume that they have in all kinds F_i one and the same common cause ψ. On this reason, every single application of this

---

[3] In this way, also the common-cause-explanation for correlated dispositions $D_1x,\ldots,D_nx$ can be reduced to common cause explanations of correlated events $R_i x t_i$ given $C_i x t_i$, where $t_1,\ldots,t_n$ are different time points in the interval Δ at which the different initial conditions have been realized). The common cause $Sx := \forall t \in \Delta: Sxt$ was present during the given temporal interval Δ and thus figures as a common cause of all the conditional events $C_i x t_i$ given $R_i x t_i$: the screening-off relation $P(C_i x t_i \wedge C_j x t_j | R_i x t_i \wedge R_j x t_j \wedge (\neg)Sx) = P(C_i x t_i | R_i x t_i \wedge (\neg)Sx) \cdot P(C_j x t_j | R_j x t_j \wedge (\neg)Sx)$ (cf. fn. 1) is satisfied for all pairs of distinct dispositions-mainfestations $i \neq j \in \{1,\ldots,n\}$.

kind of abduction introduces a *new natural kind*: the class of 'ψ-bearers' (e.g. the class of metals, the class of polar substances, etc.)

In conclusion, common cause unification has (at least) three virtues:

(1.) The *intrinsic virtue of unification*. Many elementary phenomena (statements) are explained by a few basic principles. Several philosophers, though, are inclined to think that this virtue is merely *instrumentalistic* and, hence, rather weak.

(2.) The virtue of *leading to new predictions*. This may happen in several ways. For example, if we know for some of the kinds $F_1,\ldots,F_n$, say for F*, that it possesses *some* of the dispositions, then the abduced common cause hypothesis predicts that F* will also possess all the other dispositions. Or, if we know in addition of some independent indicator G for the theoretical property ψ (i.e. $\forall x(Gx \rightarrow \psi x)$, then this knowledge together with the common cause hypothesis predicts G to be an indicator for all of the dispositions D_j. Finally, if ψ is conjectured as being sufficient and necessary for all of the D_j ["$\leftrightarrow$]", then this strengthened hypothesis predicts that all the D_j are mutually strictly correlated ($\forall i{\neq}j \in \{1,\ldots,m\}$: $D_ix \leftrightarrow D_jx$). In contrast to speculative abductions, common cause abduction are *independently testable* because of their virtue of producing new predictions.

(3.) The virtue of *discovering new (unobservable) kinds or properties* which enlarge our *causal understanding*. This is not only of theoretical, but also of *practical importance*, since knowing a disposition's cause is a necessary step for its technical utilization. Since Reichenbach's causality principle does not hold in every domain (e.g., not in quantum mechanics), there is no guarantee that the hypothetical entities postulated by common cause abduction will always have realistic reference. Nevertheless the following methodological justification can be given: wherever unobservable common causes of observable correlations exist, common cause abduction will find them, while where they don't exist, our efforts to find independent evidence for common causes will fail, and sooner or later we will adopt an instrumentalistic view of our explanatory unification attempts (see § 5).

As a further example of common cause abduction in the natural sciences, consider the formation of the theoretical kind terms *acids*, *bases*, and *salts* which was basically achieved by Glauber in the 17th century. Langley at all (1987, 196ff) have given an algorithmic reconstruction of Glauber's discovery which can be reconstructed as an abduction to an explanation of the following regularities among chemical reactions:

Abduction of the acid-base-salt-system (Glauber):

There exists a family of substances $F_1,\ldots,F_n$ and another family of substances $G_1,\ldots,G_m$ which have the following in common:
(i) Common chemical reactions:
$\forall i \in \{1,\ldots,n\}$, $j \in \{1,\ldots,m\}$: F_i und G_j reacts into a characteristic product (F_i,G_j) and water. (Chemical example: if F_1 is the acid HCl and G_1 is the lime NaOH, then (F_1,G_1) stands for NaCl, etc.)
(ii) *Common monadic properties:*
$\forall i \in \{1,\ldots,n\}$: F_i has characteristic qualities A (for example, it tastes sour)
$\forall j \in \{1,\ldots,m\}$: G_j hat characteristic qualities B (for example, it tastes bitter)
$\forall i \in \{1,\ldots,n\}, j \in \{1,\ldots,m\}$: (F_i,G_j) has qualities C (for example, it tastes salty).
==
Abductive conjectures (abduced theoretical terms: 'acid', 'base', 'salt'):
(a) Every F_i is an acid. (b) Every G_j is a lime.
(c) An acid X reacts with a lime Y into the combination (X,Y) and water, where the acid-base-combination (X,Y) is *by definition* called a salt.
(d) An acid has qualities A. (e) A lime has qualities B. (f) A salt has qualities C.

The abductive conjectures deductively entail the empirical phenomena (i) and (ii), and thereby they reduce $2\cdot n\cdot m + n + m$ empirical (elementary) laws to $n + m + 4$ theoretical (elementary) laws.

As a final example we consider *Newtonian physics*. Its the fundamental common cause abduction was the abduction of the *sum-of-all-forces* as a common cause for all kinds of accelerations, and the abduction of a universal *gravitational force* as a common cause of the different kinds of movements of bodies in the sky as well as on earth. Thereby, Newton's *qualitative* stipulation of the gravitational force as the counterbalance of the centrifugal force acting on the circulating planets was his *abductive* step, while his *quantitative* calculation of the mathematical form of the gravitational law was a deduction from Kepler's third law plus his abductive conjecture.[4]

Common cause abduction can also be applied to ordinary, non-dispositional properties or (kinds of) events which are correlated. However, in this case one has first to consider more parsimonious causal explanations which do not postulate an unobservable common cause but stipulate one of these events or properties to be the cause of the others. For example, if the three kinds of events F, G, and H (for example, eating a certain poison, having difficulties in breathing and finally dying) are strictly correlated and always occur in form of a temporal chain, then the most parsimonious conjecture is that these event-types form a causal chain. Only in the special case where two (or several) correlated event-types, say F and G, are strongly correlated, but our causal background knowledge tells us that there *cannot* exist a direct causal mechanism which connects them, then a common cause abduction is the most plausible conjecture. An example is the correlation of lightning and thunder: we know by induction from observation that light does not produce sound, and hence, we conjecture that there must exist a common cause of both of them. I call this special case a *missing link common cause abduction*.[5]

5. Probabilistic Common Cause Abduction and Statistical Factor Analysis

Statistical factor analysis is an important branch of statistical methodology whose analysis (according to my knowledge) has been neglected by philosophers of science.[6] In this section I want to show that factor analysis is a certain generalization of

---

[4] Glymour (1981, 203ff) gives illuminating details on Newton's reasoning.

[5] If the correlations between the events are not strict but merely probabilistic, then one may also use Reichenbach's *screening off* criterion (see fn. 1) to distinguish between the case were one of the events E_i causes the other ones from the case where the E_i are effects of a common cause. If the correlations are strict, Reichenbach's screening-off criterion does not work (cf. Otte 1981).

[6] Recently I have discovered an exception, namely Haig (2005). He shares my view of factor analysis as an elaboration of statistical common cause abduction.

hypothetical common cause abduction, although sometimes it may better be interpreted in a purely instrumentalistic way. For this purpose I assume that the parameters are now represented as statistical random variables X, Y, …, each of which can take several values x_i, y_j. (A random variable $X:D\rightarrow|R$ assigns to each individual d of the domain D a real-valued number X(d); a dichotomic property Fx is coded by a binary variable X_F with values 1 and 0.) The variables are assumed to be at least interval-scaled, and the statistical relations between the variables are assumed to be monotonic – only if these conditions are satisfied, the *linearity* assumption of factor analysis yields good approximations.

Let us start from the example of the previous section, where we have n empirically measurable and highly intercorrelated variables $X_1,\dots,X_n$, i.e. $cor(X_i,X_j)$ = high for all $1\leq i,j\leq n$. An example would be the scores of test persons in n different intelligence tests. We assume that none of the variables screens off the correlations between any other pair of variables ($cor(X_i,X_j|X_r) \neq 0$), so that by Reichenbach's principles (cf. fn. 1) the abductive conjecture is plausible that these n variables have a common cause, distinct from each of the variables – a theoretical factor, call it F. In our example, F would be the theoretical concept of intelligence. Computationally, the abductive conjecture asserts that for each $1\leq i\leq n$, X_i is approximated by a linear function f_i of F, $f_i(F(x)) = a_i\cdot F(x)$, for all individuals x in the given finite sample or domain D (since we assume the variables X_i to be z-standardized, the linear function f_i has no additive term "+ b_i"). The true X_i-values a scattered around values predicted by this linear function, $f_i(F)$ by a remaining random dispersion s_i; the square s_i^2 is the *remainder variance*. According to standard linear regression technique, the optimally fitting coefficients a_i are computed such as to *minimize* this remainder variance – which is mathematically equivalent to maximizing the variance of the F-values, $v(F) := \Sigma_{x\in D} (f(x) - \mu(F))^2 /|D|$.[7] Visually speaking, the X_i-values form a stretched cloud of points

[7] This follows from the fact that by the additivity of the variances it holds that the total variance $v(X_i)$ of a variable X_i equals the sum of the variance of X_i's predicted values $v(a_i\cdot F) = a_i^2\cdot v(F)$ plus the remaineder variance s_i^2 (cf. Bortz 1985, 233).

in an n-dimensional coordinate system, and F is a straight line going through the middle of the cloud such that the squared normal deviations of the points to the straight line are minimized.

So far we have described the linear-regression-statistics of the abduction of *one* factor or cause. In factor analysis one takes additionally into account that the mutually intercorrelated variables may have not only one but *several* common causes. For example, the variables may divide into two subgroups with high correlations within each subgroup, but low correlations between the two subgroups. In such a case the abductive conjecture is reasonable that there are two independent common causes F_1 and F_2, each responsible for the variables in one of the two subgroups. In fact, while Spearman had suggested in 1904 one factor being responsible for intelligence, some years later other authors, for example Burt, showed that intelligence is be better explained by several independent cognitive factors (Bortz 1985, 618; Cureton/d'Agostino 1983, 8ff). In the general picture of factor analysis there are given n empirical variables X_i which are explained by $k < n$ theoretical factors (or common causes) F_j as follows:[8]

$$X_1 = a_{11}\cdot F_1 + \ldots + a_{1k}\cdot F_k + s_1.$$

$$\ldots$$

$$X_n = a_{n1}\cdot F_1 + \ldots + a_{nk}\cdot F_k + s_n.$$

This is usually written in a matrix formulation: $\mathbf{X} = \mathbf{F}\cdot\mathbf{A}'$. While each variable X_i and factor F_j takes different values for the different individuals of the sample, the factor loadings a_{ij} are constant and represents the causal contribution of factor F_j to variable X_i. Given that also the factor variables F_j are standardized, then each *factor loading* a_{ij} expresses the correlation (covariance) between variable X_i and factor F_j, $cor(X_i,F_j)$. Since the variance of each variable X_i equals the sum of the squared factor loadings a_{ij}^2 and the remainder variance s_i, each squared factor loading a_{ij}^2 measures the

[8] For the following cf., e.g., Bortz (1985, ch. 15.1), Kline (1994, ch. 3) or Gorsuch (1983, ch. 6). I only describe the most common method of factor analysis which is used in most computer programs, without discusssing the subtle differences between different factor analytic methods.

amount of the variance of X_i 'explained' (i.e. statistically predicted) by factor F_j and coincides with the so-called *eigenvalue* λ_j of the factor f_j. The sum of all squared factor loadings divided through n measures the percentage of the total variance of the variables which is explained by the factors – this percentage is direct measure for the explanatory success of the factor-statistical analysis.

The major mathematical technique to find those $k < n$ factors which explain a maximal amount of the total variance is the so-called *principal component analysis*. Instead of any detailed mathematical explanation I confine myself to the following remarks. The k factors or axes are determined according to two criteria: (i) they are probabilistically independent (or orthogonal) to each other, and (ii) the amount of explained variance is maximized (i.e., the remainder variances are minimized). Visually speaking, the first the factor F_1 is determined as an axis going through the stretched cloud of points in the n-dimensional coordinate system; then the next factor F_2 is determined as an axis orthogonal to F_1, and so on, until the $k < n$ factor axes are determined by the system of coefficients a_{ij}. Mathematically this is done by rotations of the coordinate system in a way such that the diagonal elements of the $n \times n$ variance-covariance matrix **D** are maximized under the constraint of orthogonality; this leads to the matrix equation $(\mathbf{D} - \lambda \cdot \mathbf{I}) \cdot \mathbf{v} = \mathbf{0}$ with unknowns λ and $\mathbf{v}$ ($\mathbf{I}$ identity matrix, λ the vector of eigenvalues, so $\lambda \cdot I$ is a $n \times n$ diagonal matrix containing the eigenvalues; $\mathbf{v}$ is the rotation matrix). The first k eigenvalues λ_j and corresponding rotation vectors $\mathbf{v}_j$ obtained from solving this equation determine the system of factor loadings a_{ij}.

The success of an explanation of n variables by $k < n$ factors is the higher, the less the number k compared to n, and the higher the amount of the total variance explained by the k factors. Note that the amount of explained variance goes hand in hand with the amount of the intercorrelation between the variables which is explained by the factors: the amount of the correlation cor (X_i, X_j) explained by all factors is given as $\Sigma_{1 \leq j \leq k}\, a_{ij} \cdot a_{kj}$ (cf. Bortz 1985, 661; Kline 1994, 40). This picture fits perfectly with my account of unification of a given set of n empirical variables by a small set of k theoretical variables, as explained in §4.. While the amount of explained variance of the first factor is usually much greater than one, this amount becomes smaller and smaller when one introduces more and more factors. The trivial limiting case is given when k=n, because n empirical variables can always completely be explained by n theoretical factors. According to the Kaiser-Guttman-criterion one should introduces new factors only as long as their amount of explained variance is greater than one (cf. Bortz 1985, 662f; Kline 1994, 75). Hence, a theoretical factor only is considered as non-trivial if it explains more than the variance of just one variable and, in this sense, offers a unificatory explanation to at least *some* degree. In other word, the Kaiser-Guttman-criterion is the factor analytic counterpart of my suggested minimal criterion for hypothetical cause abduction (CU).[9]

After the principal component analysis has been performed and the factors have been standardized, the factor axes can be *rotated* without change of the amount of explained variance. So the result of a factor analysis is not unique. According to the most common *varimax* principle, the factor axes are rotated into a position in which the square loadings of the factors are, roughly speaking, either very high or very low (cf. Bortz 1985, 665-672; Kline 1994, 67f). This leads to the effect that the abduced (or 'extracted') factors can most easily interpreted in terms of certain plus-minus-combinations of the empirical variables. Fig. 3 offers an example which is taken from Bortz (1985, 672):

Variables:	*Varimax factors (their interpretations), and loadings (bold = high)*			
	F_1 (dynamics)	F_2 (emot. value)	F_3 (conciseness)	*E.V. per variable in %*
1: loud-low	**0.84**	-0.08	-0.17	73
2: harmonious-disharm.	-0.26	**0.80**	-0,22	75
3: clear-unclear	0.42	0.03	**-0,86**	91
4: fluent-haltering	0,48	0.45	-0.30	52
5: slow-quick	**-0,86**	0.29	0.07	82

9 The Kaiser-Guttman-criterion is an adequate minimnal critertion if one is interested in extracting *non-trivial* common cause factors. A stronger criterion is Catell's *scree test* which in most cases extracts factors whose explained variance is significantly greater than one (cf. Bortz 1985, 662f, Gorsuch 1983, 167ff; Kline 1994, 76). If the intercorrelations between the variables are small and one is interested in representing all correlations by common factors, then one has to extract also factors whose explained variance is smaller than 1 (see Gorsuch 1983, 163ff).

6: articulated-vague	0.28	0.24	**-0,88**	91
7: pleasant-unpleas.	-0,31	**0.86**	-0,21	88
8: activ-passiv	**0,95**	0.06	-0,23	95
9: strong-weak	**0.67**	**0.66**	-0.17	91
10: deep-high	0.41	**0.80**	0.12	81
11: confident-bashful	**0.69**	0.50	-0,30	81
12: inhibited-free	0.06	**-0,85**	0.27	80
13: quiet-lively	**-0.90**	-0.25	0.03	87
14: hesitating-pressing	**-0.94**	0.06	0.08	90
15:correct-careless	0.01	0.22	**-0.88**	82
16: engaged-tired	**0.93**	0.07	-0.11	88
17: bit-little	0.04	**0.94**	0.11	89
18: ugly-nice	0.17	**-0.84**	0.28	80
E.V. per factor:	37%	30,4%	15,0&	*Total E.V.*: 83,3%

Fig. 3: 18 empirical variables measuring the subjective evaluations of the voices of persons explained by three factors F_j . E.V. = explained amount of the variance (taken from Bortz 672).

Prima facie, hypothetical common cause abduction supports a realistic interpretation of the abduced factors. In contrast, for an *instrumentalistic* philosopher of science such as van Fraassen (e.g. 1980), the extracted factors are not taken realistically, and so the factor equations cannot be true in the realistic sense. They can only be more-or-less *empirically adequate*. For the instrumentalist an *abduction pattern* is a useful means of *discovering* an empirically adequate theory – it has an important *instrumental* value, but it does not have any *justificational* value. For judgments of empirical adequacy, an abductive inference is not needed – an *epistemic induction* principle is sufficient which infers the empirical adequacy of a theory (and hence its future empirical success) from its empirical success in the *past*.

In fact, several statisticians tend to interpret the results of a factor analysis cautiously as a merely instrumentalistic means of *data reduction* in the sense of representing a large class of intercorrelated empirical variables by a small class of independent theoretical variables. In spite of this fact I think that the properly intended interpretation of the factors of a factor analysis is their realistic interpretation as common causes, for that is how they are designed. I regard the instrumentalistic perspective as an *important warning* that not every empirically useful theoretical superstructure must correspond to an existing structure of reality (cf. also Cureton/d'Agostino 1983, 3f). This warning is already entailed by the mentioned fact that the results of a factor analysis are *non-unique modulo rotations* of the standardized factor-axes. Haig (2005, 319) considers this fact as the factor-analytic counterpart of the general situation of *empirical underdetermination* of theories by empirical evidence in science.

6. Theoretical Model Abduction

Every successful fundamental common cause abduction in the natural sciences has been a *germ f*or a new theoretical research programme in the sense of Lakatos (1970), in which scientists attempt to develop theoretical and quantitative *models* for their conjectured common cause. For example, chemical kind concepts get replaced by molecular models, or qualitative force concepts by quantitative equations. In this way, fundamental common cause abduction turns gradually into what I call *theoretical model abduction*. The capacity of producing *novel predictions* is significantly enhanced by this transformation. This concluding section of my paper provides an analysis of theoretical model abduction. Theoretical model abduction takes also place in the social sciences, for example as emerging from the results of a factor analysis. However, theories in social sciences are formulated usually only in a qualitative form, without precise equations from which quantitative predictions could be derived. I don't know whether this is only due to complexity reasons (the 'objects' of social sciences are enormously complex), or on reasons if principle – but anyhow, it seems to me that this fact reflects a major contemporary difference between theories of natural and social sciences.

Let us first consider typical examples from the natural sciences. The explanandum of a theoretical-model abduction is a well-confirmed empirical phenomenon expressed by an *empirical law* – for example, the phenomenon that wood swims in water but a stone sinks in it. The abduction is driven by an *already established* scientific theory which is usually quantitatively formulated. The abductive task consists in *find-*

ing theoretical (initial and boundary) conditions which describe the causes of the phenomenon in the theoretical language and which allow the mathematical derivation of the phenomenon from the theory.[10] Formally, these theoretical conditions are expressed by factual or lawlike statements, but their semantic content corresponds to what one typically calls a *theoretical model* for a particular kind of phenomenon within an already *given* theory, whence I speak of 'theoretical-model abduction'. Note also that with my notion of a 'model' I do not imply a particular kind of formalization of models: they can be represented by *statements* as well as by *set-theoretical* models (which in turn are characterized by statements of a set-theoretical meta-language).

The theory itself is the historical outcome of a scientific common cause abduction which identifies the qualitative kind of causes in terms of which the phenomenon has to be explained. As an example, consider Archimedes' explanation of the phenomenon of buoyancy. Here one searches for a theoretical explanation of the fact that certain substances like stones or metals sink in water while others like wood or ice swim on water, *solely in terms of mechanical and gravitational effects*. Archimedes' ingenious abductive conjecture was that the amount of water which is supplanted by the swimming or sinking body tends to lift the body upwards, with a force f_W which equals the weight of the supplanted water (see fig. 4). If this force is greater than the

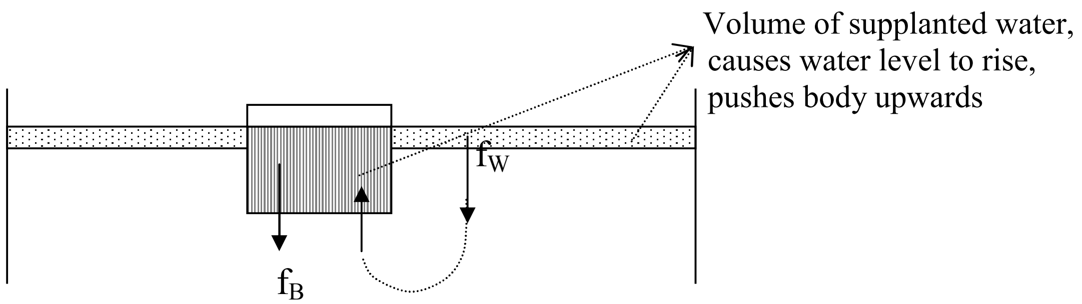

Fig. 4: Theoretical conditions which allow the mechanical derivation of the law of buoyancy.

weight of the body (f_B) the body will swim, otherwise it will sink. Since the volume

[10] Cf. Halonen/Hintikka (2005, §3), who argue that this task makes up the essential point the scientist's explanatory activity.

of supplanted water equals the volume of the part of the body which is under water, and since the weight is proportional to the mass of a body, it follows that the body will sink exactly if its density (mass per volume) is greater than the density of water.

The example shows clearly that this kind of abduction is tantamount to the formation of a *theoretical model* for a given kind of lawlike phenomenon within a *given* theory. This situation is very different from *selective* factual abductions: one does *not* face here the problem of a huge multitude of possible theoretical models or conjectures. Rather, the given theory *constrains* the space of possible causes to a small class of basic parameters (or generalized 'forces') by which the theory models the domain of phenomena which it intends to explain. In the Archimedean case, the given theory presupposes that the ultimate causes are only contact forces and gravitational forces – other ultimate causes such as intrinsic swimming capacities of bodies or invisible water creatures etc. are excluded. Therefore, the real difficulty of theoretical model-abduction does not consist in the elimination of possible explanations (this elimination is already achieved by the given theory), but to find just *one* plausible theoretical model which allows the derivation of the phenomenon to be explained. If such a theoretical model is found, this is usually celebrated as a great scientific success.

Theoretical model-abduction is the typical theoretical activity of *normal* science in the sense of Kuhn (1967), that is, the activity of extending a given theory core (or paradigm) to new application cases, rather than changing a theory core or creating a new one. If the governing theory is *classical physics*, then examples of theoretical model abductions come into hundreds, and physics text books are full of them. Examples are the theoretical models underlying

(1.) the trajectories (paths) of rigid bodies in the constant gravitational field of the earth (free fall, parabolic path of ballistic objects, gravitational pendulum, etc.);

(2.) the trajectories of cosmological objects in position-dependent gravitational fields (the elliptic orbits of planets – Kepler's laws, the moon's orbit around the earth and the lunar tides, inter-planet perturbations, etc.);

(3.) the behaviour of solid, fluid or gaseous macroscopical objects viewed as systems of more-or-less coupled mechanical atoms (the modeling of pressure, friction, viscosity, the thermodynamic explanation of heat and temperature, etc.); and finally

(4.) the explanation of electromagnetic phenomena by incorporating electromagnetic forces into classical physics (cf. Halonen/Hintikka 2005, §3).

While for all other kinds of abductions we can provided a *general* formal pattern and algorithm which by which one can *generate* a most promising explanatory hypothesis, we cannot provide such a general pattern for theoretical model abduction because here all depends on *what theory* we are in. But if the theory is specified, then such patterns can often be provided: they are very similar to what Kitcher (1981, 517) has called a schematic explanatory argument, except that the explanandum is now given and the particular explanatory premises have to be found within the framework of the given theory. For example, the abduction pattern of Newtonian particle mechanics would be something like the following:[11]

Explanandum: a kinematical process involving (a) some *moving* particles whose position, velocity and acceleration at a variable time t is an *empirical function* of their initial conditions, and (b) certain objects defining constant boundary conditions (e.g. a rigid plane on which a ball is rolling, or a large object which exerts a gravitational force, or a spring with Hooke force, etc.)

==

Generate the abduced conjecture as follows: (i) specify for each particle its mass and all non-neglectible forces acting on it in dependence on the boundary conditions and on the particle's position at the given time; (ii) insert these specifications into Newton's 2nd axiom (which says that for each particle x and time t, sum-of-all-forces-on-x-at-t = mass-of-x times acceleration-of-x-at-t); (iii) try to solve the resulting system of *differential equations*; and finally (iv) check whether the resulting time-dependent trajectories fit the empirical function mention in the explanandum − if yes, the conjecture is preliminarily confirmed; if no, then search for (perturbing) boundary conditions and/or forces which may have been overlooked.

---

[11] The suggested pattern is more general than that given by Kitcher (1981, 517) which is merely formulated for one particle under one force.

Theoretical model abduction can also be found in 'higher' sciences which are working with explicitly formulated theories. In *chemistry*, the explanations of the atomic component ratios (the chemical gross formulae) by a three-dimensional molecular structure are the results of theoretical model abductions; the given theory here is the periodic table plus Lewis' octet rule for forming chemical bonds.

Theoretical model abductions take also place in *evolutionary theory*. For example, the reconstruction of phylogenetic trees of descendance from phenotypic similarities (and other empirical data) is a typical abduction process. The basic evolution-theoretical premise here is that different biological species descend from common biological ancestors from which the have split apart by discriminative mutation and selection processes. The alternative abductive conjectures about trees of descendance explaining given phenotypic similarities can be evaluated by probability considerations. Assume three species S_1, S_2, and S_3 where both S_1 and S_2 but not S_3 have a *new* property F – in Sober's example, S_1 is sparrows, S_2 = robins, S_3 = crocs, and F = having wings (Sober 1993, 174-176). Then the tree of descendance T_1 where the common ancestor A first splits into S_3 and the common ancestor of S_1 and S_2 which has already F, requires only one mutation-driven change of non-F into F, while the alternative tree of descendance T_2 in which A first splits into S_1 and a common F-less ancestor of S_2 and S_3 requires two such mutations (see fig. 5).

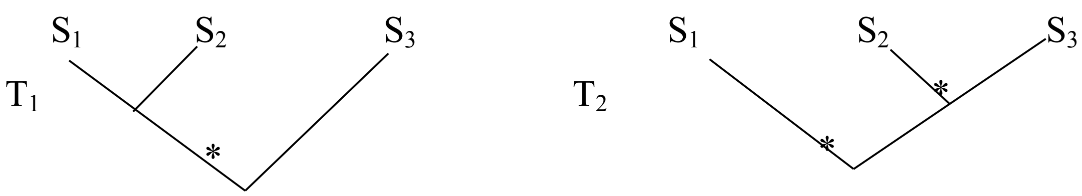

*Fig. 5: Two alternative trees of descendance; * = mutation of non-F into F.*

So probabilistically T_1 is favored as against T_2. There are some well-known examples were closeness of species due to common descent does *not* go hand in hand with closeness in terms of phenotypic similarities (e.g., *birds*, *crocs* and *lizards*): examples of this sort are recognized because there are several mutually *independent* kind of evidences which the tree of descendance must *simultaneously* explain, in particular (i) phenotypic similarities, (ii) molecular similarities, and (iii) fossil record (cf. Ridley 1993, ch. 17).

An example of qualitative model-abduction in the area of humanities is *interpretation* (an illuminating analysis is found in Gabbay/Woods 2005, §4.1). The explanandum of interpretations are the utterances, written text, or the behaviour of given persons (speakers, authors, or agents). The abduced models are conjectures about the beliefs and intentions of the given persons. The general background theory is formed by certain parts of (so-called) folk psychology, in particular the general premise of all rational explanations of actions, namely, that normally or ceteris paribus, persons act in a way which is suited to fulfill their goals given their beliefs about the given circumstances (cf. Schurz 2001, §1). More specific background assumptions are hermeneutic rationality presumptions (Davidson 1984), Grice's maxims of communicative cooperation (Grice 1975), and common contextual knowledge. Interpretative abductions may both be selective or creative: in the case of interpretations, the question whether there will be many possible interpretations and the difficulty will be their elimination, or whether it will be hard to find just one coherent interpretation, depends crucially on *what* the speaker says and *how* (s)he says it.

As a final example I discuss the theoretical model of cultural development which Inglehart has developed based on his factor-analytic analysis of the data of the *Word Value Survey (WVS) project* (see *http://wvs.isr.umich.edu/index.html*). In this project, social scientists have developed and successively improved detailed questionnaires, by which people in more than 65 countries are asked questions concerning their *attitudes* (e.g. towards religion, authority, labor, family, political organization, personal self realization, gender roles, homosexuality, liberty etc.) as well as concerning their *lifestyle* (married or single, how much time they spend for work, family, clubs, etc.).

Each answer to a question corresponds to one *empirical variable*. In order to interpret the more than 250 variables, Inglehart has performed a factor analysis. In several independent WVS-studies or 'waves' (1990, 1995, 2000, 2005), two stable (statistically independent) factors have been obtained which explain more than 50% of the total variance of the empirical variables. The factor analysis has been performed with nations as empirical units, i.e., with the mean values for each country. The two obtained factors have led Inglehart to the following theoretical interpretation of two major causes of the cultural characteristics of nations:

Factor 1: tradition-religious versus *secular-rational orientation*. This factor loads correlates strongly with: religious vs. rational-secular value-orientation, importance of family bonds vs. individual freedom, national pride high vs. low, respect vis-à-vis state authority high vs. low, birth rate high vs. low, estimation of labor high vs. low. The transition on the 1st factor axis from traditional-religious to rational-secular values is theoretically interpreted as the social transition from *agraic* to *industrial* civilizations – the so-called *process of modernization.*

Factor 2: survival values versus self-expression values. This factor correlates strongly with*:* low vs. high economic standards of life, importance of existential security & wok vs. pleasure & life-quality, low vs. high appreciation of gender equality, rejection vs. acceptance of homosexuality, intolerance vs. tolerance towards foreign immigrants, low vs. high interpersonal trust, acceptance vs. non-acceptance of authoritarian regimes, devaluation vs. appreciation of democracy. The transition on the 2nd factor axis from survival to self-expression values is theoretically interpreted as the transition from industrial to post-industrial societies with a high prosperity level and a dense social service infrastructure – the so-called *process of post-modernization*.

In his earlier books (cf. 1990), Inglehart defended a version of modernization theory in which more-or-less all nations follow in their development the same major modernization trends, going from agrarian to industrial to post-industrial societies. Because of the WVS-data, Inglehart later weakens his view and recognizes a large amount of *culture-specific path-dependence* in this development (cf. Inglehart/Baker 2000, 21). Fig. 6 shows the most recent cultural world map of nations based on the WVS-data of 2000 and 2005 (they include more nations than 1995).

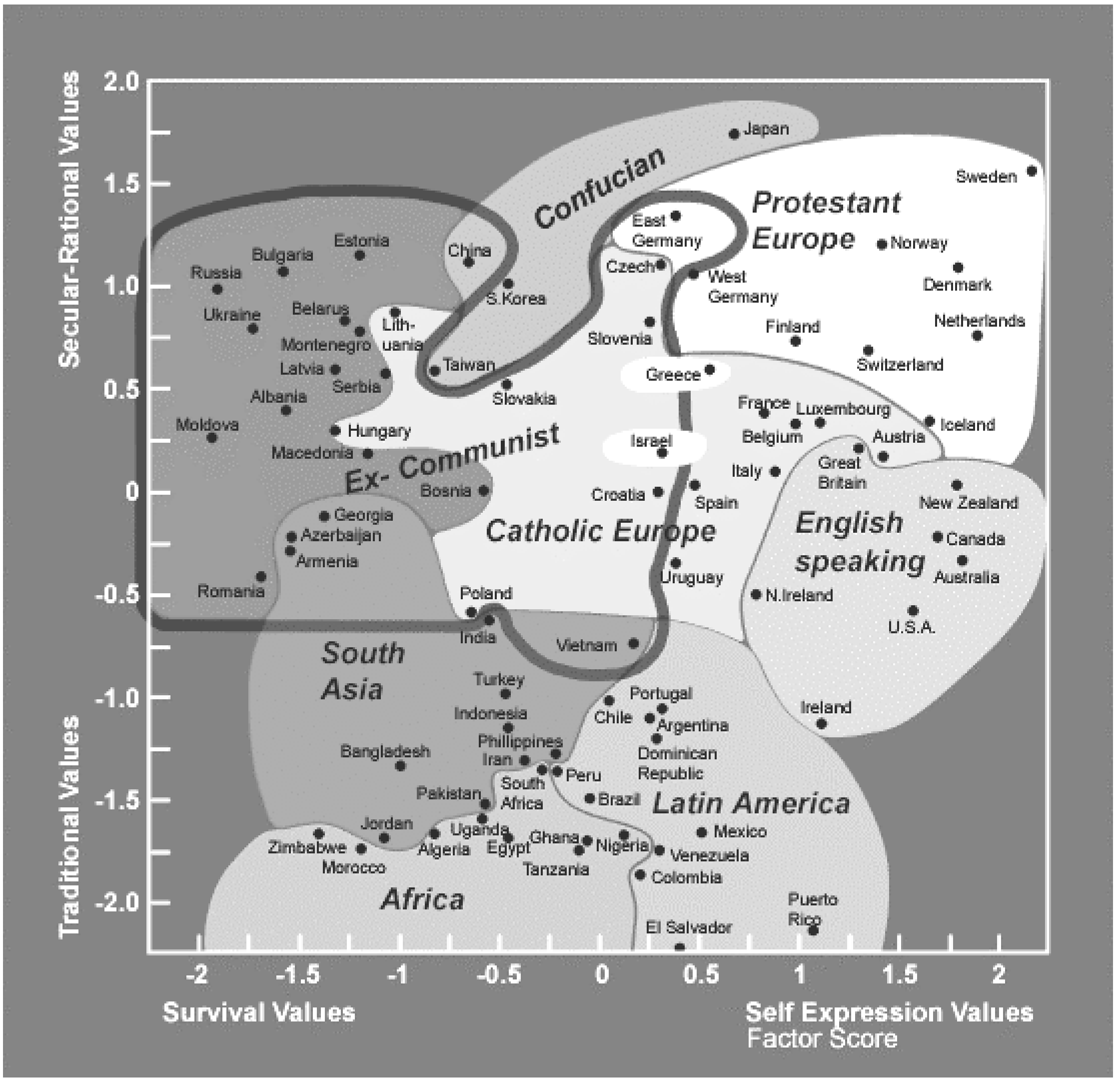

Fig. 6 The cultural world-map. Most recent data from 2000 and 2005.
(Source: http://wvs.isr.umich.edu/index.html)

Fig. 6 and related results have led to a variety of interesting questions and discoveries, which cannot be discussed here (for details cf. Inglehart/Baker 2000, Norris/Inglehart 2004, Schurz 2007a). In the perspective of this paper, Inglehart's work shows how the results of factor analysis in social science can lead to the abduction of fruitful theoretical models even in so complex areas the cultural development and cultural divergence of world-wide nations.

References

Armstrong, D.M. (1969): "Dispositions as Causes", *Analysis* 30, 23-26.

Barnes, E. (1995): "Inference to the Loveliest Explanation", *Synthese* 103, 251-277.

Bortz, J. (1985): *Lehrbuch der Statistik*, Springer, Berlin (6th ed. 2005).

Carnap, R. (1956): "The Methodological Character of Theoretical Concepts", in: Feigl, H./ Scriven, M. (eds.), *Minnesota Studies in the Philosophy of Science Vol. I*, Univ. of Minnesota Press, Minneapolis, 38-76.

Cureton, E., D'Agostino, R. (1983): *Factor Analysis: An Applied Approach*, Lawrence Erlbaum Ass., Hillsdale, N.J..

Davidson, D. (1984); *Inquiries into Truth and Interpretation*, Oxford Univ. Press.

Ducasse, J. (1974): *A Critical Examination of the Belief in a Life after Death*, Charles and Thomas, Springfield.

Earman, J. (1986): *A Primer on Determinism*, Reidel, Dordrecht.

Earman, J., 1992, *Bayes or Bust?*, MIT Press, Cambridge/MA.

Flach, P. and Kakas, A. (eds.), 2000, *Abduction and Induction*, Kluwer, Dordrecht.

Gabbay, D., and Woods, J. (2005): *The Reach of Abduction: Insight and Trial (A Practical Logic of Cognitive Systems Vol. 2)*, North-Holland, Amsterdam.

Gemes, K. (1993): "Hypothetico-Deductivism, Content, and the Natural Axiomatization of Theories", *Philosophy of Science*, 54, 477-487.

Glymour, C. (1981): *Theory and Evidence*, Princeton Univ. Press, Princeton.

Glymour, C., Spirtes, P., and Scheines, R. (1991): "Causal Inference", *Erkenntnis* 35, 151-189.

Gorsuch, R. (1983): *Factor Analysis*, Lawrence Erlbaum Ass., Hillsdale, N.J. (2nd ed.).

Grice, H. P. (1975): "Logic and Conversation", in P. Cole, J. Morgan (Hg.), *Syntax and Semantics* Vol. 3, New York 1975, 41-58.

Haig, B. (2005): "Exploratory Factor Analysis, Theory Generation, and Scientific Method", *Multivariate Behavioral Research* 40(3), 303-329.

Halonen, I., and Hintikka, J. (2005): "Towards a Theory of the Process of Explanation", *Synthese* 143/1-2, 5-61.

Hanson, N.R. (1961): "Is There a Logic of Discovery?", in: H. Feigl, G. Maxwell (eds.), *Current Issues in the Philosophy of Science* Holt, Rinehart and Winston, New York, 20-35.

Harman, G. H. (1965): 'The Inference to the Best Explanation', *Philosophical Review* 74, 173-228.

Hintikka, J. (1998): "What is Abduction? The Fundamental Problem of Contemporary Epistemology", *Transactions of the Charles Sanders Peirce Society*, Vol. XXXIV, No. 3, 503 – 533.

Howson, C., and Urbach, P. (1996): *Scientific Reasoning: The Bayesian Approach*, Open Court, Chicago (2nd ed.).

Inglehart, R. (1990): *Cultural Shift in Advanced Industrial Society*, Princeton Univ. Press, Princeton.

Inglehart, R., and Baker, W. (2000): Modernization, Cultural Change, and the Persistence of Traditional Values", *American Sociological Review* 65, 19-51.

Inglehart, R., and Norris, P. (2003): *Rising Tide: Gender Equality and Cultural Change Around the World*. Cambridge/MA: Cambridge University Press.

Josephson, J. and Josephson, S. (eds., 1994): *Abductive Inference*, Cambridge Univ. Press, New York.

Kitcher, P. (1981): "Explanatory Unification", *Philosophy of Science* 48, 507-531.

Kline, P. (1994): *An Easy Guide to Factor Analysis*, Routledge, London.

Kuhn, T.S. (1962): *The Structure of Scientific Revolutions*, Chicago Univ. Press, Chicago.

Kuipers, T. A. F. (2000): *From Instrumentalism to Constructive Realism*, Kluwer, Dordrecht.

Lakatos, I. (1970): "Falsification and the Methodology of Scientific Research Programmes", in: I. Lakatos, A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, Cambridge Univ. Press, Cambridge, 91-195.

Langley, P., et al. (1987): *Scientific Discovery. Computational Explorations of the Creative Process*, MIT Press, Cambridge/Mass.

Lipton, P., 1991, *Inference to the Best Explanation*, Routledge, London.

Magnani, L. (2001): *Abduction, Reason, and Science*, Kluwer, Dordrecht.

Mill, J. St. (1865): *System of Logic*, Parker, Son, and Bourn, London (6th ed.).

Molnar, G. (1999): "Are Dispositions Reducible?", *Philosophical Quarterly* 49, 1999, 1-19.

Moser, P.K. (1989): *Knowledge and Evidence*, Cambridge Univ. Press, Cambridge.

Mumford, S. (1998): *Dispositions*, Oxford Univ. Press, Oxford.

Niiniluoto, I. (1999): "Defending Abduction", *Philosophy of Science* 66 (Proceedings), S436 – S451.

Octoby, D.W. et al. (1999): *Modern Chemistry*, Saunders College Publ., Orlando.

Otte, R. (1981): "A Critique of Suppes' Theory of Probabilistic Causality", *Synthese* 48, 167-189.

Pap, A. (1978): "Disposition Concepts and Extensional Logic", in: Tuomela, R. (ed., 1978): *Dispositions*, Reidel, Dordrecht, 27-54.

Paul, G. (1993): "Approaches to Abductive Reasoning", *Artificial Intelligence Review* 7, 109-152.

Pearl, J. (2000): *Causality*, Cambridge Univ. Press, Cambridge.

Peirce, C.S. (1878): "Deduction, Induction, and Hypothesis", in *Collected Papers* 2.619 – 2.644.

Peirce, C. S. (1903): "Lectures on Pragmatism", in *Collected Papers* CP 5.14 – 5.212.

Peirce, C. S. (CP): *Collected Papers*, ed. by C. Hartshorne, P. Weiss, Harvard Univ. Press, Cambridge/Mass, 1931-35 .

Prior, E.W., Pargetter, R. and Jackson, F. (1982): "Three Theses about Dispositions", *American Philosophical Quarterly* 19, 251-257.

Quine, W.v.O. (1974): *Roots of Reference*, La Salle, Open Court.

Reichenbach, H. (1956): *The Direction of Time*, Univ. of California Press, Berkeley.

Salmon, W. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton Univ.

Press, Princeton.
Schurz, G. (1991): "Relevant Deduction", *Erkenntnis* 35, 391-437.
Schurz, G. (2006): "Patterns of Abduction", *TPD-Preproceedings* 2006/3, University of Duesseldorf (submitted paper).
Schurz, G. (2007): "Clash of Civilizations? An Evolution-Theoretic and Empirical Investigation of Huntington's Theses", to appear in: C. Kanzian, E. Runggaldier (eds.), *Cultures. Conflict-Analysis-Diagnosis*, Ontos Verlag, Frankfurt.
Schurz, G., and Lambert, K. (1994): "Outline of a Theory of Scientific Understanding", *Synthese* 101/1, 65-120.
Sober, E. (1993): *Philosophy of Biology*, Westview Press, Boulder.
Sperber, D. et. al (Eds., 1995): *Causal Cognition*, Clarendon Press, Oxford.
Swinburne, R. (1979):*The Existence of God*, Clarendon Press, Oxford (revised 2nd ed. 2004).