Synthese (2008) 164:201–234
DOI 10.1007/s11229-007-9223-4

Patterns of abduction

G. Schurz

Received: 4 March 2007 / Accepted: 6 July 2007 / Published online: 14 August 2007

Abstract This article describes abductions 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* (Sect. 1). A *classification of different patterns of abduction* is provided which intends to be as complete as possible (Sect. 2). An important distinction is that between *selective abductions*, which choose an optimal candidate from given multitude of possible explanations (Sects. 3–4), and *creative abductions*, which introduce new theoretical models or concepts (Sects. 5–7). While selective abduction has dominated the literature, creative abductions are rarely discussed, although they are essential in science. The article introduces several kinds of creative abductions, such as *theoretical model abduction*, *common cause abduction* and *statistical factor analysis*, and illustrates them by various real case examples. It is suggested to demarcate scientifically fruitful abductions from purely speculative abductions by the *criterion of causal unification* (Sect. 7.1).

Keywords Abduction · Inference to the best explanation · Common-cause-unification · Disposition · Theoretical concepts · Scientific discovery · Analogy · Factor analysis · Realism

1 Introduction: general characterization of abductive reasoning and IBE

In this article, I consider abductions as special patterns of inference to the best explanation (IBS) whose structure *determines* a particularly *promising* abductive conjecture

G. Schurz (✉)
Department of Philosophy, University of Duesseldorf,
Gebaeude 23.21, Universitaetsstrasse 1, Duesseldorf 40225, Germany
e-mail: gerhard.schurz@phil-fak.uni-duesseldorf.de

(conclusion) and thus serves as an *abductive search strategy*.[1] In the following sections (starting with Sect. 2) I will present a detailed reconstruction of patterns of abductions. In this introductory section I suggest three general theses, which underlie my analysis and clarify my terminology:

Thesis 1 (induction versus abduction): Peirce (e.g. 1878, 1903) has distinguished between three families of reasoning patterns: deduction, induction and abduction. Deductions are *non-ampliative* and *certain*: given the premises are true, the conclusion *must* be true. In contrast, inductions and abductions are *ampliative* and *uncertain*, which means that even if the truth of the premises is taken for granted, the conclusion may be false, and is therefore subject to *further testing*. My first thesis is that induction and abduction are two *distinct families* of ampliative reasoning kinds which are *not reducible* to each other. Thereby, I do *not* understand induction as an umbrella term for all kinds of ampliative (non-deductive) inferences (as Earman 1986; Pollock 1986, p. 42; or Ladyman 2002, p. 28); rather, I understand induction in the *narrow Humean sense* in which a property or regularity is *transferred* from the past to the future, or from the observed to the unobserved.

Inductions and abductions can be distinguished by their different *targets*. Both serve the target of extending our knowledge beyond observation—but in rather different respects. Inductions serve the goal of inferring something about the *future course of events*—which is important for planning, that is, adapting our wishful actions to the course of events. In contrast, abductions serve the goal of inferring something about the *unobserved causes* or *explanatory reasons* of the observed events—which is of central importance for manipulating the course of events, that is, adapting the course of events to our wishes (cf. also Peirce 1903, CP 5.189; Aliseda 2006, p. 35).

That abductions cannot be reduced to inductions follows from the fact that inductions cannot introduce new concepts or conceptual models; they merely transfer them to new instances. In contrast, some kinds of abductions *can* introduce new concepts (cf. Peirce 1903, CP 5.170). Following Magnani (2001, p. 20) I call abductions which introduce new concepts or models *creative*, in contrast to *selective* abductions whose task is to choose the best candidate among a given multitude of possible explanations.

That, vice versa, inductions cannot be reduced to abductions is seen as follows. Harman (1965) and Armstrong (1983, p. 78ff) have tried to reduce inductions to abductions by the following argument: the best explanation of the regularities $R(t_i)$ which we have observed at times $t_1,\ldots,t_n$ so far is that they are instances of a *universal law* $\forall t{:}R(t)$. However, I think that this argument is reasonable only if one already *presupposes* that our world is *inductively uniform*. In the absence of an inductive uniformity assumption, there is no reason why the 'true' laws of nature should not change in time, and why the infinitely many *Goodman-laws* $\forall t{:}\, R^*(t)$ (*where* $R^*(t){:}\ \leftrightarrow (t \leq t_n \wedge R(t)) \vee (t > t_n \wedge R'(t))$, for $R't$ incompatible with Rt) should not count as equally good candidates for explanation (cf. Howson 2000, p. 43ff). This shows that an independent justification of induction is needed—although I will not speak about this problem in this article.

[1] Abduction in this understanding includes not only the *discovery* but also a preliminary evaluation of explanatory hypotheses (cf. Magnani 2001, p. 25).

Thesis 2 (inference to the best available explanation may not be good enough): Most authors agree that Harman's IBE has to be modified in (at least) the following respect: nobody knows *all* possible explanations for a given phenomenon, and therefore, what one really has instead of an IBE is an inference to the *best available* explanation, in short an IBAE. However, as Lipton (1991, p. 58) and other authors have pointed out, the best available explanation is not always *good enough* to be rationally acceptable. 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 Sect. 7.1.

Summarizing, the rule IBE is not feasible, and the rule IBAE is not generally acceptable. What is needed for more satisfying versions of abduction rules are (1) *minimal criteria* for the *acceptability* of scientific abductions, and (2) *comparative criteria* for the *quality* of the abduced explanations. Concerning (2), many authors have pointed out (e.g. Niiniluoto 1999, p. S443ff) that a *unique* criterion for the quality of an explanation does not exist—we rather have several criteria which may come in mutual conflict. For example, Lipton (1991, p. 61ff) has argued that in scientific abductions we do not prefer the *likeliest* explanation (i.e., that explanatory hypothesis which is most probable), but to the *loveliest* explanation (i.e., that explanatory hypothesis which offers the best *potential* explanation in terms of explanatory strength, precision, simplicity, etc.). On the other hand, Barnes (1995) has argued that in the examples discussed by Lipton the loveliness of the explanation goes hand in hand with its likeliness, while in those cases in which loveliness and likeness go apart, we usually do not prefer the loveliest but the likeliest explanation. One result of my paper which has a direct bearing on this debate will be that *the evaluation criteria for abductions are different for different kinds of abductions*. So there is no general answer to these questions. For example, in the area of selective factual abductions, comparative plausibility criteria are important, while in the area of creative second-order existential abductions, one needs minimal acceptability criteria (etc.).

Thesis 3 (the strategical role of abductions as means for discovery): All inferences have an *justificational* (or 'inferential') and a *strategical* (or 'discovery') function, but to a different degree (see also Gabbay and Woods 2005, Sect. 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 empirical test operations, or in Hintikka's words, which stimulates new *questions* (Hintikka 1998, p. 528; Hintikka et al. 1999, Sect. 14). In deductive inferences the justificational function is maximal, because the premises guarantee the truth of the conclusion. Deductive inferences may serve also important strategical functions, because many different conclusions can be derived from given premises. In inductive inferences there is not much search strategy involved, because the inductive conclusions of a premise set are narrowly defined by the operations of generalization over instances. So the major function of inductive inferences is justificational, but their justificational value is uncertain. In contrast, 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, p. 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.

In contrast, 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, p. S442). Niiniluoto's point will be confirmed by my analysis: for many (though not all) patterns of abduction their strategical function goes hand in hand with a (weak) justificational value.

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 rules of IBE or IBAE fail completely. If you ask which explanatory conjecture you should choose for further investigation among thousands of possible conjectures, the rule IBAE 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 other words, IBE merely reflects the justificational but misses the strategical function of abductions which in fact is their *essential* function. On this reason, the rule of IB(A)E is epistemically rather uninformative (cf. Day and Kincaid 1994, p. 281).[2]

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). But Peirce did not tell us any abductive rules for conjecturing new theories; he rather explained these miraculous ability of human minds by their *abductive instincts* (CP 5.47, fn. 12; 5.172; 5.212). The crucial question seems to be whether there can exist anything like a 'logic' of discovery. I confine myself to the following remark: 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 explanatory hypotheses*—if there only *were* such rules (cf. also Hanson 1961). This paper intends to show that there *are* such

[2] Kuipers (2004) has pointed out that his version of IBE, which he calls 'inference to the best theory (IBT)', may be conceived as an abduction on the *epistemic meta-level*: from then fact that so far theory T has been empirically more successful than other competing theories one abductively infers that T is closer to the truth than its competitors. Kuipers' observation is interesting in its own right. But it does not change the fact that IBT does not give us any clue of how to *find* such a empirically successful theory T.

rules—in fact, every kind of abduction pattern presented in this article constitutes such a rule.

The majority of the recent literature on abduction has aimed at *one most general* schema of abduction (for example IBE) which matches every particular case. I do not think that good heuristic rules for generating explanatory hypotheses can be found along this route, because these rule are dependent of the *specific type* of abduction scenario, for example, on whether the abduction is mainly selective or creative (etc.). In the rest of this article, I will rather pursue a *new route to abduction*, which consists in modeling various particular schemata of abduction, each fitting to particular kind of conjectural situations. Three general results of my paper can be summarized as follows:

Result 1: There exist rather *different kinds* of abduction patterns. While some of them enjoy a broad discussion in the literature, others have been neglected, although they play an important role in science.

Result 2: I provide a classification of different kinds of abduction patterns along three dimensions. It will turn out that epistemological function and the evaluation criteria of abduction are different for different kinds of abduction patterns. No wonder that philosophers disagree about the status of abduction if they have different things in mind.

Result 3: In all cases the crucial function of a pattern of abduction or IBE consists in its function as a *search strategy* which leads us, for a given kind of *scenario*, in a reasonable time to a most promising explanatory conjecture which is then subject to further test. In selective abductions, the difficulty usually lies in the fact that the *search space* of possible conjectures is *astronomically large*. In creative abductions, however, the difficulty often consists in finding just *one* conjecture which meets the required constraints.

More important than my general theses and results are (at least for me) the *particular results* of the following sections, in which I will model each kind of abduction as a specific schema in which the most promising explanatory conjecture is structurally determined.

2 Three dimensions for classifying patterns of abduction

I will classify patterns of abduction along three dimensions:

(1) Along the kind of *hypothesis* which is abduced, i.e. which is produced as a conjecture,
(2) Along the kind of *evidence* which the abduction intends to explain, and
(3) According to the *beliefs* or *cognitive mechanisms* which *drive* the abduction.

I signify the different kinds of abduction according to the first dimension. But the three dimensions are *not* independent: the properties of an abduction pattern in the second and third dimension are in characteristic covariance with its status in the first dimension. Also, the question of *how* the evidence together with the background knowledge conveys *epistemic support* to the abduced hypothesis, and the question of

Kind of Abduction	Evidence to be explained	Abd. produces	Abd. is driven by
Factual Abduction	Singul. emp. facts	New facts (reasons/causes)	Known laws or theories
Observable-Fact-A	"	Factual reasons	Known laws
1st Order Existential A.	"	Factual reasons postulating new unknown individuals	"
Unobservable Fact-A (Historical Abduction)	"	Unobservable facts (facts in the past)	"
Law-Abduction	Empirical laws	New laws	Known laws
Theoretical-model-Abd.	General empirical phenomena (laws)	New theoretical models of these phenomena	Known theories
2nd Order Existential-Abd.	"	New laws/theories with new concepts	Theoret. b(ackgr). k(nowledge)
Micro-Part Abduction	"	Microscop. composition	Extrapol. of b.k.
Analogical Abduction	"	New laws/theories with analog. concepts	Analogy with b.k.
Hypothetical Cause Abd.	"	Hidden (unobs.) causes	(see below)
Speculative Abduction	(")	(")	Speculation
Common Cause Abd.	"	Hidden *common* causes	Causal Unification
Strict. Comm. Cause Abd.	"	New theoretical concepts	"
Statist. Factor Analysis	"	"	"
Abduction to Reality	Introspect. phenom.	Concept of extern. reality	"

Fig. 1 Classification of kinds of abduction

by which typical follow-up procedures the abduced hypotheses is put to further test, depend crucially on the kind of abduced hypothesis and require a *specific* discussion for each different pattern of abduction. Figure 1 anticipates my final classification of kinds of abduction patterns as an orientation for the reader—the listed kinds of abductions are explained in the following sections.

3 Factual abduction

In factual abductions, both the evidence to be explained and the abduced hypothesis are *singular facts*. Factual abductions are always *driven* by known implicational laws going from causes to effects, and the abduced hypotheses are found by backward reasoning, inverse to the direction of the lawlike implications. This kind of abduction may also be called 'retroduction', or 'the official Peirce abduction schema'[3] (Chisholm 1966, Ch. IV.2, speaks of "inverse induction"). It has the following structure (the double line === always indicates that the inference is uncertain and preliminary):

(FA): *Known Law:* If Cx, then Ex
Known Evidence: Ea has occurred
==========
Abduced Conjecture: Ca could be the reason.

[3] The young Peirce has formalized abduction in this way and had named it 'hypothesis' (cf. 1878). Later on he generalized abduction in the way described in Sect. 1.

Factual abductions are omnipresent in common sense reasoning, and presumably rely on inborn abductive instincts of hominids. Prototypical examples are detective stories (Sebeok and Umiker-Sebeok 1980), or more generally, all sorts of *causal interpretations of traces.* The AI-literature is focused almost exclusively on factual abductions (see Sect. 3.4). Depending on the epistemological nature of the abduced fact, one can distinguish between the following three subpatterns.

3.1 Observable-fact-abduction

Here one reasons according to schema (FA) from observed effects (Ea) to non-observed but observable causes (Ca) in the background of known laws. The follow-up test-procedure consists in the attempt to gain direct evidence for the abduced conjecture. In the example of a murderer case, such direct evidence would be given, for example, by a confession of the putative murderer to have committed the crime.

In the setting of factual abduction, the problem consists in the *combinatorial explosion* of the search space of possible causes in the presence of a rich background store of laws but in the *absence* of a rich factual knowledge. Thus, factual abductions are primarily *selective* in the sense of Magnani (2001, p. 20), and their epistemic support depends on the degree in which the background knowledge increases their *probability* in comparison to alternative possible causes. Consider the following example: if your evidence consists in the trace of the imprints of sandals on an elsewhere empty beach, then your immediate conjecture is that somebody was recently walking here. How did you arrive at this conjecture? Classical physics allows for myriads of ways of imprinting footprints into the sand of the beach, which reach from cows wearing sandals on their feet to foot-prints which are drawn into the sand, blown by the wind, or caused by radioactive decay of foot-shaped portions of the sand, etc. The majority of these physically possible abductive conjectures will never be considered by us because they are extremely improbable. The major strategic algorithm which we apply in factual abduction cases of this sort is a *probabilistic elimination technique* which usually works in an unconscious manner: our mind quickly scans through our large memory store containing millions of memorized possible scenarios and only those which have minimal plausibility pop up in our consciousness.

So, probabilistic evaluation of causes and elimination of implausible causes plays a central role in factual abductions. Of course, such a probabilistic evaluation can provide justification only to the extent that (a) these probability assertions are supported by statistical laws, and (b) our knowledge of the causal laws which may lead us to the possible causes of the explanandum via 'retroduction' is complete or, at least, does not miss some probable cause.

Fumerton (1980, p. 592f) has gone further and has argued that factual abduction can even be *reduced* to ordinary inductive-statistical inference. More precisely, he argues that the inferences pattern at the left can be reduced to the inference pattern at the right in the following way (where 'P(–)' denotes subjective-epistemic and 'p(–)' statistical probability, and 'K' expresses background knowledge):

Abductive inference:		*Inductive-statistical inference:*
L: $\forall x(Fx \rightarrow Gx)$	Fumerton's	L′: $p(Fx\|Gx) = high$
Ga	reduction:	Ga
==========	$\longrightarrow$	======== $P(Fa \mid Ga \wedge L') = high$
Fa (presupposition: $P(Fa \mid Ga \wedge L \wedge K) = high$)		Fa

Although Fumerton's reduction seems reasonable in some cases, I see two reasons why his argument is not generally correct. First, the abductive hypothesis is probabilistically evaluated not merely in the light of the evidence Ga and an inverse statistical law L′, but in the light of the entire background knowledge K. Fumerton may reply that the inference pattern at the right may be appropriately extended so that it includes background knowledge. But second and more importantly, Fumerton's proposed transformation does neither correspond to psychological reality nor would it be strategically recommendable. For every individual case (or effect) is 'different', and hence, only a small fraction of possible cause-effect-scenarios are frequently enough encountered in a human life-time to get explicitly stored by Fumerton-like conditional probabilities. For example, if you are *not* a turtle expert and you observe the trace of a turtle in the sand, then the only way in which you may arrive at the right guess that there was a turtle robbing here is by careful backward reasoning combined with elimination. Only if you are a turtle hunter you may have explicitly stored the typical sand-traces of turtles with a corresponding forward conditional of Fumerton's sort. The importance of backward reasoning and elimination is also emphasized by all experts of detective stories (cf. Sebeok and Umiker-Sebeok 1980).

3.2 First-order existential abduction

This subcase of factual abduction occurs when the antecedent of a law contains so-called *anonymous* variables, i.e. variables which are not contained in the consequent of the law. In the simplest case, the formal structure of first-order existential abduction is the following (cf. also Thagard 1988, p. 57f):

L: $\forall x \forall y(Ryx \rightarrow Hx)$	logically equivalent: $\forall x(\exists\, yRyx \rightarrow Hx)$
Ha	
==========	
Conjecture: $\exists yRya$	

Instantiating the consequent of the law with 'a' and backward chaining yields a law-antecedent in which one variable remains uninstantiated ('Rya'). In such a case, the safest abductive conjecture is that one in which we existentially quantify over this variable. We have already discussed an example of this sort in Sect. 3.1: from the footprint in the sand we abductively infer that *some* man was walking at the beach. We do not infer, as Fumerton emphasizes (1980, p. 594), that some particular person, out of million possible persons, has walked here. But note that only in some cases we will be satisfied with the existential conjecture. In other cases, in particular in criminal cases, all depends on finding out *which* individual is the one who's existence

we conjecture—who was the murderer? Here one is not satisfied with a first-order existential abduction but wants to have an proper (fully instantiated) fact-abduction.

In observable-fact-abduction the abduced hypothesis may at later stages of inquiry by confirmed by direct observation—for example, when we later meet the man who had walked yesterday at this beach. In this case, the weak epistemic support which the abductive inference conveys to the conjecture gets fully replaced by the strong epistemic support provided by the direct evidence: abduction has played an important strategic role, but it does not play any longer a justifying role. This is different, however, in all of the following patterns of abduction, in which the abductive hypothesis is not directly observable, but only indirectly confirmable via its empirical consequences.

3.3 Unobservable-fact abduction

This kind of abduction has the same formal structure as observable-fact abduction, but the abduced fact is unobservable. The *typical* case of unobservable-fact abductions are *historical-fact abductions*, in which the abduced fact is unobservable because it is located in the distant past. The abduced fact may also be unobservable in principle, because it is a theoretical fact. However, in such a case the abduction is usually not driven by simple implicational laws, but by a quantitative theory, and the abduced theoretical fact corresponds to a theoretical model of the observed phenomenon: this sort of abduction differs crucially from law-driven factual abduction and is therefore treated under the separate category of 'theoretical-model abduction' (Sect. 5).

Historical-fact abductions are of obvious importance for all historical sciences (cf. also Niiniluoto 1999, S442). Assume for example, biologists discover marine fossil records, say fish bones, in the ground of dry land. They conjecture abductively, given their background theories, that some geological time span ago there was a sea here. Their hypothesis cannot be directly verified by observations. So the biologists look for further empirical consequences which follow from the abduced conjecture plus background knowledge—for example, further geological indications such as calcium deposits, or marine shell fossils, etc. If the latter findings are observationally verified, the abductive conjecture is confirmed. Logically speaking, an unobservable-fact abduction performs a combination of abductive backward reasoning and deductive or probabilistic forward reasoning to consequences which can be put to further test. This is graphically displayed by the bold arrow in Fig. 2. If the empirical consequence Ea is verified, then both evidences Ga and Ea provide epistemic support to the abduced hypothesis Ha (modulo probabilistic considerations in the light of the background

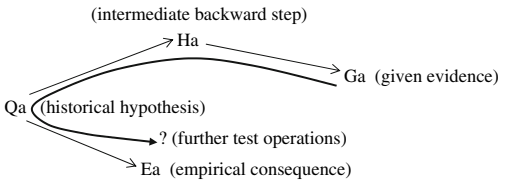

Fig. 2 Historical-fact-abduction (the *bold arrow* indicates the route of the abduction process)

knowledge). So, the initial abductive inference has not only a strategical value, but *keeps* its justificatory value.

3.4 Logical and computational aspects of factual abduction

If the background knowledge does not contain general theories but just a finite set of (causal) implicational laws, then the set of possible abductive conjectures is finite and can be generated by *backward-chaining* inference procedures. In this form, abductive inference has been studied in detail in AI research (cf. Josephson and Josephson 1994; Flach and Kakas 2000). Given is a knowledge base **K** = <**L[x]**,**F[a]**> in form of a finite set **L[x]** of monadic implicational laws going from conjunctions of open literals to literals, and a finite set **F[a]** of facts (closed literals) about the individual case a. (A literal is an atomic formula or its negation.) Given is moreover a certain fact Ga (the 'goal') which is to be explained. One is not interested just in any hypotheses which (if true) would explain the goal Ga given **K**, but only in those hypotheses which are not further potentially explainable in **K** (cf. Paul 1993, p. 133; Console et al. 1991). So formally, the candidates for abducible hypotheses are all closed literals A[a] such that A[a] is neither a fact in **F[a]**, nor is A[x] the consequent ('head') of a law, i.e., A[a] cannot possibly be further explained by other laws in **K**. The set of these possible abductive conjectures A[a] for arbitrary abduction tasks in **K** is called the set of *abducibles* **H[a]**. The *abductive task* for goal Ga is then defined as follows: find *all possible* explanations, i.e., all *minimal* sets **E[a]** of singular statements about a such that (i) **E[a]** $\subseteq$ **F[a]** $\cup$ **H[a],** (ii) **L[x]** $\cup$ **F[a]** $\cup$ **E[a]** is consistent and (iii) **L[x]** $\cup$ **E[a]** logically implies G[a] (by forward chaining). Those elements of the explanatory sets **E[a]** which are abducibles are the abductive hypotheses for G[a]. Solutions of this sort of task have been implemented, for example, in the programming language PROLOG in the form of backward-chaining with backtracking to all possible solutions.

This kind of abduction problem is graphically displayed in Fig. 3 in form of a so-called *And-Or-tree* (cf. Bratko 1986, Ch. 13; Schurz 1996). The *labeled* nodes of an And-Or-tree correspond to literals, unlabeled nodes represent conjunctions of them, and the directed edges (arrows) correspond to laws in **L[x]**. Arrows connected by an *arc* are And-connected; without an arc they are Or-connected. Written statementially, the laws underlying Fig. 3 are $\forall x(Fx \rightarrow Gx)$, $\forall x(Hx \rightarrow Gx)$, $\forall x(Q_1x \wedge Q_2x \rightarrow Gx)$, $\forall x(R_1x \wedge R_2x \rightarrow Fx)$, $\forall x(Sx \rightarrow Hx)$, $\forall x(T_1x \wedge T_2x \rightarrow Hx)$, $\forall x(Ux \rightarrow Q_1x)$, $\forall x(Vx \rightarrow Q_2x)$. Besides the goal Ga, the only known fact is T_1a.

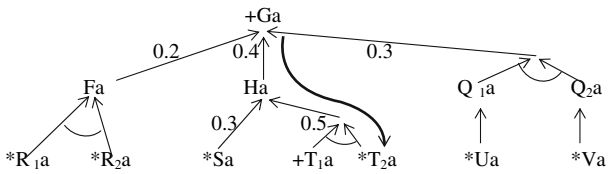

Fig. 3 Search space for a factual abduction problem. + indicates a known fact, * indicates possible abductive hypotheses. The numbers are probability values (they do not add up to 1 because of an unknown residual probability). The *bold arrow* indicates the route of a best-first search, which leads to the abductive conjecture T_2a

The task of finding *all* possible explanations has exponential complexity and, thus, is intractable (that is, the time of this task increases exponentially in the number of data and possible hypotheses). Only the complexity of finding some explanation has polynomial complexity and is tractable (cf. Josephson and Josephson 1994, Ch. 7, p. 165, th. 7.1 + 7.2). Therefore it is crucial to constrain the search space by probabilistic (or plausibilistic) evaluation methods. A simple heuristic strategy is the *best-first* search: for each Or-node one processes only that successor which has a highest plausibility value (among all successors of this node). The route of a best-first abduction search is depicted in Fig. 3 by the bold arrow. More complicated procedures update the plausibilities of chosen paths at each new layer of the search tree (cf. Bratko 1986, Ch. 13). Finding a best explanation in terms of plausibility (or probability) is polynomial only under the condition that no two Or-node-successors have the same degree of plausibility; otherwise this task is NP-hard and, hence, intractable (Josephson and Josephson 1994, Ch. 7, p. 172, th. 7.12, p. 173, th. 7.14).

A related but more general logical framework for factual abductions via backward reasoning are abductions within Beth-tableaus, which have been worked out by Hintikka et al. (1999) and in particular by Aliseda (2006). An even more general framework are abductions within Gabbay's labeled deductive systems, which are elaborated in Gabbay and Woods (2005, part III).

Besides probabilistic elimination, the second major technique of constraining the search space is *intermediate information acquiration*: not only the ultimately abduced conjectures, but also intermediate conjectures (nodes) along the chosen search path can be set out to further empirical test—or in the framework of Hintikka et al. (1999), they may stimulate further interrogative inquiry (see also Walton 2004, Ch. 6). As an example, consider again a criminal case: if backward reasoning leads to the possibility that the butler could have been the murderer, and along an independent path, that the murderer must have been left-handed, then before continuing the abductive reasoning procedure one better finds out first whether the butler is indeed left-handed. There are also some AI-abduction systems which incorporate question-asking modules. For example, the RED-system, designed for the purpose of red-cell antibody identification based on antigen-reactions of patient serum, asks intermediate question to a data-base (Josephson and Josephson 1994, p. 72f).

4 Law-abduction

In this kind of abduction, both the evidence to be explained and the abduced hypothesis is an implicational law, and the abduction is driven by one (or several) known implicational laws. Due to the latter fact, this kind of abduction is more similar to factual abductions than to theory-driven abductions which are discussed in Sect. 5. Law-abductions can already be found in Aristotle, and they correspond to what Aristotle has called the mind's power of *hitting upon the middle term* of a syllogism (An. Post., I, 34). Here is an example:

Background law:	$\forall x(Cx \rightarrow Ex)$	Whatever contains sugar tastes sweet
Emp. law to be explained:	$\forall x(Fx \rightarrow Ex)$	All pineapples taste sweet
Abduced conjecture:	$\forall x(Fx \rightarrow Cx)$	All pineapples contain sugar.

A more general example of law-abduction in qualitative chemistry is this:

All substances which contain molecular groups of the form C have property E.
All substances of empirical kind S have certain empirical properties E.
==========
Conjecture: Substances of kind S have molecular characteristics C.

In there are several causal background laws of the form $\forall x(C_i x \rightarrow Ex)$, then one has to selected the most 'plausible' one. In any case, the conclusions of law abductions are conjectural and in strong need of further support.

Flach and Kakas (2000, p. 21f) have argued that a law-abduction can be *reduced* to the following combination of a fact-abduction and an inductive generalization:

Background law:	$\forall x(Cx \rightarrow Ex)$			
Observed facts:	$Fa_i \wedge Ea_i$	$1 \leq i \leq n$	$\Rightarrow$	*Induction basis* for: $\forall x(Fx \rightarrow Ex)$
========== Factual abduction				
Abduced hypotheses:	Ca_i	$1 \leq i \leq n$		
hence:	$Fa_i \wedge Ca_i$	$1 \leq i \leq n$	$\Rightarrow$	*Induction basis* for $\forall x(Fx \rightarrow Cx)$

This decomposition, however, is somewhat artificial. Law-abductions are usually performed in one single conjectural step. We don't form the abductive hypothesis of containment of sugar for each observed pineapple, one after the other, and then generalize it, but we form the law-conjecture "pineapples contain sugar" at once.

All patterns of abduction which we have discussed so far are all driven by known qualitative implication laws, and they are mainly *selective*, i.e. their driving algorithm draws a most promising candidate from a class of possible conjectures which is very large but in principle constructible. These patterns are dominating the abduction literature. In contrast, the patterns of abductions to be discussed in the next sections are rarely discussed in the literature (except in an unspecific way under the heading 'IBE' which does inform about the underlying logical pattern and/or algorithm).[4] They are not driven by implicational laws, but either by scientific theories, or by (causal) unification procedures. They are rare in common-sense reasoning, but play a decisive role in advanced scientific reasoning. Also, they are not mainly selective but mainly *creative*, that is, the underlying operation or algorithm constructs something new, for example a new theoretical model or even a new theoretical concept.

[4] For example, Thagard's (1988) classification of abduction contains factual ('simple') abduction, (first-order) existential abduction, law ('rule') abduction, and analogical abduction which we discuss in Sect. 7.2, but nothing else.

5 Theoretical-model abduction

The explanandum of a theoretical-model abduction is typically a well-confirmed and reproducible 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 *finding 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.[5] 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).

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 weight of the body (f_B) the body will swim, otherwise it will sink. Since the volume 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 rather different from the situation of factual abductions: one does *not* face here the problem of a huge multitude of possible theoretical models or conjectures. For 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 (1962), that is, the activity of extending a given theory core (or

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

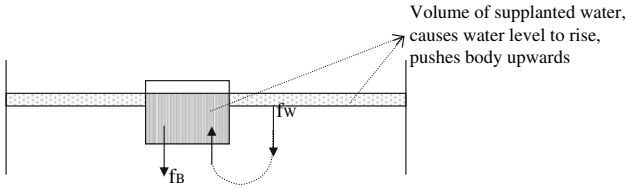

Fig. 4 Theoretical conditions which allow the mechanical derivation of the law of buoyancy

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 and Hintikka 2005, Sect. 3).

While for all other kinds of abductions we can provide 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, p. 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. Here is an example:[6]

Abduction pattern of Newtonian particle mechanics:

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

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

on the particle's position at the given time; (ii) insert these specifications into Newton's second 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 mentioned 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.

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. A computational implementation is the automatic abduction system DENDRAL (Buchanan 1969, p. 234ff), which abduces the chemical structure of organic molecules given their mass spectrum and their gross formula.

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 they 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, p. 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).

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: examples of this sort are recognized because there are several *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 and Woods 2005, Sect. 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

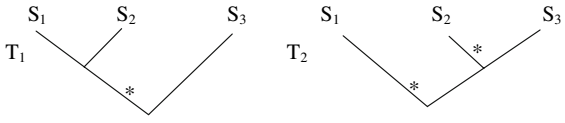

Fig. 5 Two alternative trees of descendance. * = mutation of non-F into F

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, Sect. 1). More specific background assumptions are hermeneutic rationality presumptions (Davidson 1984), Grice's maxims of communicative cooperation (Grice 1991), 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. The investigation of interpretation as abduction is also an important area in AI (cf. Hobbs et al. 1993).

What all abduction schemata discussed so far have in common is that they are driven by *known* laws or theories, and hence, they work within a *given conceptual space*. In other words, the abduction schemata discussed so far cannot introduce *new concepts*. In the next section we turn to abduction schemata which can do this: since their explanans postulates the existence of a new kind of property or relation, we call them 'second-order existential abductions'.

6 Second-order existential abduction

The explanandum of a second-order existential abduction consists, again, of one or several general empirical phenomena, or laws. What one abduces is an at least *partly* new property or kind concept governed by an at least partly new theoretical law. Depending on whether the concept is merely partly or completely new, the abduction is driven by extrapolation, analogy, or by pure unification. We discuss these kinds of abductions in the following Sects. 6.1–7.4.

6.1 Micro-part abduction

In this most harmless case of second-order existential abduction one abduces a hypothesis about the microscopic composition of observable objects in terms of micro-parts which obey the *same* laws as the observable macroscopic objects, in order to explain various observed empirical phenomena. The prototypical example is the *atomic hypothesis* which has been conjectured already in antiquity by Leucippus and Democritus and was used to explain such phenomena as the dissolution of sugar in water, or the re-sublimation of salt from 'salty air' close to the sea, etc. These philosophers have abduced a new natural kind term: *atoms*, which are too small to be observable, but otherwise obey the same mechanical laws as macroscopic bodies. So what one does here is to extrapolate from macroscopic concepts and laws to the microscopic domain—whence we may also speak here of *extrapolative* abduction. In the natural sciences after Newton, the atomic hypothesis turned out to have an enormous explanatory power. For example, Dalton's atomic hypothesis had successfully explained Avogadro's observation that equal volumes of gases contain the same number of gas particles. Dalton also postulated that all substances are composed of molecules built up from certain atoms in certain integer-valued ratios, in order to explain the laws

of constant proportions in chemical reactions (cf. Langley et al. 1987, p. 259ff). The different states of aggregation of substances (solid, fluid, and gaseous) are explained by different kinds of inter-molecular distances and interactions. We conclude our list of examples here, although many more applications of the atomic hypothesis could be mentioned.

Extrapolative micro-part abductions differ from analogical abductions insofar as the atoms are not merely viewed as 'analogical' to mechanical particles; they are literally taken as tiny mechanical particles (too small to be observable). Nevertheless one may view extrapolative abductions as some kind of 'pre-stage' of analogical abductions, which we are going to discuss now.

6.2 Analogical abduction

Here one abduces a partially new concept and at the same time new laws which connect this concept with given (empirical) concepts, in order to explain the given law-like phenomenon. The concept is only partly new because it is analogical to familiar concepts, and this is the way in which this concept was discovered. So analogical abduction is *driven* by analogy. We first consider Thagard's (1988) example of sound waves.

Background knowledge: Laws of propagation and reflection of water waves.

Phenomenon to be explained: Propagation and reflection of sound.

==========

Abductive conjecture: Sound consists of atmospheric waves in analogy to water waves.

According to Thagard (1988, p. 67) analogical abduction results from a *conceptual combination:* the already possessed concepts of wave and sound are combined into the combined concept of a sound-wave. I think that this early analysis of Thagard (1988) is too simple. In my view, the crucial process which is involved in analogical abduction is a *conceptual abstraction* based on *isomorphic* or homomorphic *mapping*. What is abduced by this analogy is not only the combined concept of sound-wave, but at the same time the *theoretical* concept of a *wave in abstracto* (also the later article of Holyak and Thagard 1989 supports this view).

A clear analysis of analogy based on mapping and conceptual abstraction has been given by Gentner (1983). According to Gentner's analysis, an analogy is a *partial isomorphic* mapping m between two relational structures, the *source* structure (D, (F_i: $1 \leq i \leq m$), (R_i: $1 \leq i \leq n$)) and the *target structure* (D*, (F^*_i: $1 \leq i \leq m^*$), (R^*_i: $1 \leq i \leq n^*$)), where the F_i are monadic predicates and the R_i are relations. Gentner argues convincingly (158f) that an analogical mapping preserves only the *relations* of the two structures (at least many of them, including second-order relations such as "being-a-cause-of"), while monadic properties are not preserved. This is it what distinguishes an *analogy* from a *literal similarity*. For example, our solar system is literally similar to the star system X12 in the Andromeda galaxy, insofar the X12 central star is bright and yellow like our sun, and surrounded by planets which are similar to our planets. Thus, our sun and the X12 star have many (monadic) properties in common. On the other hand, an atom (according to the Rutherford theory) is

merely analogical to our solar system: the positively charged nucleus is surrounded by electrons just as the sun is surrounded by planets, being governed by a structurally similar force law. But concerning its monadic properties, the atomic nucleus is very different from the sun, the electrons are different from the planets, and the electrical force between protons and electrons is different from the gravitational force between the sun and its planets. Formally, then, an analogical mapping m maps a subset D′ of D bijectively into a subset D′* of D*, and many (but not necessarily all) relations R_i, with $i \in I \subseteq \{1,\ldots,n\}$, into corresponding relations $R^*_{m(i)}$, such that for all a, b ∈ D′ and R_i with i ∈ I, aR_ib iff $m(a)R^*_{m(i)}m(b)$ holds. In this sense, the Rutherford-analogy maps "sun" into "nucleus", "planet" into "electron", "gravitational attraction" into "electrical attraction", "surrounding" into "surrounding", etc. It follows from the existence of such a partial isomorphic mapping that for every explanatory law L expressed in terms of mapping-preserved relations which holds in the D′-restricted source structure, its starred counterpart L* will hold in the D′*-restricted target structure. In this way, explanations can be transferred from the source to the target structure (which is of particular importance for Thagard 1992).

Every partial isomorphism gives rise to a *conceptual abstraction* by putting together just that parts of both structures which are isomorphically mapped into each other: the resulting structure (D′, (R_i: i ∈ I)), which is determined up to isomorphism, is interpreted in an *abstract* system-theoretic sense. In this way, the abstract model of a *central force system* arises, with a *central body*, *peripherical bodies*, a *centripedal* and a *centrifugal force* (Gentner 1983, p. 160f). So, finding an abductive analogy consists in finding the *theoretically essential* features of the source structure which can be generalized to other domains, and this goes hand-in-hand with forming the corresponding conceptual abstraction. In our example, the analogical transfer of water-waves to sound-waves can only work if the theoretically essential features of (water-) waves have been identified, namely, that waves are produced by *coupled oscillations*. The abductive conjecture of sound-waves stipulates that also sound consist of coupled oscillations of the molecules of the air. Only after this theoretical model of sound-waves is formed, a *theoretical* explanation of propagation and reflection of sound-waves becomes possible.

6.3 Hypothetical (common) cause abduction

This is the most fundamental kind of conceptually creative abduction. The explanandum 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. I will argue that only case (b) constitutes a scientifically worthwhile abduction, while (a) is a case of pure speculation. 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. This kind of abduction does not presuppose any background knowledge except knowledge about those phenomena which are in need of explanation. What drives hypothetical cause abduction is the pure search for *unification*, usually in terms of hidden or common

causes—but later on, we will meet also cases where the unifying parameters have a merely instrumentalistic interpretation. Hypothetical (common) cause abduction is such a large family of abduction patterns that we treat it in separately in the next section.

7 Hypothetical (common) cause abduction continued

Salmon (1984, p. 213ff) has emphasized the importance of finding common cause explanations for the justification of scientific realism. However, Salmon does not inform us about the crucial difference between scientific common cause abduction and speculative (cause) abduction. In the next two subsections I argue that the major criterion for this distinction is *causal unification*.

7.1 Speculative abduction versus causal unification: a minimal adequacy criterion

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, pp. 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, p. 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) *Minimal adequacy criterion for second-order abductions*: 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 and Lambert 1994, Sect. 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:	Some power wanted that John gets ill, and
$\forall x(\psi_E x \rightarrow Ex) \wedge \psi_E a$	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 (ψ_Ea) has to be postulated (cf. Schurz and Lambert 1994, p. 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 Sect. 7.2 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, p. 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, p. 54; Howson and Urbach 1996, p. 119ff; Kuipers 2000, Sect. 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: $\forall x(Fx \rightarrow \psi_D x)$	Opium has a special power (a 'virtus dormitiva')
$\wedge\forall x(\psi_D x \rightarrow Dx)$	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, Sect. 2; Ducasse 1974, Ch. 6, Sect. 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 and Lambert 1994, p. 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 $K \models S$. 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. 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, p. 130). But as it was pointed out in Sect. 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* is the following

(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).[7]

[7] Cf. Glymour et al. (1991, p. 151). A generalization of (R) is the following condition (M): "if X and Y are probabilistically dependent given Z, then either Z is a common effect of X and Y or there exists a

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 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, p. 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.

Footnote 7 continued
causal connection between X and Y which does not go through Z". The equivalence of (M) with Glymour's Markov-condition (ibid., p. 156) follows from theorems 1.2.4 and 1.2.7 of Pearl (2000, pp. 16–19). (M) implies (R), and moreover Reichenbach's *screening-off criterion* (Reichenbach 1956, p. 159), which says that direct causes screen of indirect causes from their effects, and common causes screen off their effects from each other.

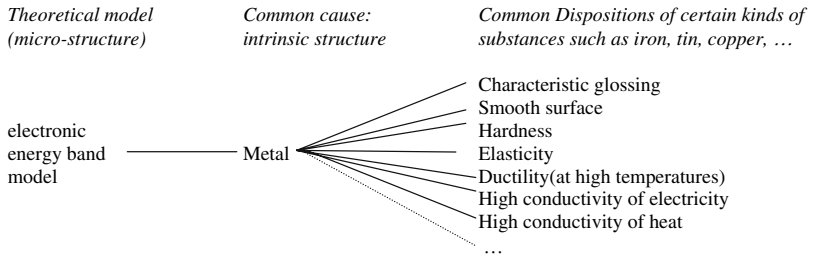

Fig. 6 Common cause abduction of the chemical kind term 'metal'

7.2 Strict common cause abduction from correlated dispositions and the discovery of new natural kinds

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 mutually 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. 6): 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 (more-or-less strictly) correlated dispositions, and they have called it *metallic character* Mx.

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*. 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 electrons can shift easily around, which offers a unifying explanation of all the common dispositions of metals (cf. Octoby et al. 1999, p. 708ff).

In the *history of chemistry*, common cause abductions from correlated dispositions were of central importance in the discovery of new (theoretical) kinds of substances.

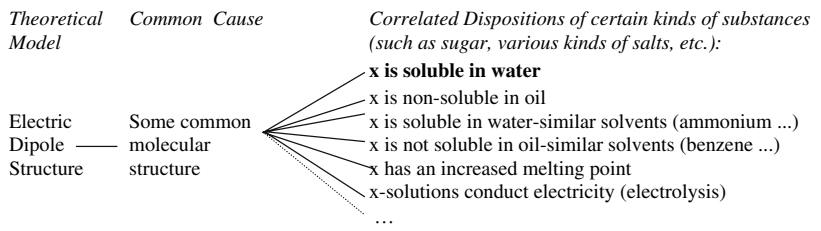

Fig. 7 Common cause abduction of the theoretical term "hydrophylic/polar" molecular structure

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. 7). 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.

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): "$D(x): \leftrightarrow p_{t\in\Delta}(Rxt|Cxt) = high$".

My conditional understanding of dispositions is in according with the 'received view' (cf. Carnap 1956, Sect. IX–X; Pap 1978, p. 44), which has been defended by Prior et al. (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, p. 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 'second-order properties', they can only be the effects of certain (categorial) causes, but cannot themselves act as causes (cf. Prior et al. 1982, p. 255; the same point has been emphasized by Ducasse 1974, Ch. 6, Sect. 2).

In contrast to this view, philosophers such as Quine (1974, Sects. 3–4), Armstrong (1969, p. 70f) and Mumford (1998, p. 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, p. 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.[8]

In conclusion, the categorial view of dispositions is not in accord with 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, 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 $\forall x \forall t \in \Delta(Sx \wedge Cxt \rightarrow_n Rxt)$.[9]

The structural pattern of the two examples (Figs. 6 and 7) can be formalized as follows:

[8] 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, p. 163).

[9] 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_ixt_i given C_ixt_i (where $t_1,\ldots,t_n$ are different time points at which the different initial conditions have been realized).

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 $D_1,\ldots,D_m$.
$\forall i \in \{1,\ldots,n\}: \forall x(F_i x \rightarrow \psi x) \wedge \forall j \in \{1,\ldots,m\}: \forall x(\psi x \rightarrow [\leftrightarrow] D_j x)$.

The abductive conjecture H logically implies E and it yields a *unification* of $n \cdot m$ 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 "[↔]"). 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 application of this 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 "[↔]", then this strengthened hypothesis predicts that all the D_j are mutually strictly correlated ($\forall i \neq j \in \{1,\ldots,m\}$: $D_i x \leftrightarrow D_j x$). 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 Sect. 7.3).

Many more examples of common cause abduction in the natural sciences could be given. For example, Glauber's discovery of the central chemical concepts of *acids*, *bases*, and *salts* in the 17th century was based on a typical common cause abduction (cf. Langley et al. 1987, p. 196ff). As a final example, the fundamental common cause abduction of Newtonian physics 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 (for details cf. Glymour 1981, p. 203ff). Generally speaking, every fundamental common cause abduction in science is 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 theoretical model abduction in the sense of Sect. 5. The capacity of producing *novel predictions* is significantly enhanced by this transformation.

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*.[10]

[10] If the correlations between the events are not strict but merely probabilistic, then one may also use Reichenbach's *screening off* criterion (see fn. 7) 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).

7.3 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.[11] In this section I want to show that factor analysis is a certain generalization of 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→ |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 $X_1,\ldots,X_n$, 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. 7) 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 given individuals x in the 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 the 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; cf. Bortz 1985, pp. 215–234). Visually speaking, the X_i-values form a stretched cloud of points 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 the general

[11] An exception is Haig (2005), who shares my view of factor analysis.

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:[12]

$$X_1 = a_{11} \cdot F_1 + \cdots + a_{1k} \cdot F_k + s_1.$$
$$\ldots$$
$$X_n = a_{n1} \cdot F_1 + \cdots + a_{nk} \cdot F_k + s_n.$$

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 represent 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 between variable X_i and factor F_j, $\mathrm{cor}(X_i,F_j)$. 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 *amount of the variance* of X_i 'explained' (i.e. statistically predicted) by factor F_j. The sum of the squared loadings of a factor F_j, $\Sigma_{1 \leq i \leq n} a_{ij}$, measures the amount of total variance of the variables which is explained by F_j, and the sum of all of the squared loadings divided through n equals the percentage of variance explained by the extracted factors, which is a 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}.

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. This 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 Sect. 7.1. 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 (in the trivial limiting case $k = n$ the amount of explained variance becomes 100%). According to the Kaiser-Guttman-criterion one should introduce new factors only as long as their amount of explained variance is greater than one (cf. Bortz 1985, p. 662f; Kline 1994, p. 75). Hence, a theoretical factor is only 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: this is the factor analytic counterpart of my suggested minimal criterion for hypothetical cause abduction (CU).

[12] For the following cf., e.g., Bortz (1985, Ch. 15.1), or Kline (1994, Ch. 3). I only describe the most common method of factor analysis, without discusssing the subtle differences between different factor analytic methods.

Table 1 Eighteen empirical variables measuring the subjective evaluations of the voices of persons explained by three factors F_j

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
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%

E.V. = explained amount of the variance (taken from Bortz 1985, 672)

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, pp. 665–672; Kline 1994, p. 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. Table 1 offers an example from Bortz (1985, p. 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 (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. By the way, 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.

7.4 Epistemological abduction to reality

The relevance of abduction for realism is usually discussed within the context of theories and theoretical-entity-realism. For many epistemologists (e.g., Chisholm 1979, p. 115; Pollock 1986, p. 44), the fundamental problem of common sense realism—the reasoning from introspective sense data to an external reality causing these perceptions—is an inference sui generis, and its justification is a problem of its own. In contrast to this position, I wish to point out that also the reasoning from introspective sense data to common sense realism is in perfect fit with the pattern of common cause abduction (cf. also Moser's IBE-analysis of this inference in 1989, p. 98). The hypothesis of external objects which cause our sensual experience yields a common cause explanation of a huge set of intercorrelations between our introspective experiences.

First, there are the *intra*-sensual intercorrelations, in particular those within our system of visual perceptions. There are potentially infinitely many two-dimensional visual images of a perceptual object, but all these two-dimensional images are strictly correlated with the position and angle at which we look at that objects; so these correlations have a common cause explanation in terms of three-dimensional external objects by the laws of perspectival projection. To be sure, these common cause abductions are mainly *unconscious* and rely on inborn computations performed by the visual cortex of our brains. What we consciously experience are the 'abduced' three-dimensional objects which make up the mind of the 'naive realist'. However, certain situations—for example the visual illusions caused by 3D-pictures—make it plain that what underlies our three-dimensional visual appearances is a complicated abductive computation process (cf. also Rock 1984). Moreover, since in our ordinary visual perceptions some objects partly conceal other objects which are behind them, our visual abductions always include the task of Gestalt complementation. Identification of three-dimensional objects based on two-dimensional projective images is also an important abductive task in the AI field of visual object recognition (Russell and Norvig 1995, Ch. 24.4). Scientifically advanced versions of visual abduction where one abduces the shape of entire objects from sparse fragments have been analyzed in the field of archaeology (Shelley 1996).

The *inter*-sensual correlation between different sensual experiences, in particular between visual perceptions and tactile perceptions, is the *second* important basis for the unconscious abduction to an outer reality—in fact, these correlations seem even to be the major fundament of our naive belief in the outer reality. If you have a visual appearance of an object, but you are unsure whether it is a mere visual illusion or not, then you will probably *go* to the object and try touch it—and if you can, then your realistic desires are satisfied. One the other hand, visual appearances which do

not correspond to tactile ones, so-called 'ghosts', have frightened the naively realistic mind and occupied its fantasy since the earliest times.

This concludes my analysis of patterns of abduction. Instead of a conclusion, I refer to the classification of abduction patterns in Fig. 1, and to my main theses and results as explained in Sect. 1, which are densely supported by the details of my analysis. For those who want to final conclusion, I propose the following: as Peirce had once remarked (CP 6.5000), the success of scientists to find true hypotheses among myriads of possible hypotheses seems to be a sheer miracle. I think that this success becomes much less miraculous if one understands the strategical role of patterns of abduction.

Acknowledgements For valable comments on earlier drafts I am indebted to Jaakko Hintikka, Ilkka Niiniluoto, Theo Kuipers, Helen Beebee, Max Kistler, Gerhard Brewka and Helmut Prendinger.

References

Aliseda, A. (2006). *Abductive reasoning*. Dordrecht: Springer.
Armstrong, D. M. (1983). *What is a law of nature?* Cambridge: Cambridge Univ. Press.
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*. Berlin: Springer [6th ed. 2005].
Bratko, I. (1986). *Prolog programming for artificial intelligence*. Reading, MA: Addison-Wesley Publ. Comp.
Buchanan, B., et al. (1969). Heuristic dendral. *Machine Intelligence, 4*, 209–254.
Carnap, R. (1956). The methodological character of theoretical concepts. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science, Vol. I* (pp. 38–76). Minneapolis: Univ. of Minnesota Press.
Chisholm, R. M. (1966). *Theory of knowledge*. Englewood Cliffs, NJ: Prentice Hall [3rd ed. 1988].
Console, L., et al. (1991). On the relationship between abduction and deduction. *Journal of Logic and Computation, 1*, 661–690.
De Regt, H. W. (2006). Wesley Salmon's complementarity thesis: Causality and Unificationism Reconciled? *International Studies in the Philosophy of Science, 20*, 129–147.
Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: Oxford Univ. Press.
Day, T., & Kincaid, H. (1994). Putting inference to the best explanation in its place. *Synthese, 98*, 271–295.
Ducasse, J. (1974). *A critical examination of the belief in a life after death*. Springfield: Charles and Thomas.
Earman, J. (1986). *A primer on determinism*. Dordrecht: Reidel.
Earman, J. (1992). *Bayes or bust?* Cambridge, MA: MIT Press.
Flach, P., & Kakas, A. (Eds.) (2000). *Abduction and induction*. Dordrecht: Kluwer.
Fumerton, R. A. (1980). Induction and reasoning to the best explanation. *Philosophy of Science, 47*, 589–600.
Gabbay, D., & Woods, J. (2005). *The reach of abduction: Insight and trial.* A practical logic of cognitive systems, Vol. 2. Amsterdam: North-Holland.
Gemes, K. (1993). Hypothetico-deductivism, content, and the natural axiomatization of theories. *Philosophy of Science, 54*, 477–487.
Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170.
Glymour, C. (1981). *Theory and evidence*. Princeton: Princeton Univ. Press.
Glymour, C., Spirtes, P., & Scheines, R. (1991). Causal inference. *Erkenntnis, 35*, 151–189.
Grice, H. P. (1991). Logic and conversation. In S. Davis (Ed.), *Pragmatics: A reader* (pp. 305–315). New York: Oxford Univ. Press [2nd print].
Haig, B. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research, 40*, 303–329.
Halonen, I., & Hintikka, J. (2005). Towards a theory of the process of explanation. *Synthese, 143*, 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* (pp. 20–35). New York: Holt, Rinehart and Winston.

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, 34*, 503–533.
Hintikka, J., Halonen, I., & Mutanen, A. (1999). Interrogative logic as a general theory of reasoning. In J. Hintikka (Ed.), *Inquiry as inquiry* (pp. 47–90). Kluwer: Dordrecht.
Hobbs, J. R., et al. (1993). Interpretation as abduction. *Artificial Intelligence Journal, 63*, 69–142.
Holyak, K., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13*, 295–355.
Howson, C. (2000). *Hume's problem: Induction and the justification of belief*. Oxford: Clarendon Press.
Howson, C., & Urbach, P. (1996). *Scientific reasoning: The Bayesian approach* (2nd ed.). Chicago: Open Court.
Josephson, J., & Josephson, S. (Eds.) (1994). *Abductive inference*. New York: Cambridge Univ. Press.
Kitcher, P. (1981). Explanatory unification. *Philosophy of Science, 48*, 507–531.
Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.
Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: Chicago Univ. Press.
Kuipers, T. A. F. (2000). *From instrumentalism to constructive realism*. Dordrecht: Kluwer.
Kuipers, T. A. F. (2004). Inference to the best theory, rather than inference to the best explanation. In F. Stadler (Ed.), *Induction and deduction in the sciences* (pp. 25–51). Dordrecht: Kluwer.
Ladyman, J. (2002). *Understanding philosophy of science*. London: Routledge.
Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–195). Cambridge: Cambridge Univ. Press.
Langley, P., et al. (1987). *Scientific discovery. Computational explorations of the creative process*. Cambridge, MA: MIT Press.
Lipton, P. (1991). *Inference to the best explanation*. London: Routledge.
Magnani, L. (2001). *Abduction, reason, and science*. Dordrecht: Kluwer.
Mill, J. St. (1865). *System of logic* (6th ed.). London: Parker, Son, and Bourn.
Molnar, G. (1999). Are dispositions reducible? *Philosophical Quarterly, 49*, 1–19.
Moser, P. K. (1989). *Knowledge and evidence*. Cambridge: Cambridge Univ. Press.
Mumford, S. (1998). *Dispositions*. Oxford: Oxford Univ. Press.
Niiniluoto, I. (1999). Defending abduction. *Philosophy of Science* (Proceedings), Vol. 66, pp. S436–S451.
Octoby, D. W., et al. (1999). *Modern chemistry*. Orlando: Saunders College Publ.
Otte, R. (1981). A critique of Suppes' theory of probabilistic causality. *Synthese, 48*, 167–189.
Pap, A. (1978). Disposition concepts and extensional logic. In R. Tuomela (Ed.), *Dispositions* (pp. 27–54). Dordrecht: Reidel.
Paul, G. (1993). Approaches to abductive reasoning. *Artificial Intelligence Review, 7*, 109–152.
Pearl, J. (2000). *Causality*. Cambridge: Cambridge Univ. Press.
Peirce, C. S. (1878). Deduction, induction, and hypothesis. In Peirce (CP) 2.619–2.644.
Peirce, C. S. (1903). *Lectures on pragmatism*. In Peirce (CP) 5.14–5.212.
Peirce, C. S. (CP). In C. Hartshorne P. Weiss (Eds.), *Collected papers* (pp. 1931–1935). Cambridge, MA: Harvard Univ. Press.
Pollock, J. (1986). *Contemporary theories of knowledge*. Maryland: Rowman & Littlefied.
Prior, E. W., Pargetter, R., & Jackson, F. (1982). Three theses about dispositions. *American Philosophical Quarterly, 19*, 251–257.
Quine, W. v. O. (1974). *Roots of reference*. Open Court: La Salle.
Reichenbach, H. (1956). *The direction of time*. Berkeley: Univ. of California Press.
Ridley, M. (1993). *Evolution*. Oxford: Blackwell Scientific Publications.
Rock, I. (1984). *Perception*. New York: Scientific American Books.
Russell, S. J., & Norvig, P. (Eds.) (1995). *Artificial intelligence*. Englewood-Cliffs: Prentice Hall.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton Univ. Press.
Schurz, G. (1991). Relevant deduction. *Erkenntnis, 35*, 391–437.
Schurz, G. (1996). The role of negation in non-monotonic logic. In H. Wansing (Ed.), *Negation* (pp. 197–231). Berlin: W. de Gruyter.
Schurz, G. (2001). What is 'normal'? An evolution-theoretic foundation of normic laws. *Philosophy of Science, 28*, 476–497.
Schurz, G., & Lambert, K. (1994). Outline of a theory of scientific understanding. *Synthese, 101/1*, 65–120.

Sebeok, T. A., & Umiker-Sebeok, J. (1980). *'You know my method'. A juxtaposition of Charles S. Peirce and Sherlock Holmes*. Bloomington, IN: Gaslight Publ.
Shelley, C. (1996). Visual abductive reasoning in archaeology. *Philosophy of Science, 63*, 278–301.
Sober, E. (1993). *Philosophy of biology*. Boulder: Westview Press.
Sperber, D., et al. (Eds.) (1995). *Causal cognition*. Oxford: Clarendon Press.
Swinburne, R. (1979). *The existence of God*. Oxford: Clarendon Press [revised 2nd ed. 2004].
Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press.
Thagard, P. (1992). *Conceptual revolution*. Princeton: Princeton Univ. Press.
Van Fraassen, B. (1980). *The scientific image*. Oxford: Clarendon Press [reprint 1990].
Walton, D. (2004). *Abductive reasoning*. Tuscaloosa: Univ. of Alabama Press.