Towards a Theory of Abduction Based on Conditionals

Rolf Pfister

Munich Center for Mathematical Philosophy (MCMP),
Ludwig-Maximilians-Universität München, Germany
rolf.pfister@posteo.de

Abstract

Abduction is considered the most powerful, but also the most controversially discussed type of inference. Based on an analysis of Peirce's retroduction, Lipton's Inference to the Best Explanation and other theories, a new theory of abduction is proposed. It considers abduction not as intrinsically explanatory but as intrinsically conditional: for a given fact, abduction allows one to infer a fact that implies it. There are three types of abduction: Selective abduction selects an already known conditional whose consequent is the given fact and infers that its antecedent is true. Conditional-creative abduction creates a new conditional in which the given fact is the consequent and a defined fact is the antecedent that implies the given fact. Propositional-conditional-creative abduction assumes that the given fact is implied by a hitherto undefined fact and thus creates a new conditional with a new proposition as antecedent. The execution of abductive inferences is specified by theory-specific patterns. Each pattern consists of a set of rules for both generating and justifying abductive conclusions and covers the complete inference process. Consequently, abductive inferences can be formalised iff the whole pattern can be formalised. The empirical consistency of the proposed theory is demonstrated by a case study of Semmelweis' research on puerperal fever.

Keywords

abduction, inference, discovery, conditional, inferentialism, inference to the best explanation

Table of Content

1 Introduction

Abduction is often described as an inference that allows one to infer a potential explanation for a given fact. However, there is no commonly accepted definition of abduction: It is the least theoretically understood type of inference and the "status of abduction is very controversial. When dealing with abductive reasoning misinterpretations and equivocations are common" (Magnani 2015, p. 313).

One reason for this is that the term 'abduction' is used by many, quite different theories: Peirce (1958; 1998) introduced the term and developed two different concepts. Harman (1965) links Peirce's abduction with his own theory of 'Inference to the Best Explanation' (IBE), which was thoroughly revised by Lipton (2004). IBE is often also called abduction (Campos 2011, pp. 419f), although many consider this to be highly misleading (cf. Park 2015, pp. 228–234; Mcauliffe 2015). Moreover, ambiguity arises as some theories interpret abduction as a logical syllogism, while others view it primarily as a computational method or as a process of epistemic change (Beirlaen and Aliseda 2014, p. 3749).

Regardless of the differences, many theories regard abduction as a cornerstone of scientific methodology (cf. Douven 2017a, sect. 1.2). It is considered the most insecure but also the most insightful kind of inference since "all the ideas of science come to it by the way of Abduction" (Peirce 1958, CP 5.145). It is the only kind of inference that allows introducing new kinds of concepts, which is also seen as the essential difference to inductive inferences (e.g. Campos 2011, p. 428; Psillos 2002, pp. 610f). For example, Psillos (2009, pp. 122, 144f) states that "no new ideas are generated by induction" since "[t]he extra content generated by induction is simply a generalisation of the content of the premises. Hence, with enumerative induction, although we may arguably gain knowledge of hitherto unobserved correlations between instances of the attributes involved, we cannot gain 'novel' knowledge, i.e., knowledge of entities and causes that operate behind the phenomena" (cf. Peirce 1958, CP 5.145, 6.475, 7.202; Minnameier 2004, pp. 78f).

In contrast, at least some kinds of abductive inferences allow for the introduction of new types of concepts. For example, Schurz (2008, p. 201) uses the common distinction between selective abductions and creative abductions, where the former ones "choose the best candidate among a given multitude of possible explanations" and the latter ones "introduce new theoretical models or theories".

When examining inferences, it is common to distinguish between the context of discovery and the context of justification. The context of discovery concerns the generation of a new hypothesis, whereas the context of justification concerns its quality. Even though the distinction is often attributed to Reichenbach, it can be found earlier in Popper, the Wiener Kreis, Husserl, Whewell, and Herschel; some trace it further back to Kant or even to Aristotle and Euclid (Hoyningen-Huene 1987, pp. 502f).

The distinction allows one to analyse the execution of inferences: The context of discovery examines how one creates a particular hypothesis. The context of justification examines conditions under which an inference is good, but it does not provide guidance on how to generate specific hypotheses. Nevertheless, although the distinction is helpful, it is arbitrary: The discovery of a new hypothesis is already influenced by justificatory considerations; otherwise, it would be very unlikely to generate a promising hypothesis by only a few trials (Peirce 1958, CP 7.220).

Many controversies in the 20th century regarding the philosophy of discovery revolved around the disagreement whether the generation of hypotheses is part of the scientific process or not (Schickore 2018, sect. 3). Some, e.g. Popper (1959, pp. 30–32) and Hempel (1966, p. 15), argue that unlike the justification of hypotheses, their generation is completely illogical and therefore not part of the scientific process. In opposition, others developed different accounts to capture the generation of hypotheses. Some accounts see discovery as a logical process, whilst others claim that it is not logical but follows

analysable patterns, is governed by a methodology, or is at least amenable to philosophical analysis (cf. Schickore 2018, sect. 6–9; Paavola 2006a, ch. 3).

Consequently, some theories see abduction as a process of generating hypotheses, while others see it as a process of evaluation or as a combination of both (cf. Beirlaen and Aliseda 2014, p. 3734; Paavola 2006b, p. 93). Still other theories leave the generation and selection of hypotheses open due to the numerous unanswered questions and focus on other aspects of abduction (cf. Woods 2011, pp. 242f).

In conclusion, although the discussion to which extent abductive inferences can be formalised is considered important (cf. Psillos 2009, p. 148), there is as yet no consensus. As Schurz (2016, p. 496) states, the major challenge is therefore to find out whether there are formally explicable rules and strategies that allow the execution of abductive inferences.

This article intends to address this challenge. The aim is to lay the foundation for a theory of abduction that overcomes the limitations of current ones and covers both the context of discovery and the context of justification. If possible, the theory should allow to formalise the process of abduction, which would allow its application in the field of computer science and artificial intelligence as well as its practical validation.

In order to achieve this goal, the article presents an approach of abduction that is based not on explanations but on conditionals. The article is divided into seven sections. Section 2 examines various important properties of abduction based on an analysis of Peirce's retroduction and Inference to the Best Explanation. Section 3 offers a discussion of conditionals and, in particular, inferentialism. Building on all this, a definition of abductive inferences founded on conditionals is given in Section 4. The different types of abductive inferences are discussed in Section 5, in which moreover the use of analogies in patterns is explored. Section 6 examines the conditions under which abductive inferences can be formalised, and finally a conclusion is drawn in section 7.

2 Properties of abductive inferences

2.1 Introduction of new concepts

In his later works, Peirce (1958, CP 5.189; 1998, EP2 p. 231) introduces his revised concept of abduction, often referred to as retroduction[1], for which he provides the following definition:

> The surprising fact, C, is observed;
> But, if A were true, C would be a matter of course.
> Hence, there is reason to suspect that A is true.

Peirce (1958, CP 5.188) regards abduction as an inference that allows new concepts to be introduced. This seems to contradict the definition above, since the concept 'A' derived by the conclusion is already given by the second premise and is therefore not new. However, as Anderson (1987, p. 25) explicates, the premise is not to be understood in the sense that it actually already contains the new concept 'A' but "in the sense that there is a logical relation between premises and conclusion". The definition only specifies the logical order, but not the temporal order. Consequently, the concept 'A' can be newly introduced in both the premise and the conclusion at the same time (p. 35).

---

[1] Peirce uses the terms 'abduction', 'hypothesis', and 'retroduction'. While 'hypothesis' is often associated with his earlier concept of abduction, 'retroduction' is more often used to name his later concept; yet, he uses all three terms in his later works as well (Paavola 2006a, pp. 40f).

The introduction of the new concept 'A' is achieved through a creative act, which Peirce (1998, EP2 p. 227) describes as follows: "The abductive suggestion comes to us like a flash. It is an act of insight, although of extremely fallible insight. It is true that the different elements of the hypothesis were in our minds before; but it is the idea of putting together what we had never before dreamed of putting together which flashes the new suggestion before our contemplation."

The origin of all abductive insights lies in perception, which is the basis of all knowledge (Rosenthal 2004, p. 193). Perception leads to perceptual judgements that are formed into abductive conjectures (Campos 2011, p. 428). However, there is no clear distinction between perceptual judgments and abductive conjectures; rather, the "abductive inference shades into perceptual judgment without any sharp line of demarcation between them; or in other words our first premises, the perceptual judgments, are to be regarded as an extreme case of abductive inferences, from which they differ in being absolutely beyond criticism" (Peirce 1998, EP2 p. 227).

Closely related to perception is imagination. Both have signs as semiotic outcomes and complement each other. As Campos (2011, p. 429) states: "When a perceptual judgment disrupts our expectations and presents us with a problem, the imagination works to form schemata or diagrams of the situation, searching for explanations. In the case of abduction, explanatory hypotheses are signs – diagrams that rearrange the relations among facts so as to explain them. Sometimes new elements (explanatory facts) are introduced into the diagrammatic hypothesis to explain the perceived, unexpected facts. 'Diagrams' or explanatory schemata may include formalized theories, equations, statistical models, figures, representations of atomic or molecular structures, and so on. The abductive insight consists in associating or relating explanatory and perceived facts in a novel way."

The capacity for abductive insight is an instinctive endowment of humans that enables them to find a correct hypothesis within a small number of guesses, despite the myriads of possible hypotheses (Peirce 1958, CP 7.220).

In summary, the creative act leading to an insight which introduces the new concept 'A' is an immanent part of abduction. For this reason, Peirce describes abduction as both an insight and a logical inference. He explicates "that abduction, although it is very little hampered by logical rules, nevertheless is logical inference, asserting its conclusion only problematically or conjecturally, [...] but nevertheless having a perfectly definite logical form".

Since Peirce's definition describes creative insights as instinctive, it does not allow for a fully formal account of abductive inferences (cf. Tschaepe 2014, pp. 121–124). In comparison, Schurz (2016, p. 494) provides the following formal structure of abductive inferences:

> Premise 1: A (singular or general) fact E that is in need of explanation.
> 'Premise' 2: A background knowledge K, which implies for a hypothesis H that H is a
>   possible and sufficiently plausible explanation for E.
> ==
> Abductive conjecture: H is true.

Similar to Peirce's account, the hypothesis 'H' of the conclusion is already referred to in the second premise. However, the background knowledge only supports the hypothesis 'H' but does not necessarily contain it itself. Besides that, unlike Peirce, Schurz does not presuppose a creative act of insight. Instead, the background knowledge 'K' can imply in a purely formal way that a hypothesis 'H' is a possible and sufficiently plausible explanation for the fact 'E'.

Consequently, Schurz's account allows for fully formalised abductive inferences that introduce new concepts in the conclusion that are not part of the premises. Nevertheless, non-formalisable abductive

inferences can also be represented: this through the background knowledge representing a non-formal process such as Peirce's intuitive creative act.

2.2 Surprisingness and observability

Peirce requires the fact 'C' to be surprising. The characteristic of surprise can trigger an abductive inference: Since surprising facts do not match our expectations, they can lead to promising new insights (Paavola 2004, p. 274). However, for the inference itself, the surprisingness of the fact does not matter, as it does not influence the generation or justification of the hypothesis. In addition, there are also many non-surprising circumstances in which abduction can be insightful, e.g. when results of a scientific experiment are to be further investigated. Therefore, even if surprisingness can be an indicator for promising investigations, its necessity should be dismissed.

Schurz (2008, p. 216; 2016, p. 495) requires for all kinds of creative abduction that the facts are observable.[2] However, abductive reasoning is desirable and used for unobservable facts, e.g. for the structure of molecules or radiation. Schurz (2008, p. 206; 2016, p. 499) also requires for all kinds of hypothetical cause abduction that the inferred hypothesis is unobservable. Yet, abduction is used for observable causes as well; for instance, one concludes that some birds fly away because a predator is approaching. It seems that the (non-)observability of a fact is relevant for the subsequent examination of a hypothesis, but not for the inference itself.

Additionally, the meaning of the fact 'C' should be understood in a broad sense. It could not only be a fact that is known to be true, but also, for example, a hypothesis of another inference or an assumption of a thought experiment.

2.3 Process of abduction

A full theory of abduction must provide a precise and complete description of how abductive inferences are performed. This is true for both the context of discovery, i.e. how a specific hypothesis is generated, and the context of justification, i.e. how the quality of a hypothesis is evaluated. In the following, I will focus mainly on Peirce's retroduction and on Inference to the Best Explanation (IBE), which are considered the most popular theories of abduction.

IBE's basic idea is that "explanatory considerations contribute to making some hypotheses more credible, and others less so" (Douven 2017a, sect. 4). Thus, given a multitude of abductive hypotheses, IBE allows one to determine which is the best hypothesis, i.e., the one most likely to be true. Different accounts of IBE suggest varying explanatory virtues that make a hypothesis preferable (cf. Cabrera 2017, pp. 1248–1250). It is still under discussion, which explanatory virtues should be considered.

In addition, it is unclear why explanatory virtues are an indicator of truth (Cabrera 2017, sect. 3). At least some of the suggested virtues, e.g. precision and scope, are non-confirmatory and only informational virtues: they do not indicate which hypotheses are true but rather which provide greater informational content and meet the goals of science (Cabrera 2017, sect. 3.3, 5.1). Hence, some (e.g.: Dawes 2013; Cabrera 2017; Jones 2018) suggest that IBE is not about justification but pursuit, that is, identifying hypotheses worthy of further investigation.

This view is also supported by practice: Darwin's hypothesis of heredity, pangenesis, fulfilled explanatory virtues but was rejected by the biological community because of missing empirical evidence. Similarly, the chromosome theory offered overwhelming explanatory power, but could not

---

[2] Likewise, Peirce speaks of the fact 'C' to be "observed". However, in his time the term had a much broader meaning. Therefore, I am uncertain whether he requires the fact 'C' to be tangible or not.

gain acceptance until both the existence and the causal power of chromosomes were demonstrated in subsequent experiments (Novick and Scholl 2020, sect. 3, 5).

Furthermore, if one considers explanatory virtues not as an indicator of truth but of informational content, one can explain why scientists accept[3] contradictory hypotheses and theories. For example, quantum mechanics and general relativity are incompatible, but both have great explanatory power (and are empirically successful). Many scientists do not believe them to be true, but accept them because both provide a solid basis for further reasoning (cf. Dawes 2013, sect. 1.3).

Peirce proposes several virtues that abductive inferences should fulfil: They should be simple, natural and plausible to us (Peirce 1958, CP 6.447) and should cost us as little effort as possible (CP. 5600, 7.220). They should explain all relevant facts (CP 7.235), have a unifying power (CP 7.221, 7.410), be licenced by existing background beliefs (Psillos 2009, p. 136) and their plausibility should be discriminated from their antecedent likelihoods (Peirce 1958, CP 5.599). Finally, hypotheses should be experimentally testable by entailing deductive and inductive predictions (CP 7.220).

Peirce argues that science is severely limited by economical constraints: "the process of verification [...] is so very costly in time, energy, and money" (CP 5.602). The suggested virtues allow one to determine which hypothesis can be tested most efficiently and should therefore be investigated further first (Peirce 1958, CP 7.220, 5.602; McKaughan 2008, pp. 452–458). As Peirce (1958, CP 1.120) states, "[t]he best hypothesis [...] is the one which can be the most readily refuted if it is false. This far outweighs the trifling merit of being likely". Thus, his proposed virtues are not about justification, but about pursuit worthiness.

To justify an inferred hypothesis, Peirce advocates determining by deduction necessary consequences that follow from it. Their truth can be tested experimentally and, by induction[4], it can be concluded that if the consequences of the hypothesis are true, then the hypothesis itself is true (Peirce 1958, CP 7.203, 7.206). However, besides that, Peirce remains rather general and does not provide specific methods or concrete conditions under which a hypothesis is considered justified. One reason for this is that Peirce (1958, CP 7.679f, 5.173, 2.753; 1998, EP2 pp. 443f) considers the human instinct to have an innate tendency "to conjecture rightly".[5] Thus, the justification is already provided by the human endowment and the correct hypothesis can be found within a few trials through experimentation. Overall, many regard Peirce's theory as one of discovery rather than justification (e.g. Minnameier 2004; Campos 2011; Douven 2017a, Supplement: Peirce on Abduction).

As far as the context of discovery is concerned, Peirce's considerations are quite detailed (cf. sect. 1.1). Yet, since the "abductive suggestion comes to us like a flash [and] is an act of insight", our explanatory suggestions "are not subject to rational self-control" (Peirce 1998, EP2 p. 227). Only once they have been created can we access them logically. Peirce thus describes the process of discovery in great detail, but he does not provide a method – indeed he rejects its possibility – by which one can deliberately

---

[3] A hypothesis or a theory is considered acceptable here in case it is worthy of commitment as a research program (cf. Cabrera 2017, pp. 1267–1270).
[4] An analysis of Peirce's understanding of induction, which includes three different types, is provided by Fann (1970, pp. 32f).
[5] Peirce (1958, CP 5.173) states "that man has a certain Insight, not strong enough to be oftener right than wrong, but strong enough not to be overwhelmingly more often wrong than right". This seems to contradict some of his other statements where he argues that "proposals for hypotheses inundate us in an overwhelming flood" (Peirce 1958, CP 5.602). In my opinion, the contradiction can be resolved by considering both aspects as two successive steps of the hypothesis generation process. When one experiences a surprising fact, there are an infinite number of possible explanations. But out of these myriads of possible hypotheses, the human instinct intuitively considers only a few promising ones, of which one becomes aware.

create abductive hypotheses. Instead, we must rely on our instinctual human endowment (Peirce 1958, CP 7.220).

IBE is primarily viewed as a theory of justification, where candidate hypotheses are usually already given (cf. Douven 2017a, Introduction; Lange 2020, p. 4). Nevertheless, there are at least some approaches that address the context of discovery. For instance, Lipton (2004, p. 59, 149–151) proposes IBE as a two-filter approach: The first filter generates a set of promising hypotheses by contrastive analysis and consideration of background knowledge. The second filter, based on explanatory virtues, selects then the best hypothesis among the generated ones. Lipton illustrates this approach with the research of Semmelweis, who investigated why cases of puerperal fever were much higher in one clinic of the Vienna maternity hospital than in the other.[6]

According to Lipton (2004, p. 83), the generation of new hypotheses begins with a contrastive analysis: For the fact to be explained, one needs a foil with a similar history, because "this sharply constrains the class of hypotheses that are worth testing". For example, Semmelweis was able to compare the conditions of the two clinics with each other as well as with those of women who had street births on the way to hospital (Semmelweis 1861, pp. 2–4, 43–46; Shorter 1984, p. 49).

As Lipton (2004, p. 149) notes, contrastive cases will never have just one difference, but many. To further reduce the number of possible hypotheses based on these differences Lipton suggests relying on background knowledge (pp. 139, 149–151). It allows considering already known explanations, determining the unificatory virtues of the hypotheses, and providing explanatory standards. For instance, Semmelweis (1861, pp. 4–10; Shorter 1984, p. 51; Scholl 2013, pp. 67–72) rejected epidemic factors and focused on endemic ones, as only the latter could explain why only one but not both clinics had high mortality rates. Moreover, Semmelweis (1861, pp. 32f) rejected the hypothesis that puerperal fever could be caused by fear of death, as this was not compatible with his background knowledge: he could not imagine how a mental state could lead to the strong physical manifestations of puerperal fever.

But even if one can further narrow down the number of potentially interesting differences, e.g. to endemic factors, there is still an infinite number left that needs to be considered. Semmelweis (1861, pp. 4–39, 51f, Shorter 1984, p. 52) considered delivery positions, exposure to a priest giving the last rites, rough examinations as well as many other differences. But despite his detailed investigation, still many more possible explanations would remain that fit well with the background knowledge: such as poisonous air from a nearby factory, inadequate cleaning of the place, or dangerous behaviour by non-examining staff. Hence, taking background knowledge into account may increase the chances of finding important differences more quickly, but it does not solve the problem of multiple differences as Lipton (2004, p. 128) intends.

Moreover, the method is highly dependent on the availability of suitable contrastive cases. Semmelweis was in the fortunate position of being able to compare two very similar clinics from the same hospital; had there been only one clinic, it would have been much more difficult to find a promising contrastive case. For other cases, e.g. the discovery of gravity or the explanation of heredity, it is not clear how to find suitable contrastive cases at all.

After many unsuccessful attempts, Semmelweis finally succeeded in identifying the cause of the increased rate of puerperal fever in one of the clinics: There, medical staff regularly performed autopsies before examining women in labour. In doing so, they transferred 'cadaverous particles'[7] that infected

---

[6] Since Semmelweis' investigation is a popular case study within the philosophy of science, I will not present it here in its entirety. The original German text of Semmelweis (1861) is available online, a short excerpt of important passages in English is provided by Shorter (1984).

[7] More precisely, puerperal fever is not caused by cadaveric matter but by bacteria living on it.

the women and caused the fever. However, Semmelweis did not reach the conclusion by comparing differences between the two clinics and identifying the performance of autopsies as a relevant one.

Instead, one of his colleagues was pricked with a knife while performing an autopsy and developed all the symptoms of puerperal fever before eventually dying. Semmelweis (1861, pp. 52–55; Shorter 1984, p. 52) was certain that the cause for his death was the autopsy knife that contaminated him with cadaverous particles. By analogy, Semmelweis concluded that the particles were also transmitted to the women in labour, via the hands of the medical staff.

Similarly, a while later there was another accumulation of cases of puerperal fever. From this, Semmelweis (1861, pp. 59f; Shorter 1984, p. 54) concluded that puerperal fever "is caused not only by cadaverous particles adhering to hands but also by ichor from living organisms". Again, the conclusion was reached by analogy and not by a contrastive analysis that reveals relevant differences.[8]

Lipton's two-filter approach suggests that once several potential explanations have been generated, one uses the second filter, based on the explanatory virtues, to determine the best, i.e. the actual explanation. Lipton (2004, p. 89f) argues: "When Semmelweis inferred the cadaveric hypothesis, it was not simply that what turned out to be the likeliest hypothesis also seemed the best explanation: Semmelweis judged that the likeliest cause of most of the cases of childbed fever in his hospital was infection by cadaveric matter because this was the best explanation of his evidence."

However, this description is not accurate: Semmelweis did not develop a range of possible explanations, evaluate their explanatory power, and take the best one. Instead, Semmelweis developed and tested one hypothesis after another over three years until he found one that could be experimentally verified. It was thus not an inference to the best explanation, but to the only one (Paavola 2006b, p. 106).

Lipton (2004, pp. 90, 149) is mindful of this discrepancy and argues that Semmelweis was in a fortunate position, but typically several candidate explanations remain and then explanatory virtues come into play. Nevertheless, Lipton is also aware of the role of experimentation and the elimination of hypotheses until only one remains. The importance of experimentation is also evident in Semmelweis' case: Semmelweis, as well as others in the scientific community, did not accept his hypothesis until he could experimentally support it in clinical interventions and in several animal experiments (Semmelweis 1861, pp. 55–58, 76–80; Scholl 2013, pp. 72–75). Other practical examples, such as the discovery of AIDS (Bird 2010, pp. 349f) or the heredity theory already mentioned (Novick and Scholl 2020), provide further support for the preference for this type of justification: In both cases, explanations were accepted not by their explanatory virtues, but by empirical verification and the elimination of all other available hypotheses.

In conclusion, both Peirce's retroduction and IBE fall short of providing a precise and complete description of how abductive inferences are performed. Peirce's retroduction does not concern the justification but only the generation of hypotheses and although the discovery is described in great detail, it remains inaccessible as it is considered as an instinctual human endowment. IBE offers methods for both generating and justifying hypotheses, but they fall short from both a theoretical and a practical perspective.

---

[8] Precisely, Semmelweis found differences with regard to the cause of childbed fever. For example, he observed that in the clinic with the higher infection rates, neighbouring patients frequently fell ill together, while in the other clinic patients fell ill in a scattered manner. However, Semmelweis was not able to use this difference to find the cause; in fact, it led him away from the correct solution: From the scattered distribution Semmelweis (1861, pp. 47f; Shorter 1984, p. 50) concluded "that puerperal fever was not a contagious disease and that the disease was not spread from bed to bed by pathogens".

2.4 Explanatoriness

Peirce, Lipton and many others state that the main purpose of abduction is to provide an explanation for a given fact. So far, however, there is no generally accepted theory of explanation. Proponents of IBE do not consider this as problematic: IBE and other abductive theories do not presuppose any particular explanatory theory, but are compatible with at least most of them (Lipton 2004, p. 2; Cabrera 2017, pp. 1250f).

However, the underlying explanatory theory does significant conceptual and justificatory work; if it is not specified, the central element of IBE is missing (Cabrera 2020, pp. 731f). For example, as long as the explanatory theory is not specified, it is not clear which hypotheses qualify as explanations and therefore, among which hypotheses the best explanation should be chosen (cf. Klärner 2003, pp. 57–61).

In addition, explanatory theories influence the coverage of IBE: For example, Lipton's (2004, pp. 30–33) theory of explanation allows only causal explanations, although non-causal explanations also exist, e.g. in mathematics, philosophy and physics. This not only makes it impossible to provide explanations for non-causal circumstances (cf. Klärner 2003, pp. 202–204), but also calls into question the applicability of IBE in general: It may be that even if causal explanations are possible, the best explanation is a non-causal one. Thus, if the set of available explanations contains only causal explanations, the best explanation may not be considered and another, wrong explanation may be chosen instead.

Moreover, many explanatory theories, such as the presently discussed counterfactual theory of explanation (Reutlinger 2018, pp. 78–81), do not provide any explanatory virtues. Yet, these virtues are required by IBE to determine which is the best explanation amongst the possible ones. IBE requires furthermore that the explanatory virtues enable comparative evaluation and, if there are several, that they can be rated against each other (cf. Klärner 2003, pp. 61–64, 117–121, 207–211).

To avoid the problem of not having a suitable explanatory theory, Cabrera (2020, pp. 744–746) suggests that IBE should not rely on a theory of explanation, but only on explanatory virtues themselves, since they do the intended justificatory work. Others question the claim that abduction is intrinsically explanatory at all, i.e. that abductive hypotheses have to be explanations. For instance, Park (2015, pp. 220–222) considers the requirement to be ill-founded and based not on theoretical but only on practical motivations, such as providing useful constraints.

Furthermore, not all types of abductively derived conclusions seem to be explanatory. Schurz (2008, pp. 230f) as well as Gabbay and Woods (2005, pp. 122f) remark that at least some kinds of abductions are implausible and purely instrumental, i.e. they provide true predictions but are unlikely to be true themselves. For instance, the action-at-a-distance equation "serves Newton's theory in a wholly instrumental sense. It allows the gravitational theory to predict observations that it would not otherwise be able to predict" (Magnani 2009, p. 77). Such purely instrumental abductions not only contradict IBE's pursuit of truth, they are also incapable of explanation as they are false. Yet, instrumental abductions are of scientific interest because they provide otherwise unobtainable predictions.

A similar kind of inference can be found in mathematics. In general, there, one reasons deductively from some given axioms to some target theorems. However, it is also possible to infer from given theorems to axioms (Easwaran 2008, pp. 383–385; cf. Niiniluoto 2018, ch. 2). As Baker (2020, sect. 2.2.2) notes, "the propositions of elementary arithmetic – '2 + 2 = 4', '7 is prime', etc. – are much more self-evident than the axioms of whatever logical or set-theoretic system one might come up with to ground them. [...] Deriving '2 + 2 = 4' from our set-theoretic axioms does not increase our confidence in the truth of

'$2 + 2 = 4$', but the fact that we can derive this antecedently known fact (and not derive other propositions which we know to be false) does increase our confidence in the truth of the axioms".

The derivation of axioms from given theorems does not aim at explanatory results (Magnani 2009, pp. 72, 122, cf. 119–139). Rather, it should make it possible to discover suitable axioms for mathematics (Magnani 2009, p. 72), to systematise uncontroversial facts, to prove further theorems (Easwaran 2008, p. 383), and to discover new theorems (Schlimm 2013, p. 40). Here, too, the conclusions are instrumental and do not necessarily lead to truth (Easwaran 2008, pp. 384f). In addition, the relevance and applicability of truth in mathematics in general are still controversial (Easwaran 2008, p. 384; Baker 2020, sect. 2.2.2). Hence, an explanatory account does not seem to be able to capture the inference of axioms in mathematics. As a possible solution, Heron (2020) proposes an account to justify axioms that relies on theoretical virtues but not on explanations.

In conclusion, it remains unclear why abductive inferences should be intrinsically explanatory. Instead, various kinds of abductively derived conclusions are instrumental, do not lead to truth, and neither should nor can explain the given fact. Thus, abductive inferences can provide explanations, and often they do, but they do not necessarily have to.

3 Conditionals as the basis of abduction

3.1 Special properties of conditionals

In fact, it seems that abductions are not intrinsically explanatory, but that for a given fact they allow one to infer another fact that implies it. Such an implication can be represented by a conditional of the form 'If A, [then] C.'. The consequent 'C' represents the given fact and the antecedent 'A' represents the to be inferred fact that implies the consequent.

In many abductive cases, the implying fact 'A' is taken to explain the implied fact 'C' - but as shown above, while this is true in most cases, it is not true in all cases. The confusion arises because explanations are often expressed through conditionals, but not all conditionals are used for an explanation. In other words, being an explanation is not an intrinsic property of an abductive conclusion but a possible application for which it can be used.

It therefore seems more promising to base abduction on conditionals. Conditionals allow one not only to infer explanations, but all kinds of preceding facts. This includes non-explanatory facts such as instrumental models and axioms, which are common conclusions in science as well.

Furthermore, conditionals have two special properties that lead to the potential of abductive reasoning:

First, conditionals are asymmetrical: a conditional and its converse version, where the antecedent and the consequent are interchanged, are not logically equivalent ('If A, then C.' $\neq$ 'If C, then A.'). Only some logical operators have this property; e.g. in classical logic, material implication is the only asymmetrical binary truth function.[9] The asymmetry of conditionals allows one to represent relations in which one proposition implies the other, but not vice versa. Such relations are common in science, where, for example, laws are represented by conditionals. Such relations are also common in reasoning and predictions to infer what follows from certain assumptions.

---

[9] The material nonimplication, the converse implication, and the converse nonimplication are also asymmetric binary truth functions, but they can be expressed as more complex versions of the material implication.

Second, conditionals allow one to infer from the truth of the antecedent to the truth of the consequent. Conditionals are not the only logical operator that allows one to infer from the truth of one proposition the truth state of the other one. For example, it follows from the exclusive disjunction 'either p or q' and 'p' that 'not q'. Yet, the exclusive disjunction as well as the alternative denial let one infer from the truth of one proposition only the falsehood of the other. In contrast, the conditional and the logical biconditional allow one to infer from the truth of one proposition the truth of the other. The ability to infer the truth rather than the falsehood of a proposition is in general more informative, as science aims to find true rather than false statements.

Due to its asymmetry, a conditional only allows one to infer with certainty from the truth of the antecedent to the truth of the consequent, but not vice versa. The reverse inference from the truth of the consequent to the truth of the antecedent, called affirmation of the consequent (Godden and Zenker 2015, pp. 88–103), is uncertain and often considered as a fallacy. This is because the consequent can be implied not only by the antecedent of the conditional but also by another fact. Thus, for a high credibility of the conclusion, it must therefore be justified that the consequent is actually implied by the antecedent and not by something else (pp. 104–120). Abduction provides this justification by combining the two special properties of conditionals: It uses the valid entailment from the truth of the antecedent to the truth of the consequent to develop a justification that allows one to infer well-justified in the opposite direction – i.e. to infer uncertainly but plausibly from the truth of the consequent to the truth of the antecedent.

3.2 Conditional theory for abduction

Material implication is a conditional theory widely used in logic, but it leads to counterintuitive results (Evans and Over 2004, ch. 2, 3). Other conditional theories include mental model theory, suppositional theories and inferentialism, of which especially the latter are currently under discussion (cf. Douven, Elqayam, Singmann and van Wijnbergen-Huitink 2018, pp. 51–53).

Suppositional theories are based on the Ramsey test (Ramsey 1929/1990, p. 155), according to which the acceptability of a conditional[10] can be determined as follows: One hypothetically assumes that the antecedent is true and adds it to one's stock of beliefs, makes minimal changes if necessary to maintain consistency, and finally assesses the acceptability of the consequent of the conditional. If the consequent is accepted, the conditional is also accepted, otherwise it is not. Suppositional theories differ in their details, e.g. with regard to the truth values of a conditional whose antecedent is false. For example, Stalnaker's (1981) possible worlds semantics regards such a conditional as true in case its consequent is true in the nearest world in which its antecedent is true. In contrast, Evans (2020, p. 62) argues that people always think about a conditional on the supposition of its antecedent, hence cases with false antecedents are irrelevant.

Inferentialism is founded on the assumption that conditionals are used to express an inferential connection between their antecedent and their consequent.[11] A conditional is considered true iff its consequent follows argumentatively from its antecedent and possibly contextually relevant background knowledge (cf. Douven 2015, pp. 35–43).[12] The inferential connection can be of various kinds and be

---

[10] Unless specified otherwise, I refer in this article only to indicative conditionals, i.e. conditionals whose antecedents are in the indicative mood. Even though some of the considerations also apply to subjunctive conditionals, they require their own analysis.

[11] Introductions to inferentialism are provided by Skovgaard-Olsen (2016) and Douven, Elqayam, Singmann and van Wijnbergen-Huitink (2018, pp. 52–54), and a general introduction to conditionals based on relevance connections is offered by Égré and Rott (2021, sect. 7).

[12] In addition, the antecedent must be deductively consistent with the background knowledge, else conditionals with a logically false antecedent could count as true (Douven 2015, p. 38).

based, for example, on a logical, heuristic or causal relationship. Accordingly, the connection may consist of a series of deductive, inductive or abductive inferential steps. A deductive connection is certain and based on logical necessities; an inductive connection is uncertain and based on statistical considerations; and an abductive connection is uncertain and based on explanatory considerations[13].

Conditionals that have an inferential connection are called connected conditionals. In contrast, in unconnected conditionals the antecedent and the consequent have no clear connection and are probabilistically independent of each other. Unconnected conditionals often seem strange or misleading, like: "If George Washington was the first president of the United States, then Paris is the capital of France.".

Nevertheless, most suppositional theories judge a conditional to be true in case both its antecedent and its consequent are true, regardless of whether it is a connected or an unconnected conditional (e.g. Evans and Over 2004, ch. 9; Baratgin, Over and Politzer 2013). Insofar as unconnected conditionals are considered strange or misleading, this is attributed to the violation of pragmatic requirements, i.e. requirements concerning the way speakers make meaningful utterances (Evans 2020, pp. 64f; cf. Skovgaard-Olsen, Singmann and Klauer 2016, p. 27).

In contrast, inferentialism regards unconnected conditionals not only as a violation of pragmatic norms but as genuinely defective. This, because they are not able to fulfil their function of expressing reason relations (Skovgaard-Olsen 2016, sect 2.2; Vidal and Baratgin 2017, p. 778). Reason relations are necessary for reasoning, prediction, and argumentation: They allow one to infer from the antecedent to the consequent and to estimate which propositions increase or decrease the probability of other propositions.

Beyond explaining the strangeness of unconnected conditionals, inferentialism is also able to match intuition about the or-to-if principle and provides a solution to Gibbard's Riverboat argument (Krzyżanowska, Wenmackers and Douven 2014). Furthermore, it is able to provide satisfying interpretations for complex cases that cannot be interpreted successfully by other conditional theories (Skovgaard-Olsen 2016, pp. 575–577).

Nevertheless, inferentialism is still under development and not all aspects have been clarified (Douven 2017b, pp. 1150–1153). For example, since it is pluralistic and allows for different types of connections, it is not yet clear which connections are permissible and which properties they must fulfil. Furthermore, it is unresolved whether conditionals can only be either true or false, or whether they can also be neither true nor false but void – which is how they are sometimes assessed in empirical studies (cf. Skovgaard-Olsen, Kellen, Krahl and Klauer 2017, p. 462).

Another unresolved issue is the determination of the probability of connected conditionals. One possibility is to use the conditional probability hypothesis P(if A, then C) = P(C|A) as suggested by many suppositional theories (e.g. Evans and Over 2004, ch. 9; Fugard, Pfeifer, Mayerhofer and Kleiter 2011; Evans 2020). Alternatively, the probability can be determined by the strength of the inferential connection. The two evaluation methods differ in the factors they take into account: The latter considers only the inherent inferential connection between the antecedent and the consequent; the former incorporates also other factors that influence the consequent.

---

[13] The term abduction is understood here in a different sense than in the rest of the article. According to Mirabile and Douven (2020, p. 5), in an abductive connection, the antecedent is best explained by the consequent, which is therefore probably true. The definition builds on IBE and as such is subject to the same criticism described in sections 2.3 and 2.4. I will discuss the relationship between this notion of abduction and the one presented in the rest of the article in detail at the end of section 4.

As an example, consider the conditional "If my neighbour throws a party, then I cannot sleep well at night.". Given that the neighbour is only every other time so loud that one cannot sleep, the probability of the conditional is 0.5 according to both evaluation methods. Now, additionally assume you cannot sleep well at night anyway because of insomnia. Then, based on the strength of the inference relation, the probability of the conditional is still 0.5, while according to the conditional probability hypothesis it becomes 1.

The conditional probability hypothesis alters the probability of uncertain conditionals in case the consequent is influenced by another, non-exclusive factor. Consequently, the probability of a conditional can change depending on other provided facts, although the inferential connection between its antecedent and its consequent remains the same. This seems incoherent with the purpose of conditionals to express a reason relation, since the probability reflects not only the relation itself but also unrelated factors. Therefore, evaluating the probability of a conditional based on the strength of the inferential connection seems preferable.

Empirically, there is evidence both for inferentialism (Douven, Elqayam, Singmann and van Wijnbergen-Huitink 2018; Mirabile and Douven 2020; Vidal and Baratgin 2017; Skovgaard-Olsen, Kellen, Hahn and Klauer 2019) as well as for suppositional theories (Over, Hadjichristidis, Evans, Handley and Sloman 2007; Fugard, Pfeifer, Mayerhofer and Kleiter 2011; Cruz and Oberauer 2014; Baratgin, Over and Politzer 2013).

However, the ambiguous results can be explained by a variety of factors (Skovgaard-Olsen, Singmann and Klauer 2016; Skovgaard-Olsen, Kellen, Hahn and Klauer 2019) and studies specifically comparing the two conditional theories provide support for inferentialism (Mirabile and Douven 2020, p. 26; Skovgaard-Olsen, Kellen, Hahn and Klauer 2019; Krzyżanowska, Collins and Hahn 2021; Nickerson, Barch and Butler 2019, pp. 61f; Krzyżanowska and Douven 2018; Douven, Elqayam and Mirabile (in submission)).

In conclusion, inferentialism is able to provide a coherent understanding of conditionals in accordance with empirical results. Moreover, it accounts for the connection between the antecedent and the consequent – which can be used in abductive reasoning to develop a justification that the given fact, which constitutes the consequent, is plausibly implied by the antecedent and not by some other, unconnected fact. Hence, understanding conditionals by means of inferentialism provides a good basis for abductive reasoning.

4 Definition of abduction

Based on the foregoing considerations, abduction is defined in this article as follows: For a given fact, an abductive inference infers a fact that implies it. The implication is represented by an inferential conditional, where the implying fact is the antecedent and the given fact is the consequent. There are several types of abduction: Selective abduction allows one to infer an antecedent for a given fact by using a known conditional. Creative abduction allows one to infer an antecedent for a given fact by creating a new conditional. Creative abduction can be further divided into two types, depending on which kind of proposition is introduced as an antecedent: Conditional-creative abduction is based on a proposition that is already defined in the theory. Propositional-conditional-creative abduction introduces a new, so far undefined proposition.

The differentiation between the three types of abduction is important from a conceptual point of view because they allow one to add different types of new knowledge to an existing theory: Selective abduction relies on a known conditional and lets one infer only the truth of the antecedent, i.e. of a fact. Creative abduction, on the other hand, lets one infer not only the truth of an implying fact but also of an

inferential connection between the given fact and the implying one. A propositional-conditional-creative abduction moreover allows one to introduce a new proposition into a theory as an antecedent. A new proposition can be formed either by a new combination of existing propositions or by the introduction of a new term that is hitherto undefined. In both cases, the new proposition expresses a new concept and is therefore the most powerful kind of inference.[14]

Similarly, the differentiation between the three types is important for performing abductive inferences: Selective abduction uses a known conditional; thus, its implementation requires only a selection process to determine which conditional of the background knowledge to use for the inference. Conditional-creative abduction introduces a new conditional with a defined proposition as its antecedent; thus, a process is required to select a proposition of the theory and to create the conditional. Propositional-conditional-creative abduction introduces a new conditional with a new proposition; therefore, a process is required to create both a proposition and a conditional.

In summary, each type represents a different kind of inference, where both the conditional and the proposition are determined by either selective or creative processes. Nevertheless, the types do not instruct how the selective and creative processes are to be carried out: Selective abduction gives no guidance as to which available conditional should be chosen; and creative abduction does not specify which proposition to consider for the to be created conditional. Each type is completely neutral in terms of its implementation.

Hence, different procedures can be used to select or create the proposition and the conditional. The procedures provide guidance on how to perform a specific abductive inference and are called patterns. A pattern consists of a set of rules for both generating and justifying an abductive conclusion and it covers the whole inference process. Justificatory rules are considered because they influence the generation process: they are intended to ensure a promising result, i.e. that the truth of the conclusion is as likely as possible.

Types and patterns are very distinct in their characteristics. There are three different types of abduction, each of which is an inferential process with selective and creative components. Moreover, types are theory-independent, i.e. they do not presuppose any particular theory. In contrast, patterns are theory-dependent because their generative and justificatory rules are based on different assumptions, e.g. on the principle of causality. Furthermore, different methods can be used to perform the selective and creative processes, e.g. simple heuristics as well as complex statistical procedures. Consequently, there is an infinite number of patterns that rely on different theories and use different methods. As a result, the various patterns differ in their applicability, efficiency and persuasiveness.

The differentiation between types and patterns has several advantages. It distinguishes between the conceptual power of types of inferences on the one hand and the generative and justificatory power of patterns on the other. Furthermore, it allows a clear distinction between selective and creative components of the inference process as well as a comparison of different patterns, e.g. of their underlying assumptions and their methods.

These considerations lead to the following formal structure of abductive inferences:

---

[14] In fact, not only abduction but also induction allows inferring a new proposition in the conclusion. However, induction allows only to introduce as an antecedent a new proposition which is a generalised version of the propositions provided in the premises. In contrast, propositional-conditional-creative abduction allows creating a new proposition that is based not only on propositions from the premises but also from the background knowledge or on so far undefined terms.

Premise 1: a fact F

Premise 2: a pattern P; i.e. a set of rules generating and justifying the conditional A →[15] F, with A being a fact that implies F

Premise 3: a background knowledge BK that is used by the pattern P

===

Conclusion: *(A → F)* ∧ A

The conclusion contains the conditional in italic, as it is only concluded in creative abduction. In selective abduction, the conditional is already known and part of the background knowledge, i.e. the premises. In creative abduction, the truth of the conditional has to be concluded since the justification of the truth of the antecedent relies on it. The conditional can be regarded either as an intermediate step to the conclusion of the antecedent or as a conclusion on its own. What is considered the main insight depends on the purpose of the inference; for instance, whether a cause or an inferential connection should be inferred.

The conclusion contains the conditional 'A → F'. In contrast, Douven (2015, p. 96) argues that in a so-called abductive conditional, the consequent best explains the antecedent, i.e. the abductive conditional has the form 'F → A'. The two conditionals are related in that the former is part of the conclusion, while the latter represents the abductive inference as a whole. Accordingly, they express two different meanings and rely on two different inferential connections.

Even though the main purpose of abduction is to identify the fact that implies the given fact, both conditionals can provide additional insights. In case one is concerned with what one can infer from the truth of the given fact, the conditional representing the abductive inference as a whole is relevant. In case one is mainly concerned with what implies the given fact, the conditional stated in the conclusion of the abductive inference is of interest.

The inferential connection of the conditional 'F → A' is based on the abductive inference process. Therefore, the more credible the abductive inference, the higher the probability of the conditional being true. For example, the abductively inferred conditional "If Paula travels from Germany to Japan, then she travels by plane." is very likely because the abductive inference can be based on the strong argument that long distances are most often travelled by plane. On the other hand, the conditional "If the car does not start, then the battery is dead." is less credible because there are many likely alternatives, such as an empty tank or a blown fuse.

5 Types and patterns of abduction

5.1 Selective abduction

Selective abduction is the best researched type of abduction (cf. Peirce 1958, CP 2.636; Psillos 2009, pp. 117–131). This is because it is rather simple: The inference starts with the given fact 'F'. Then, a pattern selects from the background knowledge a conditional in which the fact 'F' is the consequent. The inference has the formal form:

F

A → F

=====

A

---

[15] In this article, the sign '→' is used to express a conditional based on inferentialism.

The credibility of the inference depends on many different aspects, e.g. the underlying formal system as well as the number of conditionals available that have 'F' as a consequent. In case the background knowledge specified in a formal system contains every true statement and there is only one conditional that has the fact 'F' as a consequent, then the inference is certain. In case there are several suitable conditionals available, then the inference is uncertain and the pattern must provide a method to select the most likely one.

Additional uncertainty arises if the formal system is incomplete or non-monotonic: then the fact 'F' can also be realised by a fact for which the corresponding conditional is not listed in the background knowledge. This aspect illustrates the limitation of selective abduction: it can only infer antecedents that are already known to imply the given fact, but not ones for which this is not known. To infer such, creative abduction is required.

Fully formalised patterns of selective abduction are provided in computer science, e.g. by Aliseda (2006), Flach and Kakas (2000), but also discussed in psychology, e.g. by Thomas, Dougherty, Sprenger and Harbison (2008).

An illustration of selective abduction can be found in Semmelweis research on puerperal fever (cf. sect 2.3): Semmelweis (1861, p. 38f; Shorter 1984, p. 47) examines several facts that are considered to have a possible influence on puerperal fever, e.g. hyperinosis, hydremia and plethora. However, since these known facts cannot explain why puerperal fever cases only occur in one clinic but not in the other, he dismisses them and suspects another, as yet unknown reason (Semmelweis 1861, pp. 51f; Shorter 1984, p. 51).

5.2 Creative abduction

Creative abduction infers that the given fact 'F' is implied by a hitherto unrelated fact 'A'. The implication is due to an unknown inferential connection between the two facts. Creative abduction hence allows to infer not only the truth of the antecedent 'A', but also the truth of a conditional that expresses the inferential connection.

Schurz (2008, p. 218) argues that all creative abductions in science explain several mutually intercorrelated phenomena by inferring a new unobservable concept that is their common cause. Consequently, neither single nor unobservable facts can be explained nor observable causes inferred. However, these are not intrinsic limitations of creative abductive inferences, but result from the pattern used: Schurz's pattern uses statistical factor analysis and judges results by the virtue of unification (pp. 219–232). As a consequence, only causes that can explain several phenomena at once are considered worthwhile. However, also non-unifying creative abductions explaining only one fact can be scientifically insightful; for instance, in cases such as the appearance of a single fossil of an ancient fish at high altitude in the Andes or the momentary dimming of a star.

Schurz's creative abduction is also limited in that it only allows the introduction of new concepts, but not the use of already defined concepts as antecedent (Schurz 2008, pp. 216, 218; 2016, p. 495). Nevertheless, creative abductions inferring already defined concepts can be insightful as well.

In contrast, the concept of creative abduction presented here overcomes these limitations by allowing for different patterns. Consequently, it can encompass both observable and unobservable facts as well as the inference of non-unifying and defined facts.

5.3 Conditional-creative abduction

An abduction, in which a defined concept is concluded to imply the given fact 'F', is a conditional-creative abduction. It is selective concerning the implying fact and creative concerning the inferred conditional that connects the implying fact and the given fact. It has the formal form:

$$F$$
$$[A]$$
$$==========$$
$$(A \rightarrow F) \land A$$

'A' is in square brackets in the premises to indicate that it must be a defined proposition, but its truth value may be unknown.

The purpose of patterns of conditional-creative abduction is to determine which proposition available in the theory is most likely to be the antecedent of the given fact. A wide variety of methods and assumptions can be used for this.

For example, patterns based on causal Bayes nets allow one to determine a structural link based on causal power by considering interventions and known mechanisms (Chater and Oaksford 2020, pp. 121–125). Another pattern provides the search for spatio-temporal continuity: people have a strong tendency to assume a causal relationship between two events if they are no more than two seconds apart (Griffiths and Tenenbaum 2009, pp. 662, 696). Other patterns are based on the search for similarities (Magid, Sheskin and Schulz 2015, p. 101) or by comparing the characteristics of the given fact and facts that can serve as possible antecedents (pp. 103–109).

In general, theory-specific knowledge plays an important role in the selection of an appropriate proposition as antecedent: e.g. laws that explicate which types of propositions can imply which other types of propositions and thus the given fact. Hence, patterns used for conditional-creative abduction can rely on a large amount of background knowledge, which complicates their formulation.

An illustration of conditional-creative abduction is provided by Semmelweis: Having concluded that no known cause could account for the different rates of puerperal fever, Semmelweis considered facts that were known but not associated with puerperal fever so far. For instance, Semmelweis (1861, pp. 36, 51f; Shorter 1984, pp. 51f) considered the delivery position and the routes women had to take to their puerperium after giving birth.

He obtained these facts by applying various generative patterns; e.g. looking at reasons for unwellness and illness in general, or comparing the two clinics and finding differences. Nevertheless, none of the possible reasons could be substantiated. Either they could not be justified during the inference process because they did not fit with the background knowledge, or they could not be confirmed in subsequent experiments.

5.4 Propositional-conditional-creative abduction

Propositional-conditional-creative abduction assumes that the given fact 'F' is not implied by a fact defined in the theory, but by a new, hitherto undefined one. It thus infers both the truth of a new fact and the truth of an inferential connection between the new fact and the given fact. The inferential connection has to be inferred because it provides support for the truth of the implying fact. Propositional-conditional-creative abduction has the formal form:

$$F$$
$$==========$$
$$(A \rightarrow F) \land A$$

Schurz (2016, pp. 498–503) points out that there are cases where the given fact 'F' is not simple but complex, i.e. consists of a plurality of facts. For example, the given fact can state that sugar, salt, sodium carbonate and copper sulphate are all soluble in water, insoluble in oil, have a higher melting point and conduct electricity.

In addition, in some cases the multitude of facts subsumed in the given fact cannot be implied by a simple fact, but only by a complex one. For instance, using statistical factor analysis, the cultural characteristics of nations can be explained by the interplay of two main factors: the orientation between traditional-religious and secular-rational values on the one hand, and the orientation between survival and self-expression values on the other (pp. 506–508).[16] Neither factor alone would suffice to satisfactorily explain a country's cultural characteristics. In some cases, only the inference of an antecedent containing several facts leads to a satisfactory result.

Propositional-conditional-creative abduction consists of two steps: Once the number and relation of the facts of the antecedent have been determined, one must define them, i.e. express them by introducing new propositions. A new proposition can be defined either by introducing a new term or by combining already defined propositions of the underlying theory in a new way. Besides, when defining the proposition more precisely, a newly introduced proposition may turn out to be an already defined one.

It is possible to use separate subpatterns for determining the number of facts and for defining them. This is especially so since the definition of a new proposition often relies on other propositions from background knowledge and is therefore very theory-specific; whereas the inference of the number of possible facts in the antecedent is often based on more fundamental assumptions, e.g. statistical considerations.

Furthermore, both steps can be carried out independently of each other. For example, as Schurz (2016, p. 498) points out, the existence of the new proposition 'hydrophilic nature' was inferred long before the theory of atoms and molecules that allows it to be described.

Semmelweis' study of puerperal fever includes several illustrations of propositional-conditional-creative abductions. For instance, a commission suspected that the increased incidence of puerperal fever in one of the clinics was due to overly crude examinations by male students, especially foreigners (Semmelweis 1861, pp. 48f; Shorter 1984, p. 50). However, this hypothesis could not be verified in subsequent experiments.

Another propositional-conditional-creative abduction finally led Semmelweis to the solution of the increased rates of puerperal fever. As previously mentioned (cf. sect. 2.3), one of his colleagues, after being wounded with an autopsy knife, showed the same symptoms as those of puerperal fever and eventually died. Semmelweis ascribed his death to contamination with cadaverous particles in the course of the injury.

Based on this knowledge, Semmelweis inferred by analogy that the patients in the maternity ward also died from infection with cadaverous particles. However, in contrast to his colleague's case, the infection was not transmitted by an autopsy knife, but by medical staff who performed autopsies before examining the patients: Cadaverous particles remained on their hands, which were then absorbed by the patients' genitals during the examination. In conclusion, Semmelweis inferred a new, hitherto undefined fact; the transmission of cadaveric particles via hands. This new fact is considered to have an inferential connection to the given fact, i.e. patients contracting puerperal fever, and is therefore its antecedent.

---

[16] Glymour (2019) argues that the factor analysis used by Schurz is not suitable and suggests using a different pattern, which is based on other statistical methods and allows a more accurate abductive inference for this case.

Later, Semmelweis (1861, pp. 58–60; Shorter 1984, p. 54) performed two more inferences that illustrate propositional-conditional-creative abductions: First, a patient with uterine cancer was admitted and subsequently all patients in the room died. This led Semmelweis to infer that infectious matter can also be transmitted by ichor. Second, a patient was admitted with a healthy genital area but a discharging carious knee; again, most of the patients in the room subsequently died. From this, Semmelweis concluded that infectious matter can also be transmitted via air.

5.5 Analogical patterns of creative abduction

Semmelweis' research shows that the use of analogies in abduction can lead to promising hypotheses. This chapter therefore explores in more detail how analogies can contribute to the generation and justification of hypotheses in patterns. Analogies are often given in the following form (Bartha 2019, sect. 2.2; notation adapted):

P_1 is similar to P_k in certain (known) respects
P_k has some further feature Q_k
===
P_1 also has the feature Q_k, or some feature Q_1 similar to Q_k

This leads to the following formal representation:

P_1	given fact
$P_k \rightarrow Q_k$	with P_1 and P_k being similar in certain known respects
=======	
$P_1 \rightarrow Q_k$	possible conclusion 1: transfer of the same feature
$P_1 \rightarrow Q_1$	possible conclusion 2: transfer of a similar feature

In summary, an analogical inference transfers a characteristic, an inferential connection with a consequent, from one proposition to another, similar proposition.[17] Depending on the nature of the analogy, the consequent can be altered and adapted to the similar proposition. Analogical conclusions are amplifying and uncertain because the inferential connection does not necessarily apply to the similar proposition as well.

The legitimacy of analogical inferences rests on the assumption that similar conditions lead to similar results. As Mill (1974, p. 556; notation adapted) argues: "If [P_1] resembled [P_k] in all its ultimate properties, its possessing the attribute [Q_k] would be a certainty, not a probability: and every resemblance which can be shown to exist between them, places it by so much the nearer to that point. If the resemblance be in an ultimate property, there will be resemblance in all the derivative properties dependent on that ultimate property, and of these [Q_k] may be one."

Likewise, one can assume that similar results are based on similar conditions. Mill (ibid.) continues: "If the resemblance be in a derivative property, there is reason to expect resemblance in the ultimate property on which it depends, and in the other derivative properties dependent on the same ultimate property." This assumption can be used to perform an analogical abduction: Given a certain fact, one searches for a proposition which is similar and of which one knows the antecedent. One assumes that the antecedent is also that of the given fact – either in the form of the original proposition or in the form of a similar proposition adapted to the given fact. Formally, this can be expressed as follows:

---

[17] For an assessment of similarity-based arguments in the context of inferentialism, see Douven, Elqayam, Gärdenfors and Mirabile (in submission).

Q_1	given fact
$P_k \rightarrow Q_k$	with Q_1 and Q_k being similar in certain known respects
=======	
$P_k \rightarrow Q_1$	possible conclusion 1: antecedent consists of the original proposition
$P_1 \rightarrow Q_1$	possible conclusion 2: antecedent consists of a modified proposition

In case the inferred antecedent contains the original or a defined similar proposition, it is a conditional-creative abduction. In case a similar, previously undefined proposition is inferred, it is a propositional-conditional-creative abduction.

5.6 Empirical adequacy

Overall, the theory of abduction presented here provides a high degree of empirical adequacy with Semmelweis' research on puerperal fever. This does not necessarily mean that Semmelweis actually performed the processes of abduction described here – this is only an interpretation based on his writings, and there are many other interpretations of his research as well. In either case, Semmelweis' research provides an illustration of how the theory of abduction presented here could be successfully applied. It can represent all the inferences Semmelweis performed and their methods, and it can explain in detail how the solution was finally reached through the use of analogy.

Furthermore, the abductive theory can explain the order in which Semmelweis executed the research process. First, he started from facts that were considered to be related to puerperal fever or diseases in general, for example hyperinosis. When this was unsuccessful, he examined known facts such as the delivery position and tried to establish an inferential connection to puerperal fever. When this also remained unsuccessful, he tried to identify new, hitherto undefined facts that imply puerperal fever.

This order results from the fact that the selective and creative processes of abduction require different amounts of cognitive workload: Selective abduction requires only the selection of a known conditional in which the given proposition is the consequent, it is therefore the simplest type. Conditional-creative abduction uses a defined proposition, but there are usually many available and an inferential connection must be created as well. Finally, propositional-conditional-creative abduction requires not only an inferential connection but also a new proposition to be created, which requires again additional cognitive effort.

With the abductive theory presented in this article, one can also explain why some inferences were performed together, but mostly each possible implying fact was inferred for itself. The first case, the inference of several possible causes at once, happened mostly at the beginning; this because selective abduction allows several available conditionals to be selected, compared with each other and evaluated together. In creative abduction, most possible causes were inferred individually since each required its own process of generation and justification.

The proposed abductive theory shows how the virtues of IBE, such as simplicity and coherence, can be used as guidance for the generative and justificatory processes in patterns. For example, simpler solutions are preferred because they are easier to generate; and more coherent solutions are preferred because a better fit with background knowledge reduces the likelihood of contradictions.

The contrastive inference approach proposed by Lipton (cf. sect. 2.3) can be carried out in the form of a pattern using e.g. statistical factor analysis. However, the immanent problem of multiple differences becomes apparent here: The method is only successful in case the relevant data are taken into account. In Semmelweis' case, it would have been necessary to statistically compare the incidence of the autopsies with the incidence of the puerperal fever cases. But without knowing the connection, there was no reason to pay special attention to this small detail out of the myriads available. Therefore, this

pattern is only successful if a large proportion of the data can be taken into account; else, other patterns are preferable.

In the definition of abduction (cf. sect. 4), it was shown that abductive inferences not only allow one to conclude the fact 'A', but also, in the case of creative abduction, the conditional 'A → F'. In addition, the inference as a whole can be represented by the conditional 'F → A'. The different purposes of the three conclusions become apparent in Semmelweis' case: His main interest was to determine 'A', the factor causing the high rates of puerperal fever in one of the two Vienna clinics.

Besides that, the conditional 'A → F' was also of interest for him in several respects: First, he wanted to communicate it to other physicians so that they could avoid cases of puerperal fever in their own hospitals. Second, he used the conditional as a basis for further analogical abductive inferences to infer that cadaveric matter can be transmitted through ichor and air as well.

Finally, the conditional 'F → A' may be of interest in that if a case of puerperal fever appears, one can investigate whether it was caused by cadaveric matter. For instance, when Semmelweis (1861, pp. 81–85) heard of high rates of puerperal fever cases in the hospital at Pest, he suspected transmission of cadaveric matter. His subsequent investigation revealed that the examination of women in labour was carried out by physicians who had performed operations beforehand and contaminated themselves thereby.

6 Formalisation of abductive inferences

Abductive theories vary widely in their understanding of the extent to which abductive inferences can be formalised, especially concerning the context of discovery (cf. sect. 1). Formalisation is understood here in the sense that it is possible to explicitly represent all information as well as all steps in which the information is processed. This means, a formalisable inference can be completely represented in a logical system and its implementation can be expressed in form of an algorithm that is Turing-computable. A formalisable theory of abduction has the advantage that it can be implemented in computer science and used for artificial intelligence.

The theory presented here defines abduction as an inference that allows one to infer for a given fact a fact that implies it. The implication is represented by a conditional which is considered true iff there is a connection from the antecedent to the consequent. Since inferentialism is pluralistic, the connection can be of different kinds, it can be deductive, inductive or abductive. There is no unique criterion under which conditions a conditional connection is regarded as valid and thus the conditional as true. Nevertheless, it is possible to provide rules to judge the validity of a conditional connection. For example, an inductive relation can be judged valid in case there are at least ten confirming and no falsifying instances. Such rules can be formally represented either as part of the axioms of a theory or as part of the context of justification of an abductive pattern. In summary, conditionals, which form the basis of abductive inferences, as well as their truth evaluation, can be formally represented.

Structurally, an abductive inference consists of the given fact, a pattern, and background knowledge. Since both the given fact and the background knowledge can be formally represented in the form of propositions, it follows that an abductive inference is formalisable iff its pattern is formalisable. A pattern is formalisable iff every rule of the pattern, whether it concerns the generation or the justification, is formalisable and the pattern covers the complete inference process.

Fully formalisable patterns exist for both selective abduction and creative abduction. For example, in selective abduction, the background knowledge is typically searched for all conditionals that contain the given fact 'F' as a consequent. Subsequently, the available conditionals are ranked according to the joint

probability of the antecedent and the strength of the conditional.[18] Finally, the antecedent of the highest ranked conditional is considered true (cf. Aliseda 2006; Flach and Kakas 2000).

Patterns for creative abduction are more complex because they have to generate a new conditional and, depending on the type, a new proposition. Examples of patterns that allow specific kinds of creative abduction are Schurz's (2008, pp. 223–231) common cause abduction and BACON.4, which allows to search for lawful correlations in numerical data (Langley, Simon, Bradshaw and Zytkow 1987, ch. 4; cf. Jantzen 2016, sect. 3.2f).

In conclusion, some abductive inferences can be formalised. However, this does not mean that all abductive inferences are formalisable: There are patterns, e.g. Peirce's intuitive creative act (cf. sect. 2.1), which are not formalisable and which therefore preclude the formalisation of an abductive inference.

There seem to be several reasons why it is often claimed that abductions cannot be formalised: First, the underlying processes are often complex; therefore, it is difficult to explicate all rules of a pattern formally. Second, there is an infinite number of patterns because they are based on theory-specific knowledge, which makes them difficult to differentiate and capture. Third, the likelihood that the abductive conclusion is true is pattern-dependent, and many patterns yield a likelihood that is positive but not high enough to be considered feasible. Fourth, at least when real-world data is to be used as basis for abductive inferences, it is very difficult to formalise it, e.g. to determine the specific propositions – yet this is crucial for successful inferences.

7 Conclusion

The goal of the article is to lay the foundation for a theory of abduction which is complete, i.e. covering both the context of generation and the context of justification, and formalisable, which allows its application in computer science and artificial intelligence.

The theory proposed states that an abductive inference infers for a given fact a fact that implies it. By relying on conditionals, the theory stands in contrast to many other theories that consider explanations as one or even the cornerstone of abduction. Nevertheless, even though the theory does not consider abduction as intrinsically explanatory, it does not neglect the close relationship of abduction and explanation. Often abductive inferences can and do serve as explanations – but they do not have to. Relying on conditionals rather than explanations as the basis for abduction offers several advantages.

First, a theory of abduction based on conditionals allows the inference not only of explanations but of all kinds of preceding facts, which includes for example instrumental models and axioms.

Second, when using conditionals, one can rely on two special properties of conditionals: they are asymmetrical, and they allow one to infer the truth of the consequent from the truth of the antecedent. This inferential connection can be used to justify the conclusion in the opposite direction, i.e. to infer the truth of an antecedent from the truth of the consequent. This inference is uncertain since the consequent may be implied by one of several known antecedents or even by an unknown one. Nevertheless, the inferential connections from the possible antecedents to the consequent can be used as a basis to generate and justify which antecedent actually implies the consequent. This justification is provided by patterns which can be based for example on probabilistic or analogical methods.

---

[18] This approach is also empirically supported by Sebben and Ullrich (2021), who show that people tend to evaluate conditionals in this way.

Third, a theory of abduction based on conditionals does not require a theory of explanation. Since there is currently no generally accepted one, such a requirement would prevent the practical implementation and use of the abductive theory in computer science and artificial intelligence. Nevertheless, the theory presented here presupposes a theory of conditionals, which are also controversially discussed. This poses a challenge and requires further work; however, it is hoped that the open questions on conditionals – at least as far as abduction is concerned – can be resolved more easily than those on explanations.

The abductive theory presented in this article does not agree with IBE in many aspects, e.g. it is doubted that IBE's hypothesis generation is applicable and that explanatory virtues are sufficient to lead to the correct hypothesis. Nevertheless, IBE provides valuable insights.

For example, empirical studies show that people actually assign extra value to the best explanation and thereby can achieve better results (Douven 2020; Douven and Mirabile 2018). Nonetheless, further research is required. For example, the studies only address the justification but not the generation of hypotheses, and the application is intrinsically context-sensitive (Douven 2020, pp. 1, 11). Moreover, it is not clear by which explanatory virtues the quality of an explanation is to be judged – or whether non-explanatory considerations can play a role as well. Furthermore, it needs to be investigated whether a preference for the best hypothesis only occurs in abductive reasoning or also e.g. in inductive reasoning. The first case would suggest that the preference is an intrinsic part of abduction, while the second case would suggest that it is a reasoning strategy based on economic reasons and independent of abduction.

Another valuable aspect of IBE is its (explanatory) virtues, which can provide guidance as to which hypotheses are worth pursuing. Likewise, the theory discussed here incorporates components of many other theories; for example, Peirce's foundational understanding of abduction as well as methods of Schurz and others as patterns.

Thus, although the approach presented here proposes a new understanding of abduction and aims to overcome several limitations of current approaches, it also draws on them in many ways. It is hoped that the proposed theory will contribute to the ongoing discussion by providing an approach that is formalisable and computable. Additionally, it should allow all kinds of abductive inferences to be covered while being sufficiently precise by enabling the use of specific patterns.

Many open questions remain that require further research. For example, more case studies need to be performed, and patterns as well as their formalisation and application need to be explored in more detail. Similarly, the combination of the presented theory of abduction with probability theories such as Bayesianism needs to be examined. Furthermore, the properties of complex antecedents and consequents, i.e. which consist of multiple facts, need to be investigated, as does the use of nested and counterfactual conditionals.

Finally, especially for the application in computer science and artificial intelligence, a logic of abduction must be developed. The following considerations already show some possible characteristics of an abductive logic: Including probabilities, although not inherently required, allows the use of probability-based patterns as well as the determination of the likelihood of the conclusion. Non-monotonicity allows new statements to be added, e.g. experimental data that falsify previous abductive conclusions, which can lead to improved new conclusions. Other aspects, such as the derivation of additional assumptions and whether both a fact and its negation can imply a fact, are determined by the inferential conditional theory; these inferences are valid only if there is an inferential connection.

Acknowledgements

Declarations

References

Aliseda, Atocha (2006): *Abductive Reasoning. Logical Investigations into Discovery and Explanation*. Dordrecht: Springer (Synthese Library, 330).

Anderson, Douglas R. (1987): Creativity and the philosophy of C.S. Peirce. Dordrecht, Lancaster: Nijhoff (Martinus Nijhoff philosophy library, v.27).

Baker, Alan (2020): Non-Deductive Methods in Mathematics. In Edward N. Zalta (Ed.): *The Stanford Encyclopedia of Philosophy*. Summer 2020: Metaphysics Research Lab, Stanford University.

Baratgin, Jean; Over, David E.; Politzer, Guy (2013): Uncertainty and the de Finetti tables. In *Thinking & Reasoning* 19 (3-4), pp. 308–328. DOI: 10.1080/13546783.2013.809018.

Bartha, Paul (2019): Analogy and Analogical Reasoning. In Edward N. Zalta (Ed.): *The Stanford Encyclopedia of Philosophy*. Spring 2019: Metaphysics Research Lab, Stanford University.

Beirlaen, Mathieu; Aliseda, Atocha (2014): A conditional logic for abduction. In *Synthese* 191 (15), pp. 3733–3758. DOI: 10.1007/s11229-014-0496-0.

Bird, Alexander (2010): Eliminative abduction: examples from medicine. In *Studies in History and Philosophy of Science Part A* 41 (4), pp. 345–352. DOI: 10.1016/j.shpsa.2010.10.009.

Cabrera, Frank (2017): Can there be a Bayesian explanationism? On the prospects of a productive partnership. In *Synthese* 194 (4), pp. 1245–1272. DOI: 10.1007/s11229-015-0990-z.

Cabrera, Frank (2020): Does IBE Require a 'Model' of Explanation? In *The British Journal for the Philosophy of Science* 71 (2), pp. 727–750. DOI: 10.1093/bjps/axy010.

Campos, Daniel G. (2011): On the distinction between Peirce's abduction and Lipton's Inference to the best explanation. In *Synthese* 180 (3), pp. 419–442. DOI: 10.1007/s11229-009-9709-3.

Chater, Nick; Oaksford, Mike (2020): Integrating Causal Bayes Nets and inferentialism in conditional inference. In Elqayam, Shira; Douven, Igor; St. Evans, Jonathan B. T.; Cruz, Nicole; Over, D. E. (2020): *Logic and uncertainty in the human mind. A tribute to David E. Over*. London: Routledge.

Cruz, Nicole; Oberauer, Klaus (2014): Comparing the meanings of "if" and "all". In *Memory & cognition* 42 (8), pp. 1345–1356. DOI: 10.3758/s13421-014-0442-x.

Dawes, Gregory W. (2013): Belief Is Not The Issue: A Defence Of Inference To The Best Explanation. In *Ratio* 26 (1), pp. 62–78. DOI: 10.1111/j.1467-9329.2012.00537.x.

Douven, Igor (2015): *The epistemology of indicative conditionals. Formal and empirical approaches*. Cambridge, England: Cambridge University Press, 11/1/2015.

Douven, Igor (2017a): Abduction. In Edward N. Zalta (Ed.): *The Stanford Encyclopedia of Philosophy*. Summer 2017: Metaphysics Research Lab, Stanford University.

Douven, Igor (2017b): How to account for the oddness of missing-link conditionals. In *Synthese* 194 (5), pp. 1541–1554. DOI: 10.1007/s11229-015-0756-7.

Douven, Igor (2020): The ecological rationality of explanatory reasoning. In *Studies in History and Philosophy of Science Part A* 79, pp. 1–14. DOI: 10.1016/j.shpsa.2019.06.004.

Douven, Igor; Elqayam, Shira; Gärdenfors, Peter; Mirabile, Patricia (2022): Conceptual Spaces and the Strength of Similarity-based Arguments. In *Cognition* 218, p. 104951. DOI: 10.1016/j.cognition.2021.104951.

Douven, Igor; Elqayam, Shira; Mirabile, Patricia (2022): Inference strength predicts the probability of conditionals better than conditional probability does. In *Journal of Memory and Language* 123, p. 104302. DOI: 10.1016/j.jml.2021.104302.

Douven, Igor; Elqayam, Shira; Singmann, Henrik; van Wijnbergen-Huitink, Janneke (2018): Conditionals and inferential connections: A hypothetical inferential theory. In *Cognitive Psychology* 101, pp. 50–81. DOI: 10.1016/j.cogpsych.2017.09.002.

Douven, Igor; Mirabile, Patricia (2018): Best, second-best, and good-enough explanations: How they matter to reasoning. In *Journal of experimental psychology. Learning, memory, and cognition* 44 (11), pp. 1792–1813. DOI: 10.1037/xlm0000545.

Easwaran, Kenny (2008): The Role of Axioms in Mathematics. In *Erkenntnis* 68 (3), pp. 381–391. DOI: 10.1007/s10670-008-9106-1.

Égré, Paul; Rott, Hans (2021): The Logic of Conditionals. In Edward N. Zalta (Ed.): *The Stanford Encyclopedia of Philosophy*. Fall 2021: Metaphysics Research Lab, Stanford University.

Evans, Jonathan B. T. (2020): The suppositional conditional is not (just) the probability conditional. In Elqayam, Shira; Douven, Igor; St. Evans, Jonathan B. T.; Cruz, Nicole; Over, D. E. (2020): *Logic and uncertainty in the human mind. A tribute to David E. Over*. London: Routledge.

Evans, Jonathan B. T.; Over, D. E. (2004): *IF*. Oxford: Oxford University Press (Oxford cognitive science series).

Fann, Kuang Tih (1970): *Peirce's theory of abduction*. The Hague: Nijhoff.

Flach, Peter A.; Kakas, Antonis C. (2000): *Abduction and Induction*. Dordrecht: Springer Netherlands (18).

Fugard, Andrew J. B.; Pfeifer, Niki; Mayerhofer, Bastian; Kleiter, Gernot D. (2011): How people interpret conditionals: shifts toward the conditional event. In *Journal of experimental psychology. Learning, memory, and cognition* 37 (3), pp. 635–648. DOI: 10.1037/a0022329.

Gabbay, Dov M.; Woods, John (2005): *The reach of abduction insight and trial.* 1. ed. Amsterdam: Elsevier (A practical logic of cognitive systems; King's College London; Vol. 2).

Gentner, Dedre (1983): Structure-mapping: A theoretical framework for analogy. In *Cognitive Science* 7 (2), pp. 155–170. DOI: 10.1016/S0364-0213(83)80009-3.

Glymour, Clark (2019): Creative Abduction, Factor Analysis, and the Causes of Liberal Democracy. In *Kriterion - Journal of Philosophy* 33 (1), pp. 1–22.

Godden, David; Zenker, Frank (2015): Denying Antecedents and Affirming Consequents: The State of the Art. In *IL* 35 (1), p. 88. DOI: 10.22329/il.v35i1.4173.

Griffiths, Thomas L.; Tenenbaum, Joshua B. (2009): Theory-based causal induction. In *Psychological review* 116 (4), pp. 661–716. DOI: 10.1037/a0017201.

Harman, Gilbert H. (1965): The Inference to the Best Explanation. In *The Philosophical Review* 74 (1), p. 88. DOI: 10.2307/2183532.

Hempel, Carl G. (1966): *Philosophy of natural science*. [reprint]. Upper Saddle River, NJ: Prentice-Hall.

Heron, John (2020): Set-theoretic justification and the theoretical virtues. In *Synthese*, pp. 1–23. DOI: 10.1007/s11229-020-02784-z.

Hoyningen-Huene, Paul (1987): Context of discovery and context of justification. In *Studies in History and Philosophy of Science Part A* 18 (4), pp. 501–515. DOI: 10.1016/0039-3681(87)90005-7.

Jantzen, Benjamin C. (2016): Discovery without a 'logic' would be a miracle. In *Synthese* 193 (10), pp. 3209–3238. DOI: 10.1007/s11229-015-0926-7.

Jones, Nicholaos (2018): Inference to the More Robust Explanation. In *The British Journal for the Philosophy of Science* 69 (1), pp. 75–102. DOI: 10.1093/bjps/axw009.

Klärner, Holger (2003): *Der Schluß auf die beste Erklärung*. Berlin: De Gruyter (Ideen & Argumente).

Krzyżanowska, Karolina; Collins, Peter J.; Hahn, Ulrike (2021): True clauses and false connections. In *Journal of Memory and Language* 121, p. 104252. DOI: 10.1016/j.jml.2021.104252.

Krzyżanowska, Karolina; Douven, Igor (2018): Missing-link conditionals: pragmatically infelicitous or semantically defective? In *Intercultural Pragmatics* 15 (2), pp. 191–211. DOI: 10.1515/ip-2018-0004.

Krzyżanowska, Karolina; Wenmackers, Sylvia; Douven, Igor (2014): Rethinking Gibbard's Riverboat Argument. In *Stud Logica* 102 (4), pp. 771–792. DOI: 10.1007/s11225-013-9507-2.

Lange, Marc (2020): Putting explanation back into "inference to the best explanation". In *Noûs*. DOI: 10.1111/nous.12349.

Langley, Pat; Simon, Herbert A.; Bradshaw, Gary L.; Zytkow, Jan M. (1987): *Scientific discovery. Computational explorations of the creative processes*. Cambridge, Mass., London: MIT Press.

Lipton, Peter (2004): *Inference to the best explanation*. 2. ed. London: Routledge (International library of philosophy).

Magid, Rachel W.; Sheskin, Mark; Schulz, Laura E. (2015): Imagination and the generation of new ideas. In *Cognitive Development* 34, pp. 99–110. DOI: 10.1016/j.cogdev.2014.12.008.

Magnani, Lorenzo (2009): *Abductive cognition. The epistemological and eco-cognitive dimensions of hypothetical reasoning*. Berlin: Springer (Cognitive systems monographs, 3).

Magnani, Lorenzo (2015): The eco-cognitive model of abduction. In *Journal of Applied Logic* 13 (3), pp. 285–315. DOI: 10.1016/j.jal.2015.04.003.

Mcauliffe, William (2015): How did Abduction Get Confused with Inference to the Best Explanation? In *Transactions of the Charles S. Peirce Society* 51 (3), p. 300. DOI: 10.2979/trancharpeirsoc.51.3.300.

McKaughan, Daniel J. (2008): From ugly duckling to Swan: C. S. Peirce, abduction, and the pursuit of scientific theories. In *Transactions of the Charles S. Peirce Society* 44 (3), pp. 446–468.

Mill, John S. (1974): *A system of logic*. Edited by John M. Robson. S.l.: Routledge (Collected works of John Stuart Mill, vols 7 & 8).

Minnameier, Gerhard (2004): Peirce-Suit of Truth – Why Inference to the Best Explanation and Abduction Ought Not to be Confused. In *Erkenntnis* 60 (1), pp. 75–105. DOI: 10.1023/B:ERKE.0000005162.52052.7f.

Mirabile, Patricia; Douven, Igor (2020): Abductive conditionals as a test case for inferentialism. In *Cognition* 200, p. 104232. DOI: 10.1016/j.cognition.2020.104232.

Nickerson, Raymond S.; Barch, Daniel H.; Butler, Susan F. (2019): Evaluating conditional arguments with uncertain premises. In *Thinking & Reasoning* 25 (1), pp. 48–71. DOI: 10.1080/13546783.2018.1465474.

Niiniluoto, Ilkka (2018): *Truth-Seeking by Abduction* [electronic resource]. Cham: Springer (Synthese Library, vol. 400).

Novick, Aaron; Scholl, Raphael (2020): Presume It Not: True Causes in the Search for the Basis of Heredity. In *The British Journal for the Philosophy of Science* 71 (1), pp. 59–86. DOI: 10.1093/bjps/axy001.

Over, David E.; Hadjichristidis, Constantinos; Evans, Jonathan B. T.; Handley, Simon J.; Sloman, Steven A. (2007): The probability of causal conditionals. In *Cognitive Psychology* 54 (1), pp. 62–97. DOI: 10.1016/j.cogpsych.2006.05.002.

Paavola, Sami (2004): Abduction as a Logic and Methodology of Discovery: the Importance of Strategies. In *Foundations of Science* 9 (3), pp. 267–283. DOI: 10.1023/B:FODA.0000042843.48932.25.

Paavola, Sami (2006a): *On the origin of ideas. An abductivist approach to discovery*. Helsinki: Department of Philosophy, University of Helsinki (Philosophical studies from the University of Helsinki, 15).

Paavola, Sami (2006b): Hansonian and Harmanian Abduction as Models of Discovery. In *International Studies in the Philosophy of Science* 20 (1), pp. 93–108. DOI: 10.1080/02698590600641065.

Park, Woosuk (2015): On classifying abduction. In *Journal of Applied Logic* 13 (3), pp. 215–238. DOI: 10.1016/j.jal.2015.04.001.

Peirce, Charles S. (1958): *Collected Papers of Charles Sanders Peirce*. 8 volumes. Edited by C. Hartshorne, P. (volumes 1-6) Weiss, volumes 7-8 (Burks, A.). Cambridge MA: Belknap Press.

Peirce, Charles S. (1998): *The essential Peirce. Selected philosophical writings*. Vol. 2, 1893-1913 / edited by the Peirce Edition Project. Bloomington, Ind., London: Indiana University Press.

Popper, Karl R. (1959): *The logic of scientific discovery*. New York: Basic Books, Inc.

Psillos, Stathis (2002): Simply the Best: A Case for Abduction. In Robert Kowalski, Antonis C. Kakas, Fariba Sadri (Eds.): *Computational logic. Logic programming and beyond: essays in honour of Robert A. Kowalski*. Part 2, vol. 2408. Berlin, London: Springer, pp. 605–625.

Psillos, Stathis (2009): An Explorer upon Untrodden Ground. In: Gabbay, Woods et al. (Ed.) 2009, *Handbook of the history*, vol. 10, pp. 117–151.

Ramsey, Frank (1929/1990): General propositions and causality. In Mellor, D. H. (Ed.), *Philosophical papers*, pp. 145–163. Cambridge: Cambridge University Press.

Reutlinger, Alexander (2018): Extending the Counterfactual Theory of Explanation. In: Reutlinger, Alexander; Saatsi, Juha (2018): *Explanation Beyond Causation. Philosophical Perspectives on Non-Causal Explanations*: Oxford University Press, pp. 74–95

Rosenthal, Sandra (2004): Peirce's Pragmatic account of Perception. In C. J. Misak (Ed.): *The Cambridge companion to Peirce*. Cambridge: Cambridge University Press, pp. 193–213.

Schickore, Jutta (2018): Scientific Discovery. In Edward N. Zalta (Ed.): *The Stanford Encyclopedia of Philo*sophy. Summer 2018: Metaphysics Research Lab, Stanford University.

Schlimm, Dirk (2008): Two Ways of Analogy: Extending the Study of Analogies to Mathematical Domains*. In *Philosophy of Science* 75 (2), pp. 178–200. DOI: 10.1086/590198.

Scholl, Raphael (2013): Causal inference, mechanisms, and the Semmelweis case. In *Studies in History and Philosophy of Science Part A* 44 (1), pp. 66–76. DOI: 10.1016/j.shpsa.2012.04.002.

Schurz, Gerhard (2008): Patterns of abduction. In *Synthese* 164 (2), pp. 201–234. DOI: 10.1007/s11229-007-9223-4.

Schurz, Gerhard (2016): Common cause abduction: The formation of theoretical concepts and models in science. In *Logic Jnl IGPL* 24 (4), pp. 494–509. DOI: 10.1093/jigpal/jzw029.

Sebben, Simone; Ullrich, Johannes (2021): Can conditionals explain explanations? A modus ponens model of B because A. In *Cognition* 215, p. 104812. DOI: 10.1016/j.cognition.2021.104812.

Semmelweis, Ignaz (1861): *Die Aetiologie, der Begriff und die Prophylaxis des Kindbettfiebers*. Pest a.o.: Hartleben. Available online at https://www.deutschestextarchiv.de/book/show/semmelweis_kindbettfieber_1861.

Shorter, Edward (1984): Ignaz Semmelweis: The etiology, concept, and prophylaxis of childbed fever. In *Medical history* 28.

Skovgaard-Olsen, Niels (2016): Motivating the Relevance Approach to Conditionals. In *Mind Lang* 31 (5), pp. 555–579. DOI: 10.1111/mila.12120.

Skovgaard-Olsen, Niels; Kellen, David; Hahn, Ulrike; Klauer, Karl Christoph (2019): Norm conflicts and conditionals. In *Psychological review* 126 (5), pp. 611–633. DOI: 10.1037/rev0000150.

Skovgaard-Olsen, Niels; Kellen, David; Krahl, Hannes; Klauer, Karl Christoph (2017): Relevance differently affects the truth, acceptability, and probability evaluations of "and", "but", "therefore", and "if–then". In *Thinking & Reasoning* 23 (4), pp. 449–482. DOI: 10.1080/13546783.2017.1374306.

Skovgaard-Olsen, Niels; Singmann, Henrik; Klauer, Karl Christoph (2016): The relevance effect and conditionals. In *Cognition* 150, pp. 26–36. DOI: 10.1016/j.cognition.2015.12.017.

Stalnaker, Robert C. (1981): A Theory of Conditionals. In William L. Harper, Robert Stalnaker, Glenn Pearce (Eds.): *IFS: Conditionals, Belief, Decision, Chance and Time*. Dordrecht: Springer Netherlands, pp. 41–55.

Thomas, Rick P.; Dougherty, Michael R.; Sprenger, Amber M.; Harbison, J. Isaiah (2008): Diagnostic hypothesis generation and human judgment. In *Psychological review* 115 (1), pp. 155–185. DOI: 10.1037/0033-295X.115.1.155.

Tschaepe (2014): Guessing and Abduction. In *Transactions of the Charles S. Peirce Society* 50 (1), p. 115. DOI: 10.2979/trancharpeirsoc.50.1.115.

Vidal, Mathieu; Baratgin, Jean (2017): A psychological study of unconnected conditionals. In *Journal of Cognitive Psychology* 29 (6), pp. 769–781. DOI: 10.1080/20445911.2017.1305388.

Woods, John (2011): Recent developments in abductive logic. In *Studies in History and Philosophy of Science Part A* 42 (1), pp. 240–244. DOI: 10.1016/j.shpsa.2010.11.016.