A causal approach to analogy[1]

Wolfgang Pietsch, wolfgang.pietsch@tum.de

Abstract: Analogical reasoning addresses the question how evidence from various phenomena can be combined and made relevant for theory development and prediction. In the first part of my contribution, I review some influential accounts of analogical reasoning, both historical and contemporary, focusing in particular on Keynes, Carnap, Hesse, and more recently Bartha. In the second part, I sketch a general framework. To this purpose, a distinction between a predictive and a conceptual type of analogical reasoning is introduced. I then take up a common intuition according to which (predictive) analogical inferences hold if the differences between source and target concern only irrelevant circumstances. I attempt to make this idea more precise by addressing possible objections and in particular by specifying a notion of causal irrelevance based on difference making in homogeneous contexts.

1. Introduction

When evidence from different phenomena is combined in order to predict, to explain or to develop a conceptual framework, this can often be understood in terms of analogical reasoning. After all, analogical inferences, according to a typical explication, are inferences based on similarity: If two phenomena, source A and target A*, are similar and A has a

[1] Published in *Journal for General Philosophy of Science*, https://doi.org/10.1007/s10838-019-09463-9

characteristic C, then under certain circumstances it is plausible or probable to assume that A* has characteristic C as well.

At all times in history, scientists have stressed the epistemological significance of analogy, including such luminaries as William Gilbert, Johannes Kepler, Joseph Priestley, or James Clerk Maxwell. Johannes Kepler, for example, wrote in his *Opticks*: “I cherish more than anything else the Analogies, my most trustworthy masters. They know all the secrets of nature.” (Kepler 1604; cited in Polya 1954, p. 12) And indeed, analogical reasoning was a major source of creativity in Kepler’s scientific method. In his analysis of the solar system, he crucially relied on the analogy between the emission of light and the propagation of what he called the *anima motrix*, i.e. the spirit that moves the planets around the sun:

> “Let us suppose, then, as is highly probable, that motion is dispensed by the Sun in the same proportion as light. Now the ratio in which light spreading out from a center is weakened is stated by the opticians. For the amount of light in a small circle is the same as the amount of light or of the solar rays in the great one. Hence, as it is more concentrated in the small circle, and more thinly spread in the great one, the measure of this thinning out must be sought in the actual ratio of the circles, both for light and for the moving power [motrice virtute].” (Kepler, 1596/1981, p. 201, cited in Gentner et al. 1997, p. 414-415)

Thus, the analogy suggests that the anima motrix, just as light, constitutes a conserved quantity acting according to an inverse square law. The example demonstrates well, how evidence from two different sources, i.e. the theory of optical phenomena and phenomenological knowledge about the solar system, can be combined in order to develop a model concerning the interaction of material bodies in the solar system.

As another example of analogical reasoning in the sciences, consider animal models that are used in medicine and pharmacology to determine the efficacy of a treatment in human beings. Again, evidence from disparate phenomena, here mice and human beings, is amalgamated to further the knowledge about these phenomena. I will argue later that this example is different from the previous one in important respects. Most importantly, it aims at prediction, while Kepler was primarily concerned with theory or model development.

In view of these examples of successful scientific practice, it is remarkable that influential authors have questioned, whether there are any universal rules governing analogical inferences.[2] For example, Paul Bartha, who has written the most extensive modern-day treatise on analogical reasoning (2010), states: “Despite the confidence with which particular analogical arguments are advanced, nobody has ever formulated an acceptable rule, or set of rules, for valid analogical inferences. There is not even a plausible candidate.“ (Bartha 2013, Sec. 2.4) In a similar vein, Patrick Maher writes: „Argument by analogy is a generally accepted form of inductive reasoning and many think that inductive reasoning can be represented using the probability calculus. From these facts one might expect that there would

[2] Note that the absence of universal rules for analogical inferences does not necessarily imply that such inferences cannot be reliable. An interesting proposal in this regard is John Norton’s *material theory of induction* (cp. Norton 2011 and references therein). A critique of Norton’s theory of induction is beyond the scope of this paper.

be accepted probability models that can represent inference by analogy, but no such model exists.” (Maher 2001, p. 183) As an example from the statistics and computer science literature, Henri Prade and Gilles Richard write in their recent overview of the field: “analogical reasoning is not amenable to a formal framework in a straightforward manner due to the brittleness of its conclusions.” (2014, 5)

The most pressing and interesting epistemological problem with respect to reasoning by analogy therefore is how to bring these two aspects together, on the one hand, the ubiquitous use of analogy in scientific practice and, on the other hand, the widespread belief that a formal framework for analogical inferences does not exist.

In the next section, relying on short case studies from the history of scientific method, I argue for three interrelated points. First, I briefly present Carnap’s framework of induction, building mainly on enumerative induction. While he tries to implement analogical reasoning in his approach, he fails to find a convincing manner to do so. This situation leads me to argue for a general failure of enumerative approaches to implement analogical reasoning. Instead, eliminative approaches, focusing on the variation of circumstances rather than the repetition of instances as in enumerative induction, are much more amenable to analogical reasoning, as the second case study on Keynes’ approach to induction shows. Third, I introduce two influential contemporary frameworks by Mary Hesse and Paul Bartha, which address one of the major problems of Keynes’ approach, the so-called counting problem. To this purpose, they develop a two-dimensional framework, which takes into account the ‘horizontal’ similarities between different phenomena, but also the ‘vertical’ nature of the relations between similarities and differences.

In the third section, a distinction between two types of analogical reasoning is introduced, namely conceptual and predictive analogies. These differ in their epistemic aim, the nature of the vertical relations, the criteria of evaluation, and the methodological framework. I argue that the widespread skepticism concerning analogical inferences partly results from a failure to recognize this distinction. While conceptual analogies indeed are not amenable to a formal framework to determine the truth or probability of such inferences, this is not the case for predictive analogies.

In section four, I then sketch a framework for predictive analogies building on the intuition that ‘a predictive analogical inference holds, if the differences between source and target are irrelevant to the prediction’. I discuss some preliminary objections and argue that irrelevance must be understood in causal terms. Examining different explications of the notion of causal irrelevance from the literature, I find none of them suitable for the context of analogical reasoning. My own proposal construes causal irrelevance in terms of difference making in a given background context.

Since the framework that was developed so far is intended for deterministic situations, I briefly address in section five, how it can be extended to include probabilistic analogical inferences. While there are straightforward ways to implement probability, a crucial problem remains regarding the interpretation of probability in this context.

2. Three historical perspectives

The history of methodological thinking about analogy is quite rich. In the following, I concentrate on three more recent episodes or case studies how methodologists have approached analogical reasoning. These will provide the groundwork for the approach to be outlined later in the article.

2a. Carnap and the inadequacy of enumerative approaches

Rudolf Carnap developed one of the most extensive and detailed inductive frameworks in the 20th century, in which he explicitly aimed to include considerations of analogy. Carnap's approach is based on a confirmation function c(h|e), which designates the confidence in a hypothesis h based on some evidence e. As is well-known, Carnap was a dualist about probability, distinguishing an empirical and a logical role of probability—the former regarding relative frequencies while the latter is usually identified with rational degree of belief in a hypothesis based on some evidence.

Carnap construes analogical inferences as inductive inferences from one individual to another based on their known similarity, much in line with the general understanding that was presented in the introduction: "The evidence known to us is the fact that individuals b and c agree in certain properties and, in addition, that b has a further property; thereupon we consider the hypothesis that c too has this property." (1950, p. 569)

Carnap's general approach to induction is based on what is often called the 'straight rule' of induction: Given a family of predicates P, i.e. a mutually exclusive but exhaustive group of predicates that applies to a number of individuals, the degree of confirmation corresponds to the relative frequency s_j/s of a property P_j in the first s individuals. In other words, the straight rule of induction is just ordinary enumerative induction. Carnap recognizes the deficiencies of this simple rule and consequently extends it to 'a continuum of inductive methods' which is determined by a number of additional parameters. There are several versions in his writing over the course of his life, the best known being the so-called λ-γ system developed in his mature, posthumously published *Basic System of Inductive Logic* (1971, 1980) with a confirmation function

$$c_j(s_1, \dots, s_k) = \frac{s_j + \lambda\gamma_j}{s + \lambda}.$$

Here, $s = s_1 + \dots + s_k$ can be interpreted as the number of real individuals and λ the number of virtual individuals. Among the former s_j have the property P_j, among the latter $\lambda\gamma_j$. This confirmation function can be rewritten in terms of an empirical and a logical part:

$$c_j(s_1, \dots, s_k) = \frac{s}{s+\lambda}\frac{s_j}{s} + \frac{\lambda}{s+\lambda}\gamma_j$$

For large s, the empirical part dominates, for small s, the logical part. Thus, the logical part can be interpreted as an *a priori* contribution to the confirmation function.

In general, analogical influence is considered to belong to this logical part. Carnap specifies several kinds of analogical influence. First, he draws a distinction between similarity influence, which takes into account the distance between properties, and proximity influence referring to the distance between individuals—presupposing in both cases that an adequate metric exists. With respect to the former, Carnap further distinguishes between analogical influence within one predicate family and that between different predicate families. While he acknowledges that the latter is much more common than the former, he mainly addresses in the *Basic System* analogical influence within one predicate family, presumably because it is the simpler problem (for a very brief discussion of analogical influence between different predicate families, see Carnap 1950, §110 D). Furthermore, Carnap's analysis of analogy is restricted to individuals which have certain properties in common, while in typical analogical inferences individuals are also known to differ in certain other properties—a critique spelled out in some detail by Mary Hesse (1964).

Carnap suggests treating analogical inferences in terms of the mentioned γ corresponding to the width (or weight) of properties and an additional η corresponding to the distance between properties. If two properties P_1 and P_2 are sufficiently similar, i.e. are close in terms of the distance measure, then the relative frequency of P_1 will influence the confirmation function for P_2 and vice versa. Naturally, the width also has to be taken into account: basically, the more weight a property has, the greater its influence. According to Carnap, such analogy influence "is usually very small", it "decreases with increasing [evidence in terms of number of individuals] s", and therefore can "be practically neglected" if s is large (1980, p. 41). To repeat, this is because analogy influence belongs to the logical and a priori part of the confirmation function, which can be neglected for $s >> \lambda$. As an example, Carnap uses the color space to illustrate the concepts of width, essentially the range or variation subsumed under a specific color, and distance, i.e. the perceived similarity between different colors. Both are determined by the chosen metric of the color space (1980, Sec. 14.A).

Carnap's treatment of analogy remains brief and fragmentary—in contrast to his very detailed treatment of induction in general—and this situation may already cast doubt over the suitability of enumerative approaches to analogy, i.e. essentially those approaches that are based on some version of the straight rule. There have since been a number of attempts to integrate analogical reasoning within an essentially Carnapian approach to inductive logic (e.g. Hesse 1964, Kuipers 1984, Romeijn 2006, Maher 2001). It seems fair to say that no agreement has been reached (for a helpful overview, see Huttegger forthcoming). Many decades after Carnap published his approach to inductive logic, it continues to be doubtful whether his framework is capable to cover analogical reasoning in a sensible manner.

One strain of criticism attacks the use of additional parameters such as γ or η which must be derived from a metric over properties, which rarely is explicitly available. These parameters seem considerably ad hoc as is well illustrated by the example of the color space for which a wide variety of representations are possible (Reibe & Steinle 2002). In fact, this situation has led Wolfgang Stegmüller, a close collaborator of Carnap, to suggest that Carnap is really talking about subjective rather than logical probability (Stegmüller 1973, 514)—which would further undermine any attempt to justify reliable predictions based on analogical reasoning, even though these are ubiquitous in the sciences, as the examples from the introduction suggest.

In the end, what seems the most problematic aspect about Carnap's approach is its focus on the straight rule and on relative frequencies as the core concepts for confirmation—automatically confining analogy to prior considerations, which wash out as increasing evidence in terms of instances is gathered.[3] After all, scientific practice suggests otherwise: relative frequencies are generally a bad indicator for confirmation, while analogies can often provide highly reliable evidence. The lesson from the case study on Carnap's treatment of analogy thus seems to be that just as enumerative approaches to induction in general, enumerative approaches to analogy, confining analogy to prior considerations, run into deep and presumably unsolvable problems.

2b. Keynes and the ubiquity of analogical reasoning

It is often thought that the essence of inductive reasoning lies in the multiplication of instances and Carnap's approach with its reliance on the straight rule and on relative frequencies attempts to formalize this intuition. However, there has been for many centuries an alternative tradition of inductive reasoning which focuses on the variation of circumstances rather than on the number of instances. Proponents of this later tradition, which is sometimes referred to as eliminative induction, are among others Francis Bacon, John Stuart Mill and more recently John Maynard Keynes. It turns out that its basic inductive framework is much more amenable to analogical reasoning. After all, an analogical inference concludes from one instance with certain circumstances to another with different circumstances. Indeed, proponents of eliminative induction have often considered analogical inference as the core of inductive reasoning. The best example in this regard is John Maynard Keynes, who in his *Treatise on Probability* lays out a general framework for induction based on analogy:

> „In an inductive argument, therefore, we start with a number of instances similar in some respects AB, dissimilar in others C. We pick out one or more respects A in which the instances are similar, and argue that some of the other respects B in which they are also similar are likely to be associated with the characteristics A in other unexamined cases. The more comprehensive the essential characteristics A, the greater the variety amongst the non-essential characteristics C, and the less comprehensive the characteristics B which we seek to associate with A, the stronger is the likelihood or probability of the generalisation we seek to establish." (Keynes 1921, 219-220)

Note again that Keynes's description closely resembles what we had defined as an analogical argument in the introduction, while he considers it the fundamental form of an inductive argument. Keynes introduces some terminology that has since become standard in the literature on analogical reasoning. The *positive analogy* concerns those properties which source and target have in common, the *negative analogy* those properties in which source and target differ, and the *unknown analogy* those properties of which it is yet unknown whether they belong to the positive or negative analogy. Finally, the *hypothetical analogy* concerns

[3] Note that Bayesian approaches to confirmation often assume a similar role for analogy as being confined to prior considerations (e.g. Salmon 1990): "I suspect that the use of arguments by analogy in science is almost always aimed at establishing prior probabilities. […] The moral I would draw concerning prior probabilities is that they can be understood as our best estimates of the frequencies with which certain kinds of hypotheses succeed. These estimates are rough and inexact..." (186-187).

those properties which are known of the source phenomenon and predicted of the target phenomenon (see also Bartha 2013).

Keynes' approach to induction turns the Carnapian view upside down.[4] While for Carnap enumerative induction in the form of the straight rule is central and analogy is confined to prior considerations that wash out with increasing evidence, for Keynes, analogical inferences are fundamental and enumerative induction only plays a subordinate role by controlling for circumstances whose influence thus far has not been explicitly considered:

> "The object of increasing the number of instances arises out of the fact that we are nearly always aware of *some* difference between the instances, and that even where the known difference is insignificant we may suspect, especially when our knowledge of the instances is very incomplete, that there may be more. Every new instance *may* diminish the unessential resemblances between the instances and by introducing a new difference increase the Negative Analogy. For this reason, and for this reason only, new instances are valuable.“ (Keynes 1921, 233)

Relatedly, Keynes denies that relative frequencies can be used to determine probabilities along the lines of the straight rule. The reason is that instances vary in different ways regarding their circumstances and thus there is usually no reason to count them with equal weight as the straight rule presupposes:

> „I do not myself believe that there is any direct and simple method by which we can make the transition from an observed numerical frequency to a numerical measure of probability." (Keynes 1921, 367)

In summary, Carnap's system implements a clear distinction between enumerative induction and analogy, it confines analogical influence to a priori considerations, and it endorses a principle of instantial relevance ("one of the basic characteristics of customary inductive reasoning", Carnap 1971, 161), according to which any positive instance strictly increases the confirmation function that the next instance is positive as well.[5] All this is incompatible with Keynes's approach, who argues that all induction basically relies on analogy, even seeming applications of enumerative induction actually aim at increasing the negative analogy. He rejects any simple frequentist approach to confirmation, which quantifies confirmation based on some variant of the straight rule. Relatedly, he rejects the principle of instantial relevance: in particular, if two instances are fully identical in all their relevant circumstances, then the additional instance does not confirm at all (1921, 233).

Unfortunately, the shift away from enumerative induction to an inductive framework based on analogy, while conceptually sensible, eliminates the most obvious candidates for a measure of confirmation, namely the number of positive instances or relative frequencies. Instead, a quantitative measure could consist in a weighted comparison between positive and negative analogy. Bartha suggests the following characterization of this widespread intuition:

---

[4] i.e. systematically speaking, historically of course Keynes was prior to Carnap.

[5] Carnap qualifies that the strict inequality only holds if the original confirmation function is not zero or one.

> "Suppose S and T are the source and target domains. Suppose $P_1, \ldots, P_n$ (with $n \geq 1$) represents the positive analogy, $A_1, \ldots, A_r$ and $\neg B_1, \ldots, \neg B_s$ represent the (possibly vacuous) negative analogy, and Q represents the hypothetical analogy. In the absence of reasons for thinking otherwise, infer that Q^* holds in the target domain with degree of support $p > 0$, where p is an increasing function of n and a decreasing function of r and s." (2013, Sec. 2.4)

But, as many authors including Bartha have stressed, this approach leads to the notorious *counting problem*. While counting instances in enumerative induction seems straight-forward, counting properties in analogical reasoning is not. If two instances have property 'color' in common, but differ in property 'size', how possibly should one compare color and size? It appears impossible to formulate general rules for this task, which has led many to conclude that analogical reasoning is necessarily contextual. As a result, Keynes' approach remains almost entirely qualitative, which may have contributed to the fact that it is barely used in contemporary science.

Still, Keynes does derive some general guidelines for analogical reasoning. Inductive arguments which conclude from a number of examined instances to a generalization can be strengthened by the following means:

- "by reducing the resemblances known to be common to all the instances, but ignored as unessential by the generalization,
- by increasing the differences known to exist between the instances,
- by diminishing the sub-analogies or unessential resemblances known to be common to some of the instances and not known to be false of any." (Keynes 1921, 231-232)

For this, either new instances have to be examined or the knowledge of familiar instances has to be extended. Most standard treatments of analogical reasoning propose similar qualitative guidelines (see Bartha 2013, Sect. 3.1 for a comprehensive list of commonsense guidelines).

In summary, Keynes' framework bases inductive reasoning on analogical inferences, i.e. every inductive inference is conceived as an inference based on similarity. While this is conceptually plausible, proponents have largely failed to come up with a quantitative confirmation measure for such an approach.[6]

2c. Hesse, Bartha and the two-dimensional approach

No solution to the counting problem seems to be forthcoming. Apparently, how properties are counted very much depends on the specific context. There is, however, one crucial insight that has occasionally been pointed out in discussions of analogical reasoning, but that was most forcefully stressed by Mary Hesse and more recently by John Norton and Paul Bartha. For analogical reasoning it is important to not only consider the similarity and differences in

[6] While the approach proposed in this essay builds on Keynes's ideas in many ways, one of the advantages with respect to Keynes is that to some extent it is quantitative. In particular, a sufficient and necessary criterion is given for analogical inferences in deterministic contexts. Thus, analogical inferences fulfilling this criterion are valid with probability 1. In Section 5, an extension of the proposed framework is briefly sketched, under which circumstances one can meaningfully assign a probability to a prediction based on an analogical reasoning.

properties between source and target, but also the nature of the relation between these properties:

> „Under what circumstances can we argue from, for example, the presence of human beings on the earth to their presence on the moon? The validity of such an argument will depend, first, on the extent of the positive analogy compared with the negative (for example, it is stronger for Venus than for the moon, since Venus is more similar to the earth) and, second, on the relation between the new property and the properties already known to be parts of the positive or negative analogy, respectively. If we have reason to think that the properties in the positive analogy are causally related, in a favorable sense, to the presence of humans on the earth, the argument will be strong. If, on the other hand, the properties of the moon which are parts of the negative analogy tend causally to prevent the presence of humans on the moon the argument will be weak or invalid." (Hesse 1966, 58-59; cited in Norton 2011, 8)

In other words, Hesse proposes a two-dimensional model, where the horizontal relations concern the similarity between source and target, i.e. the identity or difference in properties, and the vertical relations concern the relations between properties, which Hesse believes to be causal in most cases. Simply comparing the negative and the positive analogy thus will not do, but rather the nature of the relationship between the properties in the positive and the negative analogy with the properties in the hypothetical analogy has to be taken into account.

In his recent influential work on analogical reasoning, Paul Bartha very much builds on Hesse's two-dimensional account (Bartha 2010, briefly summarized in 2013, Section 3.5.2). He classifies different types of analogical reasoning in terms of different vertical relations, e.g. logical, causal, or statistical. Bartha's *principle of prior association* then demands that some kind of connection between the positive analogy and the hypothetical analogy has to be established, taking into account the negative analogy as well. Bartha's second principle, the *principle of potential for generalization*, requires that there should be reason to expect that the relationship between positive and hypothetical analogy in the source obtains for the target as well. In particular, there should be no "critical disanalogy" between source and target.

Let me emphasize again that these modern authors have established that any reasonable approach to analogy has to take into account both similarity in properties between source and target as well as the relations between these properties and the hypothetical analogy. The proposal in this essay builds on this important idea, a more detailed critique of both Hesse and Bartha unfortunately is beyond the scope of this paper.

3. Predictive and conceptual analogies

In the following, I introduce a distinction between predictive and conceptual analogies, which differ in various respects: concerning the epistemic aim, the nature of the vertical relations, the criteria of evaluation, and the methodological framework.[7] Arguably, the failure to clearly

[7] The proposal is embedded within a broader distinction between phenomenological science on the one hand and abstract or theoretical science on the other hand. Perhaps the most important difference between

hold these types of analogical reasoning apart has led to considerable confusion in the debate on analogical reasoning. Maybe most importantly, only for conceptual analogies the role of analogical reasoning is primarily heuristic, while predictive analogies aim at true or at least probable inferences. As argued in section 2b, when discussing Keynes' approach, the latter type of analogies constitutes the core of inductive and causal reasoning.

An example of a predictive analogy is the use of animal models such as the mouse model in pharmacology to determine the effectiveness of certain medication to cure diseases in human beings. But as will become clear in the course of this paper and as was already emphasized by Keynes, any inductive inference from one instance to another instance, when aiming at truth or at least probability, can be construed as a predictive analogy. Further examples to be discussed in Section 4e concern predictions of the period of a pendulum or inferences to life on other planets.

Predictive analogies aim to establish reliable prediction or effective intervention. Consequently, the relevant vertical relationships must be of causal nature. This follows from a view of causation in the sciences as the crucial concept to distinguish between effective and ineffective strategies—as developed by Nancy Cartwright and others (especially Cartwright 1979). Only if there is some causal link between administering the medication and recovery both in the mouse and in the human being, the analogical inference is reliable.

More exactly, a strategy how to effectively intervene in a phenomenon has to be based on a direct causal relationship between some circumstances in the positive analogy and the hypothetical analogy. Similarly, a reliable prediction must be based on some causal connection, which however need not consist in a direct causal link, but can also result from a common cause structure. In particular, an analogical inference aiming at prediction may infer from a correlation between two variables with a common cause in the source phenomenon to a similar correlation in the target phenomenon. By contrast, a merely accidental correlation that does not result from some causal connection cannot be used either for prediction or for intervention.

One might worry that the above argument presupposes Reichenbach's principle of common cause (1956, 157–159), which is controversial (e.g. Sober 1988). Colloquially, this principle can be formulated as follows: 'If there is a correlation between two events, then this correlation must be either due to a direct causal connection between the correlated events or due to a common cause.'[8] Clearly, in the above argument such a principle of common cause is

---

phenomenological and theoretical science concerns the aim: the former is mainly interested in reliable prediction and successful manipulation, the latter in the development of a conceptual and explanatory framework. Thus, predictive analogies fit well with phenomenological science, conceptual analogies fit well with theoretical science. There are a number of further characteristics that both distinctions share, for example whether the laws that are used are causal or not. Some of the claims in this section can only be understood from the perspective of this broader distinction between phenomenological and abstract science, for which unfortunately I cannot argue here due to lack of space. Notable scholars, who have made and argued for the distinction, include Duhem (1954) and Cartwright (1983).

[8] This is a typical formulation (e.g. Sober 2001, 331). Reichenbach was somewhat more cautious: "the *principle of the common cause* […] can be stated in the form: *If an improbable coincidence has occurred, there must exist a common cause.* […] Chance coincidences, of course, are not impossible […] The existence of a common cause is therefore […] only probable. This probability is greatly increased if coincidences occur repeatedly." (1956, 157-158)

not assumed, since accidental correlations are possible, which by definition do not result from a common cause. Elliott Sober's well-known example of a correlation between Venetian sea levels and British bread prices is a plausible candidate for an accidental correlation.

However, such an accidental correlation is not a *reliable* correlation and thus cannot be used for *reliable* prediction. Therefore, accidental correlations cannot support sound analogical arguments. A *reliable* correlation, as I understand it here, requires that some reason exists, why the correlation holds and continues to hold. Such a reason could be a direct causal link, a common cause, or a definitional or normative relationship between variables.[9] For accidental correlations, more or less by definition, such a reason does not exist. Therefore these cannot be employed for *reliable* prediction, even though a prediction based on an accidental correlation may very well turn out to be true by chance. Definitional or normative relationships between variables hold by stipulation and therefore cannot ground predictive analogical inferences, which concern empirical relationships.

In summary, no matter whether they aim at effective intervention or at reliable prediction, predictive analogies always have to establish a causal relationship in the target phenomenon based on some knowledge about a corresponding causal relationship in the source phenomenon.

Predictive analogies are evaluated by verifying whether an intervention works, which is suggested by the analogy, or whether a prediction turns out to be true. After all, there is a matter of fact, whether a medication that cures a disease in a mouse will also lead to recovery in a human being afflicted by a similar disease. Of course, as this example demonstrates, such predictive analogies will in general not be deterministic, but statistical, i.e. they will only hold with a certain probability. Thus, methodological frameworks for predictive analogies try to determine the truth or at least probability for analogical inferences. Both Carnap's and Keynes' approaches to analogy, as delineated in the previous sections, are examples of such probabilistic frameworks for analogical reasoning—covering chiefly predictive analogies.

An example for a conceptual analogy is the analogy between the transfer of heat and interaction in electromagnetic phenomena as it was elaborated in great detail by William Thomson and James Maxwell towards the end of the 19th century—resulting in the modern particle-field theory of classical electrodynamics:

> "The laws of the conduction of heat in uniform media appear at first sight among the most different in their physical relations from those relating to attractions. The quantities which enter into them are *temperature*, *flow of heat*, *conductivity*. The word *force* is foreign to the subject. Yet we find that the mathematical laws of the uniform motion of heat in homogeneous media are identical in form with those of attractions varying inversely as the square of the distances. We have only to substitute *source of heat* for *centre of attraction*, *flow of heat* for *accelerating effect of attraction* at any

[9] Due to lack of space, we cannot address here certain interesting, but controversial cases, such as correlations due to indeterministic relationships, which arise for example in connection with the Bell Inequalities, or correlations due to conservation laws.

> point, and *temperature* for *potential*, and the solution of a problem in attractions is transformed into that of a problem in heat. […]
>
> It is by the use of analogies of this kind that I have attempted to bring before the mind, in a convenient and manageable form, those mathematical ideas which are necessary to the study of the phenomena of electricity." (Maxwell 1855/56, 157)

As is clear from this quote, Maxwell's aim in developing the analogy between heat and electricity is not primarily prediction or intervention. Rather, Maxwell wants to develop a conceptual framework for electromagnetic phenomena based on another framework that was more familiar and much better developed at the time, namely the theory of heat. Such reasoning facilitates transferring certain results and solutions from one field to the other.

Since the primary aim is neither prediction nor intervention, the relevant vertical relationships in such conceptual analogies are in general not causal—arguing again with a Cartwrightian concept of causation as sketched above. In the example of classical electrodynamics, there are good reasons to assume that the considered relationships are to considerable extent definitional or conventional. In particular, this perspective is in accordance with a standard view on the nature of axioms and laws of fundamental scientific theories—interpreting these as implicit definitions of basic theoretical terms. Certainly, it cannot be the place here to defend this view, but typical arguments range from underdetermination of abstract theory to the observation that the laws in fundamental theories are too abstract to have themselves considerable empirical content. Only when supplemented by further assumptions, e.g. bridge principles according to the classic syntactic view of scientific theories, do these laws acquire empirical meaning. This observation alone might suffice to establish the non-causal nature of the fundamental laws of abstract scientific theories.

Relatedly, conceptual analogies are evaluated by whether they play a fruitful role in transferring established solutions and results from one field to another rather than in terms of truth and probability. While in predictive analogies, one can verify whether an analogical inference corresponds to a matter of fact, e.g. whether a prediction turns out true or not, this is in general not possible for conceptual analogies. To verify, whether a Poisson equation for the electric potential holds, when postulated in analogy to the Poisson equation for temperature in the theory of heat, is certainly not as simple as verifying predictive analogies. One reason lies in the considerable underdetermination of abstract conceptual frameworks. Indeed, Maxwell stressed the underdetermination of classical electrodynamics insisting that there exists considerable flexibility how to formulate the fundamental laws. For example, a choice between action at a distance and field theory in electrodynamics remains possible (Pietsch 2012).

Thus, conceptual analogies are a creative endeavor. Whether they hold, is not so much a matter of truth and probability but to considerable extent depends on the ingenuity of the scientists—whether they are successful in mapping (part of) the fundamental structure from one phenomenon to the other. Consequently, such analogies cannot be treated in terms of probabilistic frameworks like those of Carnap or Keynes. Approaches to analogical reasoning based on structure mapping, such as from the work of Dedre Gentner (1983), seem much more adequate. Gentner’s framework relies on a classification of various entities, attributes and relations as well as a quite sophisticated set of inference rules. Analogies are evaluated according to a *systematicity principle*, essentially that those analogies are more plausible that result from a mapping of mutually connected higher order relations compared with those mapping only isolated properties. Note that this main criterion of the structure mapping theory can hardly be translated into probabilities and consequently, Gentner’s theory, while well suited for conceptual analogies, seems unable to serve as a framework for predictive analogies. Rune Nyrup's pursuit worthiness account is another example of an approach intended for conceptual analogies (2016).

From yet another perspective, a relatively sharp criterion to distinguish between predictive and conceptual analogies concerns a difference in epistemic attitude when formulating analogies. In the case of predictive analogies, one is primarily interested in whether the respective inferences turn out true or not. By contrast in the case of conceptual analogies, one is prepared to engage in considerable conceptual reevaluation trying to reframe and redefine relevant notions in order to make the analogy work. A conceptual analogy, thus, is never a simple prediction but rather presupposes a substantial willingness of the scientist to try to make the analogy fit the facts.

To some extent, conceptual analogies also aim at truth in that the conceptual framework, which is developed on the basis of such analogies, is at some point used to make predictions about the phenomena, for which the framework is intended. However, the primary focus is on developing a simple, but fruitful conceptual basis with considerable explanatory power, while the truth or probability of any predictions is only an indirect or secondary aim. In particular, true or probable predictions based on the conceptual framework become important only at a later stage, once the framework is sufficiently developed.

While the difference in attitude provides a relatively sharp criterion to distinguish between predictive and conceptual analogies in scientific practice, one and the same analogy can still be framed as either predictive or conceptual depending on the respective attitude of the person formulating the analogy. For example, a scientist could use a mouse model to predict the efficacy of a medication in humans, but could also use the same mouse model in order to develop an understanding of how specific phenomena in the human body work.

Predictive and conceptual analogies are the main types of analogical reasoning in the empirical sciences. Whether there are other types, for example in mathematics, is a difficult question. Given the non-empirical nature of mathematics, predictive analogical inferences in the above sense will not play a role in this field. With respect to conceptual analogical reasoning, some variant is likely to be used in mathematics, not least in view of the substantial similarities between mathematics and theoretical physics. Whether other types of analogical inferences are employed, ultimately depends on the epistemological status that is attributed to mathematics. However, that question leads us far away from the actual topic of the essay (see e.g. Bartha 2010, Ch. 5).

4. A deterministic framework for predictive analogies

4a. A first suggestion

There exists a core intuition about valid analogical reasoning that can be found across the literature and that is in line with the two-dimensional model sketched in Section 2c. This intuition is for example incorporated in Bartha's second principle that for valid analogical inferences no essential disanalogy between source and target should exist. The basic idea is the following (PA):

> A (predictive) analogical inference holds, i.e. the hypothetical analogy is true for the target,[10] if and only if the negative analogy concerns only *causally irrelevant* circumstances.

Note that in line with the distinction introduced in Section 3, the vertical relations of interest are causal in nature since the focus lies on predictive inferences. To repeat, this insight stems from a Cartwrightian understanding of causation, the core feature of which is to draw a distinction between effective and ineffective strategies, including between reliable and unreliable prediction.

I will in the following suggest a methodology for predictive analogical inferences that builds on the core intuition (PA). Before discussing the crucial notion of causal irrelevance, let me briefly point out some possible objections against the proposed approach which are then mostly addressed later on in order to refine (PA). A first issue concerns situations, where an analogical inference is valid even though some circumstances in the negative analogy are causally relevant—i.e. (PA) is not a necessary condition for predictive analogical inferences.[11] Notably a factor may be causally relevant, but may play no role in the considered analogy, because other contributing factors are not instantiated, e.g. the burning match does not cause a fire since there is no combustible material present. Also, the influences of some causally relevant circumstances could exactly cancel each other. For example, one might infer from the acceleration that a stone receives on the earth to the acceleration that a stone of the same mass receives on the moon. The acceleration is indeed the same, if the difference in gravitational field is exactly compensated by an acceleration of the system of reference on the moon. Similarly, the same effect can be due to alternative causes, e.g. the acceleration of a body may be caused by gravitational or by electromagnetic fields. An analogical inference may still be valid even if in various instances different alternative causes are active, if the effects of these different causes add up to the same result.

---

[10] One might object that the validity of an analogical inference should not be confused with whether a prediction turns out true or not. Notably, it has been argued that valid inferences are those that adhere to commonly accepted methodological conventions, largely independently of empirical success. However, in the case of (predictive) analogical inferences, a necessary and sufficient criterion for empirical success can be stated. Under these circumstances, it seems adequate to identify valid (predictive) analogical inferences with those that obey the criterion.

[11] In this category falls a well-known example concerning the inference that there is life on Mars based on the existence of life on Earth, even though there apparently are relevant differences between both planets. How to deal with such examples will be outlined in Section 4c.

Secondly, certain cases suggest that (PA) is not a sufficient condition for predictive analogical inferences. In particular, predictive analogical inferences may sometimes be based on relationships other than causal relevance, e.g. on mere correlations. More exactly, even if the negative analogy is causally irrelevant, the analogical inference could nevertheless fail to hold due to mere correlations between some circumstances in the negative analogy and the hypothetical analogy. To use a well-known example, one might infer that the bread price in London this year will be the same as the bread price in London last year because the negative analogy is causally irrelevant. However, upon reading an essay by Elliott Sober (2001), one discovers that there exists a strong correlation between London bread prices and Venetian sea levels. In addition, there are clear indications that the sea level in Venice this year is much higher than last year. This seems to provide substantial evidence to infer that the analogy fails, i.e. bread prices in London will not remain the same. Even though Venetian sea levels presumably are causally irrelevant to London bread prices, a change in the former may suggest a change in the latter due to the mentioned non-causal correlation—contradicting the claim that only causally relevant circumstances are important for analogical inferences.

Relatedly, predictions are sometimes based on definitional relations. This can again result in situations where analogical inferences fail to hold even though the causal structure has not changed. For example, an analogical inference from the gravitational field in one location to another at the same distance from the earth could fail just because the concept of a gravitational field is understood differently in both situations.

A third point concerns the distinction between properties (which are ‘one-place’) and relations (which are ‘many-place’). While the Keynesian terminology of positive and negative analogy suggests a focus on properties rather than relations, many scholars insist that analogy is less about a supposedly superficial similarity in terms of common properties of source and target, but rather about similarity in terms of relations. For example, in the analogy between heat and electricity, the essential similarity is not between corresponding terms such as temperature and electric potential or source of heat and charge. Rather it concerns relations between these terms, e.g. that they obey a Poisson equation.

To resolve this issue, note first that relations always link properties with each other. Thus, it would be wrong to think that one could exclusively focus on relations neglecting properties altogether. The Poisson equation, for instance, relates temperature and sources of heat as well as charges and electric potential. Furthermore, the proposed approach (PA) obviously takes into account relationships as well, by examining the causal relevance or irrelevance of certain properties for others.

It might still be questionable, whether complex analogies can be formulated in terms of positive and negative analogies. After all, it does not appear straightforward how to compare concepts like temperature and electric potential in terms of differences and similarities? In response, it should be stressed that if shared relations exist one can always formulate shared properties corresponding to these relations. For example, both temperature and electric potential share the abstract property that they serve as potentials which by means of corresponding forces lead to the distribution of certain quantities. By contrast, electric potential and temperature differ in terms of the nature of the potential, in particular regarding the quantity on which it acts, namely either charged matter or heat. In this manner, positive and negative analogy can be distinguished. With sufficient ingenuity, this is always possible.

Fourth and last, there are substantial worries concerning the notion of causal irrelevance. For example, it is far from certain, whether causal irrelevance can ever be established at all. After all, a circumstance that is normally considered irrelevant may suddenly become causally relevant in some obscure situation. The constellation of the stars at birth is usually not considered relevant to the fate of a person, but in some contrived story it might have an impact. For example, the person may be superstitious and the astrological prediction of a psychic may be so scaring that it becomes a self-fulfilling prophecy. The ultimate lesson to draw from such counterexamples is that causal irrelevance is context-dependent and that in an explication of analogical reasoning this must be taken into account. Such context-dependence is of course not surprising to anyone familiar with the philosophical debate on causation. It was stressed in particular by John Mackie, who in his approach to causation introduced the crucial notion of a causal field, to which all causal statements are relative (1980).

Whether the basic intuition (PA) has merits or not, crucially depends on the construal of the notion of causal irrelevance. To this issue we will turn now.

4b. The notion of causal irrelevance

In the following, I discuss several suggestions from the literature how to define causal irrelevance and based on these will later lay out my own proposal. All in all, it seems fair to say that the notion of causal irrelevance has not played a central role especially in philosophical accounts of causation, which are almost exclusively focused on the notion of cause in a positive sense, i.e. on causal relevance. Therefore, the following overview can be rather brief.

First, one might try to define causal irrelevance based on statistical independence.[12] The most straightforward connection between both notions originates within a probabilistic approach to causation (see e.g. Hitchcock 2016 for a useful overview). If, broadly speaking, causal relevance of an event C to another event E is identified with the increase or decrease of the conditional probability $P(E|C) \lesseqgtr P(E|\neg C)$, it seems natural to define causal irrelevance of C to E in terms of an unchanged probability $P(E|C) = P(E|\neg C)$. As mentioned, most accounts of probabilistic causation do not explicitly address the notion of causal irrelevance in much detail. A notable exception in this regard is Ellery Eells who distinguishes positive, negative, mixed, and neutral causal relevance—the latter corresponding to causal irrelevance (Eells 1991).

One important problem for a definition of causal relevance and irrelevance along these lines are common cause structures, where a correlation between two variables F and G does not result from a direct causal link between them, but rather from a common cause H that is causally relevant to both variables. Let us assume in the following for reasons of simplicity that all variables are binary. Even though no direct causal relevance between the variables F and G exists, the conditional probability changes, e.g. $P(G|F) \neq P(G|\neg F)$. However, it is well

[12] While I will eventually seek a deterministic notion of causal irrelevance, it is nevertheless helpful to first also look at related suggestions, including statistical notions.

known that common causes shield off such correlations—i.e. while P(G|F) ≠ P(G|¬F), we have P(G|F&H) = P(G|¬F&H) when conditionalising on H. Thus, one needs to control for common causes in order to identify the true relations of causal relevance or irrelevance.

For his definition of causal irrelevance, Eells suggests considering the probabilistic impact of a potential cause X on a potential effect Y in various causal background contexts. In each causal background context, all factors $F_1, \ldots, F_n$ that are causally relevant to Y, independently of X[13], are held fixed. Only if the probability of Y is not changed by X in *all possible* contexts, should one speak of causal irrelevance (Eells 1991, 86). This condition is often called *contextual unanimity*. Note that Eells' definition of causal irrelevance is circular to some extent since the definiens itself employs the notion of causal relevance in that it requires all causally relevant factors to be held fixed in causal background contexts. However, he argues that the circularity is not vicious, since the definiens refers to the causal relevance of factors other than X, of which the irrelevance is examined (87). Eells further relativizes causal relevance and irrelevance to "a particular population, as well as to a kind that the token population exemplifies" (87). In part, this is required in order for a probability distribution to exist at all. Certainly, causal and probabilistic dependencies will differ between populations and kinds of populations.

The definition of causal irrelevance (CI) proposed below in Section 4c is in many ways similar to Eells's approach, but it is deterministic and introduces context-dependence in a somewhat different manner. Broadly speaking, context-dependence becomes simpler, since in a deterministic situation the existence of a probability distribution does not have to be ensured. As should be obvious, these changes with respect to Eells's account are necessary for correctly interpreting the role of causal irrelevance in (PA).

In recent years, a link between causality and probabilistic independence has been elaborated in the context of causal modeling on the basis of directed acyclic graphs satisfying the causal Markov condition—such graphs are often referred to as causal Bayes nets. The causal Markov condition implies a range of probabilistic independency relations. In particular, the probabilities for all nodes must be probabilistically independent when conditionalising on all parents PT of the nodes in the graph:

$$\mathrm{P}(X_1, X_2, \ldots, X_n) = \Pi_i\, \mathrm{P}(X_i \mid \mathrm{PT}(X_i))$$

Conditions like faithfulness or minimality further restrict the range of possible causal models. Faithfulness, for example, states that both conditional and unconditional probabilistic independencies in a graph must follow from the causal Markov condition. In particular, if two variables are probabilistically independent there should be no causal link between them.

The faithfulness condition illustrates well the difficulties that arise when building causal models merely from statistical relationships. On the one hand, premises like faithfulness are indispensable to reduce the number of possible models to a manageable amount. On the other hand, a range of counterexamples shows that faithfulness and related conditions can be little

[13] i.e. factors that are causally relevant to Y but to which X is not causally relevant—excluding in particular factors that lie on a causal chain from X to Y.

more than pragmatic and fallible tools to develop causal models. As an example, the faithfulness condition cannot account for causal relationships that exactly cancel each other.

Generally speaking, statistical independence is neither a sufficient nor a necessary criterion for causal irrelevance. As mentioned, when causal influences between two variables exactly cancel each other, there is presumably a causal link between these variables even though they are probabilistically independent. Also, two variables may be probabilistically independent, but in a number of instances of measure zero, there may nevertheless be causal relevance. Such cases show that probabilistic independence is not sufficient for causal irrelevance. But probabilistic independence is not necessary either. After all, as is elaborated in the following, there are methods that determine causal irrelevance in deterministic situations, i.e. in situations in which evidence in terms of probabilistic independence may be entirely absent, for example because the relevant probability distributions are not even defined. Furthermore, probabilistic independence can never be fully established empirically as fluctuations will always lead to some small dependence.

As a second approach, let us take a look at deterministic definitions of causal relevance, i.e. definitions that do not refer to probability distributions. A typical version is given by Christopher Hitchcock:

> "X is *causally relevant* to Y, if and only if there is some set of variables, and some set of values of those variables, such that when we intervene to set all those variables to those values, at least some interventions on the value of X will lead to different values of Y." (2009, 305)

In a similar vein, Michael Baumgartner and Gerd Grasshoff, who advocate a sophisticated regularity view of causation, suggest:

> „Factor A is causally relevant for the occurrence of an effect B, if and only if there exists at least one causal process, in which an event of type A (partly) causes the occurrence of an event of type B." (Baumgartner & Grasshoff 2004, 49; my translation)

While most authors discussing causal relevance do not bother to explicitly define causal irrelevance, it can be construed as complementary to causal relevance. Starting from the above definitions, causal irrelevance would essentially require that *no* intervention can lead to a change in values of the effect variable or that *no* process exists where an event of type A at least partly causes the occurrence of an event of type B.

Such an approach to define causal irrelevance turns out inadequate for an analysis of analogical inferences based on intuition (PA). After all, it may well happen that a circumstance is causally relevant in some situation, while for the considered analogical inference it plays no role, for example because other contributing causal factors are missing or because there is a counteracting cause (cf. the first objection in Section 4a). Thus, circumstances that are causally relevant according to the above definitions may change from source to target, while the analogical inference may still be valid—contradicting (PA).

Indeed, very few circumstances will turn out causally irrelevant according to the above definitions, because in some obscure situation these might all be causally relevant (cp. the fourth objection in Section 4a). For this very reason, Baumgartner and Grasshoff largely reject the notion of causal irrelevance (2004, 212). One main lesson to draw from these attempts to define causal irrelevance is that context-dependence is not taken into account in an adequate manner. Exceptions to causal irrelevance in more or less obscure situations should be discounted on the basis that they occur within a context that differs from the one that is employed in the analogical inference.

David Galles and Judea Pearl belong to the small number of authors, who in an influential article (1997) explicitly define deterministic causal irrelevance and carefully implement context dependence:

> "A variable X is causally irrelevant to Y, given Z […] if, for every set W disjoint of X ∪ Y ∪ Z, we have
>
> $$\forall(u, z, x, x', w), \quad Y_{xzw}(u) = Y_{x'zw}(u)$$
>
> where x and x' are two distinct values of X." (reproduced in Pearl 2000, 235-6)

Here, u are the values of the background or exogenous variables of the model. According to Pearl, this definition captures the intuition that "if X is causally irrelevant to Y, then X cannot affect Y under any circumstance u or under any modification of the model that includes do(Z=z)." (Pearl, 236)

It may be possible to use this definition for an approach to predictive analogies based on (PA). However, this would turn out unnecessarily complicated. The first reason concerns model dependence. Galles and Pearl relativize their definition to a specific causal model that is determined by a number of exogenous or background variables U, a number of endogenous variables, and functions that determine each endogenous variable based on the other variables. The type of background dependence to be sketched in Section 4c is much simpler than such rather sophisticated model dependence. Secondly, by relying on an interventionist account of causation, Galles and Pearl subscribe to a substantial distinction between interventions and observations, which leads them to introduce the do-calculus for formally handling interventions. However, the notion of intervention plays no major role in analogical reasoning, neither in predictive nor in conceptual analogies, which suggests that an interventionist framework might not be the first choice for explicating analogy.[14]

Note finally that while (PA) suggests looking for a deterministic explication of causal irrelevance, this raises the question how to deal with indeterministic contexts and with situations, in which the evidence allows to formulate only probabilistic dependencies—an issue that will be briefly addressed in Section 5.

[14] I have emphasized repeatedly Cartwright's point that causation allows for implementing effective strategies. Note that this does not necessarily imply an interventionist take on causation. Instead, I favor an understanding in terms of difference making. The latter is more general and implies less ontological commitments compared with the interventionist approach.

4c. A necessary and sufficient criterion

Let me in the following sketch an account of causal irrelevance based on difference-making in context. While we cannot ultimately defend the proposed definitions here, they should get some initial plausibility from their close resemblance to the method of difference, which is arguably the most successful rule in scientific method to determine causal dependence.[15]

Causal relevance shall be defined in the following manner (CR):

> A is *causally relevant* to C in a context B, if and only if (i) an instance exists, where A and C occur in the context B, (ii) a second instance exists, where neither A nor C occur in the same context B, and (iii) B guarantees homogeneity.

Note again that this definition largely corresponds to the method of difference as given in particular by John Stuart Mill. Note further that causal relevance of A to C with respect to B implies that a change in A within a context B *always* leads to a change of C—in contrast to the definitions by Hitchcock as well as Baumgartner and Grasshoff given in the previous section. Causal irrelevance can then be defined as the complementary[16] notion (CI):

> A is *causally irrelevant* to C in a context B, if and only if (i) an instance exists, where A and C occur in context B, (ii) a second instance exists, where A does not but C still occurs in the same context B, and (iii) B guarantees homogeneity.

Causal irrelevance of A to C with respect to B implies that a change in A within context B *never* leads to a change in C. For example, a switch is causally irrelevant to a light given two instances, one, in which both switch and light are on, and another, in which the switch is off but the light still on, while nothing else that could be relevant to the light has changed—the last premise essentially corresponding to homogeneity. By contrast, the switch is causally relevant, if in the second instance, the light is off.[17,18]

The context or background B is constituted on the one hand by a set of circumstances that are allowed to change and on the other hand by a set of circumstances that must remain constant. Homogeneity, which was already invoked by Mill in his formulation of the method of difference, essentially captures the intuition that factors in the background that are causally relevant to the examined phenomenon may not change. It is a concept that is used both in counterfactual approaches to causation such as by Rubin and Holland (e.g. Holland 1986) and

---

[15] For a more extensive argument in favor of the proposed framework, compare Pietsch (2016).

[16] As in Eells's approach, there are mixed cases, in which a circumstance is neither relevant nor irrelevant with respect to a given context.

[17] Note that the definition has some seemingly counterintuitive implications. If, for example, a light is controlled by two switches A and A*, where the light is on if at least one of the switches is on, and if it is presupposed as part of the background conditions that A* is on, then A will be classified as causally irrelevant according to the definition (CI). While this sounds counterintuitive, the definitions above are intended as refinements or improvements of our everyday notions in order to make causal language more precise and avoid contradictions. Eventually, these seemingly counterintuitive implications will allow to resolve the first group of problems for (PA) as discussed in Section 4a.

[18] Building on the example of footnote 16, A is causally relevant to C, if it is part of the background conditions that A* is always off, while A is causally irrelevant to C, if it is part of the background conditions that A* is always on. Again, this seeming contradiction only underlines the need to always relativize causal dependencies to a background.

also in sophisticated regularity approaches such as Baumgartner and Grasshoff (2004). The latter provide an extensive discussion (2004, 208).

While homogeneity[19] is usually defined that all causally *relevant* factors must *remain constant*, I prefer the complementary perspective that only causally *irrelevant* circumstances are *allowed to change*. In combination with the definitions discussed in the previous Section 4b, this has some subtle implications. Most importantly, the explication of homogeneity given in the following is less demanding in that more circumstances are allowed to change. Notably, some circumstances, e.g. causal factors that are only active in certain contexts, may be identified as causally irrelevant based on (CI), while they are causally relevant according to conventional definitions, such as those of Baumgartner and Grasshoff. Thus, these would have to remain constant to ensure homogeneity according to Baumgartner and Grasshoff, while they are allowed to change according to the following explication of homogeneity (H):

> Context B guarantees homogeneity with respect to the relationship between A and C, if and only if only circumstances that are causally irrelevant to C can change, (i) except for A and (ii) except for circumstances that are causally relevant to C in virtue of A being causally relevant to C.

The second exception allows for circumstances to change that lie on a causal chain through A to C or that are effects of circumstances that lie on this causal chain.[20] Clearly, the above explication implements the before-mentioned intuition behind the notion of homogeneity that factors in the background B that are causally relevant to the examined phenomenon C may not change.

Let me now briefly address how to deal with the problems that were raised in Section 4a. Concerning the first objection, consider for example cases where two influences A and B exactly cancel each other and therefore the analogical prediction remains valid even though causally relevant circumstances change. In response, let me specify that for valid analogical inferences it is not required that every property in the negative analogy *taken by itself* must be causally irrelevant, but strictly speaking only all properties in the negative analogy *taken in conjunction*. If A and B exactly compensate each other, then a change from $A \wedge B$ to $\neg A \wedge \neg B$ is irrelevant. Similarly, if A and B are alternative causes for a phenomenon C, then a change from $A \wedge \neg B$ to $\neg A \wedge B$ is irrelevant for the phenomenon C. In the first case, A and B taken in conjunction are causally irrelevant to C, as are ¬A and ¬B. In the second case, A and ¬B taken in conjunction are causally irrelevant to C, as are ¬A and B.

In general, causal irrelevance of circumstances taken in conjunction can be defined in the following way (CI'):

> A number of factors $A_1, A_2, \ldots, A_N$ *taken in conjunction* is causally irrelevant to a hypothetical analogy C with respect to a context B, if and only if (i) an instance exists,

[19] Homogeneity broadly corresponds to context-unanimity in Eells' account.

[20] The notion of 'causal relevance in virtue of' cannot be discussed here in further detail due to lack of space. An exact explication is: "A condition X is causally relevant to C in virtue of A being causally relevant to C with respect to a background B, iff in all contexts within B, in which X is causally relevant to C, A is causally relevant to C as well (but not necessarily vice versa)."

where $A_1, A_2, \ldots, A_N$ and C occur in context B, (ii) a second instance exists, where $\neg A_1, \neg A_2, \ldots, \neg A_N$ and C occur in the same context B, and (iii) B guarantees homogeneity.

Thus, in order to determine causal irrelevance of a number of factors taken in conjunction, one does not need to test the causal irrelevance of each factor individually or the potentially huge number of all possible combinations of those factors. Instead, it suffices to establish the two situations mentioned above.[21]

Note further that if a factor which is commonly considered a cause fails to be relevant for the considered analogy because other contributing factors are not instantiated, e.g. the burning match does not cause a fire since there is no combustible material present, such a factor is identified as causally *irrelevant* with respect to the considered context according to the proposed definition (CI)—in contrast to all other definitions of causal irrelevance discussed in the previous section. Therefore, only the proposed definition of causal irrelevance (CI) in combination with the intuition (PA) correctly classifies such analogical inferences as valid. As an example, one might conclude from a barn without fire that another barn is not on fire either notwithstanding the presence of a burning match, just because combustible material is absent in the second barn.

The second problem raised in Section 4a concerned analogies based on correlations. Consider again the example that a circumstance F changes from source to target, which is causally irrelevant to the hypothetical analogy G, but which is correlated with it and therefore leads to the failure of the analogical inference. It turns out that such situations are precluded in the sketched approach. Indeed, according to the view of causation introduced in Section 3, any meaningful correlation between variables *must* result from a common cause, i.e. any correlation that leads to a *reliable* prediction. Therefore, in order for an analogical inference to fail in the described manner, the corresponding common cause variable must change. However, such a change in common cause variables is precluded by (PA), since these are not causally irrelevant to the hypothetical analogy.

In response to the problem that analogies may be based on definitional, instead of causal relevance, one might be tempted to restrict predictive analogies to causal vertical relationships only. However, this runs into problems with familiar epistemological issues such as confirmational holism and relatedly the lack of a clear distinction between empirical and definitional statements. Instead, I broadly suggest to integrate definitional relevance in the framework which should be rather easily done since definitional relevance can be defined in much the same manner as causal relevance—given that the main difference merely lies in the nature of the necessity between antecedent and consequent.[22]

---

[21] Requiring irrelevance for all possible combinations of variables or for each variable individually would be much too strong such that (PA') as stated below would not be a necessary criterion. For example, analogical inferences, in which two causally relevant factors exactly cancel each other, would be wrongly identified as invalid.

[22] Basically, one needs to replace in (PA), (CI), (CR), and (H) "causally irrelevant" with "causally and definitionally irrelevant" as well as "causally relevant" with "causally or definitionally relevant". Note that for predictive analogies, all changes in the definitions of relevant terms must be known in advance in order to clearly

With respect to the two other issues that were raised in Section 4a, the distinction between properties and relations was already discussed. Concerning the notion of irrelevance, (CI) in combination with (H) is supposed to yield an adequate explication. In particular, by introducing strict context-dependence, (CI) aims to avoid the problem that was pointed out in Section 4a, namely that causal irrelevance is practically inexistent, since any circumstance can be relevant in some obscure situation.

But a crucial question remains, namely what exactly should be chosen as an adequate context for the statement of irrelevance in the basic intuition (PA). Remember that a context consists of circumstances that are allowed to change and others that must remain constant. Since the impact of all circumstances that change (i.e. the negative analogy) is explicitly considered as antecedent, these cannot be ascribed to the context. What is left to account for are thus all circumstances that remain constant, i.e. the positive analogy. Apparently, these then constitute the context.

In summary, the proposed deterministic approach to predictive analogical inferences is given by the following explication (PA') in combination with the definitions of causal irrelevance (CI) and homogeneity (H):

> Predictive analogical inferences from a source instance to a target instance are valid, if and only if the negative analogy (taken in conjunction) is causally (and definitionally) irrelevant to the hypothetical analogy with respect to a context constituted by the constancy of the positive analogy.

It is important to emphasize that in (PA') the negative analogy refers to the *complete* negative analogy, i.e. comprises all circumstances that differ between the source instance and the target instance. Similarly, the positive analogy refers to the *complete* positive analogy, i.e. all circumstances that are the same for source and target instance. In both cases, the considered circumstances thus include known as well as unknown circumstances. Furthermore, it is reasonable to restrict the considered circumstances to those in the past as well as in the present of the event denoted by the hypothetical analogy. Sub specie aeternitatis, one may need to also take into account circumstances in the future, but a discussion of this question is far beyond the scope of this paper.

Let me give an example to illustrate how (PA') is applied. Regarding the question, whether the existence of life on Earth allows inferring the existence of life on Mars, one would start from two instances or situations, the first (1) on Earth and the second (2) on Mars. Let us assume that it is known: (a) that the hypothetical analogy Q is true for (1), i.e. that life indeed exists on Earth; (b) what the positive and what the negative analogy between both situations is; (c) that the negative analogy taken in conjunction is irrelevant to the hypothetical analogy with respect to a background constituted by the positive analogy. Then, it can be concluded that the hypothetical analogy is true for situation (2), i.e. that life exists on Mars.

In a very simple illustration, it may be known that Earth and Mars differ only in temperature, which thus constitutes the negative analogy. Temperature is also the negative analogy taken in

---

distinguish predictive from conceptual analogies according to the criterion that only for conceptual analogies one is prepared to engage in conceptual work.

conjunction, since there is only one variable (cp. the definition in section 4b). All other variables constitute the positive analogy, by assumption these then have the same value for Earth and for Mars. Furthermore, one knows from some experiment relying on the method of difference that temperature is irrelevant to the existence of life, with respect to a context constituted by the positive analogy. For example, an ingenious scientist may succeed in briefly lowering the temperature on Earth to the temperature that is typically found on Mars, while at least some kind of life survives the temperature change. Carrying out the experiment on Earth ensures that the context is the same as in the examined analogical inference. Under these conditions, one can conclude from the existence of life on Earth to the existence of life on Mars.

4d. Conceptual derivation

(PA’) can be reformulated in the following manner (PA’’):

> Given (i) a source instance P & N_1 & … & N_k, of which it is known that C is the case, and (ii) a target instance P & ¬N_1 & … & ¬N_k, of which it is not known whether C is the case, where P denotes the positive analogy, N_1, …, N_k the negative analogy and C the hypothetical analogy, the following holds:
>
> if and only if (iii) N_1 & … & N_k taken in conjunction is causally and definitionally irrelevant to C with respect to context P,
>
> then the analogical inference that C is the case for the target instance holds.

Note that for the sake of clarity and without loss of generality, all circumstances in the negative analogy are formulated in a positive way for the source instance and in a negative way for the target instance.

A proof of (PA’’) proceeds as follows:

(I) To show that premise (iii) is a sufficient criterion, assume that (iii) is true.

- ⇨ It follows from the definition (CI’) in combination with premise (i) of (PA’) that the following instances exist: (1) P & N_1 & … & N_k & C and (2) P & ¬N_1 & … & ¬N_k & C. Note that so far, instance (2) need not coincide with the target instance.
- ⇨ By definition, P denotes the *complete* positive analogy and N_1 & … & N_k denotes the *complete* negative analogy with respect to the hypothetical analogy C. Therefore, in a deterministic setting, for which (PA’’) is intended, the circumstances P & ¬N_1 & … & ¬N_k must uniquely determine, whether C is the case or not.
- ⇨ Therefore, since in the target instance the state of the circumstances P & ¬N_1 & … & ¬N_k is the same as in instance (2) and it is known that C is the case in instance (2), it follows that for the target instance the hypothetical analogy C must also be the case.
- ⇨ Thus, the analogical inference is correct.

(II) To show that premise (iii) is a necessary criterion, assume that (iii) is false.

- ⇨ It follows that one of the premises in (CI') must be false, i.e. one of the premises (a) that an instance P & N_1 & … & N_k & C exists or (b) that an instance P & ¬N_1 & … & ¬N_k & C exists or (c) that the context P guarantees homogeneity.
- ⇨ By premise (i) of (PA''), an instance P & N_1 & … & N_k & C exists. Also, homogeneity is trivially fulfilled, because the context consists only of circumstances P, which by assumption all remain constant. Thus, from premise (iii) of (PA'') being false follows that the remaining premise (b) of (CI') must be false, i.e. there may not be an instance P & ¬N_1 & … & ¬N_k & C.
- ⇨ According to premise (ii) of (PA''), there exists a target instance with the circumstances P & ¬N_1 & … & ¬N_k. Furthermore, as explained under (I) above, the state of the circumstances P & ¬N_1 & … & ¬N_k must uniquely determine whether C is the case or not. Since an instance P & ¬N_1 & … & ¬N_k & C may not exist, ¬C must be the case for the target instance.
- ⇨ Thus, the analogical inference is false.

From (I) and (II) follows (PA'') and therefore (PA').

4e. Applicability

In the following two sections, several points of criticisms with respect to the proposal of the previous section are discussed. In this first section, various issues concerning the applicability of (PA') are addressed. For example, one might doubt whether it is always possible to identify the positive and the negative analogy, even if sufficient evidence is available. In particular, one could object that analogical inferences are actually based on similarity, while such similarity cannot necessarily be spelled out in terms of a constant positive analogy and changes in a negative analogy.

As an example, consider again the question, whether the existence of life on Mars can be inferred from the existence of life on Earth. For instance, both Earth and Mars have an atmosphere, but it is not at all clear, whether having an atmosphere belongs to the positive or the negative analogy. After all, both planets have an atmosphere, but it also differs in important respects. Is it only possible to state a similarity between Earth and Mars with respect to having an atmosphere, while this similarity cannot be made explicit in terms of differences and conformities? In the end, the problem is that the chosen description is inadequate for the question at hand, because it is too coarse. Using sufficiently detailed terminology can resolve the issue. For example, the positive analogy might be that both planets have an atmosphere containing oxygen and carbon dioxide, while the negative analogy is that the concentrations of these components differ and that the atmosphere on Mars or Earth may contain traces of other gases, which are not present in the atmosphere of the other planet.

While differentiating the negative and the positive analogy will often be challenging, I do not see any reason, why it should not be possible in principle, if sufficient evidence is available. The following procedure corroborates this claim. Start with a sufficiently detailed description of the first phenomenon. All characteristics that need to be changed, taken away or added in order to arrive at a sufficiently detailed description of the second phenomenon constitute the negative analogy, while all properties that remain the same constitute the positive analogy. In summary, it is an implicit premise of the proposed approach that any statement of similarity can be broken down into a positive and a negative analogy, but there are reasons to believe this should always be possible, at least in principle.

One might also worry about *practical* applicability. Even if it is possible to always clearly differentiate the positive and negative analogies in principle, it may still be the case that in real-world situations there is rarely sufficient evidence to apply (PA'). As a first remark, let me point out that in practice we often have quite reliable intuitions that a large range of circumstances are irrelevant and which circumstances may potentially be relevant. To further assess the issue at hand, let me also recall the main insight drawn from the discussion of Keynes' approach in section 2b, which was that all inductive inferences are to some extent based on analogy.

Consider for example a physicist, who examines two pendulums in order to determine whether their periods are the same, i.e. the time for a complete cycle. For this purpose, the physicist consults the relevant formula, which tells her that at least for small amplitudes, the period depends only on the length of the pendulum and on the acceleration of gravity at the location of the pendulum. In other words, the formula tells the physicist which variables are relevant, so that she can examine whether those relevant variables have the same value for both pendulums (or whether their deviations mutually compensate each other). If the lengths of the pendulums are the same and they are located in places with the same acceleration of gravity, then one can infer by analogy relying on (PA') from the period of one pendulum that the period of the other pendulum must be the same.

Whenever an inference is made for a specific situation based on a phenomenological scientific law that is empirically well established, this inference can be construed according to the above pattern as an analogical inference with respect to some of the evidence that was used to establish that law. In particular, the relevant experiments and observations, which were used to establish the law, allow to determine which variables are causally relevant to the considered phenomenon, usually by some application of the method of difference or the method of concomitant variation. Also, the evidence will include instances for which the hypothetical analogy is identical or at least very similar compared with the specific instance, which is predicted. In this way, the application of a phenomenological law can be interpreted as an implicit analogical inference according to (PA') with respect to such instances in the relevant evidence. Note further that analogical inferences can be considered on different levels of coarse-graining. In the above-mentioned example, one might infer the explicit value of the period, merely a value range to which the period belongs or even only the fact that pendulums have a constant period for subsequent oscillations. This shows that applications of (PA') are quite wide-spread.

As a further worry regarding applicability, many inferences that are considered typical analogical inferences do not seem to be covered by (PA'). In particular, (PA') does not guarantee that C holds in just any situation where P holds, as should be obvious at least from the equivalent formulation (PA'') given above. For example, for a successful predictive analogical inference according to (PA''), two instances must exist: a source instance (1) P &

N_1 & … & N_k & C and a target instance (2) P & ¬N_1 & … & ¬N_k & C. However, (PA’’) says nothing whether C is the case for instances (3) for which P holds and at least one of the N_1, …, N_k is true, while at least one of the N_1, …, N_k is false. However, inferences from instances (1) and (2) to instances (3) seem to be typical analogical inferences.

At first, this appears to render (PA’’) and thus (PA’) practically useless. However, (PA’’) can be applied in the above-mentioned evidence situation if additional assumptions are made, in particular if one can somehow reduce the number of circumstances of which the irrelevance in conjunction needs to be determined, instead of looking at the complete negative analogy. For example, it suffices to look at a smaller number of circumstances, if it can somehow be established that this smaller number of circumstances fully determines the truth or falsity of C, irrespective of whether the other circumstances are true or false in all those combinations that are possible in a given context.

Based on this insight, (PA’’’) can be formulated as a further version which is more directly applicable to practical examples, but which provides only a sufficient criterion for a predictive analogical inference to hold:

> Given (i) a source instance P & N_1 & … & N_k, of which it is known that C is the case, and (ii) target instances P & ¬N_1 & … & ¬N_m with $m < k$, of which it is not known whether C is the case and for which the remaining circumstances N with indices m+1, …, k can take on arbitrary combinations of values within a given range, the following holds:
>
> if (iii) N_1 & … & N_m taken in conjunction is causally and definitionally irrelevant to C with respect to a context determined by the constancy of P and by the circumstances N_{m+1}, …, N_k being allowed to take on any combination of values within the given range,
>
> then the analogical inference that C is the case for the target instances holds.

Obviously, for applying (PA’’’), it suffices to know the state of the circumstances P & N_1 & … & N_m for the source instance, while with respect to the other circumstances it must only be guaranteed that they are within the given range. Strictly speaking, P is not the complete positive analogy and circumstances N_1 & … & N_m are not the complete negative analogy. After all, when specific source and target instances are considered, the circumstances N_{m+1}, …, N_k can but need not vary and may thus belong either to the negative or to the positive analogy. More exactly, the positive analogy between a specific source and a specific target instance consists of P and the respective circumstances N with indices m+1, …, k, which remain constant between both instances, and the negative analogy consists of the circumstances N_1, …, N_k as well as those circumstances N with indices m+1, …, k which change.

An equivalent formulation of (PA’’’) somewhat similar to (PA’) reads as follows:

> An analogical inference from a source instance to a range of target instances holds, if a first part of all circumstances taken in conjunction is causally and definitionally irrelevant to the hypothetical analogy with respect to a context determined by the

constancy of a second part of all circumstances and possible variations of the remaining circumstances within a given range.

Typically, (PA''') can be employed, if the fraction N_{m+1}, …, N_k of all circumstances is causally related to C only by means of the circumstances N_1, …, N_m for the given range of possible variations of N_{m+1}, …, N_k. This could be the case for example in the following situations:

i) the circumstances N_1, …, N_m and C may be causally unrelated to the circumstances N_{m+1}, …, N_k for the given range of possible variations, i.e. colloquially speaking the two groups of circumstances belong to different 'patches' of the world;

ii) the circumstances N_{m+1}, …, N_k may be causally related to C only by means of common causes of those circumstances and C, where the common causes can be expressed in terms of the circumstances N_1, …, N_m;

iii) the circumstances N_{m+1}, …, N_k may lie on causal chains from the circumstances N_1, …, N_m or combinations thereof to C, i.e. they may either be mediating circumstances between the circumstances N_1, …, N_m and C or they may be causes of C, which act on C only by mediating circumstances N_1, …, N_m. Essentially, circumstances on causal chains co-vary and therefore do not have to be considered individually.

Of course, for most applications, combinations of the above situations will be the case. While these situations typically cannot be established with certainty, it is straightforward to show that increasing variational evidence of the type championed by Bacon, Mill or Keynes can continuously improve the reliability of such assumptions. Assumptions of the above type are wide-spread in the sciences and epistemology, e.g. various locality assumptions familiar from physics or from philosophical debates on causation.

Obviously, (PA''') solves the problem, how to approach those analogical inferences, in which one infers from evidence in terms of instances such as (1) P & N_1 & … & N_k & C and (2) P & ¬N_1 & … & ¬N_k & C to other instances (3) for which P holds and at least one of the N_1, …, N_k is true, while at least one of the N_1, …, N_k is false. But note that from a logical perspective (PA'') and thus also (PA') remain necessary and sufficient criteria for predictive analogical inferences to hold.

To illustrate (PA''') with an example, let us return once more to the life on Mars analogy. Suppose in a still very artificial version of the Mars analogy that two factors are known to differ between Earth and Mars, namely warm temperature (N_1) and the presence of two small moons (N_2). For Earth, we have N_1 & ¬N_2 and life exists (C). In addition, an exoplanet Alpha is known with N_1 & N_2 and where life exists as well as a further exoplanet Beta with ¬N_1 & ¬N_2 and where life also exists. How can an analogical inference to life on Mars with ¬N_1 & N_2 be established based on this evidence?[23]

[23] I am grateful to one of the referees for this example, which I have quite shamelessly copied almost verbatim from the report.

Clearly, neither (PA’) nor (PA’’’) provide an adequate basis for such an analogical inference without any further assumptions. However, if homogeneity of context can be established for all four instances, i.e. essentially that all other circumstances that may vary between the instances are causally related to C only via N_1 and N_2, and if some additional assumption about the circumstances N_1 and N_2 can be established, then the requirements of (PA’’’) can be met. For example, it may be possible to show that N_1 and C are causally unrelated to N_2, i.e. these two groups of variables belong to different ‘patches’ of the world, e.g. because other causal evidence proves the irrelevance of the number of moons for the existence of life. Or it may be possible to show that N_1 and N_2 act only independently from each other on C, if they are causally related to C at all. Under this latter assumption, e.g. the instantial evidence of Earth and Alpha proves the irrelevance of N_2 for C and the instantial evidence of Beta can then be used to analogically infer the existence of life on Mars relying on (PA’’’).

In summary, while from a logical point of view (PA’) constitutes a necessary and sufficient condition for predictive analogical inferences, the application of (PA’) to actual phenomena generally requires an adequate modeling of the phenomena as well as a host of further assumptions, which can be corroborated by variational evidence but which in principle always remain fallible.

4f. Further discussion

A number of further objections that could be raised against (PA’) are addressed in the following. A first worry concerns the epistemological status of the complete negative analogy (and similarly of the complete positive analogy). Since it is not plausible that all circumstances which differ between two instances can ever be fully known, (PA’) may appear to be merely a *metaphysical* rule with little practical import.

However, (PA’) including the notion of a complete negative analogy is intended primarily as a *logical* rule. Without committing to any particular view on the nature of logic, several similarities between (PA’) and other logical inference rules can be pointed out. All of these may be debatable and would actually deserve a much more detailed discussion, for which unfortunately there is not enough space. One crucial characteristic underpinning the logical nature of (PA’) is that (PA’) can be considered as truth-preserving. As with other logical rules, when the conclusion of (PA’) fails to be true, one can always put the blame on one of the assumptions of (PA’) being false rather than putting the blame on (PA’) itself. Note that truth preservation is generally held to be a characteristic of deductive logic, while analogical inferences form part of inductive logic. Inductive rules are often thought to allow only for relationships between assumptions and conclusion that are somewhat weaker than truth-preserving.

As a further point, logical concepts typically are not directly applicable to the phenomena as has been discussed in much detail for (PA’) in the previous section. Instead, the phenomena first have to be modeled in adequate ways. This is largely due to the abstractness of most logical concepts, which in turn can be seen to enable their universality. However, if the phenomena are adequately modeled, abstract logical concepts can become immensely useful. As an example, syllogistic inferences are among the most potent inferences in deductive logic, even though it remains unclear whether there are any true universal statements in many of those fields, where syllogistic inferences are successfully employed. Who knows, whether all men are indeed mortal. Similarly, (PA') refers to all different circumstances between two instances. Whether these can ever be completely known is at best doubtful. Thus, as was pointed out in the previous section, (PA') requires some further modeling assumptions, which reduce the number of differences to a manageable amount.

Stressing the logical nature of (PA') may somewhat alleviate the worry that the concept of a complete negative analogy is too open-ended. After all, this worry relates to the differences between actual instances in the world, while (PA') is primarily a rule in an abstract conceptual framework that aims for consistency and precision. In a way, (PA') may be considered as a vanishing or limiting point, which can never be fully reached but which can be approximated further and further by collecting the appropriate kind of evidence.

A second issue that is somewhat related to questions of applicability arises regarding the notion of causal irrelevance (CI). It may well be that strictly speaking, no circumstance fully satisfies (CI) and thus that no circumstance is truly irrelevant to a phenomenon.[24] Specifically, there might always be some however faint causal influence, by which the circumstance and the phenomenon are connected. It thus appears that (PA') can never be fulfilled, because the negative analogy is never entirely irrelevant to the phenomenon.

The conclusion to draw from this objection is that causal irrelevance is a contextual notion, which depends on the amount of coarse-graining that is assumed for the relevant variables. In the pendulum example of the previous section, an inference that the periods of two pendulums are *exactly* the same does not make much sense, since in the empirical sciences, any value can only be determined up to some degree of accuracy. While the negative analogy will presumably always be causally relevant to some extent, the crucial question is, whether the effect of this causal relevance is larger than the amount of coarse-graining that is assumed for the variables in the hypothetical analogy. If the deviation is small enough *for the purpose at hand*, the variables in questions must be considered causally irrelevant *for that purpose at hand*. Consequently, (PA') can still be considered to hold.

A third issue which is also related to the notion of causal irrelevance is that the definitions (CI') and (H) suffer from a circularity in that (CI') presupposes (H) and (H) presupposes (CI'). However, this objection can be overcome in a similar manner as proposed by Eells' for an analogous circularity arising in his account (cp. Section 4b). In particular, definition (H) refers to the causal irrelevance of other circumstances, namely circumstances of the context, than the circumstances, whose causal irrelevance is explicitly examined in (CI').

As a fourth criticism, many analogical inferences do not follow the rationale of (PA') by examining the causal irrelevance of the negative analogy, but are based on other kinds of evidence, for example on correlations, on statistical laws, on retrodictions, etc. Such analogical inferences often enough turn out valid and there are in many cases good reasons to be fairly confident in them. One may want to call *strong predictive analogical inferences* those that are based on evidence in terms of the causal irrelevance of the negative analogy,

[24] I am grateful to one of the referees for bringing up this issue.

while *weak predictive analogical inferences* are based on other types of evidence.[25] Strong analogical inferences are certain if sufficient evidence is available, while weak analogical inferences mostly convey only a degree of probability or even only plausibility. Strong analogical inferences are based on causal relationships and therefore imply knowledge about interventions, while this is not necessarily the case for weak analogical inferences.

From a fundamental epistemological perspective, any weak analogical inference, which proves reliable, must be based on a corresponding strong analogy, which underlies the weak inference but is at least partly unknown to the person drawing the inference. Thus, all types of evidence mentioned above in connection with weak analogical inferences are only useful to the extent that they indicate the existence of an underlying strong analogy. For example, an analogical inference based on correlations is only reliable, if the correlated variables are adequate proxies for underlying causal variables, e.g. in terms of common causes. In retrodictions, the future variables should be suitable proxies for past causal variables, which may for example be connected to the future variables by means of deterministic laws.

Finally, not all relevant variables may be known such that one has to rely on statistical causal relationships. The analogical inference will then be valid only with a certain probability. This case, which has considerable practical significance, will be briefly addressed in the next section 5.

According to a fifth point of criticism, (PA') presupposes determinism with respect to the variables of the hypothetical analogy, i.e. that those variables are fully causally determined by their respective circumstances. In the case of indeterminism, (PA') may be fulfilled, but the analogical inference may nevertheless fail to hold because pure chance interferes. One could of course alter (PA') such that it also accounts for indeterministic cases. Essentially, in indeterministic situations the analogical inference holds only *with the corresponding objective probability*. Furthermore, the status of the indeterministic hypothesis, i.e. the claim that the world is to some extent indeterministic, remains uncertain even with respect to microphysics, as supposedly deterministic interpretations of quantum mechanics show, in particular Bohmian mechanics. And even if the micro-realm is indeterministic, many macro phenomena are deterministic, e.g. those treated by much of classical physics and engineering, and thus (PA') in the deterministic version of the previous section applies to these phenomena.

A sixth issue concerns the so-called problem of induction. According to a wide-spread consensus in epistemology and philosophy of science, inductive inferences are always fallible. We can never know for certain, whether an empirical prediction based on an inductive inference will turn out true. Some tension seems to exist between this insight and the claim of the previous section that (PA') constitutes a necessary and sufficient criterion for the truth of predictive analogical inferences.

In this context, two assumptions should be distinguished. The first regards whether a consistent logic of induction exists. The second assumption regards whether we can ever be completely certain that the premises for a valid inductive inference are fulfilled, even if a

[25] I am grateful to one of the referees for suggesting this helpful distinction.

consistent inductive logic is available. The general fallibility of inductive inferences may result from a failure of either the first or the second assumption.

While many scholars doubt the existence of an inductive logic, the first assumption is much more controversial than the more general point that inductive inferences are in principle fallible. Certainly, influential scholars have in the past attempted to formulate a consistent inductive logic and seem to have believed that such a program is feasible at least in principle, e.g. Carnap and Keynes (cp. Section 2a and 2b). Equally, in the essay at hand, I start from the assumption that eliminative induction can provide a consistent framework of inductive logic, into which (PA') can be embedded.

It is the failure of the second assumption that leads to the general fallibility of inductive inferences. With respect to predictive analogical inferences, it turns out impossible to fully verify empirically, whether all premises required for (PA') are fulfilled. Most importantly, while we often have strong intuitions which circumstances may be relevant, it is impossible to know with absolute certainty. After all, among the myriads of circumstances that change between two instances, there may always be some circumstance, however far-fetched and remote it may appear, that has not yet been taken into account, but which eventually might turn out relevant. Thus, judgments of causal irrelevance are always fallible and consequently inferences based on (PA').

5. Analogy and probability

Thus far, we have only addressed deterministic analogical inferences that hold with certainty. Of course, analogical inferences are often only probabilistic: Given a certain known positive, known negative and unknown analogy, what is the probability that the hypothetical analogy is the case for the target phenomenon?

An extension of the approach delineated in Section 4c to cover such probabilistic inferences is straightforward. Let me briefly discuss the most important cases: (i) first, there may be an unknown analogy, which is causally relevant; (ii) second, there may be a negative analogy, which is causally relevant only with a certain probability; (iii) there may be situations of indeterminism.

Regarding the first case, we have thus far only considered ideal situations, in which every circumstance is known to belong either to the positive or to the negative analogy. Now, as Keynes has rightly pointed out, in actual situations it is usually unknown of a number of circumstances whether they belong to the positive or negative analogy. Assume for the sake of simplicity that the unknown analogy consists of only a single circumstance which is causally relevant in the respective context. Then, the analogical inference is valid with the probability that this circumstance belongs to the positive rather than the negative analogy, if otherwise (PA') holds. Equally, if there is more than one factor in the unknown analogy, one has to determine the combinations which are causally relevant and then add up the respective probabilities belonging to those combinations.

In the second case, one may be uncertain whether circumstances that belong to the negative analogy are causally irrelevant. Thus, the analogical inference is valid with the probability that these circumstances taken in conjunction are causally (and definitionally) irrelevant. Finally, in cases of indeterminism, i.e. in cases where the circumstances determine the hypothetical analogy only up to a certain probability, analogical inferences are valid with that probability. Of course, in such situations, causal relevance has to be interpreted in a probabilistic manner determining a probability distribution over states and not the state itself.

These are the principal cases, how probabilities may enter the assessment of analogical inferences. Of course, various combinations are possible, for example there may be an unknown analogy of which it is unknown whether it is causally relevant. While one has to carefully keep track of the corresponding probabilities, these complications do not add any substantial conceptual problems.

At this point, one may worry again about applicability. But in line with the Keynesian insight that analogical reasoning is essential not only for deterministic, but in particular also for probabilistic inferences, it turns out that the framework delineated above underlies many forms of inductive reasoning in statistics and probability theory. For instance, whenever one infers from a representative sample to a new individual, the above framework is employed. Let us assume that the probability to get lung cancer is 30% in a representative sample of a given population, e.g. all smokers. From this, one concludes that a certain further individual, who belongs to the same population of smokers, will get lung cancer with a probability of 30%. Essentially, the population is defined by the extent that certain factors are allowed to vary while others remain constant, e.g. genetic factors, living conditions, habits etc. These factors determine the positive and negative analogy between the various individuals of the population. Of these factors, it is often not known, whether they are present or absent in specific individuals and/or whether they are causally relevant to smoking. This is the reason why we rely on representative samples in the first place. Now, the 30% probability basically expresses that

- in average, combinations of factors which are relevant to lung cancer are present with a probability of 30 % in each individual of the population or
- in average, combinations of factors which are present in the population are relevant to lung cancer with a probability of 30%.

Basically, these correspond to the cases (i) and (ii) as introduced above. Thus, an inference to the probability for a new individual relies on the discussed framework, if we know that the individual belongs to the same population as a representative sample, for which the respective probability is known. Note finally that such probabilistic inferences are *predictive* analogical inferences as long as the person making them is not prepared to implement definitional changes.

One crucial question, however, which goes far beyond the present article, concerns the interpretation of probability in such probabilistic analogical inferences (cp. Pietsch 2015). On the one hand, the interpretation presumably needs to be objective as prediction and intervention concern matters of facts rather than subjective credence. On the other hand, the most common objective interpretation, the frequentist approach, is not an adequate interpretation since it belongs to the tradition of what was called above enumerative approaches to induction, which are generally hostile to analogical reasoning.

Acknowledgments

I am grateful to two anonymous referees for extremely helpful comments and suggestions.

References

Bartha, Paul. 2010. *By parallel reasoning: The construction and evaluation of analogical arguments*. New York: Oxford University Press.

Bartha, Paul. 2013. "Analogy and analogical reasoning." *Stanford Encyclopedia of Philosophy (Fall 2013 Edition).* http://plato.stanford.edu/archives/fall2013/entries/reasoning-analogy/

Baumgartner, Michael and Gerd Graßhoff. 2004. *Kausalität und kausales Schließen.* Norderstedt: Books on Demand.

Carnap, R. 1950. *Logical Foundations of Probability.* Chicago, Il: University of Chicago Press.

Carnap, R. 1971. "A Basic System of Inductive Logic Part I." In Richard Jeffrey & Rudolf Carnap (eds.), *Studies in Inductive Logic and Probability*, 34-165. Berkeley, CA: University of California Press.

Carnap, R. 1980. "A Basic System of Inductive Logic Part II." In Richard Jeffrey & Rudolf Carnap (eds.), *Studies in Inductive Logic and Probability*, Vol. 2, 7–155. Berkeley, CA: University of California Press.

Cartwright, Nancy. 1979. "Causal Laws and Effective Strategies." *Noûs* 13(4): 419-437.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie.* Oxford: Oxford University Press.

Duhem, Pierre. 1954. *The Aim and Structure of Physical Theory.* Princeton, MA: Princeton University Press.

Eells, Ellery. 1991. *Probabilistic Causality*. Cambridge, UK: Cambridge University Press.

Galles, David, and Judea Pearl. 1997. "Axioms of Causal Relevance." *Artificial Intelligence* 97: 9- 43.

Gentner, Dedre. 1983. "Structure-Mapping: A Theoretical Framework for Analogy." *Cognitive Science* 7: 155–70.

Gentner, Dedre, Sarah Brem, Ron Ferguson, Phillip Wolff, Arthur B. Markman, and Kenneth D. Forbus. 1997. „Analogy and creativity in the works of Johannes Kepler." In Thomas B. Ward, Steven M. Smith, and Jyotsna Vaid (eds.), *Creative thought: An investigation of conceptual structures and processes.* Washington, D.C.: American Psychological Association.

Hesse, Mary B. 1964. "Analogy and Confirmation Theory." *Philosophy of Science* 31:319–327.

Hesse, Mary. 1966. *Models and analogies in science.* South Bend, Il: Notre Dame University Press.

Hitchcock, Christopher. 2009. “Causal Modelling.” In Helen Beebee, Christopher Hitchcock, and Peter Menzies (eds.), *The Oxford Handbook of Causation*, 299-314.

Hitchcock, Christopher. 2016. “Probabilistic Causation.” *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). https://plato.stanford.edu/archives/win2016/entries/causation-probabilistic/

Holland, Paul W. 1986. “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81(396): 945-960

Huttegger, Simon. Forthcoming. “Analogical predictive probabilities.” *Mind*. Preprint: http://faculty.sites.uci.edu/shuttegg/files/2011/03/AnalogicalPredictionJuly2016.pdf

Kepler, J. 1596/1981. *Mysterium cosmographicum* I, II (A. M. Duncan, Transl.). (2nd ed.). New York: Abaris Books.

Keynes, John M. 1921. *A Treatise on Probability*. London: Macmillan.

Kuipers, T. 1984. “Two types of inductive analogy by similarity.” *Erkenntnis* 21:63-87.

Mackie, John L. 1980. *The Cement of the Universe*. Oxford: Oxford University Press.

Maher, P. 2001. “Probabilities for multiple properties: The models of Hesse and Carnap and Kemeny.” *Erkenntnis* 55:183-216.

Maxwell, James Clerk. 1855/56. “On Faraday’s Lines of Force.” *Transactions of the Cambridge Philosophical Society* X.1: 155-229.

Norton, John. 2011. “Analogy.” Draft chapter of a book on inductive inference. http://www.pitt.edu/~jdnorton/papers/material_theory/Analogy.pdf

Nyrup, R. (2016). “A pursuit worthiness account of analogies in science.” http://philsci-archive.pitt.edu/12577/

Pearl, Judea. 2000. *Causality*. New York: Cambridge University Press.

Pietsch, Wolfgang. 2012. “Hidden Underdetermination: A Case Study in Classical Electrodynamics.” *International Studies in the Philosophy of Science* 26(2): 125-151.

Pietsch, W. (2015). “Causal interpretations of probability.” http://philsci-archive.pitt.edu/11315/

Pietsch, Wolfgang. 2016. “A Difference-Making Account of Causation.” http://philsci-archive.pitt.edu/11913/

Polya, G. 1954. *Induction and Analogy in Mathematics*. Princeton, NJ: Princeton University Press.

Prade, Henri and Gilles Richard. 2014. *Computational Approaches to Analogical Reasoning: Current Trends.* Heidelberg: Springer.

Reibe, Neil and Friedrich Steinle. 2002. “Exploratory Experimentation: Goethe, Land, and Color Theory.” *Physics Today* 55(7):43-49.

Reichenbach, Hans. 1956. *The Direction of Time*. Berkeley, CA: University of Los Angeles Press.

Romeijn, Jan-Willem. 2006. “Analogical Predictions for Explicit Simmilarity.” *Erkenntnis* 64(2):253-280.

Salmon, Wesley. 1990. “Rationality and objectivity in science or Tom Kuhn meets Tom Bayes.” In C. Wade Savage (ed.), *Scientific Theories*, 175. University of Minnesota Press.

Sober, Elliott. 1988. “The Principle of the Common Cause.” In James Fetzer (ed.), *Probability and Causality*, 211-229. Dordrecht: Reidel.

Sober, Elliott. 2001. “Venetian Sea Levels, British Bread Prices, and the Principle of the Common Cause.” *British Journal for the Philosophy of Science* 52:331-346.

Stegmüller, Wolfgang. 1973. *Carnap II: Normative Theorie des induktiven Räsonierens.* Berlin: Springer.