A difference-making account of causation[1]

Wolfgang Pietsch[2], Munich Center for Technology in Society, Technische Universität München, Arcisstr. 21, 80333 München, Germany

A difference-making account of causality is proposed that is based on a counterfactual definition, but differs from traditional counterfactual approaches to causation in a number of crucial respects: (i) it introduces a notion of causal irrelevance; (ii) it evaluates the truth-value of counterfactual statements in terms of difference-making; (iii) it renders causal statements background-dependent. On the basis of the fundamental notions 'causal relevance' and 'causal irrelevance', further causal concepts are defined including causal factors, alternative causes, and importantly inus-conditions. Problems and advantages of the proposed account are discussed. Finally, it is shown how the account can shed new light on three classic problems in epistemology, the problem of induction, the logic of analogy, and the framing of eliminative induction.

[1] Draft, I am grateful for comments and criticism.

[2] pietsch@cvl-a.tum.de

1. Introduction

It is a stunning fact about scientific methodology that some of the most fundamental concepts seem also the least understood and the most controversial. Causation is a case in point. At times, it has been strongly criticized, both in philosophy and in the sciences. For example, Bertrand Russell notoriously proclaimed: “The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.” (Russell, 1913, p. 1) Karl Pearson, one of the founders of modern statistics, concurred: “Beyond such discarded fundamentals as ‘matter’ and ‘force’ lies still another fetish amidst the inscrutable arcana of modern science, namely, the category of cause and effect.” (Pearson, 1911, p. iv) By contrast, other not less prominent figures have argued for the exact opposite of such views, for example David Hume: “these [Resemblance, Contiguity, and Causation] are the only ties of our thoughts, they are really to us *the cement of the universe,* and all the operations of the mind must, in a great measure, depend on them.” (cited in Mackie 1980, p. v; my italics)

At least in part, the controversy results from the continuing lack of an adequate conceptual analysis of causation. It seems fair to say that all major philosophical accounts suffer from serious shortcomings. Naïve regularity approaches cannot ground the distinction between correlation and causation, contemporary counterfactual accounts rely on the dubious idea of similarity between possible worlds, manipulability and interventionist accounts are implausible when working with observational data, and mechanistic or process theories have failed to convincingly explicate the notion of a causal mechanism such that cases like causation by omission can be covered. In a way it seems that all the essential elements are on the table, but thus far a coherent picture has not been arranged from these pieces of the conceptual puzzle. In the present essay, I will suggest a somewhat novel difference-making account that is chiefly based on the counterfactual idea, but exhibits some crucial differences with respect to other approaches in that tradition.

In Section 2, the basic ingredients of the difference-making account are introduced. First, counterfactual definitions for the fundamental notions of causal relevance and causal irrelevance are given. Then a difference-making account of counterfactuals is proposed, according to which the truth or falsity of a counterfactual statement can be determined by showing that it belongs to a class of statements with the same truth-value, of which at least one is realized in the actual world. The main advantage of this account with respect to other prominent approaches—in particular the semantic approach based on possible worlds due to Robert Stalnaker and David Lewis and the metalinguistic view due chiefly to Nelson Goodman—is that less vagueness enters into the evaluation of counterfactuals, thereby rendering causation more objective. In its deflationist spirit, the difference-making account of causation, as outlined in this essay, is quite similar to Federica Russo’s ‘invariance across changes’ account (2014), which tries to extend basic interventionist ideas to situations where manipulations play no obvious role. The proposed account also bears considerable resemblance to sophisticated regularity theories like those of Mill (1886), Mackie (1980), or Baumgartner and Graßhoff (2004). With all of them, it shares the focus on eliminative induction and like Mackie, it emphasizes the importance of background dependence and of counterfactual reasoning. At the end of the section, causal ordering and the direction of causation are briefly discussed.

In Section 3, further causal concepts, in particular causal factors and alternative causes, are defined on the basis of the two fundamental notions, i.e. causal relevance and causal irrelevance. A link to the inus-logic of causal conditions is established. It is also shown that the difference-making approach can account for some crucial features of the pragmatics of causation, most importantly that causal relations can be formulated at various levels of coarse-graining and that they fulfill some basic intuitions about transitivity. Finally, I discuss how functional relationships can be grounded in the logic of relevant and irrelevant conditions, mitigating one of the classic objections against inus-accounts of causation that these can allegedly only deal with discrete relationships concerning the absence and presence of factors.

In Section 4, a number of conceivable objections are rehearsed. Also, the role of mechanisms and interventions for the difference-making account is briefly addressed. In summary, I argue that the proposed account manages fairly well to establish an objective approach to causation as appropriate and desired when considering the role of causal knowledge in the sciences.

Finally, in Section 5, some methodological issues concerning causal inference are addressed. A reframing of Mill's methods is suggested, based on only two fundamental methods, which corresponds directly to the conceptual analysis of the difference-making account. We will then take a look at two long-standing epistemological problems that are connected with causal inference. First, the problem of induction both in its Humean version and according to Goodman's new formulation will be discussed from the perspective of the difference-making account and of the corresponding methodology of eliminative induction. I will argue that the problem of induction that arises is largely distinct from traditional versions of this issue. The second epistemological problem concerns the formulation and justification of a formal framework for analogical reasoning. Here, I will argue that analogical inferences can be naturally integrated into eliminative inductive approaches in the tradition of Mill's methods.

2. The difference-making account[3]

In the following, the main elements of the difference-making account will be introduced: First, the notions of causal relevance and causal irrelevance are defined in counterfactual terms—broadly in line with the original counterfactual definition of causation as provided by David Hume. Second, the main conceptual problem of counterfactual approaches will be addressed, namely how truth-values of counterfactual statements are evaluated. It seems fair to say that all suggestions made in the past for this purpose face serious objections. I will therefore introduce a further proposal taking Mill's method of difference[4] as a guideline. The main reason, why this approach has not been developed in the past, arguably lies in the fact that the feasibility of a coherent notion of causal irrelevance has generally been dismissed.

[3] A preliminary sketch of the account was published in Pietsch (2015, Sec. 4a).

[4] Which actually is not Mill's, but was formulated before him by various known and unknown methodologists, including William Ockham, Francis Bacon, and John Herschel.

Third, the concept of a causal background or context will be introduced in the tradition of John Anderson and John Mackie. According to the suggested view, all causal statements will be rendered background- or context-dependent, i.e. without exception they hold only ceteris paribus.

2a. Causal relevance and causal irrelevance

As is well known, Hume proposed two distinct definitions of causation that he somewhat mysteriously equated although their relation is anything but clear: "We may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. Or, in other words, *where, if the first object had not been, the second never had existed*." (Hume 1748, Section VII). In fact, Hume elaborated merely the first definition leading to his regularity account which focuses on constant conjunction and notoriously results in the subjectivization of causal relationships. The second definition, relying on counterfactuals, was developed into a full account of causation only in the 20th century by David Lewis. But Hume still deserves the merit of having formulated the core idea: in order to determine a causal dependence between two events—or whatever one takes the causal relata to be—both events must be realized in the actual world in a certain instance and the truth of the corresponding counterfactual conditional must be somehow established.

To introduce the fundamental concepts of the difference-making account it is useful to recall Lewis's counterfactual approach for comparison, which according to Paul Horwich (1987) combines four basic elements. First, Lewis introduces a counterfactual definition of 'causal dependence' closely following Hume's view. According to Lewis, it is problematic that such causal dependence turns out not to be transitive. Typical counterexamples are cases of preemption[5], where causal dependence holds for the individual steps in a causal chain, but not for the relation between end and starting points (cp. Menzies 2014, Sec. 2.3). He therefore defines 'causation' in terms of causal chains as the transitive complement to causal dependence: "C caused E *if and only if* there was a sequence of events X_1, X_2, …, X_n such that: if C had not occurred, than X_1 would not have occurred; if X_1 had not occurred, then X_2 would not have occurred, … if X_n had not occurred, then E would not have occurred." (Lewis 1973, quoted in Horwich 1987, p. 167) Obviously, if there is causal dependence between two events, there is causation, but not vice versa.

As a second crucial element of his account, Lewis needs to specify how the truth-values of the counterfactual conditionals are determined, which trivially cannot be directly observed. To this purpose, he introduces his celebrated semantic approach relying on possible worlds and the notion of similarity between these worlds: "'If C were true, then A would also be true' is true (at a world w), iff either (1) there are no possible C-worlds, or (2) some C-world where A holds is closer to the actual world than is any C-world where A does not hold." (Lewis 1973, 560) This issue will be discussed in detail in Section 2b.

Given that the notion of similarity between possible worlds appears in Lewis's account of counterfactuals, the next difficulty consists in formulating a suitable account of similarity. It seems fair to say that Lewis never managed to find a satisfying solution for this task. He

[5] I will come back to this issue in Section 4b.

determines the main criteria for measuring similarity, namely resemblance with respect to laws and with respect to matters of facts (1979). But what is lacking is a detailed account providing a set of rules how to combine these two fundamental aspects in concrete situations. Rather, he seems to be constantly adjusting his approach in hindsight in order to accommodate certain examples for which clear intuitions about causal dependencies exist. Lewis has repeatedly stressed that this failure correctly reflects the in his view essentially subjective nature of causation, but this is clearly at odds with the fairly objective nature of causal knowledge in the sciences. Think of in particular the engineering sciences, where based on causal knowledge, complicated and highly reliable technical artefacts can be built like bridges, airplanes, computers etc.

Finally, the fourth element of Lewis's account concerns the time asymmetry of causation. For Lewis, the asymmetry consists mainly in the fact that counterfactuals do not generally hold in both directions: if the present were different, the future would change, but the past would generally not have changed. This topic will be briefly elaborated in Section 2c.

Let me now present the basic ingredients of the difference-making account, which, just as Lewis's approach, relies on a counterfactual definition of causation. However, it takes a different route to evaluating the truth-values of counterfactual propositions, one that gets by without the slippery notion of similarity between possible worlds. Rather it refers to situations in the actual world that differ only in terms of irrelevant circumstances. Thus, a notion of causal irrelevance is required as a complement to the notion of causal relevance.

The difference-making account defines causal relevance and causal irrelevance in the following manner:

> *In a context B, in which a condition A and a phenomenon C occur, A is causally relevant to C, in short $A \mathcal{R} C \mid B$, iff the following counterfactual holds: if A had not occurred, C would also not have occurred.*[6]
>
> *In a context B, in which a condition A and a phenomenon C occur, A is causally irrelevant to C, in short $A \mathcal{I} C \mid B$, iff the following counterfactual holds: if A had not occurred, C would still have occurred.*

Here and in the following, capital letters denote specific states, and lowercase letters denote variables that generally allow for a range of different states. Thus x=X means that variable x is in state X and x=¬X means that it is in state ¬X.

From the treatment of counterfactual propositions as elaborated in Section 2b, it is easily seen that the following relations hold:

$$A \mathcal{R} C \mid B \leftrightarrow \neg A \mathcal{R} \neg C \mid B \qquad \langle 1 \rangle$$

$$A \mathcal{I} C \mid B \leftrightarrow \neg A \mathcal{I} C \mid B \qquad \langle 2 \rangle$$

[6] Note that the meaning of causal relevance as defined here is not identical with that in the context of probabilistic causation, where causal relevance merely indicates the increase of probability under conditionalization.

Here, negation ¬ is understood in the sense of classical logic, in particular $\neg(X \wedge Y) = \neg X \vee \neg Y$ as well as $\neg(X \vee Y) = \neg X \wedge \neg Y$.

The proposed terminology is supposed to work both for the token and for the type level. To begin with, 'conditions' and 'phenomena' in the above definitions refer to types, i.e. they can all be instantiated more than once. In particular, variables as well as specific states of variables must be interpreted as types. Finally, the background is thought to be constituted by a large number of conditions, therefore it is equally a type-concept. Thus, causal relationships as identified by the above definitions are generally situated on the type level. However, sometimes the combined conditions determining a specific event may be so strong that it turns out in practice not repeatable, i.e. a token.

In contrast to Lewis's counterfactual approach, the difference-making account does not need to distinguish between causal dependence and causation, since it will deal differently with both transitivity and preemption (see Sections 3c and 4b). Note further that according to conventional terminology (e.g. Baumgartner & Graßhoff 2004, Ch. 3.2), causal relevance requires that a condition makes a difference *in at least one situation*, while according to the above framework it must *always* make a difference *in a specified context*. Correspondingly, causal irrelevance is conventionally taken to require that a condition *never* makes a difference, while in the framework proposed above it must *never* make a difference *with respect to a specified context*. Obviously, according to the conventional perspective, causal irrelevance is in practice impossible to establish since nobody can ever claim to have examined all possible contexts, which has led to the general rejection of this notion in the literature. Conversely, a meaningful notion of causal irrelevance is only possible when requiring universal background dependence of causal statements.

Thus, one notable feature of the difference-making account is that it attributes a central role to the notion of causal irrelevance—which arguably distinguishes it from all other contemporary accounts of causation.[7] In particular, causal irrelevance will turn out crucial for the novel manner to evaluate counterfactuals that will be outlined in Section 2b. Roughly, a counterfactual statement is true if an instance is realized in the actual world that differs from the examined counterfactual instance only in terms of irrelevant circumstances.

A further important characteristic is that according to the difference-making account causal relations are universally rendered context-dependent, i.e. they are always defined with respect to a background or context of further conditions that are held constant if potentially causally relevant or that are allowed to vary if causally irrelevant.[8] More precisely:

[7] A notion of causal irrelevance is discussed in Galles and Pearl (1997, Sec. 4). However, these authors, working within an interventionist rather than a counterfactual approach, do not ascribe a central role to irrelevance in their conceptual framework. In particular, they do not acknowledge the fundamental role of causal irrelevance for the evaluation of counterfactual statements, as discussed in Section 2b. Tilman Sauer, in a yet unpublished manuscript, provides an interesting case study from physics regarding the analysis of the cosmic microwave background, where conclusions to irrelevance are frequently made and an explicit statistical methodology for detecting causal irrelevance is developed.

[8] Context dependence is of course stressed in Mackie's account of causation, while the basic idea dates back at least to his teacher John Anderson.

> *A context shall consist in (i) certain conditions $A_1, \ldots, A_n$ that are considered potentially causally relevant to a number of phenomena $C_1, \ldots, C_m$ and whose impact on these phenomena is explicitly examined; (ii) conditions that co-vary with some of the As, e.g. because they lie on causal chains leading through at least one of the As to at least one of the Cs; (iii) further conditions that may vary and that are assumed to be causally irrelevant to the considered phenomena; (iv) further conditions that remain or are explicitly held constant, some of which may be causally relevant to the considered phenomena.*
>
> *More narrowly, a background shall comprise conditions (iii) and (iv).*

As an example, consider an experiment to examine the causal structure of a simple pendulum. (i) Typical conditions A that are explicitly varied are the length of the rod, the mass of the weight, or the initial displacement of the weight. A phenomenon C in this example could be the period of the pendulum. (ii) A condition that is not explicitly considered as one of the As but co-varies with them is the gravitational force on the weight which changes with its mass and thereby leads to a change in period. In other words, the variable gravitational force lies on a causal chain between the variables weight and period. (iii) Certainly, there are myriads of conditions that might change even during a simple pendulum experiment, e.g. the position of the earth with respect to the sun, the formation of clouds in the sky, the thoughts of the president of the United States etc. All of these conditions are generally assumed to be irrelevant for the causal structure of the considered set-up. But note that it may occasionally happen that this third category comprises conditions that eventually turn out relevant to the examined causal structure, which would then require a readjustment of the context. (iv) Finally, in the last category are all those conditions that stay constant in the considered set-up, either because they are explicitly held fix or because they coincidentally happen to remain constant. This category usually comprises some conditions that are known to be relevant or potentially relevant for the set-up, such as in the discussed example the mass of the earth or the position of other large bodies in the immediate vicinity of the pendulum.

As a matter of terminology a change in background will be noted in the following manner: $B \wedge X$, i.e. $x=X$ is held constant in addition to the requirements determined by B; $B \wedge x$, i.e. variable x can take on any possible value, e.g. X or $\neg X$. We speak of a background within B if it is at least as restrictive as B, i.e. if it imposes at least the restrictions of B concerning constancy of conditions.

Let me also provide an example to point out some consequences of the definitions of causal relevance and irrelevance. In 1871, a terrible fire almost completely destroyed the city of Chicago. Relatively quickly, the location was found where the fire had originated, namely in a barn in the southwest of the city. Soon afterwards, a newspaper published a story, according to which a cow belonging to the owners of the barn Patrick and Catherine O'Leary had kicked over a lantern and started the fire. As is well-known, the logic of such causal conditions has been analyzed by John Mackie, constituting his most important contribution to the literature of causation. According to Mackie, a cause is an inus condition, i.e. an *i*nsufficient but *n*on-redundant part of an *u*nnecessary but *s*ufficient condition. Clearly, the kicked-over lantern is such an inus condition since the general expression for the cause of a barn fire may be something like: (kicked-over lantern $\wedge$ hay on the floor) $\vee$ lightning cause a barn fire with respect to a background of other conditions B*.

Let us briefly examine in the example of the Chicago fire, which events or conditions are causally relevant or irrelevant. As we will see, some of the consequences are somewhat counter-intuitive. For example, the kicked-over lantern is only relevant with respect to certain backgrounds, e.g. with respect to background B* $\wedge$ hay $\wedge$ no lightning. However, it is causally irrelevant with respect to B* $\wedge$ no hay $\wedge$ no lightning. Moreover, it is causally irrelevant with respect to a background, in which another factor causes the fire, e.g. with respect to B* $\wedge$ hay $\wedge$ lightning. A further consequence is that in some contexts the question of relevance or irrelevance is underdetermined, for example with respect to B* $\wedge$ no lightning. After all, both may be the case depending on the presence or absence of hay. The concepts of causal factor and alternative cause required for adequately describing such situations will be introduced in Section 3a.

Let me stress again that some of this may sound counter-intuitive at first, since according to conventional terminology a condition is considered causally relevant to an event just in case that there exists *some* background under which it makes a difference to the event. Thus, the kicked-over lantern is always considered causally relevant to a barn fire, no matter if combustible material is present or not. But again, it is essentially this conceptual choice, which renders it more or less impossible to define a meaningful concept of causal irrelevance. Also, it is exactly this move that is at the origin of many contradictions and inconsistencies in classical approaches to causation, e.g. concerning overdetermination and preemption. Arguably, the only way to avoid these problems is to carefully keep track of the background with respect to which causal relevance or irrelevance can be stated.

From the perspective of the sketched framework, causal knowledge can be acquired and improved in a process of variation of conditions or circumstances.[9] In particular, causal relevance and irrelevance of conditions must be examined in different contexts. By increasing the set of known irrelevant conditions for a certain phenomenon the generalizability of causal relationships can be determined. But note again that by this procedure, one never arrives at strictly universal relationships, the ceteris-paribus character implied by the background-dependence will always remain.[10]

Note finally that in distinction to Mackie's approach the fundamental terminology of the difference-making account is not in terms of necessary and sufficient conditions, but in terms of causally relevant and irrelevant conditions with respect to a context. Crucially, the difference-making approach to counterfactuals which will now be outlined is not accessible to the wide-spread terminology of necessity and sufficiency.

[9] This variational rationale, which is in spirit opposed to Humean regularity views, can be found with a variety of authors including Bacon, Mill, and Keynes. In recent literature, it has been most forcefully defended by Federica Russo (e.g. 2009).

[10] Thus, strictly universal relationships, as they can for example be found in physics, cannot be fully causal, but must have a decisive conventional element. This constitutes yet another answer to the ongoing debate, to what extent causality plays a role for physics.

2b. A difference-making account of causal counterfactuals

As mentioned, the main challenge for the counterfactual analysis of causation concerns the evaluation of the truth-values of the counterfactual conditionals. Once this problem is solved, many features of causation follow directly, e.g., that causation implies some kind of necessity and therefore can establish claims about prediction and manipulation. In the literature, one finds two main accounts of counterfactuals, an older one often called the *metalinguistic framework* that was chiefly developed by Nelson Goodman and a newer one mostly referred to as *possible worlds* or *semantic view* due mainly to Robert Stalnaker and David Lewis (cp. e.g. Psillos 2015, Reutlinger 2012, Sec. 4).

The basic idea of the metalinguistic framework (e.g. Goodman 1983, Ch. I) when evaluating a counterfactual propositions such as "if A had been true, B would also have been true" is to examine under which additional premises in terms of laws and matters of facts, A would actually entail B. Thus, the fundamental problem for the metalinguistic approach is to determine which premises are 'cotenable' with A and which are not, i.e. what should be admitted in addition to A in order to determine if B is the case and thereby the truth value of the counterfactual. It turns out that this issue has never been satisfactorily resolved.

As a consequence, Mackie developed what is sometimes called the *supposition view*, which stands broadly in the tradition of the metalinguistic framework but stresses a strong contextuality of counterfactual statements: to assert a counterfactual statement like that given above is to claim that B broadly belongs to the implications, if A is supposed. Thus, the truth value of the counterfactual proposition much depends on the attitude and context in which it is expressed. Cotenable are those propositions that are naturally presupposed in a certain context along with A. (Psillos 2015, 88-89)

The other main approach to counterfactuals that has dominated discussions about causation in the past decades attempts to determine the truth value of counterfactual statements by referring to possible worlds and the similarity between them. While fundamental ideas are due to Robert Stalnaker, it was mainly David Lewis who turned these into a full-blown and sophisticated account (Lewis 2001).

The basic idea in Lewis's approach is the following: "If A were true, then C would also be true" is true (at a world w), iff either (1) there are no possible A-worlds, or (2) some A-world where C holds is closer to the actual world than is any A-world where C does not hold (Lewis 1973, 560). Here, an A-world just refers to a possible world in which A is realized. The first part (1) amounts to a definition for specific, rather extraordinary situations and thus, the second part (2) is certainly more interesting. Obviously, the main challenge to this approach is to come up with a proper account of possible worlds and especially with a plausible notion of similarity.

Lewis's remarks concerning similarity are quite vague and general: "Overall similarity among worlds is some sort of resultant of similarities and differences of many different kinds, and I have not said what system of weights and priorities should be used to squeeze these down into a single relation of overall similarity. I count that a virtue. Counterfactuals are both vague and various. Different resolutions of the vagueness of overall similarity are appropriate in different contexts." (Lewis 1979, 465) As also required in the metalinguistic account, Lewis needs to balance considerations of laws and of matters of facts when judging counterfactual statements. The feeling remains that his account can reconstruct many causal analyses in hindsight, but that it mostly fails to guide intuitions in difficult cases. Notorious and much-discussed counterexamples to Lewis's approach concern situations in which a single antecedent has enormous consequences, for example pressing the infamous red button to start a nuclear war.

Thus, the two most influential approaches suffer from considerable vagueness which admittedly may sometimes be a property of counterfactual statements. However, judging from common intuitions about causal knowledge, the subjectivity seems overemphasized at least for those counterfactuals corresponding to causal dependencies. After all, for a large number of phenomena the causal structure is fairly well understood and more or less unambiguous. Think for example of all those achievements that the engineering sciences have contributed to human knowledge. A further issue is that both traditional approaches refer to the notion of law when evaluating counterfactual statements—which introduces at least the threat of circularity, since often such laws will themselves be of causal nature.

In view of these problems, let me argue for yet another way to evaluate the truth of counterfactual statements. On the positive side, it will not be plagued by the vagueness and subjectivity of the above approaches and it will not refer to laws. On the negative side, the account is not meant as a full-blown account for all counterfactual statements. For example, it cannot plausibly answer what the truth-value is of strongly hypothetical propositions that differ substantially from familiar phenomena, e.g. describing the consequences had Julius Caesar led the UN-forces in the Korean War, to cite an oft-used example from the literature. Rather, the account will be limited to counterfactual statements that occur in causal contexts. It is explicitly intended as an account of causal counterfactuals.

The starting point of the proposed account is that in order to determine the truth-value of a counterfactual proposition two things have to be established: (i) it must be shown that the counterfactual statement belongs to a class of propositions with the same truth value; (ii) and at least one of the propositions in this class must describe an instance which is either realized in the actual world implying that the examined counterfactual proposition is true, or the negation of which is realized implying that the counterfactual proposition is false. Note once more that the proposed account does not need to construe possible worlds, but refers only to instances which are realized in the actual world.

A guiding idea is to use the method of difference as a reference point to fill in the details. The most influential formulation of this method is of course due to John Stuart Mill: "If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon." (1848, 256) There are several obvious problems. First of all, it seems impossible to ever change only a single factor. At least, the causes and effects of a considered factor will always change with it. In addition, there will always be myriads of presumably unrelated conditions in the universe that will be different. Furthermore, there are a number of situations in which the method does not yield results in accordance with usual intuitions. For example, it fails to identify necessary, but insufficient factors, when additional necessary conditions are not instantiated, although we habitually speak of those as causes. Furthermore, the method has problems with cases of overdetermination and preemption. Many authors have inferred from this rather devastating survey that the method of difference can only serve as a heuristic rule to identify candidate causal factors—a conclusion, which certainly contributed to the wide-spread view of causation as a subjective concept.

Here, I will take a different route arguing that the mentioned problems are solvable by conceptually refining the method of difference. The main trick will consist in introducing the notion of causal irrelevance as a complement to causal relevance and to require that the instances compared in the method of difference differ only in terms of irrelevant circumstances with some exceptions that will be specified further below. This condition will be called homogeneity in the following. Thus:

> *'If A were not the case, C would not be the case' is true with respect to an instance in which both A and C occur in a context B, if (1) at least one instance is realized in the actual world in which neither A nor C occurs in the same context B and (2) if B guarantees homogeneity.*

As a next step, homogeneity needs to be defined:

> *Context B guarantees homogeneity, iff only conditions that are causally irrelevant to C (and ¬C) can change, (i) except for A and (ii) conditions that are causally relevant to C in virtue of A being causally relevant to C.*

Note that, strictly speaking, irrelevance in this definition of homogeneity must again be understood with respect to a specified background, which turns out, in fact, to be also B excluding of course the considered irrelevant condition itself as well as other conditions that vary in virtue of changes of the considered condition. Calling to mind the definition of a context as provided in the previous Section 2a, homogeneity requires that all other conditions A_1, …, A_n with the exception of the considered condition A are causally irrelevant to C. Homogeneity conditions have occasionally been evoked in the literature on causality, for example by Baumgartner and Graßhoff (2004, Sec. 2.4) who formulate a quite sophisticated version of this requirement in the context of a regularity approach to causation or by Holland (1986, p. 948) who employs it in the context of a counterfactual theory.

Holland defines homogeneity of two instances that they exhibit the same value for the dependent variable given the same value of the independent variable. By contrast, Baumgartner and Graßhoff define homogeneity in terms of causal relevance[11], essentially that all causally relevant conditions must remain constant between instances. A serious problem for the latter account is that the authors are working with conventional terminology, according to which, as already stated, a condition is causally relevant basically if it makes a difference to a phenomenon *in some context*. Consequently, many conditions are identified as causally

[11] i.e. not causal irrelevance as in the definition given in the present article

relevant to a phenomenon that count as causally irrelevant with respect to specific contexts according to the definitions in the present essay. Even worse, the danger looms large that according to conventional terminology almost everything is causally relevant to everything else since one can always construct complex stories where one thing leads to another. For example, the current position of Jupiter might be used by a psychic to scare some poor person to an extent that she commits suicide confirming the very astrological prediction. It seems to follow that the position of Jupiter has to be held fix to fulfill homogeneity when examining causes of suicides. This renders Baumgartner and Graßhoff's version of the homogeneity condition a much stronger requirement than the context-dependent one proposed in this article. In fact, the former may even turn out inapplicable.

The definition of homogeneity suggested here in terms of causal irrelevance is also superior since it is closer to scientific practice. Judgments of irrelevance seem to be ubiquitous, e.g. when the range of validity of a causal relationship is extended to novel contexts. Indeed, one important manner of improving causal knowledge is by showing that a causal relation that has originally been established for a very specific context continues to be stable under the change of an increasing number of conditions, i.e. that these further conditions are irrelevant to the considered phenomenon. Thus, a general theory of causation without a notion of irrelevance appears not viable.

Finally, we need to explicate a further concept that appears in the definition of homogeneity:

> *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).*

Remember that contexts within B are all those contexts that are at least as restrictive as B, i.e. in which some further conditions might be fixed in comparison with B. Conditions X include those that lie on a causal chain through A to C (both prior or posterior to A) or are epiphenomena of such conditions. From the difference-making account as presented above we can immediately derive relation <1>. That causal relevance of A to C in a context B implies causal relevance of ¬A to ¬C in the same context B, trivially follows from the truth condition for the counterfactual which requires two situations to exist in the same context B, one where both A and C are present and one where both A and C are absent.

Let me give an example for the evaluation of causal counterfactuals using again the story of the Chicago fire. In order to say that the kicked-over lantern is causally relevant to the barn fire with respect to a background B* ∧ hay ∧ no lightning, we first have to know that on the evening in question the cow kicked over the lantern and a fire started. Second we have to evaluate the truth-value of the causal counterfactual "if the lantern had not been kicked over the fire would not have started". For this, an instance must be found with the same background B* ∧ hay ∧ no lightning which differs only in circumstances that are causally irrelevant to the barn fire, except that the lantern is not kicked over, and the causal conditions leading to the kicked-over lantern may be different as well as the conditions leading from the kicked-over lantern to the barn-fire. And indeed, the night before, everything was at the same place in the barn except that the cow did not raise its hoof and no fire happened. Abundant experience suggests that all other circumstances that differed in the two nights were irrelevant to the barn fire, e.g. what the neighbors did on those evenings, the chicken running around in front of the barn etc.

The causal counterfactual that appears in the definition of causal irrelevance is evaluated in a fully analogous manner:

> *'If A were not the case, C would still be the case' is true with respect to an instance in which both A and C occur in a context B, if (1) there exists at least one instance in which A does not occur but C still occurs in the same context B and (2) B guarantees homogeneity.*

Note that with respect to the definition of homogeneity, there trivially are no conditions that are causally relevant to C in virtue of A being causally relevant to C. Obviously, causal irrelevance of A to C with respect to B immediately implies causal irrelevance of ¬A to C with respect to B (relation <2> of the previous section). One might suspect that one can also infer causal irrelevance of A and ¬A to ¬C with respect to B. But this is clearly not the case for the simple reason that by definition ¬C cannot be realized with respect to B if A is causally irrelevant to C and B guarantees homogeneity.

Let us briefly point out some of the characteristics of the discussed approach to counterfactuals. First, the approach can only account for the small class of counterfactuals which were referred to as causal counterfactuals. In particular, it cannot determine the truth value of counterfactual statements if no instance exists in which A is not realized with respect to the same background B. Sometimes, idealizations and simplifications may help, for example when determining the causal nature of the gravitational interaction of the planets in the solar system. But in many situations, this is impossible. The above account just cannot provide any clue, whether New York would exist, if the Neandertals had survived, or who would have won the Korean War, if Caesar had led the UN-forces. In a way, the difference-making approach to counterfactuals singles out those cases in which an objective evaluation is possible modulo the specific problem of induction discussed in Section 5.

Second, one might be tempted to think that the presented account for evaluating counterfactuals is just a special case of Lewis's possible world approach. But this is clearly not true. For one thing, the given account evaluates counterfactuals based on actual phenomena occurring in the one and only world accessible to human experience and not with respect to possible worlds.[12] Also, the measure of comparison between different situations is not at all a similarity measure of the kind as David Lewis had envisaged it. In the difference-making approach, the 'measure' is only a two-valued function: either homogeneity is fulfilled for two situations, i.e. the two situations differ only in terms of irrelevant circumstances, or not. Moreover, this measure depends not only on the two situations or states that are compared, but also on the phenomenon C and the potential cause A as is clear from the definition of homogeneity. By contrast, similarity is usually framed as a gradual and universal measure that only depends on the compared instances, not on the epistemological aim, i.e. in

[12] Given that the proposed account refers to the actual world, one might think that it really is about indicative conditionals. But obviously, this is not the case since the target of analysis clearly is a counterfactual proposition.

this case the causal relation which is examined. For example, Lewis's similarity measure between possible worlds only depends on these worlds and not on the events mentioned in the considered counterfactual statement. Finally, let me emphasize again that the measure of comparison in the difference-making approach does not exhibit the subjectivity and vagueness that Lewis associated with similarity.

Third, the suggested account of causation prioritizes singular causation, since the definition of causal relevance refers to a specific instance in a certain context with respect to which counterfactual questions are asked. Clearly, a causal relationship can be established without any actual or even potential regularity, based only on two instances as required in the definition of causal relevance. On the other hand, causal regularities may follow if the relevant conditions under which the causal relationship has been established are in combination so weak that the relationship is repeatable. But note that such causal regularities always remain context-dependent, they preserve a distinct ceteris-paribus character in that all relevant conditions can never be made explicit in total.[13]

Finally, a crucial restriction of the account as outlined thus far is that it works only for deterministic contexts. For example, in a situation of indeterminism, conditions might be identified as irrelevant even though they are statistically relevant for a phenomenon. But an extension of the framework to reasonable cases of indeterminism is quite straightforward as will be outlined in Section 5b.

2c. Remarks on the asymmetry of causation

A perennial problem plaguing the debate on causation concerns the nature and origin of the causal asymmetry in particular in relation to the temporal asymmetry and various connected asymmetries (Hausmann 1998; Price and Weslake 2009; Frisch 2014, Ch. 5). It is clear that the account given so far cannot be sufficient. Most importantly, the definition of causal relevance is almost—though not fully—symmetric with respect to the role of A and C, while causal relevance certainly is not a symmetric concept. In most cases, causal relevance of X to Y does not imply causal relevance of Y to X.

Let us therefore focus briefly on the asymmetric aspects in the definition of causal relevance. There is an asymmetry in terms of the context dependence in that the conditions that are held constant in the background are usually required not to be causally and temporally posterior to the examined phenomenon. This holds both for causal relevance and for causal irrelevance. But note that there is a further asymmetry for the latter in that even without background dependence the situation is not symmetric with respect to A and C. After all, causal irrelevance of A to C requires two situations to be observed, namely A and C co-occurring as well as ¬A and C co-occurring, while by definition ¬C cannot occur with respect to background B.

[13] With respect to universal regularities that are not context-dependent like the axioms of physical theories, I would argue that these do not have an exclusively empirical character but are rather of conventional or definitional nature, in line with conventionalist ideas brought forward in particular in the second half of the 19th century by Poincaré, Duhem, Mach, and others.

One crucial question is whether time coordinates must be introduced for all events in order to account for the asymmetry of causation. While time coordinates may be pragmatically useful, I believe that sub specie aeterni, they should be dispensable—if assuming in the tradition of Leibniz that time is ultimately a relational concept that describes how phenomena evolve in relation to other phenomena. Thus, if all changes in the world were of causal nature, then the asymmetry of time should be reducible to asymmetries of causation.

I will in the following briefly sketch an argument, according to which both temporal ordering and the temporal asymmetry can be linked to the notion of causal chains. It belongs to our basic human experiences that there are processes, which closely link several events C_1, …, C_N by means of causal relevance with respect to a relatively broad and stable context B*: any C_i is causally relevant for any C_j with respect to B* (at least for $i < j$). It is also the case that we can often interfere with such systems in a targeted manner. Let us therefore further introduce N interruption events I_1, …, I_N and as many stimulation events S_1, …, S_N: an interruption event I_i prevents C_i from occurring, thus basically is an inhibiting cause of C_i in the presence of C_{i-1}. A stimulation event S_i is an event that starts or upholds a causal process, i.e. an alternative cause of C_i instead of C_{i-1}. Both interruption and stimulation events can be thought of as external interventions in the system comprising C_1, …, C_N. Obviously, B* as framed before must include the absence of interruption and stimulation events to ensure causal relevance between the members of the causal chain.

Let me give two examples: first, a chain of toothed wheels, where each toothed wheel moves a subsequent toothed wheel. Here, event C_x is the rotation of wheel number x. In the second example, billiard balls collide elastically on a linear track. Here, event C_x is the setting in motion of ball x by the previous ball x-1. Both examples concern extremely stable processes, where one event regularly leads to the next with respect to a wide variety of backgrounds. The crucial difference concerns the fact that in the first example the interaction is instantaneous, in the second it is not.

It is easy to see that causal chains introduce a definite causal ordering through the notion of interruption and external stimulation. Consider the chain of toothed wheels. In the presence of an interruption, two groups of wheels result, such that relations of causal relevance remain that link every element with every other within each group, while all wheels belonging to different groups are causally irrelevant to each other. In a causal chain linking N events, there are N-1 possibilities to interrupt leading to N-1 different groupings, which in turn imply a definite causal ordering of the Cs. After all, the groupings can be arranged in a way that neighboring groupings differ only in terms of one of the Cs changing sides. In this manner, causal ordering can be constructed for a large class of phenomena from experience.

As already mentioned, the main difference between the two examples is that the chain of billiard balls introduces a causal asymmetry, i.e. a specific direction in the chain. After all, in the example of the toothed wheels, we have for any stimulation event S that they are possible alternative causes for *all* C. The situation is asymmetric in the example of the billiard balls in that only for one end of the chain all S are possible alternative causes or more generally, for C_i only all S_j with $j \leq i$ are possible alternative causes. This introduces a causal asymmetry in such processes that can be identified with the temporal asymmetry. The asymmetry is in accordance, of course, with the basic experience that while we can influence the future, we cannot influence the past. Thus, the basic suggestion is that the time asymmetry derives (at least partly) from the experience that there are processes which exhibit the described causal asymmetry with respect to interventions and stimulations.

Thus, by means of the notion of causal chains, both an ordering and an asymmetry can be established. Arguably, this causal ordering and asymmetry are at the basis of our spatial and temporal ordering of the world. Spatial relationships correlate closely with the question how easily one can causally interact with certain events and I very tentatively tend to believe that space should eventually be reducible to this feature of causation. A point that was not addressed thus far is why the ordering relations and the asymmetry should be universal, i.e. imply a universal spatial and temporal ordering as well as a universal time direction for all human experience. The brief answer is that none of the causal chains depicted above exists in isolation. Rather, every phenomenon is embedded in a large web of causal relations. Through this web, the ordering can be transferred and compared with other events. Indeed, if the empirical world fell apart in two or more causally separated regions, the sketched perspective would imply that there could be no spatial and temporal relations between these regions.

3. A causal calculus

3a. Causal factors and alternative causes

Causal relevance and causal irrelevance are the fundamental building blocks of the difference-making account. Other causal notions can be defined on this basis by what one might call a causal signature, which states in which specific contexts a factor is causally relevant and in which it is causally irrelevant. Most importantly, by this approach, we will be able in the following to reproduce the inus-logic of causal factors that arguably constituted John L. Mackie’s most important contribution to the debate on causation.

A crucial concept is that of a *causal factor* A for phenomenon C with respect to a background B. A causal factor is not itself sufficient in the context B to produce the phenomenon C, but requires other factors X to be instantiated at the same time. The definition of causal factors can be stated as follows:

> *A is a causal factor for phenomenon C with respect to background B, iff there exists an X such that A is causally relevant to C with respect to B∧X and irrelevant to ¬C with respect to B∧¬X (i.e. C is always absent in B∧¬X).*

As is easy to see, it follows: X is relevant to C with respect to B∧A and causally irrelevant to ¬C with respect to B∧¬A, i.e. X is also a causal factor for C with respect to B. After all, according to the above definition, the states A∧X∧C∧B, ¬A∧X∧¬C∧B, A∧¬X∧¬C∧B, ¬A∧¬X∧¬C∧B must exist. We can combine the first and the third state to show that X is relevant to C with respect to B∧A and the second and fourth state to show that X is causally irrelevant to ¬C with respect to B∧¬A.

Note further that X cannot be redundant with respect to A, i.e. A cannot itself be causally relevant to C with respect to B. Equally, A cannot be redundant with respect to X. In other words, the definition already implies, what Baumgartner has called the principle of non-redundancy with respect to causal factors (2013).

From the definition follows theorem <I>:

> *If and only if A is a causal factor, then $A \wedge X$ is causally relevant to C with respect to B.*
> *($A \wedge X\ \mathfrak{R}\ C \mid B \leftrightarrow A\ \mathfrak{R}\ C \mid B \wedge X$ and $A\ \mathfrak{J}\ \neg C \mid B \wedge \neg X$)*[14]

After all, according to the definition of causal relevance, we need to observe $A \wedge X$ as well as $\neg(A \wedge X) = \neg A \vee \neg X$. If we introduce the additional requirement[15] that in the case of ∨-connections all possible combinations of conditions must be realizable, we thus need to observe the following four instances $A \wedge X \wedge C \wedge B$, $\neg A \wedge X \wedge \neg C \wedge B$, $A \wedge \neg X \wedge \neg C \wedge B$, $\neg A \wedge \neg X \wedge \neg C \wedge B$, which are exactly the ones mentioned before when introducing the notion of a causal factor.[16]

In the example of the Chicago fire of the previous section, we had identified (kicked-over lantern ∧ hay) ∨ lightning as possible causes in some context B*. Thus, the kicked-over lantern is a causal factor A of the barn fire C with respect to a background B* ∧ no lightning with presence of hay being the complementary factor X.

The second concept required for establishing the inus-logic is that of an *alternative cause*:

> *A is an alternative cause to C with respect to background B iff there exists an X such that A is causally relevant to C with respect to a background $B \wedge \neg X$, but causally irrelevant to C with respect to a background $B \wedge X$ (i.e. C is always present in $B \wedge X$).*

It immediately follows that X is causally relevant to C with respect to a background $B \wedge \neg A$, and causally irrelevant to C with respect to a background $B \wedge A$, i.e. X is also an alternative cause to C with respect to B. After all, the definition of an alternative cause requires the following states: $A \wedge \neg X \wedge C \wedge B$, $\neg A \wedge \neg X \wedge \neg C \wedge B$, $A \wedge X \wedge C \wedge B$, $\neg A \wedge X \wedge C \wedge B$. Again, we can recombine these: the second and fourth to show that X is causally relevant to C with respect to a background $B \wedge \neg A$, the first and third to show that X is causally irrelevant to C with respect to a background $B \wedge A$. An example in the Chicago fire case is lightning as alternative cause to (kicked-over lantern ∧ hay).

Again, the definition implies Baumgartner's principle of non-redundancy, now with respect to alternative causes. Essentially, non-redundancy for the difference-making account amounts to the requirement that all possible combinations determined by the ∨-connector, if it appears in positive and/or negative instances actually occur.

It follows theorem <II>:

[14] Here, X refers to the same quantity as in the definition of causal factor.

[15] which is implicit in the definition of alternative causes.

[16] From the second and third instance taken together nothing follows, the first and fourth can be combined as $A \lozenge X\ \mathfrak{R}\ C \mid B$ according to terminology introduced in Section 3c. According to a theorem derived there, it follows from the first and fourth state that there is an inus-complex drawing on A and/or X which is causally relevant to C.

> *If and only if A is an alternative cause, then A∨X is causally relevant to C with respect to B.* ($A \vee X\ \mathfrak{R}\ C \mid B \leftrightarrow A\ \mathfrak{R}\ C \mid B \wedge \neg X$ *and* $A\ \mathfrak{J}\ C \mid B \wedge X$)[17]

This is again easy to see. Causal relevance requires that $(A \vee X) \wedge C$ as well as $\neg(A \vee X) \wedge \neg C = \neg A \wedge \neg X \wedge \neg C$ are instantiated. Keeping in mind that in case of the ∨-connection all possible combinations must occur, this results in the four combinations listed after the definition of alternative causes.

Theorems <I> and <II> allow for a pragmatic way to deal with several causal factors and several alternative causes, since according to them every condition can itself be considered a more complex operator of several alternative causes and causal factors.

Finally, let me introduce a related notion of alternative causes relying on an exclusive ($\underline{\vee}$) rather than a non-exclusive (∨) or and which shall be referred to as *substitute cause* in the following. Regarding the state $A \wedge X \wedge B$ several options are possible. For example, such a state may be impossible or two instantiations of C could result.[18] The latter case will be taken up when discussing concomitant variations in Section 3d.

We now have the conceptual resources to define the crucial notions of an *inus-condition* and of an *inus-complex* based on causal factors and alternative causes:

> *A factor X is an inus-condition for C with respect to a background B, iff X is a causal factor of a condition A that is causally relevant to C with respect to a background (B plus absence of all alternative causes for C, of which there is at least one).* [19]

The notion is largely identical with Mackie's concept of an inus-condition, i.e. it constitutes a direct translation of what an "insufficient, but non-redundant part of an unnecessary, but sufficient condition" amounts to in terms of alternative causes and causal factors. As we had seen, the required non-redundancy is implied already by the definitions of causal factors and alternative causes, though in Section 3c one slight deviation from Mackie's understanding of non-redundancy will be introduced in order to render the proposed concept of an inus-condition transitive. Furthermore:

> *A complex of conditions of the form $(X \wedge Y) \vee Z$ shall be called an inus-complex for C with respect to B, if it is causally relevant to C with respect to B. Note that each condition X, Y, or Z may itself be an inus-complex. By definition, it may be the case that $Y=1$ and/or $Z=0$.*

Essentially, an inus-complex is a condition that is itself causally relevant and consists of an arbitrary number of alternative causes that are each composed of an arbitrary number of causal factors. In the example, the kicked-over lantern would be an inus-condition, and (kicker-over lantern ∧ hay) ∨ lightning would be an inus-complex.

[17] Here, X refers to the same quantity as in the definition of an alternative cause.

[18] Correspondingly, there could be special types of causal factors, where the absence of both A and X is not instantiated.

[19] There is a slight ambiguity, since one might want to distinguish alternative and substitute causes, here. But nothing much hinges on this choice.

Let me also introduce the notion of an *inhibiting factor*:

> *Condition A is an inhibiting factor counteracting a causally relevant X for C with respect to background B iff X is causally relevant to C with respect to* $B\wedge\neg A$*; X is causally irrelevant to* $\neg C$ *with respect to a background* $B\wedge A$*.*

According to this definition, we have the following states $\neg A\wedge C\wedge X\wedge B$, $\neg A\wedge\neg C\wedge\neg X\wedge B$, $A\wedge\neg C\wedge X\wedge B$, $A\wedge\neg C\wedge\neg X\wedge B$. A recombination of these yields the following: A is causally relevant to $\neg C$ with respect to $B\wedge X$; A is causally irrelevant to $\neg C$ with respect to a background $B\wedge\neg X$. Note that an inhibiting cause A is not a causal factor for C, but $\neg A$ fulfills the definition. Thus, A is an inhibiting factor counteracting X for C with respect to background B, if and only if $\neg A$ is a causal factor for C with respect to background B.

Finally, a further notion is that of a *common cause*: A is a common cause to X and Y with respect to a background B, if in all backgrounds within B, in which $X \mathfrak{R} Y$ and/or $Y \mathfrak{R} X$, we also have $A \mathfrak{R} X$ and $A \mathfrak{R} Y$. There are certain difficulties distinguishing various causal structures: common cause, a causal chain A-X-Y or A-Y-X, a situation, where X and Y are definitionally related etc. For lack of space, we cannot go into details, here. Essentially, what is required for a common cause structure is that there are backgrounds within B, where X is causally irrelevant to Y and/or vice versa, because X and/or Y are caused by conditions other than A. In a way, the approach to common causes as outlined above is a deterministic version of Reichenbach's screening-off condition, which essentially states that when there is correlation between two effects this can be attributed to the common cause. Here, the analogous statement holds that when there is causal relevance between effects, it can be completely attributed to the causal relevance of the common cause. The probabilistic version should follow given an appropriate account of causal probability—a topic which leads us too far astray at this point.

At the end of this section, let me elaborate in some additional depth on the notion of causal irrelevance. First, we have theorem <III> regarding the extension or widening of backgrounds:

> $A \mathfrak{I} C \mid B\wedge X$ *and* $A \mathfrak{I} C \mid B\wedge\neg X \rightarrow A \mathfrak{I} C \mid B\wedge x$*, if* X, $\neg X$ *are all possible values of* x*.*[20]
>
> $A \mathfrak{R} C \mid B\wedge X$ *and* $A \mathfrak{R} C \mid B\wedge\neg X \rightarrow A \mathfrak{R} C \mid B\wedge x$*, if* X, $\neg X$ *are all possible values of* x*.*

This follows in a straightforward manner from the definitions of causal relevance and irrelevance. Since variables with more than two possible values can always be expressed in terms of binary variables, an extension to discrete variables in general is straightforward.

As a matter of terminology, let me also introduce the notion of irrelevance*, which applies to contexts in which relevant factors are allowed to change (theorem <III*>):

> $A \mathfrak{I} C \mid B\wedge X$ *and* $A \mathfrak{I} \neg C \mid B\wedge\neg X \rightarrow a \mathfrak{I}^* c \mid B\wedge x$*, if* A, $\neg A$ *are all possible values of* a*;* C, $\neg C$ *all possible values of* c*; and* X, $\neg X$ *all possible values of* x*.*

[20] There is no double arrow, since it is generally not required that all possible combinations in the background must actually occur.

In other words, a circumstance is irrelevant* to a phenomenon, if it is separately irrelevant (without *) to the phenomenon with respect to all different backgrounds in which only irrelevant conditions can change. From $A\ \mathscr{I}\ C \mid B\wedge X$ and $A\ \mathscr{I}\ \neg C \mid B\wedge\neg X$ follow four equivalent expressions: $A\ \mathscr{I}^*\ C \mid B\wedge x$; $A\ \mathscr{I}^*\ \neg C \mid B\wedge x$; $\neg A\ \mathscr{I}^*\ C \mid B\wedge x$; $\neg A\ \mathscr{I}^*\ \neg C \mid B\wedge x$. In short, this shall be expressed as $a\ \mathscr{I}^*\ c \mid B\wedge x$. Again, the notion irrelevance* is required, when irrelevance of conditions is stated in contexts where relevant conditions may change. In particular, when we required in the explication of homogeneity that all conditions in the background that can vary are causally irrelevant, this should actually have been irrelevant*.

Finally, we have theorem <IV>:

> *$A\vee D\ \mathscr{I}\ C \mid B \rightarrow A\ \mathscr{I}\ C \mid B\wedge d$ and $D\ \mathscr{I}\ C \mid B\wedge a$, where d can take on the values D or $\neg D$ and a the values A or $\neg A$.*
>
> *Furthermore, $A\vee D\ \mathscr{I}\ C \mid B \leftrightarrow A\wedge D\ \mathscr{I}\ C \mid B$.*

The second part follows from the definition of causal irrelevance plus the requirement that for $\vee$-connections, all possible combinations must be realized. It is then easy to show that for both $A\vee D$ and $A\wedge D$ the following states must be realized: $D\wedge A\wedge C\wedge B$, $\neg D\wedge A\wedge C\wedge B$, $D\wedge\neg A\wedge C\wedge B$, $\neg D\wedge\neg A\wedge C\wedge B$ leading to the stated equivalence. The first part follows from these states using theorem <III>. Here, the converse arrow does not hold since again it is not required that all possible combinations in the background are instantiated.

3b. Effect factors and alternative effects

While the logical structure of the causes in terms of inus-conditions is generally addressed in various accounts of causation, the logical structure of the effects is rarely examined. In the following, I propose the analogous notions of alternative effects and effect factors, which will become important in the next section when causal hierarchies are considered.

An *effect factor* can be defined in the following way:

> *C is an effect factor of the cause A with respect to a background B iff there exists an X such that A is causally relevant to $C\wedge X$ with respect to background B.*

This implies the following situations due to $\neg(C\wedge X) = \neg C\vee\neg X$: $A\wedge X\wedge C\wedge B$, $\neg A\wedge\neg X\wedge C\wedge B$, $\neg A\wedge X\wedge\neg C\wedge B$, and $\neg A\wedge\neg X\wedge\neg C\wedge B$. It follows that $A\ \mathscr{R}\ X \mid B\wedge C$ and $A\ \mathscr{R}\ C \mid B\wedge X$. Note that here the additional conditions in the background are effects and not causal conditions as was always assumed up to now. However, one could read this notation merely as shorthand for the requirement that the causal conditions determining these respective effects are held constant. As an example for the notion of an effect factor, consider how, during that night in Chicago in October 1871, the barn fire first spread to two immediately adjacent buildings. It follows that if the barn fire had been different that night, one or both of the buildings would not have caught fire or at least would have burnt in a different way.

While this may sound plausible, the notion of an effect factor is fraught with one additional difficulty. The four states listed above suggest a conclusion, which stands in contradiction with the combined premises of homogeneity and determinism: namely that a change in the effects may happen without a change in possible causes, e.g. when comparing $\neg A\wedge\neg X\wedge C\wedge B$ and $\neg A\wedge X\wedge\neg C\wedge B$. The somewhat obvious solution is to assume an inner causal structure for A according to which the change from $\neg X\wedge C$ to $X\wedge\neg C$ is determined. Thus, we have to presuppose that given a Boolean structure of the effect, the cause must have a corresponding structure that is at least as detailed. Let us call this *the causal structure assumption*.

An *alternative effect* can be defined in an analogous way:

> *C is an alternative effect of the cause A with respect to a background B iff there exists an X such that A is causally relevant to $C\vee X$ with respect to background B.*

Due to $\neg(C\vee X) = \neg C\wedge\neg X$, this implies the situations: $A\wedge X\wedge C\wedge B$, $A\wedge\neg X\wedge C\wedge B$, $A\wedge X\wedge\neg C\wedge B$, and $\neg A\wedge\neg X\wedge\neg C\wedge B$. It follows $A\ \mathscr{R}\ X \mid B\wedge\neg C$ and $A\ \mathscr{R}\ C \mid B\wedge\neg X$. As an example consider an arsonist walking around late at night with a single torch A. He sets either barn C or barn X on fire, but it may well happen that in the end both barns burn down. Again, the additional difficulty of a possible violation of determinism arises and is once more resolved by presupposing the causal structure assumption.

A further important notion is that of a *substitute effect*:

> *C is a substitute effect of the cause A with respect to a background B iff there exists an X such that A is causally relevant to $(C\underline{\vee}X)$ with respect to background B, where $X\wedge C$ cannot occur with respect to B under condition of homogeneity.*

This concept plays an important role in the analysis of concomitant variations where a member of a certain class of causes can individually cause any one but only one member of a certain class of effects. We have: $A\ \mathscr{R}\ X \mid B\wedge\neg C$ and $A\ \mathscr{R}\ C \mid B\wedge\neg X$. And again we have to assume a corresponding Boolean structure for the cause to avoid contradictions.

For causal irrelevance the following holds regarding effect factors:

$$A\ \mathscr{I}\ C\wedge X \mid B \leftrightarrow A\ \mathscr{I}\ C \mid B\wedge X \leftrightarrow A\ \mathscr{I}\ X \mid B\wedge C$$

After all, for all three expressions the following situations are required: $A\wedge X\wedge C\wedge B$, $\neg A\wedge X\wedge C\wedge B$. For alternative effects $C\vee X$ possible states are $C\wedge X$, $C\wedge\neg X$, and $\neg C\wedge X$, i.e. changes in effects are feasible and one therefore has to employ the notion of irrelevance*. Generally, if $a\ \mathscr{I}^*\ c \mid B\wedge x$ and $a\ \mathscr{I}^*\ x \mid B\wedge c$ then a will be irrelevant* to all possible inus-complexes consisting of C, X, and/or their negatives.

3c. Causal hierarchies and transitivity

A crucial requirement regarding the pragmatics of causation is that one can look at causal relations at a variety of different resolutions, both horizontally, i.e. when considering causal chains, and vertically, i.e. when looking at the formulation of causes and effects at various levels of coarse-graining, e.g. macro- and micro-levels.

Let us discuss the vertical relations first. Certainly, causal factors can be formulated in different levels of detail: for example, we might generally claim that the kicked-over lantern caused the fire while actually referring to a bundle of conditions that includes the kicked-over lantern itself, the spilled oil, the presence of hay, the absence of rain etc. Some basic rules for the coarse-graining of causes were already derived in Section 3a in terms of theorems <I>, <II>, and <IV>, demonstrating that an inus-complex can itself be considered causally relevant. Equally, in Section 3b, some results concerning the coarse-graining of effects were presented.[21]

However, often the exact inus-conditions are unknown. Rather, one may be confronted with just a bundle of conditions which apparently have an impact on a certain phenomenon. The following theorem <V> is supposed to cover those cases:

> *If one observes under homogeneity the following two states: $A_1, \dots, A_n$ present with C as well as $\neg A_1, \dots, \neg A_n$ present with ¬C (notation for this evidence situation: $A_1 \Diamond A_2 \Diamond \dots \Diamond A_n \ \mathcal{R}\ C \mid B$), then there is an inus-complex among the A that is causally relevant to C with respect to B.*

Obviously, the conclusion follows directly from the assumptions of homogeneity and determinism. Note that the invoked inus-condition may include both positive A's and negations of the A's. The theorem is important, because the set-up corresponds to a typical situation that arises in scientific practice, in which we observe presence and absence of a larger number of conditions but do not yet have any clue about the exact causal structure. In particular, it may very well be the case that some of the A are actually irrelevant to C.

To illustrate the theorem, consider the simple situation that we observe A_1 and A_2 present with C and $\neg A_1$ and $\neg A_2$ present with ¬C, then there are the following possibilities: i) $A_1\ \mathcal{R}\ C$ and $A_2\ \mathcal{I}\ C$; ii) $A_1\ \mathcal{I}\ C$ and $A_2\ \mathcal{R}\ C$; iii) $A_1 \vee A_2\ \mathcal{R}\ C$; iv) $A_1 \wedge A_2\ \mathcal{R}\ C$. This is an exhaustive set, since there are exactly four possibilities to choose C for the remaining situations $\neg A_1 \wedge A_2$ and $A_1 \wedge \neg A_2$.[22] Generalizations to a larger number of conditions are straightforward. Note that a statement analogous to theorem <V> does not hold for causal irrelevance. After all, there may for example be inhibiting causes among the A.

Of course, it should also be possible that we consider a phenomenon on various levels of coarse-graining on the side of the effects:

> *If one observes under homogeneity the following two states: A present with $C_1, \dots, C_n$ as well as ¬A present with $\neg C_1, \dots, \neg C_n$ (notation: $A\ \mathcal{R}\ C_1 \Diamond C_2 \Diamond \dots \Diamond C_n \mid B$), then given determinism there is an inus-complex that must include all C and to which A is causally relevant. According to the causal structure assumption there must be a corresponding inus-structure hidden in A.*

If we observe under determinism that A and ¬A are both present with $C_1, \dots, C_n$ under condition of homogeneity, we can of course conclude that A is irrelevant to all individual C with respect to a background that includes the presence of all other C's, respectively. While

[21] Note that statistical methods like randomized control trials or propensity score matching essentially rely on coarse graining as well, more specifically the construction of higher-level concepts, in this case populations, for which a causal analysis is feasible.

[22] It may of course also be the case that some of these situations are not observable or even impossible, which would leave the causal structure underdetermined.

we have thus far examined coarse-graining in the causes and the effects, separately, a combination of both brings no further difficulties.

Let us conclude the part about causal hierarchies by commenting on an issue that is widely discussed in the literature. According to the proposed account causal relationships may exist between various levels of coarse-graining as long as there is general consistency. In particular, there can be causation from the micro- to the macro-level and vice versa.

Let us next examine transitivity, which is closely connected with the feasibility of different horizontal resolutions. Many authors have considered transitivity a fundamental requirement of causal relationships, while others completely deny that it constitutes a property of causation, mainly in reaction to some pertinent counterexamples, which will be discussed below. Essentially, I will broadly follow David Lewis's suggestion that some causal notions are transitive while others are not. Notably, in the difference-making account causal relevance is a transitive concept, while causal irrelevance is not:

> *If X is causally relevant to Y with respect to B and Y is causally relevant to Z with respect to B, then X is causally relevant to Z with respect to B.*

This is just a consequence of the definition of causal relevance. As defined in Section 2a, causal relevance amounts to an "iff …, then …" relation (plus the causal asymmetry), which is trivially transitive. It follows:

> *If X is causally irrelevant to Z wrt B, then there is no Y such that X is causally relevant to Y wrt B and Y is causally relevant to Z wrt B.*

But note that causal irrelevance is not a transitive notion, i.e. it does not hold: If X is causally irrelevant to Y and Y is causally irrelevant to Z, then X is causally irrelevant to Z. Certainly, shooting at someone is irrelevant for the sun rising at that very moment and the sunrise will be irrelevant to the death of the person, but the shot is still very relevant to the death.

By contrast, the notions of causal factor, alternative cause and inus-condition are also transitive:[23]

> *If X is a causal factor (alternative cause, inus-condition) for Y wrt B and Y is a causal factor (alternative cause, inus-condition) for Z wrt B, then X is a causal factor (alternative cause, inus-condition) for Z wrt B.*[24]

This again follows from the respective definitions of causal factors, alternative causes, and inus-conditions. Let me provide the sketch of a proof for the transitivity of inus-conditions. Given that X is inus-condition for Y, i.e. there exist a D and E such that 'if and only if $(X \wedge D) \vee E$ then Y' and Y is inus-condition for Z, i.e. there exist an F and G such that 'if and only if $(Y \wedge F) \vee G$ then Z'. It follows that 'if and only if $(X \wedge D \wedge F) \vee (E \wedge F) \vee G$ then Z'. While this

[23] I am much indebted to Michael Baumgartner for apt criticism of earlier versions of my claim concerning the transitivity of inus-conditions. Among other things, I owe him the reference to the switching examples that are discussed further below.

[24] One should exclude well-defined cases of the type as the potassium salt-fire-death example discussed below.

expression already looks much as if X were an inus-condition for Z, it remains to be shown that X is indeed a non-redundant part of an alternative cause of Z.

Thus, assume that X is non-redundant for Y with respect to D and that Y is non-redundant for Z with respect to F. It must be proven that X is non-redundant for Z with respect to $D\wedge F$. Two problematic situations can arise: X can be redundant for Z with respect to $D\wedge F$, if either (i) X is fully contained as a factor in $D\wedge F$ or if (ii) X determines only such aspects of Y that are irrelevant for Z or (iii) a mixture of both cases. Since the third situation adds nothing essentially novel, it suffices to consider the first two.

Ad (i), there must be substantial overlap between the factors Y and F (and possibly between X and D). In this situation, $D\wedge F$ contains factors that have the same effect as X, which could either be X itself or an alternative to X with the same effect. In both cases X remains an inus-condition of F.

Ad (ii), this corresponds to the example of the potassium salts and the fire, which is discussed further below. In such cases, it must be ensured that all components of a sufficient condition have a causal function, in particular those factors that become explicit only through earlier links in a causal chain. Thus, the more precisely one knows what exactly is causally relevant for each link in a causal chain, the more plausible that inus-conditions will turn out transitive.

The transitivity of causal factors and of alternative causes follows in exactly the same way, if E and G are set to zero or D and F are set to one, respectively. In the case of causal factors it again has to be ascertained that causal factors earlier in the chain remain relevant to later links.

At this point, we need to address what some take to be the classic counterexample against the transitivity of inus-conditions, namely so-called switching structures (see in particular Baumgartner 2013, Fig. 3). Consider a train that travels from A to B either passing through C if a switch is in position E or passing through D if the switch is in position $\neg E$. Now, $A\wedge E$ is at least inus for C and C is at least inus for B, but E is not at least inus for B, since $(A\wedge E)\vee(A\wedge\neg E)=A$ is a sufficient and necessary condition for B. In other words, E does not seem part of a *minimal* sufficient condition for B, which by Mackie is defined as: “none of its conjuncts is redundant, no part of it […] is itself sufficient” (1980, 62). Rather, E in $A\wedge E$ seems to be redundant, since A is already sufficient for B.

In such cases, I would reply that one should qualify the type of redundancy involved here. After all, it is not the case that A is complemented with an arbitrary $\wedge(E\vee\neg E)$, which would certainly be absurd. Rather, E and $\neg E$ stand for different causal paths how B can be reached from A. One might want to speak of *non-redundancy post factum* since the additional factor designates the exact causal pathway on which a certain effect is reached. The importance of distinguishing pathways is further underlined when the causal structure is slightly altered. For example, one could specify the events B according to their causal history as B_C and B_D. Then E and $\neg E$ are non-redundant for B_C and B_D, respectively. Or, when an interference factor breaks one of the pathways either through C or D, then E or $\neg E$, respectively, will become non-redundant for B. Thus, when claiming the transitivity of inus-conditions, the possibility of such non-redundancy post factum must be taken into account as well. In this respect, the notion of non-redundancy used in the present article differs from Mackie’s viewpoint and that of many contemporary authors, especially those, who have argued for the intransitivity of inus.

Let us take a look at some further stock examples brought forward against the transitivity of causation (see e.g. Paul and Hall, 2013, Ch. 5). One of them is the following: A man is hiking in the mountains. Suddenly, a rock is set in motion and starts rolling towards him. He crouches down and survives. In this specific situation, it could seem that the falling rock somehow is causally responsible for the survival of the hiker, which certainly sounds counterintuitive.

The first important remark is that according to the account of causation presented here, transitivity only holds with respect to a fixed context or background. In many alleged counterexamples to transitivity, including the one just outlined, background-dependence is not guaranteed. This neglect is at least partly responsible for some of the putative paradoxes of transitivity.

Let us depict the formal structure of the mentioned example, which is in fact analogous to that of many similar counterexamples. The background B* broadly concerns someone hiking in the mountains on a trail with rocky grounds above him. The various events are: the falling of a rock R from somewhere above the hiker, the crouching C of the hiker, some additional conditions X under which the hiker crouches in case of a falling rock, in particular that he realizes it early enough, and finally either death D or survival $S = \neg D$ of the hiker. The following relations hold:

$R \,\mathcal{R}\, D \mid B^* \wedge \neg C$ <3>

$R \,\mathcal{I}\, S \mid B^* \wedge C$ <4>

$R \wedge X \,\mathcal{R}\, C \mid B^* \;\rightarrow\; R \,\mathcal{R}\, C \mid B^* \wedge X$ <5>

$C \,\mathcal{R}\, S \mid B^* \wedge R$ <6>

$C \,\mathcal{I}\, S \mid B^* \wedge \neg R$ <7>

We can already deduce from <3>, <5>, and <6> that there is no transitivity in terms of causal relevance. However, this is consistent with the general claims above since the background changes, while transitivity of causal relevance holds only in the case of constant background. It follows using theorems <I> and <II>:

$R \wedge \neg C \,\mathcal{R}\, D \mid B^* \;\leftrightarrow\; \neg R \vee C \,\mathcal{R}\, S \mid B^*$ (from 3, 4, 6, 7) <8>

$\rightarrow\; \neg R \vee (R \wedge X) \,\mathcal{R}\, S \mid B^*$ (from 5) <9>

The double arrow in <8> also results from negation <1>. According to <9>, R is formally an inus condition of S with respect to background B* corroborating the transitivity of the (at least) inus-relation with respect to a constant background. However, it is a strange kind of inus-condition. After all, it follows from <9> via the definition of alternative causes:

$$R\wedge X\ \mathscr{R}\ S \mid B^*\wedge R \qquad \langle 10\rangle$$

Obviously, the first R is redundant, since it is already fixed by the background. In fact, in <9> the second R can be dropped without changing the content of the statement.[25] This redundancy directly correlates with the fact that no background within B* exists, in which R is causally relevant to S, only backgrounds in which ¬R is causally relevant to S. Certainly, this situation implicitly contributes to the wide-spread refusal to see R as a cause for the survival. Again, the justification for seeing R as non-redundant is that it designates a specific causal path, which requires $R\wedge X$ in order for the crouching C to happen.

The sketched formal analysis in terms of the difference-making account can clarify to what extent transitivity holds and why R is such a peculiar inus-condition in this example. By the way, Lewis presents a closely related analysis of an analogous case (2000, 194-195) claiming that our reluctance to accept R as a causal factor stems from several issues: (i) R and ¬R both appear as at least inus-conditions (Lewis is not using the term), which may lead to confusion, but in principle there is nothing wrong with this double role; (ii) in most cases, the falling of a rock leads to death and not to survival, i.e. judging by the number of possible realizations, the presence of R all in all increases the probability of death for the hiker; (iii) falling rocks do not matter for a careful hiker, who will always survive: if $X = 1$, then <9> becomes $\neg R \vee R = 1\ \mathscr{R}\ S \mid B^*$, i.e. the hiker always survives.

There are other supposed counter-examples to transitivity. A well-known one is the following that was already alluded to: suppose that potassium salts P put into the fire place are a cause of the fire turning purple, and that the fire F is a cause of the eventual death D of a person. Are the potassium salts then a cause for the death? Plausibly not. The seeming paradox can be easily dissolved within the framework of the difference-making account. In the case that we are largely ignorant about the working of fires, we may find that purple fires are indeed causally relevant to the death of a person, i.e. $P\lozenge F\ \mathscr{R}\ D \mid B^*$. However, by taking into account further instances, it can be established that it was not the color of the fire that killed but other properties: $P\ \mathscr{I}\ D \mid B^*\wedge F$ and $P\ \mathscr{I}\ \neg D\ \ |B^*\wedge\neg F$, therefore $p\ \mathscr{I}^*\ d \mid B^*\wedge f$ by theorem <III*>. The addition of potassium is not an inus-condition for the fire as a cause of death, except in the artificial non-causal sense: $(P\wedge F)\vee(\neg P\wedge F)\ \mathscr{R}\ D \mid B^*$. Crucially, we do not have non-redundancy post factum in this case since P and ¬P do not constitute different pathways leading to D, i.e. pathways that can be independently interrupted. In summary, there is no transitivity, because the potassium salts are causally irrelevant to the death.

A further class of counterexamples concerns long chains of inus-conditions. For instance, several authors have pointed out that the birth of a person is formally an inus-condition to his/her death, which seems awkward. Several aspects play a role to resolve this apparent paradox. First, the relationship between birth and death is to some extent of definitional character: one can only die, if one was born, and those who are born must all eventually die. Second, for long causal chains in terms of inus-conditions the causal nature eventually wears off. After all, the range of backgrounds with respect to which the remote condition is causally relevant becomes increasingly smaller as the following argument shows. Given $A\vee(C\wedge D)\ \mathscr{R}\ E$

[25] Note also that the requirement stated in Section 3a that all feasible combinations must be realized is clearly not possible for the expression in <9>, e.g. $\neg R \wedge (R\wedge X)$ is not realizable.

$| B$ and $F \vee (E \wedge G)\ \mathcal{R}\ X | B$, then $C\ \mathcal{R}\ E | B \wedge \neg A \wedge D$ and $C\ \mathcal{R}\ X | B \wedge \neg A \wedge D \wedge \neg F \wedge G$. Clearly, the further down the causal chain we move, the more restrictive the backgrounds become, where an inus-condition remains causally relevant. Consequently, specific circumstances at the birth of a person are only in a very restricted way causally relevant to specific aspects of the death of a person. Note that such an argument cannot be construed for chains purely in terms of causal relevance, but those almost never exist, not least due to possible external interventions.

This discussion is closely related with an important theme, which we can only briefly touch upon in the current article, namely the pragmatics of causation regarding in particular the question under which circumstances inus-conditions are colloquially singled out as “causes”. Certainly, stability with respect to background, manipulability of the causal conditions and available contrast classes are among the principal criteria, why sometimes certain events or properties are distinguished as causes and others that also fulfill the formal criteria are not.

3d. Functional dependencies

When causation came under attack towards the end of the 19th century, one of the main objections concerned what Ernst Mach called the pharmaceutical character of causation, where a piece of this leads to a piece of that. Historical approaches to causation did not seem suited to account for the causal character of some of the most fundamental laws of science, for example the axioms of mechanics, which obviously are not formulated in terms of the presence or absence of certain factors, but rather as functional dependencies. Bertrand Russell, in his famous critique of the notion of cause (1913), concurred that the old concept of causation should be replaced by functional laws, in particular by differential equations.

Such criticism is not entirely fair, since many approaches to causation in the 19th century and before included methods to account for the causal character of functional relationships. Maybe, the most important is the method of concomitant variations in Mill’s approach: “Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation.” (1886, 263) Bacon’s table of degrees can be seen as a precursor to this method.

Admittedly, the quantitative method of concomitant variations sits somewhat apart from the other methods in Mill’s inductive framework, which are qualitative concerning the presence and absence of factors. Several authors have therefore tried to establish a closer connection between quantitative and qualitative induction (e.g. von Wright 1951, 154; Skyrms 2000, Sec. V.9). I will broadly follow this line of approach to show how the method of concomitant variations can be understood as a special case of the method of difference.

Consider two finite sets of events of the same size n, a set of cause events $\{A_1, \ldots, A_n\}$ and a set of effect events $\{C_1, \ldots, C_n\}$. These could for example be distance elements, velocity elements or small amounts of a solid or liquid etc. Let there be a natural ordering among the C, but not among the A. One can then construct integral events $\mathbf{C}_m = \Lambda_{i=1,\ldots,m} C_i \wedge \Lambda_{i=m+1,\ldots,n} \neg C_i$ and $\mathbf{A}_{m,\{m \text{ from } n\}} = \Lambda_{i=\{m \text{ from } n\}} A_i \wedge \Lambda_{i=K(\{m \text{ from } n\})} \neg A_i$,where {m from n} denotes a specific subset of size m and K({m from n}) denotes the complement of this subset.[26] These integral events are aggregated distances, velocities, amounts of stuff etc. We have some rudimentary version of concomitant variation if $\mathbf{A}_{m,\{m \text{ from } n\}}\ \boldsymbol{\mathcal{R}}\ \mathbf{C}_m \mid B,\ \forall\ m$ (i.e. relevance of at least one subset {m from n} for all m).

Now, let this relation hold for any subset, i.e. the A_i are completely interchangeable regarding their effect on the phenomenon C. With $\mathbf{A}_m=\underline{V}_{\text{all possible subsets}}\mathbf{A}_{m,\{m \text{ from } n\}}$ we have: $\mathbf{A}_m\ \boldsymbol{\mathcal{R}}\ \mathbf{C}_m \mid B$.[27] In reasonable cases, this family of causal relations can be idealized as a continuous linear function based on the indices m. A simple example is the supposedly linear relation between the expended amount of fuel and the distance covered by an (idealized) car. In principle, it does not matter, which part of the fuel is spent for driving which distance. Note that there may be interchangeability on the side of the effects as well, but this brings no additional difficulties.

Several complications are conceivable. For example, at first glance the account just presented seems to allow only for linear relationships $m(A) \rightarrow m(C)$, where m denotes a measure ascribed to the phenomena on the basis of the respective indices. But of course, arbitrary functional relationships are feasible by allowing for transformations of the measures for both the source and the target phenomena essentially corresponding to a change in indices. For discrete phenomena, there exists a natural measure by just counting the number of elementary causes and effects. By contrast, for continuous phenomena, such a natural measure does not normally exist and the measure is usually fixed by pragmatic considerations aiming at overall simplicity of various interconnected functional relationships.

Furthermore, the A_i may not be strictly interchangeable, i.e. not every A_i may be able to cause a specific C_j. For example, a certain A_i may appear only after other A's are already present. But those cases can easily be covered by the given framework. The same holds for functions with several variables. Finally, various limiting processes may occur. First, the number of cause and effect events may be infinite. Second and more importantly, the cause events and the effect events may become infinitesimally small. The latter process can be understood on the basis of the outlined discrete framework if the individual A's and C's are themselves considered as aggregate quantities etc. Since human experience is in general finite, limits involving infinity should be interpreted as idealizations.

In summary, the method of concomitant variations is a special case of the method of difference applied to certain classes of cause and effect events. As already von Wright has stressed, quantitative laws thus turn out a subspecies of qualitative laws (1951, 83). As a final remark, note that the resulting functional relationships convey more information than the mathematical laws known from the sciences, which generally lack a causal interpretation, most importantly a distinction between cause and effect variables.

[26] Λ, V, and $\underline{V}$ shall be understood in analogy to the $\sum$-sign denoting a sum over several elements, e.g. $\Lambda_{i=1,2} C_i = C_1 \wedge C_2$. Remember that $\underline{V}$ denotes an exclusive or, where only one option can be realized at a time.

[27] For m=2, we have: $\neg A_1 \wedge \neg A_2\ \boldsymbol{\mathcal{R}}\ \neg C_1 \wedge \neg C_2 \mid B$; $(\neg A_1 \wedge A_2)\ \underline{\vee}\ (A_1 \wedge \neg A_2)\ \boldsymbol{\mathcal{R}}\ C_1 \wedge \neg C_2 \mid B$; $A_1 \wedge A_2\ \boldsymbol{\mathcal{R}}\ C_1 \wedge C_2 \mid B$

4. Objections and criticism

I will in the following discuss a number of conceivable objections against the difference-making account. Given that the proposed approach to causal counterfactuals relies on the method of difference, the usual criticism of Mill's methods has to be considered. It is addressed mostly by offering various refinements to these methods. A second group of objections concerns those traditionally raised against counterfactual accounts. This includes the plethora of counterexamples that are discussed in the contemporary literature on causation, e.g. cases of preemption or overdetermination. Furthermore, since the difference-making account aims to be a monistic approach to causation, it has to make sense of the core intuitions that are considered fundamental in other approaches. In particular, it has to take a stance regarding the role of interventions and of mechanisms in causal reasoning. Finally, some problems arise due to certain peculiarities of the difference-making account itself, in particular in connection with the notion of causal irrelevance.

4a. Objections to Mill's method of difference

A general worry that may come to mind with respect to the outlined account is that it is guilty of mixing up metaphysics and methodology of causation, i.e. the question what causation is with the question how causal relationships can be discovered in the world. In contrast, I would argue that these issues should not and cannot be fully separated. The definition of causation must be informed by methodological considerations and conversely, the definition should to some extent imply the methodology.[28] Indeed, the proposed difference-making account of causation constitutes a direct counterpart to the account of induction broadly outlined in Pietsch (2014), which stands in the tradition of eliminative induction and Mill's methods type of reasoning. The method of difference and the strict method of agreement as formulated there respectively determine causal relevance and irrelevance according to the definitions given in Section 2a.[29]

Consequently, a first important class of objections concerns those traditionally raised against eliminative induction and in particular the method of difference, most of which apply to the strict method of agreement as well. These objections are usually targeted at Mill's original formulation of the method of difference, which continues to be the most influential: "If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon." (1886, 256)

One serious worry concerns the applicability of this method. After all, it seems generally impossible to vary only a single one of the circumstances. Rather, always a large, plausibly infinite number of circumstances changes along with the one whose influence is explicitly

[28] Compare Nancy Cartwright's viewpoint: "If causal claims are to play a central role in social science and in policy – as they should – we need to answer three related questions about them: What do they mean? How do we confirm them? What use can we make of them? The starting point for the chapters in this collection is that these three questions must go together. For a long time we have tended to leave the first to the philosopher, the second to the methodologist and the last to the policy consultant. That, I urge, is a mistake. Metaphysics, methods and use must march hand in hand." (2007, 1)

[29] This topic will be taken up in Section 5a.

examined. Thus, strictly speaking Mill's construal of the method of difference turns out inapplicable. However, the refined account of causal relevance outlined in the present essay resolves this issue by specifying which circumstances may change, namely irrelevant circumstances and those that lie on a causal chain through the purported cause to the phenomenon (cf. the definition of homogeneity in Section 2b; cp. also Section 5a). These requirements, although they can never be definitely established in concrete applications, are nonetheless much more viable than Mill's rather crude premise.

Another aspect that is already alluded to in Mill's own writings is that the method of difference according to his formulation can only handle situations in which a single circumstance is causally responsible for a phenomenon. For example, it cannot identify necessary but insufficient factors, since these generally do not make a difference, e.g. a short-circuit does not cause a fire in the absence of inflammable material. Relatedly, Mill's formulation of the method of difference cannot identify causal factors in the presence of inhibiting causes. For example, lightning should plausibly count as a cause for burnt-down houses even though fire-extinguishers sometimes prevent fires from destroying buildings.

All these difficulties are resolved in the difference-making account by introducing a refined terminology that carefully distinguishes various types of causal notions, in particular causal relevance, causal irrelevance, causal factors, and alternative causes, and that is capable of replicating the inus-logic of causal conditions.

4b. Objections against counterfactual accounts

A further class of possible objections concerns those that are traditionally raised against counterfactual approaches like that of David Lewis. One topic that Lewis discusses extensively regards those infamous examples of plural causation that had already turned out problematic for regularity accounts, in particular instances of overdetermination and of preemption.

In cases of overdetermination, there are (at least) two causes present that are independently sufficient for a phenomenon. A classic example is the simultaneous assassination by two marksmen. Why this situation is problematic for counterfactual accounts can be easily grasped. If one of the marksmen had not pulled the trigger, the poor fellow would still have died from the shot of the other. Therefore a basic counterfactual analysis would not identify each of the shots as a cause of death, which seems counterintuitive.

In cases of preemption, two causes are again independently sufficient for a phenomenon, only that now one of them prevents the other from becoming causally relevant. Lewis further distinguishes early from late preemption (1986, Sec. E). In the former, one of the causes is prevented from being relevant, before the phenomenon actually happens. A typical example concerns the desert traveler with two mortal enemies. One of them poisons the drinking water, the other later pours the water out. Both of these deeds would kill the traveler, but the pouring out of the water led to his dying of thirst, while at the same time preventing the traveler from being poisoned.

In late preemption, one of the causes is hindered from being relevant by the fact that the considered phenomenon has already been produced by the other cause. Consider two kids throwing rocks at a bottle. The first throw shatters the bottle, while the second would have shattered it as well, if the first had not already done so. Note that for reasons beyond the scope of the present article, the distinction between early and late preemption is especially important in Lewis's account. In fact, while his solution for early preemption is rather convincing, he always struggled with late preemption. Recall also that preemption cases constitute one of the reasons—along with ensuring transitivity—why Lewis's definition of causation employs causal chains.

Within the framework of the difference-making account, all these cases of plural causation can be easily treated. In a way, the mentioned problems only arise when not being sufficiently precise about the type of causal dependence that is considered. For instance, preemption can be accounted for on the basis of alternative causes, where one of the alternative causes (or a causal factor of it) constitutes an inhibiting factor of the other alternative cause (or a causal factor of it). In the case of the desert traveler, who either dies D from poisoning P or from pouring out the water W, we clearly have such an alternative-cause structure according to the definition of Section 3a: $P \mathcal{R} D \mid B \wedge \neg W$; $W \mathcal{R} D \mid B \wedge \neg P$; $P \mathcal{I} D \mid B \wedge W$; $W \mathcal{I} D \mid B \wedge P$. To see the presence of an inhibiting cause, subscripts need to be introduced distinguishing different types of death, i.e. by poisoning D_P and by thirst D_W: $P \mathcal{R} D_P \mid B \wedge \neg W$; $P \mathcal{I} \neg D_P \mid B \wedge W$. According to the corresponding definition in Section 3a, W constitutes an inhibiting factor counteracting P with respect to D_P. An analogous causal structure can be identified in the example of the two kids throwing rocks at a bottle. Thus, late and early preemption can be treated in largely the same manner, essentially because, in contrast to Lewis's approach, the analysis does not rely on the notion of causal chains. Finally, in situations of overdetermination we are also dealing with alternative causes except that none constitutes an inhibiting factor for the other—which implies that both types of events may co-occur, e.g. being shot to death by both marksmen at the same time.

While cases of overdetermination and preemption allegedly show that counterfactual definitions are not necessary for causality, other authors have claimed that they are in fact not sufficient either. Notably, Jaegwon Kim has argued that various kinds of non-causal relations fulfill the basic counterfactual definition as well—in particular logical or analytical dependence (if George had not been born in 1950, he would not have reached the age of 21 in 1971), parthood (if I had not written 'r' twice in succession, I would not have written 'Larry'), interdependent actions (if I had not turned the knob, I would not have opened the door), and non-causal determination (if my sister had not given birth at t, I would not have become an uncle at t) (Kim 1973). A typical rejoinder has been to specify the kind of relata figuring in causal relations. For example, Menzies requires that causally connected events are distinct, i.e. "not identical, neither is part of the other, and neither implies the other" (Menzies 2014, § 2.1).

Let me briefly sketch a different solution. The difference-making account can indeed be applied to various kinds of necessity, in particular to the physical necessity (but logical contingency) of causal relations and also to the conventional necessity of definitional relations. In fact, all the above-mentioned counterexamples seem to some extent of definitional nature, although not much hinges on allowing for further kinds of necessity.[30]

Now, the fact that the difference-making approach works for both causal and definitional necessity should be seen as a virtue rather than a vice. After all, it fits with well-established skepticism concerning a definite analytic-synthetic distinction. Indeed, in many parts of science and everyday knowledge, definitional and empirical relations cannot clearly be held apart, which essentially is implied by the Duhem-Quine thesis of confirmational holism. To a certain extent, it is always possible to shift the empirical content from one part of the relevant regularities to another. For some ornithologists, it may be part of the definition that all ravens are black, for others it may be a matter of empirical fact. Of course, such flexibility in assigning empirical import would not be possible, if the formal structure of definitional and of causal-empirical relations would be radically different.

A further problem for counterfactual approaches consists in the fact that most of them presuppose determinism, i.e. that all considered phenomena are fully determined by their circumstances. At least in simple versions, counterfactual approaches therefore fail to make sense of statistical causal relationships like ‘smoking causes lung cancer’. After all, it is not universally true that if someone had not smoked, he/she would not have died of lung cancer. This issue is too complex to be addressed in the space available here. Elsewhere, I have argued that statistical and even indeterministic relationships can be covered by the difference-making account of causation if it is extended by a causal interpretation of probability (see also Sec. 5b).

4c. Mechanisms and interventions

The difference-making account aims to be a monistic account of causation. Therefore, it should be able to make sense of typical intuitions that are the focus of other approaches, in particular: (a) Causal relationships can be corroborated by evidence concerning the underlying mechanisms linking causes with effects; (b) causal knowledge must be acquired experimentally by examining the impact of interventions on a phenomenon. Obviously, these claims correspond to two major classes of interpretations of causality, namely mechanistic or process accounts, on the one hand, and the recently very popular interventionist accounts, on the other hand.

The prominent role of interventions for determining causal relationships can easily be understood from the perspective of the difference-making account. After all, controlled interventions are ideally suited to yield the type of evidence required for establishing causal relationships according to the difference-making account. They generate in close temporal succession two situations which usually differ in a relatively small number of circumstances and hopefully only in irrelevant circumstances except those that were explicitly changed by the intervention.

While accounting for the evidential power of interventions, the difference-making approach avoids the most troubling objections against interventionist accounts. Most importantly, it

[30] Note that in particular relations of parthood need not be definitional, but can also be of causal nature.

does not *require* interventions to determine causal relationships and thus evades the problem that such approaches by definition fail in contexts where interventions are impossible, e.g. to establish the supposedly causal nature of the regularities governing the movement of stars and planets. In comparison, the difference-making account allows for a much broader range of evidence for causal relationships.[31] While the notion of an intervention apparently presupposes that the two instances which are compared are temporally and spatially close and that they belong to one and the same system, according to the difference-making account the compared instances can be arbitrarily far away from each other both in temporal and in spatial terms and can belong to different systems. For the proposed account it does not matter whether the required variational evidence results from intervention or mere observation.

A further problem concerns the question how exactly interventions should be conceptually framed. Several authors have suggested an anthropomorphic account resulting in unsurmountable difficulties. Are, for example, only humans able to effectuate interventions or are animals as well, not to speak of technical artefacts or even natural objects? In the face of such irresolvable challenges, influential authors like Woodward (2003) and Pearl (2000) have introduced less demanding definitions in terms of "'surgical' change": "the proposals variously speak of an intervention on X as breaking the causal connection between X and its causes while leaving other causal mechanisms intact or as not affecting [the target variable] Y via a causal route that does not go through X." (Woodward 2008, §5-6) Pearl (2000) develops a special calculus that models interventions in this sense, defining the "causal effect" of X on Y in terms of the value of Y given that X is explicitly set to x: P(y|do(X=x)).

Still, both Woodward and Pearl rely on a special ontological category of an intervention to establish their notion of causation. As mentioned, one crucial feature is that an intervention on a variable X cuts off other possible influences for this variable as well as causal influences on the phenomenon Y that are not mediated by X. But this somewhat peculiar requirement contradicts basic human experiences. When intervening in the world, it rarely seems the case that all other possible causal influences are disrupted. Rather, even in simple situations, e.g. when kicking a ball, all kinds of causal influences remain intact, including the pull of gravity, air pressure etc.[32] Thus, in addition to general worries about ontological parsimony, even Woodward's and Pearl's weaker reading of interventions in terms of surgical change does not seem realistic for many contexts of application.

The difference-making account can make sense of interventions, while not relying on the controversial idea that causal connections are broken. Indeed, it does not have to introduce a distinct ontological category at all. The complex technical features of an intervention, as for example laid out by Woodward (2013, §6), can nevertheless be reproduced. They essentially follow from the requirement of homogeneity of context. Conceptually speaking, external interventions in a system can be modeled within the difference-making account in terms of changes in the background such that some previous relations of causal relevance or irrelevance cease to be valid. For instance, the disruption of a causal influence can be accounted for if the corresponding variable is held constant in the causal background or if the

[31] In spirit, this perspective is quite similar to Federica Russo's 'invariance under changes' approach (2014).
[32] I owe this point to Mathias Frisch.

background is restricted in a way that the variable becomes causally irrelevant. Note further that the proposed account can address issues of external validity by examining a phenomenon with respect to various backgrounds, while typical interventionist approaches that lack the concept of a causal background must generally and wrongly conclude that a causal relationship once established will continue to hold in all contexts.

Accounts in the interventionist tradition mostly use two kinds of representations for causal relationships, structural equations and DAGs, i.e. directed acyclic graphs (cf. in particular Pearl 2000). While these representations are compatible with the difference-making account, the Boolean logic of circumstances employed in the latter is more general. As argued in Pietsch (2015), this logic works well in non-parametric contexts in which (exact) structural equations cannot exist since by definition the relation between cause and effect cannot be described with a finite amount of parameters. As another example, directed acyclic graphs presuppose a static picture of causal relationships and thus cannot adequately deal with strong context-dependence in complex phenomena. Also, the graph-theoretic framework fails to account for the inus-logic of causal conditions. In particular, by relying solely on directed links between variables, it does not distinguish between causal factors and alternative causes.[33]

In summary, interventionist approaches introduce a problematic distinction between situations, where variables are observed, and others, where they are set by intervention. By contrast, the difference-making account gets along without a sharp division, while still acknowledging the extraordinary epistemic power of interventions for detecting causal relationships.

Mechanisms constitute another central concept for causal reasoning. In fact, several authors go as far as arguing for a dual nature of causation, part interventionist and part mechanistic. By contrast, I will now argue that mechanisms can be integrated into the difference-making account in straightforward manner. Here, they can broadly be framed along the following lines: “A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon.” (Illari and Williamson 2012, 120) For the present discussion, I would add that mechanisms should be understood in terms of causal (maybe also definitional) relations and introduce context- or background-dependence. Essentially: “A mechanism for a causal relationship between a set of circumstances and a phenomenon in a certain context consists in more fine-grained causal dependencies that together account for this relationship.”

Such causal mechanisms can be interpreted as filling in the details of a more coarse-grained causal picture. Consider for example a causal dependence between a circumstance A and a phenomenon C with respect to a background B. A simple mechanism would consist in a number of intermediary circumstances A_1, …, A_N, such that A is causally relevant to A_1, A_1 to A_2, ..., A_N to C, all with respect to the same background B. Of course, one can easily imagine more complex dependencies relying on various causal factors and alternative causes.

[33] Baumgartner and Graßhoff (2004) suggest an extension of causal graphs for this purpose.

From the perspective of the difference-making account, the individual dependencies constituting the mechanism would again be ordinary causal relationships that themselves can be accounted for in terms of difference making. This explains the most important lesson about mechanisms in the context of causal reasoning, namely that knowledge of mechanisms can to some extent improve the confidence in coarse-grained causal relationships, essentially because it allows taking into account further evidence, namely evidence for the intermediary steps. This is particularly effective in the case of physical mechanisms since the corresponding mechanical laws are so well established. Note finally that sometimes there may be no independent evidence for the coarse-grained relationship itself. Then, the confidence in the latter derives fully from the evidence for the intermediary steps.

In contrast to mechanistic approaches, the difference-making account takes the notion of cause as fundamental and the notion of mechanism as parasitic on causation instead of the other way around. Therefore, the latter is not susceptible to the familiar objections against mechanistic approaches. Most importantly, the usual problems with the concepts of process and mechanism do not arise. A causal mechanism as outlined above does not have to transmit structure, marks, or signals, and it does not have to be reducible to the laws of physics. Furthermore, a mechanism in the proposed sense need not be local, while nevertheless the fruitful role of locality for inductive reasoning can be accounted for (cf. Section 5b). Also, there can be mechanisms for causation by omission, since the absence of a circumstance may well be causally relevant to a phenomenon in the sense of difference making and one can easily imagine a more fine-grained picture corroborating such a causal relationship. Finally, the difference-making account does not face the foundationalist problem of mechanistic approaches that regularities at the most fundamental ontological level cannot themselves be accounted for in terms of mechanisms and thus they cannot be causal. After all, it is difference making that is essential for causality, rather than the existence of mechanisms.

Thus, like in the case of interventions, the approach proposed in this essay can account for the immensely fruitful evidential role of mechanisms, while not being affected by the problems that result from putting the notion of mechanism at the core of an interpretation of causality.

4d. Objections specific to the difference-making account

Let us finally address a number of objections that are more or less specific to the difference-making account. Crucially, it exhibits a circularity in the definitions of causal relevance and irrelevance, which consists in the fact that these very notions again figure in the requirement of homogeneity of background. More exactly, the two instances that are compared when evaluating causal relevance and irrelevance should differ only in circumstances that are themselves causally irrelevant except for those that lie on a causal chain through the examined circumstance to the phenomenon. However, the latter relevance and irrelevance claims concern causal relationships that are different from the one that is explicitly examined. Therefore, the supposed circularity just comes down to a consistency requirement. At least in principle, causal claims can be established merely on the basis of an (often very large) number of instances with varying circumstances.

Thus, the difference-making account allows reducing causal claims to non-causal evidence in terms of instances under varying conditions. This constitutes a crucial difference compared with interventionist approaches, which also exhibit a circularity in the definition of causation in that intervention is itself a causally tainted term. For example, in both Pearl's and Woodward's versions of interventionism, the notion of intervention itself draws upon various causal concepts, e.g. the interruption of causal influences on the variable X whose influence is examined or the causal relation between the intervening variable I and variable X (cf. Section 4c). In contrast to what I have argued above for the difference-making account, interventionist approaches are generally non-reductive due to this circularity, i.e. they do not allow "the possibility of a reduction of causal claims to claims that are non-causal" (Woodward 2008, §1).

A further difficulty of the difference-making account regards the fact that in certain contexts, circumstances will be identified as causally relevant that one would not ordinarily consider as such. This typically happens in common cause structures, which are discussed extensively by Mackie (1980, 83-86). Consider for example two different effects of a disease, e.g. a fever C and a rash D as symptoms of roseola A. Now, in specific contexts B, there may always be covariation of fever and rash in the presence of the roseola virus with, say, the fever preceding the rash. It seems that one is forced to accept the counterintuitive conclusion that with respect to such B's the fever is causally relevant to the rash.

At this point, it helps to recall the notion of a common cause as explicated in Section 3a. Roughly, A is a common cause of C and D with respect to a background B, (i) if in all backgrounds within B, in which C $\mathcal{R}$ D (assuming that C is always temporally prior), we have A $\mathcal{R}$ C and A $\mathcal{R}$ D, and (ii) if in all backgrounds within B, in which C is caused by conditions other than A, then C $\mathcal{I}$ D and/or C $\mathcal{I}$ ¬D.[34]

Consider the following set-up, which fulfills these requirements: let conditions A and E be the only alternative causes for C, condition A shall be causally relevant to phenomenon D, and E causally irrelevant to D, all with respect to a background B. Assume further that C is temporally prior to D. It follows that A is causally relevant to C and D with respect to B∧¬E, that C is causally relevant to D with respect to B∧¬E, but that C is irrelevant to ¬D with respect to B∧¬A. In the example, E could be another virus, e.g. the flu, which also leads to fever, but not to a rash. In the absence of roseola, fever is irrelevant to the rash, while in the absence of flu, fever is causally relevant to the rash.

Again, the latter is a strange result, since intuitively causal relevance between different effects should not exist in the case of a common cause structure. In particular, this consequence seems to contradict the conception that causal relationships can be employed to manipulate phenomena since in general one effect cannot be used to change another effect. However, strictly speaking with respect to background B∧¬E, by manipulating C phenomenon D is also changed—essentially because C can only be manipulated via A. The actual common cause structure is then established by showing that in various other contexts, when C and D are due to other conditions than A, they are irrelevant for each other.

[34] Assuming that there is only one common cause.

In fact, this example illustrates well the development of causal language when increasing evidence is gathered. There is nothing contradictory in the fact that causal structure changes when the background is extended (e.g. from the causal relevance of a condition to being an alternative cause) and that with respect to certain evidence the causal structure may be underdetermined. In complex situations like Mackie's notorious example, in which the sounding of the Manchester hooters could seem a cause for the Londoners to return from work, the causal structure can only be resolved by taking into account a sufficiently large context of other causal relations. With very limited evidence, it may seem that there is causal relevance, but when we learn about the spatial distance separating the two events, the various temporal relationships, the causal laws regarding sound propagation and so on, a causal relevance between the Manchester hooters and the London workers can be excluded with respect to sufficiently general backgrounds.

Another objection is specific to the notion of causal irrelevance of the difference-making account, namely that the latter obviously depends on measurement accuracy. Indeed, there may be causal relevance that is just not detectable due to the small size of the influence. My suggestion is to bite the bullet on this one. Arguably, this feature constitutes another aspect of the pragmatics of causation. Indeed, causal irrelevance to some extent depends on the choice of representation, and measurement accuracy should be considered one aspect of this. Certainly, paradoxical situations can arise: for example when several undetectable contributions add up to a measureable effect. In such cases the representation just needs to be adjusted to recover consistency. Obviously, a certain amount of simplification and idealization will always be required for the causal representation of empirical phenomena, since otherwise it may well turn out that everything depends on everything else.

A final remark concerns the requirement of background dependence in the difference-making account. Since this implies a ceteris-paribus character for *all* causal laws, those scientific laws that claim to be truly universal cannot have a distinct causal nature. This holds in particular for the fundamental axioms of physical theories like Newton's laws of mechanics. The suggestion is that such axioms acquire a causal character only when they are restricted and applied in a specific context. While this may sound strange to some working scientists, it fits well with ideas of Mach, Poincaré, Duhem, and others, who have argued that the fundamental laws of physics are conventions or implicit definitions rather than empirical (and thus causal) statements. Recall also that there has been an ongoing debate in philosophy to what extent causation plays a role in physics at all. Doubts have been raised from various perspectives. For example, some interventionists have argued that interventions are impossible when fundamental physical theories apply to the whole universe (Woodward 2013, §12), presumably because the notion of interrupting causal influences does not make sense anymore. Russell and Mach have argued that the laws of physics cannot be causal since they are functional and this does not go well with regularity accounts requiring constant conjunction of conditions. The difference-making account gives yet another answer, namely that the laws of physics cannot be causal since they do not exhibit a ceteris-paribus character and thus they cannot be derived in a process of eliminative induction which according to the proposed account constitutes the only way to generate causal knowledge.

5. A fresh look on old problems in epistemology

5a. Mill's methods revisited

As already emphasized, the difference-making account of causation should be seen as a direct counterpart to the framework of eliminative induction broadly outlined in Pietsch (2014), which stands in the tradition of Mill's methods type of reasoning. Notably, the two fundamental inductive methods delineated there, the method of difference and the strict method of agreement, directly correspond to the definitions of causal relevance and irrelevance, respectively, as given in Section 2a.

Thus, we now turn from the conceptual question, what causation is to the methodological question how causal relationships can be determined via inductive methods. The proposal is to base causal inferences exclusively on two fundamental methods, the *method of difference* to determine causal relevance and the *strict method of agreement* to determine causal irrelevance:

> *Method of difference: If two instances with the same background B are observed, which differ in a potentially relevant circumstance A and in a change of phenomenon C, then A is causally relevant to C with respect to background B, if (i) B guarantees homogeneity and (ii) if A is causally prior or simultaneous to C.*
>
> *Strict method of agreement: If two instances with the same background B are observed, which differ in a potentially relevant circumstance A, while the phenomenon C remains unchanged, then A is causally irrelevant to C with respect to background B, if (i) B guarantees homogeneity and if (ii) A is causally prior or simultaneous to C.*

These two methods exhaust the fundamental methodology of causal inference. This is obvious for the case when homogeneity is fulfilled, since then the method of difference and the strict method of agreement are exactly complementary. Otherwise, there may be relationships between A and C with respect to non-homogeneous backgrounds, which constitute neither causal relevance nor causal irrelevance (cf. Section 2b). But such situations can be covered by introducing further causal terminology that is itself defined entirely on the basis of causal relevance and irrelevance, including most importantly causal factors, alternative causes, and inus conditions. With these further notions, all types of causal dependencies can be covered under two assumptions: that the considered set-up is deterministic and that one is not yet dealing with functional relationships, but only with the presence or absence of factors. However, extensions of the framework covering statistical and functional relationships are straightforward (cp. Sections 3d and 5b).

Let me briefly compare the proposed approach with other prominent expositions of eliminative induction. The most popular and most cited account until today remains Mill's own formulation, which is usually referred to as Mill's methods. Most of these methods had already been known at least since the middle ages and there were previous influential systematizations, most notably Bacon's and Herschel's. However, it seems fair to say that Mill's presentation is more systematic and concise, introducing the following five methods or canons of induction: the method of agreement, the method of difference, the joint method of agreement and difference, the method of residues, and the method of concomitant variations.

Mill's method of difference, which was already stated in Section 4a, is quite similar to the formulation above, which only adds some refinements, in particular the introduction of background-dependence and the requirement of homogeneity.

Mill's method of agreement is reinterpreted as a method for determining causal irrelevance. Nevertheless, Mill's original reading as a method for determining potential causes can be recovered as well: "If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon." (1886, 255) Indeed, with the framework of eliminative induction proposed in the beginning of this section, the exact premises can be determined, under which the method of agreement is valid. For example, it must be the case that there are no alternative causes among the varying circumstances. After all, the phenomenon could be due to different causes in both instances, while the only common circumstance is in fact irrelevant.

Somewhat more reliable is the joint method of agreement and of difference: "If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances in which it does not occur have nothing in common save the absence of that circumstance; the circumstance in which alone the two sets of instances differ is the effect, or cause, or an indispensable part of the cause, of the phenomenon." (1886, 259) But again, the presence of alternative causes may undermine such conclusions. However, based on the proposed framework of eliminative induction, the irrelevance of co-varying circumstances can be established such that inferences based on the joint method are corroborated.

The method of residues can also be derived based on causal relevance and irrelevance: "Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents." (1886, 260) Again, there is no difficulty in determining the specific premises, under which such inferences are valid. In particular, ensuring homogeneity is crucial. Finally, we have already discussed in much detail in Section 3d, how the method of concomitant variations determining functional relationships can be based on judgments of causal relevance regarding the presence and absence of factors.

A relatively small number of novel formulations of eliminative induction have been proposed in the 20th century, most notably by von Wright (1951, Ch. 4), Skyrms (2000, Ch. 5), and Mackie (1980, Appendix), but none of these seems to have eclipsed Mill's original framework. In a way, these modern approaches are somewhat more systematic, since they start from an explicit logic of necessary and sufficient conditions and then try to reformulate in particular Mill's methods of agreement, of difference, and the joint method. Unfortunately, the resulting methodology is quite complex. Mackie's framework, for example, introduces a complicated classification scheme for the basic methodology that depends both on the specific method one uses, e.g. method of difference or method of agreement, and on the type of causal condition one is after, i.e. whether the suspected cause is a single factor, a negation of a factor, a conjunction of factors, a disjunctions of factors, etc. Presumably, the complexity of such frameworks based on a logic of necessary and sufficient conditions is one of the main reasons why they have not caught on.

In summary, I claim that the proposed framework of eliminative induction, which relies on causal relevance and irrelevance instead of necessary and sufficient conditions and which reinterprets the method of agreement as a method determining irrelevance, has the crucial advantage of simplicity in comparison to all other accounts of eliminative induction that have been suggested until today, being based on only two complementary fundamental methods.

5b. Hume's problem of induction

Several epistemologists, including such authorities as Bacon, Mill, or Keynes, have in the past criticized enumerative induction as a primitive methodology, i.e. inferences from a number of particular observations of As that are also C to the general law that all As are C (e.g. Vickers 2014, Sec. 1). And indeed eliminative induction draws a much more realistic picture of the inductive process. For example, it can account for Mill's observation that sometimes we can reliably infer from only a few instances—ideally the method of difference needs only two—while in other situations nothing can be learned even from a large number of instances (Mill 1843, Ch. III.III). Indeed, eliminative induction points to the right kind of evidence required for causal knowledge, namely variational evidence, i.e. instances of a phenomenon under varying circumstances, instead of a mere repetition of instances (cf. also Russo 2009). This fits well with the observation that reliable causal knowledge is often generated in experimental contexts, i.e. in laboratory environments, which are distinguished by an excellent control over potentially relevant circumstances. Another crucial difference is that in the case of eliminative induction, one can show how increasing variational evidence steadily increases the quality of causal knowledge, while nothing analogous can be claimed for enumerative induction.

Every inductive method has its own problem of induction essentially consisting in the question under which premises reliable inferences result. The traditional problem of induction, notably in Hume's influential formulation, refers mostly to enumerative induction. Usually, enumerative inferences are taken to presuppose some principle of the uniformity of nature, which Hume states as follows: "that instances of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same." (Hume 2009, 62; cited in Vickers 2014, §2) This is indeed impossible to justify on a general level—essentially due to the vicious circularity that nature is only uniform in those aspects for which inductive inferences are valid.

The problem of eliminative induction is completely distinct from Hume's problem of enumerative induction. In particular, eliminative induction does not presuppose an indefensible uniformity of nature, but rather the following principles that are much more plausible if anything but trivial: (i) the principle of causality that every phenomenon is fully determined by conditions; (ii) an adequate causal language; (iii) constancy of background; (iv) repeatability of background and causal conditions (Pietsch 2014, Sec. 3.6).

Item (ii) will be discussed in the next section, since it is directly related to Goodman's new riddle that explicitly addresses the linguistic dimension of induction. While the development of causal categories that are adequate for a given phenomenon constitutes a considerable challenge, this issue illustrates well some realistic aspects of eliminative induction with respect to scientific practice. Most importantly, conceptual development always goes hand in hand with the formulation of empirical laws. The framework of eliminative induction and the corresponding difference-making account of causation are well-suited to trace how a refinement of language may take place during the inductive process, when increasing evidence is gathered.

Regarding item (iii), the requirement of a constant background or context, one can never be absolutely certain that it is fulfilled. But again, the framework outlined in this article delineates a procedure how causal knowledge can be improved in this regard by accumulating the right kind of evidence. In particular, it can be shown that circumstances, which were originally only postulated or believed to be irrelevant, really are irrelevant to a phenomenon with respect to a well-defined background.

Finally, item (iv) is only required if predictions are to be made on the basis of causal knowledge, not for the analysis of causal relations as such. Therefore, it need not always be fulfilled, rather it should be considered a matter of contingent fact whether it is the case or not. This fits well with the role of causal analysis for example in historical or juridical contexts, where repeatability is often not granted. The historical conditions leading to the Second World War are certainly not repeatable nor are the political circumstances leading to the creation of the European Union. Note that of course repeatability depends on the level of coarse-graining of the description.

Let us now turn to the principle of causality (i) which arguably constitutes the crucial premise for eliminative inferences. Why this principle is required is easy to see by constructing indeterministic situations, where the method of difference and the strict method of agreement fail, and thus also the definitions of causal relevance and irrelevance do not apply. For example, a certain indeterministic phenomenon may change due to pure chance while at the same time a circumstance varies that is in fact completely irrelevant to the phenomenon. Obviously, the method of difference applying the definition of causal relevance from Section 2a would lead to a wrong conclusion in this case.

From a modern perspective, one might be inclined to think that any method that presupposes the principle of causality cannot be taken seriously given that many sciences, from quantum mechanics to sociology, nowadays assume the ubiquity of indeterminism. By contrast, I will argue now that eliminative induction and thus the difference-making account also work for weakened versions of the principle of causality which are compatible with some amount of 'tame' indeterminism, while in the case of 'wild' and unrestricted indeterminism science would not be possible anyways.

A number of authors have pointed out that a trivial sort of determinism is generally irrefutable and thus cannot be in contradiction to the mentioned developments in science. For our purposes, the following version of such trivial determinism is suitable: Any distinct event is individualizable in terms of conditions or circumstances. Such determinism is quite plausible and intuitive given that pretty much every event is individualizable at least in terms of spatial and temporal coordinates, which according to the argument sketched in Section 2c are just shorthand for describing complex sets of causal relationships with causally prior or simultaneous events.

Given this trivial determinism, every event is indeed causally fixed, at the very least by the full list of prior or simultaneous circumstances. Certainly, this is not what scientists or everyday people are after. Even if nature were fully determined in this sense, our epistemological situation would deny us access to the full list of causes. Furthermore, these causes may not explain much or may not even ground reliable predictions.

Still, the feasibility of trivial determinism implies that eliminative induction as a scientific method cannot be falsified, as opposed to enumerative induction, which in fact keeps being refuted time and again. After all, it often occurs that a constant conjunction is observed for a large number of times only to fail later at a further trial. By contrast, in the case of eliminative induction, an analogous situation cannot happen: if the same phenomenon fails to occur under seemingly the same circumstances with respect to the same background, one has to conclude that a causally relevant circumstance was ignored, or alternatively that the chosen language was inappropriate. This move is always possible given the mentioned trivial determinism. In fact, the difference-making account relies on a principle of uniformity of nature that is true by definition: „Future events are similar to those of the past with respect to causal inferences if they differ only in terms of circumstances that are irrelevant for the inference." From this perspective, eliminative induction is consistent and irrefutable.

If aiming at predictions and explanations, one has to assume in addition that the world is to some extent orderly, i.e. that certain individual events can be grouped together and can be shown to be causally relevant for other groups of events in certain contexts. John Maynard Keynes has formulated a requirement to that purpose that he termed principle of limited independent variety: "the objects in the field, over which our generalisations extend, do not have an infinite number of independent qualities; that, in other words, their characteristics, however numerous, cohere together in groups of invariable connection, which are finite in number." (1921, Ch. 22) This leads us back to premise (iv). While endorsing the spirit and motivation of Keynes' principle, I would suggest a slightly different criterion that might be called the principle of repeatability of relevant circumstances: at least for some causal relationships, background and relevant conditions should be so unrestrictive that they can and will reoccur. As indicated before, this principle is not fulfilled for many phenomena.

There are a number of old and venerable principles in discussions on scientific method that work towards restricting the number of potentially relevant conditions. The most important is the principle of locality according to which an event can only be (directly) causally relevant to other events in its immediate spatial and temporal vicinity. Apparently, this principle of locality excludes a large number of circumstances from the list of potential direct causes. Assuming locality should not be seen as a strictly binding ontological constraint, but rather it has significant pragmatic connotations enabling eliminative induction in the first place. An argument for both temporal and spatial locality could start from the claim that our space-time geometry may be fully supervenient on causal dependencies (cp. Section 2c). Broadly speaking, it is not only the case that events can interact causally because they are close, but also that events are considered close because they can causally interact.

Another old acquaintance from the history of philosophy regards the principle of (qualitative and quantitative) equality of cause and effect. Without going into details, one should equally consider it a pragmatic constraint, rather than an ontological necessity. In the case of repeated and ubiquitous violation of this principle to exorbitant extent, science would just not be possible anymore.

Even if the outlined trivial determinism is not defeasible, it may sometimes be useful to work with various versions of tame indeterminism. The most innocuous variant of indeterminism is the kind one encounters in quantum mechanics according to the orthodox Copenhagen interpretation, where the circumstances do not fix the specific event that is happening, but only a probability distribution over related events. Such situations are relatively harmless and fully accessible to eliminative induction since determinism is restored on a coarse-grained level, apart from the additional conceptual difficulty how to determine relevance or irrelevance with respect to probability distributions, i.e. essentially with respect to ensembles of instances. Indeed, orthodox quantum mechanics is deterministic at the level of wave functions, i.e. probability distributions, whose evolution is fully determined by the Schrödinger Equation.

One could imagine a somewhat stronger indeterminism, where not even the probability distribution is fixed by circumstances, but only a range of events. If both the range of effects and the range of causes are sufficiently constrained, one can still define summary events. On the coarse-grained level of such summary events a deterministic description once again becomes feasible. Such indeterminism can become increasingly wild and complex, i.e. a large variety of events may follow arbitrarily after a large variety of circumstances. The world would become increasingly chaotic and unpredictable. In the limit of the wildest indeterminism imaginable, where anything can happen after anything, science becomes impossible. Note that eliminative induction may still formally work, if all events are individualizable in terms of spatiotemporal coordinates, but predictability is entirely lost.

In summary, indeterminism can only be handled scientifically, if determinism can be restored on a coarse-grained level of description involving probability distributions and/or summary events. The restriction to deterministic contexts in the difference-making account neither narrows down the range of application nor does it prevent the framework from being applied to the usual cases of indeterminism familiar from the sciences.

5c. The new riddle of induction

In the 20th century, there was one further influential attempt due to Nelson Goodman to formulate a problem of induction (1983, Sec. III). While the basic thrust is similar to Hume's problem, Goodman's new riddle of induction more clearly sheds doubt on probabilistic inferences and it puts the focus on linguistic aspects asking which predicates are projectable and which not.

Goodman's new riddle is often framed in the following manner: suppose a number of emeralds have been observed in the past which all exhibit the property green. What is one to infer from these instances? Is one justified to predict that the next emerald will be green as well, instead of for example blue? Goodman argues against this obvious conclusion on the following basis: instead of the usual predicate pair green/blue, one can construct an alternative pair grue/bleen with grue being green until some moment t in the near future and green afterwards, and bleen being blue until t and green afterwards.

Goodman claims that no fundamental reason exists for preferring the pair green/blue rather than grue/bleen on empirical grounds. One might want to argue that grue and bleen should be rejected because they are defined quantities that furthermore include in their definition a reference to some moment t. But green and blue can just as well be taken as being defined with green as grue until time t and bleen afterwards and blue being bleen until time t and grue afterwards. Also, attempts at a metaphysical solution of Goodman's riddle fail that green and blue constitute natural kinds as opposed to bleen and grue since we lack empirical access to any natural-kind property for predicates.

As already mentioned, the new riddle of induction is distinct from Hume's problem of induction by focusing on the projectability of predicates rather than on direct predictions. Also, it aggravates Hume's problem in that probabilistic accounts of induction are more directly affected. The latter can easily be seen in that the same evidence affords a possibly infinite number of conclusions: All emeralds are green / grue / gred / grack etc. According to enumerative induction all these inferences are based on the same number of observations, i.e. the same evidence, and should thus be ascribed the same probability, e.g. on the basis of a prescription like Laplace's rule of succession. Even when presupposing only a small range of possible colors, the axioms of probability imply that the probability for any prediction with a specific color must be small which obviously contradicts scientific practices.

Let us briefly discuss the new riddle in the framework of the difference-making account. Suppose there is a lamp and a switch that can make the lamp change color: if the switch is on (S=1) then the lamp is green (L=G), if it is off (S=0) then the lamp is blue (L=B). Now, someone using the predicates grue and bleen will discover the following: the lamp changes color from grue to bleen at time t without a cause in sight, while at all other instances S covaries with L. Given some confidence in the homogeneity of background and in determinism, this should prompt a reevaluation of the predicates grue and bleen resulting in the definitions green and blue to avoid the causal anomaly at time t. Of course, there may be other reasons to stick with grue and bleen, if these predicates have previously shown to be successful in other contexts.

Goodman's grue-problem provides a useful illustration, how eliminative induction and conceptual development are closely related and generally go hand in hand. Many more examples could be given. Indeed, in most situations, one has a choice either to interpret a recalcitrant observation linguistically or empirically, i.e. to either reevaluate language or to look for wrongly ignored circumstances (or even to bite the bullet and accept indeterminism). As will be elaborated later on, the more entrenched certain terms the less viable is the first option. By contrast, the better established homogeneity of background and determinism, the less viable is the second option. Thus, the process of eliminative induction not only leads to causal relationships but also to appropriate causal kinds, which are partly chosen with an eye on pragmatic criteria, most importantly simplicity.

But the real challenge of Goodman's argument concerned the projectability of predicates. Why are we almost always epistemically successful by relying on inferences containing the predicates green/blue and almost never by using grue/bleen? From the perspective of eliminative induction and the difference-making account an answer can be given that is closely related to Goodman's own solution in terms of entrenchment.

A first important remark is that eliminative induction presupposes some primitive properties on which other properties can be defined. Otherwise, the whole inductive business would become impossible since everything would resemble everything else, i.e. every event could be classified with every other with equal justification. Clearly, science would become futile. But fortunately, human experience of the world comes in terms of primitive properties that are determined by the sensual apparatus. Furthermore, it presumably results from our evolutionary development that these properties can afford at least some causal grip on the world.

A second remark concerns the time-dependence in the definitions of grue and bleen. Here, the insight is relevant that, according to the conventionalist perspective sketched in Section 2c, any reference to a point in time is just shorthand for the conditions that were present at that particular moment or prior to it.

With these insights in mind, consider a situation in which the constant conjunction of certain properties is observed until some time shortly before t: C with X, and D with Y. Now, grue-type predicates can be constructed in the familiar manner, in particular XY, which is X until t and Y thereafter, as well as YX, i.e. Y until t and X thereafter. Taking into account only the described evidence of constant conjunctions, Goodman's problem what to predict for a time after t is indeed not solvable. However, both the conditions C and D as well as the predicates X and Y generally appear in other contexts as well. They are all integrated in a huge web of causal relationships. For example, the colors blue and green are related to physical knowledge about wavelengths, about processes of color generation or of color change, or to phenomenological knowledge concerning the colors of various objects in the world etc.

So, it generally does not occur that with the introduction of a grue-type predicate like XY or YX only one causal relationship is implicated, but rather a large number of other relationships is affected as well. Thus, from the usage of grue-type predicates a rupture line in the conceptual web necessarily emerges. In the case of grue emeralds, for example, we have to ask, whether the wavelengths of green and blue are exchanged after instant t, whether certain atomic spectra and emission lines will shift accordingly, or whether trees will turn blue etc. If all this happens, i.e. if everywhere in the conceptual web the role of green and blue is just exchanged at t, then from a phenomenological perspective grue is just another word for green and there is no riddle at all. So, if Goodman's riddle is to be a genuine problem for induction then only some of this should happen, e.g. that emeralds turn blue, while trees remain green.

From the perspective of eliminative induction, such changes should normally happen only in combination with a modification of relevant circumstances. Therefore, assuming determinism and fairly well-established homogeneity of background, it should not occur that emeralds turn blue without any change in circumstances, while trees remain green without any change in circumstances or, equivalently, emeralds remain grue without any change in circumstances, while trees turn from grue to bleen without any change in circumstances. Thus, there must be a condition that is responsible for either the change between green and blue or that between grue and bleen. This decides between the predicate pairs blue/green and grue/bleen.

The better entrenched a predicate, the more tightly knit the causal web, to which it belongs, i.e. essentially the better a causal relationship involving this predicate is established by a process of eliminative induction testing it against a wide variety of contexts, the less likely that we should experience any surprises in terms of grue-type predicates in sufficiently similar contexts.[35] More exactly, grue-type predicates should not occur without a corresponding change in relevant circumstances for well entrenched predicates in well examined contexts. Obviously, this solves the problem of grue only if the issue of determinism as discussed in the previous section has found a satisfactory answer. Thus, the perspective from eliminative induction defuses Goodman's riddle in that it yields a clear criterion in terms of causal entrenchment, why blue/green should be preferred to grue/bleen when formulating predictions. Still, it remains to concede as a general moral of Goodman's riddle that the whole business of induction depends crucially on language and that sometimes surprises may happen.

5d. The logic of analogy

Another major epistemological problem in connection with inductive inferences, although it is not often acknowledged as such, concerns the nature and justification of analogical inferences. In fact, there is wide-spread skepticism, whether analogical reasoning provides more than a heuristic tool for hypothesis generation. For example, Paul Bartha, author of the most extensive modern treatise on analogy in philosophy of science (2010), writes in one essay: "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." (2013, §2.4) Historically, the main attempts to formulate a logic of analogical reasoning are due to Carnap and Keynes, but both frameworks have a number of well-known deficiencies. Nevertheless, the approach outlined in the following owes much to Keynes's work (1921).

While for obvious reasons analogical inferences are not covered by enumerative induction, from the perspective of eliminative induction, a sharp distinction between analogy and (enumerative) induction seems misconstrued. Strictly speaking, every inference according to eliminative induction is an analogical inference because generally no two instances are alike in all respects, i.e. in all circumstances. Thus, the difference-making approach to causation naturally supplies an epistemological basis for analogical reasoning and in particular implies a similarity measure that I would claim is superior to those proposed in the literature so far.

[35] Note again that grue-typeness can only be identified for well-entrenched predicates with respect to large classes of phenomena that remain green.

Let me set up a formal framework. Consider a source phenomenon S and a target phenomenon T. John Maynard Keynes once introduced the following useful terminology (1921, Ch. XIX; cp. also Bartha 2013, §2.2): the *positive analogy* consists in all those circumstances or properties that both source and target have in common. The *negative analogy* concerns all those circumstances in which source and target differ. The *unknown analogy* regards all those circumstances of which it is unknown whether they are shared by source and target. Finally, the *hypothetical analogy* consists in those circumstances within the unknown analogy that are instantiated in the source and are predicted for the target. I will work with a two-dimensional framework for analogical reasoning as endorsed by several authors, introducing the crucial distinction between horizontal relations accounting for similarities between source and target in terms of circumstances, and vertical relations concerning different types of relationships between those circumstances, e.g. causal or deductive (Hesse 1966, Bartha 2010, Norton 2011).

Arguably, the main epistemological challenge of analogical reasoning consists in establishing an adequate measure of similarity between target and source phenomena that translates into a probability for the hypothetical analogy based on the positive, negative, and unknown analogy. This issue lies at the heart of the wide-spread skepticism concerning the reliability of analogical inferences. In fact, many authors have argued that the unstable nature of such inferences directly derives from the contextual and subjective nature of similarity itself. Remarkably, all major accounts of analogical reasoning differ with respect to the similarity measure that they employ. Sometimes, the number of corresponding properties is counted and weighted against the number of differing properties. The influential structure-mapping theory of Dedre Gentner suggests a measure in terms of structural similarity. As a third example, Paul Bartha requires different similarity measures for different types of analogies, e.g. the causal analogies discussed in the following are treated with Paul Humphrey's framework of 'aleatory explanation' counting contributing and counteracting causal factors (Bartha 2010, Ch. 4.5).

Let me now briefly argue that the measure of similarity implied by the difference-making account is superior to these and other approaches at least in causal contexts. The crucial advantage is that it builds on well-defined notions of relevance and irrelevance that allow for a substantial amount of objectivity. In its simplest version, the difference-making account works under the assumption of full determinism, in which case we have:

> *An analogical inference holds if the negative analogy (i.e. the ◊-conjunction of all circumstances therein) is causally irrelevant to the hypothetical analogy with respect to a background B* constituted by the positive analogy.*

Of course, it does not follow that all circumstances in the negative analogy themselves need to be causally irrelevant with respect to B*. Certainly, there may be causally relevant influences in the negative analogy, if they completely cancel each other.

Various complications can arise. For example, it may not be fully known whether the negative analogy is irrelevant. Then, the analogical inference will be valid only with the probability that the negative analogy is indeed irrelevant. If there is just one circumstance in the negative analogy of which it is unknown if it is irrelevant, then the whole analogical inference will hold with the probability that this circumstance is irrelevant. To guarantee a certain amount of objectivity for such probabilistic inferences, a notion of probability should be employed that itself warrants some objectivity. Without being able to cover the details, I would suggest to rely on a causal notion of probability in the tradition of writings by Cournot, Mill, and von Kries as well as modern authors like Strevens, Rosenthal, and Abrams among others. This notion fits well with the difference-making account (Pietsch manuscript).

Furthermore, there may be an unknown analogy. If the unknown analogy and the negative analogy are both known to be irrelevant with respect to B, then the analogical inference will of course be valid. If the negative analogy is irrelevant, but the unknown analogy is known to be relevant, then the analogical inference will be valid with the probability that the unknown analogy (or at least the relevant circumstances therein) belongs to the positive analogy. Finally, there may be an unknown analogy of which it is unknown if it is relevant or not, but a probabilistic treatment of such cases is again straightforward.

Thus, eliminative induction provides a consistent logic for analogical reasoning. The apparent unreliability and ambiguity of many analogical inferences then must have a different origin. First, in many cases the respective probability measures are unknown. Therefore, the corresponding probabilities cannot be determined objectively and consequently the strength of the analogical inferences cannot be quantified. In such situations of ignorance, which are the rule rather than the exception, analogical reasoning indeed becomes rather heuristic.

Second, not all analogical inferences are geared at predictions. Rather, an important class of analogical reasoning concerns the development of structural analogies between different phenomena, when for example Thomson and Maxwell developed electrodynamics in analogy to hydrodynamics or when the theory of harmonic oscillators is applied in a wide variety of scientific disciplines. In fact, the difference-making approach can account for this crucial distinction between predictive and structural analogies in part on the basis of an observation that was discussed already in Section 4b: the logic of the difference-making approach works just as well for empirical as for definitional relationships. Indeed, structural analogies employ the underdetermination that exists especially in more abstract sciences to develop a conceptual structure that allows transferring empirical results from one phenomenon to the other. Since one is dealing with at least partially definitional or conventional relationships, these analogies cannot be evaluated in terms of truth or probability. Such reasoning therefore always has a heuristic dimension being largely guided by pragmatic considerations, in particular a pronounced familiarity with the source phenomenon.

Thus, while there are significant heuristic modes of analogical reasoning, when analogical inferences are geared at prediction, eliminative induction and the difference-making account provide an adequate and rigorous formal framework. All inferences based on eliminative induction are analogical inferences.

6. Conclusion

A difference-making account of causation was proposed and its prospects examined. The difference-making account broadly stands in the counterfactual tradition, but has certain characteristics that help it avoid some of the most troubling objections with relatively little conceptual and metaphysical burden. Arguably, the difference-making account fares better than other accounts in establishing a notion of causation that exhibits little vagueness and subjectivity and that thus fits well with the role of causation in actual scientific practice.

Acknowledgments

I am very grateful to Michael Baumgartner and Federica Russo for comments on the manuscript as well as to Mathias Frisch for discussions. Audiences at PSA 2014, the MCMP in Munich, and EPSA 2015 have provided helpful feedback and criticism. Finally, I thank the students of my philosophy of causation seminar in the winter semester 2015/16 for many insightful discussions.

References

Bacon, Francis. 1620/1994. *Novum Organum*. Chicago, IL: Open Court.

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. 2013. “A Regularity Theoretic Approach to Actual Causation.” *Erkenntnis* 78: 85-109.

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

Cartwright, Nancy. 2007. *Hunting Causes and Using Them.* Cambridge: Cambridge University Press.

Frisch, Mathias. 2014. *Causal Reasoning in Physics.* Cambridge, UK: Cambridge University Press.

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

Goodman, Nelson. 1983. *Fact, Fiction, and Forecast,* Cambridge, Mass.: Harvard University Press.

Hausman, Daniel M. 1998. *Causal Asymmetries*, Cambridge: Cambridge University Press.

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

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

Horwich, Paul. 1987. *Asymmetries in Time,* Cambridge, Mass: MIT Press.

Hume, David. 1748. *An Enquiry concerning Human Understanding*.

Hume, David. 2009. *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hüttemann, Andreas. 2013. *Ursachen.* Berlin: de Gruyter.

Illari, Phyllis and Jon Williamson. 2012. "What is a mechanism? Thinking about mechanisms across the sciences." European Journal for Philosophy of Science 2(1):119-135.

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

Kim, Jaegwon. 1973. "Causes and Counterfactuals." *Journal of Philosophy* 70: 570-572.

Lewis, David. 1973. "Causation." *Journal of Philosophy* 70, pp. 556-67.

Lewis, David. 1979. "Counterfactual Dependence and Time's Arrow", *Noûs* 13: 455–76.

Lewis, David. 1986. "Postscripts to 'Causation'." In Philosophical Papers, Volume II. Oxford: Oxford University Press, pp. 172-213.

Lewis, David. 2000. "Causation as Influence", *Journal of Philosophy*, 97: 182–97

Lewis, David. 2001. *Counterfactuals*. Oxford: Blackwell.

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

Menzies, Peter. 2014. "Counterfactual Theories of Causation." *Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). Available online: http://plato.stanford.edu/entries/causation-counterfactual/

Mill, John S. 1886. *System of Logic Ratiocinative and Inductive*. London: Longmans, Green, and Co.

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

Paul, L.A. and Ned Hall. 2013. *Causation: A User's Guide*, Oxford: Oxford University Press.

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

Pearson, Karl. 1911. *The Grammar of Science.* (3rd ed.). New York: Meridian Books.

Pietsch, Wolfgang 2014. "The Structure of Causal Evidence Based on Eliminative Induction." *Topoi* 33(2):421-435.

Pietsch, Wolfgang. 2015. "The Causal Nature of Modeling with Big Data." *Philosophy & Technology*. DOI: 10.1007/s13347-015-0202-2.

Pietsch, Wolfgang. Manuscript. "Causal Interpretations of Probability." http://philsci-archive.pitt.edu/11315/

Price, Huw and Brad Weslake. 2009. "The Time-Asymmetry of Causation", in H. Beebee, C. Hitchcock, and P. Menzies, *The Oxford Handbook of Causation*, Oxford: Oxford University Press, pp. 414–43.

Psillos, Stathis. 2015. "Counterfactual Reasoning, Qualitative: Philosophical Aspects." In James Wright (ed.), *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam: Elsevier, pp. 87-94.

Reutlinger, Alexander. 2012. "Getting Rid of Interventions." *Studies in the History and Philosophy of Science Part C* 43 (4):787-795.

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

Russell, Bertrand. 1913. "On the Notion of Cause." *Proceedings of the Aristotelian Society* 13:1-26.

Russo, Federica. 2009. *Causality and Causal Modelling in the Social Sciences. Measuring Variations.* New York: Springer.

Russo, Federica. 2014. “What Invariance is and How to Test for it.” *International Studies in the Philosophy of Science* 28(2):157-183.

Scholl, Raphael 2013. “Causal Inference, Mechanisms, and the Semmelweis Case.” *Studies in History and Philosophy of Science* 44(1):66–76.

Scholl, Raphael and Tim Räz. 2013. “Modeling Causal Structures: Volterra’s Struggle and Darwin’s Success.” *European Journal for Philosophy of Science* 3(1):115–132.

Skyrms, Brian. 2000. *Choice and Chance*. Belmont, CA: Wadsworth.

von Wright, Georg H. 1951. *A Treatise on Induction and Probability*. New York, NY: Routledge.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, James. 2013. “Causation and Manipulability.” *Stanford Encyclopedia of Philosophy (Winter 2013 Edition)*. http://plato.stanford.edu/archives/win2013/entries/causation-mani/