A Formal Ontological Approach to CAUSALITY Embedded in the Top-Level Ontology of GFO (General Formal Ontology)

Von der Fakultät für Mathematik und Informatik
der Universität Leipzig
angenommene

DISSERTATION

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM
(Dr. rer. nat.)

im Fachgebiet
Informatik

vorgelegt

von Magister Artium Hannes Michalek
geboren am 09.09.1977 in Schopfheim.

Die Annahme der Dissertation haben empfohlen:
1. Prof. Dr. John A. Bateman, Universität Bremen
2. Prof. Dr. Gerhard Brewka, Universität Leipzig
3. Prof. Dr. Heinrich Herre, Universität Leipzig

Die Verleihung des akademischen Grades erfolgt auf Beschluß
des Rates der Fakultät für Mathematik und Informatik vom
14.12.2009 mit dem Gesamtprädikat *magna cum laude*.

Epitome

In this thesis, a formal ontological theory of causality is developed that is conceptually based on the intuitions of regularity and counterfactual dependency (covering manipulability in the process as well). Both relations are introduced as extensions of the General Formal Ontology (GFO), and they are defined on coincidence pairs of presentials and include a probabilistic aspect. While regularity covers statistical dependency on universals' instances, counterfactual dependency is about supportive/undermining causally contrastive clusters of coincidence pairs, taking their relative distance to actuality as a reference cluster into account. Based on GFO's relations between presentials and processes, the basic causal relation is extended to cover different kinds of causal relations between processes. The quartet of interaction being a nice example for this extension's modeling capability. With respect to epistemics, the theory is able to explain our general ability to empirically discover causal relationships and in which ways it is limited. A reconstruction of methods used in performing experiments in general, and in clinical trials in particular, shows the epistemic adequacy of the theory developed.

Acknowledgements
or:

Mea Culpa

While I was working on the subject of causality in the context of the ontology of GFO, many people kindly tried to help me overcome my shortcomings. Here's the place to make clear that

. . . if I ever felt lost in any aspect of theoretical computer science (especially in formal logics), or missed the necessary scientific feedback despite having Professor Heinrich Herre, Professor Barbara Heller (from whose enthusiasm I benefited much too shortly), Frank Loebe, Robert Höhndorf, Patryk Burek and the other members of the Onto-Med group in Leipzig on my side,

. . . if I stumbled in the fields of clinical trials and statistics even after being instructed by Professors Dirk Hasenclever, Korbinian Strimmer and Nihat Ay,

. . . if my conceptual analyses should be flawed even after Monika Dullstein commented on a draft and even though I had excellent teachers at the Munich School of Philosophy, especially Professor Godehard Brüntrup, to whom I am grateful for introducing me to the field of metaphysics in the first place,

. . . if my contact to the department of computer science was rather eclectic even though it hosts many gentle people like Professor Brewka who agreed to support me in the course of the graduation procedure,

. . . if there should still be English language mistakes in this text even though Andrew Colville and Mathias Schäfer spent their precious time ironing them out,

. . . and finally, if I ever quarrelled with my fate despite growing up in the family in which I thankfully did and despite having my own little family around me every day now,

. . . it is entirely my fault.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Limiting the scope

It is not a new insight that terms like “causes” or “causal relation” are ambiguous. In fact, there are so many meanings and connotations (depending on context and the speakers’ intentions) that the discussion of causality even brought about the request to get rid of “causes” in general:

> “[T]he word ’cause’ is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable [. . .].”
>
> (RUSSELL, 1910, p. 180)

Nevertheless, causal knowledge plays a vital role in various fields, so we will (pace RUSSELL) follow this general strategy: we limit the scope by making explicit what fields (and, thus, what interpretations) are outside the focus of this thesis. Additionally, we will try to illustrate our idea of causality as well as we can. So even if you disagree that what we take as causality “really is” causality, we hope that you can still benefit from our considerations, interpreting them as “causality in the sense of this thesis”, whatever technical term you may use for it.

Causality as “Physically Making Happen”

Let us plunge directly into *medias res* with some statements that refer to causal relations:

(S 1) A thunderstorm will occur because the air pressure is dropping rapidly.

(S 2) Mary caught a cold because she visited her friend Sue, who already suffered from that disease.

(S 3) The window pane broke into pieces because Mary’s ball hit it.

Up to here, there is no problem. Each of the sentences (S 1) to (S 3) refers to a cause and its effect. But then there are statements that look pretty much the same:

(S 4) Joe is a bachelor because he is an unmarried young man.

(S 5) 5 is prime as it is a natural number which has exactly two distinct natural number divisors: 1 and itself.

In our view, however, these sentences do not refer to causal connections. So we shall take some time to explain why, in order to make you familiar with how we will understand and use terms like “cause”, “causal relation”, and so forth, in the course of this thesis.

The shortest description of what we think that “causes” – as in “A causes B” – means, is: “*makes happen*” or “*physically brings about*”. In sentence (S 1) e.g. , it is the drop in air pressure that physically brings about the thunderstorm. But with sentence (S 4) things are different. It would be strange to claim that being an unmarried young man “physically brings about” being a bachelor. The same holds for (S 5): being divisible by only 1 and itself does not “physically bring about” being a prime number. The connection that the latter sentences refer to is more of a conceptual or logical inference/consequence than a causal relation: It is part of the meaning of “bachelor” to be “unmarried”. And being divisible by only 1 and itself is the definition of “being prime”. But why are we tricked by the surface structure? Isn’t there anything that all these sentences have in common?

Indeed, there is. They all refer to *explanations*. They give reasons in order to answer the question about why something is the case. But – and this is what needs to be stressed – they differ by the kind of reason they give. The first three do give *causal explanations*, while the latter give *conceptual* or *logical* explanations.

Causality as Physical Causality

The expression of “physically bringing about” already shows that we will restrict our analysis to *physical causality*, i.e. to causal relations between physical objects and whatever is connected to them (like qualities or processes they take part in). The ontology of e.g. the mind, however, and of the mind’s relations to the physical world (like psycho-physical interactions, or the so-called “Mind-Body” problem) is a huge field of its own that we cannot try to cover here, so you will not find a proposal for modeling statements such as “My anger made me crush that glass” or similar ones.

For the same reason, we do not cover relations from the social realm (or “stratum”, as it is e.g. called by HERRE ET AL. (2007, chapt. 4)[1]). We do not deal with judgments like e.g. “The industrial revolution led to the early socialist movement.”

Nevertheless, we are convinced that if any causal notion is to be used in these or other fields, it will share the main conceptual features of *regularity* and *counterfactual dependency* that we will develop in the course of our investigations. But everything more concrete (in terms of building an ontological theory), like the question of the ontological nature of the causal relata, and the connections to time, may well look very different there than it does with physical causality.

[1] HERRE ET AL. (2007) follow POLI (2001), here, who, in turn, rests his framework on HARTMANN (1964).

No Causal Pluralism in Physical Causality

One strategy in order to tackle the numerous accounts of causality is the approach of *causal pluralism*, which defends the idea that there are several different notions of cause.[2] We do not follow this approach, here: as far as the realm of physics is concerned, we believe that there is only one kind of relation that may be called causal.

"A" Cause, not "the" Cause

As the last clarification on what kind of causality we are discussing, let us make clear that we do not identify "a cause" with being "the cause":

> The careless tossing of a lit cigarette, the recent drought, the presence of oxygen in the atmosphere; these all count among the causes of the forest fire. [...] Which cause we single out will depend upon context and the interests of the speakers [...].
>
> (HITCHCOCK, 2003, p. 5)

So when something is identified as a cause by our theory, this does not mean that there are no other causes as well.

1.2 Computer Science and Philosophy Overlapping

When looking at the table of contents, you will find that sections with philosophical content play a relevant role in this thesis. The following section will explain why, and it will do so from the perspective of both, philosophiers and computer scientists.

1.2.1 Why is Causality an AI Problem? (An Introduction for Philosophers)

In the early days of computer usage, the term "electronic data processing" was coined. And even if it seems outdated by now, it proves handy for illustrating why certain computer scientists care about the nature of causality.

1.2.1.1 Machines Follow Rules

"Electronic processing" means that an electronic *machine* is used to process the data, which has an important implication: being a physical entity, a machine first of all follows the laws of nature. Additionally we as the machine's creators can add extra *rules* of how the machine should behave, e.g. what it should do under certain conditions. Indeed, we must say that a machine cannot do

[2] For overview and defense of causal pluralism cf. HITCHCOCK (2003); GODFREY-SMITH (2006)

anything else but to follow (more or less complex) rules. In computers, most of these rules are typically given by software. And even if one might be tempted to think that by adding e.g. a randomization device the rule based character could be transcended, all it does imply is that we may probably not be able to foretell its behaviour. Yet we do know that once it gets the data from the randomization device, it processes that data according to the rules given.

1.2.1.2 Computers Manipulate Symbols

Regarding computers, what is the "data" that is being processed, and what does "processing" mean with respect to this data? In modern PCs, of course, it is the bits, physically stored on disks and memory chips, that the machine works on, which means that the processor takes them, performs some (mathematical) operations on them, and returns bits as output which then are stored again or e.g. presented on a display. Computers, thus, belong to what two of the godfathers of artificial intelligence, Allen NEWELL and Herbert A. SIMON, called a "physical symbol system"(NEWELL and SIMON, 1976) which comprises the following elements:

> [...] a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). Thus, a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another). [...] Besides these structures, the system also contains a collection of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction and destruction.
>
> (NEWELL and SIMON, 1976, p. 116)

In short what the computer works on (the "data" in "data processing") are physical symbols, i.e. patterns built of physical bits.

1.2.1.3 Knowledge Representation

The consequence of computers following rules, and being symbol systems, is that whenever we want to make use of computers, we must find a way to "translate" the problem into (or: "encode" by) symbols and rules on how to handle these symbols. An example might be of help, here.

Let us say we want a computer to sort a list of professions alphabetically (cf. fig. 1.1). In order to do so, we may encode the strings (as sequences of characters) into sequences of bits. Then we have the computer sort the bit sequences following certain rules, e.g. comparing the sequences bit by bit firstly grouping "0..." and "1..." sequences, then grouping within those groups to get "00...", "01...", "10...", "11..." sequences, and so on. And once we have got the binary sequences sorted, we translate them back into characters and strings.

This procedure exemplifies the steps necessary for all computer problem solving:

- Translate the problem into symbols the computer can work with.

String	*Binary Code*	*Sorted Binary*	*Sorted String*
`Chauffeur`	01000011[...]	01000010[...]	`Barber`
`Miller`	01001101[...]	01000011[...]	`Chauffeur`
`Teacher`	01010100[...]	01001101[...]	`Miller`
`Vet`	01010110[...]	01010100[...]	`Teacher`
`Barber`	01000010[...]	01010110[...]	`Vet`

Table 1.1: *Sorting a list of professions alphabetically*

- Define rules on how to manipulate the symbols.
- Interpret the symbols that are created during the manipulation.

This string sorting task is trivial for a computer, of course, as the binary representation we used here already encodes the alphabetic order of characters, i.e. it assigns smaller binary "numbers" to characters that occur earlier in the alphabet. However, is it easy in yet another sense: the information that is needed to perform the task (the alphabetic order) is already present in the characters themselves. You don't have to know *what* a barber is in order to sort the list. If someone gave you words written in an alphabet you do not know, but gave you the rules for sorting them as well, you could work out a perfectly sorted list without understanding any of the words. And this, in fact, is what the computer does.

But for very many other tasks, it is exactly the knowledge you have on e.g. barbers that must be used: say, we want the professions not to be sorted alphabetically but grouped by whether they require an academic degree (cf. fig. 1.2). There is no way to solve this problem by using strings of characters (like `Barber`) alone. In order to enable a computer to solve this problem, we must find a way to encode our *knowledge* about professions in a computer-readable way. Further, we must encode the rules of how to process this knowledge e.g. to draw the inferences necessary for this specific task.

Professions	Grouped Professions
Chauffeur	Chauffeur
Miller	Miller
Teacher	Barber
Vet	Teacher
Barber	Vet

Table 1.2: *Professions grouped by whether they require attending university*

And this is the reason why computer scientists are faced with the philosophical problem of the nature of causality: they need to model causal relations and causal knowledge in a way that can be processed in computers. This requires a discussion of what causality *is*. It is the question of *conceptual adequacy* of models that computers may work with.

Formal ontology (which we shall hear more about in sect. 3.2), is one way of modeling (encoding) knowledge in a formal language computers can process. And it is the way that we will deal with in this thesis.

1.2.2 Why is Philosophy Relevant for the Problem of Causality? (An Introduction for Computer Scientists)

Now that we know that knowledge modeling is necessary to make certain problems accessible for computers, the computer scientist's question probably is: what has philosophy got to do with it?

To answer that question, we must note that there are two kinds of knowledge concerning causality: first, there is knowledge about what entities are causally connected to each other, a question that clearly belongs to the subject area of a domain expert rather than a philosopher. The other, however, is: *what is* a causal relation? What does it mean when we state that "A causes B"? What conceptual inferences can be drawn from a statement containing a certain concept? How does a causal relation differ from other kinds of relations? In other words, it is *concept analysis* that is asked for. And concept analysis exactly is what many contemporary philosophers (often referred to as "*analytic* philosophers") would regard as their main scientific interest: What makes a judgment containing a certain concept true? Is causality reducible to other, more fundamental relations or concepts? How is causality connected to other concepts, e.g. the concept of responsibility or guilt?[3]

And indeed, there is more complexity to the concept of causality than may be obvious. Here is an example of how a commercial[4], large-scale knowledge base combined with a reasoning engine does approach causality in a way that may easily come to mind when trying to model causality.

Many causal statements have the form "A, therefore B". Here are some examples:

(S 6) The temperature dropped below 0°C, therefore the puddle is frozen.

(S 7) There is water in the fuel, therefore the engine misfires.[5]

---

[3] For the latter cf. e.g. LEHMANN (2003), which is an excellent example for how the question of conceptual adequacy of modeling causality is relevant in even more fields besides philosophy and computer science. In this case, it is the realm of law.
For a discussion of the concept of causality as it is relevant to the field of medicine, cf. GROSS and LÖFFLER (1998).

[4] It should be mentioned that the "Cyc" project described here has a free-of-costs sibling called "OpenCyc": `http://www.opencyc.org/`

[5] For this example and the following critique on Cyc, cf. COPELAND (1997).

Being somewhat familiar with logics, this may sound like the material implication of $A \rightarrow B$. And indeed, this way of representing causality has been used in a (multi-million dollar) system called "*Cyc*" (cf. GUHA and LENAT, 1990; GUHA, 1990).

But then, this approach runs into serious problems concerning conceptual adequacy, as a material implication is true if the antecedent is false (a feature called "ex falso quodlibet" or the "principle of explosion"). In the Cyc model, the following sentences would be true, as well:

(S 8) 4 is prime, therefore the puddle is frozen.

(S 9) 4 is prime, therefore the engine misfires.

But clearly, neither a puddle's freezing nor an engine's misfiring is causally connected to a number being prime or not. Thus, the Cyc concept of causality contains a grave flaw with regard to conceptual adequacy.[6]

1.2.3 The Common Ground

The last two sections should have made it clear that conceptual adequacy is the point where computer science and philosophy meet when the question of causality is concerned – it is simply about how to model a concept correctly (for computational reasons), which involves (philosphically) analyzing what that concept means.

Computer science has the advantage that the normative aspect of concept analysis (i.e. examining what causality "really" is) can be put aside more easily by focusing on the descriptive facet. Computer scientists can concentrate on checking whether certain (given) causal knowledge (statements, conclusions, facts) can be modeled (mostly) without discussing whether the applied causal concepts are used correctly.

But even then, the formalisms, axioms, or even full logics developed, need to be checked with respect to their conceptual presuppositions and commitments. The philosophical part being inseparably entangled with computer science's task.

1.3 Structure in Content

The structure of this thesis is straightforward: While the present section gave an introduction on both, the philosophical and the computer science view taken on causality, chapter 2 introduces the central theories of the philosophical (conceptual) analysis. We will discuss these approaches in more detail before moving on to our own conceptual theory of the causal relation.

[6] We will present other difficulties of the Cyc approach later, when discussing several computer science theories/models of causality (cf. sect. 3.2.2).

Equipped with the necessary terms and concepts to evaluate models of causality (w.r.t. conceptual adequacy), chapter 3 is the place to present and criticize some of the ways in which computer science has dealt with causality so far.

We will then introduce our theory of causality, which is done in two steps: chapter 4 describes the top level ontology called General Formal Ontology (GFO), as introduced in HELLER and HERRE (2003, 2004b)[7], which will be used as the basis for ontologically modeling all non-causality-related parts.

Chapter 5, finally presents the new approach, starting with the question of the ontological nature of the causal relata. In a next step, a formalisation of both regularity and counterfactual dependency is presented to cover what will be called the basic causal relation. This basic theory is then extended to cover causal relations between processes.

The concluding chapter 6 then is dedicated to the epistemic implications of our theory and its application to natural science performing experiments with clinical trials as a special case.

[7] The most recent version being HERRE ET AL. (2007)

2 Philosophical Theories of Causality

The history of philosophy has seen a wide range of different theories on causality (cf. SCHAFFER, 2003). Most of them can be aligned into three main branches based on central intuitions that our common sense concept of causality is committed to:

- Regularity
- Counterfactual dependency
- Manipulability

This chapter will provide an overview of these intuitions, showing that in certain cases we have certain commitments/ideas about what it *means* that something is a cause of a certain effect. Additionally, regularity and counterfactual dependency are analyzed in order to create a coherent GFO based theory of causality. This will include explaining why abovementioned manipulability is not part of this list.

2.1 Regularity

Imagine the following situation (*"Security Alarm"*): You are leaving a shopping centre and just when you pass the door, an alarm is set off. A member of the security service approaches you and discretely asks you to pass the security gate again when the ringing has stopped. Since you know that you haven't bought (or stolen) anything, you follow his orders, and this time, there is no alarm. The guard apologises and you leave the store without any further inconvenience.

Why did the security officer let you go? The answer is simple, of course: If you *had* stolen something, the alarm should have started again. When it did not start, the officer concluded that you were not causing the alarm. His decision was based on a simple intuition on causality: If something causes something else, this connection should be reproducible. If there really is a causal relation, a reconstruction of the assumed cause (you, passing the door) should yield the same effect (alarm is set off). If it does not, we do not have a cause–effect–relation. We call this the intuition of *regularity*.

Within the development of philosophical theories on causality, regularity plays a very important role. The first "modern" approach on this topic, written by DAVID HUME, is based on exactly this intuition:

> [. . .] 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.* [8]
>
> (HUME, 1748, p. his emphasis)

This approach is very useful as it explains several peculiarities of causality (many of which are already covered in HUME'S 1748 "Enquiry"): first, we cannot *perceive* causality. Even in very obvious case of two billiard balls colliding, all we can actually see is the movement of balls, their changing direction and speed. We can hear the clicking sound when they touch. And we may probably feel the vibration they produce when moving on the cloth. But there is no human sense that covers causality. It is only by comparing our observations to similar other cases that we are able to *conclude* a causal relation.

Secondly, regularity makes the scientific concept of *falsification* make sense: Imagine a group of scientists claiming to have made a certain discovery, say, that oxygen produces a loud sound of 440Hz if it is cooled down to exactly -244°C. They call it the "cryophon" effect and publish their results in a scientific journal. Rival researchers will now try to reproduce this effect, and if they cool oxygen down to -244°C without detecting the sound, the cryophon theory (which is a theory about a certain causal relation) must be rejected. Rejection by falsification is based on the same inferences as in the initial example of the security guard's decision to let you go. If there really is a causal connection, it must be reproducible.

The cryophon example shows another interesting feature of the regularity condition: It presupposes that causality is about "kinds" or "families" or "groups" of *similar* situations. If the causal claim was restricted to the single occurrence in the first research group's laboratory, no reconstruction could falsify the cryophon theory. Thus, a causal claim is about something more general than just a unique occurrence.

Following these examples, it is tempting to *identify* or equate causality with regularity. But here the problems begin, as several counterexamples have been developed that try to show that there are cases in which there is regularity without causality, or in which there is causality

[8] Please note that the term "object" may be misleading, here. Being the empiricist that he was, HUME makes this claim focussing on *experiences* not objects. Three sentences later he gives another definition: "[. . .] *an object followed by another,* and *whose appearance always conveys the thought of that other.*" [his emphasis]. His idea of causality is based on a relation between *thoughts* that are aroused by experience (and through anticipation).

without regularity.[9]

We begin with a well known example which we shall refer to as *Barometer–Storm*: Imagine a causal connection between a rapid drop of air pressure and the arrival of a storm with the air pressure drop having a second effect: The barometer reading falls. And now – so the counterexample goes – let us apply the regularity condition to the barometer reading and the storm's arrival, isn't there regularity between the barometer reading and the storm? Indeed there is. Everytime the reading falls (i.e. the air pressure falls), a storm arrives. Equating regularity with causality would force us to call this relation causal. But we know this is wrong. The barometer does not cause the storm. Thus, we have regularity between barometer and storm without them being connected as cause and effect.

This counterexample belongs to a whole family of examples built upon what may be called "subsequent effects of a common cause". In Barometer–Storm, the air pressure causes both the reading's falling and the storm's arrival. Other members of this family of arguments may e.g. be construed on subsequent symptoms of a single desease: A common cold often starts with sneezing, and is later accompanied by a mild fever. Again, regularity holds between sneezing and the fever, but it would be wrong to conclude that the sneezing causes the fever. In the light of these counterexamples we must accept that regularity is not sufficient for causality, i.e. it does not logically imply causality.

This is not the only restriction: Some counterexamples claim to show that there is even causality without regularity, which might directly render the inferences made by the guard in our Security Alarm example (and by cryophon's second research group's falsification attempts) invalid. If causality does not always entail regularity, the alarm not being set off (when you are stepping through the gate again) does not exclude you from being its cause any more.

Let us take a look at one of these examples, which again refers to a common cold. Following basic medical knowledge (cf. NHS CHOICES, 2009), it is viruses that cause a cold, and these viruses are spread by a certain mechanism (airborn droplets, coughing / sneezing, hand contact) which might make you catch a cold by e.g. visiting a contagious patient. But - and here is the crucial point - you do not always catch a cold when visiting a contagious patient, even when viruses are spread. In short, catching a cold involves *chance*. A cold *can* be caused by visiting a patient, but visiting not always causes a cold. In this case, causality does not imply regularity.

[9] Even if the following sections are dedicated to the difference between regularity and causality, this does *not* mean that regularity is somehow unimportant or that knowledge on regularity is second-rate. Knowing about regularities has a value of its own as it e.g. allows you to predict parts of the future so you can take precaution or otherwise adapt your actions to the expected events. If an animal shows a certain behaviour prior to heavy rain, you can make use of that knowledge without believing that the animal "makes" (i.e. causes) the rain.

2.1.1 Probabilistic Causality

Later, we will explain how regularity theory can easily be adjusted to cover cases of chancy causation as well. But let us dwell on the subject of probabilistic causation for a moment, as it is tempting for many people to assume that every causal connection is determinate "in the end", and that probabilistic aspects are just results of limited knowledge about the relevant mechanisms. Nevertheless, we believe that genuine probabilistic causation is a sensible and consistent concept that a theory of causality should be able to deal with and will present three lines of argument to justify this assumption based on the consistency of the concept of magic, on real word epistemical restrictions and finally on indeterminacy in modern physics.

It is very common to conceptually connect causality with non-probabilistic physical laws like the Newtonian laws of motion. This may well be the reason why we tend to think that once we understand the world in an ideal way, we will recognise that everything is following strict determinate mechanisms.[10] On the other hand, there is the concept of *magic*, e.g. in literature and other fields of fiction. Magic (if not understood as in "legerdemain" or "magic trick") means bringing about something in a way that opposed to common scientific laws. And despite the fact that magic appears to be restricted to the realm of fiction, this example shows that we can conceptually distinguish between causation as such and causation by physical laws. And once we distinguish between causality and physics, there seems to be nothing problematic about probabilistic causation:

> So finally, Merlin felt impelled to cast this most dangerous spell, which *might* save his fellows' lives. Although he well knew that indeed there was a reason why it was never written down in any of the known and of the many more forgotten languages of this world but only passed on from master to apprentice as its outcome could not be foreseen and in only one out of a thousand casts did not lead to plain disaster.
>
> (author's invention)

In less elegiac words, the sentence "Morgana cast a spell that with a chance of 30% would transform her victim into a mouse at midnight." is not "falsified" by there being cases of cast spells that did not succeed in any transformation. It is indeterminate. But *if* there was a transformation at midnight, there is no reason not to regard Morgana's as spell the cause of this

[10] In the words of 19th century mathematician and astronomer Jean-Pierre de Laplace: "An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes." (de LaPlace, 1814, as translated in Hoefer, 2005).

transformation. As we have seen, the concept of (pure) causality – as opposed to identifying causality with mechanisms of physics – does no longer rule out probabilistic notions.

The second argument comes from scientific research[11]: We know that smoking causes lung cancer, but obviously not all smokers develop a cancer. Further, we know of no definite set of circumstances where smoking is invariably followed by lung cancer. “Rather, what we observe is that smokers develop lung cancer at much higher rates than non-smokers; this is the prima facie evidence that leads us to think that smoking causes lung cancer.” (HITCHCOCK, 2002). Thus, the concept of probabilistic causation is pragmatically unavoidable e.g. in clinical contexts.

The third indication for the legitimacy of causal relations with a probabilistic aspect is that modern physics theories, following the so called “standard (or Copenhagen) interpretation” of quantum mechanics, make use of probability as a basic concept (cf. SCHLEGEL, 1970). This led to heavy discussions within physics[12] and was plainly rejected e.g. by Albert EINSTEIN[13], which led to rival approaches like the BOHMIAN interpretation of quantum mechanics (which is a specimen of so-called “hidden variables” theories), which tries to evade indeterminism (cf. PINCH, 1979). Today, the “standard interpretation” is widely accepted, and since we certainly cannot judge on any of these approaches, we should not rule out probabilistic causal relations *a priori*. Instead, we accept that the world might be indeterministic in the following sense: “[...] there are actual events that might have failed to occur without violation of any actual laws.” (RAMACHANDRAN, 2004).

2.1.2 Probabilistic Regularity

We were led to chancy causation because we started with identifying regularity and causality, which implies two claims: whenever there is causality, there is regularity, and whenever there is regularity, there is causality. The first one was refuted by examples of chancy causation, the second by examples like Barometer–Storm.

In this section, we will show how a refined version of regularity could deal with the counterexamples of the first kind, so as to keep regularity as a necessary condition of causality: whenever there is causality, there is indeed (a certain, probabilistic, kind of) regularity.

Our solution for the problematic second claim is rather different.We accept Barometer–Storm as a counterexample to the identification of causality and regularity. In our view, causality does

[11] Argument taken from HITCHCOCK (2002).

[12] Cf. COMBOURIEU (1992) for a very emotional interview with KARL R. POPPER on the debates in quantum physics, which includes POPPER’S statement that he “gave up Physics because of Bohr” who was one of the central figures of the Copenhagen interpretation, because “ [Bohr] annihilated me [Popper] with a torrent of words!”

[13] As expressed in his famous quote “I, at any rate, am convinced that He [God] does not throw dice.” (cf. CLARK, 1972).

not only depend on regularity, but also on an additional condition that will be called *counterfactual dependency* (cf. sect. 2.2).

What, then, should a condition look like that, the one hand, supports our intuition and reasoning regarding the regularity aspect of causality and, on the other hand, allows for probabilistic relations as well? As we see it, the solution lies in understanding common sense (i.e. strict) regularity as an extreme case of probabilistic causality in which a certain cause is followed by a certain effect *with a probability of one-hundred percent*. This certainly fits the Security Alarm example and the preconditions of falsification if – like in the cryophon example – the claim at stake does not include any chances. But in order to cover connections like catching a cold when visiting a patient, or developing lung cancer after smoking, we need to reformulate the regularity condition in the following way: *Regularity between A and B means that the occurrence of A heightens the chance of B happening.*[14][15]

This is trivially true in cases where A is *always* followed by B. But it holds for more relations, as e.g. for the relation between visiting a contagious patient and catching a cold. Without visiting the patient, you have a certain general chance to catch a cold. But when visiting, this chance is definitely heightened. This relation fulfils the new regularity condition and is, thus, no longer (erroneously) excluded from being a genuine cause–effect relation.[16]

Let us now come back to the kind of counterexamples that did not rely on chancy causation (causality without strict regularity) but on cases of regularity without causality, as in Barometer–Storm.

2.2 Counterfactual Dependency

There are several possibilities for why a certain A[17] is regularly followed by a certain B even if A and B are not related as cause and effect. In Barometer–Storm, A and B are consecutive

[14] As e.g. MACKIE puts it: “We could say that A tends to produce P not only where A conjoined with some set of other factors is always followed by P, but also where there is an indeterministic, statistical, law to the effect that most, or some, instances of A, or some definite percentage of such instances, are followed by P, or perhaps where an A has a certain objective chance of being followed by a P.” (MACKIE, 1980, p. 76).

[15] In the following we will use the term “regularity” in this probabilistic way. The “old” interpretation will be circumscribed as “strict regularity”, “100%-regularity” or the like.

[16] Some philosphers try to show that even this revised regularity theory fails, as there are – so they claim – causal relations where the cause does not heighten the probability of its effect, but in fact lowers it. We will briefly discuss those arguments after introducing our formalization of regularity (cf. sect. 5.1.2.5).

[17] The question on the nature of the causal relata, i.e. “What kind of entities are connected by causality?” will be addressed in sect. 5.1.1.

effects of the same cause, and the same holds (as we have seen) for certain successive symptoms of a single disease. Additionally, A and B might not be effects of literally one and the same cause, but only share some cause within the line of causes and effects that eventually lead to A and B. Think for example of two clocks that were synchronised in the factory they were built in. So they show the same time with strict regularity but obviously without being directly causally related. It is not the first clock that makes the second show a certain time. It is the mechanics that makes it show the time. But the mechanics was synchronized by the same mechanism, i.e. the synchronicity is somehow based on a common cause. We may extend this example to clocks from different factories or to clocks that use very different mechanisms to "calculate" the actual time, the alleged common cause becoming increasingly fuzzy while the clocks are still synchronised, i.e. while regularity holds. Moreover, we can easily think of events that happen in the same intervals without having *any* causal connection, e.g. ("Cyclist's Watch") the position of the valve of a quite slowly moving bicycle's tyre and the position of the sweep hand of the cyclist's watch. If the cyclist is not in any way made to cycle at this specific slow speed, the regularity between the positions of hand and valve can be completely *accidental*.

A well founded theory of causality has to be able to discriminate between mere regularities and regularities that are results of a cause–effect relation.[18] In our approach, regularity is one of two conditions for causality, i.e. we add the condition of counterfactual dependency to rule out cases of regularity without causality. The following sections will extensively deal with it.

Let us again start with Barometer–Storm. We know that there is no cause–effect relation between the barometer reading and the storm, despite them being connected by regularity. But what makes the difference? Most notably the following: the storm could not have been prevented by fixating the barometer's needle.

And this line of thinking is very common to rule out alleged causes. Say, you notice that the glasses in your kitchen start to clink every time your neighbour listens to loud techno music. You talk to your neighbour and he agrees to stop the noise. But while it instantly becomes very quiet in your kitchen, the glasses still clink. This convinces you that the clinking would have taken place even if the music had not been playing. You therefore infer that whatever the reason for the vibration was, it surely wasn't the music.[19] The general "rule" behind these inferences is the following: if the "effect" would have taken place even if the (alleged) "cause" had not taken place then there is no cause–effect relationship.

[18] Or in terms of sequences: "[. . .] what is our concept of causal as opposed to non-causal sequences [. . .]"(cf. MACKIE, 1980, p. 29).

[19] It might be, e.g. , that your neighbour uses to turn on the music in order to drown out his washing machine's noise. If this is the case (note once again how naturally the counterfactual analysis fits our intuitions on causes and effects), stopping the washing machine should stop the clinking.

Interestingly, HUME had the very same insight, too, as his famous claim, with which we started our reflections on regularity, in fact gives two definitions:

> [. . .] *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, p. his emphasis)

He obviously did not regard this second definition as being different from the first, however, and it was not until the 20^{th} century that DAVID LEWIS proposed a theory of causality that is explicitly founded on this very idea: "if the first object had not been, the second never had existed". The impact of his theory was enormous, and "helped to turn the tide" (COLLINS ET AL., 2004b, p. 1) against regularity theories, which were known to have their drawbacks but nevertheless "dominate[d] the philosophy of causation" (LEWIS, 1973, p. 556) until then. Today, counterfactual analysis has become one of the most important philosophical theories on the topic of causality.

How, then, does this analysis work? And how does it solve our problem with spurious regularities? To give a quick, but not irresponsibly short overview, we have to introduce two central concepts: *possible worlds* and the relation of *comparative similarity* between them. Roughly speaking, possible worlds are ways our world could have been if things had taken a different turn. This includes worlds that look pretty much like ours with very slight differences like a world where you started reading this sentence a second later than you actually did. The differences may also be more extensive, like Latin still being the *lingua franca* in Europe, or e.g. the Neanderthals never having become extinct. There are worlds in which the laws of physics are different to a minor or major extent, e.g. one in which natural constants differ slightly from their actual values possibly rendering carbon based (human) life impossible (cf. HAWKING, 2006). You can compare possible worlds to the worlds created in films, in books, plays or any other kind of fiction, their common feature being a difference from our actual world. That, of course, is where the term "counterfactual" comes from: being different to "the factual", to the way things actually are.

The examples of possible worlds given above already show what comparative similarity between possible worlds is supposed to mean: the possible world which differs from our actual world only by you not reading this text right now (but a second later) is certainly more similar to the actual world than one in which there is no living being in the first place.[20] LEWIS also uses

[20] The judgement obviously is unproblematic in this extreme case, but there may well arise difficulties in closer cases. We will add some remarks concerning the problem of comparative similarity in the formal discussion of counterfactual analysis (cf. sect. 5.1.3.5) but in the end accept that a certain vagueness cannot be overcome.

the expression "being closer to" for "being more similar to".[21] The "one-second-later-world" can then be called closer to actuality than a world in which the natural laws themselves are different from ours.

2.2.1 Possible Worlds Analysis of Counterfactuals

Now that we have introduced the concepts of possible worlds and comparative similarity, the question is: how do they come into play when counterfactual intuition is concerned? How do they help with non-causal regularities? Let us have a closer look at Barometer–Storm: we rule out the barometer reading to be a cause because the storm would have happened anyway, even if we prevented the barometer needle from moving. This *judgement* contains the central aspect. With respect to the storm, what we do know by *experience* are simply the things that actually happen: the pressure drops, the barometer reading falls, the storm arrives. In this very situation, we have no empirical knowledge of "what would have happened if the barometer reading did not drop", as this refers to a counterfactual situation, and as a principle, we cannot measure counterfactual data in the actual world.[22] But – and this is the crucial point – we can *infer*, what would probably have happened. To do so, we imagine possible variations of the situation in question. This is, where possible worlds (and their relative distance, as we shall see in a minute) come into play. We compare the actual setting with variations in which the barometer reading does not drop. Following our counterfactual intuition on causality, we should expect the storm to arrive nonetheless, as there is no causal connection between barometer and storm. But a look at the following examples shows that this is not the case in every possible world:

1. The barometer is broken, everything else is just as in the actual world: air pressure drops:
    - Barometer reading does not drop
    - There is a storm

    Undermines the causal claim (as alleged cause is not present, but effect is).

2. The pressure drops, but the physical laws concerning both the barometer's mechanics and the weather are different to our world:
    - Barometer reading does not drop

[21] Formally, as LEWIS makes explicit, comparative similarity is a weak ordering of worlds where ties are permitted, but any two worlds are comparable, and (additionally) the actual world should be closest to actuality (cf. LEWIS, 1973, p. 560).

[22] Of course, if things had gone differently, we *would have measured something different*. But then the measurement takes place in the counterfactual situation, not in our actual world. Measurement is restricted to the very world where the measurement system actually exists and the measurement actually takes place.

- There is *no* storm

Supports the causal claim.

3. The pressure falls, but the earth is destroyed before the reading can drop:

   - Barometer reading does not drop
   - There is *no* storm

   Supports the causal claim.

Now we have a counterfactual situation where the storm arrives, and we have two counterfactual situations where there is no storm. The barometer reading does not drop in any of them. It is only the first variation that could help to show that there is merely regularity between barometer and storm, but not causality. The storm arrives despite the barometer staying still. The other examples imply the opposite, which leads to the following problem: counterfactuals can rule out alleged causal connections which are in fact mere regularities, but they rely on possible worlds, and possible worlds seem to be too "liberal". They contain worlds which no longer do what we need counterfactuals for – ruling out spurious regularities. Does that mean that the counterfactual approach fails?

Fortunately, it has a second component: comparative similarity, the relative distance/difference between possible worlds, and in particular their distance to actuality. Let us examine this aspect in the examples above: The first world differs only with respect to the barometer. If you were, for example, transported to this world, you would probably never experience a difference.[23] This is no longer true in the second possible world. If the physical laws that both barometer and weather rely upon (behaviour of liquids like mercury or water, deformation of flexible boxes, granulation in alcoholic solution of camphor) were different, this difference would show up in many other devices and situations in everyday life or industrial or scientific use. If you were transferred to this world, you could not avoid noticing the difference. The same holds for the third example, where our planet is destroyed – provided that you are able to experience anything before you vanish.

It may well be a matter of argument whether world number two or world number three is more distant to actuality, but in any case the first example is closer to actuality than the others. And – as we have seen – this very possible world is one which we *can* use to single out spurious regularities. So this is the way to solve our problem: we do make use of possible worlds, but we take their difference from actuality into account. If there is a causal relation between A and B,

[23] Of course, there might be a longer trail of consequences of this single difference: there may be someone who is ordered to repair the barometer, or there might be an air plane accident because the pilot relies on the barometer, all of which does not happen in the actual world. But we believe that this is still a minor alteration compared to the massive changes in the other examples.

the counterfactual "B had not been if A had not been" must hold and it does if there is a world in which A and B both fail and which is closer to actuality than any other world in which A fails but B takes place. In David Lewis' words (in which a "true antecedent" corresponds to our "A fails to take place", and "true consequent" corresponds to "B fails to take place"):

> [...] it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent.
>
> (Lewis, 1973, p. 560[24])

In Barometer–Storm, we saw that this condition is not fulfilled. Worlds two and three in which both events (barometer reading's falling and storm) fail to take place, are more distant from actuality than a world in which the storm arrives notwithstanding the barometer's reading. The counterfactual condition rules out the alleged effect (cf. table 2.1).

World (ordered by closeness)	*supportive / undermining*	*Counterfactual holds*
World 1: Barometer broken	undermining	No
World 2: Physical laws differ	supportive	
World 3: Earth is destroyed	supportive	

Table 2.1: *Counterfactual analysis of "Barometer–Storm"*

When it comes to the connection between air pressure and storm, the result of the counterfactual analysis should be different. To become more familiar with how it works, we will analyse this case, too, before continuing with the ontological analysis of this additional condition of causality.

The claim at stake is: the fall of air pressure caused the storm. The corresponding counterfactual is: if the air pressure had not fallen, the storm would not have arrived. Hence, we need to find possible worlds in which the air pressure does not fall. In some of them, the storm arrives, in others, it doesn't. The counterfactual condition now demands that worlds in which the storm happens despite the air pressure's stability must be more distant to actuality than a world in which stability goes together with no storm.

1. The high pressure area is more stable and lasts longer than in the actual world:
   - The air pressure does not fall
   - There is *no* storm

   Supports the causal claim.

2. The physical laws governing the weather are different. Storms happen at high air pressure:
   - The air pressure does not fall

- There is a storm

Undermines the causal claim.

3. Magic is real and a wizard casts a storm spell:

    - The air pressure does not fall
    - There is a storm

    Undermines the causal claim.

It is not difficult to see that the first possible world is less different to actuality than the others. Both real magic and different laws of nature, would affect much more than just the weather. Using LEWIS' notions, the world in which the antecedent (air pressure does not fall) is true along with the consequent (there is no storm) is closer to actuality than worlds in which the antecedent is true without the consequent. This means that the counterfactual condition is fulfilled and air pressure is "confirmed" as being a genuine cause of the storm (cf. table 2.2).[25]

World (ordered by closeness)	*supportive / undermining*	*Counterfactual holds*
World 1: High pressure stays longer	supportive	Yes
World 2: Different physical laws	undermining	
World 3: Wizard casts storm spell	undermining	

Table 2.2: *Counterfactual analysis of "Air pressure–Storm"*

Using comparative similarity we are now able to select those possible worlds which the counterfactual condition "B had not happened if A had not taken place" relies upon: we now have a second criterion for causal relations which allows us to refute alleged causes that "slip through" the condition of regularity.

As yet, our theory of causality does not contain possible worlds and the relation of comparative similarity, so we will have to add them. But before we do so, these notions have to be analyzed ontologically: what kind of entities are possible worlds? And: do we have epistemic access to them in order to evaluate them being supportive/undermining?

[25] Strictly speaking, these examples do not show that *every* such world in which the consequent is false is more distant to actuality than one in which antecedent and consequent both are true, even if it is true for the given examples.
Please note that the examples rely on our loose introduction to possible worlds. In sect. 2.2.3, we will present our theory of the nature of the possible worlds in question, which is much more restrictive than declaring everything imaginable a "possible world".
Within this approach, then, the aforementioned problem becomes a epistemological one, and we will admint that we simply never can access *all* alternative situations. So our inferences on counterfactual dependency are never fail-save.

2.2.2 Theories of Possible Worlds

Advocating the counterfactual analysis of causality, DAVID LEWIS proposed a theory of possible worlds (cf. LEWIS, 1986) which – put dangerously succinctly – says that possible worlds are just as real as our world but spatio-temporally distant, and not accessible from our world (cf. fig. 2.1). If you were an inhabitant of a certain world w, that world is the actual world to you, just like our world α is actual to us. Viewed from w, the world α is simply one of the myriad of possibilities, but in no way something special. This theory, which treats all possible worlds as equal, is often referred to as *possibilism* (or *modal realism*).

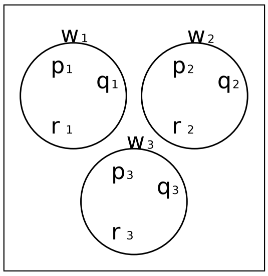


Figure 2.1: *Possible worlds in possibilism: All worlds are equal and they are not connected. Each of them is actual to their inhabitants.*

In contrast to possibilism, ALVIN PLANTINGA has developed a theory in which there is just one – our – actual world. In a nutshell, our actual world is a (certain) set of states of affairs that do obtain (take place). Other possible worlds are made of the same kind of entities (states of affairs), but may consist of obtaining as well as non-obtaining states of affairs, or only of states of affairs that do not obtain (cf. fig. 2.2). In short, possible worlds may share some states of affairs with actuality (and with each other, for that matter), but many do not (cf. PLANTINGA, 1974). To illustrate this point, let us consider two states of affairs "Stephen Hawkings' writing of *A Brief History of Time*", and "the four dragons' attack on Westminster Abbey". While both states of affairs exist, they differ in that the first obtains while the second does not. Just like all other possible worlds, the actual world is made up of states of affairs, but the difference is that it consists of all the states of affairs that do obtain. This theory, which says that there is only one actual world (a world that is something special), is commonly called *actualism*.[26]

Finally there is a third theory proposed by Nicholas RESCHER (cf. RESCHER and PARKS, 1973; RESCHER, 1979, 1999) which is based on yet another idea: possible worlds exist only in people's minds (cf. fig. 2.3). They are ideas or thoughts about how the world could be different.[27] In this view, possible worlds are neither of the same kind as the actual world (possibilism) nor

[26] For a detailed analysis of PLANTINGA'S theory in contrast to DAVID LEWIS, cf. MICHALEK (2002) (in German).

[27] In slightly more detail: possible worlds consist of possible objects which are variants of actual individuals' essential properties (cf. RESCHER, 1979).

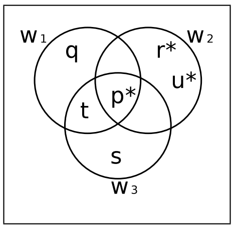


Figure 2.2: *Possible worlds in actualism: The worlds are "made up" of the same entities (states of affairs) which differ in whether they obtain (starred characters) or not. The actual world consists of exactly those states of affairs that do obtain (w_2).*

made up of the same entities as the actual word (possibilism/actualism). They exist in a very different way that is bound to subjects with the capacity of thought. This approach might (for our purposes) be called *subjectivism*.

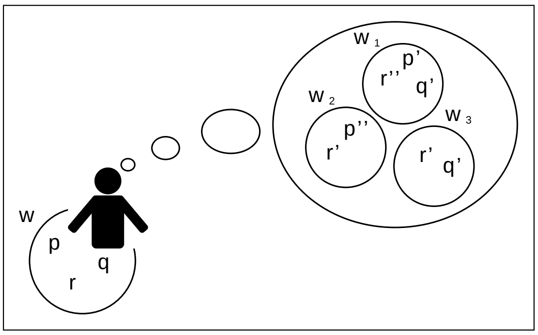


Figure 2.3: *Possible worlds in subjectivism: The worlds are variations of actuality created by a mind.*

We will not discuss these theories at length, but we must mention what major problematic consequences we would face if we simply adopted any of these theories. After all, it is due to these issues that we decided to create our own theory, which is not a theory of possible worlds in general (generality being the route of the troubles) but a theory of just the kind of possible worlds that is relevant for the counterfactual condition of our theory of causality. It will be presented in the following section.

Let us take a look at the shortcomings. In possibilism, the sentence "I could have died" is true because a person who is my counterpart in another world dies in that possible world. This raises two problems: firstly, what makes this very person be my counterpart? In other words, how can trans-world-identity be understood? And secondly, isn't there a fundamental difference between my possible death and some other person's *exitus*?

Actualism does not have these problems. There is no counterpart of mine involved. "I could have died" is about me (although in a different set of states of affairs than the actual world), not about some counterpart. That is why it is relevant to me. But if we adopted PLANTINGA'S theory, we would have to face heavy discussion in fields like the nature of states of affairs, and on PLANTINGA'S notion of "essence" (as his theory makes use of "essential properties"). Again we cannot present those problems for reasons of brevity, but we must state that several outcomes of the actualist's theory are highly controversial (cf. MICHALEK, 2002).

So what about the subjectivist approach? Firstly, it entails all the general problems of subjectivism: possibilities are mind-dependent, so they rise and fall together with their thinkers. Calling something "possible" now depends on choosing a certain mind to refer to. Also, no two minds can think of one and the same possibility, as they produce two thoughts. And secondly, the ontological framework of GFO (General Formal Ontology, cf. chapter 4), which our theory of causality shall become part of, does not incorporate explicitly subjectivist concepts, so we should not introduce them unless absolutely necessary. And indeed we believe they are not, as shall be shown in the following.

2.2.3 Causal Counterfactuals

As noted above, we believe that the aforementioned problems arise because these theories on possible worlds are very general, trying to cope with possibility as such. All *we* need is a theory of the very kind of possible worlds that is (only) relevant, and in fact necessary, for our counterfactual theory of causality. To build a theory that meets our needs we should recall the role it plays in counterfactual analysis: counterfactual dependency holds iff worlds in which the effect takes place without its cause are more different from actuality than worlds in which neither cause nor effect occur. Worlds in which both do not happen must be "stranger" than worlds in which the effect exists alone.

Earlier (cf. sect. 2.2.1) we said that we cannot measure what happens in possible worlds, but we can *infer* what would have happened if things had been different. This requires the ability to compare possible worlds. In possibilism, we would have to have (mental or sensorial) access to those worlds which are *ex hypothesis* not spatio-temporally connected to our world. This is not an attractive solution for sure. In actualism, we only need access to the sets of states of affairs the possible worlds are made of. This is not so grave a consequence, and it suits the idea that we compare situations which are variations of the actual incidents. Since, in actualism, possible worlds are made of a set of states of affairs, we can (mentally) replace some of these while keeping others, which yields more or less similar situations in terms of comparative similarity. In subjectivism, possible worlds are already products of our mind, so we don't need any special means to access them. Thus, actualism and subjectivism allow the comparison of worlds in thought, i.e. using our (mental) capacity of reason. But – and this is where we part with both theories – possible worlds are still not open for empirical methods. The judgement whether the relevant counterfactuals hold is left to mind-equipped subjects with access to either states of affairs or the part of mental activities that construes possible worlds.

In our view this can be avoided, and in order to explain our theory of *alternative situations* (or: *alternatives*) [28], we once again take a step back, this time to the point where we said that we are able to infer counterfactuals to hold. Instead of simply stating that this includes comparison of alternative situations let us take a closer look at what exactly we compare the actual situation to. In our view, this comparison is no different to what happens all the time when we are trying to find our way through the world: we take our *experiences* into consideration. We compare the present situation to situations that have already happened. This means that we do not compare "our world", understood as containing the history of literally everything there is, was and will be to another "world" with its own alternative history of everything that existed, exists and will exist there. What we compare[29] is an actual situation with past situations, all taken from the history of *our* world (cf. fig. 2.4).

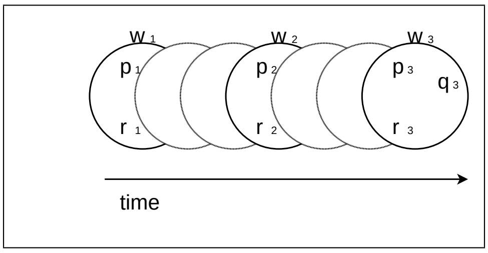


Figure 2.4: *Possible worlds in our account: Snapshots from the history of actuality, i.e. from the history of our world.*

Leaving the epistemic implications aside (we will deal with them later, cf. sect. 6), this approach solves the problem of how we access alternative situations: it works the very same way we get in contact with the actual world. No special mental abilities are required.

Let us summarise our account of counterfactual dependency. Besides regularity, causality is up to counterfactual dependency, which means that if there is a causal relation between A and B, and A and B both happen in the actual situation, *there must be situations (in the actual world) similar to the actual one, and for these the following must hold: situations in which only the cause (A) is missing (which are called* "undermining*" in the following) must be more distant to actuality than at least one situation in which both, A and B are absent (these we call* "supportive*").*[30]

[28] In the following we will use these terms instead of "possible worlds" to make clear that our approach is no full-fledged account of possibilities in general but just about the kind of possibilities used in counterfactuals as applied in our theory of causality.

[29] The mental ability to compare past situations should not be mixed up with the subjectivist claim that the alternative situations exist in our minds, only.

[30] In our aproach, we will model "A and B happen/fail" by the existence or non-existence of a certain (cause or effect) universal's instance, cf. sect. 5.1.3.

2.2.3.1 *Exkursus:* Do the Alternative Situations Suffice?

It is worth noting that compared to how we introduced alternative situations in the first place, our theory on counterfactuals is highly restrictive in at least two ways. First of all, an alternative is no longer a complete space-time worm of a full universe but restricted to temporally flat parts of our world's spatio-temporal extension. Furthermore, an alternative is a situation in the actual world, which excludes (as far as we know) magic, strongly altered physical laws and the existence of entities that never show (or showed, or will show) up in our world. It is this fundamental restriction that made us choose the term "alternative situation" instead of "possible world".

There might be problem with our restrictive notion of alternatives. Some of the examples we used to illustrate the intuition of relative similarity between possible worlds (cf. sect. 2.2.1) made use of worlds that are now excluded (e.g. physical laws are drastically altered, wizards cast storm spells). But since relative similarity is necessary for counterfactual analyses, we shall provide some examples of *how different* alternative situations can be, even if they are restricted in our sense, i.e. to situations in our world.

In fact, our first example (Barometer–Storm) does not need too many adjustments. The first alternative is one in which the barometer was broken, which has undoubtedly happened many times in our actual world. The third is one in which earth is being destroyed. Even if we might feel uneasy about it, we cannot rule this event out for the future of our world like we most probably can with the second possibility of radically different physical laws. But for those who want to exclude both, the destruction of the Earth and altered physical laws, we may add another alternative situation, in which (as in the first one) the air pressure falls and the barometer is broken, but which also features a highly sophisticated (and – at the time of writing – yet to be developed) kind of cloud seeding that prevents rain, thunder and lightning so that there is nothing we would call a storm in the first place. If we compare this alternative to the first one (barometer broken, everything else unchanged), there is no doubt that an additional advanced cloud seeding renders this alternative more distant from actuality. In other words, the first alternative situation (no barometer drop, but storm arrives) is still closer to actuality than alternatives in which both barometer drop and storm are absent. The barometer is thus rightly eliminated as a cause of the storm (cf. table 2.3).

World (ordered by closeness)	*supportive / undermining*	*Counterfactual holds*
World 1: Barometer broken	undermining	No
World 2: Cloud seeding	supportive	

Table 2.3: *Counterfactual analysis of modified "Barometer–Storm"*

The two "more distant" alternative situations in the second example (air pressure and storm), however, are both excluded from our theory of alternatives. Presumably, neither different physical laws nor “real” magic have ever been or will ever be part of our world. To show that counterfactual analysis is still valid, we will present another alternative which is allowed in our approach: an alternative in which the air pressure does not fall, but a storm is created artificially (say, with the help of steam-and-tension-producing – again, at the time of writing not yet developed – devices). Clearly, this world be more distant to actuality than the alternative situation in which there simply is no fall in air pressure and no storm. Again, compared to the constant air pressure going together with no storm, this alternative situation is more distant to actuality. The conditions of counterfactual dependency are fulfilled, so air pressure is correctly *not* ruled out as a cause of the storm (cf. table 2.4).

World (ordered by closeness)	*supportive / undermining*	*Counterfactual holds*
World 1: High pressure stays longer	supportive	Yes
World 2: Artificial storm	undermining	

Table 2.4: *Counterfactual analysis of modified “Air pressure–Storm”*

We hope that these examples show that even our highly restrictive concept of possible worlds leaves enough room for significant differences in relative similarity as required for counterfactual analysis.

2.2.4 Preemption

Now that we have added counterfactual dependency to our theory, we cannot continue until we have dealt with a major issue that has been raised against it: *preemption*.[31]

It actually caused LEWIS to partially revise his initial theory (cf. LEWIS, 2000b). We will now give a quick overview of this problem and will dedicate the next section to our solution.[32]

The most basic case of preemption (in a slight variation of HALL (2001)) works as follows: imagine two bored students in a maths class who try to pass the time with a rather special competition. They crumble up sheets of paper, and once the professor turns to write formulas on the blackboard, the two students aim their paper balls at an open window. Both balls succeed in passing the window frame, but unfortunately they destroy a spider’s web that was woven within the frame. For the sake of this argument, we assume that the first student’s ball reached the cobweb earlier, so it was *his* ball that destroyed the web.

From a counterfactual perspective, this conclusion is problematic because the web would have been destroyed (by the second ball) even if the first ball would not have been thrown. Thus, the counterfactual between the first ball and the web does not hold. We seem to have causality

[31] Cf. SCHAFFER (2000); NOORDHOF (1999)

[32] For a critique of the adjusted theory, e.g. that it “generates a great number of spurious instances of causation”, cf. MENZIES (2001).

without counterfactual dependency. However, we believe that a careful reconstruction of the situation enables us to keep counterfactual dependency as a necessary condition for causality.

First of all, note that the counterfactual dependency only fails "in the long run".[33] If the first paper ball had not been thrown, the web still would have been destroyed *in the end*. But we should certainly be more interested in what happens *right after* the first ball hits the web. Since the second ball reaches the window later than the first, the web would not have been destroyed at that point (or, for that matter, at any other point until the second ball's arrival). Hence, if our causal theory is suitably fine-grained with respect to time, the counterfactual *does* hold. We will deal with the nature of causal relata in sect. 5.1.1, and as we will argue for (temporally) most fine-grained entities – called *presentials* – as primary causal relata, we can safely keep the condition of counterfactual dependency.

Our current solution to the problem of preemption is based on the time difference between the first and the second ball of paper.[34] What happens if the two balls reached the cobweb at the same time? First of all, this is no longer preemption since no potential cause is cut off by another. But we may still call it *overdetermination* as either of the balls would have been sufficient to destroy the web. Secondly, it is no longer entirely clear what the "right" result of a causal analysis should be. If only one ball touches the web, it definitely is the cause of its destruction, but in this case we have two balls. We may agree that at least the two balls together (taken as one entity) can be regarded as the cause, but when it comes to each individual ball, would we say that both constitute separable causes? Or just one of them? Or neither of them on its own? We seem to lack causal intuitions regarding the separate balls. Then, overdetermination cannot be used against counterfactual analysis, here. It does remain a relevant topic, though, and we will address it in a later section (cf. sect. 5.3).

2.2.5 Background Chances as a Challenge to Counterfactual Dependency

When discussing regularity (cf. sect. 2.1), we argued for probabilistic/chancy causation and explained how to include chances in the regularity condition. In this section, we will do the same for counterfactual analysis.

At first glance we would probably not consider the concept of chance a problem for counterfactual analysis. Let us assume that an effect E does counterfactually depend on cause C, i.e. E would not have happened if C had not happened. Following RAMACHANDRAN (2004), one might say that this is perfectly consistent with the assumption that C does not always lead to E.

[33] This argument was first introduced by LAURIE A PAUL to overcome another possible solution: that the spider web would have been destroyed differently if the seconnd ball would have hit him. It turned out that the so called "fragility" of events was itself a very fragile concept (cf. PAUL, 1998).

[34] For further discussion on the importance of the time difference, cf. MACKIE (1992).

Even in a situation that involves chance, E would not have happened without C.

However, problems arise when "E has a *background chance* of occurring" (RAMACHANDRAN, 2004, p. 388 his emphasis), which means that even if the cause is absent, E can happen more or less spontaneously. It is the "would not have happened" bit of counterfactual analysis which seems to raise doubts and as mentioned above, we will solve this problem by introducing chances in counterfactual analysis, as well.

2.2.6 Probabilistic Counterfactual Dependency

Just as we adopted the regularity condition from "Every time C happens, E happens" to "the chance of E happening is higher given that C happened", we need to change the counterfactual expression "E would not have happened if C had not happened" to "the probability of E would have been lower if C had not happened". That said, this idea has to be spelled out in the right way. It must not be read as a regularity statement about "not-C" , like e.g. "the probability of the effect is lowered, if not-C is given", which would hardly add anything other to causality than what regularity already implies.[35] In particular, alternative situations – that build the core of any counterfactual – would not play *any* role in this understanding.

In order to arrive at a more reasonable interpretation of probabilistic counterfactuals, let us examine where the 100% connection comes into play in the counterfactual theory we have developed up to here: it is in clustering the alternative situations into supportive and undermining ones. And as long as we take only single alternatives into account, there is no way to include chances: a *contrastive* situation (where the alleged cause is missing) either does contain the effect, or not. Additionally, the point about the alternatives (once they were sorted into supportive and undermining ones) was their distance to actuality. Our task, then, is, to find a replacement for single alternative situations that allows for both, expressing chances and a comparison with respect to actuality.

Our proposal is the following: where we hitherto spoke of single alternatives, we now introduce *clusters* of alternative situations. I.e. sets of alternative situations which are similar to each other.[36] Within these clusters, we can evaluate the probability of an effect, which now no longer is binary ("does happen" or "does not happen"), but has some value between 0% and 100%.

The first important cluster is that which stems from the initial situation. It will play the role of a reference cluster. Within this cluster, we can evaluate the *initial probability* of the effect. More precisely, we do not take all the similar situations of that cluster into account (as the cluster

[35] In fact the following holds:
$P(A|B) > P(A) \rightarrow P(A|\bar{B}) < P(A)$
i.e. if the cause heightens the effect (regularity), then it's absence lowers it. (Cf. appendix A)

[36] "Similarity" will be modeled by universals' instances. For details, please cf. the formal descriptions in sect. 5.1.

may include situations that are indeed similar to the initial situation, but yet do not contain the cause, i.e. the similarity may not rely on the alleged cause being present), but only those which do contain the alleged cause.

An example may illustrate this idea: say we are investigating whether using a certain switch lights up a given bulb. What could be a similarity cluster around this situation? Actually, there are unlimited possibilities, but let us take this one: “situations in which some switch is used”. In order to evaluate the initial probability, we restrict this cluster to those situations, in which the very switch is used that we are interested in (as this is the alleged cause).[37] Within these, the probabiliy of that very bulb to light up is (depending on the switch’s reliability) e.g. 99.999%.

The same kind of clustering like with the initial situation is done to the alternative situations: instead of sorting them into supportive and undermining *situations*, we cluster them by similarity and call *the clusters* supportive or undermining depending on whether the absence of the cause does lower the probability of the effect (compared to the initial probability) or not; the probabilities being evaluated *within* their respective clusters.

Concerning distance to reality, it is now the clusters we compare and probabilistic couterfactual dependency holds iff there is a supportive cluster that is closer to the initial (or: reference) cluster than any undermining one. Figure 2.5 gives an overview.

2.2.6.1 Example: Catching the Flu

Let us apply this analysis to an example, and (in case the claims are controversial: just for the sake of this argument) let us assume that people indeed can catch the flu when visiting a patient that is already suffering from the flu, that there are other factors which might produce flu (like swimming in cold water), and that there is also a chance to develop the flu without external triggers. The initial situation is one where someone visits a patient and catches the flu. The alternative situations contain visits as well as non-visits and flu-catching/developings as well as non-catchings/developings.

Next we cluster these situations by similarity, which gives the following clusters:

1. A healthy person visits a patient who suffers from the flu.

2. A healthy person visits a patient who does not suffer from the flu.

3. A person with a compromised immune system takes part in an autumnal triathlon and does not visit a patient.

[37] Both, what cluster to evaluate, and what to take as the alleged cause may be choosen differently to how it was done in this example. The conclusion whether counterfactual dependency holds or not is relative to this choice. Cf. sect. 5.1

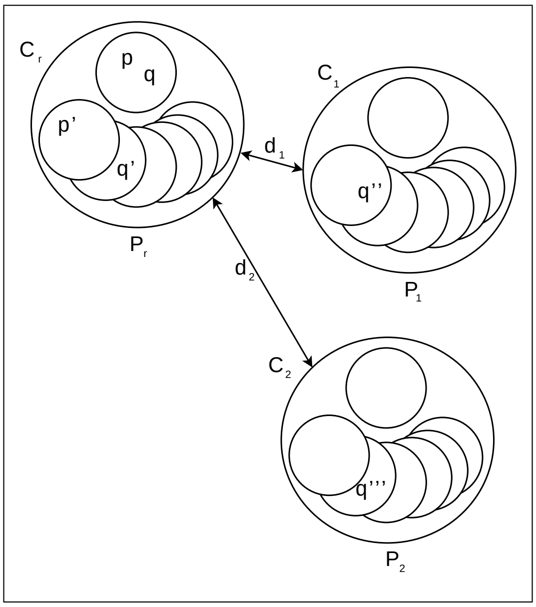


Figure 2.5: *Clusters of alternative situations: the reference cluster c_r with the effect's probabiliy of P_r, two contrastive clusters of alternatives with the probabilities effect's probabilities of P_1 and P_2. Additionally, the distance of the alternatives to the reference cluster is depicted by the length of arrows d_1 and d_2. If C_2 is supportive and if it is closer to C_r than any undermining one, counterfactual dependency holds.*

Evaluating the onsets of the flu in each cluster gives the probabilities P_1, P_2, P_3 for catching/developing the flu in cluster 1 (initial probability), cluster 2 or cluster 3, respectively. And we know that P_2 will be lower than P_1, which makes the second cluster a supportive one (cause is missing, probability of effect becomes smaller). P_3, on the other hand, may well be – and for the sake of this argument, we assume that it is – greater than the initial probability which makes it an undermining cluster (cause is missing, but probability of effect is even higher).

Obviously, cluster 2 is closer to the initial situation's cluster than the undermining cluster 3 that has nothing to do with any visits. So in the end, the probabilistic counterfactual holds (cf. table 2.5).

World (ordered by closeness)	*supportive / undermining*	*Counterfactual holds*
Cluster 1: Visit non-flu patient	supportive	Yes
Cluster 2: Autumnal triathlon	undermining	

Table 2.5: *Probabilistic counterfactual analysis of "Catching the flu"*

2.3 Manipulability

A very important aspect of causality is that we use causal knowledge to change the world: in ancient times, ways to keep people dry by living in caves or building primitive huts were highly important, as were methods for starting a fire. Later on, mankind found out how to prevent strong wind to damage our homes, and how to use a metallic rod to resist lightning strikes. Building machines was another very important step for human development, and again it relied on causal knowledge used to manipulate the world. And looking at our time, sciences that do *not* seem to have immediate benefits in terms of applicability often have a hard time explaining why they should be valuable at all.

This "pragmatic" value of causal knowledge gave rise to another branch of causal theories that take *manipulability* as the core characteristic of any causal relation. The meaning of "manipulability", however is multifarious: early theories (cf. VON WRIGHT, 1971; COLLINGWOOD, 1940) took manipulation as changes, brought about by human action. In COLLINGWOODS words: "[...] the cause of an event in nature is the handle, so to speak, by which human beings can manipulate it." (COLLINGWOOD, 1940, 296). This view, however, gave rise to criticism centring around their apparent *anthropocentricity* and the reductive status of the theories was questioned, accusing manipulation theory of *circularity* (both aspects will be discussed in the following). Later theories (cf. SPIRTES ET AL., 1993; HAUSMAN, 1998; WOODWARD, 1997, 2000) tried to find characteristics of certain (not necessarily human) interactions that qualify as manipulations in the sense of a manipulation theory. To distinguish between the two approaches, we will use the term *intervention* for those that do not depend on human interaction, as is rather common in this field. However – as we see it – the intervention theories, while avoiding anthropocentricity, still fail in being reductive analyses of causality.

2.3.1 Anthropocentricity

If causality is closely tied to *actual* human interaction (i.e. human interaction that does take place right now, or took place in the past), the existence of causal relationships relies on there being humans that perform actions. This would imply several highly counterintuitive claims, two of which are: (1) there were no causal relations until human beings evolved – which was the case for the major part of the universe's history, and (2) there cannot be any causal relations in places where no human being has yet been – which excludes not only nearly the whole universe, but (just think of the deep sea) also a major part of the planet we inhabit.

But even if causality is based not only on actual, but also on the *possibility* of human action, those causal relations are ruled out, where humans cannot take any action at all – which excludes causal relationships in environments where humans cannot exist (think of processes within the sun or within it's immediate proximity, or processes that took place in the uncomfortable environment close to the Big Bang). And even if humans may be able to exist, in many cases, they do not have any influence either because they are too big (as e.g. in the world of nano particles) or because they are too small (think e.g. of earth tectonics or of the movement of celestial bodies).

2.3.2 Circularity

If formulated as reductive theories, manipulation theories face another problem: "human *interaction*" might well be a causal concept itself. But if it is the approach to "translate" the concept of causality to something that already is a causal concept is no reduction at all.

What about the interventionist theories? How do they describe interventions if not connected to human interaction? Here is an example[38]:

> Such an intervention I must meet the following requirements (M1)-(M4):
>
> (M1) I must be the only cause of X; i.e., as with Pearl, the intervention must completely disrupt the causal relationship between X and its previous causes so that the value of X is set entirely by I
>
> (M2) I must not directly cause Y via a route that does not go through X as in the placebo example
>
> (M3) I should not itself be caused by any cause that affects Y via a route that does not go through X, and
>
> (M4) I leaves the values taken by any causes of Y except those that are on the directed path from I to X to Y (should this exist) unchanged.
>
> (WOODWARD, 2001, p.)

As we see, there are quite a lot of requirements, and *all* of them presuppose causal relations, some of which must not hold between certain elements, and some of which must hold in a certain way. So the concept of *intervention* indeed is built upon the concept of causality. Just like with the manipulation approaches, interventional theories (at least as presented here) are not of a reductive kind. [39]

[38] For illustration cf. the graphs in sect. 3.1.1. The conditions (M1) to (M4) that will be given in the following, describe how an intervention like in fig. 3.2 must be designed.

[39] We won't go into more detail, here, but should note that the problem of circularity understood as non-reductivity is an issue, manipulist theorists are completely aware of (cf. WOODWARD and HITCHCOCK, 2003, p. 14). Their answers to this challenge range from declaring human interaction to something that differs from ordinary causality –
"The connection between an action and its result is intrinsic, logical and not causal (extrinsic).[...] It is a bad mistake to think of the act(ion) itself as a cause of its result. " (VON WRIGHT, 1971, p. 67-68).)
– to explaining how "[...] a theory can be non-reductive without being trivial or uninformative." (WOODWARD and HITCHCOCK, 2003, p. 15)

2.3.3 Twofold Manipulability

Acknowledging the aforementioned critics of manipulation theory, we do not adopt manipulation as an additional condition on causal relations. But this does not mean that we neglect the relevance of the manipulation aspect that causal knowledge has. As we will show in the following, the manipulationist's intuition is indeed composed of two components: an ontological and an epistemic one. The first consists of manipulability being a necessary condition for causality, while the latter focuses on *finding* causal relations (or their direction) by means of manipulation.[40] We will come to the epistemological part later (cf. chapter 6) where we straightforwardly accept manipulations as valuable means to identify causal relations.

The ontological content of manipulability on the other hand is, as we will argue in the following, already covered by the two building blocks of our theory, i.e. by regularity and counterfactual dependency.

2.3.4 Manipulation: What is left?

Putting the epistemological question of how to identify causal relations aside, manipulability means that (some) changes in the cause yield changes in the effect. Moreover, if one consequence of manipulability is that we can affect the effect according to our intentions, there should be "predictable" changes in the effect (given specific changes in the cause). Taken this way, manipulability relies on a *regular* connection that holds between cause and effect. But this relation obviously is, what regularity and counterfactual dependency[41] already do provide, so there's no need to add manipulability as a separate criterion.

However, manipulability may shed light on an aspect of regularity hitherto not made explicit as we have not yet focussed on the question of the ontological nature of the causal relata. As this is a cental topic of a later section (cf. sect. 5.1.1), we shall not discuss it here, but some remarks might be admissible. The kind of entity that is probably most easily related to changes or manipulations are *properties* (or whatever your ontology provides as an appropriate surrogate): changing the tension of a bow's string, or manipulating the initial direction of the arrow allows the archer to make the arrow hit a certain position on the target, for example. In order to cover such cases of manipulability, we must take care that our theory – where manipulability is not explicitly included – must (nevertheless) be able to connect properties in an appropriate way. Yet, it is not only the "difference in properties" that may be used in manipulations, but also what

[40] So in an "ontological" reading it is
causality $\Rightarrow$ *manipulability* (manipulability is a necessary condition for causality), while in the epistemiological reading, the direction of the inference changes:
manipulability $\Rightarrow$ *causality* (if manipulability is found, we may infer causality).

[41] In this section, we will simply speak of regularity. All remarks concerning the causal relata of regularity do also apply to counterfactual dependency.

could be called “difference in existence”. We can prevent some effects not only by altering the cause, but also by complete *elimination* of the cause.

One way to combine both kinds of manipulations is to take existence as just another property, a proposal, whose discussion has an honorable history of its own.[42] But even if we do not follow this route, what we do stay committed to (in order to cover the manipulationist intuition) is that out theory does not only have to provide means to connect properties in the right way, but also for properly connecting the *existence and nonexistence* of entities.[43]

[42] A historical reference is Immanuel KANT (cf. KANT, 1787, p. 401). For an overview of the issue cf. NAKHNIKIAN and SALMON (1957); LEJEWSKI (1954).

[43] We will see that the GFO theory of causality takes exactly the opposite route w.r.t. unifying properties and existence. Rather than taking existence as a property; the condition on properties is understood as a condition of existence: changing a property’s value means that the old property value (as entity) is no longer existent, while the new value (as entity) comes into existence. Cf. sect. 5.1.4 for more details.

3 Computer Science Theories of Causality

Looking at computer science, there are very different ways in which causality is dealt with.[44] Starting with the field of statistics, we will present and discuss the causal models developed by Judea PEARL that rest on directed acyclic graphs (DAGs). Then we will concentrate on ontological approaches (DOLCE, Cyc, and the work of John SOWA), as our theory will be based upon an ontology too (called GFO: General Formal Ontology). Indeed, it can be seen as an extension to that ontology.

3.1 Statistics

3.1.1 Directed Acyclic Graphs (J.Pearl)

The probably best known approach to formal causal representation (also called "causal modeling") is the one connected to Judea PEARL and his causal interpretation of BAYESIAN nets.[45]

Those nets (cf. fig. 3.1) consists of two components: a directed acyclic graph (DAG) consisting of vertices V and edges E, plus the local mechanisms[46] $k^v(x_{pa(v)}; x_v)$ – associated to each vertex v – that generate/compute the output values of the x_v given the values of the parents of x_v (where "generating" may include stochastic mechanisms): $\mathcal{B} = (V, E, (k^v)_{v \in V})$

In figure 3.1 there is a very simple DAG with four nodes and four directed arcs. To each node there is a local mechanism assigned. For this net's joint distribution, the following holds:

$$p(x_1, x_2, x_3, x_4) = k^1(x_1)k^2(x_1; x2)k^3(x_1; x_3)k^4(x_2, x_3; x_4)$$

with the general joint distribution of any DAG configuration being

$$p = p(\mathcal{B}) = \prod_{v \in V} k^v(x_{pa(v)}; x_v)$$

[44] We will roughly follow the distinction between "Numerical" and "Symbolic" approaches as introduced in LEHMANN (2003).

[45] This introduction follows PEARL (1993, 1994, 2000)

[46] The idea of using "mechanisms" to get a grasp on causality goes back to SIMON (1977), cf. PEARL (1994)

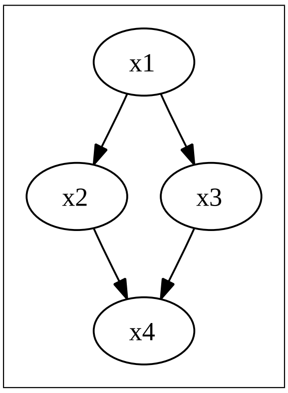


Figure 3.1: *Simple DAG with four nodes ("Diamond")*

Up to here, the DAGs can be interpreted as "carriers of independence assumptions" (PEARL, 1994, 4). However, the picture (literally) changes, if we do not stick to the *observational* variables x_v but allow for *interventional* ones. Figure 3.2 shows the diamond with an intervention on x_2 that solely sets the value of x_2 breaking all other parental connections of x_2 (i.e. the edge between x_1 and x_2).

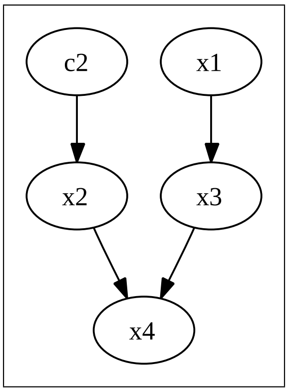


Figure 3.2: *The diamond DAG with an intervention on x_2*

As the mechanisms stay the same (the nodes are "modular" i.e. "it is conceivable to change one such relationship without changing the others" (PEARL, 2000, p. 22)), the joint distribution under the intervention c_2 can be calculated as

$$p(x_1, x_2, x_3, x_4 \mid\mid c_2) = k^1(x_1)\delta_{c_2}(x_2)k^3(x_1; x_3)k^4(x_2, x_3; x_4)$$

with $\delta_{c_2}(x_2)$ replacing $k^2(x_1; x2)$ from the original diamond's formula. Under the intervention, we find that x_3, for example, is no longer dependent on x_2. There is no link of "causal influence" (cf. PEARL, 1994, 3) between them.

The difference in content is this: while Bayesian networks contain information about observable distributions of the vertices' values, the causal DAGs tell us, what observables would change, if an intervention were to take place. In PEARLS words:

> A joint distribution tells us how probable events are and how probabilities would change with subsequent observations, but a causal model also tells us how these probabilities would change as a result of external interventions – such as those encountered by policy analysis, treatment management, or planning everyday activity.
>
> (PEARL, 2000, p. 22)

3.1.1.1 Analysis

First of all, we must state that PEARL is not explicitly interested in *what* causality is, but aims his analysis at two epistemological questions: "(1) What empirical evidence is required for legitimate inference of cause–effect relations? (2) Given that we are willing to accept causal information about a phenomenon, what inferences can we draw from such information, and how?" (PEARL, 2000, p. xiii).

But clearly, what lies conceptually (i.e. putting aside the remarkable analysis of causal DAGs as mathematical objects) behind modeling causality by causal DAGs is the manipulation account of causality: a relation is causally relevant, if the alleged effect is dependent on the alleged cause under an "atomic" (PEARL, 2000, p. 70) intervention (on the cause). Furthermore, he calls "influence, manipulation and control" the "more basic notions associated to causation" (PEARL, 1994, 5).

3.2 Ontology

Although we shall not go any further into detail about what an ontology is[47], we will collect some constituents, that at least the following approaches have in common:

- Categories are used to structure the (knowledge) content in question
- Hierarchies are used to structure the order of categories
- A formal language (based on mathematical set theory and first order logics) is used for machine readable representation and
- Natural language sentences are provided to help the reader understand the concepts in question.

[47] In computer science ontological literature GRUBER'S definition plays the role of a classical dictum: "An ontology is an explicit specification of a conceptualization." (GRUBER, 1993, p. 1), but in almost any field of research where ontologies are considered as playing an important role, people have developed their own understanding of the term. For an impressive overview cf. GUARINO (1998).

These constituents nicely depict the double nature of formal ontology as touching both, philosophical (categories, concepts and their order) and computer science issues (machine readable formal representations).

3.2.1 DOLCE

Developed as a module[48] of the "WonderWeb[49] Foundational Ontologies Library" the aim of the "Descriptive Ontology for Linguistic and Cognitive Engineering", DOLCE, is to "capture the ontological categories underlying natural language and human common-sense", which the authors call a "clear *cognitive bias*", as they are not interested in the "intrinsic nature of the world", but in the "cognitive artifacts ultimately depending on human perception, cultural imprints and social conventions". (MASOLO ET AL., 2003, p. 13 their emphasis).

This "cognitive bias", however, does not mean that DOLCE's choice of basic categories (cf. fig. 3.3) is profoundly different to other, rather realistically oriented, top-level ontologies. Roughly spoken, the difference is not about how to conceptualise the world, but about what "world" is to be conceptualised. In case of DOLCE, it is the world of our cognition and language – independently of how it may correspond to an external reality.

The DOLCE theory of causality was presented in LEHMANN ET AL. (2004) which will be the main reference for this section's content.

3.2.1.1 The DOLCE Theory

We will start our quick, informal overview of the relevant concepts of the DOLCE theory of causality with those entities that are not causality related (cf. LEHMANN ET AL., 2004, sect. 4.1):

- *Physical endurant* Located in space and time, wholly present at any time it is present (no temporal parts). Examples are: a car, Barack Obama, the K2, an amount of gold
- *Perdurant/Event* Temporally extended entity. The authors give reaching the summit of K2, a conference and eating as examples.[50]
- *Physical/Temporal quality* This entails "'aspects' of entities that can be perceived and measured like shapes, colors, lengths, speeds and energies" as well as temporal locations

[48] Alongside OCHRE and BFO (cf. MASOLO ET AL., 2003; SCHNEIDER, 2003b,a; GRENON, 2003).

[49] Cf. `http://wonderweb.semanticweb.org/index.shtml` (as of 2007/06/29); according to this site, the WonderWeb project officially finished in Juli 2004. The final report is HORROCKS (2005).

[50] Just like the authors, we will use "perdurant" and "event" synonymously.

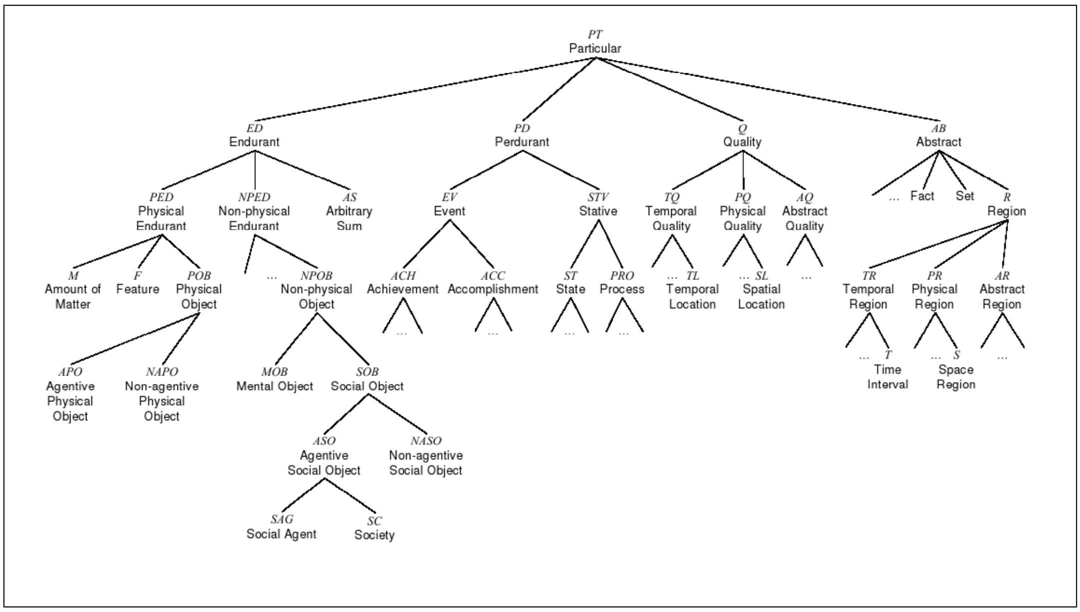


Figure 3.3: *Taxonomy of* DOLCE *basic categories (taken from* MASOLO ET AL.*, 2003, p. 14)*

(of perdurants) and spacial locations (of physical endurants).

- *Physical/Temporal region/quale* Following GOODMAN (1951); GÄRDENFORS (2000) and the philosophical concept of "tropes" (cf. BACON, 2002), DOLCE distinguishes between qualities and qualia (singular: quale). In short, a quale is an individual quality's position in "quality space". Having the same (single) quale e.g. justifies speaking of two roses having the same color, i.e. their (distinct, individual) qualities have the same quale.

- *Participation of an endurant to a perdurant* Endurants can participate in perdurants during the full event, or just at certain times.

- *Temporal inclusion/coincidence* Temporal coincidence of perdurants (roughly) means that both entities exist for the same time interval.

In addition to this part of the DOLCE ontology, the authors firstly introduce the (non-causal) concepts of "unique participation"[51] and "common quality change" to define the central category of "basic quality change" (cf. LEHMANN ET AL., 2004, sect. 4.3):

- *Unique participation* At every time there is no other endurant than x participating to a certain event e.

[51] Note that the subsequent expressions "unique participation" and "common quality change" are not used in LEHMANN ET AL. (2004), where the authors define the corresponding predicates ($UPC_C(x, e)$ and $BQC^*(e, x, PQ_i)$) without providing a descriptive term. We introduce them for convenience reasons.

- *Common quality change* There is an endurant x uniquely participating to an event e, and a certain physical property of x has different qualia at different times.

- *Basic quality change* "Perdurant capturing the change of an endurant along just one aspect/quality type". This means that it fulfils the following conditions:
  1. There is unique participation between the (changing) endurant and the change (as an event).
  2. The endurant has a common quality change with respect to a certain pyhsical quality.
  3. There is no part of the event, that temporally coincides with the event (like a horizontal layer) and has a common quality change in whatever physical quality.[52]

In a further step, the authors introduce dependency relations between basic quality changes. With respect to the "temporal relations between quality changes and of the identity relation between their participants" they introduce the following "three different kinds of generic *existential dependence* [. . .] that individuate sets of quality changes" (LEHMANN ET AL., 2004, sect. 4.4, their emphasis; cf. that section for the following as well):

- *Synchronic dependence* Eg. if the shape changes, the spatial locations changes simultaneously. This does not hold vice versa.

- *Backward dependence* This relation covers the idea that some changes in one entity necessitate different changes of other entities that are temporally prior. If, for example, a change in shape takes place, some change in the spatial location must have taken place.

- *Forward dependence* This is the "opposite" of backward dependency expressing that a certain change is to be followed by another certain change.

For covering different simultaneous quality changes, the notions of *multiple forward/backward and mixed back/forward dependencies* are introduced.

With these expressions at hand, the first causal expression is introduced in a tern of more restricted synchronous/forward/backward dependencies:

- *Structural dependence* This relation holds, if there is a synchronous dependency between the basic quality changes of the same object. These represent very general laws based on its "structure (ontological characteristics)" (LEHMANN ET AL., 2004, sect. 4.4).

- *Causality dependence* These dependencies hold between types of quality changes non-synchronically occurring on distinct objects.

[52] We think that the corresponding definitions D3-D5 of LEHMANN ET AL. (2004) should express "in any physical quality different to the one that the basic quality change is about" instead of "in whatever physical quality". A difference, however, that is not of relevance, here.

- *Circumstantial dependence* These are connections between quality changes that take the qualia into consideration:
  - *Intrinsic dependence* with respect to *how* the qualia change (e.g. if one increases, the other decreases).
  - *Relational dependence* comparing the temporal or physical qualia of the participants' qualities, e.g. spatial location at a given time.

Note that *causality* is meant to refer to the relation between types of quality changes, while the authors use *causation* for concrete, individual causal relations. And here is what they understand as causation: Given a set of events that is *structurally closed* (i.e. satisfies the structural constraints of the system), the *causation relation* holds, if at least one of the following conditions hold (cf. LEHMANN ET AL., 2004, sect. 4.5):

1. A forward or backward dependency holds between the two relatas' basic quality change types (or between a the set of events, synchronously dependent on the relata)

2. There are events that are synchronously dependent on the relata and condition 1 holds for them

3. There is an (intermediate) additional event that is on the one hand connected with the first relatum (under the conditions given above) and on the other hand connected to the second relatum, forming a kind of "transitive mediator".

3.2.1.2 Analysis

From an ontological perspective, the DOLCE theory has two important characteristics:

1. The relata are "basic quality changes" (or kinds of "basic quality changes").

2. Causality is handled as a kind of constraint.

The first aspect refers to the ontological nature of the causal relata, which we did not yet discuss. However, we will address this question here, without anticipating too much of later discussion (cf. sect. 5.1.1).

The second aspect locates the DOLCE theory within the regularity theories[53], as the constraints are expressed by dependencies that are defined in terms of existential conditionals.[54]

[53] The authors are aware of that: "We proposed to look at these constraints as forms of dependencies among event types that cover physical laws". (cf. LEHMANN ET AL., 2004, sect. 6).

[54] As for example the definition of "synchronic dependence" (sect. *4.4*, notation slightly changed): $sQD(\alpha, \beta) =_{df} \exists(\alpha(x)) \wedge \forall x(\alpha(x) \rightarrow \exists y(\beta(y) \wedge CN_T(x, y) \wedge pc(x) = pc(y)))$

Thus, it is not only regularity, but what we introduced as “strict” or “100%” regularity (cf. sect. 2.1.2).

This leads us to the following problematic aspects of the DOLCE theory’s basic assumptions:

- Some causes *prevent* changes (i.e. they prevent change processes/events from happening) or account for stability, like e.g. a break prevents a vehicle from moving, or like a sprinkler system prevents a fire or like the “strong force” in physics keeps quarks and gluons together to form protons and neutrons. These causes are not covered in DOLCE
- Looking closely at changes, one may question whether it really is the change that causes anything, and not the “final state” at the end of the change. If two changes lead to the same “final state”, wouldn’t they have the same causal consequences? If you agree to this, it seems like it is not the change that is causally relevant, but the “final state” (however it is evoked).
- The first problematic consequence of relying on regularity[55] is that (subsequent) effects of a common cause erroneously are identified as cause–effect pairs.
- Secondly, strict regularity does not cover probabilistic causal relations, which are very common in e.g. medicine.

However, although we do not agree to these basic assumptions, we acknowledge that the DOLCE theory’s details (like the concept and the kinds of basic property changes, and the connection to the non-causal parts of DOLCE) are impressingly comprehensive, and in fact the most extensive formal ontological analysis up to now.

3.2.2 Cyc

Founded in 1984 with an initial budget of US$ 50 million (cf. COPELAND, 1997), the large-scale knowledge base of Cyc – according to their developers – currently contains “nearly two hundred thousand terms and several dozen hand-entered assertions about/involving each term.”(CYCORP INC., 2008). And as we already mentioned in the introduction (cf. sect. 1.2.2), the makers of Cyc decided to model causality by material implication.

3.2.2.1 Analysis

We already saw that material implication (taken as means to model causality) has the strange consequence that literally everything causally follows from a wrong assertion: “5 is prime” thus becomes the cause of “there is a thunderstorm”. But this is not the only shortcoming.

55 As discussed in sect. 2.1.

Material implication does not contain information about time. So this approach allows for the implication's consequence to be be prior in time than the antecedens. This allows for causal relations that go backwards in time. A view many people would disagree with.

And finally Cyc falls into the trap of taking regularity as equivalent to causality. Every case of non-causally connected regularities we've mentioned so far (Barometer–Storm, subsequent symptoms of an underlying disease, etc.) would be understood as a cause–effect relation.

It should be noted that when searching the OpenCyc[56] concept browser online, you will find that the developers obviously have taken the early critics like those mentioned COPELAND (1997) into account as they added some notes on the "causes" relation that connects propositions in order to make the difference to the "implies" relation explicit. This contains the temporal order (effect proposition must not precede cause proposition), as well as the following: "a Causes Prop Prop sentence presumes an underlying mechanism of causation." (CYC FOUNDATION, 2008). We assume that this "mechanism of causation" is meant to prevent the causal relation from falling into the shortcomings mentioned above, however, OpenCyc does not tell us, how this is done. Actually, no information on these "mechanisms" is given.

3.2.3 Sowa's Theory

3.2.3.1 Continuous Processes

The most basic concept in the ontological theory of John SOWA (as laid out in (SOWA, 2000c)[57], which is the main reference for this section) is that of a mathematical function as used in physics, on which SOWA then relies when defining "continuous processes"[58]:

> "A *continuous process* P is a pair (F, M) consisting of a collection F of differentiable functions defined on a four-dimensional manifold M.
>
> - Every point p of M has an open *neighborhood* U that is homeomorphic to some subset of four-dimensional Euclidean space, E^4. The homeomorphism at p determines a coordinate system x_1, x_2, x_3, x_4 over the neighborhood U.
> - A *path* through M is the image of a continuous map m from a real interval $[a, b]$ into M. The point $m(a)$ is called the *beginning*, and $m(b)$ is called the *ending* of the path.
> - The coordinate x_4 of a point p, which may also be represented as $t(p)$, is called *time*.

[56] OpenCyc is a restricted open source version of Cyc, cf. `http://www.opencyc.org`.

[57] According to (SOWA, 2000a), this text is based on contents of (SOWA, 2000b) plus some additional material.

[58] Formally, the function is the most basic concept, but how to tell which "collection" of functions counts as a process? Every collection? If not, then whatever accounts for the process-identity is even more basic.

(SOWA, 2000c, sect. 2.1, his emphasis)

Conceptually, continuous processes are a certain kind of process (cf. fig. 3.4), whose "incremental changes take place continuously" (SOWA, 2000c, sect. 1), and are thus opposed to "discrete processes", where changes "occur in discrete steps called *events*, which are interleaved with periods of inactivity called *states*." Depending on whether the beginning or ending are of concern, continuous process may be divided into "Initiations" (without ending), "continuations" (wihtout beginning and ending) and "cessations" (without beginning).

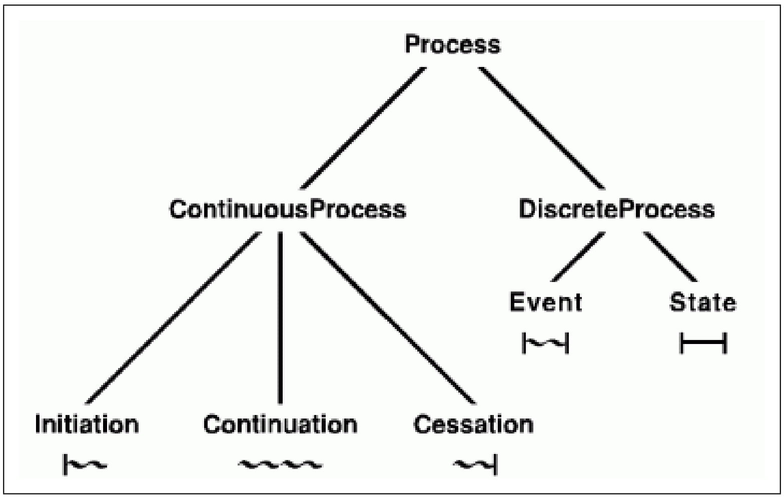


Figure 3.4: SOWA'S *process hierarchy (taken from* SOWA, *2000c, sect. 1)*

With processes and functions, SOWA introduces the first causal notion – that of "causally equivalent" functions:

Let $P = (F, M)$ be a continuous process [...].

- Two functions f and g in F are said to be *causally equivalent* with respect to a point p in M if for any point q in the past with respect to p, $f(q) = g(q)$.

(SOWA, 2000c, sect. 2.6, his emphasis)

In a next step, the author introduces *constraints* on continuous processes:

"A *constraint* on a continuous process $P = (F, M)$ is a predicate $C : 2^F \times M \to \{true, false\}$. If S is any subset of functions in F and U is any open neighborhood of M for which $C(S, p)$ is true for all p in U, then C is said to *constrain* the functions in S on the neighborhood U."

(SOWA, 2000c, sect. 2.4, his emphasis)

The concept of a constraint, thus, is to restrict the values of a function's processes. This includes very detailed constraints (as e.g. in theoretical physics) as well as mere rule of thumb in everyday life (SOWA, 2000c, sect. 2.4).

Certain constraints are then called "causal constraints", which is the core causal notion in SOWA'S theory:

> Let $P = (F, M)$ be a continuous process, and C a constraint on P.[...]
>
> - The constraint C is said to be a *causal constraint* if for any point p, the truth of C at p is unchanged when any function f in F is replaced by another function that is causally equivalent to f with respect to p.
>
> (SOWA, 2000c, sect. 2.6, his emphasis)

The notion of a causal constraint then is a means to discriminate between "law governed", "random" and "deterministic" processes depending on whether there is a causal constraint on some or all functions of the process (on some neighborhood), or whether there is no such constraint, or whether the process is not only law governed, but even more constrained such that the future values of the process's functions are uniquely determined by values in the past.

The theory about continuous processes, however, is just one part of SOWAS causal theory. It tries to cover discrete (i.e. step-by-step) connections as well.

3.2.3.2 Discrete Processes

SOWA introduces discrete processes as a directed, acyclic, bipartite graph consisting of two kinds of nodes ("states" and "events") and ordered pairs of nodes, the "arcs". Depending on what nodes are connected by arcs, the following terms are introduced: Whenever an arc connects two nodes one of which is a state and the other an event, it is said that the first node has a "causal influence" on the second. If the first node of a causal influence arc is a state (connected to an event), the arc is called "input arc" and the state "input state". If the first node of a causal influence arc is an event, the arc is called "output arc" and the second (state) node "output state". The causal influence is defined as transitive (cf. SOWA, 2000c, sect. 3.1).

With the intermediate step of introducing "event [and state] types" (SOWA, 2000c, sect. 3.2), the "preconditions" (which are certain input state types of an event type) and "postconditions" (which are certain output state types of an event type) are called the "signature" of an event (SOWA, 2000c, sect. 3.3).

The immediate causal interpretation of this model comes from understanding the pre- and postconditions as causes and results of the event.

A second way, causality is covered by this theory is that there are axioms that connect the universes of discrete and continuous processes (cf. SOWA, 2000c, sect. 3.3):

Refinement A discrete process can be refined by replacing some state or event node by a new discrete process.

Embedding Maps states and events of a discrete process into points in the continuous processes manifold.

Approximation Whenever there is a discrete process, mapped into a continuous one, there exists another discrete process, which is a) a refinement of the first discrete process, b) also embedded into the continuous process, and c) a better approximation to the continuous one (compared by terms of *error* in *prediction*

3.2.3.3 Analysis

SOWA'S theory is obviously rooted in two scientific domains: The definition of continuous processes as functions on three-dimensional manifolds is very similar to how (theoretical) physics describes the (causal) world, and his treamtent of discrete processes clearly stems from insights of distributed systems modeling (e.g. like with Petri nets) in computer science.[59]

Both, typically, do not wear their causal content on their sleeves. In physics, it is all about functions and the distribution of their values, while Petri nets are usually described as modeling systems with concurrency and resource sharing (cf. PETRI NETS WORLD, 2007; DESEL ET AL., 2004). Researchers in both fields, however, tend to use causal notions (like "A makes B do / become / act like C") in informal settings, which is not surprising, as – as laid out in the introduction (cf. sect. 1.2) – they are modeling parts of the world that belong to the realm of what we call connected by causality. SOWA'S aim, therefore, is what may either be called giving the scientific models a causal interpretation, or it may be termed rooting the concept of causality in the universe of natural sciences' findings. This, undeniably, is an important part of any causal theory that does not want to find itself opposed to natural sciences, which are indeed our best way to discover causal relations.

This being said, what is the conceptual content of SOWAS causal theory? In the case of concrete processes, causes and effects are certain types of pre- and postconditions of event types (the pre-and postconditions being states). The states and events are then connected by arcs of causal influence.

Just like in DOLCE, SOWA seems to rely heavily on regularity: it is *types* of pre- and postconditions of event *types*, which means that similar events (given similar preconditions) by causal influence are tied to similar outcomes. And again like DOLCE, causality is tightly related to changes (which are not as elaborated as in DOLCE, but still are *the* essential criterion to discriminate between states and events).

[59] Another case is the use of numerical methods to approximate e.g. differential equations.

In the case of continuous processes, we have “causal equivalence” and “causal constraint”. The first refers to two processes sharing a history of past values (compared to a point), the second to a certain kind of condition that restricts the values of a process’s functions. Again, it is the idea of regularity that can be found in these concepts. “Causal equivalence” is strict regularity between values in the past, while “causal constraints” are regularities that pertain on a process’s function in general.

Is it strict (100%) regularity, that SOWA is committed to? In the field of discrete processes, the *types* seems to refer to strict regularity, but what about the “causal constraints” in continuous processes? This is not easy to answer, as SOWA does not say much about what these constraints may look like. All that he says is that they might be of different coarseness varying between the “fundamental laws of electrodynamics or derived laws that relate averaged functions, such as temperature, pressure, and heat” and “ ‘People can’t run much farther than a mile in 4 minutes’ or ‘People can’t spontaneously metamorphose into ducks or tomatoes”’ (SOWA, 2000c, sect. 2.8, emphasis removed). Mentioning the laws of thermodynamics may indicate that statistical expressions might be part of the constraint, which might protect this theory from falling into the first major pitfall of regularity based theories.

But how about the second one, i.e. regularity does not necessitate causality? In our view, the notion of a law, i.e. a causal constraint is too liberal, so it fails in non-causal but regular cases. Every regularity may be regarded as a constraint, and those which fulfill certain conditions (e.g. they are limited to “a region called the *light cone*” (SOWA, 2000c, sect. 2.4, his emphasis)) may be called causal constraints. Constraints that refer to the two effects of a common cause would fall under this concept, just as real causal relations.

4 General Formal Ontology: GFO

4.1 GFO and the Project of GOL (General Ontological Language)

Development on the top-level ontology of GFO (General Formal Ontology) started in 1999 at the University of Leipzig, as the central part of a project called "General Ontological Language" (GOL, cf. HELLER and HERRE (2003, 2004b)). Over the years, however, several directions of research have been followed that split the initial project into various fields, such that GFO is now regarded as one (though important) component of a larger ontology–based framework for knowledge representation (cf. ONTO-MED RESEARCH GROUP, 2008).

4.2 GFO Basics

Our first task is quickly to introduce those parts of GFO that are either directly connected to our theory of causality, or necessary to get a grasp of the underlying "GFO-spirit". For the sake of brevity and readability, this overview will not present a finicky description of the relevant concepts within the GFO-concept hierarchy, covering *all* the reasons and problems, but we shall use a more narrative style, which – according to our experiences – is easier to follow, and is not that much in danger of distracting our concentration from the main topic, which is causality.[60]

Starting with some background information, we should firstly be aware that GFO takes a rather "realistic" point of view, when it comes to the entities captured. This is a major difference to e.g. DOLCE, which is – as the authors point out – cognitively biased (cf. MASOLO ET AL., 2003, p. 13). So we will speak of modeling the world (or a domain) instead of modeling an agent's view on the world. In the words of William J. CLANCEY: "The primary concern of knowledge engineering is modeling systems in the world, not replicating how people think [. . .]." (CLANCEY, 1993).[61] Secondly, calling it a "Top-Level Ontology" means that it is concerned with those concepts that are domain-independent i.e. they are needed in almost every specific domain.

[60] For a detailed overview on GFO cf. HELLER and HERRE (2004a) and HERRE ET AL. (2007).

[61] A discussion of various realistic/cognitive/constructivist approaches and their relationship to truth (understood as correspondence) can be found in GUARINO (1995).

Summarising, GFO aims at covering those entities in the world, that are so general that they are of use in almost any concrete domain. And this is a good point to start our survey.

4.2.1 Time and Space

A nice example of highly domain independent concepts are those which express temporal and spacial relations: A public library needs to know, where a certain book is at any given moment, while biologists and medics for example are interested in certain processes that happen in certain places at certain times.

The GFO theory of time is a *glass continuum* (cf. HAYES, 1996), and indeed, a glass stick might be a good analogue to the temporal entities in GFO. Such a stick is spatially extended and always has two endings: a and a'. And because it is all solid glass, there is no structure to be seen within. But once you break it, two new endings, b and b' are created out of the same point of the old stick. So now you have two smaller sticks one with the endings a and b and the other with a' and b'.

The GFO theory of time starts with temporally extended entities called *chronoids* that have exactly two (extremal) *time-boundaries*. And just like the stick can be broken at (nearly) any position, a chronoid can be split up everywhere. In other words: a chronoid has an infinite number of inner time-boundaries that would become extremal boundaries of the resulting new chronoids. Note that just as it does not make sense to call a stick the sum of all the endings any possible breaking would create, a chronoid is not the sum of all its (inner) time-boundaries. Coming back to the splitting, we find that it creates two boundaries out of "the same point" within a chronoid. So there should not be a temporal difference between such a pair of boundaries. GFO introduces the notion of *coincidence* to indicate that such a pair is so tightly connected that there is no temporal gap between them. They are, in a sense, "at the same time", while still being different entities. Using the concept of coincidence, GFO allows for seamlessly "(re-) connecting" two chronoids with the old endings becoming a pair of coinciding time-boundaries, that belong to the inner time-boundaries of the new, bigger, chronoid.

4.2.1.1 Summary

- There are two basic temporal entities: chronoids and time-boundaries.
- Every chronoid has two extremal and infinitely many inner time-boundaries (which are extremal boundaries of sub-chronoids).
- Every part of a chronoid is a chronoid itself.
- A chronoid is not the sum of all its time boundaries.
- Pairs of time-boundaries (one right, and one left time-boundary) may coincide.

A very similar route is taken in modeling space: The basic entities are *topoids*, which have boundaries, as well. The only difference is, that there are topoids of different dimensions, while time is seen as a one-dimensional “line”.

4.2.2 Individuals

The aforementioned entities that *constitute* time and space are relevant in order to establish temporal or spacial relations between entities that are *in* time and space. The latter are called *individuals* in GFO, and they share some characteristics of what philosophical literature calls “particular” or “concrete” (cf. GRACIA, 1995; BUTCHVAROV, 1995).

4.2.2.1 Processes

As the basic temporal entity is a chronoid, we begin with those individuals that are extended in time, like a 100-meter sprint, a series of lectures or the pumping of a person’s heart. GFO calls those entities *processes*, and assigns a chronoid to each process, such that the chronoid *frames* exactly that amount of time the process unfolds in. Another GFO expression for the special relation between a process and its temporal extension is that the process is *projected onto a chronoid*.

The strong connection to chronoids leads to other features of GFO’s processes: Parts of processes are processes themselves, and processes have boundaries, too, which can coincide, if the processes *meet*.[62] Those boundaries will be subject to the next section.

4.2.2.2 Presentials

Processes cannot only be projected onto chronoids, they can be *projected on time-boundaries*, too. The result is a process boundary, and the entities found there are called *presentials*, as they are *not* extended in time. Another way to put this, is, that presentials have no temporal parts, or that they fully exist at single time-boundaries.

An example would be a bottle at a certain time t_1. It simply is a bottle. But if we take a “snapshot” of a 100m-sprint, it is no longer a 100-meter sprint.

4.2.2.3 Summary

- Processes are entities that are extended in time or unfold in time.
- Processes can be projected onto their framing chronoids.

[62] With the backup of GFO’s theory on chronoids, the meeting relation between processes is as powerful as the “meet” relation of ALLEN (1989), so all the other relations of Allen and Hayes (“BEFORE”, “STARTS”, “DURING”, etc.) can easily be defined, too.

- Processes can be projected onto time-boundaries, yielding presentials.
- Parts of processes are processes themselves.
- Process boundaries coincide if their chronoids extremal time-boundaries coincide. The processes are then called meeting.
- Presentials are – in a temporal sense – the opposite of processes, because they exist fully at a single time-boundary.

4.2.3 Universals

Having a look at presentials, a very basic relation between them is similarity. Some presentials (bottles, cells, bones) can be grouped together by certain similarities. So we believe, that this, again, is something, a top-level ontology should be prepared to deal with.

GFO uses the classic notion of *universals*, here, and introduces the *instantiation* relation between an abstract universal and the concrete *instance*. The bottles on my table are similar (as are all bottles), because they are instances of the same bottle universal. Being "abstract" means that unlike individuals (and unlike time and space entities themselves), it does not make any sense, to make temporal or spatial claims about universals. They do not exist at a certain time, or at a certain space: they are in a very fundamental way out of time and space.[63]

To avoid confusion, it should be added, that two distinct universals may have the same instances (*extension*), as in the well known example of "human" and "featherless biped".

4.2.3.1 Summary

- Universals are abstract entities that can be instantiated.
- The (concrete) instances of a universal share similarities in some respect. So universals group similar entities together.
- Unlike sets, two universals are not necessarily identical if they have the same extension.

4.2.4 Properties, Qualities, Values

If we have another look at the bottle on my desk, we find that it has certain characteristics, like a particular colour or a certain weight. And having such characteristics is surely not specific to a domain: neurons have specific shapes, newspapers have a certain layout, and singers' voices have a specific pitch. Again, this is something, a top-level ontology should contain.

[63] To stress this fundamental difference, JUBIEN (1997) calls the separation of abstract and concrete entities the "Great Line of Being".

Relying on the abstract/concrete distinction as introduced in section 4.2.3, we find that some features should be abstract, making the following claim possible: “The two bottles on my table have the same colour”. If “the same colour” was a concrete entity, it could not be in two different places at the same time. It could not be literally “the same” if there are two distinct concrete objects. This abstract feature is called *property* in GFO.

On the other hand, we may refer to the concrete colour of one of the bottles. This entity, called *quality* in GFO, is something different to the colour of the other bottle, or the colour of some car driving by.

The abstract/concrete distinction was based on the relation to time and space. And its easy to see, that properties/qualities indeed differ in this respect. Speaking about the lung of a smoker having the same colour as asphalt, it does not make sense to ask whether this abstract property is left or right, or before, or after another property. With the concrete qualities, we can: The lung’s particular grey existed only after several years of heavy smoking.

In order to give the full picture on properties et. al., we should note another distinction. The one between “the lung has a colour” and “the lung is red”, or between “this table has a particular height” and “this table’s particular size is 0.95cm”. It is the difference between properties and *property values*, or qualities and *quality values*, respectively.

Taking all these kinds of entities together, the “full”[64] picture of, say, a rose being red, involves the following entities :

1. The rose, which is a presential.
2. The abstract property colour.
3. The abstract property’s value “Redness”.
4. The concrete quality: the colour of that specific rose.
5. The concrete quality value: the particular redness of that specific rose.

These entities are related by several relations:

1. The abstract property and the concrete quality are connected by *instantiation*, the property being a universal.
2. Quality and presential are connected by *inherence*.
3. Properties (and qualities) and their values are connected by another relation called *value_of*.

[64] Depending on what you are about to model, you will not need all of these entities, of course.

4.2.4.1 Summary

- Properties are universals of certain characteristics.
- Qualities are instances of properties.
- Qualities and their bearers are connected by inherence.
- Properties have abstract property values, while qualities have concrete quality values.

5 A GFO Theory of Causality

After quite a lot of necessary background information and discussion on the main topics of causality analysis, the time has come to introduce the GFO proposal concerning causality.

It is split into three parts: first, the most fundamental relation cause (x, y) is formally introduced with discussion of its components.[65] Secondly, the basic causal relation is extended in numerous ways to cover processes as causal relata. And thirdly, cases of parallel causal relations are discussed.

5.1 The Basic Causal Relation

5.1.1 Presentials as Primary Causal Relata

If the question comes to the nature of the causal relata, the philosophical repertory is overwhelming (cf. SCHAFFER, 2003), but if the discussion is not directly focussed on the relata, they are usually assumed to be events. As LEHMANN ET AL. (2004) puts it: "[...] events have a strong causal flavor, due to their tight relationship with the notions of change and time, and this makes them appealing causal relata."

And we agree that everyday language prefers events (which we will call processes, as introduced in 4.2.2.1) as causal relata. Yet we think that serious problems may well arise, if we take everyday language to express an ontological theory rather than being a pragmatically justified abbreviation of an underlying ontology that is shared by the speaker's community. In other words: While the surface structure seems to presuppose processes to be the primary causal relata, fine-grained analysis might show that another kind of entity does play that role without changing the surface structure. And – after presenting the problems that arise, if we treat processes as being primary – we will present such an analysis in the following, starting with presentials that are connected by the basic causal relation cause (x, y), a relation which then can easily be extended to cover processes (and claims about causally connected processes) as well.

65 Parts of this section's arguments and results were first presented in MICHALEK (2005).

5.1.1.1 Problematic Processes: Causal Relevance

Imagine a billiard ball running towards another ball which rests on the cloth. The second ball is hit, and begins moving (while the first ball might change speed and angle of its movement).

Analysed in terms of processes, we would identify two of them, meeting at the two balls' collision. And now, the puzzle begins:[66] What part of the first process is relevant to the second?[67] Let's divide process P_1 in its two halves yielding $P_{1.1}$ and $P_{1.2}$ (cf. fig. 5.1). The question now becomes: What part of P_1 is relevant to the second process P_2?

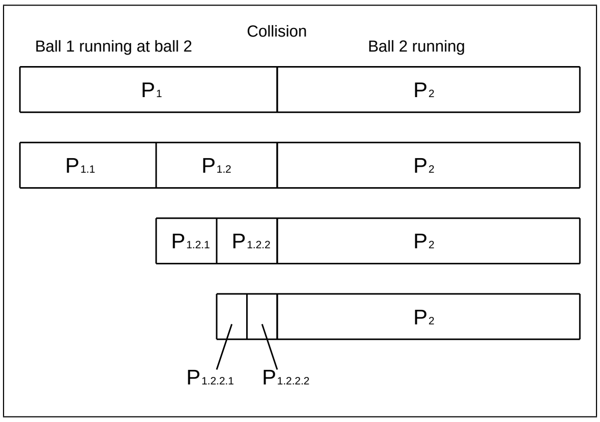


Figure 5.1: *What part of process P_1 is relevant to process P_2?*

Take the first half, i.e. $P_{1.1}$, alone: It does not contain the collision and there is a temporal gap between $P_{1.1}$ and P_2. So its status of being the one that causes P_2 is quite questionable.[68] The second half, i.e. $P_{1.2}$ that includes the collision, is definitely more promising, because given $P_{1.2}$ (alone, or even together with a different $P_{1.1}$), the result would be the same as in the unmodified situation: P_2 would be the same. Indeed, $P_{1.2}$ seems to bear *all* the causal power of the first process (with respect to the effects on the second process).

But if we disregard the first half, and concentrate on the second alone, we might raise the same question again: Which of the two halves of $P_{1.2}$ is relevant to P_2? The first half – $P_{1.2.1}$ – has the same problem as $P_{1.1}$ before, it does not contain the collision, and there is a temporal gap between $P_{1.2.1}$ and P_2. The causal power seems, again, to lie in the second half, where the same question will lead to the same answer: it always is the last part of every new last part, that

[66] The following argument is based on a thought experiment in JUBIEN (1997) which actually explicates an idea of RUSSELL (1910) (cf. footnote 68, below).

[67] Some readers might be tempted to deny that this question makes sense, as it might always be the *whole* process at stake, if it causes something. We will come to this objection at the end of the actual section (p. 57).

[68] As RUSSELL (1910, p. 184) puts it when discussing one of several definitions of causality: "earlier parts are not contiguous to the effect and therefore (by the definition [introduced before]) cannot influence the effect".

is causally relevant to the effect.

This leads us to the assumption, that it is the situation or state of affairs (both in a non-technical reading, here) *at the very end of the first process*, that is causally relevant to the second process. And looking at GFO, the very last piece is the presential at the processes' end.

Objection: No Splitting The above argument relied on splitting the process that is the alleged cause, and it might be objected, that this argument fails because it presupposes that it is a part of the process that contains the causal power. The rival thesis would then be: it is always the *whole* process, that is causally relevant, and not one of its parts; least of all a single time-slice (i.e. a presential).

We believe, that this does not work for important parts of sciences that deal with causal relationships. Think of physicists testing the predictions of a certain theory. They will proceed by creating the initial conditions, the theory is about, and then check for the expected results. However if the initial conditions contain, a certain low temperature for example, the scientists are free in choosing the way of cooling. All that matters is generating the right (presential) conditions. Without regard to the kind of process that comes up with these conditions.

The same holds for the billiard balls. It is not relevant, *how* the first ball got its speed or angle of movement. It may as well be struck by the queue, as be hit by another ball, or by some fancy automata. The effect, i.e. the movement of the second ball would be exactly the same, as long as the situation at the very moment of touching is the same.

5.1.1.2 Problematic Processes: Temporal Connection

Taking processes as primary causal relata leads to another difficulty: How should cause and effect be directly connected?[69] Typically, the temporal extension of processes is modelled by intervals of real numbers, but this is where the problem arises. Intervals may be open or closed, so we get the following combinations (cf. fig. 5.2):

- The first interval is right-closed, the second left-closed. This includes two possibilities:
  - The two intervals do not overlap. Because of the nature of the real numbers, this immediately leads to the conclusion that there is a temporal gap between the two processes. Thus the connection is not immediate.[70]

[69] It might well be that there are causal chains, where the (first) cause and the (last) effect are not immediately connected, but in this case, at least some intermediate elements have to be directly connected.

[70] The problem of a temporal gap is nicely depicted in RUSSELL (1910, p. 187): "I put my penny in the slot, but before I can draw out my ticket, there is an earthquake which upsets the machine and my calculations."

 - The two intervals overlap. This, again, would not be immediate succession as we wanted the connection to be. Furthermore, the time-structure between cause and effect would be confusing. Parts of the cause would be before the effect, some would be synchronous to the effect, and some even after (several) parts of the effect.

- One of the two intervals is closed while the other is open. The problem with these solutions is the ontological interpretation of an open interval, i.e. a process without a definite endpoint. Either the cause has no definite ending, or the effect has no definite beginning. In case of the billiard balls, this means that there is no definite time point, where the first ball stops, or changes its speed or angle.
  If we take our considerations from the last section into account: no point of an open interval is able to carry the causal relevance, as there is always some other point being closer to the point of "connection".

- Both intervals are open. Here, the same problems arise, only this time in both, cause and effect simultaneously.

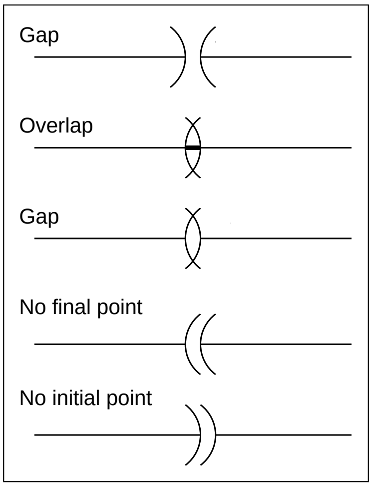


Figure 5.2: *Variations of connecting time intervals*

If we take the GFO model of time (as introduced in sect. 4.2.1), these problems do not appear. We simply have two chronoids whose time boundaries coincide, i.e. there is no temporal gap between the boundaries, yet the boundaries stay two distinct entitites. Hence we have true initial and ending points.

The final picture is as follows: processes are temporally framed by chronoids. Two chronoids are temporally immediately connected by their extremal boundaries coinciding. The projection of time-boundaries onto processes are presentials and it is precisely the presentials at the coinciding time-boundaries that are connected by causality in our theory.[71]

A1. $\text{cause}\,(x,y) \rightarrow \text{Pres}\,(x) \wedge \text{Pres}\,(y) \wedge \exists t_1, t_2(\text{at}\,(x,t_1) \wedge \text{at}\,(y,t_2) \wedge \text{coinc}\,(t_1,t_2))$

(*The causal relata are presentials at coinciding time-boundaries.*)

Looking at axiom A1, we find that the basic causal relation is *trivially transitive* (cf. C1 below) because there is no entity that can play the role of the connecting variable (b in C1) in transitivity's antecedent. Playing that role would require being both, the second participant in one causal relation and the first participant in another causal relation. This, however would imply that the connecting variable (b) is both a left time-boundary presential and a right time-boundary presential, a distinction that that GFO explicitly introduces as being exclusive.

C1. $\text{cause}\,(a,b) \wedge \text{cause}\,(b,c) \rightarrow \text{cause}\,(a,c)$

(*Transitivity, trivial*)

We call the causal relation's transitivity "trivial" to stress that although the formal condition for transitivity is fulfilled, this is the case only because there are no entities to which transitivity can be applied. In short: transitivity holds, but can not be used (or: can not be of any use) in a logical deduction.

5.1.1.3 Processes Do Still Belong to the Full Picture

The previous arguments against the use of processes as primary causal relata should not compromise the vital role processes play with regard to presentials. The latter only exist as projections of processes onto time boundaries, so there is no presential without its underlying process. Additionally, the process it depends on might even be necessary for the presential to be able to have certain properties[72]

All that is argued for, is that causality does not hold directly between processes, but only by means of their presentials which could be called an *indirect* causal connection. Section 5.2 deals with causal relations within and between processes.

5.1.2 Regularity

Now that we have the relata, their relation is the next crucial point. Following our considerations in the introduction, we will begin by covering the idea of regularity.

[71] In the following formulae, axioms, definitions and corollaries (understood as immediate consequences following from axioms and definitions) are marked with "A", "D", and "C", respectively.

[72] Like, e.g. an object having a certain velocity. Having a velocity is only possible if the object (here: the presential) takes part in a process.

Recall that regularity calls for a number of *similar* causes (and effects) to be connected in a certain way. Similarity, however, is why universals had been introduced in GFO, so we may use them here, too: the causes (and the effects) must be grouped by being instances of universals.

Additionally, a *statistical dependency* must hold between (the existence of) instances of the cause and (the existence of) instances of the effect universals: The existence of the former must heighten the probability of the existence of the latter.

5.1.2.1 Coincidence Pairs

To formalise regularity, we shall begin by introducing *coincidence pairs* of presentials, i.e. presentials that exist at coinciding time boundaries:

D1. $\text{coincPair}\,(x, y) =_{df} \text{Pres}\,(x) \wedge \text{Pres}\,(y) \wedge \exists t_1, t_2(\text{at}\,(x, t_1) \wedge \text{at}\,(y, t_2) \wedge \text{coinc}\,(t_1, t_2))$

(*Coincidence pair: presentials at coinciding time boundaries*)

The collection of all the coincidence pairs gives the uncountable universe (sample space) Ω_{cp_r} of the subsequent probability considerations[73]

D2. $\Omega_{cp_r} =_{df} \{(x, y) \mid \text{coincPair}\,(x, y)\}$

(*Universe of coincidence pairs*)

5.1.2.2 Probabilistics

Following the standard textbook account on probability in uncountable sample spaces (cf. Chung and AitSalhlia, 2003), we introduce a non-empty set S_r as a σ-algebra over subsets of Ω_{cp_r}:

A2. $A \in S_r \rightarrow \bar{A} \in S_r$

(*Closed under complements*)

A3. $A_1, A_2, \ldots \in S_r \rightarrow \bigcup_i A_i \in S_r$

(*Closed under countable unions*)

This includes that S_r contains the empty set and is (via DeMorgan's Law) also closed under countable intersections.

Next, we introduce a function P_r assigning real numbers (probabilities) to members of S_r, with P_r fulfilling the Kolmogorov conditions:

[73] The index r indicates that we deal with regularity, here.

A4. $P_r(A) \geq 0 \quad \text{for} \quad A \in S_r$

(*No negative values*)

A5. $P_r(\Omega_{cp_r}) = 1$

(*Total measure one*)

A6. $P_r(\bigcup_i A_i) = \sum_i P_r(A_i) \quad \text{for} \quad A_i \in S_r$ with A_i pairwise disjoint, i countable

(σ*-additivity*)

The triple $(\Omega_{cp_r}, S_r, P_r)$ is a *probability space* where we can use the common expression of *statistical dependence* between two events[74] C and D, i.e. "Probability of C, given that D" denoted by "$P_r(C \mid D)$". But what do C and D refer to, ontologically, if regularity is concerned?

5.1.2.3 Probabilistic Regularity

With the abbreviations

D3. coincPairUf $(x, y, U) =_{df}$ coincPair $(x, y) \wedge x :: U$

(*Coincidence pair with universal's instance in first participant*)

D4. coincPairUs $(x, y, U) =_{df}$ coincPair $(x, y) \wedge y :: U$

(*Coincidence pair with universal's instance in second participant*)

and the corresponding subsets of Ω_{cp_r}

D5. $E_f\,(U) =_{df} \{(x, y) \mid$ coincPairUf $(x, y, U)\}$

(*Coincidence pairs with an instance of U as the first participant*)

D6. $E_s\,(U) =_{df} \{(x, y) \mid$ coincPairUs $(x, y, U)\}$

(*Coincidence pairs with an instance of U as the second participant*)

we can formally express a statistical dependency between instances of universals:

D7. statDepU $(U_1, U_2) =_{df} P(E_s\,(U_2)) < P(E_s\,(U_2) \mid E_f\,(U_1))$

(*Statistical dependence between universals, mediated by instances*)

[74] The term "event" is used here as common in probability theory. No ontological connotation intended.

Definition D7 now captures the idea that the probability to "find" an instance of the effect universal is higher, if there already is an instance of the cause universal.[75] Finally, we are able to summarise the regularity condition on causality:

D8. statDepP $(x, y, U_c, U_e) =_{df} x :: U_c \wedge y :: U_e \wedge$ statDepU (U_c, U_e)

(*Statistical dependency between presentials w.r.t. universals* U_e, U_c)

A7. cause $(x, y) \rightarrow \exists U_c, U_e($statDepP $(x, y, U_c, U_e))$

(*Regularity Axiom*)

5.1.2.4 Objection: Is Immanence at Stake?

Causal theories may be divided by their answer to the following questions: "Are the causal relata immanent, or transcendent? That is, are they concrete and located in spacetime, or abstract and non-spatiotemporal?" (SCHAFFER, 2003, sect. 1.1).

Our theory makes use of universals which clearly belong to the realm of the abstract. So let us take some time to check whether this might turn out to be a flaw. The argument we will briefly go into is that abstract entities cannot interact, or as an opponent to immanence puts it (when defending abstract "facts" as causal relata): "Some people have objected that facts are not the sort of item that can cause anything. A fact is a true proposition (they say); it is not something in the world but is rather something *about* the world, which makes it categorically wrong for the role of a puller and shover and twister and bender." (BENNETT, 1988, p. 22, his emphasis) Translated to our approach the question is: as universals are not part of the (spatiotemporal) world, are they categorically unsuitable for the role of pullers and shovers and twisters and benders?

In order to answer this question, let us reconsider what role universals play in our theory. To begin with, they are not the basic causal relata. The basic causal relation is defined on coincidence pairs of presentials. And as presentials are a perfectly immanent kind of entities, there should be no disagreement that they indeed can interact, or - in BENNETT'S words – "behave like elbows in the ribs" (BENNETT, 1988, p. 22). Universals come into play to group

[75] "Higher": compared to the probability of finding an instance of the effect universal in some arbitrary coincidence pair of the full universe Ω_{cp_r}.
Note that this is not the same as the probability being higher compared to the probability of finding an instance of the effect universal in an coincidence pair where the cause-universal is absent like for example HITCHCOCK understands probability raising: "[...] A causes B if and only if $P(B|A) > P(B|\bar{A})$." (HITCHCOCK, 2002). Adjusted to his formulae, our approach would demand that $P(B|A) > P(B)$. While the latter expression entails HITCHCOCK'S, they are not equivalent, cf. appendix A.

similar presentials for comparison, as comparability (or: similarity) is a necessary constituent of the concept of regularity. But even if we defined a relation of statistical dependency between universals (cf. D7), this relation is defined upon the universals' instances. So in the end, we find our causal relation being grounded in the realm of the concrete, spatio-temporally located, immanent world.

5.1.2.5 Objection: May Causes Lower the Probability of Their Effects?

In the discussion on regularity we announced that we would check our theory against the claim that there are situations in which the cause may lower its effect's probability. So how would such an argument go? Here is an example:

> Pam throws a brick through the window. Meanwhile, Bob (a more reliable vandal), holds his throw on seeing Pam in action, though had Pam not thrown Bob would have. [...] Pam's throw is obviously a cause of the window shattering. But her throw is a probability lowerer of the shattering: since Bob is a more reliable vandal, the window's chances would have been worse with Bob in action. Thus probability raising is not necessary for causation.
>
> (SCHAFFER, 2001, p. 79)

Let us make it clear, where this analysis differs from what we said up to here. SCHAFFER compares the actual situation to situations where not Pam, but Bob has thrown his stone. So, say, when Pam throws the brick, it hits its target in 10% of the cases, while Bobs has a success rate of 90%. As Pam's throw prevented Bob from throwing his brick, the window's chances of not breaking were lowered from 90% to 10%.

But this is not the only way to compare probabilities, here. Another reasonable way would be comparing Pam's throw with Bob's not throwing to other situations *where Bob does not throw*. There we still have the probability of the window's being destroyed by Pam's stone (10%) which is definitely higher than the window breaking on it's own. Given this comparison, our initial claim about causes raising their effects probabilities still holds.

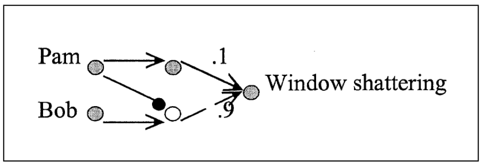


Figure 5.3: *Does the cause lower the effect's probability from 90% to 10%? (Taken from* SCHAFFER *(2001, p. 79).)*

Let's take SCHAFFER'S diagram to illustrate the difference (cf. fig. 5.3). Advancing in time, we find Pam's throw first to prevent Bob's throw, then shattering the window. So the actual situation may be compared to either the situation before Bob's throw is prevented, or to the situation after the prevention. SCHAFFER chooses the first option: if Pam had not thrown, Bob would have. When taking the situation after Bob's throw is prevented (i.e. Bob does not throw), Pam's throw becomes a probability raiser again and our approach is still valid. And when carefully reviewing our approach, we find that there is good reason to *not* follow SCHAFFER'S analysis: our basic causal relation is defined on coincidence pairs of *presentials*. They are not extended in time. If the window's shattering is addressed as an effect of the stone, the causal setting is about the stone right when touching the window pane. And this clearly is after Bob's throw was prevented, so in the end, we might state that our theory is not affected by examples of this kind. [76]

5.1.3 Counterfactual Dependency

As presented in section 2.2, counterfactual dependency has the following conceptual constituents:

- Alternative situations and the relation of *similarity* between them
- *Clusters of similar situations* needed for the probabilistic aspects. The cluster around the initial situation, e.g. gives the initial probability of the effect.
- *Causally similar* and *causally contrastive* alternative situations – causally similar ones contain the cause, contrastive ones don't.
- *Supportive* and *undermining* causally contrastive alternative situations
  - Non-probabilistic: undermining situations contain the effect (although the cause is missing), supportive ones don't.
  - Probabilistic: undermining clusters of situations are those, where (the cause is absent and) the probability of the effect is lower or equal to the initial probability. In supportive clusters, it is higher. (This covers the non-probabilistic variant as a special case.)
- A notion of *distance* with respect to a reference cluster that allows for comparing clusters.

So far we loosely spoke of "situations", but as our considerations on the causal relata (cf. sect. 5.1.1) have shown, it is presentials that are causally relevant in the world of concrete entities.

[76] Secondly, the argument relies on the causal relation being transitive (Pam's throw causing Bob not to throw which causes the window not being shattered by Bob's brick). We, however, do believe that although there is some kind of transitivity in certain circumstances, the basic causal relation cannot be used transitively. (In more detail: it is trivially transitive, but there cannot be any b such that cause $(a, b) \wedge$ cause (b, c), cf. C1 on p.59).

So we need to adjust the concepts above. And we do so by replacing the "situations" by two presentials that build a coincidence pair. So what we need now becomes:

- Coincidence pairs of presentials, and a notion of similarity between presentials.
- Clusters of similar coincidence pairs, initial probability
- Causally similar and causally contrastive clusters of coincidence pairs – causally similar are those, whose first participants are similar to each other and contain an instance of the cause universal, while contrastive coincidence pairs are similar to each other but do not contain an instance of the cause universal in the first participant.
- Supportive and undermining clusters or causally contrastive coincidence pairs
  - Non-probabilistic: undermining coincidence pairs' second participants are instances of the effect universal (although the cause is absent in the first participant of this pair), in supportive coincidence pairs, there is no instance of the effect universal in the second participant (i.e. the effect is missing as the cause is missing).
  - Probabilistic: in undermining clusters of coincidence pairs, the probability of the effect (as instance of the effect universal in the second participant) is higher or equal to the initial probability (although the cause is absent in this pair). In supportive pairs, it is lower (so an absent cause lowers the effect's chance).
  - A (distance) relation that orders clusters of coincidence pairs with respect to a reference cluster.

5.1.3.1 Similarity and Contrast

Similarity and *contrast* between presentials is modelled by instances of universals[77]. The universal might be made explicit or not:

D9. similarPres $(x, y) =_{df} Pres(x) \wedge Pres(y) \wedge \exists U(Univ(U) \wedge x :: U \wedge y :: U)$

(*Ordinary similarity between presentials*)

D10. similarPresU $(x, y, U) =_{df} Pres(x) \wedge Pres(y) \wedge x :: U \wedge y :: U$

(*Similar presentials w.r.t. a universal*)

D11. contrastPresU $(x, y, U) =_{df} Pres(x) \wedge Pres(y) \wedge \neg(x :: U) \wedge y :: U$

(*Contrastive presentials w.r.t. a universal*)

[77] We will use the GFO symbol :: to express instantiation.
So "a is an instance of U" will be formalized as $a :: U$ (cf. HERRE ET AL., 2007, p. 53).

This implies the following:

C2. $Pres(x) \wedge x :: U \rightarrow$ similarPresU (x, x, U)

(*Reflexive "self-similarity" of a presential*)

C3. similarPresU $(x, y, U) \leftrightarrow$ similarPresU (y, x, U)

(*Symmetry on fixed universal*)

C4. similarPresU $(x, y, U) \wedge$ similarPresU $(y, z, U) \rightarrow$ similarPresU (x, z, U)

(*Transitivity on fixed universal*)

C5. $\neg$contrastPresU (x, x, U)

(*Irreflexivity*)

C6. contrastPresU $(x, y, U) \leftrightarrow \neg$contrastPresU (y, x, U)

(*Asymmetry on fixed universal*)

5.1.3.2 Clusters of Presentials and of Coincidence Pairs

The relation of similarity with respect to a certain universal is the basis for defining clusters of presentials on which clusters of coincidence pairs (what we hitherto called "similar situations") do rely:

D12. SimilarPresU $(U_s) =_{df} \{x \mid x :: U_s\}$

(U_s*-cluster of similar presentials centered around a certain universal*)

D13. SimilarCpU $(U_s) =_{df} \{(x, y) \mid$ coincPair $(x, y) \wedge x :: U_s\}$

(*Similar coincidence pairs whose first participants are* U_s*-clustered*)

D14. ContrastCpU $(U_s) =_{df} \{(x, y) \mid$ coincPair $(x, y) \wedge \neg(x :: U_s)\}$

(*Contrastive coincidence pairs whose first participants are outside the* U_s*-cluster*)

Within the clusters of coincidence pairs there are those that have (or explicitly do not have) an instance of a certain universal as the first participant . These clusters – we labeled them "causally similar / causally contrastive" – build the basis for counterfactual dependency and all its probabilistic aspects[78]:

D15. CsimilarCpU $(U_s, U_c) =_{df} \{(x, y) \mid x :: U_s \wedge x :: U_c\}$

(U_s-cluster of causally similar coincidence pairs, i.e. with first participant being an instance of (the cause) universal U_c)

D16. CcontrastCpU $(U_s, U_c) =_{df} \{(x, y) \mid x :: U_s \wedge \neg(x :: U_c)\}$

(U_s-cluster of causally contrastive coincidence pairs, i.e. the first participant is not an instance of (the cause) universal U_c)

D17. $\Omega_i =_{df}$ CsimilarCpU (i)
with $i \in \{U_s, U_c \mid$ Univ $(U_s) \wedge$ Univ $(U_c)\}$

(For abbreviation purposes; Universe of (causally similar) coincidence pairs restricted to U_s and U_c)

D18. $\Omega_{\bar{i}} =_{df}$ CcontrastCpU $(\bar{i})$
with $i \in \{U_s, U_c \mid$ Univ $(U_s) \wedge$ Univ $(U_c)\}$

(For abbreviation purposes; Universe of (causally contrastive) coincidence pairs restricted to U_s and contrastive w.r.t. U_c)

5.1.3.3 Probabilistics

Following the strategy we used for covering probabilistic regularity (cf. sect. 5.1.2) we define non-empty sets S_i as σ-algebrae over subsets of clusters of coincidence pairs Ω_i:

A8. $A \in S_i \rightarrow \bar{A} \in S_i$

(Closed under complements)

A9. $A_1, A_2, \ldots \in S_i \rightarrow \bigcup_j A_j \in S_i$

(Closed under countable unions)

78 You will note that the set *SimilarCpU* from definition D12 in the subsequent formulae, is the same as E_f, defined before in D5. This probably calls for an explanation: in the section on regularity, the focus was on whether the first or the second participant of a coincidence pair was an instance of some universal; therefore we introduced E_f and E_s. In the present section, however, the focus lies on similarity and contrast (*SimilarCpU* and *ContrastCpU*), so we took the freedom to define an equivalent expression in order to symbolically (i.e. with respect to the sets' "names") support the line of logically constructing our account of counterfactual dependency on similariy.

Each of the S_i contains the empty set and is closed under countable intersections.

The functions P_i, then, assign real numbers (probabilities) to members of S_i, with P_i fulfilling the KOLMOGOROV conditions:

A10. $P_i(A) \geq 0 \quad \text{for} \quad A \in S_i$

(*No negative values*)

A11. $P_i(\Omega_i) = 1$

(*Total measure one*)

A12. $P_i(\sum_j A_j) = \sum_j P_i(A_j) \quad \text{for} \quad A_j \in S_i$ with A_j pairwise disjoint, j countable

(σ*-additivity*)

The triples (Ω_i, S_i, P_i) are *probability spaces* (defined over causally similar clusters) we will use for expressing the *initial probability* as well as *supportive* and *undermining* clusters of coincidence pairs.

Note that the same apparatus can be applied to the $\Omega_{\bar{i}}$ (i.e. causally contrastive clusters of coincidence pairs that explicitly lack an instance of the cause universal) yielding the probability spaces $(\Omega_{\bar{i}}, S_{\bar{i}}, P_{\bar{i}})$, respectively.

Let us now introduce two more abbreviations to express the probability of the effect – i.e. of "finding" an instance of the effect universal – in the second participant of coincidence pairs that belong to either causally similar or causally contrastive clusters:

D19. $P_i(U_e) =_{df} P_i(\{(x,y) \mid \text{CsimilarCpU}\,(i) \wedge y :: U_e\})$
with $i \in \{U_s, U_c \mid \text{Univ}\,(U_s) \wedge \text{Univ}\,(U_c)\}$

(*Probability of effect e in causally similar cluster U_s*)

D20. $P_{\bar{i}}(U_e) =_{df} P_{\bar{i}}(\{(x,y) \mid \text{CcontrastCpU}\,(\bar{i}) \wedge \neg(y :: U_e)\})$
with $i \in \{U_s, U_c \mid \text{Univ}\,(U_s) \wedge \text{Univ}\,(U_c)\}$

(*Probability of effect e in causally contrastive cluster U_s*)

5.1.3.4 Supportive and Undermining Clusters

Compared to a reference cluster where the cause is present, a second cluster is called *supportive*[79], if it does not contain the cause, and the probability of the effect is lower than in the

[79] I.e it supports the causal claim one might make if only observing the reference cluster.

reference cluster. If the probability of the effect is higher or equally high (even though the cause is missing), it is called *undermining*:

D21. supportiveCluster $(U_s, U_r, U_c, U_e) =_{df} P_{U_r,U_c}(U_e) > P_{\overline{U_s,U_c}}(U_e)$

(*U-cluster supportive w.r.t. U_r-cluster and cause/effect universals*)

D22. underminingCluster $(U_u, U_r, U_c, U_e) =_{df} P_{U_r,U_c}(U_e) \leq P_{\overline{U_u,U_c}}(U_e)$

(*U-cluster undermining w.r.t. U_r-cluster and cause/effect universals*)

5.1.3.5 Distance Between Clusters

The next element needed is *distance* between causally contrastive clusters with respect to a reference cluster.

LEWIS gives some rough ideas of what the distance between possible worlds can rely on, like "similarities in matters of particular fact trade off against similarities of law" (LEWIS, 1973, p. 560). However, he accepts that the vagueness of what he calls "comparative overall similarity" (LEWIS, 1973, p. 559) (which take several of these "rules" and assigns different weights to each of them) cannot be overcome as it simply is part of causality (cf. LEWIS, 1973, 560).

We follow his analysis and accept this limitation of our theory.[80] The distance thus is introduced as a primitive relation between three clusters of coincidence pairs that means "cluster U_1 is closer to cluster U_r (the reference cluster) than U_2 is to U_r":

A13. closerToThan $(U_1, U_r, U_2) \rightarrow$ Univ $(U_1) \wedge$ Univ $(U_r) \wedge$ Univ (U_2)

(*Based on universals around which the clusters are centered*)

A14. closerToThan $(U_1, U_r, U_2) \wedge$ closerToThan $(U_2, U_r, U_3) \rightarrow$ closerToThan (U_1, U_r, U_3)

(*Transitivity w.r.t. reference universal*)

A15. closerToThan $(U_1, U_r, U_2) \rightarrow \neg$closerToThan (U_2, U_r, U_1)

(*Asymmetry*)

A16. $\neg$closerToThan (U_1, U_r, U_1)

(*Irreflexivity*)

5.1.3.6 Probabilistic Counterfactual Dependency

Counterfactual dependency now holds if there is a supportive cluster that is closer to a reference cluster (representing the actual situation in which both, cause and effect took place) than every

[80] It should be noted, though, that *pragmatically*, science indeed has developed ways to distinguish between sensible and far fetched alternatives when performing experiments.

underminig one:

D23. $\text{counterfactDep}\,(U_c, U_e, U_r) =_{df} \exists U(U \neq U_r \wedge \text{supportiveCluster}\,(U, U_r, U_c, U_e)$
$\wedge \forall U_u(\text{underminingCluster}\,(U_u, U_r, U_c, U_e) \rightarrow \text{closerToThan}\,(U, U_r, U_u)))$

(*Counterfactual dependency with respect to a reference universal*)

On the level of presentials, counterfactual dependency demands that the presentials are instances of universals that are connected by counterfactual dependency as defined above:

D24. $\text{counterfactDepP}\,(c, e, U_c, U_e, U_r) =_{df} \text{coincPair}\,(c, e) \wedge c :: U_c \wedge e :: U_e$
$\wedge\ \text{counterfactDep}\,(U_c, U_e, U_r)$

(*Counterfactual dependency between presentials, mediated by cause, effect and reference universal*)

This leads to the final formulation of the counterfactual condition on causality:

A17. $\text{cause}\,(c, e) \rightarrow \text{coincPair}\,(c, e) \wedge \exists U_c, U_e, U_r(\text{counterfactDepP}\,(c, e, U_c, U_e, U_r))$

(*Axiom of counterfactual dependency*)

5.1.3.7 Sufficient Conditions

Axioms A8 and A17 presented the two necessary conditions of regularity and counterfactual dependency that causality relies on. In our view, the conjunction of these conditions (based on the same cause and effect universals) are in turn sufficient for causality:

D25. $\text{cause}_{expl}(c, e, U_c, U_e, U_r) =_{df} \text{coincPair}\,(c, e) \wedge$
$c :: U_c \wedge e :: U_e \wedge \text{statDepU}\,(U_c, U_e) \wedge \text{counterfactDep}\,(U_c, U_e, U_r)$

(*Causal relation with cause and effect universals made explicit*)

A18. $\text{cause}_{expl}(c, e, U_c, U_e, U_r) \rightarrow \text{cause}\,(c, e)$

(*Explicit causal relation implies basic causal relation*)

5.1.4 Manipulability Recreated

As explained in sect. 2.3.4, we believe that the manipulability intuition (in short: if there is a causal relation, the effect must be modifiable by manipulating the cause) is covered by regularity and counterfactual dependency as introduced above – and thus does not need to be introduced separately. Now that we have our theory at hand, we shall show how it covers what we figured out to be the manipulationist's main points:

- An effect's properties may be changed by changing the cause's properties.
- Effects may be changed by the non-existence of the cause.

The first point can be made easily: the basic causal relation holds between presentials, and this is exactly, what an object's properties *are*. The color of a rose, or the weight of a stone are not extended in time: at any point in time, we can say what color the rose has, or what weight a stone has. This means that causal relations between properties are perfectly covered by our theory.

Now it may well be that some properties require a process taking place like a bullet's velocity or a billiard ball's momentum. Moreover, if we had a look at some flying bullet or a moving billiard ball at a single time boundary, it will not move, apparently so it might be tempting to say that there are properties which are not presentials. However, we know that this bullet or billiard ball is (ontologically) different to non-moving bullets or billiard balls. Even if this difference is not visible at the single time boundary. We can still model it by a presential.[81]

Now, having a look at how GFO models properties and their values (cf. sect. 4.2.4), we find that the change of a property's value is modeled by one value being removed by an (ontologically) different value. So changing properties already includes one entity coming into existence, while the other is no longer there. Additionally, if whole objects should disappear, this will immediately change the clusters of alternative situations that both, regularity and counterfactual dependency depend on. In the end we can conclude that the relevance of both, changing a cause's properties and removing a cause completely, indeed is contained in our theory.

5.2 Extending the Basic Relation: Processes

Starting with the basic relation *cause* (x, y) relating presentials as defined above, we can now go on to extend it to cover processes as causal relata as well. This does not add to the concept of causality as introduced up to here (it still is all about regularity and counterfactual dependency between presentials), but simply provides means to connect the basic causal relation to a wider range of GFO ontological categories.

5.2.1 Processes and Presentials

Recall the two projection relations of GFO. The first, prt (P, C), connects a process P and a chronoid C (which is roughly the time-interval, the process takes place in; cf. fig. 5.4). The second, prt (P, t, p), projects a chronoid's time-boundary t on the process yielding a presential p (cf. fig. 5.5). If C frames P and if the time-boundaries are exactly the extremal left and right

[81] Additionally: manipulability does not rely on *every* property of the effect being manipulable by alteration of the cause.

time-boundary of C, we can call the corresponding presentials *presential at the left/right end of process* P (cf. fig. 5.6).[82]

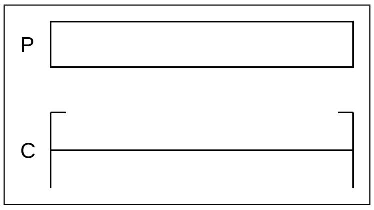


Figure 5.4: *Process P projected onto chronoid C with its left and right boundary.*

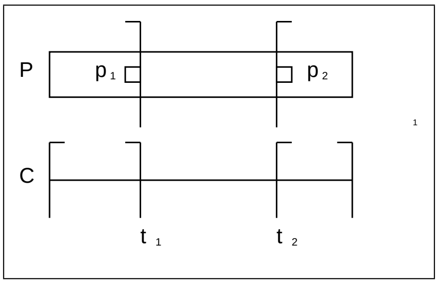


Figure 5.5: *Process P projected onto a right and a left (inner) time-boundary t_1 and t_2, respectively, yielding presentials p_1 and p_2.*

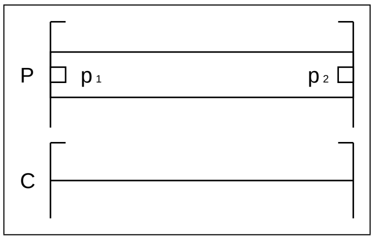


Figure 5.6: *Process P with PaLp (p_1, P) and PaRp (p_2, P).*

D26. PaLp $(p, P) =_{df} \exists C, t(\text{Proc}\,(P) \wedge \text{Chron}\,(C) \wedge \text{prt}\,(P, C) \wedge \text{lb}\,(t, C) \wedge \text{prt}\,(P, t, p))$

(*Presential at left end of process*)

D27. PaRp $(p, P) =_{df} \exists C, t(\text{Proc}\,(P) \wedge \text{Chron}\,(C) \wedge \text{prt}\,(P, C) \wedge \text{rb}\,(t, C) \wedge \text{prt}\,(P, t, p))$

(*Presential at right end of process*)

We will now explore different possible extensions of the basic causal relation, starting with merely technical definitions (i.e. they are of rather marginal modeling use) which will lead to the

[82] An overview of the symbols used in the following diagrams can be found in appendix B (p. III).

notion of causal adhesion which – in our belief – is *the* fundamental relation from a modeling point of view.

5.2.2 Dual-Boundary Causality

5.2.2.1 Heterogeneous Causality

Combining the basic causal relation with PaRp and PaLp allows us to introduce a notion of *heterogeneous causality* between a presential and a process. The main idea being that the basic causal relation holds between the presential and the PaRp /PaLp of the process in question (cf. fig. 5.7 and 5.8).

D28. $\text{cause}_{hetPres}(p_1, P) =_{df} \exists p_2(\text{PaLp}\,(p_2, P) \wedge \text{cause}\,(p_1, p_2))$

(*Heterogeneous causation between presential and process*)

D29. $\text{cause}_{hetProc}(P, p_2) =_{df} \exists p_1(\text{PaRp}\,(p_1, P) \wedge \text{cause}\,(p_1, p_2))$

(*Heterogeneous causation between process and presential*)

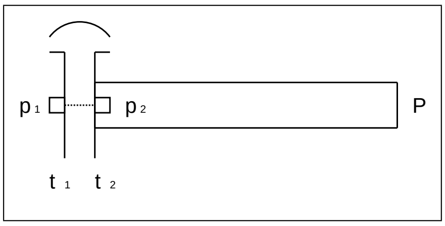


Figure 5.7: *Heterogeneous causation connecting presential p_1 and process P.*

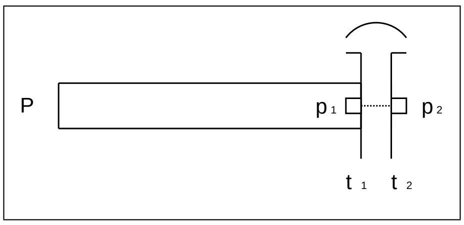


Figure 5.8: *Heterogeneous causation connecting process P and presential p_2.*

5.2.2.2 Sequential Causality

Starting with heterogeneous causation, it is not difficult to *sequentially* connect two processes causally. We use the same mechanism as in the heterogeneous cases above, but "in both directions", using PaLp and PaRp , respectively (cf. fig. 5.9).

D30. cause $_{seqProc}(P, Q) =_{df} \exists p_1, p_2($PaRp $(p_1, P) \wedge$ PaLp (p_2, Q)
$\wedge$cause $(p_1, p_2)))$

(*Sequential process causation*)

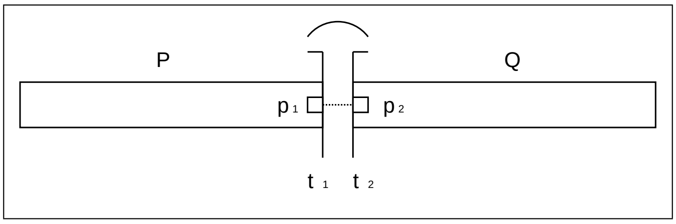


Figure 5.9: *Sequential causality between processes P and Q via the causal relation between PaRp (p_1, P) and PaLp (p_2, Q), that exist on coinciding time-boundaries t_1 and t_2.*

5.2.3 Multi-Boundary/Continuous Causality

The characteristic feature of the following causal relations is, that there are no longer only two time-boundaries involved (and thus not only two presentials) but infinitely many. This is due to the fact that a chronoid in GFO has infinitely many inner boundaries, which are the boundaries of sub-chronoids (i.e. proper temporal parts of a chronoid).

5.2.3.1 Causal Cohesion

The main aim of *causal cohesion* – besides introducing the core idea of multi-boundary continuous causality – is to cover the difference between processes that have an internal causal structure, while others lack it. This difference can be modeled in terms of our approach by stating that the following holds within the process: *every pair of presentials at coinciding (inner) time-boundaries is connected by the basic causal relation* (cf. fig. 5.10 and 5.11).

D31. cause $_{coh}(P) =_{df} \exists C($Proc $(P) \wedge$ Chron $(C) \wedge$ prt $(P, C) \wedge$
$\forall t_1, t_2(($innertb $(t_1, C) \wedge$ innertb $(t_2, C) \wedge$ coinc $(t_1, t_2))$
$\rightarrow \exists p_1, p_2($prt $(P, t_1, p_1) \wedge$ prt $(P, t_2, p_2) \wedge$ cause $(p_1, p_2))))$

(*Causally cohesive process*)

Examples for causally coherent processes are the rotation of all the cogs and springs in a mechanical watch, and the movement of the planets in the solar system. The following description, on the other hand, refers to a process that clearly lacks causal cohesion:[83]

[83] If you are interested in a broader discussion of the so called problem of "causal processes", i.e. whether processes that are not causally coherent should be called processes at all, DOWE (2004) gives an overview.

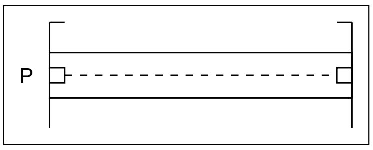


Figure 5.10: *Causal cohesion within (the full length of) process P. (Detailed view on what happens within P is given in fig. 5.11.)*

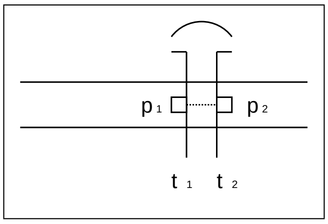


Figure 5.11: *Causal cohesion in detail: causally connected presentials p_1 and p_2 at every inner pair of coinciding boundaries t_1 and t_2 within P.*

> "The Moon's umbral shadow first touched down on Earth at 0836 GMT (0936 BST), at sunrise on the east coast of Brazil. *It then raced across the Atlantic Ocean* before making African landfall in Ghana at 0908 GMT (1008 BST), where residents of the capital Accra filled the streets to view the event."
>
> (BBC NEWS WEBSITE, 2006, emphasis added)

The "racing" of the moon's shadow may indeed be described as a process, but the shadows we get at this process's time-boundaries do not stand in cause–effect relations to each other as causal cohesion would require.

5.2.3.2 Causal Adhesion

While causal cohesion addressed a single process, *causal adhesion* is our expression for temporally overlapping processes, that are causally connected throughout this overlap (cf. fig. 5.12):

D32. $\text{cause}_{adh}(P, Q) =_{df} \text{Proc}(P) \wedge \text{Proc}(Q) \wedge$
$\exists C(\text{Chron}(C) \wedge \text{prt}(P, C) \wedge \text{prt}(Q, C) \wedge$
$\forall t_1, t_2((\text{innertb}(t_1, C) \wedge \text{innertb}(t_2, C) \wedge \text{coinc}(t_1, t_2))$
$\rightarrow \exists p_1, p_2(\text{prt}(P, t_1, p_1) \wedge \text{prt}(Q, t_2, p_2) \wedge \text{cause}(p_1, p_2))))$

(*Processes continuously and entirely connected by causal adhesion*)

The concept of causal adhesion may come in various special ways, some of which are presented in the following.

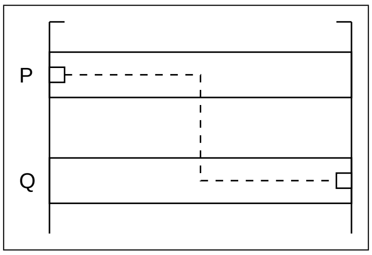


Figure 5.12: *Processes P and Q, continuously and entirely connected by causal adhesion. (Detailed view on what happens within is given in fig. 5.13.)*

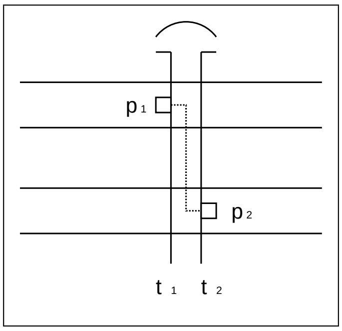


Figure 5.13: *Causal adhesion in detail: Causally connected presentials p_1 and p_2 at every pair of coinciding time-boundaries t_1 and t_2. p_1 and p_2 belong to processes P and Q, respectively.*

5.2.3.3 Adhesive Overlap

The probably most relevant causal relation between processes is that of *adhesive overlap*, which means that two processes overlap in time and are connected by causal adhesion throughout the overlap (cf. fig. 5.14).

D33. cause$_{ov}(P,Q) =_{df} \exists P_1, Q_1($procpart$(P_1,P) \wedge$ procpart$(Q_1,Q) \wedge$ cause$_{coh}(P_1,Q_1))$

(*Causally adhesive overlap*)

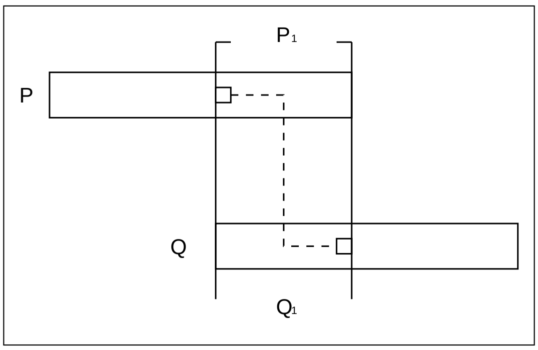


Figure 5.14: *Processes P and Q with overlapping parts P_1 and Q_1 that are connected by causal adhesion.*

Example: Central Elastic Collision (Billiard Balls' Pulse Transmission) This is the place for a more realistic model of the billard balls' collision, whose slimmed down brother (where the momentum is exchanged instantaneously) helped us in the beginning.

What would the "full" picture look like, now? The semi-natural language description (enriched by basic physics) is:

- Ball 1 is moving towards ball 2. Ball 2 rests on the cloth.

- The balls. Ball 1 comes to a full stop, while ball two is accelerated to it's final velocity.

- Ball 1 rests on the cloth (assuming a central collision with balls of equal size and weight), while ball 2 moves with constant velocity (no friction, here).

This, very naturally, calls for two processes that represent the movement of the first ball (including the deceleration) and the movement of the second ball (including the acceleration). As deceleration and acceleration take place over the same period of time, the two processes overlap. But does the overlap fulfil the condition that is posed on adhesively overlapping processes?

The condition (as given in D33) is: At every pair of coinciding boundaries during the overlap, there must be a presential at the left boundary of this pair (belonging to the "first" process) and there must be a presential (belonging to the "second" process) at the right boundary. And these presentials must be connected by the basic causal relation. And indeed, every pair consists of presential billiard balls with their respective momentum, and these pairs fulfil the conditions for causal connection: the momenta are connected by regularity (i.e. the law of conservation of momentum) and the second ball's movement would look very different, if the first ball would not hit it.

Example: Periodical Stimulation (Pushing the Swing) Another example where (causally) adhesively connected parts of processes play a role is when, say, a child is siting on a swing, and her friend helps her to swing really high by pushing her forward whenever the first girl reaches the lowest point of the swing's movement. With respect to causality, we have two processes again: the girl's swinging, and the movements of her friend. And at regular intervals, those two are connected by causal adhesion as depicted in fig. 5.15.

You may feel uneasy about this model, as obviously, the "real" causal relationship is not one-way as depicted here, but goes into both directions. From pusher to swing, and from swing to pusher. It is, in short, an interaction. And we shall deal with interactions in the following.

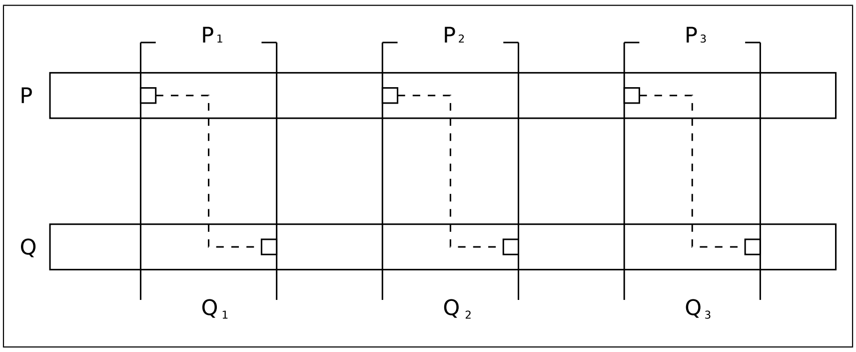


Figure 5.15: *Processes periodically overlapping causally cohesively.*

5.2.4 Reciprocity (Quartet of Interaction)

Physics tells us that forces always come in pairs[84], and we certainly do find this reciprocity in many causal relations. Again, consider the stone that is thrown at the glass pane. Not only does the stone shatter the window, but the window influences the stone's flight (by slowing it down, for example). In modeling this situation, we may want to explicitly express the reciprocity without losing the difference between causes and effects (an entity must not be both cause and effect). How may this be modeled in our framework?

The interesting part of the whole story obviously is the time between the stone touching the pane and the stone leaving the (destroyed) window. It's not difficult to see that the connection between stone and pane can be modeled by adhesive overlap (just like the billiard balls in sect. 5.2.3.3). But how is the "reaction", i.e. the window influencing the stone's flight, to be modeled?

Again, we use the notion of adhesive overlap: There are two processes (window, stone), and for every pair of coincident time-boundaries, we find

- a (window) presential at the right time-boundary,
- a (stone) presential at the left time-boundary, and
- a relation between those presentials, that fulfils the conditions of the basic causal relation.

This makes the window a cause for the stone's way of movement (deceleration, change of direction, etc.). So window–stone is a case of causal overlap, too.

Additionally, we have two cases of causal cohesion: the movement of the stone is causally related to "itself" while the window (e.g. its structure) is causally related to how it behaves after the ball touches it. So the final picture is that of an *interaction quartet*, as shown in figures 5.16 (schematic diagram) and 5.17.

---

[84] Newton's third law of motion: "Actioni contrariam semper & aequalem esse reactionem [...]." (NEWTON, 1686, p. 13) typically translated as: "To every action there is an equal and opposite reaction."

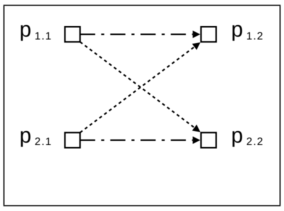


Figure 5.16: *Interaction quartet.* $p_{1.1}$ *and* $p_{1.2}$ *depict the stone before and after the collision, while* $p_{2.1}$ *and* $p_{2.2}$ *represent the window. The dotted arrows represent causal adhesion (ball–window and window–ball), the others causal cohesion (ball–ball, window–window).*

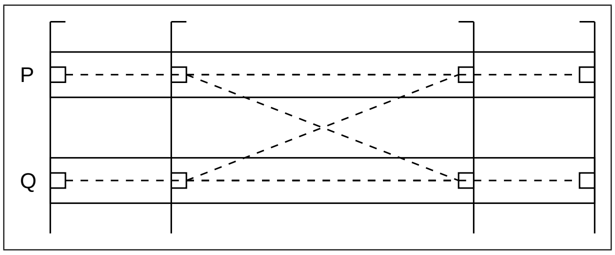


Figure 5.17: *Interaction quartet. Causal adherence within processes* P *and* Q*, and causal adhesion between* P *and* Q *over a certain interval.*

Note, that although expressing reciprocity, our model does not give up the difference between cause and effect (no presential is both cause and effect) and it maintains the cause *preceding* the effect.

The complexity of this model might be objected, but (following the four possible ways from the left to the right in figure 5.17) it already covers the causal relations relevant to the following statements:

- The ball's flight makes the window shatter (cf. fig. 5.18).
- The window slows the ball down and changes its flight's direction (cf. fig. 5.19).
- The initial velocity of the ball influences the final velocity of the ball (cf. fig. 5.20).
- The structure of the glass pane determines the way the window shatters (cf. fig. 5.21).

5.2.4.1 Causally Coherent Transition

Sometimes, it is useful to refer to the beginning and the end of a causally coherent process, so we introduce the relation of a *causally coherent transition*, which *connects the presentials at the first with those at the last boundary of a causally coherent process*:

D34. cause$_{transition}(p_1, p_2) =_{df} \exists P(\text{cause}_{coh}(P) \wedge \text{PaLp}\,(p_1, P) \wedge \text{PaRp}\,(p_2, P))$

(*Causal Transition between presentials at start/end of Process*)

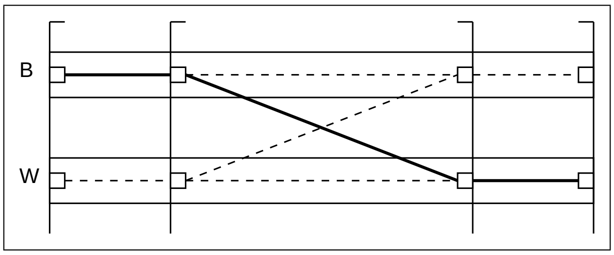


Figure 5.18: *Interaction quartet. B represents the ball's process, W the window's. Solid black line: effect of ball on window.*

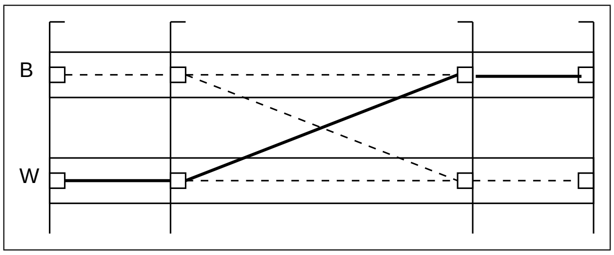


Figure 5.19: *Interaction quartet. Solid black line depicts effect of window on ball.*

Note that the presentials at stake are not themselves connected by causality. They are simply at the beginning or end of a causally coherent process.

Causal Propagation (Causality Extended in Time) Starting with a causally coherent transition, it's tempting to think of such a transition in which the coherent process *propagates* causation such that the presentials at the processes beginning and at its end are themselves connected by regularity and counterfactual dependency.

However, we do not follow this idea, here, for when analyzed carefully, it appears that causal propagation cannot be introduced as an extension of our basic causal relation but is a rather different relation:

- The time-boundaries of the presentials connected by causal propagation do not coincide (as required by the basic causal relation).
- Regularity and counterfactual dependency are themselves defined on coinciding time-

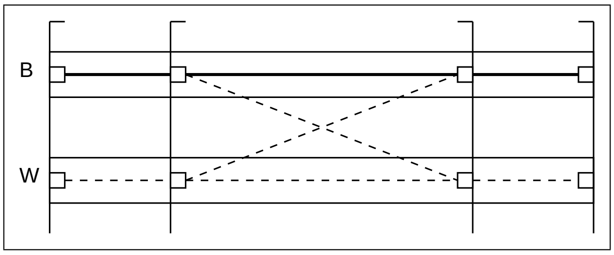


Figure 5.20: *Interaction quartet. Solid black line depicts effect of ball on itself.*

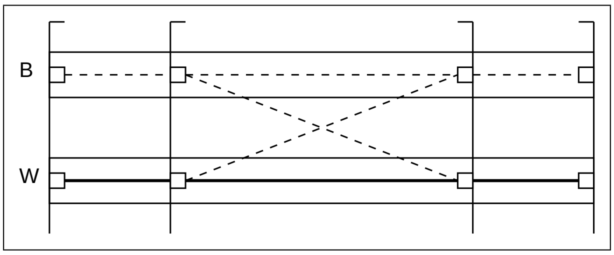


Figure 5.21: *Interaction quartet. Solid black line depicts effect of window on itself.*

boundaries, so new relations would need to be introduced.

- The basic causal relation is defined on a pair of right-boundary and left-boundary presentials while causal propagation requires a pair of left-boudary to right-boundary.

5.3 Parallel Causal Relations

Even though many of the examples used in the sections above did not make it explicit (besides when preemption was concerned, cf. sect. 2.2.4), our theory allows for an entity being causally related to more than one other entity. This includes:

1. Several causal relations holding at the same time, sharing the "cause entity", e.g. the air pressure's dropping that causes both the dropping of the barometer reading, and the thunderstorm.

2. Several causal relations holding at the same time, sharing the "effect entity", like e.g. two billiard balls hitting a third at the same time.

3. A causal relation where – on second sight – only a part of the "cause entity" actually does cause the effect like e.g. a mixture of drugs A and B that cures some disease. Even if B is ineffective to that disease, there is a causal relation between administering this mixture and the cure.

We will now take some time to explain variants two and three (assuming that variant one is not problematic) on the basis of the billiard balls example. Here we have one causal relation spanning from the first upcoming ball (B_1) to the one being hit (B_3). A second relation connects the second ball (B_2) to that very same ball B_3. Our first task, thus, is to discuss how effects of different causes may "sum up". Additionally, we think that there is a causal relation between balls B_1 and B_2 – taken as a single entity – and ball B_3 which will be discussed next.

5.3.1 Summing Up Effects

The GFO theory of causality allows for two cause entities being causally related to the same effect entity. However, we believe that the question of how exactly these effects may sum up (or: interact) is not a question of conceptual analysis, but belongs to the realm of empirical science.

Some examples may be admissible, though:

- In the billiard balls example, ball B_3's kinetic energy is a *scalar sum* of the ones transferred from balls B_1 and B_2.
- The velocities, on the other hand, are summed up by *vector addition*.
- Some effects *mutually exclude* each other: A cat is either alive or dead[85]. So if the effects "lives" and "is dead" sum up, only one of them will show up, suppressing the other.
- In cases of *overdetermination*, the combination of two causes can lead to the same result as each of the causes alone. Think of the two students throwing paper balls on a cobweb, which is destroyed.
- As in the destructive interference of waves (also called: wave subtraction), two effects might interfere in a way that none of them shows up.

5.3.2 Collated Causes

As a consequence of effects summing up in various ways, the following is possible: there is a causal relation between presential p_1 and presential p_2. But p_1 can be decomposed into two parts $p_{1.1}$ and $p_{1.2}$ with only $p_{1.1}$ being causally related to p_2. In other words, only some part of the cause presential "actually" influences the effect.[86]

Note that knowing about the inner structure does not render the initial relation. The one between $p_{1.1}$ and p_2 is just an additional parallel relation besides the one connecting p_1 and p_2. Both fulfil the conditions on regularity and counterfactual dependency. It is only that these conditions do not rely on the same clusters of alternative situations.

Take e.g. the causally contrastive alternatives that counterfactual analysis depends on. With respect to the first relation (p_1, p_2), the contrast would lie in p_1 not taking place. With respect to the second relation ($p_{1.1}$, p_2), the contrast would be that $p_{1.1}$ does not take place. So while the first relation needs alternatives as close as possible to p_1 taking place, the second needs alternatives as close as possible to only some part of p_1 (i.e. $p_{1.1}$) taking place.

[85] Leaving "Schrödinger's cat", i.e. the Copenhagen interpretation of quantum mechanics' concept of superposition, aside.

[86] This issue will play a role in the following chapter on epistemics (cf. sect. 6.2.4).

It is, thus, perfectly fine to both identify some cause via regularity and counterfactual dependency and to find that only some part of it is relevant to the effect. As stated in the introduction: we are not interested in identifying *the* cause. We are interested in what physically makes something happen. And even if we know that $p_{1.1}$ is the causally relevant part of p_1, it is still true that p_1 makes p_2 happen.

6 Epistemics and Application

In this section we will discuss the epistemic consequences of our theory, i.e. the question of: in what sense can we (or a machine) find out causal relations using our senses (extended by measuring devices) and our cognitive apparatus. We will start by collecting all the elements that our theory consists of and describe whether (and to what extent) these elements are within epistemic reach.

Then we will show the *epistemic adequacy* of our theory by reconstructing the (apparently successful) procedures used in the natural sciences in terms of the GFO theory of causation. This, actually, means applying our theory to the realm of experiments and clinical trials.

6.1 Epistemic Status of our Theory's Ontological Constituents

Let us quickly summarise the ontological building blocks of our theory:

- Presentials
- Coincidence pairs of presentials
- Universals
- A presential being an instance of a universal
- Clusters of (similar) coincidence pairs
- Clusters of causally similar coincidence pairs
- Probabilistics on clusters of coincidence pairs
- Distance between alternative situations

To describe the epistemic status of our theory, we need to show which of the elements listed above are subject to experience, to measurements or to detections (and, probably, to what extent) - and which may not be.

Presentials Objects and properties are examples of presentials, and many of them are in reach of our experiences. Although it may be debatable, in which way e.g. a certain property can be operationalised.

Examples of presentials we can measure/detect:

- the kinetic energy of a bullet
- the temperature of water in a container
- a fibre in a cobweb being ripped
- a glass pane being in good order

Status: *Partially accessible.*

Coincidence pairs of presentials Besides being able to detect an object (or a property) we need to tell the time of the measurement. Again, this might be problematic in detail, but in very many cases, there is no doubt about the possibility of measuring time.

Status: *Partially accessible.*

Universals Universals might not be directly accessible, but what indeed is needed is epistemic access to its extension (cf. the next entry)

Status: *Not of relevance.*

A presential being an instance of a universal While it may not be possible for all universals, there are those where we can identify whether a presential is an instance of that universal, or not.

Examples for presentials where we can measure/detect whether they belong to some universal's extension:

- a human being that instantiates the universal "woman"
- an animal that is an instance of the universal "hedgehog"
- a colour that belongs to "redness"
- an instance of "physical object"

Status: *Partially accessible.*

Clusters of (similar) coincidence pairs In case we know how to identify a universal's instance, we can collect all presentials (we know of) that indeed are an instance of that universal.

Status: *Partially accessible (through universals' instances).*

Clusters of causally similar coincidence pairs This just means clustering inside a cluster.

Status: *Partially accessible (through universals' instances).*

Probabilistics on cluster of coincidence pairs Once we have collected the causally similar coincidence pairs, we can calculate the ratios, our theory depends on. (We classify it as "partially accessible" due to the limitations of our experience; we cannot access *all* the relevant alternative situations).

Status: *Partially accessible.*

Distance between alternative situations As we do not have identified the criteria for closeness in our theory, we cannot say much, here. But again there are cases, in which we are sure about the "is closer to actuality than" relation.
Examples were given in section 2.2.1 where counterfactual analysis was introduced.

Status: *Partially accessible.*

It is important to see what "partially accessible" means: it implies that there are limitations in what concrete causal relations we can find out and that there are limitations on how much we can rely on our results, but none of these elements of causality is hidden to our experiences (and experiments) on principle.

One particularly relevant practical restriction (as mentioned above with respect to the probabilistic aspects of the causal relation) is that *we cannot access all relevant alternative situations*. However, we can extend the limits of, say, a single person's experience by:

- Communicating our experiences to others and hearing or reading about theirs.

- Increase the accessed alternative situations by intentionally *creating* them, i.e. by performing experiments or studies.

Especially the last point is the key to natural science being as successful an undertaking for identifying causal relations as in fact it is. In the following section we will thus have a closer look at their methods of performing experiments or trials, and we will find that these are perfectly backed up by our causal theory.

6.2 Experiments, Studies, Trials

We are strongly convinced that modern science (through its historical development and success) pragmatically is the best way to find causal relations. So any theory of causality that claims to be not only conceptually, but *epistemically justified* must be able to interpret scientific practice in this theory's terms, i.e. the theory must be able to "reconstruct" scientists' procedures, and it must be able to explain why these procedures are indeed (valuable) ways to find causal relationships. Our claim is that the GFO theory hitherto developed is indeed able to do this, and we will demonstrate this by applying our theory to the techniques used in performing experiments in general, and to clinical trials in particular.

6.2.1 Experiments

6.2.1.1 Basic Characteristics

What are the main elements of performing an experiment? We think they are the following:

Explicit Operation Procedure To perform an experiment means to create an environment and perform actions in this environment in a (for the relevant parts) explicitly specified way. This includes the involved objects, some of their properties, and the processes to take part.

Measurement Part of this explicit specification is about what measurements are taken, and in which way.

Repetition In most cases, the experiment will be run several times.

6.2.1.2 Reconstruction in the GFO Theory of Causality

Explicit Operation Procedure The explicitly specified procedure, i.e. the description of the objects and processes that will be involved, allows for creating a class of *similar* experiments. Taken this way, the specification describes the universals whose instances are relevant for drawing causal conclusions. More specifically, the procedure describes the operation, not the outcome, so it is the universals related to the cause that the operation procedure mainly is about.

Measurement Here, the outcome of the experiment is captured. In order to check that the experiment came out in a certain way, the measurements have to be interpreted (like "expected effect took place"). In terms of our theory, this interpretation means: checking whether the result belongs to, i.e. is an instance of, a certain (effect) universal.

Repetition Performing an experiment repeatedly allows for statistical analysis of the results. In terms of the GFO theory, repetition means creating alternative situations (whose similarity is generated by following the operation procedure) that then can be used for regularity and couterfactual analyses.

6.2.1.3 Summary

Given the reconstruction above, we can conclude that performing experiments incorporates all elements that a causal relation needs:

- To make sure that the alternative situations are clustered around certain cause universals, an operation procedure is to be followed.
- Effect universals then support the measurement, or more precisely: the interpretation of the raw measurement.

- And finally, the experiment is repeated to allow for analysis of probabilistic regularity and counterfactual dependency.

6.2.2 Clinical Trials

Even if scientists of different professions share the approach of performing experiments, they have developed certain methods specific to their particular fields of research.[87] To demonstrate that these methods fit our theory as well as the generic case discussed above, we will dwell on the medical field and go into some more detail concerning prospective, randomised clinical trials.

6.2.2.1 Basic Characteristics

Performed on Groups of Patients Clinical trials are not about single, individual cases, but are performed on groups of patients.

Inclusion/Exclusion There are strict criteria on what patients are included in the study. Inclusion may be based on sex, age, kind and severeness of a disease, and many other parameters.

Blocks, Branches of Treatment, Control Groups and Operation Procedures As the main idea of a clinical trial is to compare different treatments (which includes comparing some treatment with a controlled non-treated control group), there is an explicit specification of how the treatments in the different branches are to be performed.

Randomisation The included patients are assigned to the different treatments by randomisation procedures.

Collecting Results A very important part of the trial's specification is how the immediate results are interpreted. E.g. when does a patient count as cured? How to interpret if patients decease within a six month period after the treatment, or within a two year period?

Analysis Using statistical tools, the effect of the treatment (in comparison to other treatments or against a non-treatment) is calculated.[88]

[87] Cf. SELWYN (1996); COBB (1997); DEAN (1999)

[88] It is important to note that our analysis does not touch the relevance and use of the statistical methods used in analyzing trials. What we give is a conceptual background of how the trial is to be understood ontologically.

6.2.2.2 Clinical Trials: Reconstruction in the GFO Theory of Causality

Performed on Groups of Patients In the GFO view, both, regularity and counterfactual dependency rely on groups (clusters) of alternative situations. So, observing many treatments – each individual treatment creating an alternative situation, a treatment branch creating a cluster of similar treatments – is necessary in order to discover causal relations

Inclusion/Exclusion In terms of our theory, inclusion and exclusion make sure that the alternative situations are indeed clusters of similar alternative situations. In a sense they "define" the universals that the clusters' similarity relies on.

Branches of Treatment, Control Groups and Operation Procedures Branches of treatments (or explicit non-treatments in case of control groups) lead to groups of alternative situations within those which are similar through the inclusion criteria. In terms of our causal theory, these branches create causally similar (or causally contrastive) clusters that do (or explicitly do not) contain the cause. It is these clusters, then, that are relevant for evaluating the result (with respect to regularity and counterfactual dependency).

Randomization Randomisation is a technique to equally distribute variables that are not observed in the trial. In terms of the GFO theory randomisation takes additional care of the clusters of alternative situations being similar.

Collecting Results The interpretation of the results means deciding to which class an outcome belongs. This ontologically corresponds to the question of whether the result universal is instantiated or not.

Analysis The statistical methods for analysis take two things into account: counting the results within the treatment branches, and then comparing the treatments to each other. In terms of our theory this refers to the probabilities within the clusters, and to the comparison between the clusters (for regularity and couterfactual dependency analyses).

Just as a prospective, randomised clinical trial is a more elaborated version of what we have simply called "experiment" above, we find that the reconstruction is just as well elaborated. But still the GFO theory is able to cover all the presented aspects. And as we will see, they apply to even more of the actual performance of such a trial.

6.2.3 Reconstruction of Epistemic Difficulties

It is not just the successful natural science procedures that our theory is able to describe, but also its deficiencies and restrictions. Let us have a brief look at how each of the steps of a trial (as listed above) may fail, and how these restrictions can be understood as consequences of the nature of causality as we have introduced it:

Performed on Groups of Patients Sometimes, e.g. for very rare diseases, it may be difficult to get enough patients who are willing to participate in a trial. This makes evaluation difficult, both when determining the outcome of each treatment group (e.g. the success rate) and when determining the differences between the treatments.

In our theory, this is a direct implication of regularity and counterfactual dependency relying on clusters of alternative situations.

Inclusion/Exclusion If inclusion and exclusion do not follow strict criteria, a trial can easily become void. In our view, the reason for this is that similarity then is in question, which in turn bears the task of clustering alternative situations. Additionally, judging on the relative distance between the clusters may well become impossible.

But there is another effect of wide and narrow criteria: the more narrow the inclusion criteria are, the less people take part in the trial, which (as noted above) undermines analysis of the effect. However, if the inclusion criteria are too wide, applying the result to an individual (for reasons of treatment) is difficult. In our theory, the reason is that including a wide range of different situations potentially undermines their similarity.

Branches of Treatment, Control Groups and Operation Procedures Just like with inclusion and exclusion, a controlled study must make sure that the operation procedure is indeed being followed. It must be clear, which individuals were treated in what way, and which were not treated. In case of a trial on caffeine, for example, the members of the control group must not drink coffee during the trial.

In terms of our theory, not following the procedure (in our example: not making sure the that the control group does not drink coffee) prevents the necessary clustering of the individual treatments by similarity and prevents telling causally similar clusters apart from causally contrastive ones.

Randomisation Randomisation is introduced to have unobserved factors uniformly distributed among the groups. In cases where the procedure is e.g. connected to cause or effect itself (think of a doctor who finds himself not being able to deprive severely ill patients from a promising treatment – although the randomisation would have put them into the control group) this may "taint" the clusters.[89] We cannot grasp *all* alternative situations, so we must make sure they do not show a significantly different behaviour than the "average" elements of that cluster.

[89] Another unsatisfying way would be relying on people choosing a number "at random". Here's how DEAN (1999, p. 4) summarizes a study reported in the *Royal Statistical Society News and Notes (January 1988)*: "The study [. . .] asked students to pick a number at random between 0 and 9. The numbers 3 and 7 were selected by about 40% of the students. This is twice as many as would be expected if the numbers were truly selected at random."

Collecting Results Non-specific or vague interpretation criteria for the trials' outcomes (e.g. patient counts as cured) can render a trial useless.

The reason (in terms of our theory) is, that in this case, we are not able to tell supportive from undermining clusters.

Analysis A first implication of our theory (which again conforms to scientific practice) is that regularity may indicate causality, but a researcher may well go wrong if she takes regularity as implying causality. For us, it is only one of two conditions.

But there is another implication we will have a closer look at in the next section.

6.2.4 Testing Parts of a Cause as Creating Closer Alternatives

There is a certain peculiarity about the result of clinical trials we will introduce by the following example, before reconstructing/modeling it in the terms of our causal theory:

In order to test whether drug A does cure disease B, a trial might be set up as follows:

- Treatment group: patients are given a certain portion of some substance in a particular manner e.g. they need to visit their doctor twice a week and swallow some pills.
- Control group: patients are not treated in any way similar to the above.

The outcome then might be that 30% of the treatment group's patients is cured, while there was no cure in the control group. From this, it may be reasonably inferred that the drug works as expected, i.e. drug A cures disease B.

However, another trial might be set up with an additional branch, which takes the so called placebo effect into account.

- Placebo group: The patients do follow the same treatment procedure as in the treatment group, but their pills do not contain drug A (instead they swallow a substance known to have no effect on B).

And now it might turn out that this "treatment" has a success ratio of 40% (which is more than with the real drug). While the first trial leads us to thinking that the drug indeed cures the disease, the second trial, now, persuades us to accept the opposite result: A does not cure B; at the same time indicating the treatment procedure being the cause for the healing. And if asked about the efficacy of A on B, we would deny A any effect. But does this mean that the first trial's outcome is somehow rejected or denied? It clearly must have *some* significance.

To solve what seems to be a dilemma, let us first examine what lead us to our inference: The difference between the first and the second trial is that the first took drug A and a certain treatment procedure (P, say) as a single, united, cause. The result, then, was that this cause indeed is causally related to disease B. The second trial, however, was not about the same hypothesis ("A and P have effect on B"), but split off P which then lead to the result that A without P does not cure B. The final picture, then, is the following: while the first trial tried to test the effect of A on B it actually tested $A+P$ on B. And re-formulated in this way, it matches the result of the second trial: $A + P$ does have an effect, as it contains the "real" cause P.

This situation can be described very easily using our causal theory: In the first trial, there was one contrastive alternative situation and it proved to be supportive so counterfactual dependency was inferred (cf. fig. 6.1).

World (ordered by closeness)	*supportive / undermining*	*Counterfactual holds*
World 1: No drug	supportive	Yes

Table 6.1: *Counterfactual analysis of first trial*

But in the light of the second trial's result, we find that what the first trial simply referred to "drug A" actually consisted of A *and* P, so the situations becomes like depicted in table 6.2

World (ordered by closeness)	*supportive / undermining*	*Counterfactual holds*
World 1: No drug, no procedure	supportive	Yes

Table 6.2: *Counterfactual analysis of first trial, re-interpreted after the second trial*

For testing the drug without the procedure, the second trial created contrastive situations that were closer to "A and B cause C" than the one in the first trial. And that closer cluster proved to be undermining (cf. fig. 6.3).

World (ordered by closeness)	*supportive / undermining*	*Counterfactual holds*
World 2: No drug; but procedure	undermining	No
World 1: No drug; no procedure	supportive	

Table 6.3: *Counterfactual analysis of clinical trial of drug trial including placebo*

As we are not able to test *all* alternative situations, it can never be ruled out that new experiments/trials may reveal that only part of what we discovered as a cause is relevant for the effect to take place. This, as said before (cf. sect. 5.3.2), does not imply that the "old" causal relation no longer holds. We just discovered another (additional) causal relation.

Summarizing, we can say that experiments or clinical trials can ontologically be understood as ways of *creating custom-made alternative situations* on which regularity and counterfactual dependency can then be evaluated (probabilistically). This approach allows for reconstructing the way that natural sciences' methods succeed in discovering causal relations. Additionally, it allows for ontologically illustrating the various ways in which experiments and clinical trials may go wrong.

7 How to Move On

In the preceding chapters, we have discussed what the concept of causality is about, how it can be modeled with the inventory of GFO, and finally how adequate our theory is with respect to epistemics. Starting from each of these steps, there are several areas that the present work may be extended to: concept analysis, formal ontological modeling, and relations to other fields of research. Here are some examples:

Concept Analysis

- Our investigations were restricted to physical causality. What does “make happen” mean to other fields, like in “this made me laugh”?
- Do other kinds of causality still rely on regularity and counterfactual dependency? Or do we need to discriminate between several kinds of causality?
- We explicitly refrained from the question of how to identify *the* cause as opposed to stating that something is *a* cause. Is there a way to conceptually single out the one cause from the causes that our theory is about? At least for a given context and a certain interest of the speaker?
- We did not discuss possible causal relations. How would we approach the question of dispositions from our theory’s point of view?
- The notion of “distance” between clusters of alternative situations was introduced as a primitive relation. But we already saw that when performing experiments, scientists do share some pragmatic idea of what rules for achieving relevance are. Could a general theory of distance be derived from that knowledge?

Ontological Modeling

- Our theory is based on presentials and has been extended to cover processes as well. Is it possible to make it cover even more kinds of entities?
- If we leave the realm of physical entities, what takes the place of presentials, then? And whatever the replacement is, what is its relation to time?
- We allowed for effects to “sum up”, or “interact”. Is there a theory that covers these interactions? Are there families of properties, for example, that share the way they are summed up?

- The basic causal relation relies on coincidence pairs, which are characteristic of GFO compared to other top-level ontologies. Does that mean that causal theories based on these other ontologies are flawed right from the start? Could these theories be adjusted to other areas (with respect to the nature of the relata for example) to account for that?
- For modeling counterfactual dependency, we relied on a restricted variant of possible worlds, i.e. possible worlds understood as alternative situations of the actual world's history and future. Is this a feasible approach to possibilities as they are discussed outside the causal context? Can our alternative situations be used in full-fledged theories of necessity?

Relations to Other Fields

- We have seen that the GFO theory of causality supports the practice of clinical trials. How, then, does it fit in with ontologies in exactly that field, like the "Ontology of Clinical Research"(cf. CARINI ET AL., 2009) for example?
- When discussing the epistemic status of our theory, we did not discriminate sharply between a constituent of causality being accessible to human senses and that constituent being detectable by a machine's sensors. But does a machine really have access to all the relevant entities? How would universals need to be understood in this case? As a list of detectable parameters?
- If "detecting causality" (following our theory, that is) is open to machines, could our theory of causality contribute to "the automation of science", as KING ET AL. (2009) call the aim of their project of a "Robot Scientist" that is said to identify "genes encoding orphan enzymes"?

  Or, the other way round, starting from observing causal relations, could our theory play a role in machines' "Distilling Free-Form Natural Laws from Experimental Data" as presented in SCHMIDT and LIPSON (2009)?

We started our investigations by intentionally ignoring RUSSELL'S advice to remove causality from the scientific vocabulary because of its "misleading associations" (RUSSELL, 1910, p. 180). Instead, we have succeeded in giving it a clearly defined meaning, at least for the physical world. Our conclusion therefore is as different to RUSSELL'S as it can be. We do not only explicitly include causal relations into our own ontology, but we also believe that several other fields would benefit from introducing causality to their ontological models as well.

A Proofs on Conditional Probabiliy

Proposition 1.

$$P(A|B) > P(A) \quad \rightarrow \quad P(A|\bar{B}) < P(A) \tag{Prop1}$$

Proof.

$$\begin{aligned}
P(A) &= P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) && \text{[Law of total probability]} && \text{(A.1)}\\
P(A) &= P(A)P(B) + P(A)P(\bar{B}) && [P(B) + P(\bar{B}) = 1] && \text{(A.2)}\\
0 &= [P(A|B) - P(A)]P(B) + [P(A|\bar{B}) - P(A)]P(\bar{B}) && \text{[A.1 = A.2]} && \text{(A.3)}
\end{aligned}$$

Examining (A.3), we know that $P(B)$ and $P(\bar{B})$ are nonnegative. And, given the antecedent of (Prop1), the same holds for $[P(A|B) - P(A)]$. So $[P(A|\bar{B}) - P(A)]$ of (A.3) must be negative, which is the consequent of (Prop1). □

Starting from (Prop1) the following holds as well:

Corollary 1.

$$P(A|B) > P(A) \quad \rightarrow \quad P(A|B) > P(A|\bar{B}) \tag{Cor1}$$

Proof.

$$\begin{aligned}
P(A|B) &> P(A) && \text{[Antecedent of (Cor1)]} && \text{(A.4)}\\
P(A) &> P(A|\bar{B}) && \text{[A.4 and (Prop1)]} && \text{(A.5)}\\
P(A|B) &> P(A|\bar{B}) && \text{[A.4, A.5]} && \text{(A.6)}
\end{aligned}$$

□

Interchanging antecedent and consequent, however, leads to a contradiction, so the expressions $P(A|B) > P(A)$ and $P(A|B) > P(A|\bar{B})$ are not equivalent:

Corollary 2.

$$\neg\,[\;P(A|B) > P(A|\bar{B}) \quad \rightarrow \quad P(A|B) > P(A)\;] \tag{Cor2}$$

Proof.

$$P(A|B) > P(A|\bar{B}) \quad \rightarrow \quad P(A|B) > P(A) \qquad \text{[Negating (Cor2)]} \tag{A.7}$$

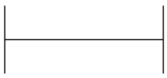
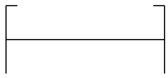
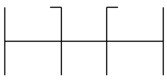
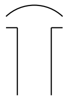
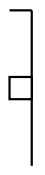
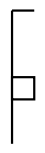
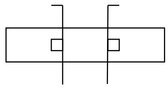
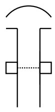
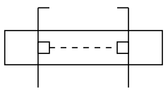
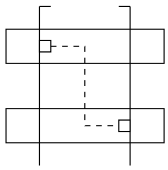
$$P(A|\bar{B}) > P(A) \qquad \text{[A.7]} \tag{A.8}$$

$$P(A|\bar{B}) > P(A|B) \qquad \text{[A.8, (Cor1)]} \tag{A.9}$$

$$\text{Contradiction!} \qquad \text{[A.7, A.9]} \tag{A.10}$$

□

B Keys for Ontological Diagrams

 Chronoid with its two extremal boundaries	 Chronoid's external boundaries explicitly marked as left and right boundary.
 Chronoid with inner time-boundaries explicitly marked as right and left boundary	 Coinciding right and left time-boundary
 Presential at a right time-boundary	 Presential at a left time-boundary
 Process	 Process with presentials at inner time-boundaries (projected on process boundaries)
 Causal relation between two presentials	 Coincidence pair (Presential at coinciding time-boundaries which are connected by causality)
 Causally cohesive part of a single process	 Causally adhesive parts of different processes

Index

Bibliography

ALLEN, Patrick J., James F. & Hayes (1989). “Moments and points in an interval-based temporal logic.” In: *Computational Intelligence*, 5:225–238.

BACON, John (2002). “Tropes.” In: *The Stanford Encyclopedia of Philosophy* (edited by Edward N. Zalta), URL `http://plato.stanford.edu/archives/fall2002/entries/tropes/`, fall 2002 Edition, link checked 2009/04/05.

BBC NEWS WEBSITE (2006). “World marvels at total eclipse.” URL `http://news.bbc.co.uk/2/hi/science/nature/4849224.stm`, link checked 2009/02/05.

BEEBEE, Helen, HITCHCOCK, Christopher, MENZIES, Peter (editors) (to appear). *The Oxford Handbook of Causation*. "To appear" as of 2009/02/05 (cf. `http://www.oup.co.uk/academic/socsci/handbooks/philosophy/`)Cyc.

BENNETT, Jonathan (1988). *Events and their Names*. Hackett Publishers, Indianapolis.

BUTCHVAROV, Panayot (1995). “Metaphysics.” In: *The Cambridge Dictionary of Philosophy* (edited by Robert Audi), pp. 489–491, Cambridge University Press.

CARINI, Simona, HARRIS, Steve, MACCALLUM, Peter, RECTOR, Alan, SIM, Ida, TOUJILOV, Igor, TU, Samson (2009). “The ontology of clinical research (ocre).” URL `http://rctbank.ucsf.edu/home/ocre.html`, link checked 2009/04/03.

CHUNG, Kai Lai, AITSALHLIA, Farid (2003). *Elementary Probability Theory. With Stochastic Processes and an Introduction to Mathematical Finance*. Springer, 4th edition.

CLANCEY, William J. (1993). “The Knowledge Level Reinterpreted: Modeling Socio-Technical Systems.” In: *Knowledge Acquisition as Modeling* (edited by K. M. Ford, J. M. Bradshaw), pp. 33–50, John Wiley & Sons, New York.

CLARK, Ronald William (1972). *Einstein : the life and times*. Avon Books, New York, N.Y.

COBB, George W. (1997). *Introduction to Design and Analysis of Experiments*. Textbooks in mathematical sciences, Springer, New York, Berlin, Heidelberg.

COLLINGWOOD, Robin George (1940). *An Essay on Metaphysics*. Clarendon Press, Oxford.

COLLINS, John, HALL, Ned, PAUL, Laurie Ann (2004a). *Causation and Counterfactuals*. MIT Press, Cambridge, Mass.

— (2004b). "Counterfactuals and causation: History, problems, prospects." In: *Causation and Counterfactuals* (edited by Ned Collins, John an dHall, L. A. Paul), The MIT Press, Cambridge, Massachusetts. / London, England.

COMBOURIEU, Marie-Christine (1992). "Karl R. Popper, 1992: About the EPR controversy." In: *Foundations of Physics*, V22(10):1303–1323, URL `http://dx.doi.org/10.1007/BF01889715`, link checked 2009/02/05.

COPELAND, B. Jack (1997). "Cyc: A case study in ontological engineering." In: *The Electronic Journal of Analytic Philosophy*, 5, URL `http://ejap.louisiana.edu/EJAP/1997.spring/copeland976.2.html`, link checked 2009/02/05.

CYC FOUNDATION (2008). "Opencyc concept browser: causePropProp." URL `http://www.cycfoundation.org/concepts/causes-PropProp`, link checked 2009/02/05.

CYCORP INC. (2008). "the Cyc Knowledge Base^{TM}." URL `http://www.cyc.com/cyc/technology/whatiscyc_dir/technology/technology/whatiscyc_dir/whatsincyc`, link checked 2009/02/05.

DE LAPLACE, Pierre Simon (1814). *Essai philosophique sur les probabilités*. Paris.

DEAN, Daniel, Angela & Voss (editor) (1999). *Design and Analysis of Experiments*. Springer Texts in Statistics, Springer, New York, Berlin, Heidelberg.

DESEL, Jörg, REISIG, Wolfgang, ROZENBERG, Grzegorz (editors) (2004). *Lectures on Concurrency and Petri Nets*. Springer, Berlin.

DOWE, Phil (2004). "Causal Processes." In: *The Stanford Encyclopedia of Philosophy* (edited by Edward N. Zalta), URL `http://plato.stanford.edu/archives/win2004/entries/causation-process/`, winter 2004 edition, link checked 2009/02/05.

GÄRDENFORS, Peter (2000). *Conceptual Spaces: The Geometry of Thought*. The MIT Press, Cambridge, MA, USA.

GODFREY-SMITH, Peter (2006). "Causal pluralism." In: *Oxford Handbook of Causation*, URL `http://www.people.fas.harvard.edu/~pgs/OnlinePapers/CausalPluralismOx06E.pdf`, to appear in BEEBEE ET AL. (to appear), link checked 2008/10/04.

GOODMAN, Nelson (1951). *The Structure of Appearance*. Harvard University Press, Cambridge, MA.

GRACIA, Jorge J. E. (1995). "Individuation." In: *The Cambridge Dictionary of Philosophy* (edited by Robert Audi), pp. 367–368, Cambridge University Press, Cambridge, UK.

GRENON, Pierre (2003). "BFO in a nutshell." Technical Report 06, Institute for Formal Ontology and Medical Information Science, Leipzig.

GROSS, Rudolf, LÖFFLER, Markus (1998). *Prinzipien der Medizin. Eine Übersicht ihrer Grundlagen und Methoden*. Springer, Berlin, Heidelberg, [1]1997.

GRUBER, Thomas R. (1993). "A translation approach to portable ontology specifications." In: *Knowledge Acquisition*, 5(2):199–220.

GUARINO, Nicola (1995). "Formal Ontology, Conceptual Analysis and Knowledge Representation." In: *International Journal of Human-Computer Studies*, 43(5/6):625–640.

— (1998). "Formal ontology and information systems." In: *Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998*, pp. 3–15, IOS Press, Amsterdam, URL `http://www.loa-cnr.it/Papers/FOIS98.pdf`, link checked 2009/02/05.

GUHA, Ramanathan V. (1990). "The representation of defaults in cyc." In: *AAAI*, pp. 608–614.

GUHA, R.V., LENAT, D.B. (1990). "Cyc: A mid-term report." URL `http://www.cyc.com/doc/articles/midterm_report_1990.pdf`, link checked 2009/02/05.

HALL, Ned (2001). "Two Concepts of Causation." In: COLLINS ET AL. (2004a).

HARTMANN, Nicolai (1964). *Der Aufbau der realen Welt. Grundriß der allgemeinen Kategorienlehre.* Walter de Gruyter and Co., Berlin.

HAUSMAN, D. (1998). *Causal Asymmeries*. Cambridge University Press, Cambridge.

HAWKING, Steven (2006). "Life in the universe." URL `http://hawking.org.uk/index.php?option=com_content&view=article&id=65`, link checked 2009/02/05.

HAYES, Pat (1996). "A catalog of temporal theories." Technical Report UIUC-BI-AI-96-01, University of Illinois.

HELLER, Barbara, HERRE, Heinrich (2003). "Formal ontology and principles of gol. onto-med report nr. 1." Technical report, Research Group Ontologies in Medicine (Onto-Med), University of Leipzig, Leipzig, Germany.

— (2004a). "General Ontological Language (GOL) – A Formal Framework for Building and Representing Ontologies. Version 1.0." In: *Onto-Med Report (University of Leipzig)*, (7).

— (2004b). “Ontological categories in gol.” In: *Axiomathes*, 1(14):57–76.

HERRE, Heinrich, HELLER, Barbara, HOEHNDORF, Robert, LOEBE, Frank, MICHALEK, Hannes (2007). “General formal ontology (gfo): A foundational ontology integrating objects and processes. part i: Basic principles (version 1.0.1).” Technical report, Research Group Ontologies in Medicine (Onto-Med), University of Leipzig.

HITCHCOCK, Christopher (2002). “Probabilistic Causation.” In: *The Stanford Encyclopedia of Philosophy* (edited by Edward N. Zalta), URL `http://plato.stanford.edu/archives/fall2002/entries/causation-probabilistic/`, fall 2002 edition, link checked 2009/02/05.

— (2003). “Of Humean bondage.” In: *British Journal for the Philosophy of Science*, 54:1–25, URL `http://bjps.oxfordjournals.org/cgi/reprint/54/1/1`, link checked 2009/02/11.

HOEFER, Carl (2005). “Causal determinism.” In: *The Stanford Encyclopedia of Philosophy* (edited by Edward N. Zalta), URL `http://plato.stanford.edu/archives/sum2005/entries/determinism-causal/`, summer 2005 edition, link checked 2009/02/05.

HORROCKS, Ian (2005). “D29: Final report.” URL `http://wonderweb.semanticweb.org/deliverables/D29.shtml`, link checked 2007/06/29.

HUME, David (1748). *An Enquiry concerning Human Understanding.*

JUBIEN, Michael (1997). *Contemporary Metaphysics*. Blackwell Publishers, Oxford, Malden (Mass.).

KANT, Immanuel (1787). *Kritik der reinen Vernunft*. Riga, 2nd ed.

KING, Ross D., ROWLAND, Jem, OLIVER, Stephen G., YOUNG, Michael, AUBREY, Wayne, BYRNE, Emma, LIAKATA, Maria, MARKHAM, Magdalena, PIR, Pinar, SOLDATOVA, Larisa N., SPARKES, Andrew, WHELAN, Kenneth E., CLARE, Amanda (2009). “The automation of science.” In: *Science*, 324(5923):85–89, doi:10.1126/science.1165620, URL `http://www.sciencemag.org/cgi/content/abstract/324/5923/85`.

LEHMANN, Jos (2003). *Causation in Artificial Intelligence and Law*. SIKS Dissertation Series, Amsterdam.

LEHMANN, Jos, BORGO, Stefan, GANGEMI, Aldo, MASOLO, Claudio (2004). “Causality and Causation in DOLCE.” In: *Formal Ontology in Information Systems, Proceedings of the International Conference FOIS 2004*.

LEJEWSKI, Czesław (1954). "Logic and existence." In: *The British Journal for the Philosophy of Science*, 5.

LEWIS, David K. (1973). "Causation." In: *Journal of Philosophy*, 70:556–67.

— (1986). *On the Plurality of Worlds*. Blackwell Publishing, Oxford, UK (and elsewhere).

— (2000a). "Causation as Influence." In: COLLINS ET AL. (2004a).

— (2000b). "Causation as Influence." In: *Journal of Philosophy*, 97:182–197, URL `http://www.jstor.org/stable/2678389`, link checked 2009/02/05.

MACKIE, J. L. (1980). *The Cement of the Universe*. Oxford University Press, Oxford.

MACKIE, PENELOPE (1992). "Causing, delaying, and hastening: Do rains cause fires?" In: *Mind*, 101(403):483–500, doi:10.1093/mind/101.403.483, URL `http://mind.oxfordjournals.org`, link checked 2009/02/05.

MASOLO, Claudio, BORGO, Stefano, GANGEMI, Aldo, GUARINO, Nicola, OLTRAMARI, Allessandro (2003). "WonderWeb Deliverable D 18. Ontology Library (final)." Technical report, Laboratory For Applied Ontology - ISTC-CNR.

MENZIES, Peter (2001). "Counterfactual Theories of Causation." In: *The Stanford Encyclopedia of Philosophy* (edited by Edward N. Zalta), URL `http://plato.stanford.edu/archives/spr2001/entries/causation-counterfactual/`.

MICHALEK, Hannes (2002). *Alvin Plantingas Theorie der Notwendigkeit (English: Alvin Plantinga's Theory of Necessity)*. Master's thesis, Hochschule für Philosophie, Philosophische Fakultät S.J., München (Munich School of Philosophy), München, URL `http://www.informatik.uni-leipzig.de/~michalek/download/Magister.pdf`, link checked 2009/02/05.

— (2005). "A causal relation based on regularity and manipulability." In: *Proceedings of the EDOC International Workshop on Vocabularies, Ontologies and Rules for the Enterprise (VORTE'05), Enschede, The Netherlands, Sep 20*, pp. 1–7, CTIT, URL `http://www.ctit.utwente.nl/library/proceedings/vorte.pdf`, link checked 2009/02/05.

NAKHNIKIAN, George, SALMON, Wesley C. (1957). "'exist' as a predicate." In: *Philosophical Review*, 66.

NEWELL, Allen, SIMON, A., Herbert (1976). "Computer Science as Empirical Iquiry: Symbols and Search." In: *Communications of the ACM*, 19(3):113–126.

NEWTON, Sir, Isaac (1686). *Philoosphiae Naturalis Principia Mathematica*. Jussu Societas Regiae ac Typis Josephi Streater, Prostant Venales apud Sam. Smith . . . , London, URL http://books.google.com/books/pdf/Philosophiae_naturalis_principia_mathema.pdf?id=XJwx0lnKvOgC&output=pdf&sig=ACfU3U3lC0pvW1b5skErPqWoNSmDRbbBlg&source=gbs_summary_r&cad=0, link checked 2009/02/05.

NHS CHOICES (2009). "Cold, common." URL http://www.nhs.uk/conditions/Cold-common/Pages/Introduction.aspx?url=Pages/what-is-it.aspx, link checked 2009/02/05.

NOORDHOF, Paul (1999). "Probabilistic causation, preemption and counterfactuals." In: *Mind*, 108(429):95–125, ISSN 00264423, URL http://www.jstor.org/stable/2659904, link checked 2009/02/05.

ONTO-MED RESEARCH GROUP (2008). "General formal ontology. Brief history." online, URL http://www.onto-med.de/en/theories/gfo/index.html#hist, link checked 2009/02/05.

PAUL, L.A. (1998). "Keeping track of the time: Emending the counterfactual analysis of causation." In: *Analysis*, 58(3):191–198, URL http://dx.doi.org/10.1111/1467-8284.00121, link checked 2009/02/05.

PEARL, Judea (1993). "Aspects of graphical models connected with causality." In: *Proceedings of the 49th Session of the International Statistical Institute*, volume 55, pp. 399–401, Florence, Italy, URL http://ftp.cs.ucla.edu/pub/stat_ser/R195-LL.pdf, UCLA Cognitive Systems Laboratory, Technical Report (R-195-LL), link checked 2009/02/05.

— (1994). "From bayesian networks to causal networks." In: *Bayesian Networks and Probabilistic Reasoning* (edited by A. Gammerman), pp. 1–31, Alfred Walter Ltd., London, URL http://ftp.cs.ucla.edu/pub/stat_ser/R195-LLL.pdf, UCLA Cognitive Systems Laboratory, Technical Report (R-195-LLL), link checked 2009/02/05.

— (2000). *Causality. Models, Reasoning, and Inference*. Cambridge University Press, New York, 2005 (5th reprint).

PETRI NETS WORLD (2007). "Petri nets world - faq." URL http://www.informatik.uni-hamburg.de/TGI/PetriNets/faq/, link checked 2009/02/05.

PINCH, Trevor J (1979). "The hidden-variables controversy in quantum physics." In: *Physics Education*, 14(1):48–52, URL http://stacks.iop.org/0031-9120/14/48.

PLANTINGA, Alvin (1974). *The Nature of Necessity*. Oxford University Press, Oxford [a.o.].

POLI, Roberto (2001). "The Basic Problem of the Theory of Levels of Reality." In: *Axiomathes*, (12):261–283.

RAMACHANDRAN, Murali (2004). "A counterfactual analysis of indeterministic causation." In: COLLINS ET AL. (2004a).

RESCHER, Nicholas (1979). "The ontology of the possible." In: *The Possible and the Actual: Readings in the Metaphysics of Modality* (edited by Michael J. Loux), Cornell University Press, Ithaka.

— (1999). "How many possible worlds are there?" In: *Philosophy and Phenomenological Research*, 59(2):403–420, ISSN 00318205, URL `http://www.jstor.org/stable/2653678`, link checked 2009/02/05.

RESCHER, Nicholas, PARKS, Zane (1973). "Possible individuals, trans-world identity, and quantified modal logic." In: *Noûs*, 7(4):330–350, ISSN 00294624, URL `http://www.jstor.org/stable/2214360`, link checked 2009/02/05.

RUSSELL, Bertrand (1910). "On the notion of cause." In: *Mysticism and Logic. And other Essays.*, George Allen & Unwin Ltd., London, 11 (1959) edition, first published as "Philosphical Essays". Second edition as "Mysticism and Logic".

SCHAFFER, Jonathan (2000). "Trumping preemption." In: *The Journal of Philosophy*, 97(4):165–181, ISSN 0022362X, URL `http://www.jstor.org/stable/2678388`, link checked 2009/02/05.

— (2001). "Causes as probability raisers of processes." In: *The Journal of Philosophy*, 98(2):75–92, ISSN 0022362X, URL `http://www.jstor.org/stable/2678483`, link checked 2009/02/05.

— (2003). "The Metaphysics of Causation." In: *The Stanford Encyclopedia of Philosophy* (edited by Edward N. Zalta), URL `http://plato.stanford.edu/archives/spr2003/entries/causation-metaphysics/`, spring 2003 edition, link checked 2009/02/05.

SCHLEGEL, Richard (1970). "Statistical explanation in physics: The copenhagen interpretation." In: *Synthese*, V21(1):65–82, URL `http://dx.doi.org/10.1007/BF00414188`, link checked 2009/02/05.

SCHMIDT, Michael, LIPSON, Hod (2009). "Distilling free-form natural laws from experimental data." In: *Science*, 324(5923):81–85, doi:10.1126/science.1165893, URL `http://www.sciencemag.org/cgi/content/abstract/324/5923/81`.

SCHNEIDER, Luc (2003a). "Designing foundational ontologies. the object-centered high-level reference ontology ochre as a case study." In: *ER 2003. 22nd International Conference on Conceptual Modeling*.

— (2003b). "How to build a foundational ontology. the object-centered high-level reference ontology ochre." In: *KI 2003 ? 26th German Conference on Artificial Intelligence (Lecture Notes in Computer Science 2821)* (edited by A. Günter, R. Kruse, Neumann B.), Springer, Berlin.

SELWYN, Murray R. (1996). *Principles of Experimental Design for the Life Sciences*. CRC Press, Boca Raton, Florida.

SIMON, H. A. (1977). *Models of Discovery: and Other Topics in the Methods of Science*. D. Reidel, Dordrecht, Holland.

SOWA, John F. (2000a). "Causality and causation." Posted on suo.ieee.org mailing list, URL `http://suo.ieee.org/email/msg00043.html`, link checked 2009/02/05.

— (2000b). *Knowledge Representation. Logial, Philosophical and Computational Foundations*. Brooks/Cole, Pacific Grove, CA, USA.

— (2000c). "Processes and Causality." URL `http://www.jfsowa.com/ontology/causal.htm`, (as of: 14.10.2003), link checked 2009/02/05.

SPIRTES, P., GLYMOUR, C, SCHEINES, R. (1993). *Causation, Prediction and Search*. Springer-Verlag, New York.

VON WRIGHT, Georg Henrik (1971). *Explanation and Understanding*. Cornell University Press, Ithica, New York.

WOODWARD, James (1997). "Explanation, Invariance and Intervention." In: *PSA 1996*, 2:26–41.

— (2000). "Explanation and Invariance in the Special Sciences." In: *British Journal for the Philosophy of Science*, 51:197–254.

— (2001). "Causation and Manipulability." In: *The Stanford Encyclopedia of Philosophy (Fall 2001 Edition)* (edited by Edward N. Zalta), URL `http://plato.stanford.edu/archives/fall2001/entries/causation-mani/`, fall 2001 edition, link checked 2009/02/05.

WOODWARD, James, HITCHCOCK, Christopher (2003). "Explanatory Generalizations, Part I: A Counterfactual Account." In: *Nous*, 37(1).

Formalities

Wissenschaftlicher Werdegang

1997	**Abitur**
	Theodor-Heuss-Gymnasium Schopfheim
1998 - 2000	Grundstudium Philosophie, Magister
	Hochschule für Philosophie, Philosophische Fakultät SJ, München
2000	**Bakkalaureat**
	(Philosophie)
2000 - 2002	Hauptstudium Philosophie, Magister
	Hochschule für Philosophie, Philosophische Fakultät SJ, München
2000 - 2001	Grundstudium Physik, Diplom
	Ludwig-Maximilians-Universität München
2001 - 2003	Grundstudium Physik, Diplom
	Universität Leipzig
2002	**Magister Artium**
	(Metaphysik, Grundlegung der Sozialwissenschaften,
	Wissenschaftstheorie, Allgemeine Philosophie)
2002 - 2003	*Wiss. Hilfskraft* am Institut für Theoretische Physik
	Universität Leipzig
2003 - 2005	*Wiss. Hilfskraft* am Institut für Informatik
	Universität Leipzig
Seit 2004	Promotionsstudium Informatik
	Universität Leipzig

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachte Dienstleistungen als solche gekennzeichnet.

Leipzig,