

ON COUNTERFACTUAL REASONING

Abstract:

Counterfactual reasoning has always played a role in human life. We ask questions like, “Could it have been different?”, “Under which conditions might/would it have been different?”, “What would have happened if...?” If we don’t find an answer, i.e. what we accept as an answer, we may start *reasoning*. Reasoning means introducing still new information/assumptions, new questions, new answers to new questions etc. From a formal point of view, it may be compared with stepwise moving towards a destination in a path-system, in which you never fully have an overview. In this way, reasoning is an activity, with its own rationale, which will be studied from the agent’s own perspective. Questions include: What are the conditions where asking that specific question, or introducing this information/assumption, etc. will count as a reasonable step or progress towards the answer of the initial question? What makes this step more reasonable than another?

Introduction

It was Goodman who in 1947 coined the term ‘counterfactual’, and defined a counterfactual conditional as a conditional with an if-clause which is contrary to fact. Since Lewis (1973) and Stalnaker (1968) modeled counterfactuals on the possible world semantics of modal logic, the subject has increasingly engaged philosophers (Nozick 1981), (Ginsberg 1986), (Pearl 2000), (Bennett 2003), just to name some of the most influential theorists). The basic questions discussed are: What is the meaning of a counterfactual conditional sentence? What are its truth-conditions – if there are any? The epistemology? What is the relation between counterfactual conditionals and the meaning of causal claims?

This paper does not – at least not immediately – enter any of these discourses, let alone add a new theory. The very topic is different: I will not study counterfactuals, or “counterfactuality”, but counterfactual *reasoning*. By reasoning, I shall here understand the sequence of *considerations* – assertions, assumptions, rejections, questions, answers etc. – that you make when you progress towards an answer of a question that you cannot immediately answer. The individual considerations I shall call steps, and the analysis of reasoning that I want to make is concerned about the principles that you must follow, and indeed *do* follow, when you step by step progress from an initial question – via other questions, and answers, and the introduction of new premises – towards the answer of that question. Indeed, there is no linear connection; it’s rather like moving towards a destination in a path-system of which you never have a clear overview.

The steps are *not* sentences, since they are performed, but they will be linguistically expressed by sentences, starting out with a sentence in the interrogative, further including sentences in the indicative and, of course, other interrogatives.

When reasoning specifically is counterfactual, it will further include counterfactual sentences, not only counterfactual conditionals (“If A had been the case, C would have been the case”), but also the

“simple” conditional (“X – not being the case – might have been the case”), and the initial question will be a counterfactual interrogative e.g. “Is it possible that X – not being the case – might/could have been the case?”, “What are the conditions, if there are any, under which X would have been the case?”, “Would/might X be the case, if Y had been the case?” Etc.

1. Elements of Counterfactual Reasoning

Counterfactual reasoning [CFR] is a type of dialectics we all practise from time to time, be it with other persons or on our own, and probably more often than we used to think. Counterfactuality may operate down to the very details of our vocabulary – not only in verbs in the subjunctive mode: When we say about a person that he “arrived *at the last moment* at the platform to catch the train”, the meaning is about the same as “had he arrived a moment later he would have missed the train”. Sometimes the factual cannot be proposed without the assumption of the counterfactual: When we say that the jacket protected Peter from getting wet, we assume that had he not worn the jacket, he would have become wet. In the description of the factual there are already counterfactual assumptions that can’t simply be left out.

As reasoning it always begins with questions such as: “Could it have been different?”, “Under which conditions might/would it have been different? “What would have happened if...? Subsequently, it develops in sequences of answering and asking again.

Since CFR is something we *do* we may ask: What is the point of doing it? When does such an activity make sense? When and why is it reasonable to make CFR? In history for instance it obviously makes sense – it can’t even be avoided (although some historians think they should try to avoid it.) When historians maintain that the “explanation” why Arabs didn’t develop a separation of church and state is that they didn’t have an Enlightenment, then a counterfactual point is presupposed: If there had been an Enlightenment, then they might/would have developed a separation of church and state (and maybe had a democracy in our modern sense). But in our own lives counterfactual reasoning often feels strange: The things we reason about are things we after all cannot do anything about. Since things we consider in so far as we might have done something about them, are things that we might have wanted to do something about, our CFR particularly often deals with what is unwanted or even painful, “negative”: “Could the conflict have been avoided?”, “If NN had taken the medicine M, would he have become well again?”, etc. Yet they may also be positive: “Had I not gone to that recital, would I ever have met my wife?” Why - or when – is it reasonable to reason in counterfactual terms? After all: “Glücklich ist, wer vergisst, was nicht mehr zu ändern ist.” (Johan Strauss, *Die Fledermaus*). Is it so?

These and similar questions are actually what triggered my interest for CFR, but in the following I shall stick to another – derived – issue: The internal structure of *reasonable* CFR. What are the elements of and principles for a reasonable way to reason in counterfactual terms?

All reasoning as a starting point takes something for granted – assumed, known – before it starts out asking questions, finding answers and so on. In CFR the starting point is: Something *p* is the case. Reasoning then, in the simplest mode, starts with a question: Is it possible that *p* might not have been the case? What then are the conditions under which *p* might/would not have been the case?¹ And what would have happened if *p* had not been the case?

¹ On ‘would’ versus ‘might’ subjunctive conditionals: The stronger version ‘ $B \square \rightarrow A$ ’ versus – i.e. from which follows – the weaker version ‘ $B \diamond \rightarrow A$ ’. This is an important distinction. However, I shall not care so much about it in these notes, since I don’t believe that it might/would(!) make any significant change in my analysis. In cases where I have to choose

Only facts may be objects of CFR. And basically all facts i.e. whatever is “*factual*” may *qua* facts be objects of *counterfactual* reasoning. The term ‘fact’ is derived from the Latin perfect participle of ‘facio’, i.e. ‘factus’: “*something done*”. And obviously CFR are typically performed about actions/events that were performed/happened in the past. But the “done” should be thought of in roughly the same way as we in mathematics talk about something “given” without an implicit act of ‘giving’: as a “done” without an implicit act of “doing”. The subject of CFR need not be something that took place or is the result of something that took place. The fact we are reasoning counterfactually about may be “timeless”: It is a fact that water expands when it freezes and turns into ice. Had it not been so, then life on Earth in the form we know it would have been impossible.² Or it may be a “future fact”: Xmas Day 2020 falls on a Monday. Had it not been so, then the shops would have been open on Xmas Eve that year.

The starting point in CFR put into a sentence is a question, not a proposition. But strictly speaking, propositions do not at all occur in CFR: there is proposing, not propositions. Not p but “ p ”, for instance: answering a question in the affirmative or the negative. And also: We are dealing with questions as a matter of asking them, of “questioning”, not with questions as sentences. Not $p?$, but “ $p?$ ”. CFR is *performed*. For a collective term I shall use the term ‘utterance’. To stress this I shall – at least in principle - put quotation marks around sentences when they are embedded/functioning in utterances. ‘Utterance’ is for the purpose the best term. But it is not quite appropriate, since the function of the utterance in CFR is not dependent on its being uttered. It may just be “uttered to you yourself”, just “thought”, or even better “lived” (see p. 5-6). The function of performing an utterance in CFR is to take a step towards answering the original counterfactual question. In our analysis we shall accordingly ask (p. 10-17): (1) What does a sequence of utterances (asking and answering) look like in case it really *is* a CFR i.e. make a series of well-formed/valid steps? And (p. 18-19): (2) What are the principles of performing it well? What are the criteria for good – reasonable – CFR?

2. Counterfactual Reasoning: An Example

Throughout this paper I shall make use of the following example:

On 8 November 1973 the Danish politician Erhard Jakobsen, due to a faulty fuel gage, ran out of gas on the way to parliament. He did not participate in the vote on the Assessment Act at 11:00. The government found itself in a minority and had to call an election. The outcome of this election changed the composition of parliament radically - a new party called Fremskridtspartiet (“The Future Party”) got in with 28 seats. Two other new parties got in, and two old parties who had been out of parliament since 1960 came in again. More than a third of the members were replaced.

I shall usually stick to the ‘would’-version rather than the ‘might’-version. Two of the reasons are: First, you often say ‘might’ rather than ‘would’ out of epistemic caution, not for reasons supposed to be found “in reality”. I want to avoid that ambiguity. Secondly – and as far as I understand it - modal logicians consider the would-version ‘ $B \square \rightarrow A$ ’ to be more important than the might-version ‘ $B \diamond \rightarrow A$ ’ and have devoted more work to it. (Lewis, 1973; Bennett, Jonathan, 2003)

² Had it constricted, then it would have fallen to the bottom of sea, lakes, rivers etc. Accordingly it would not have any isolating effect on the surface and there would have been bottom freezing even at temperatures only slightly under freezing point.

According to the established version he did it voluntarily. He was a member of the ruling party, but unhappy with the performance of its policy. However, for the sake of the example I shall believe the best in him and say that it was not his own choice. I shall also change the story in the way that he actually tried to arrive at the parliament in due time – but failed. Here, then, is a piece of counterfactual reasoning. (The arrows indicate – supposedly well-formed and reasonable - steps from one utterance to another).

- 1 He didn't reach the meeting at 11:00.
 1→2 Having run out of gas, is it still possible that he might have reached the meeting?
 2→3 Are there non-actualised conditions under which he might/would have reached the meeting?
 3 → 4 If, for instance, he had taken a taxi, might/would he then have reached the meeting?
 4 → 5 The distance from the place on the motor road where he ran out of gas to the parliament is 8 km, there was still one hour, and the usual speed of taxis is so and so etc. – so yes: In case he had taken a taxi, he might/would have reached the meeting.
 4 → 6 There had been three car accidents on the motor road and for all traffic there was a delay of one hour – so no: Even if he had taken a taxi, he might/would not have reached the meeting.
 5 → 7 But was it possible at all to catch a taxi on the motor road within a reasonable period of time that day?
 7 → 8 The meeting took place during the petrol strike, and save for taxis, public transport in general etc., car-driving was simply not allowed. But there were plenty of taxis all around, and he, being a member of the parliament, had primary access to taxis – so yes: It was possible for him to catch a taxi within a reasonable period of time on the motor road that day.
 7 → 9 All taxis were striking that day – so no: It was not possible to catch a taxi on the motor road within a reasonable time that day.
 5 → 10 But notice, that morning was not a usual one: there had been a huge car accident on the motor road, and traffic nearly stopped for one hour – so no: Even if the distance from the place on the motor road where he ran out of gas to the parliament is 8 km, *and* there was still one hour, *and* the usual speed of taxis is so and so etc.: in case he had taken a taxi, he would still *not* have reached the meeting.
 6 → 11 All right, but couldn't he simply have walked?
 Etc.

This sequence is exploring *conditions* under which something that is the case (“he didn't reach the meeting”) might/would not have been the case. Such conditions will appear as antecedents in a counterfactual conditional, and I shall call this kind of dialectics *antecedential counterfactual reasoning*.

Another option is exploring the *consequences* in case something that is the case (“he didn't reach the meeting”) had not been the case. This I shall call *consequential counterfactual reasoning*.

- 1 → 2' What would have happened in case he had reached the meeting? (What are the differences between what actually happened and what might/would have happened?)
 2' → 3' In case he had reached the meeting one more cup of coffee would have been sold in the café.
 2' → 4' In case he had reached the meeting, the government would not have found itself in a minority and would not have called an election.

- 4' → 5' In case the government had not found itself in a minority and had not called an election, the composition of the parliament would not have changed.
- 2' → 6' One more member of the government would also have voted against the law, so the government would still have found itself in a minority and would have called an election.

As a rule, the relevance of performing the antecedential version will *ceteris paribus* increase, the *smaller* the difference is between the factual scenario and the imagined counterfactual scenarios in the antecedent, whereas the relevance of performing the consequential version *ceteris paribus* will increase, the *bigger* the difference is between the factual scenario and the imagined counterfactual scenarios in the consequent. (“This small difference might have made such a big difference.”)³ However, for the sake of brevity I shall only analyse the antecedential version in this paper.

3. Interrogative Sentences

As already emphasized, in CFR questions are not sentences, but “questionings” i.e. they are something performed. However, in the performance of asking questions we make use of interrogative sentences, and for a start we will take look at such sentences. Basically they occur in two grammatical forms:

1) In the simplest case the questions stands in a simple relationship to a positive or negative proposition: Is the coat in the wardrobe? The coat is (or is not) in the wardrobe. Did he reach the meeting? He reached (or did not reach) the meeting. If we take the indicative to be the “natural” or “basic” form of a sentence (whatever that is supposed to mean), then the interrogative is derived from the indicative by a simple operation.⁴

2) From all propositions, positive or negative, you may derive a corresponding question. But not all questions are derivable from a proposition. In the second variant the questions start with an interrogative pronoun, typically a “wh-question” (when, where, why, who, what): Where is my coat? The coat is in the wardrobe. When did he arrive at the meeting? He arrived at the meeting at 2:00 p.m. (Or maybe: He did not arrive in time.) Here the question is rather related to the answer as a variable to a value of that variable. The question is so to speak a variable “calling” for a value: ‘my coat is on location x: identify the value of x!’

On questions, asking questions, performing or living with questions *as* questions: We as analysts have to stick to a verbal formulation, i.e. to sentences. But living with the question need not be derived from or even connected to a sentence. To perform a question means to be subsumed under a directive: ‘Find out whether *p* or $\neg p$!’, ‘Find the value of that variable *x*!’. The performance of a question, then, has two steps. You consider it *as* a question, and you try to find an answer – succeeding or failing. The utterance is rather an *expression* of the question as a piece of lived life. (Kühl, 1976)

Typically you consider a question as a question simply because it *is* a question. If I am going out and need my coat but cannot find it, then it *is* a question where my coat is, before – and even independently of whether – I *ask* it. It is a mode of a situation or of being in a

³ The dichotomy antecedential-consequential *counterfactual reasoning* should not be confused with the dichotomy backward-forward counterfactual *conditionals* as introduced by David Lewis (1973).

⁴ In Spanish even words and word order may be the same: ‘llegó a la reunión’ (= he reached the meeting) and ‘¿llegó a la reunión?’ (= did he reach the meeting?). Context, and in written language punctuation, in spoken language intonation will account for the difference.

situation so-and-so. Incidentally, the original meaning of ‘question’ (lat. ‘questio’) is not the linguistic one, but “agenda” (“matter”, “issue”). Cf. “To be or not to be, that is the question”: that is what matters. There is a picture painted by the Skagen painter Michael Ancher “Klarer de pynten?” (“Will he round the point?”). A few hardy older fishermen standing on the beach in the storm and looking seriously, more thoughtfully than timidly, out to sea. One of them is pointing. The title's "he" is not among them, and he is not seen in the picture. He is on the lake, on the way in, and in a moment he will perform a dangerous operation to come ashore. The title takes the form of a question: Will he succeed or fail? It may be a matter of life and death. No sentence in the indicative would have been more fitting!

The “answer” may be (a) a full answer, (b) a provisional answer, or (c) not really an answer but a new question, the answer of which will bring you closer to a full answer: Take again “where is my coat?” as performed by somebody in a specific situation. The answer may be: (a) “... hanging in the wardrobe.” The answer here is of a type and may even have a precision needed for going to where the coat is and taking it out (“it is *hanging* in the wardrobe”). (b) “... (somewhere) in my house”. The answer does not permit you to go directly to the place where the coat is, but narrows down the options. Or (c) “Where did you leave it the last time you wore it?”

4. Sentence Types in CFR

The steps in CFR are not sentences, but (what I have decided to call) ‘utterances’ (see p. 3). But wherever there is in this way an *utterance*, whether explicitly uttered by somebody or not, we (the analysts) have to formulate an *uttered sentence*. In the performance of CFR we apply a number of sentence types, all of which may be seen as logico-grammatical inflections of a proposition.

As already noted, on the sentence level a question may (in the simplest cases) be seen as an inflexion of a proposition:

- (a) He reached the meeting. (Or negative: He did not reach the meeting.)
- (b) Did he reach the meeting?

Another inflexion will introduce counterfactuality:

- (c) He might/could have reached the meeting. (Or negatively: He could not have reached the meeting.)

This is what I shall call a *simple counterfactual proposition*: It is possible that *p* – not *being* the case – *might* have been the case. Typified {1}.

Next:

- (d) Is it possible that he might have reached the meeting?

The simple counterfactual proposition has here been further inflected to a *simple counterfactual question*: Is it possible that *p* – not *being* the case – *might* have been the case?

Now, corresponding to (a)-(d) we get the conditional types:

- (a') Since he reached the meeting, he took a taxi.

- (b') Is it the case, that he reached the meeting, since he took a taxi?
 (c') If he had taken a taxi, he might/would have reached the meeting.

The latter is what I shall call a *simple counterfactual conditional proposition*: If q had been the case, then p might/would have been the case. Typified {3}.

The counterfactual conditional proposition may be further inflected to:

- (d') If he had taken a taxi, would he then have reached the meeting?

This I shall call a *simple counterfactual conditional question*: Is it true that if q had been the case, then p might/would have been the case?

However, from (c') we may also derive this form:

- (c'') There are (or in the negative: there are not) currently unfulfilled conditions, such that had they been fulfilled, then he would have reached the meeting.

which I shall call an *existential counterfactual conditional proposition*: There is an x such that had x been the case, then p might/would have been the case. Typified {2}.

And correspondingly, from (d') we may derive:

- (d'') Are there (currently unfulfilled) conditions such that had they been fulfilled, then he would have reached the meeting?

This I shall call an *existential counterfactual conditional question*. It is fairly close to a counterfactual wh-question: What are the conditions – if there are any – under which he might/would have reached the meeting? However, I shall not make use of that form.

5. A Closer Look at Counterfactual Sentences

We have introduced three types of counterfactual propositions. There is a simple counterfactual proposition {1}, a simple counterfactual conditional proposition {3}, and an existential counterfactual conditional proposition {2}. It will later be demonstrated that all steps used in CFR-utterances may be derived from these three types of propositions, occurring in modes of assertion (asserting that something is the case), question (asking whether or not it is the case), and rejection (rejecting that it is the case). They will then be applied in the order {1}-{2}-{3}, but since the existential counterfactual must be constructed on the basis of the simple counterfactual, they have here been introduced in the order {1}-{3}-{2}, and we shall now take a closer look at them in that order.

One thing will be given as a starting point:

- (i) It is *not* the case that A. (= It is a fact that not-A.)⁵
 For instance: He did not reach the meeting.

⁵ The reason why an expression in the form 'not-A' rather than an expression in the form 'A' has been taken as a starting point is a matter of simplicity in the presentation. This will later become evident.

Later on this may be added, still in the mode ‘given’:

(ii) Further premise: T

For instance: There were plenty of taxis all around, and he, being a member of the parliament, had primary access. Or: There was an enormous traffic jam on the motor road.

Simple counterfactual

$$\diamond A = \text{type } \{1\}$$

‘It might/could have been the case that A’.

(= ‘It is possible that A – as *is* not the case – *might/could* have been the case.’)

For instance: It might/could have been the case that he reached the meeting at 2:00 p.m.

Simple counterfactual conditional

$$B > A = \text{type } \{3\}^6$$

‘If B had been the case, then A would have been the case’.

For instance: If he had taken a cab, then he would have reached the meeting.

Comments and specifications:

(i) It was given that A is not the case. Insofar as the formula ‘ $B > A$ ’ is well-formed, B is not the case either.

(ii) In daily usage expressions having the form ‘If B had been the case, then A would have been the case’ may have what I shall call a *weak* and a *strong* reading. Take the example (a) ‘If Mr. Jakobsen had taken a cab, then he would have reached the meeting’. In the *weak* reading you just imagine a scenario in which a taxi successfully is bringing Mr. Jakobsen from the place on the motor road where he took it to the meeting in the parliament. But maybe there were no taxis to get hold of, say, within a reasonable time? If the question was ‘Could Mr. Jakobsen have reached the meeting?’, then (a) won’t do for an affirmative answer unless there is added (b) ‘It would have been possible to get hold of a taxi.’

But you may also consider the process of getting access to the taxi as being embedded in the concept of “taking a taxi”. The project of taking a taxi starts out with the project of getting hold of a taxi. This represents the *strong* reading.

In some contexts we are obviously – maybe even explicitly – dealing with the weak reading: “All infected people in the village Guago in Guinea would have been saved from ebola if they had been vaccinated with Zmapp.” In other cases we are clearly dealing with the strong reading: “If he had extinguished the candle there would have been no fire.” Here the possibility that he could have put out the candle is implicitly assumed. Or take the question: “Would Peter have survived the disease, had he taken the medicine M?” Definitely an important question for the purpose of clarifying the quality and potency of the medicine M. But if it is meant to function as a step towards answering the question ‘Could Peter possibly have survived the disease?’, then the first question must be supplemented by another question: “Could he possibly have taken that medicine?” For instance, was it possible at all to get hold of it?

⁶ For the sake of simplicity I shall apply the notation ‘ $B > A$ ’ instead of the authorised ‘ $B \square \rightarrow A$ ’.

Definition: In our formalisation the expression ‘ $B > A$ ’ represents the *weak reading* of the counterfactual conditional (NB unlike possible world semantics), whereas the *strong* reading will be expressed ‘ $(B > A) \ \& \ \diamond B$ ’.

Simple counterfactual conditional on premises

T: $B > A$ = type {3’}

‘Under the explicit, specific *actual/factual* condition T:

If B had been the case, then A would have been the case.’

There was no traffic jam, the distance was 8 km, there was still one hour, and the usual speed of taxis is so and so etc. (so): If he had taken a cab, then he would have reached the meeting.

Comments and specifications:

In practice, all CFR is performed “on premises”. In our main example above lots of premises are introduced in the story told before the question is asked. What should/might count as “explicit” or “implicit” is not clear, and maybe will never be clear, but the very distinction cannot be skipped. Since there are always such premises, we shall in our analysis only make explicit premises that are *added* during the sequence.

Existential counterfactual conditional

$\exists x(x > A)$ = type {2}

‘There is an x such that if x had been the case then A would have been the case.’

There is a condition (or a set of conditions) such that had it (they) been fulfilled then he would have reached the meeting.]

There are conditions under which he would have reached the meeting.

Comments and specifications:

(i) {2} is constructed on the basis of {3} and was accordingly presented before {3}. But from a logical point of view {2} is weaker than {3}. And that will be decisive when it comes to developing the generative principles of CFR (p. 9f).

(ii) Similar to {3}: It was given that A is not the case. Insofar as the formula ‘ $x > A$ ’ is well-formed, x is not the case either.

(iii) Similar to {3} the expression represents the *weak* reading of the counterfactual conditional. It is *not* implicit/assumed that x is possible.

Existential counterfactual conditional on premises

T: $\exists x(x > A)$ = type {2’}

‘Under the explicit specific *actual* condition T:

There is an x such that if x had been the case, then A would have been the case.’

There was no traffic jam, the distance was 8 km, there was still one hour, and the usual speed of taxis is so and so etc. (so): There are (further) conditions under which he would have reached the meeting.

Each of the expressions {1-3} may occur in three modes of utterances: in *asserting* “{1!}”, “{2!}”, “{3!}”, in *asking* a question “{1?}”, “{2?}”, “{3?}”, and in *rejecting* “{1-!}”, “{2-!}”, “{3-!}”.⁷ All these nine expressions should in principle always occur in quotation marks to remind us that they are somebody’s “thought”, but since they are there all the time, we may rather leave them implicit.⁸

6. The *fcnc*-clause (“*ficonoci*”) in the counterfactual conditional

When counterfactual conditionals are performed in CFR, two things are implicitly assumed:

First, other things as far as possible remain unchanged, be it conditions that you know of or conditions that you don’t know of. Mr. Jakobsen taking a taxi, and Mr Jakobsen reaching the meeting should *in principle* be the only difference between the factual scenario and the imagined, counterfactual scenario.

And *second*, it is implicitly assumed that things in general behave in the “normal” way, i.e. as they usually do, as they are meant to do, etc. In Mr. Jakobsen’s case we are dealing with normal taxis, or with normal taxis on a normal day in that town, or on a normal Thursday, or on a normal Thursday morning, etc.

So in the performance of a conditional counterfactual utterance there is always a clause: *If* B/x had been the case, *then* - *fixing the context* (be it the part of it that I know of or the part that I do not know of), *and* assuming the presence of (what I take to be) *normal circumstances* – A would have been the case. I notate this clause ‘[*fcnc*]’ as a short-hand for “fix the contextnormalcircumstances and read it “*ficonoci*” (cf. Fibonacci!):

$$B \text{ [fcnc]} > A$$

But since it is operative throughout this presentation I leave it out of the notation, in the same way as I do with the quotation marks around utterances.

Of course, what exactly is incorporated in the clause - how much I know about the context and what I incorporate in the “assumed normal circumstances”- is deemed to be imprecise and vague. My claim is that we *have* to fix the context and assume normal circumstances when we perform this kind of practical reasoning.

⁷ I write “(B > A)-!” and not “-(B > -A)!” to remind that the utterance may be an *answer*. And it contributes to notational simplicity and clarity. Without further specification it would not be clear whether “-(B > -A)!” should be read “-(B > -A)!” (i.e. not asserting that (B > -A)) is true, or “-((B > -A)!)” (i.e. asserting that (B > -A) is not true).

⁸ Quotation marks as double inverted commas mark the sentence as representing an utterance. Single inverted commas, following the convention, mark the sentence as being mentioned, not used.

7. Steps in CFR. Elaborating the Complete Conclusive Sequence

My task in the following will be to identify principles for well-formed *steps* i.e. steps that – according to their form – “make sense”, are reasonable, as potentially progressing towards a conclusion in CFR. In this way we talk about ‘steps’ meaning steps of reasoning *taken* from one utterance to another. However, the term ‘step’ is ambiguous. The *positions* from/to which steps in this sense are taken may also be termed ‘steps’. This will hardly lead to any confusion. We speak in the same way about staircases: We take steps on the steps i.e. from step to step.

Reasoning means a *sequence* of “thoughts”. Reasoning is performed. And I shall try to present a generative system of well-formed steps in antecedential CFR, first (p. 9-14) with two arguments A and B, later (p. 15-16) with more than two arguments. Finally (p. 17-18) I will briefly introduce some principles according to which some well-formed steps are more reasonable than others.

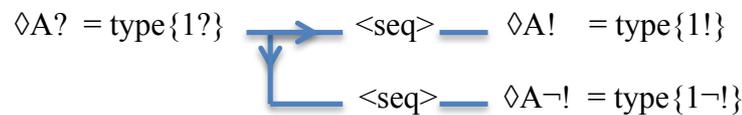


Fig.1

Fig. 1 is called a *CCS complete conclusive sequence*. It starts out with a question in type $\{1\}$ and, following the arrows through a sequence, it ends with either a positive or a negative answer to that question, i.e. an assertion or a rejection in type $\{1\}$. What really matters is, of course, what may happen in between.

The path system on p. 20 presents all possible ways to perform a CCS sequence with just two counterfactual arguments A and B. Moving in the scheme means reasoning. Read the scheme in the following way:

- (i) Steps in the horizontal direction are always left-right.
- (ii) Steps up-down or down-up are either marked with an arrow, or they are prolongations of an already given horizontal route.
- (iii) Where two possible routes start from the same point this will be marked with arrows.
- (iv) Successive steps in the scheme may be performed successively in that order, or in the same moment i.e. as one move, but never in reverse order.

And again, remember that the expressions are not sentences, but the *performance of utterances* in modes of asserting (positive and negative), asking and rejecting. Quotation marks are still implicit.

Arrows indicate steps of reasoning.⁹

In the following I shall put comments on each move in the scheme. Each formula is indexed with a minuscule in superscript to make it easier to find the corresponding location in the scheme. Rightmost on each line the type of the utterance is made explicit: simple counterfactual, existential counterfactual conditional, or simple counterfactual conditional, and each occurring in the modes of asserting, asking, or rejecting.

⁹ Some of the steps make valid inferences, occurring insofar as they are drawn, not insofar as they are valid. Others are just reasonable.

Given: $A \neg$
 Question: $\Diamond A?^a$ {1?}

In order that “ $\Diamond A?$ ” may become a question to be reasoned about, one step has to be taken initially:

1. $\Diamond A?^a \rightarrow \exists x(x > A)?^b$ {1?} \rightarrow {2?}

If something that isn't the case *might* have been the case, then something else that isn't the case *would* have been the case too. If A *might* have been the case, then there are conditions under which – when fulfilled – A *would* have been the case. Maybe you can have counterfactual truths without counterfactual conditions though I find it difficult to make sense of it. But if we don't accept as an axiom ‘ $\Diamond A \supset \exists x(x > A)$ ’ there cannot be counterfactual reasoning.

Let each of the examples (i), (ii), (iii) and (iv) be a part of the story as told on page 3-4.

(i) He ran out of petrol at 11:05.

Now, there are by mathematical necessity no conditions under which he would have reached the meeting at 11:00. If something is true in mathematics, it couldn't possibly have been false. Conclusion: {2 \neg !} and further {1 \neg !}.

(ii) He ran out of petrol at 10:50.

Somewhere on a scale between practical and physical necessity there cannot be any conditions under which A might have been the case. For all practical matters it may be considered unreasonable to start reasoning. Again we get conclusion {2 \neg !} and further {1 \neg !}.

“Conclusion” here means something *performed* by somebody – let's call him “the agent”. In (a) the agent in the performance of CFR immediately concludes negatively and doesn't start out reasoning, because he has realised that the counterfactual can't be true. In (b) he makes the same conclusion and doesn't start out reasoning because he can't see that anything less than a miracle would do.

(iii) He ran out of petrol at 9:30.

Now the question {2?} has become a “good question”, and counterfactual reasoning begins. The agent is not answering the question yes rather than no. He is not answering the question at all. He proceeds i.e. he lets the question *be* i.e. function as a question.

(iv) He ran out of petrol the day before.

Now the question in a sense is not a real question any longer. (The expression ‘reaching a meeting’ makes no sense, where there is plenty of time i.e. when no effort to reach it is called for.) Rather another question will show up: Why didn't he show up at the meeting?

Go back to 1. Is it true, then, that there are instances of x ? Then you must find one. Is B such an instance?

2. $\exists x(x > A)?^b \rightarrow (B > A)?^c$ {2?} \rightarrow {3?}

Of course you may jump directly from {1?} to {3?}. Most often we do so, but not always. And in case we separate them, then – for sheer logical reasons – the inverse order is impossible.

Logic cannot help us in making suggestions to such conditions, but it is possible to develop a certain methodology for searching and trying out antecedents. More about that p. 18-19.

From “(B > A)?” there are three options for a next step. One of them consists in answering the question in the *negative*:

$$3. \quad (B > A)?^c \rightarrow (B > A)\neg!^d \quad \{3?\} \rightarrow \{3\neg!\}$$

Even if he had taken a taxi he would not have reached the meeting. But from “(B > A)\neg!” you cannot move further on towards any conclusion. You cannot *infer* that “ $\exists x(x > A)\neg!$ ” and accordingly “ $\neg A!$ ”), but you may *stop* reasoning, definitely or provisionally:

$$4. \quad (B > A)\neg!^d \rightarrow [\text{Stop}]^e$$

Another move from “(B > A)\neg!”^d would be to try to find/introduce a new antecedent to be substituted for B:

$$5. \quad (B > A)\neg!^d \rightarrow \exists x(x \neq B \ \& \ x > A)?^f \quad \{3\neg!\} \rightarrow \{2?\}$$

The taxi would not do. Could there be other possibilities? Would Mr. Jakobsen e.g. have reached the meeting in case he had used his bike in the trunk of his car? Or maybe simply walked? (Inserting a specific value for x amounts to introducing a third argument. I shall return to that option later.)

The second option for a next step from “(B > A)?”^c consists in answering the question *affirmatively*: “(B > A)!^g” (“Yes, if he had taken a cab he would have reached the meeting.”)

$$6. \quad (B > A)?^c \rightarrow (B > A)!^g \quad \{3?\} \rightarrow \{3!\}$$

However, as already mentioned it is possible to ask whether A would have been the case if B had been the case, *without* relating to the question whether B is possible at all. That constitutes the *weak* reading of ‘If B had been the case, then A would have been the case’, and is the reading intended in our formalisation. So from “(B > A)!^g” we cannot infer/conclude: “ $\diamond A!$ ” We need further to take this step (considering this question):

$$7. \quad (B > A)!^g \rightarrow \diamond B?^h \quad \{3!\} \rightarrow \{1?\}$$

The third option for a next step from “(B > A)?” consists in *suspending* the question whether he would have reached the meeting, had he made the trip by taxi. Maybe there was no taxi to get hold of? Check that first:

$$8. \quad (B > A)?^c \rightarrow \diamond B?^1 \quad \{3?\} \rightarrow \{1?\}$$

From “ $\diamond B?$ ” the reasoning proceeds on exactly the same conditions i.e. with formally the same possible steps as “ $\diamond A?$ ”, since they are both of type {1?}, and there is opened up a new counterfactual sequence. That sequence may be brought to a negative conclusion (at least a provisional conclusion): “ $\diamond B\neg!$ ”

$$9. \quad \diamond B?^h \langle seq \rangle \diamond B\neg!^i$$

And (similar to step no. 4 above) you may stop reasoning, definitely or provisionally.

Or (similar to step no. 5) you may ask for another antecedent to be substituted for B:

$$10. \quad \diamond B\neg!^i \rightarrow \exists x(x \neq B \ \& \ x > A)?^f \quad \{1\neg!\} \rightarrow \{2?\}$$

Or the sequence starting out from “ $\diamond B?$ ” may be brought to a positive conclusion. (It runs slightly differently depending on whether we move from $\diamond B?^h$ or $\diamond B?^1$. Let it here be $\diamond B?^h$.):

$$11. \quad \diamond B?^h \langle seq \rangle \diamond B!^j \quad \{1?\} \langle seq \rangle \{1!\}$$

from which a further step may be taken to the affirmative answer of the original question:

12. $\diamond B!^j \rightarrow \diamond A!^k$ $\{1!\} \rightarrow \{1!\}$

Since ‘ $(B > A) \& \diamond B \Rightarrow \diamond A$ ’ is a logical truth, step no.12 amounts to the performance of a valid inference. (All steps carry with them, as premises, all propositions asserted on the route up to the point from where the actual step is taken.)

As mentioned earlier: The performance of a question has two moments: You *identify/consider* the question as being a question, and you *try to find an answer* – succeeding or failing. As regards the questions “ $(B > A)?$ ” and “ $\diamond B?$ ” you may identify/consider them at the same time (corresponding to the strong reading of the counterfactual conditional) or in the noted order, but you cannot identify them in the reverse order. That would imply that you first make a list of possible facts, and then insert them one by one as the antecedent in “ $\exists x(x > A)?$ ”.¹⁰ On the other hand, the order of further performance of searching for an answer may be arbitrary i.e. you yourself may choose it: In case one of the two is highly improbable there is no reason at the first place to elaborate the other. In case a helicopter is suggested, the problem is not whether Mr. Jakobsen would have reached the meeting flying by it, but whether he would have got hold of it. In case of walking to the parliament it is most likely not a question whether or not he actually could make that walk, but whether he would have made it in due time. In case you search for and find an affirmative answer of the question “ $(B > A)?$ ” before you start out with the question “ $\diamond B?$ ”, then your route will be the one indexed c-g-h-j-k. Alternatively, in case you search for and find an affirmative answer of the question “ $\diamond B?$ ” before you start out with the question “ $(B > A)?$ ”, then your route will be indexed c-l-m-n-k.

Summary: As we have seen there are three basic expressions in CFR: the simple counterfactual ‘ $\diamond A$ ’ i.e. type {1}, the simple counterfactual conditional ‘ $(B > A)$ ’ i.e. type {3}, and the existential counterfactual conditional, ‘ $\exists x(x > A)$ ’ i.e. type {2}. Each of the expressions may occur in three forms of utterances: in *asserting* “{1!}”, “{2!}”, “{3!}”, in *asking* a question “{1?}”, “{2?}”, “{3?}”, and in *rejecting* “{1-!}”, “{2-!}”, “{3-!}”. If we follow the routes in the scheme [stepwise or in leap] it turns out that only the transitions in fig. 2 have been performed. But even when we introduce more arguments (see the next section), the steps in fig. 2 will return again and again, and (with slight modifications) they will be the only possible steps.

¹⁰ Cf. Hempel’s *Raven Paradox* and the two practical strategies for confirming an A-judgement.

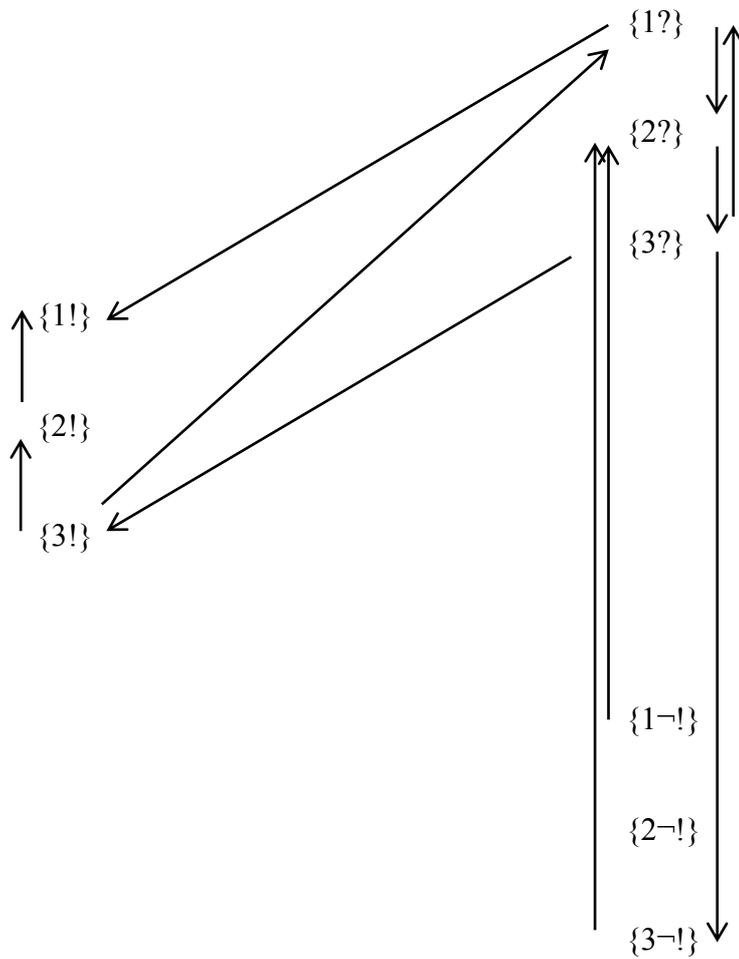


Fig. 2

Steps according to the scheme of elementary counterfactual reasoning at p. 20.

CFR with three or more arguments

So far only CFR with two arguments at a time (here: A and B) has been analysed. We shall now take a look at steps in the dialectics where a third argument is introduced, thereby opening the procedure for adding as many arguments as you like. Once again look at the scheme p. 20, and take $(B > A) \neg!$ ^d or $\diamond B \neg!$ ⁱ as the starting point. From here, as we have seen, further reasoning on the question $\diamond A?$ may simply stop or – on a general level - ask for an alternative antecedent $\exists x(x \neq B \ \& \ x > A)?$ ^f. Now one option may now consist in actually trying out a *specific alternative counterfactual antecedent*. There are two ways:

$$13. \quad \exists x(x \neq B \ \& \ x > A)?^f \rightarrow (C > A)?^o \quad \{2?\} \rightarrow \{3?\}$$

Say it has been rejected that ‘If he had taken the medicine M1, he would have been cured’, and consider ‘If he had taken the medicine M2, he would have been cured’. C has been *substituted* for B.

Or you may try *adding* a specific counterfactual antecedent to B:

$$14. \quad \exists x(x \neq B \ \& \ (x \wedge B) > A)?^p \rightarrow ((C \wedge B) > A)?^q \quad \{2?\} \rightarrow \{3?\}$$

Say it is not true that ‘If he had taken the medicine M1, he would have been cured’. Still it may be true that ‘If he had taken the medicines M1 *and* M2, he would have been cured’. Since capital letters for antecedents should indicate conditions that may or should be thought of as *sets* of conditions, then step 14 from a formal point of view is just a variant of step 13. Step 14, however, deserves a point of its own because what it actually does is to increase the tolerance for the principle *Fix the Context* (see p. 10) - “fix the context *as much as you can*” - in adding one more argument to the antecedent of the counterfactual conditional.

Anyway, the same sort of motivation that made you start reasoning counterfactually is operative whether you carry on substituting or adding antecedents. And in both cases counterfactual reasoning will just proceed from a type $\{3?\}$ position in the path system.

However, from the starting points $(B > A) \neg!$ ^d or $\diamond B \neg!$ ⁱ further reasoning may also proceed in quite a different way by *adding a factual premise to the conditional counterfactual as a whole*. Instead of modifying the *counterfactuality*, you may add more “*factuality*” in the specific performances of counterfactual reasoning. This option is more complicated, but also more important already from a merely quantitative point of view. Say e.g. you are reminded or come to know that the meeting in the parliament in which Mr. Jakobsen was meant to participate took place during the petrol strike, and save for taxis, public transport in general etc., car-driving was simply not allowed. And there were plenty of taxis. Let ‘T¹’ mean something like “8 November 1973 was a day with privileged conditions for taxi traffic in Copenhagen”, and you may carry on reasoning this way:

$$15. \quad (B > A) \neg!$$
^d or $\diamond B \neg!$ ⁱ $\rightarrow T^1: (B > A)?^r \quad \{3 \neg!\} \text{ or } \{1 \neg!\} \rightarrow \{3'?\}$

What you have got here was introduced page 8 as a simple counterfactual conditional *on premises*, here in the mode of a question, and you have moved to a position that is formally identical to $(B > A)?^c$. The addition of an explicit factual premise will make no change to the structure of well-formed steps in the path system. Adding ‘T’ as a prefix to each of the steps you may proceed within the path system.

Obviously the addition of a factual premise may also interfere with reasoning on its way towards an affirmative conclusion. Say you have affirmed $(B > A)!$ ^g and/or $\diamond B!$ ^h. Then it comes to your mind, or you are told, that on that day there was an enormous traffic jam on the motor road. Say ‘T²’ means “there was a delay on the motor road of more than one hour”. Reasonable reasoning may now consist in taking this step:

16. $(B > A)!^g$ and/or $\diamond B!^h \rightarrow T^2: (B > A)?^r$ $\{3!\}$ and/or $\{1!\} \rightarrow \{3'?\}$

Like step 15 you proceed your considerations asking a question in terms of a counterfactual conditional on premises and thereby to a position that is formally identical to $(B > A)?^d$.

In case the sequence leads you to answer the question in 16 negatively (i.e. $T^2: (B > A) \neg!$ ^d) you may either stop reasoning: [stop]^e. Or you may carry on asking: $T^2: \exists x(x \neq B \ \& \ x > A)?^f$, or $T^2: \exists x(x \neq B \ \& \ (x \wedge B) > A)?^p$. Or you may add yet another factual premise: $(T^2 \wedge T^3: B > A)?^f$.¹¹

Since the only difference between these positions and the positions indexed with the same minuscule in superscript indexes in the original CFR-scheme consists in the content of the added premises, further reasoning simply amounts to new trips through the path system making something explicit that was implicit (or even wrongly assumed), producing still higher evidence for a conclusive answer of the original question ‘ $\diamond A?$ ’.

A third possibility is, of course, that the added premise makes no difference as to the result of your reasoning. For instance, even in a traffic jam he would have reached the meeting by taking a taxi i.e. ‘ $T^2: (B > A)!$ ’. Or even on the premise that there were privileged conditions for taxi traffic, he would *not* have reached the meeting. But the evidence for the conclusion has increased.

Now, maybe you did explicitly assume that there was no traffic jam, and now you have been corrected. But much more likely you did not make any assumptions at all concerning road navigability. When thinking of Jakobsen’s possible taxi-trip you simply and implicitly assumed that his car driving would have taken place in the way car driving was meant to work and used to work. You adhere to some *Principle of Normality* (see p. 10). Actually there are three such principles, or (better) the principle works on three different levels:

(A) The (in a Kantian sense) *transcendental* version/level: (a) The world is so that there is something we may call “the way things used to be”, according to their nature, according to rules of occurrence, principles of causation etc., and (b) our assumptions as to the way things used to be are usually valid/true. The validity of this principle is an indispensable conditional for human/practical life, language etc. Since life, language etc. in fact exist, the principle must be valid – whether arranged by God or not.

(B) From the transcendental version we derive a *general* version: I assume, explicitly or implicitly, this or that generality to be a normality, and usually it turns out that it actually is so. (I used to find my key where I left it, cab-trips from this to that place usually take less than one hour, etc.)

(C) And finally I assume the conditions for the *specific* case to be instances of a general version of the principle, and usually it turns out that it actually is so. (The conditions for Mr. Jakobsen’s possible cab driving on the motor road November 8, 1973 is thought to be an instance of the normality that cab-trips from this to that place usually take less than one hour.)

Since the world is infinitely complicated, and our knowledge is limited, we cannot incorporate more than a small part of possibly relevant arguments into our reasoning, be it counterfactual or not. Still new explicit factualities/premises may enter again and again as you get new information, remember things etc. But most importantly, and typically, they enter when implicit assumptions are made explicit – and *negated*: You discover, or you are reminded about, your implicit assumptions about something being a fact, only to realise that it *is* not a fact.¹²

¹¹ It will of course make no difference whether this position is developed from the question in 14 rather than in 15.

¹² If somebody asked you, whether you presupposed that there were no traffic jam on that day, you would answer: “Oh yes, certainly, I never thought about it, I just took it for granted – but now you tell me that I was wrong!”

Criteria for reasonable choices of antecedents or reasonably adding explicit factual premises

The main topic in this article has been well-formed steps in CFR and the attempt to develop a system of such steps. Finally I shall sketch some criteria for making this or that step not merely well-formed, but also more or less *reasonable* or *reasonably preferable* to another step. Obviously, such criteria will merely hold *ceteris paribus* and in practical performance of CFR they may often conflict. But they do represent the ways (or most of the ways) we actually *can* make preferences.

The first group of criteria will include considerations on modifying the antecedent:

(a) Antecedents that are assumed to be *more* likely to occur should *ceteris paribus* be considered *before* antecedents that are *less* likely to occur. (The opportunity of being brought to the parliament by taxi is more likely to occur than the opportunity of being brought there by helicopter.)

(b) Antecedents that are assumed to be *more* likely to make the counterfactual implication true should *ceteris paribus* be considered *before* antecedents that are *less* likely to do so. (It is more likely that he would have reached the meeting had he been picked up by a helicopter rather than by a cab.)

(c) Antecedents for which (a) and/or (b) are *easier to decide*, should *ceteris paribus* be considered *before* antecedents that are *more* difficult to decide. (One never knows about taxi and taxi drivers in that town. It is easier to decide what would have happened had he done it on foot.)

(d) A and B should preferably be the only counterfactual elements (cf. again the principle Fix the Context). But this, of course, is impossible. A world in which Mr. Jakobsen took a cab and reached the meeting (which didn't happen) would also be a world in which another person did *not* take that cab (which actually happened). Maybe this would make no further difference in the present piece of CFR. Or maybe it would? If Mr Jakobsen had taken a cab, there might have been another person who did not take that cab and maybe not reached that meeting etc. But every time way we add new propositions to the antecedent of the counterfactual, we make the reasoning less and less "relevant". Considerations on counterfactual conditionals with a shorter list of counterfactual elements in the antecedent should be preferred to considerations on counterfactual conditionals with a longer list. The longer the list of unfixed context, the less reasonable is the reasoning.

The second group of criteria will include considerations on reasonably adding (more) factual premises:

(e) First there is a simple principle: The *longer* the list of added factials, the bigger the evidence *ceteris paribus*. The more explicit premises the better. That holds for all sorts of reasoning.

(f) If the fact that T_m depends on [presupposes] the fact that T_n , then add T_m and not T_n . (Add "He had money to pay" and not "There is a monetary system". And add "There were no, or few, or many cabs to get hold of" and not "There was no earthquake yesterday".) There is nothing wrong in introducing T_n as well as not T_m but it is superfluous. And it is less reasonable to do something superfluous than not to do it.

(g) However, in the practical performance of CFR another principle is more important. The original question in CFR took the form " $\diamond A$ ", so we are still searching for arguments that, as antecedent in the assertion " $x > A$ ", will make that assertion true. But the factual premises $T_1 \dots T_n$ which we may *add* (" $T_1 \dots T_n: x > A$ ") should not be selected according to whether they draw towards a *positive* conclusion ("a cab was standing right next to his own car", "the streets were empty" – so yes: "if he had so no: even if he had taken a cab, he would not have reached the meeting") or a *negative* conclusion ("only 5 cabs were in the streets that afternoon", "there was traffic jam"- so no: even if he had taken a cab he would not have reached the meeting). Factual premises should be searched for, and prioritised, not according to *which* direction they draw but how *strong* they draw.

(f) This leads to the idea in all CFR of searching for “*knock-out arguments*”. A knock-out argument is a *factuality* that, if added to (the latest version of) the counterfactual implication, but without the addition of further *counterfactualities*, is considered to make the counterfactual implication (i) definitely undeniable (“all taxi-drivers went on strike”) or (ii) conclusively assertable (“the chairman of the parliament offered Mr. Jakobsen to send a special cab and to adjourn the meeting until his arrival”).

(g) Corresponding to the epistemic principle (c) for modifying the antecedent there will be an epistemic principle (g) for the choice of factual truths: Factualities for which there is more evidence should be preferred to factualities with less evidence.¹³

REFERENCES

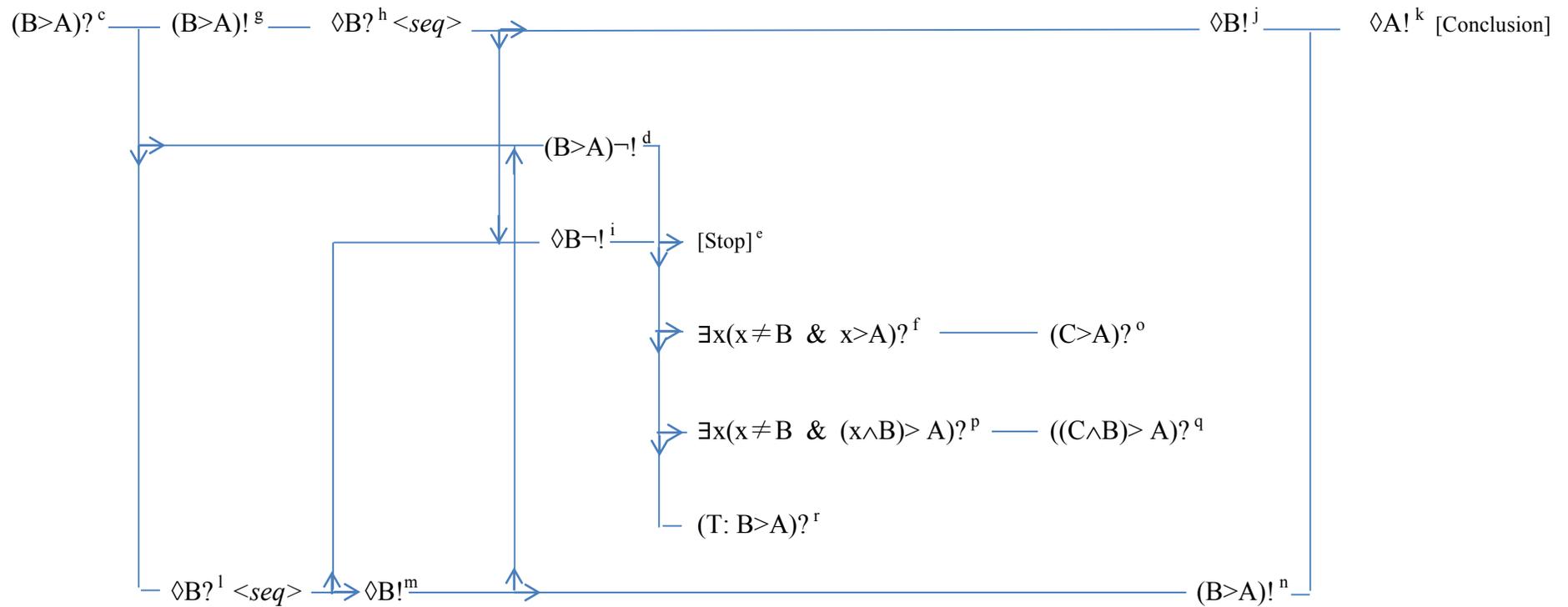
- Ginsberg, M. L. (1986). "Counterfactuals". *Artificial Intelligence*, 30: 35-79
- Goodman, N. (1947). "The Problem of Counterfactual Conditionals", *The Journal of Philosophy*, Vol.44, No.5, 113-128.
- Hempel, C.G. (1945). “Studies in the Logic of Confirmation”. *Mind*, 54: 1-26.
- Lewis, D. *Counterfactuals*, Oxford: Blackwell (1973).
- Nozick, R. *Philosophical Explanations*, Oxford: Oxford University Press (1981)
- Bennett, J. *A Philosophical Guide to Conditionals*. Oxford University Press (2003).
- Bennett, J. (2001). “On Forward and Backward Counterfactual Conditionals”. In Gerhard Preyer (ed.), *Reality and Humean Supervenience: Essays on the Philosophy of David Lewis*. Rowman and Littlefield 177-202.
- Pearl, J. *Causality: Models, Reasoning, and Inference*. Cambridge University Press (2000)

¹³ The principles (c) and (g) are simply variants of a more general principle for all reasoning whatsoever: Prefer reasoning with arguments with higher evidence to reasoning with arguments with less evidence.

Elementary Counterfactual Reasoning (ECR).

Scheme

The use of two arguments and the addition of a third one.

Given: $\neg A$ \Downarrow
 $\Diamond A?^a$ \Downarrow
 $\exists x(x > A)?^b$ 

Explanation for the ECR Scheme above.

A *conclusion* [the act of concluding] means something performed. And it means the outcome of the performance i.e. it is *drawn*. An answer has been uttered to the original question. Reasoning has come to an end.

A *stop* is also performed. It has no outcome. Reasoning may have come to a part of an answer, but does not proceed. In the scheme: B won't do for a counterfactual antecedent. Maybe there are alternatives, maybe not.

Capital letters for antecedents B, C, etc. should indicate conditions that actually are *sets* of conditions. So B may be different from C and yet be a part of it.