FRANKFURT-STYLE COUNTEREXAMPLES TO INFLUENCE THEORY OF CAUSATION

MILAN Z. JOVANOVIĆ[1]

Abstract: Two prominent counterexamples to Lewis's Influence theory of causation (Schaffer 2001, Hall 2004) happen to be structurally very similar to so-called Frankfurt cases. This should come as a surprise since Lewis explicitly addresses Frankfurt cases while formulating his theory, and claims that theory deals with cases like that successfully (Lewis 2000). Hence, a good question to ask is – whether these two counterexamples are indeed plausible and valid objections despite their structural similarity to the Frankfurt cases. In this paper, I offer an analysis of two mentioned counterexamples in order to answer this question. On the one hand, in agreeing with Noordhof (Noordhof 2001), I will try to show that Schaffer's counterexample can indeed be accommodated and explained by the Influence theory. On the other hand, I will try to maintain that, even if we accept Lewis's premises, the counterexample offered by Ned Hall is still plausible – due to a certain feature that differentiates it from both: Frankfurt cases and Schaffer's counterexample. While the latter two are cases of early preemption, Hall's Smart Rock scenario doesn't exhibit that – from the perspective of Lewis's theory – convenient causal pattern in which we can find stepwise influence (which is enough for the theory to get these cases right). This result, as I believe, shows why we should regard Hall's counterexample as a better and more plausible argument (than Schaffer's counterexample) against the Influence theory.

Key words: Influence theory of causation, preemption, Frankfurt cases, ancestralization, counterfactuals.

[1] University of Niš, Faculty of Philosophy, milan.jovanovic@filfak.ni.ac.rs

Introduction*

Faced with the seemingly unsolvable problem of late preemption, David Lewis had abandoned his original counterfactual analysis (Lewis 1973, 1986c) and formulated a new theory of causation (Lewis 2000). This new theory – the Influence theory of causation (ITC) – has also been strongly criticized. Interestingly, the counterexamples (CEs) offered against ITC typically involve, again, cases of preemption.

An even more interesting fact is that two prominent CEs offered against the theory (Hall 2004, Schaffer 2000) bear a striking structural resemblance to well-known Frankfurt cases (Frankfurt 1969). This should come as a surprise since Lewis explicitly addresses Frankfurt cases (FCs) while formulating his theory (Lewis 2000) and claims that the theory deals with cases like that successfully.

So, the question is whether these CEs are plausible, and Lewis is simply wrong about FCs, or, conversely, Lewis is right that ITC can explain Frankfurt-style examples, while Schaffer's and Hall's CEs are flawed? Or there is even some more refined resolution, questioning perhaps the similarity between offered CEs, or between them and FC?

Those are the questions that I am going to pursue in this paper. Before turning to them, I will present ITC and show how FCs – according to Lewis – can be accommodated within ITC. After presenting and analyzing given CEs, I will briefly point out the results and summarize the paper.

Influence theory of causation

Belonging to the same theoretical background as its predecessor – Lewis's counterfactual theory of causation (CTC) – ITC is an analysis of the singular causal statements about actual events (and omissions).[2] It is

* I am very grateful to Mihai Rusu (USAMV Cluj-Napoca/Babeș-Bolyai University) for his comments on the earlier version of this paper. Also, I would like to thank all the participants of *Topics in Analytic Philosophy 3* workshop (TAP3, 19-20 April, Bucharest) for the interesting, provoking and helpful discussions.

[2] See Lewis 1986b for the discussion on omissions (whether they are events and should they be accepted as the causal relata).

concerned with the non-discriminatory notion of a cause,[3] and it is restricted to deterministic worlds.

The central notion of ITC is that of *influence*. Instead of analyzing causation in terms of simple counterfactual dependence between actual events (as in CTC),[4] in this new theory, Lewis proposes analysis is based on the relation of influence which is defined as a pattern of counterfactual dependences between various (actual and non-actual) *alterations* of two given events.

"Alteration" is a technical term in ITC. An alteration of a given event is an extremely fragile version or variation of that event, i.e. version of the event with maximally specified conditions (time and manner) of its occurrence. If we look *in that way* on some given actual event, with fully specified conditions of its occurrence, it is itself an alteration. But naturally, the rest of the alterations of that event are non-actualized, but merely possible events which are at least slightly different from the actualized alteration of the given event.

For example, if we take some actual throwing of the rock (call it event a) in some specific moment t, with a specific mass of the rock m, and force of the throw f, at a specific angle θ, and so on… the alterations of a (alongside the alteration that is actualized) would be possible events (a_1, a_2, a_3, a_i…) which differs (at least slightly) from the actual throwing with respect to time, force or angle of the throw, or with respect to the mass or the shape of the rock, and so on.

Lewis defines influence as follows:

> "Where c and e are distinct actual events, let us say that c influences e iff there is a substantial range c_1, c_2, . . . of different not-too-distant alterations of c (including the actual alteration of c) and there is a range e_1, e_2, . . . of alterations of e, at least some of which differ, such that if c_1 had occurred, e_1 would have occurred,

[3] Meaning, it is not an analysis of *the cause*, *main* or the *prominent* cause, but rather a theory about what should count as *a cause*, without further pragmatic considerations.

[4] To be more precise, causation is in CTC defined as the causal dependence between distinct events, which is in turn defined as a counterfactual dependence between the proposition about the occurrence of those events. Also, causation is ancestral of causal dependence

> and if c_2 had occurred, e_2 would have occurred, and so on." (Lewis 2000 [2004], 91)

Back to the example of the rock-throwing. Let us imagine a very simple scenario: Suzy throws a rock (event s) and breaks the bottle (event x). Surely, there is a substantial range of alterations of Suzy's throw such that – if those specific events (i.e. alterations) had taken place instead of her actual throw, some other alterations of bottle breaking, different than the actual one would have taken place as well.[5] So, if Suzy were to throw the rock earlier or at a different angle (which would constitute different alterations s_1, s_2, and so on), the bottle would break earlier or with the glass flying off somehow differently than it actually did (the alterations x_1, x_2, and so on).

According to the definition given above,[6] Suzy's throw *does* influence the bottle breaking. And, since her throwing of the rock is clearly a cause of the bottle being broken, that is a good result if we are to analyze causation using the influence relation.

This *almost* completes ITC.

ITC and preemption

In the introduction, I have mentioned the cases of preemption as an insurmountable obstacle to Lewis's CTC. Let us take a closer look at these cases since they are important for the rest of the paper and, moreover, they could help us understand ITC better.

Cases of preemption are asymmetrical cases of overdetermination (or redundant causation). In all cases of redundant causation, we have more than one event (say: a, b, c…) that "overdetermine" some effect, i.e. more than one event, such that each is sufficient for the effect (e) in the absence of others. In *asymmetrical* redundant causation, we can, in addition, clearly identify one among those overdetermining causes (say: c) as a cause, and others as merely the backups.

---

[5] I will – for the sake of convenience – often use the term *mapping* for this counterfactual relation between the alterations of cause and the alterations of the effect.

[6] For now, we can leave aside the vague phrases in the definition: "substantial range" and "not-too-distant".

The problem these cases pose to the CTC is immediately obvious – there is no counterfactual dependence between the given cause (*c*) and the effect (*e*) since it is not true that: if *c* had not occurred, *e* wouldn't have occurred. Because, it would, caused by the *b* or *a*, in the absence of *c*.

To illustrate an ordinary case of preemption, let us add to the Suzy's scenario another person – Billy, as in famous Lewis's examples (Lewis 1986c) – who throws another rock, aiming for the same bottle. So, we have Billy and Suzy throwing the rocks in order to break the bottle. Suzy's rock arrives first and breaks the bottle, just a millisecond later Billy's rock flies through space where the bottle had been. However, if Suzy's throw had been absent, Billy's rock would have broken the bottle, in almost the same time and manner.

ITC, as mentioned, was offered with the promise of solving the problem of preemption. So, how ITC works in these cases? There are clearly many alterations of Suzy's throw which could make a difference to the bottle breaking. Change the angle or the force of her throw (or the mass and shape of the rock) and, correspondingly, the effect would be different. However, there are also plenty of alterations of Billy's throw that map onto alterations of the bottle breaking – although, admittedly, not any alteration would do.[7] If Billy were to throw his rock earlier (enough), it would reach the bottle first and break it. Or, if he would have thrown the rock faster than he actually did, his rock would get there before Suzy's rock and would break the bottle. Moreover, every alteration of that sort (in which Billy's rock reaches the bottle before Suzy's rock), with further differences in force, mass, angle, and so on, would be mapped onto different alterations of the effect.

Should we conclude that both of these throws influence the breaking of the bottle? The answer is *no*, and the reason for that answer is present in the definition of influence. Vague phrases "substantial" and "not too distant", which acts as the restrictions on the *type* and the *volume* of the alterations – are there to ensure that we don't get this result (that both: preempting and preempted cause have the influence on the effect).

---

[7] We should, however, acknowledge that not any alteration of Suzy's throw will do, either. The alterations of her throw which *delays* the collision of her rock with the bottle (so that Billy's rock hit the bottle first) don't map onto different alterations of the bottle breaking.

So, in the case of Suzy and Billy, the right resolution of these explicit vague qualifications would be the one that allows as not-too-distant only those (unactualized) alterations that differ from the actualized alterations so slightly that:

- Such alterations of Suzy's throw do stand in the counterfactual relation with the different alterations of the effect.

- On the other hand, such alterations of Billy's throw do not map onto different alterations of the effect.

The central idea with this vague and case-sensitive threshold for influence rests on the observation that preempting cause has an advantage (over the preempted cause) that is relevant to influence and hence theoretically exploitable. To put it simply, preempting cause is in (somehow) a more delicate relation with the effect (than the preempted one is) and it is always possible to find really subtle "wiggle" that will be enough for the preempting cause to make difference to the effect, but not enough for the preempted cause to do the same.[8]

It looks like ITC *works* well with cases of preemption. And this, again, *almost* completes ITC. Just one more important thing…

ITC and Frankfurt cases

For all we have seen in this paper, causation in ITC can simply be identified with this (complexly defined) relation of influence. But *it isn't*. Like in the formulation of his CTC, Lewis again uses the maneuver of *ancestralization*, and here defines causation as the ancestral of influence. The precise definition is: "… *c* causes *e* iff there is a chain of stepwise influence from *c* to *e*." (Lewis 2000 [2004], 91)

In other words, two events (*c*, *e*) *can* stand in the causal relation although there is no direct influence between them, provided that there is a chain of influence(s) leading from *c* to *e*, i.e. provided there are some intermediate events (say) d_1, d_2, d_3… d_n such that *c* influences d_1, d_1 influences d_2, d_2 influences d_3…, and d_n influences *e*.

[8] The reader should have in mind that what I have presented is a charitable and streamlined reading of ITC. What is the right interpretation of that theory, or the most adequate one, is an open issue addressed in almost all papers discussing the theory: (e.g. Kvart 2001, Strevens 2003, Choi 2005, Maslen 2004).

But, do we need this further addition to theory? And if so – why? For which cases it would prove necessary to admit even a stepwise influence? To motivate this amendment to the theory, Lewis explicitly cites famous Frankfurt cases (FCs), as the perfect example of causation without direct, but with indirect (stepwise) influence.

In his classical paper, Harry Frankfurt offers a scenario that shows how a person could be morally responsible for some action even when she could not have acted differently (Frankfurt 1969: 835-836). The scenario involves Jones who is about to make some decision and perform some action, and Black who wants Jones to perform the exact action *e*, and who has the means of ensuring that outcome (let's say that he is a neuroscientist – as it is commonly assumed – who can control Jones's brain). Black monitors Jones in the process of deliberation and makes a decision whether or not to intervene, depending on the decision Jones had come to. Nevertheless, in the course of events, Black didn't have to do anything since Jones decided to do and did exactly what Black had wanted.

It is uncontroversial that Jones caused his consequent action.[9] But, an interesting question for us is – how can ITC account for this result? There is clearly no influence between his decision and the final outcome – due to the central feature of the given scenario; the outcome has no alternatives, it could not have been different, no matter what Jones had decided.

Lewis acknowledges both these claims: Jones's decision *was* a cause of his consequent action, and there is *no* influence between those two events. Nevertheless, he does not regard FCs as the CE to ITC, but rather as a clear example of *why* causation should be defined as an ancestral of influence relation.

How, according to Lewis, ancestralization helps in FCs? The explanation he gave is detailed and informative:

---

[9] Although the scenario in question is not the most simple and clear case of causation, this is an intuitive verdict, commonly and widely accepted in literature, and it is an important thesis within Frankfurt's argument. In their seminal, textbook-like work on free will, Fischer (Fischer et al., 2014: 54-61), Pereboom (Fischer et al., 2014: 87-90) and Kane (Fischer et al., 2014: 167-170), all speak about the Jones's action using causal locutions.

> "Let *c* be Jones's initial brain state; let *e* be the desired behavior. Consider a time after the neuroscientist has read Jones's brain, but before she would have seized control if the reading had been different. Let *d* combine Jones's brain state at that time with the neuroscientist's decision not to intervene. We have a two-step chain of influence from *c* to *d* to *e*. But *c* does not influence *e*."[10]

So, he regarded this scenario as another case of preemption.[11] One way of putting this suggested causal structure into a Lewis-style diagram is given in *diagram 1*.

Circles in this diagram represent events. Simple arrows should depict influence relation; the arrow with the rounded tip marks the inhibitory connection – the one that stops some event from being actualized; the arrow with dotted line represents potential and unactualized causal relation – if the event at the beginning of that arrow had been actualized, it would cause the event on the end of it. Shaded circles represent actualized events, while the non-shaded one is unactualized (since it is inhibited).

Alongside the events Lewis explicitly mentioned, I added a few more that are implicit in his explanation, in order to have a more elaborate and informative diagram. I also added the indicated moment *t* with the vertical dotted line cutting

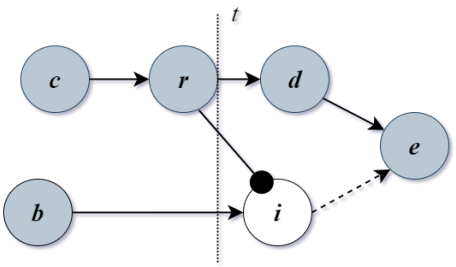

Diagram 1: Frankfurt's scenario

[10] One detail in this passage looks problematic. Namely, the Lewis's formulation that *d* combines Jones's brain state and Black's decision not to intervene, is either misleading or – in the worst case – goes against Lewis's argument here. I tried to avoid potential problems by keeping these two events simultaneous but distinct and separate. One such problem would be that if Black's decision not to intervene is on the main causal route, then it is also one of the causes of the given effect. And since it is (itself) the effect of Black's intentions and his monitoring, we then have a case of joint causation, rather than a case of preemption.

[11] Also, within the philosophy of free will – as the field from which this scenario originates – it is a commonplace to regard FCs as the case of preemption (see, for example, Funkhauser 2009, for the discussion on FCs and overdetermination)

through the course of events. Black with his intention and readiness to get the desired behavior from Jones is represented with *b*. Circle with *r* stands for Jones's brain state in the time of the crucial reading – reading that will determine whether Black intervenes or not. Black's intervention (on Jones's brain) that would happen if the reading had been different is depicted by *i* and it is inhibited or prevented by the actual reading.

Parallel to Lewis's explanation, in the diagram we have Jones's initial brain state (*c*) influencing his brain state at the moment of crucial reading (*r*), which in turn influences his decision (*d*).[12] We also have the reading (*r*) preventing Black's intervention (*i*), by deeming it redundant or unnecessary from Black's point of view. As expected, we have Jones's decision (*d*) influencing his behavior (*e*). And, finally, we have a potential causal relation between Black's intervention (*i*) and Jones's behavior (*e*), that is *merely* potential since Black's intervention wasn't needed and didn't happen.

To summarize: Lewis regards FCs, as many other philosophers do, simply as cases of preemption. Although different than worrisome case with Suzy and Billy, these cases are still accountable for within ITC, thanks to the ancestralization move (that allows stepwise influence to count as causation even without direct influence).

This completes ITC.

Early and late preemption

Before turning to the analysis of the CEs, there is one more important distinction to be drawn. A careful reader probably noted that there is an important structural difference between the example with Suzy and Billy and Frankfurt's scenario. Although both of these cases are the cases of asymmetric redundant causation, and hence both are rightly called preemption, they nevertheless structurally differ.

[12] The reader could probably see that *c* also influences *d*. That should not come as a surprise. Although influence is not transitive relation (as this exact example shows) it does not mean that it is antitransitive so that we can never have three events *a*, *b* and *c*, such that: *a* influences *b*, *b* influences *c*, and *a* influences *c*. Quite contrary, typically we would have just that.

When we take a closer look at Suzy and Billy case, we can see that what prevented Billy's rock from breaking the bottle was the sole event of bottle breaking – caused by Suzy's throw. In other words, there is nothing along the causal route leading from Suzy's throw to bottle breaking that inhibits (cuts, or stops in anyway) the causal route leading from Billy's rock to the same effect, except the occurrence of the effect itself.

Frankfurt's example is different in that respect. The event that stops Black from intervening happens early on in the scenario – certainly before the Jones has carried out his action. It is the crucial reading of Jones's brain (*r*) that prevents Black from taking over the control, manipulating Jones's brain and hence causing the effect himself. And that is apparent in *diagram 1*.

On the contrary, in the Billy and Suzy case, we have two sequences of events – one starting with Suzy throwing a rock (*s*) and another with Billy throwing a rock (*b*) – that proceed towards the effect, without interfering with each other, until the end – i.e. until the occurrence of the effect. The shattering of the bottle itself (*e*) prevents Billy's rock from breaking the bottle, by preventing some event antecedent to it, consisting of (say) Billy's rock making the initial contact with the glass of the bottle (g_2). That can be represented as in *diagram 2*.

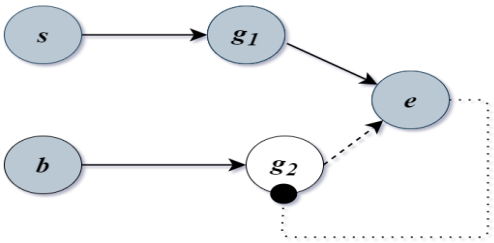

Diagram 2: Late preemption scenario with Suzy and Billy

When we compare *diagram 1* to *diagram 2*, we can see that the structural difference between them is concerned with – when this cutting of the alternative causal sequence happens. In the case of Jones and Black, it happens *early* – meaning, before the preempting causal sequence has reached the effect. Cases like this Lewis calls *early preemption*.

On the other hand, in the case of Suzy and Billy cutting happens *late*. It is the effect that functions as an inhibitor, so the cutting of the preempted causal chain happens in the moment of (or in some moment after) its occurrence. Cases of preemption with this feature Lewis calls *late preemption*.

Two counterexamples: Button vs. Switchboard and Smart Rock

Back to the counterexamples. The literature about ITC is filled with different CEs to the theory (a list of notable ones would certainly include: Kvart 2001, Schaffer 2001, Strevens 2003, Hall 2004, Bigaj 2012). Some of them are directed towards proving that influence is not necessary for causation, i.e. that we can have causation without influence. Some others aim to prove that we can have influence between two distinct events without having causation between them. Some even aim for both.

In this section, I will present and analyze two prominent CEs, which allegedly prove that influence is neither necessary nor sufficient for causation. As it will become apparent shortly, those two CEs are structurally very similar to each other, and even more interestingly, very similar to – just discussed – Frankfurt cases. It is exactly this latter similarity that should make us suspicious with regards to their plausibility and effectiveness against the theory.

Let us start with the CE offered by Schaffer, which he conveniently called *Button vs. Switchboard counterexample* (BvS).

> "The set-up: Pam is locked in a room which contains a single button. Bob is locked in a room which contains a vast switchboard. Vic is covered with electrodes and strapped to a chair. The story: Pam presses the button. Bob just watches. Vic is electrocuted." (Schaffer 2001, 12) [Additional story] "Bob is in fact a preempted backup who will jump in if Pam delays for even a millisecond… and Pam's wiring was only just set up at the time she actually pressed, so that had she hastened even a millisecond her button would not have worked and Bob would then have done the deed at the time and in the manner Pam actually did." (Schaffer 2001, 15)[13]

The moral of the story is this: we are inclined to consider Pam's pressing the button as a cause of Vic's electrocution, but that event has no influence whatsoever on the effect. Take any alteration of her

[13] Schaffer develops throughout the paper different CEs based on the same basic set up – that he offers at the beginning. That is why this specific CE is segmented and has two parts.

pressing the button – in which she presses the button earlier, later, harder, with right or left finger …, or even the alteration where she doesn't press the button at all – the corresponding effect is the same, Vic is getting electrocuted (at the same time and in the same manner).

On the other hand, Bob is only a preempted back up, and we wouldn't count him as a cause of the electrocution. Nevertheless, as Schaffer claims, we would be justified in asserting that Bob's watching influences the electrocution, since he has a vast switchboard for manipulating Vic's electrocution.[14]

To summarize, according to Schaffer, in BvS we have a case in which a cause has no influence on the effect; and, moreover, we have the event which does have the influence on the effect but is still not a cause of that effect, but merely a preempted alternative.

Structural similarity to the Frankfurt cases should be apparent.[15] In both cases, we have one person (Jones/Pam) ready to perform some action, and the other person (Black/Bob) monitoring the process and ready to step in if the first person diverges from the course leading to that desired behavior. In both cases, that other person does not intervene in the end, but acts as a preempted back up, ensuring that there is no alternative to the actual effect.

---

[14] Bob, according to the scenario and thanks to the aforementioned switchboard, can deliver Vic's electrocution in a different time and different manner: varying in the power of electricity, frequency, etc. That guarantees a very reach range of alterations (of Bob's behavior) that map onto different alterations of Vic's electrocution. It is, indeed, objectionable – as Noordhof points out (Noordhof 2001) – that those alterations are not-too-distant since we need different sorts of using the switchboard to be the alterations of Bob's *just watching*. But that is hardly an objection that would bother Schaffer. All he needs is that preempting cause doesn't have the advantage (of the kind described earlier) over preempted cause, so no resolution of vague restrictions "not-too-distant" and "substantial" can go in favor of preempting cause.

[15] There is, still, one important difference. In Schaffer CE we have embedded contrast between what Pam did and could have done, and what Bob did and could have done. The idea behind that is, of course, to provide Bob with a rich range of alternatives that map onto the effect, while restricting the same for Pam. There is no such strong contrast in Frankfurt cases. However, this is not relevant to the overall structure of the scenario and to the question of the applicability of Lewis's ancestralization maneuver.

Interestingly, Schaffer himself does not mention the Frankfurt cases, nor he addresses this similarity in any way. Nevertheless, that similarity poses a serious threat to the plausibility of his CE. If BvS can be framed in early preemption pattern in the same way Frankfurt cases can, then we would be right to dismiss it as an effective CE for ITC.

Before addressing this question, let us take a closer look at another CE with a similar structure – again. Namely, we are going to examine the Smart Rock CE, offered by Ned Hall, although he credits it to Steven Yablo (Hall 2004, 237). The aim is, again, twofold, but the stress in the paper is on the claim that influence is not necessary for causation. To show that, Hall (or rather Yablo) proposes a slightly altered story about Suzy and Billy:

> "This time, Billy throws a Smart Rock, equipped with an on-board computer, exquisitely designed sensors, a lightning-fast propulsion system – and instructions to make sure that the bottle shatters in exactly the way it does, at exactly the time it does. In fact, the Smart Rock doesn't need to intervene, since Suzy's throw is just right. But had it been any different – indeed, had her rock's trajectory differed in the slightest, at any point – the Smart Rock would have swooped in to make sure the job was done properly." (Hall 2004, 237)

Again, we have clear intuition that Suzy throw is a cause of bottle breaking, but due to the preempted backup – that is ready to intervene and ensure the effect is exactly the same, no matter which alteration of the cause had been actualized – there is no influence between Suzy's throw and the bottle shattering. On the other hand, one could insist that throwing the Smart Rock does influence the effect since its different settings could make a difference to the breaking of the bottle.

The analogy with the FS and BvS is, I believe, easy to spot. We have Jones, Pam, and Suzy, all three bringing about some effect, but in the complex environment which overdetermines that effect with the backup alternatives. Those backup alternatives – namely: Black, Bob, and Billy – are all idle in the actuality, but nevertheless rob their respective preempting causes of any influence on the effect.

Problems for the two CEs?

Does this structural similarity between BvS and SR (on the one side), and FCs (on the other), constitute a problem for these two CEs? The reason for this worry is simple: since we have seen that ITC has the means to deal with FCs, and we have recently maintained that there is a relevant similarity between FCs and two given CEs, it is natural to ask whether ITC can explain away, in the same manner, BvS and SR? Or, even more precisely, isn't it a case that we could have a stepwise influence in the alleged CEs offered by Schaffer and Hall?

We can start answering this question by analyzing BvS. A quick recap of this CE: we have Pam pressing the button (which is pressable only at the given moment and cannot be pressed before or after that), we have Bob, who is monitoring the signal leading from Pam's button to Vic's electrodes, with the intention and the means of ensuring that Vic is getting electrocuted in the same time and manner that he actually was (due to the Pam's pressing of the button).

This scenario is easily adjustable to the diagram I have offered earlier for FCs. It has all the characteristics of the early preemption cases. In arguing that, I am agreeing with Noordhof (Noordhof 2001) and further support his claims by offering the diagram parallel to the one constructed for FCs.

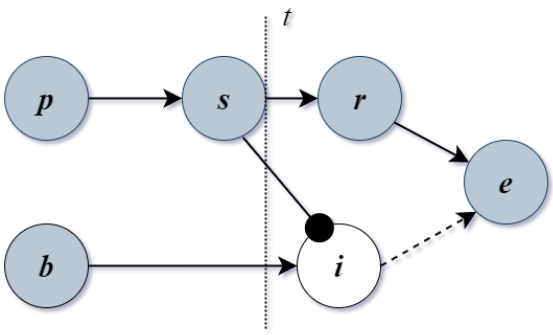

Diagram 3: Button vs. Switchboard CE

In this diagram (*diagram 3*), we have *b* standing for Bob (with his intentions) who starts monitoring what happens with the signal; then, we have *p* that marks the event of Pam's pressing the button, which causes the signal to go through the part of the wire that Bob is monitoring (*s*). This *s* event has two further effects: on the one side, it causes the signal to continue its journey through the rest of the wire (*r*) which leads to Vic's electrocution (*e*); and on the other side, it

inhibits Bob's intervention (*i*), since Bob – after seeing the signal – decides not to intervene. I also added the moment *t*, at which Bob finishes his monitoring, and depending on it decides what to do.

If this diagram is correct and parallel to the one offered for the FCs, then we should have that *p* influences *s*. But that could seem controversial – and Schaffer would probably disagree with it (Schaffer 2001: 16). No matter how Pam presses the button (gently or hard, with her left or right hand, with the index finger or the thumb…) as long as it is at the given time, the signal is the same (and we have the same alteration of *s*, which is the actual alteration – call it s_a). On the other hand, according to the scenario, since the button is only responsive in that given moment (in which Pam actually pressed it), all other potential pressings of the button, before or after that given time, simply map onto the one and the same unactualized alteration of the event *s* – namely, the alteration in which the signal doesn't go through the wire (let us call that alteration s_b). Pam's not pressing the button at all – if that should count as a not-too-distant alteration of her pressing the button – also maps onto the alteration s_b.

So, bottom line, we have only two different alterations of the effect, onto which all the different alterations of the cause are mapped. Is this enough for influence? Do we have a substantial range of not-too-distant alterations of *s* that maps onto different alterations of the effect *t*?

The answer has to be – yes. The definition of influence grants that. In it, it is explicitly stated that the alterations of the cause should map onto alterations of the effect such that *at least some of them differ*. It is the peculiarity of the case under consideration that makes the potential range of different alterations of the effect sparse. Hence, the *richness* and *variety* in the mapping that is required for influence also have to be proportionally moderate in this case. If we don't allow for this kind of reading of the vague phrases in the definition of influence, then it is easy to have even less complex CEs to ITC than the ones we have considered above.

Back to the diagram. Now, after this worrisome first step, other steps are pretty straight forward. Had the signal in the monitored part of the wire been different,[16] it would certainly be different later on trough

[16] In analyzing the pattern of counterfactual dependence between these two events, we are free to entertain even those possibilities that are not consistent with what the

the given wire, as well. So, we have influence between *s* and *t*; and, similarly, all the different alterations of the signal in the later part of the wire stand in the counterfactual relation with different alterations of Vic's electrocution.

To summarize: although we do not have direct influence between *p* and *e*, it could be shown – as I have tried to maintain in the passages above – that there is an indirect, stepwise influence between those two events.

Now, can we do the same for Hall's SR scenario? A quick recap: we have a regular rock thrown by Suzy, which breaks the bottle; in addition, we have Billy throwing a smart rock which monitors the trajectory of Suzy's rock, and which is able to step in at any moment and ensure that effect happens the same way it actually happened. Admittedly, this scenario looks structurally very similar to the one just analyzed. So, is SR another case of early preemption in which we do have a stepwise influence between the cause and the effect?

No, SR is not a case of early preemption at all. Although it does look similar to FCs and BvS, it still has one important feature that is different, and that feature deprives it of being an instance of early preemption.

In both cases, FCs and BvS, we hypostasized some moment *t* in which the monitoring happens (or rather ends), and after which – depending on the reading – backup cause can step in and ensure the effect. The events that happen on the main causal route after that moment are crucial for the early preemption since any of them can serve as an intermediate event – one which depends on the cause, and on which in turn depends effect, although there is no direct dependence between the cause and the effect.

---

scenario tells us about Pam's pressing of the button. That is out of the picture now and – contrary to what Schaffer seems to think (Schaffer 2001, 16) – we don't need to ask: "well, how the signal could be there earlier if it is impossible for Pam to press the button earlier", or "how it could be stronger (the signal) when Pam only has a button without any controllers", and so on. What we would assume in doing that is the truth of some so-called backtracking conditionals (e.g. "If the signal were to be different, then it would have to be the case that Pam had a switchboard rather than simple button"). But these conditionals are peculiar and problematic – they seem to state how the things that are earlier depend on things that happened later on; Lewis denies that they can be true in the common contexts (Lewis 1979, 457-8; 1986c, 169).

But the reader should note that in the case of SR we don't have such a moment and, furthermore, we can't add it consistently to the story. Instead of an *early* moment in which the monitoring ends, and after which the potential intervention of the backup alternative happens, in this scenario, we have monitoring as an ongoing process that ultimately ends only with the occurrence of the effect.[17] And that is, as we have seen, a distinctive mark of late preemption.

Simply, in SR – and cases similar to it – it would be impossible to find an intermediate event, which is crucial for the implementation of stepwise influence maneuver. Hence, we cannot frame these cases into a diagram similar to those for FCs and BvS.

Consequently, with this stepwise strategy unavailable, and without direct influence between the cause and the effect in the SR scenario that Hall offers, ITC is surely in big trouble. As it turns out, SR can't be that easily disarmed (as BvS was).

Concluding remarks

In this paper, I have analyzed two prominent counterexamples to the Influence theory of causation, with a special interest in their similarity to Frankfurt cases. As I tried to show, that similarity poses a threat to the plausibility of the given CEs.

And indeed, by insisting on the structural similarity of FCs and BvS, and hence accepting Noordhof's critique of that Schaffer's CE (together with offering a further elaboration of that critique) I aimed to show that Button vs. Switchboard CE is not a valid CE to the Lewis's latest theory of causation.

---

[17] Somebody is maybe inclined to object that this shows that the scenario in question is not a viable one, that it simply cannot be real. The objection would proceed by stating that in order to intervene, even the Smart Rock would require some finite amount of time. So, the monitoring process should end, if not earlier, then right before the time of the effect occurrence (how much before? – the same amount of time needed for its intervention). This objection has some plausibility, but it crucially depends on the (*a posteriori*) physical restrictions and deals with the physical impossibility. That, however, is too restrictive when we seek a conceptual analysis of causation. If the analysis is successful in catching the notion of causation, it should be general enough to "work" even with different physical laws.

However, it would be wrong to conclude the same with regards to Hall's CE. Due to a slight but important structural difference between SR and BvS, the maneuver used to disarm BvS is not applicable to the CE Hall had proposed.

More precisely, Lewis's ancestralization maneuver, i.e. his introduction of a stepwise influence to the theory, was only meant for dealing with early preemption, and SR is not an instance of early preemption, as we have seen. On the other hand, although ITC was formulated with the hope of solving the late preemption problem, clearly there are still cases (different than typical late preemption cases) for which the theory doesn't give the right answers.

The analysis offered in this paper serves to show that SR is a plausible and successful CE, and, consequently, ITC is not a fully adequate theory of causation. Also, that analysis, I hope, provide us with good reason to favor Hall's CE (over the one Schaffer had offered) as the compelling argument against the theory.

REFERENCES

Bigaj, T. 2012. "Causation Without Influence". *Erkenntnis* 76 (1): 1-22.

Choi, S. 2005. "Understanding the Influence Theory of Causation: A Critique of Strevens". *Erkenntnis* 63**,** 101-118.

Collins, J., Hall, N. and Paul, L.A. (eds.). 2004. *Causation and Counterfactuals*. Cambridge, Massachusetts: MIT Press.

Dowe, P. 2000. *Is Causation Influence?*. Unpublished manuscript, available at: http://fitelson.org/269/Dowe_ICI.pdf.

Frankfurt, H. 1969. "Alternate possibilities and moral responsibility". *Journal of Philosophy* 66 (23): 829-839.

Fischer, J.M., Kane, R., Pereboom, D. and Vargas, M. 2014. *Four Views on Free Will*. Malden (Ma); Oxford; Victoria: Blackwell.

Funkhouser, E. 2009. "Frankfurt Cases and Overdetermination". *Canadian Journal of Philosophy* 39(3), 341-369.

Hall, N. 2004. "Two concepts of causation". In: J. Collins *et al*. (ed.). 2004. pp. 225-276.

Kvart, I. 2001. "Lewis's 'causation as influence'". *Australasian Journal of Philosophy* 79 (3), 409-421.

Lewis, D. 1973. "Causation". *Journal of Philosophy* 70: 556-567.

—— 1979. "Counterfactual Dependence and Time's Arrow", *Noûs* 13: 455-476.

—— 1986a. *Philosophical Papers II*. Oxford: Oxford University Press.

—— 1986b. "Events". In: Lewis 1986a.

—— 1986c. "Postscripts to 'Causation'". In: Lewis 1986a.

—— 2000. "Causation as Influence". *Journal of Philosophy* 97/4: 182-197. Reprinted in: J. Collins *et al*. (ed.). 2004. Pp. 75-106.

Noordhof, P. 2001. "In Defence of Influence". *Analysis* 61 (4): 323-327.

Maslen, C. 2004. "Degrees of influence and the problem of pre-emption". *Australasian Journal of Philosophy* 82 (4): 577-594.

McDermott M. 2002. "Causation: Influence versus Sufficiency", *The Journal of Philosophy* 99 (2): 84-101.

Schaffer, J. 2001. "Causation, Influence, and Effluence". *Analysis* 61.1, 11-19.

Strevens, M. 2003, "Against Lewis's New Theory of Causation: A Story with Three Morals". *Pacific Philosophical Quarterly* 84: 398-412.