Interlevel causation and external causes

By Marco Totolo

Interlevel causation, also called bottom-up or top-down causation, is a type of relation between entities defined at different levels of aggregations.

Apparently, when we say that an entity at a lower level causes one at a higher level of which the former is a part of, we have a problem of circularity: the part causes the whole, but as the whole is also composed by that part, the part ends up being a cause of itself. The same pattern applies to top-down cases: the whole causes one of its parts and therefore can be said to be a cause of itself.

The part-whole relationship that arises in a system with different levels can be described as the relation between a mechanism and its parts. A mechanism can be defined as "a structure performing a function in virtue of its component parts, component operation and their organisation. The orchestrated functioning is responsible for one or more phenomena" (Bechtel & Abrahamsen 2005). The part-whole relation instantiated by a mechanism is therefore always local and relative to the specific activities carried out by the mechanism and by its parts.

The components of a mechanism are hence themselves entity/activity pairs that, in virtue of their mutual interactions, give rise to the entity/activity pair identified with the whole. The nature of the relationship between the parts and the whole is of critical importance. There have been a few attempts to define necessary and sufficient conditions for parts to be constitutive of a mechanism, see Woodward's mark transmission approach (Woodward 2003) and Craver's ideal intervention notion (Craver 2007). The qualitative definition given above will suffice for the purpose of this essay.

Craver and Bechtel 2007 describe two examples of apparent bottom-up causation. In the first one a virus infects a general and we can say that the virus is the cause of the general's infection. In the second example the general has a heart attack, which causes her death. We will see that the latter can't be said to be a genuine case of causation. In the next paragraph I will merge the two examples and relate Craver and Bechtel's (C&B) account.

Let's suppose that a virus infects the general's body and sets in motion a causal chain of

events leading to the symptoms of the general's death by heart attack. Those symptoms constitute the general's death, in the sense that the death can't be said to be caused by the absence of psychological functions, it is just that absence. The viral infection and the general's biological functions and states are located at the lower level, while the general's death is at the upper level. We say that the heart attack caused the general's death, but what we really mean is that a hybrid kind of relation obtains: strictly causal (the virus causing the cessation of the general's psychological activity through the stopping of the heart) and constitutive (the absence of psychological activity constituting the general's death). Using this strategy C&B aim to explain any type of apparent interlevel causation: bottom-up and top-down cases can be treated in analogous fashion.

An important notion relative to the concept of levels used is that the levels themselves are defined locally. The general's biological functions and her death are on different levels because the latter is constituted by the former. On C&B's view they can only interact constitutively. The virus on the other hand, despite its existing at a very different magnitude scale than that of the general's, can be said to cause the general's death because it doesn't belong to any of the general's internal parts or of the definition of her death. This means that entities of very different sizes can interact causally even if they belong to very different size groups. What can't causally interact are entities at different levels of the same system.

Using the same notation as in Craver 2007, we can define a mechanism S that performs the activity ψ as Sψ, which denotes S ψ-ing. The mechanism is made of a number of parts ($X_1 \ldots X_n$) each performing an activity ($\varphi_1 \ldots \varphi_n$). $X_i\varphi_i$ is at a lower level with respect to Sψ iff $X_i\varphi_i$ is a component of the mechanism Sψ.

In the general's heart attack example, Sψ is the general's death and ($X_1\varphi_1 \ldots X_n\varphi_n$) are the lower level components of the general's biology that make up her death. Yχ is the virus infecting the general. Yχ isn't a lower level component of Sψ. Yχ causes one or more of the conditions $X_i\varphi_i$, say $X_k\varphi_k$, which in turn causes all the other ($X_1\varphi_1 \ldots X_n\varphi_n$). Sψ is caused by Yχ through the constitutive relation between the ($X_1\varphi_1 \ldots X_n\varphi_n$) and Sψ. This means that when we say that the virus caused the general's death, we are implying a hybrid kind of causation.

C&B's view is very appealing and provides an elegant and concise way to explain (away) apparent cases of interlevel causation. In what follows I will argue that there can be cases of causation that violate this view and hence must be dealt with before fully embracing C&B's hybrid causation view. I will propose a way to accommodate those cases.

Suppose that the general is holding a vial containing the virus and accidentally lets the vial

drop to the ground. The vial breaks, the virus is released and causes the general's death, as outlined in the example above.

Now suppose that the accidental drop of the vial was caused by the activity of one of the biological functions involved in the general's subsequent death, for example the unresponsiveness of a certain muscle, which was in this case caused by something other than the virus. Here we have the lower level cause that directly causes the events leading to the breaking of the vial, which in turn cause the general's death. This new chain of events starts from the failing of the muscle and ultimately causes the general's death. This can be considered as a counterexample of C&B's view, since the component of a mechanism is causally related to the mechanism itself. Note that the failing of the muscle to respond is one and the same component/activity pair that's involved in the dropping of the vial and in the general's death, at least if we want to consider mechanisms and their components as types and not tokens. I assume that mechanisms are types: my heart pumping blood is one and the same mechanism as anybody else's heart pumping blood.

One might object that the failing of that specific muscle isn't a necessary condition for the general's death by heart attack. This might be the case for example if, for some reason, the general didn't have that muscle to begin with. However, here I'm considering the death of the general as the whole and the failing of the muscle as a symptom of death (assuming the general actually has that muscle) even though it's not essential to a heart attack (of which the only essential thing would be the stopping of the heart itself.) I understand that there is a clash here between type and token instantiations of mechanisms. As I've written mechanisms must be types because my pumping of the heart is intuitively the same mechanism as your pumping of the heart. But mechanisms are also defined locally when we're dealing with the part-whole relation. In that case it's a token relation. But I don't think it's a contradiction: the interlevel relationship is tokenised, while the mechanism as a whole can be typified.

The upshot is that while the failing of the muscle is not a necessary condition for death (or death by heart attack), it's still a part of the mechanism (part-whole relationship) I'm proposing as an example.

Using the previous notation, we have $X_j\varphi_j$, which is one of the ($X_1\varphi_1$ … $X_n\varphi_n$), causing $Y\chi$, which in turn causes $X_k\varphi_k$ and all the other ($X_1\varphi_1$ … $X_n\varphi_n$) constituting $S\psi$. The relationship between $X_j\varphi_j$ and $S\psi$ is double: causal and constitutive.

It seems that that same issue would arise each time that a lower level component of a mechanism is a cause of an event taking place outside the mechanism, which in turn influences the

mechanism as a whole from the outside. Clearly, the role of $X_j\varphi_j$ is different when it causes $Y\chi$ and when it is a constitutive part of $S\psi$, but we need a way to tell the difference between the two cases without reference to the two different token instantiations: the $X_j\varphi_j$ type is one and the same in both cases.

One possible strategy to sidestep the above scenario is to introduce a time constraint. I assume that the constitutive relation between a part and the system which it is a constituent of is instantaneous: when any one of the $(X_1\varphi_1 \ldots X_n\varphi_n)$ changes we have by definition a change in $S\psi$. On the other hand, a causal relation as described by mechanistic causation can always be thought to act in a definite time span. We can expect that the $X\varphi$ happening simultaneously with $S\psi$ be in a constitutive relationship, and the $X\varphi$ happening before $S\psi$ be in a causal relationship.

We can introduce the notation $S\psi t$, meaning S ψ-ing at time t. In the accidental vial drop example we have the following chain:

$X_j\varphi_j t_{j1} \rightarrow Y\chi t_\chi \rightarrow X_k\varphi_k t_k \rightarrow (X_1\varphi_1 t_1 \ldots X_n\varphi_n t_n)$ with $t_{j1} < t_\chi < t_k < \min(t_1..t_n) < \max(t_1..t_n) = t_\psi$

And, for the sake of clarity, here's a list of the time indexes in play:

t_{j1}: the muscle fails to respond (inside the mechanism, lower level);

t_χ: the vial drop (outside the mechanism);

t_k: the first symptom of the general's death (inside the mechanism, lower level);

$t_1 \ldots t_n$: all the other symptoms of the general's death (inside the mechanism, lower level);

t_ψ: the general's death (inside the mechanism, upper level).

In order to avoid the $X_j\varphi_j$ being a cause of itself, I'm supposing that between the time t_{j1} and t_k the X_j stops φ_j-ing, and starts again at time t_{j2} (between $\min(t_1..t_n)$ and $\max(t_1..t_n)$) as a consequence of the causal events started by $X_k\varphi_k t_k$.

At the time t_ψ, when the last of the $(X_1\varphi_1 \ldots X_n\varphi_n)$ starts φ-ing, S starts ψ-ing and the constitutive relation is realised. At any time earlier than t_ψ, $X_j\varphi_j$ is only potentially a part of $S\psi$ because S is still not ψ-ing. This means that for any $t < t_\psi$, $X_j\varphi_j$ can be a cause of $S\psi$. In fact, without $S\psi$ we can't say that $X_j\varphi_j$ is a part of it.

Our final analysis is therefore the following: at $t = t_{j1}$ and any later $t < t_{j2}$, $X_j\varphi_j$ is a cause of $S\psi$ without raising any circularity problem, since for that time span $S\psi$ still doesn't exist. At $t = t_\psi$ and all the $t > t_\psi$ until S stops ψ-ing, $X_j\varphi_j$ stands in a constitutive relation with $S\psi$. Note that the above analysis has been applied to bottom-up cases but could be easily extended to top-down cases.

Summarising, it seems that in order to apply C&B's hybrid causation we have two options: either we avoid the ambiguity of saying that $X_j\varphi_j$ is a part of $S\psi$ when S still isn't ψ-ing or we assign a time index to each component/activity pair at the lower and upper level of the mechanism.

Bibliography

Woodward, J. 2003: *Making Things Happen: A Theory of Causal Explanation,* Chapter 2. New York: Oxford University Press.

Craver, C. & Bechtel, W. 2007: "Top-down causation without top-down causes". *Biology and Philosophy*, 22, 547-563.

Craver, C. 2007: *Explaining the brain*, Chapter 4. Oxford.: Oxford University Press.

Bechtel, W. & Abrahamsen, A. 2005: "Explanation: a mechanist alternative". *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, 36, 421-441.