Can Program Explanations Save the Causal Efficacy of Beliefs?

Sven Walter, Saarbrücken

Frank Jackson and Philip Pettit offered the "program explanation account" (*PEA*) in order to vindicate the causal relevance of mental states such as beliefs. According to J&P, a property *F* of a cause-event *c* (potentially a mental property) can be causally relevant for an effect-event *e*'s having property *G* because "*e* had *G* because *c* had *F*" is an informative, non-redundant *program explanation*. If *PEA* succeeded, the causal relevance of beliefs would be vindicated and mental property epiphenomenalism would be avoided.[1] However, it doesn't succeed.

A "program explanation identifies a condition such that its realization is enough to ensure that there will be causes to produce the event explained" (J&P 1992, 119). Since the instantiation of a mental property *M ensures* the instantiation of a physical property *P*, an explanation of *e*'s having *G* in terms of *M* ("*e* had *G* because *c* had *M*") is informative and *M*'s instantiation *programs for* the existence of a property, *P*, with causal powers suitable to bring about *e*'s having *G*.

> The realization of [*M*] ensures … that a crucial productive property is realized and … that the [effect] event … occurs. [*M*] does not figure in the productive process leading to the event but it more or less ensures that a property-instance which is required for that process does figure. … [I]ts realization programs for the appearance of the productive property and … for the event produced. (J&P 1990a, 114)

Suppose the temperature of the water in a closed flask is raised until the flask cracks. The salient program explanation is "The flask shattered because the temperature of the water reached boiling point". The rise of the temperature did not itself cause the shattering, which was caused by the impact of molecules on the walls of the flask. Still, we "properly count citing the increase in temperature as explaining the shattering, for the increase programmes the shattering" (J&P 1988, 395). Therefore, although what is causally relevant is the impact of molecules, the increase in temperature is still causally relevant$_{JP}$, and this is enough, J&P claim, to reject mental property epiphenomenalism.

This suggests the following notion of causal relevance$_{JP}$:

> **Definition 1:** *c*'s having *F* is causally relevant$_{JP}$ for *e*'s having *G* iff (i) *c*'s having *F* is causally relevant for *e*'s having *G*, or (ii) *c*'s having *F* is causally irrelevant for *e*'s having *G* but ensures the instantiation of a property *H* of *c*, distinct from *F*, such that *c*'s having *H* is causally relevant for *e*'s having *G*.

But D1 is too weak. Suppose Hannah's aluminum ladder touches an electric wire and Hannah dies. The ladder has the dispositional property of *being a good thermal conductor*, which supervenes upon its categorical basis, a property of the cloud of free electrons that permeates the metal. The ladder's being a good thermal conductor thus ensures the instantiation of a property (the categorical basis), which is causally relevant for Hannah's dying, but the ladder's being a good *thermal* conductor was certainly not relevant for Hannah's dying.

If *one* lower-level property is the categorical basis of *two* distinct dispositional properties (like *being a good* thermal *conductor* and *being a good* electrical *conductor*), ensuring the presence of a causally relevant lower-level property is *not sufficient* for causal relevance$_{JP}$.

One might want to add that *F*, in addition to ensuring the presence of a causally relevant property, must figure in an informative program explanation. The result would be:

> **Definition 2:** *c*'s having *F* is causally relevant$_{JP}$ for *e*'s having *G* iff (i) *c*'s having *F* is causally relevant for *e*'s having *G*, or (ii) *c*'s having *F* is causally irrelevant for *e*'s having *G* but ensures the instantiation of a property *H* of *c*, distinct from *F*, such that *c*'s having *H* is causally relevant for *e*'s having *G and* "*e* is *G* because *c* was *F*" is an informative program explanation.

The above counterexample can then be avoided since "Hannah died because her ladder was a good *electrical* conductor" *is* an informative program explanation, while "Hannah died because her ladder was a good *thermal* conductor" is *not*. But *why* is the former, in contrast to the latter, an informative program explanation?

Here's a test of significance for any program explanation. Suppose that we have a program explanation of an event *e* by reference to an antecedent *P*, and that *P* explains *e* because its realization effectively ensures that some factor of type *F* occurs. Imagine now that we identify the *F*-factor in operation. *A useful test for the significance of the original program explanation is to ask whether it offers any information not available, at least under ordinary assumptions, to someone possessed of the F-explanation*. (J&P 1992, 124; italics S.W.)

But this test remains silent about why "Hannah died because her ladder is a good electrical conductor" *is* an informative program explanation but "Hannah died because her ladder is a good thermal conductor" is *not*. Explanations in terms of the thermal conductivity of Hannah's ladder certainly provide information not available to someone possessed solely of explanations in terms of clouds of free electrons. "Hannah died because her ladder is a good thermal conductor" is an informative program explanation, according to J&P's test.

It will not do to argue that "Hannah died because her ladder is a good electrical conductor" is informative because it mentions a property *relevant* to Hannah's dying. This would be tantamount to saying that those program explanations are informative that appeal to relevant properties, and obviously the attempt to ground the informativeness of program explanations in their appeal to relevant properties is hopeless if the causal relevance$_{JP}$ of a property is supposed to be grounded in its aptness to figure in informative program explanations.

Adding to D2 the requirement that the property in question figures in an informative program explanation does not lead anywhere, unless we can specify which program explanations are informative.

[1] J&P distinguish between causally *efficacious* properties (properties doing 'real' causal work) and causally *relevant* properties (properties which are either causally efficacious, or causally inefficacious but figure in program explanations). I will use 'causal relevance' instead of J&P's 'causal efficacy', and 'causal relevance$_{JP}$' instead of their 'causal relevance'. J&P claim, then, that a causally irrelevant mental property may nevertheless be causally relevant$_{JP}$ and that this suffices to avoid mental property epiphenomenalism.

In another paper, J&P have offered an account based on the idea of what they have called "invariance of effect under variation of realization" (J&P 1990b, 202).

> We can express the basic idea behind a program explanation in terms of what remains constant under variation. Suppose state *a* caused state *b*. Variations on *a*, say, *a*', *a*'', … would have caused variations on *b*, say *b*', *b*'', …, respectively. It may be that if the a^i share a property *P*, the b^i would share a property *Q*: keep *P* constant among the actual and possible causes, and *Q* remains constant among the actual and possible effects. If you like, *Q tracks P*. Our point is that in such a case *P* causally explains *Q* by programming it, even though it may be that *P* does not produce *Q*. (J&P 1988, 394)

The idea underlying this relatively abstract formulation is simple. In cases of informative program explanations, the same effect (e.g. the shattering of the flask) would, in different possible situations, have been produced by different lower-level realizers of the higher-level property cited in the program explanation. This higher-level property is the property in common to all those possible situations, in each of which the effect would have been produced by a realizer and in one of which it has actually been produced.

If *any* of the realizers of the property cited in the program explanation, the actual one as well as the possible ones, would have brought about the effect – if there is *invariance of effect under variation of realization* – then this property is causally relevant$_{JP}$.[2] If, however, there are realizers of the higher-level property that fail to bring about the effect – if there is *variance* of effect under variation of realization – then this property is causally irrelevant$_{JP}$. This is why Hannah's ladder's being a good electrical conductor differs from its being a good thermal conductor (or its opacity, which is J&P's example):

> The reason being a good conductor of electricity is causally relevant to [Hannah's] death is that it did not matter … what the categorical basis of that disposition was, for provided the causal role definitive of good electrical conductivity was occupied by a state of the ladder she would have died. … And, of course, the reason opacity, say, is not causally relevant to her dying is that it might easily have been realized without her dying – as would, for instance, have been the case had the ladder been wooden. (J&P 1990b, 205)

The notion of causal relevance$_{JP}$ based on the idea of "invariance of effect under variation of realization" can be captured as follows:

> **Definition 3:** *c*'s having *F* is causally relevant$_{JP}$ for *e*'s having *G* iff for any property *P* of *c* such that *P* realizes *F*, if *c* has *P* in some world *w*, then *c*'s having *P* is causally relevant in *w* to *e*'s having *G*.

Yet, D3 is neither necessary nor sufficient for causal relevance$_{JP}$.

To see how the above example can be modified to yield a counterexample to D3, note that the causal irrelevance$_{JP}$ of thermal conductivity was due to there being realizers of thermal conductivity that are not causally relevant for Hannah's dying, namely those *that do not in addition realize electrical conductivity*. The existence of those realizers is responsible for the variance of effect in the case of thermal conductivity. This shows that all we need is a causally relevant$_{JP}$ property *F* and a causally irrelevant$_{JP}$ property *H*, such that *H* has *no* realizers that are not also realizers of *F*: In this case, if *F* is causally relevant$_{JP}$ to *e*'s being *G* according to D3, then all realizers of *F* will bring about *e*'s being *G*, but so will all realizers of *H*, because the latter are just a subset of the former. Hence, any property *H* for which $\Box\forall R\ (R \in \Pi_H \supset R \in \Pi_F)$ will also be causally relevant$_{JP}$ to *e*'s being *G* (where "$R \in \Pi_H$" means "*R* is a realizer of *H*" and the necessity operator '□' has whatever modal force is operative in our definition of realization).[3] Thus, we need an example such that:

(a) *c*'s being *F* is causally relevant$_{JP}$ to *e*'s being *G*;

(b) *c*'s being *H* is causally irrelevant$_{JP}$ for *e*'s being *G*;

(c) $\Box\ \forall R\ (R \in \Pi_H \supset R \in \Pi_F)$ and

(d) $H \notin \Pi_F$.

Suppose the worlds relevant to the notion of realization are the nomologically possible worlds. Then $\Box\forall R\ (R \in \Pi_H \supset R \in \Pi_F)$ means that $\Pi_H \subset \Pi_F$ in all nomologically possible worlds. Examples of properties *F* and *H* such that *H* does not realize *F* and $\Pi_H \subset \Pi_F$ in all nomologically possible worlds are easy to find since nomological dependence between properties is weaker than realization. Suppose a strict law of the form "all *H*s are *F*s" connects *H* and *F*. This does not entail that *H* realizes *F*. There may be a significant ontological independence between the instantiation of *H* and *F*, and we might not think of *H* as constituting *F* or as amounting to *F*, as we would in the case of realization. Nevertheless, if all *H*s are *F*s, $\Pi_H \subset \Pi_F$. And of course there is nothing incoherent in the claim that *H* and *F* are connected by a strict law and yet *c*'s having *F* is causally relevant$_{JP}$ for an effect while *c*'s having *H* is not.

Assume it is a strict law that all planets move on circular orbits. There is no reason why *being a planet* should be thought of as a realizer of *moving on a circular orbit*. Suppose Hannah has to prepare a list of all objects which have trajectories through space congruent with the ancient Greek's favorite geometrical figure. Hannah includes Pluto on her list. According to D3 Pluto's being a planet is causally relevant$_{JP}$ for its being included, since any way of realizing the former (within the range of nomologically possible worlds) would have resulted in the latter. But Pluto's being a planet is in no sense relevant to its ending up on Hannah's list; it does not matter whether Pluto is a planet. Had it been a spaceship or a satellite it would have been on the list as well.

This shows that D3 and with it J&P's last attempt to ground *PEA* fails, since D3 is not *sufficient* for causal relevance$_{JP}$. Moreover, D3 is not even *necessary* for causal relevance$_{JP}$.

Any account of the causal relevance of irreducible mental properties must account for Fred's desire's being a desire for beer being causally relevant to Fred's going to the fridge to get some beer, since no one can seriously accept that Fred's desire's being a desire for beer is causally irrelevant$_{JP}$. Nevertheless, there are circumstances where D3 renders *being a desire for beer* causally irrelevant$_{JP}$. Suppose Fred has desire for beer, and he knows that there is some *Corona* in the fridge, some *Canadian Dry* in the basement and some *Budweiser* in his study. There is no *invariation of effect under variation of realization*: not any realizer of *being a desire for beer* will

[2] Given that the counterexample to D1 was based upon the fact that all that mattered was the presence of the *actual* realizer, taking into account other possible realizers naturally suggests itself.

[3] This abstract remark captures a straightforward idea: if *F* is causally relevant$_{JP}$ because each of its realizers is causally relevant to *e*'s having *G*, then the instantiation of any property *H* such that $\Pi_H \subset \Pi_F$ will also entail the presence of a realizer property that is causally relevant for *e*'s having *G*.

eventuate in Fred's trip to the fridge. Some realizers of *being a desire for beer* will cause Fred to go into the basement or his study. *Being a desire for beer* does not satisfy D3 and hence fails to be causally relevant$_{JP}$, which I take to be an intolerable consequence, close to a *reductio* of D3.

J&P's attempts to define a weaker notion of causal relevance$_{JP}$ that can be attributed to mental properties even if the only causally relevant properties are their physical realizers fails. If mental property epiphenomenalism is false, it cannot be because *PEA* is true.

Literature

Jackson, F. and Pettit, P. 1988 "Functionalism and Broad Content", *Mind* 97, 381-400.

Jackson, F. and Pettit, P. 1990a "Program Explanation: A General Perspective", *Analysis* 50, 107-117.

Jackson, F. and Pettit, P. 1990b "Causation and the Philosophy of Mind", *Philosophy and Phenomenological Research* 50, 195-214.

Jackson, F. and Pettit, P. 1992 "Structural Explanation in Social Theory", in D. Charles and K. Lennon (eds.), *Reduction, Explanation, and Realism,* Oxford: Clarendon Press, 97-131.