Causal equivalence as a basis for the specification of neural correlates

Uwe Meixner, University of Augsburg

Abstract: The paper defines causal equivalence and uses it to explicate the notion of neural correlate. It is seen that many issues of the mind-body problem can be fruitfully discussed in terms of causal equivalence, or in relation to it. The paper ends by proposing a dilemma that cannot be rationally resolved.

1. Two concepts of causal equivalence

Let E and E′ be events:
(a) E′ is (simpliciter) *causally equivalent* to E if, and only if, (*i*) everything caused by E is also caused by E′, and *vice versa* [substitute the statement obtained by interchanging "E" and "E′" in the previous statement], and (*ii*) everything that causes E also causes E′, and vice versa.
In other words: E′ is causally equivalent to E if, and only if, E′ has the same effects and the same causes as E.
(b) E′ is *event-causally equivalent* to E if, and only if, (*i*) every event caused by E is also caused by E′, and vice versa, and (*ii*) every event that causes E also causes E′, and vice versa.
In other words: E′ is event-causally equivalent to E if, and only if, E′ has the same event-effects as E and the same event-causes as E.

Notes:
(i) *Causal equivalence* entails *event-causal equivalence*, but *not* vice versa – unless, of course, *causation* coincides with *event-causation*; this is believed by many, but, to date, not proven. In what follows, I shall stick to the simple concept of causal equivalence, without argumentatively deciding whether it coincides with event-causal equivalence or not. (Being a believer in agent-causation, I do believe that the two concepts of causal equivalence do not coincide.)
(ii) Every event is causally equivalent to itself; and if an event E′ is causally equivalent to an event E, then E is also causally equivalent to E′; and if an event E′′ is causally equivalent to

an event E′, which is in turn causally equivalent to an event E, then E′′ is also causally equivalent to E. Thus, the relation of causal equivalence is reflexive, symmetrical, and transitive among events, in other words: it is an equivalence relation among events.

2. Two theses of causal equivalence regarding mental events

(I): For every actual mental event E: there is some wholly physical actual event E′ such that E′ is causally equivalent to E.
(II): For every actual mental event E: there is some (at least) partly physical actual event E′ such that E′ is causally equivalent to E.

Note:
Being wholly physical entails *being (at least) partly physical*, but not vice versa. Hence: thesis (I) entails thesis (II), but not vice versa.

3. Two theses of psychophysical dualism

(1) Some actual mental event is wholly non-physical.
(2) Some actual mental event is not wholly physical.

Notes:
(i) Thesis (1) entails thesis (2), but not vice versa.
(ii) *Wholly non-physical* is logically equivalent to *not partly physical*; *not wholly physical* is logically equivalent to *partly non-physical*.

4. Two theses of causal *non*-equivalence regarding mental events

non-(I): For some actual mental event E: there is *no* wholly physical actual event E′ such that E′ is causally equivalent to E.
non-(II): For some actual mental event E: there is *no* partly physical actual event E′ such that E′ is causally equivalent to E.

Notes:
(i) Thesis *non*-(II) entails thesis *non*-(I), but not vice versa.
(ii) Thesis *non*-(I) entails the dualistic thesis (2) (that is: *Some actual mental event is not wholly physical*).
(iii) Thesis *non*-(II) entails the dualistic thesis (1) (that is: *Some actual mental event is not partly physical*, i.e., is *wholly non-physical*).

5. Two research programs for cognitive neuroscience

(A) Seek to *corroborate* thesis (I) (and hence also thesis (II)) to the point of *definitely establishing* thesis (I) (and hence also thesis (II)) – that is, in the sense a scientific hypothesis can be said to be "definitely established". Or, in default of reaching this goal, seek to *corroborate* thesis *non*-(I) to the point of definitely establishing it; if this other effort is crowned with success, the dualistic thesis (2) will have been established.

(B) Seek to *corroborate* thesis (II) to the point of definitely establishing it. Or, in default of reaching this goal, seek to *corroborate* thesis *non*-(II) to the point of definitely establishing it; if this other effort is crowned with success, the dualistic thesis (1) will have been established (and also thesis *non*-(I)).

Notes:
(i) Obviously, each of these two research programs is very significant for the advancement of human knowledge.
(ii) Even if thesis (I) were definitely established, this does not mean that dualism stands refuted, not even in the stronger version represented by thesis (1). The reason for this is this: even if, for a given actual mental event E, there is a wholly physical actual event E´ that is causally equivalent to E (as must be the case according to thesis (I)), it does not follow that E is at least partly physical; E may still be wholly non-physical.
(iii) Matters would be different if one had reason to assume that *causally equivalent actual events are identical*. But whereas it is very plausible to assume that causally equivalent *wholly physical* actual events are identical, it amounts to begging the question against dualism if one assumes that an actual mental event E and an actual wholly physical E´ are identical if they

are causally equivalent. On the contrary, it seems that there are non-causal, "inner" properties – properties that E has (qua mental event), but E´ has not (qua wholly physical event) – which distinguish E from E´ even if E´ is causally equivalent to E.
(iv) Since causally equivalent wholly physical actual events are – *very plausibly* – identical, there cannot be *more than one* wholly physical actual event that is causally equivalent to a given actual mental event. (For suppose that both E´ and E´´ are wholly physical actual events and are *both* causally equivalent to the actual mental event E; hence they are also causally equivalent to each other – the causal equivalence of events being a symmetrical and transitive relation – and *therefore* E´ and E´´ are identical, according to the invoked principle of identity for wholly physical actual events.)
(v) It may also be true that all causally equivalent actual *mental* events are identical; but perhaps some intrinsically different actual mental events are, though non-identical, nevertheless causally equivalent. I leave it as an open question which of these two conflicting hypotheses is true.

6. Neural correlates as causal equivalents?

As a matter of fact, scientists appear to be working already on research program (A). Given an actual mental event E (a visual experience, say), they are trying to find a – in fact: *the* – wholly physical actual event E´ – a brain-event, to be specific – which is *correlated* with E: E´ is called "the neural correlate of E". The methods used in this scientific enterprise are still very coarse, but they may improve with time. What should interest us *here and now* is the following question: what, precisely, is the nature of the invoked *correlation* between E´ and E? Which fact of correlation makes a wholly physical actual event *the neural correlate* of an actual mental event?

A wholly physical actual event E´ is selected as *the neural correlate* of an actual mental event E on the basis of three criteria: first, *spatial location*: E´ is in that region of the brain of *the person with* E where the physical data connected with E are ultimately processed; second, *temporal location*: E´ is simultaneous – or at least approximately simultaneous – with E; third, *unique prominence*: E´ stands out – against all other physical events that are also simultaneous with E and are also located in the relevant brain-region – in a manner that links E´ uniquely with E.

Now, these three criteria for a wholly physical actual event E´ being selected as *the neural correlate* of an actual mental event E all point in the direction of the neural correlate of E being nothing else than the wholly physical actual event that is causally equivalent to E:

the neural correlate of E = *the wholly physical actual event that is causally equivalent to* E.

In fact, the truth of this identity statement is guaranteed if E has *some* neural correlate *and* if the following principle is adopted as a (partial) analysis of the predicate "X is a neural correlate of E":

Principle of Neural Correlation (PNC)
For all actual mental events E:
For all X: X *is a neural correlate of* E *if, and only if,* X *is a wholly physical actual event that is causally equivalent to* E.[1]

PNC seems attractive, one reason for its attractiveness being its relative clearness. But it has some not entirely obvious consequences that do not fit some assumptions that are often made regarding actual mental events and their causal relations to wholly physical actual events. *Let* E *be an actual mental event; then we have*:

Firstly, according to PNC: no neural correlate of E causes E. For suppose X were a neural correlate of E *and* caused E; it follows by PNC that X is causally equivalent to E, which entails, according to the definition of causal equivalence (see sect. 1), that everything that causes E also causes X. But, according to supposition, X causes E. Therefore:, X causes X – which is impossible. Thus we have Theorem 1: *Neural correlates of actual mental events do not cause the mental events of which they are neural correlates*.

Secondly, according to PNC: if there is a neural correlate of E, then E causes some wholly physical event. For suppose X were a neural correlate of E and E caused no physical event; it follows by PNC that X is a wholly physical actual event that is causally equivalent to E, which entails, according to the definition of causal equivalence, that everything caused by X is also caused by E. But, according to supposition, E causes no wholly physical event.

[1] The identity statement (preceding PNC) follows from PNC and the assumption that (some) X is a neural correlate of the actual mental event E as follows: Suppose X is a neural correlate of the actual mental event E. Hence by PNC: X is a wholly physical actual event that is causally equivalent to E. Hence (see sect. 5, note (iv)): *only* X is a wholly physical actual event that is causally equivalent to E. Hence by PNC: *only* X is a neural correlate of E. Hence: X is *the* wholly physical actual event that is causally equivalent to E, and X is *the* neural correlate of E. And therefore: *the neural correlate of* E = *the wholly physical actual event that is causally equivalent to* E.

Therefore, X causes no wholly physical event – which is certainly false, seeing that X is a wholly physical actual event: *every* such event causes some wholly physical event. Thus we have Theorem 2: *No actual mental event with a neural correlate is causally epiphenomenal with regard to wholly physical events*.

Thirdly, according to PNC: if E has a neural correlate and is itself a wholly non-physical event, then this results in a case of genuine causal over-determination. For suppose X is a neural correlate of E and E is a wholly non-physical event; it follows by PNC that X is a wholly physical actual event that is causally equivalent to E. And therefore: since X causes some wholly physical event (see the previous paragraph), E causes that same event, too – a situation which certainly constitutes a case of genuine causal over-determination, *seeing* that X is a wholly physical event and E a wholly non-physical one. Thus we have Theorem 3: *Wholly non-physical actual mental events with neural correlates give rise to genuine causal over-determination*.

Fourthly, perhaps there are no cases of genuine causal over-determination, although one can certainly not exclude its occurrence a priori. In any case, on the basis of Theorem 3 we have Theorem 4: *If there is no genuine causal over-determination, then actual mental events with neural correlates are not wholly non-physical* (but at least partly physical).

Fifthly, we have – and this is perhaps the most significant result – Theorem 5: *If every event that causes a wholly physical event is itself wholly physical, then every actual mental event with a neural correlate is identical to that correlate*.[2] Theorem 5 is the basis of one of the great unresolved philosophical dilemmas of our time: If it is true that every event that causes a wholly physical event is wholly physical, then it cannot also be true that *some actual mental event with a neural correlate is not identical to that correlate* – no matter how much this proposition of *non-identity* may *seem* to be true. If, on the other hand, it is true that some actual mental with a neural correlate is not identical to that correlate, then it cannot also be true that *every event that causes a wholly physical event is itself wholly physical* – no matter how fervently one may wish this proposition of *physical causal closure* to be true from a

[2] The proof is easy: Suppose every event that causes a wholly physical event is itself wholly physical, and suppose E is an actual mental event with the neural correlate X. Hence by PNC: X is a wholly physical actual event that is causally equivalent to E. Since *every wholly physical actual event causes some wholly physical event* (see the deduction of Theorem 2), X causes a wholly physical event Y. Since X is causally equivalent to E, not only X but also E causes the wholly physically event Y, and therefore, according to supposition, E is a wholly physical event. But *wholly physical actual events that are causally equivalent are identical* (see note (iv) of sect. 5). Therefore: E is identical to its neural correlate X.

monistic point of view. You have the philosophical choice; I submit, there is nothing that will rationally determine how you *should* decide. So take your pick.[3]

[3] More on causal representation on the basis of causal equivalence – its relata being *causal representatives of each other* – and a demonstration of the relevancy of these concepts for the mind-body problem can be found in my book *The Two Sides of Being. A Reassessment of Psycho-Physical Dualism* (Paderborn: mentis 2004). There I defend, among other things, a position I call *interactionist parallelism*.