Lowe's Dualist Construal of Mental Causation

Max Kistler

Université Paris 1 Panthéon-Sorbonne
IHPST (Institut d'Histoire et de Philosophie des Sciences et des Techniques)
13 rue du Four
75006 Paris, France
mkistler@univ-paris1.fr

Abstract

It is shown that four of E.J. Lowe's arguments for property dualism, which are based on his analysis of mental causation, establish weak but not strong property dualism. According to the former, which is compatible with physicalism, the property of an event to be a decision differs from its neural property of being constituted by a specific pattern of neural activity. According to the latter, which is incompatible with physicalism, mental properties are, in addition, irreducible to neural properties. The representation of causal relations in the framework of structural equations provides the means to clarify how mental properties differ from neural properties: a precise meaning can be given to the claim that mental causes are *specific* for bodily movements constitutive of actions, whereas neural causes are not. This framework also provides the means to interpret Lowe's thesis that decisions are uncaused, as saying that no cause of a decision is a specific cause.

Keywords

Dualism, physicalism, causation, mental causation, specific cause, action, structural equation, intentionality.

This paper proposes a critical evaluation of a series of arguments of E.J. Lowe's for property dualism that are based on his analysis of mental causation. According to property dualism, some objects, such as human persons, have properties that belong to (at least) two different types: mental and physical. Property dualism comes in a strong and a weak form. Strong property dualism holds, whereas weak property dualism doesn't, that the mental cannot be reduced to the physical. We will get back to what is meant by saying that mental properties of humans are not only different from their physical properties but also "irreducible" to them. However, in whatever way the claim of the irreducibility of the mental

and physical properties is interpreted, the resulting doctrine of property dualism is weaker than substance dualism. Property dualism follows from substance dualism but not vice versa. Lowe has also developed and defended a version of substance dualism. According to his "Non-Cartesian Substance Dualism" (NCSD), "*persons* or *selves* – that is to say, self-conscious subjects of experience and agents of intentional actions – are distinct from their organic physical bodies and any parts of those bodies, such as their brains or central nervous systems" (Lowe 2006, p. 5; italics Lowe's). Contrary to Cartesian substance dualism, NCSD "is a dualism not of *minds* and bodies, but of *persons* – or, more generally, of *subjects of experience* – and their 'organized' bodies" (Lowe 2010, p. 439; italics Lowe's). In his arguments for this doctrine (Lowe 2006; 2010; 2013), Lowe shows in particular that it does not fall prey to the main difficulties of Cartesian dualism, which seems to be in principle incapable of accounting for the causal interactions between the mind or "thinking substance" and the body or "extended substance". No puzzle arises from the causal interaction of persons with their material environment because persons have physical properties themselves.

Here we will take a closer look at some of Lowe's arguments for property dualism. These arguments are also crucial for Lowe's defence of NCSD because property dualism is presupposed in the argument that shows that NCSD can overcome the main weakness of Cartesian dualism and account for the causal interaction between persons and their physical environment. Even if NCSD makes the causal interaction between persons and physical events conceivable in principle, it must be supplemented with the thesis of the causal efficacy of the mental properties of persons. It is indeed essential for our conception of ourselves as persons that some of these interactions, in particular our actions, are due at least in part to our thoughts, and thus our mental properties, and not only to our physical properties. Now, such efficacy is controversial if and only if mental properties are irreducible[1].

E.J. Lowe offers four arguments for property dualism. According to the first, neural causes of several bodily movements constituting actions that are executed in parallel can be entangled, but not their mental causes. Thus, the mental causes cannot be identical with the neural causes. According to the second argument, bodily movements constituting actions are no coincidences. However, if they had only neural causes, they would be coincidences: there are always several independent chains of neural events that converge so as to contribute to

[1] Kim (1998) has made a strong case that mental causation, i.e. the idea that persons can change their physical environment by virtue of their mental properties, is incompatible with a set of plausible metaphysical hypotheses, such as the so-called causal closure of the physical domain, according to which every physical event has, at each time preceding it, a complete exclusively physical cause. On the causal closure principle, see Lowe (2000).

cause a given bodily movement. Such a convergence of independent causal chains is the mark of a coincidence. According to the third argument, decisions and their neural underpinnings have different causal profiles. Therefore, they cannot be identical. According to the fourth argument, decisions are intentional, in the sense that they cause an action as an intentionally characterized type of event. Neural events are incapable of that kind of causation and can only cause the action as the particular event of the corresponding bodily movement. Therefore decisions cannot be identical with any neural event or set of neural events.

In what follows, I will look more closely at each of these arguments and show that they are indeed successful in establishing a form of property dualism. However, it will turn out that the form of dualism that is justified by the premises of Lowe's arguments is weaker than the form of dualism Lowe himself claims to have established. The crucial difference between the weak form of dualism that follows from the premises and the strong form that Lowe tries to establish is that the weak form is, whereas the strong form is not, compatible with reduction, i.e. with the possibility of finding a scientific explanation of the possession of the mental properties, on the basis of neural and other scientific properties. Though weaker than what Lowe himself claims, the result of Lowe's arguments is substantial and important. Indeed, weak property dualism is incompatible with eliminativism and type identity. We will see that Lowe's arguments can be expressed in the framework of the recently elaborated theory of causation in terms of structural equations. This framework allows giving a precise meaning to claims to the effect that certain properties are causally relevant; and it is connected to empirical methods that can be used to justify such claims on the basis of experiments. The conceptual tools of the structural equations framework may also help clarify another strong and important thesis of Lowe's: that decisions are uncaused.

Lowe's first argument for property dualism is best presented with the help of the diagram that accompanies his own presentation.

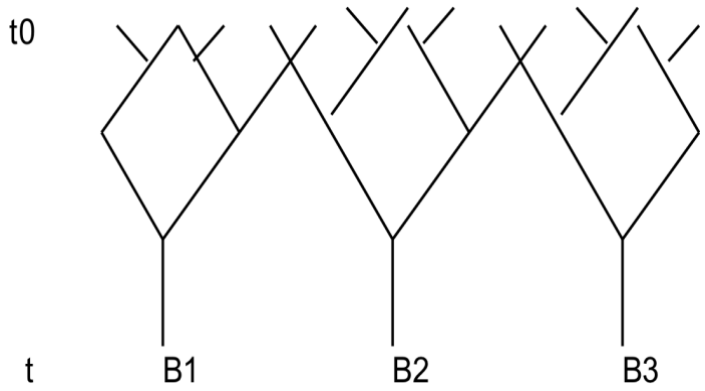

Figure 1. Schema of entanglement of neural causes of 3 simultaneous actions. From Lowe (1993, p. 639; 1996, p. 85)

In fig. 1, the vertical axis represents time, the direction of time pointing downwards. Lowe suggests the following thought experiment. At time t, a person simultaneously executes three different actions B_1, B_2, and B_3, i.e. more precisely, the bodily movements that are constitutive of those actions.

Now the following situation seems to be empirically possible. If we trace back in time the neural causes of three different voluntary bodily movements B_1, B_2, B_3, which are executed at the same time t, we may find that, as we follow the paths of all the causal chains converging on each of the bodily movements B_i in the reverse direction of time, we will get to a certain time t_0 earlier than t, at which the causal paths are entangled. More precisely, the empirical possibility we are contemplating consists in the fact that some single-neuron events occurring at t_0 or before are part of the causal antecedents of more than one of the movements B_i. If this is indeed the case, it is plausible to suppose that the entanglement gets the more widespread the more we go back in time. The earlier the time t_i before t_0, the more single-neuron events will be part of the complete neural causes of all three bodily movements. Let us call a "complete cause of B_i at t_i" the set of all events that take place at t_i and lie on some causal path or other that leads to B_i at t. Then the sketched situation is such that at all times t_i earlier than t_0, the complete physical causes of B_1, B_2 and B_3 at t_i overlap, i.e. have common parts. Now, says Lowe, given that B_i are bodily movements constituting actions, they must have mental causes. It is part of the concept of an action to be caused by a decision (or a volition) of a subject. However, the mental causes of B_1, B_2 and B_3, i.e. the decisions or volitions to execute them, do not overlap. Decisions cannot overlap or share parts with other decisions, simply because decisions have no "parts" at all. Lowe accepts the hypothesis, which constitutes an additional premise of the argument, that, for each of the bodily

movements B_i, the decision D_i that is its mental cause takes place at some time t_i earlier than t_0. Then it follows for each B_i that the decision (at t_i) to act by bodily movement B_i (at t) is not identical with the complete physical cause (at t_i) of B_i. "We could discover that whereas the putative mental causes of token bodily movements are distinct and separable, the neural causes are inextricably entangled" (Lowe 1993, p. 639; Lowe 1996, p. 84).

Without providing any further argument, Lowe then draws a stronger conclusion. He claims that the argument just sketched not only shows that the decision D_i for B_i, which takes place at t_i, is not identical with the complete physical (or neural) cause of B_i, but also that the causal powers of this decision "are not wholly grounded in (or 'causally reducible to') the causal powers of those elements of the system which produced it" (Lowe 1993, p. 636/7; 1996, p. 82). The claim that the decision is not only different from the underlying set of neural events but also irreducible to it suggests that it is independent of them.

Lowe's argument is open to the following two objections (Kistler 2005). One objection is directed against the premise according to which the model sketched in fig. 1 is realistic, i.e. that there are real situations that share the causal structure of this model. The other objection is that the premises do not entail Lowe's stronger conclusion that the mental cause of an action is irreducible to its complete physical cause.

According to the first objection, it is implausible that there are situations involving real human subjects that have the structure of the model sketched in figure 1. The following empirical considerations raise doubts as to whether the model ever applies to any real human decisions. Psychological research shows that there are limitations to the human capacity of performing two or more tasks simultaneously. "Is it possible that with increased practice and skill we can do as many things as we like at the same time? The answer is probably 'no'" (Smyth et al. 1994, p. 153). The extent to which it is possible to do two things at the same time depends on many factors, such as the similarity of the tasks, the difficulty of the tasks, and the amount of automaticity that has been reached by a subject in performing the tasks. No doubt, subjects can learn to perform tasks simultaneously although initially the execution of one task disrupts the execution of the others. Beginning drivers have difficulties in driving and talking simultaneously without suffering from interference in each of these activities, while more expert drivers have acquired that capacity. However, learning to perform (what appears to be) two things simultaneously seems to be possible only in two ways, neither of which complies with Lowe's model.

Either the causal pathways leading to the execution of what is initially described as two tasks are entangled but the subject really performs one complex task rather than two

independent tasks, or there are really two tasks performed in parallel but the causal pathways leading to their execution are not entangled. Here is a case of the first type. A piano player learns to execute qualitatively different movements with her two hands. The movements may in particular differ in terms of rhythm. She accomplishes this by merging "the two elements into an integrated whole so that there is no longer a dual-task situation" (Smyth et al. 1994, p. 129). Several studies of the capacity of expert musicians to execute different rhythms with the two hands have led to the hypothesis that "a central counting mechanism is required from which the complex counting of the skilled musician must be derived" (Smyth et al. 1994, p. 130). If this correctly describes situations in which a subject executes what looks like two actions, there really is just one integrated action executed by the subject. It is one action with two parts, containing two different subroutines that control the movement of each hand.

Here is a case of the second type, in which two actions are really carried out in parallel without getting integrated into a single complex action. "Tasks can be carried out at the same time provided it is possible to prevent each processing system (sometimes known as a module) from picking up interfering by-products from other processing activities" (Smyth et al. 1994, p. 152). Doing two things at a time requires separating the cognitive resources used in executing the two actions. The expert chess player who chooses her move in a "blitz" game in a second has learnt to organize her capacities efficiently: she draws directly on her memory of similar positions instead of going through lengthy calculations of alternative possibilities. This frees her processing capacities for other simultaneous tasks (Smyth et al. 1994, p. 150). If a subject executes two actions in parallel in such a way that at least one of the actions is controlled by a modular mechanism[2], so that its execution is controlled automatically and unconsciously, the causes of that automatic action are not entangled with the causes of other actions executed in parallel. It is part of the concept of a cognitive module that the execution of its tasks is "encapsulated", i.e. causally isolated from processes going on outside the module.

In conclusion, it seems that neither of these ways in which a person is able to do "two things at once", complies with Lowe's model. Either the causes are entangled but those causes belong to the history of a single complex action which contains two subroutines or the subject really carries out two different actions in parallel but at least one of these actions is executed by a modular mechanism that precludes entanglement of the action's causes with any causes of the actions carried out in parallel.

[2] For the definition of a cognitive module, see Fodor (1983).

Let us now put this first objection to one side and accept the hypothesis that there really are situations that have the structure sketched in figure 1. The second objection against Lowe's argument is that even if some cases of human simultaneous actions had the structure of Lowe's model, it wouldn't follow that decisions are not grounded in neural events, i.e. are not reducible[3] to neural events. To show this, consider the opposition Lowe draws between the irreducible and "strongly emergent" (Lowe 1993, p. 636) causal powers of the decision to carry out some action and the "weakly emergent" causal powers of a chemical substance, such as "the causal powers of liquid water" (Lowe 1993, p. 636; 1996, p. 81). To say that the causal powers of water are emergent means that they belong to macroscopic samples of water but not to its molecular parts: Individual H_2O molecules are neither transparent nor even liquid (at room temperature and standard pressure)[4]. To say that these causal powers of a macroscopic sample are only *weakly* emergent means that they are novel with respect to its molecular components but nevertheless "explicable in terms of the causal powers and relations of its constituent molecules" (Lowe, *ibid.*). The macroscopic powers are grounded on the powers of the microscopic constituents. Together with a sufficiently strong theory of the macroscopic and microscopic powers and the dependence of the former on the latter, the former can be reductively explained in terms of the latter. Now the question is this. Do we have reasons to think that the causal powers of decisions are more strongly emergent with respect to the powers of the underlying neural events than the powers of water with respect to the powers of its component molecules? Saying that an event (or a property of an event) is strongly emergent means that it is 1) emergent and 2) cannot be explained in a reductive way.

The premises of Lowe's argument, i.e. the hypothesis that there are real situations in which a human subject executes several simultaneous actions B_1, B_2, B_3, and that its neural but not its mental causes are entangled, are compatible with the hypothesis that each of the decisions for the actions B_i is *weakly* emergent with respect to the underlying pattern of neural events. In that case, the decision is not identical with any pattern of neural events but nevertheless reducible to it. To say that one property or event is reducible to another means that there is a reductive explanation showing that the former is a consequence of the latter,

[3] Lowe himself doesn't distinguish the two claims that decisions are not grounded on sets of neural events and that they are not reducible to them. Given that there are several models of both reduction (van Riel and van Gulick 2019) and grounding (Bliss and Trogdon 2016), it is a substantial and controversial claim that both claims stand and fall together. In what follows, I will take grounding to be necessary but not sufficient for reduction. Reducibility is epistemological whereas grounding refers to the metaphysical dependence relation between two properties or laws, which makes the reductive explanation of one in terms of the other possible.

[4] Needham (2017, p. 135-7) argues that not only H_2O molecules but even their parts are water, as long as they are within a macroscopic sample of water, against Quine (1960, p. 98) who holds that the atomic parts of H_2O molecules are not water.

given a theory of the lawful relations between the two domains, in this case between neural events and processes and the corresponding psychological events and processes. An explanation is a human activity but it can only be true if there exists a natural relation of determination between the events referred to in the explanans and the event referred to by the explanandum. A property *P* of a complex object is naturally determined by the properties of its parts and relations among its parts, if and only if there are laws of nature to the effect that every complex object whose parts have such properties and stand in such relations, has *P*. This equivalence can also be expressed in terms of events rather than properties: A complex event is naturally determined by certain events that are its parts and by certain relations among these parts, if and only if there are laws of nature to the effect that every complex event that has such parts standing in such relations, has *P*. Knowledge of those laws can then be used to provide reductive explanations of the former in terms of the latter.

In our hypothetical situation, this means that the premises of Lowe's argument are compatible with the hypothesis that 1) the decisions D_1 for B_1 and D_2 for B_2 differ from each other, 2) that each of these decisions can be reductively explained by a set of neural events together with psychophysical laws, and 3) that the reduction basis of D_1 – the conjunction of neural events underlying D_1 – overlaps, i.e. shares a part with the reduction basis of D_2 – the conjunction of neural events underlying D_2. In this situation, the decisions D_1 and D_2 are different from the sets of neural events on which they are grounded (and which overlap) but they are only weakly emergent with respect to them.

Here is an analogous situation in physics. Metals can conduct both electricity and heat. The electrical conductivity σ is a different property of a piece of metal than its thermal conductivity κ; these properties correspond to different causal powers of the metal. Its having a certain electrical conductivity gives a piece of metal the power to let, in appropriate circumstances (if a voltage gradient exists across the metal) current flow through it. Its having a certain thermal conductivity gives it the power to let, in appropriate circumstances (if a temperature gradient exists across the metal, together with a heat source and a heat sink on opposite sides) heat flow through it. Both are naturally (non-causally) determined by (or grounded on) overlapping aspects of the microphysical structure (electronic configuration of component atoms). More precisely, the electrical conductivity σ is determined by (among other things) the unit electric charge of electrons, but not by (what determines) the temperature of the metal, whereas the thermal conductivity κ is determined by (among other things) (what determines) the temperature of the metal, but not by the unit electric charge of

electrons. Some microscopic features enter in the determination of both: n (the number of free electrons per cubic centimetre), τ (the relaxation time or mean free time of the free electrons, i.e. the average time interval between two collisions), and m (the mass of an electron). This situation corresponds to weak emergence (determination of the properties σ and κ of the macroscopic metal by properties of its microscopic constituents, together with reducibility) of two different characteristics of a macroscopic object, where the underlying microscopic bases to which these characteristics can be reduced overlap, i.e. share a common part.

Given that the physical situation shares the structure sketched in fig. 1 with the situation in which two decisions are grounded on overlapping sets of neural events, and given that a situation with this structure is compatible with reduction in the physical case, we would need to be given additional reasons for which the situation of the same structure is not compatible with reduction in the psychophysical case.

To conclude the analysis of Lowe's first argument for property dualism, the argument shows that mental properties such as decisions for actions are "weakly emergent". This is a substantial result. It is indeed incompatible both with eliminativism and with the type-identity of mental properties with the underlying neurophysiological properties. However, the argument fails to show that mental events are emergent in any stronger sense than the transparency of water, or the electric or thermal conductivity of metals. The argument establishes that decisions are emergent but not that they are strongly emergent, in the sense of being irreducible.

Here is Lowe's second argument for property dualism. Bodily movements constituting actions are not "coincidental". To see what it means to say that a given event is a coincidence, consider a single action, consisting in bodily movement *B*. Let us accept the plausible hypothesis that the neural causes of *B* are structured in a tree-like pattern as sketched in fig. 2, where the tree is upside-down, with roots on top and trunk pointing downwards.

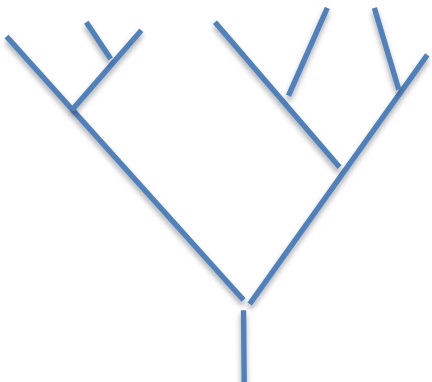

Fig. 2: Part of fig. 1. Sketch of the structure of neural causes of the bodily movements constitutive of an action B, with the direction of time pointing downwards.

In the model sketched in fig. 2, several paths (chains of events) converge so as to cause *B* at *t*. Event *B* is said to be a coincidence if and only if there are, at a given time t_i earlier than *t*, several events that are causes of *B* but independent of each other, in the sense that they do not have any common causes at any time t_j earlier than t_i. "An event occurs by coincidence (…), when two or more events co-occur and cause that event, but those causes are themselves causally independent, in the sense of having no common cause among their various causes" (Lowe 1999, p. 230/1). An event *B* whose causal history has the structure sketched in figure 2 has a causal explanation at the neural level, but this causal explanation makes it appear as a coincidence. However, a bodily movement constituting an action is no coincidence. It follows that it must have another additional complete cause, which is in itself sufficient for *B*, and which is part of a unique chain that has, at a given time, no parts that are independent of each other. Decisions and volitions are mental causes that satisfy this condition for making their effects non-coincidental. "A mental state causes some physical event" by "rendering that event non-coincidental, which it can do by rendering non-independent the causal histories of that event's immediate physical causes" (Lowe 1999, 236).

There are (at least) two ideas here. One is that B is not a coincidence if it is caused by a decision although it might wrongly appear to be a coincidence to someone who ignores B's mental cause and only considers its neural causes. The second idea is that the decision "renders non-independent the causal histories of B's immediate physical causes". Lowe doesn't say how exactly the mental cause can not only cause *B* in such a way that *B* is no coincidence but also somehow modify the pattern of *B*'s neural causes, so that the different neural causes of *B* cease to be independent of each other. Whether or not such an explanation

can be found, the argument shows that every bodily movement *B* constitutive of an action has a mental cause, occurring at time t_i, that is not identical with the set of *B*'s physical causes occurring at time t_i. However, nothing in this argument justifies the claim that this mental cause is irreducible to the underlying pattern of neural events.

Here is Lowe's third argument for property dualism. Let us accept the hypothesis that bodily movements constituting actions have both neural and mental causes, i.e. decisions or volitions. Let us say that *B* occurs at *t* and has at t_0 (earlier than *t*) both a mental cause *D* and a neural cause *N*, which may consist of a set of neural events, where nothing is presupposed about whether N and D are different or identical. Lowe argues that the decision *D* is not identical with *N* because these events have different causal powers and make true different counterfactuals. The causal powers of *D* can be analysed in terms of a set of counterfactuals indicating what would happen if *D* occurred and if *D* did not occur; the same holds for the causal powers of *N*. Intuitively, there seems to be a difference between a counterfactual possible world that is as close as possible to the actual world but in which *N* doesn't occur, and a counterfactual possible world that is as close as possible to the actual world but in which *D* doesn't occur. The difference is that intuitively, the closest possible world to the actual world in which *N* is lacking is very close to the actual world because *N* is a neural event with very specific identity conditions. That possible world differs only in a very minimal way from the actual world containing *N*, i.e. it contains a neural event *N'*, which resembles *N* very much. The difference between *N* and *N'* may be the activation of a single neuron, which is active in *N* but not in *N'*, or vice versa. Now it is plausible on empirical grounds that most bodily movements are not very sensible with respect to the fine detail of their neural causes. It is plausible from a biological point of view that a mechanism with a certain amount of redundancy is better adapted than an extremely sensitive mechanism in which a difference in activation of a single neuron makes a difference to the bodily movement that is performed. To say that the neural mechanism that causes *B* is not very sensitive means that a very small difference at the level of neural activity, such as a difference with respect to the activity of a single neuron, doesn't make a difference to the execution of *B*. To express the same idea in counterfactual terms: If *N* had not happened, an act of type *B* would have happened nevertheless. This means that, in the closest possible world lacking *N*, *B* still occurs. If *B* is my action of raising my arm, in order to, e.g., hail a taxi, "if *N* had not occurred, my arm would still have risen in almost exactly the same way as it actually did" (Lowe 2006, p. 14; similarly Lowe 2010, p. 456). However, the identity conditions of the mental cause *D* of

the same bodily movement *B* are very different. Mental causes such as decisions are not insensitive to psychological differences. The difference between the decision to raise my arm slowly and the decision to raise it quickly may be a small difference but it is enough to make a difference in its behavioural effect. Contrary to neural causes, psychological causes are difference makers for their behavioural effects. They lack redundancy. In counterfactual terms, this means that if I had not taken decision *D*, I would have acted in a way very different from *B*. The closest world without *D* is far more distant from the actual world than the closest possible world without *N*. It follows that *D* and *N* differ by their causal (and explanatory) profile. Therefore, *N* and *D* are not identical.

One way to express the idea that the neural cause *N* and the psychological cause *D* differ in sensitivity, in the sense that small changes in *N* are biologically and behaviourally indifferent (make no difference to behavior) whereas any small change in *D* makes a behavioural difference, is by saying that *D* describes the state of the person in a way that is more abstract than the way the neural predicate describes the state of the person (Yablo 1992). This conception of the respective roles of the predicates *N* and *D* makes it possible to conceive both of them as referring to causes, occurring at the same time t_0, of the same bodily movement *B* at *t*. Both *N* and *D* refer to a property that the person possesses at t_0 and which can be used in a causal explanation of *B*. However, these explanations are not equivalent. Here is Lowe's way of spelling out the difference. The neural explanation in terms of the predicate *N* causally explains *B as a particular event*, whereas the psychological explanation in terms of the psychological predicate *D* provides a causal explanation of a more abstract kind, which explains the bodily movement as a movement of a certain abstract kind that corresponds to a type of action. "Mental states (….) provide causal explanations of certain general physical states of affairs and not merely of particular physical events." (Lowe 1999, p. 236) Psychological explanations in terms of psychological predicates have a specific utility. "A mental cause can (….) explain why an effect *of that kind* occurred, not merely why *that particular event* occurred" (1999, 236; italics Lowe's). Neural explanations are more fine-grained in the sense that they can explain why the precise bodily movement *B* occurred. However, this is in general not what we are interested in explaining: our explanatory interest typically lies in a movement of a certain abstract type, i.e. the explanandum is the fact that the movement belongs to a certain type of movement, say a movement of raising one's arm. The neural explanation of *B* may explain something more detailed, e.g. why *B* consists in raising the right arm and not the left arm. But this is a too fine-grained explanation with respect to

our request for explanation. The psychological explanation is typically geared exactly to the level of abstraction that corresponds to our conception of the effect. To say that the psychological causal explanation is at the same level of abstraction as the conception of the explanandum behaviour *B* means that *D* is not only sufficient but also necessary for *B*. If *D* had not occurred, I would not have done *B* but nothing or something of a different type of behaviour. *N* on the other hand is sufficient but not necessary. *N* describes the state of the person at the moment of the decision in less abstract and more detailed terms, so that a behaviour of type *B* would have taken place even if the detailed neural cause had not occurred.

The fourth argument for property dualism highlights one particularly important aspect of the specificity of mental causes that distinguishes them from the underlying neural events. Mental causation is intentional causation. The effect of a decision is the occurrence of a behaviour of a certain kind; and this kind is the intentional object of the decision. Only psychological causes can have intentional objects[5]. Lowe accepts that the behaviour *B* occurring at *t*, which is caused by a decision occurring at t_0, also has a neural cause *N* occurring at the same time t_0. However, as a neural cause, it doesn't have any intentional content. It cannot cause *B as a bodily movement characteristic of a certain type of action*. Rather, the neural event *N* underlying the decision *D* causes movement *B* as a particular bodily movement (Lowe 2006, p. 16; 2010, p. 458/9).

If we ask for an explanation of the behaviour as being of a certain psychologically described type of action, rather than being this particular instance of the type with its contingent detailed properties, the appropriate cause to mention is the fact that the person decided to do *B*. Lowe offers the following example to illustrate the difference between the two sorts of explanation. Consider the last stroke in a snooker game, which results in the complex event of all the remaining balls on the table falling into pockets. If more than one ball fall into pockets, we may describe the complex event *E* of all remaining balls falling into pockets, as the fusion of the partial events E_i, where E_i represents the event consisting in the i-th remaining ball falling into a pocket. The intention of a typical snooker player is to bring about an event of the abstract type "to pot all the coloured balls remaining on the table" (Lowe 1999, p. 237). When we look for a causal explanation of why an event of type *E* has

[5] Davidson (1970) has famously defined mental predicates in terms of intentionality. Whether or not all mental events are intentional, as Brentano thought, is controversial. It is also controversial whether there is a naturalistic theory of intentionality, as Millikan (1984) and Dretske (1988) claim. In the context of evaluating Lowe's argument, we only need the premise that some mental events, such as decisions and volitions, have an intentional content.

happened, we are looking for an explanation that is geared to the level of abstraction of the type *E*, in the sense of being both necessary and sufficient for the occurrence of an event of this type. The decision to pot all the remaining balls is a cause of just that level of abstraction. The intentional content of the decision corresponds exactly to the level of abstraction at which we conceive the explanandum. The fact that the type of event that plays the role of the cause lies at the same level of abstraction as the type of event that plays the role of the effect makes the explanation robust: "The way in which a mental cause interconnects chains of physical causation is such as to ensure that the common effect of such chains is (….) *robust*" (Lowe 1999, p. 237; italics Lowe's). Saying that the effect is robust means that the occurrence of an event of type *E* does not depend on the fine details of the cause event. Even if some of the physical causal chains leading to some of the partial events E_i had been a little different, or if there had been less balls, the particular event *E* would not have occurred but an event of the same kind "all remaining balls fall into pockets" would still have occurred.

Dennett (1981) has constructed a thought experiment in which intelligent Martians do not interpret the movements of our bodies as parts of rational actions that can be understood and explained in terms of reasons. Instead, Dennett's Martians adopt, with respect to the humans they encounter and whose movements they intend to explain and predict, the "physical stance". They might succeed in calculating the exact trajectory of each elementary particle constituting the body of a human following some course of action. They might be able to provide a scientific explanation of the movements of the person's body in purely physical terms, on the basis of the fact that the evolution of the particles constituting her body conforms to the laws of physics. In a sense, the Martian causal explanation of, say, my raising my arm, is certainly worse than the explanation you may give in terms of my intention to hail a taxi. The Martian explanation is too fine-grained and not robust enough to be applicable in other similar situations, whereas your explanation has that quality. In Dennett's words, by adopting the physical stance and explaining our behaviour in terms of purely physical causes, the Martians would miss a "real pattern" (Dennett 1990). When we adopt the intentional stance, we explain a given behaviour as belonging to a psychological type of event, which may recur and allow for an explanation of the same type. With respect to the example of the last stroke of the snooker game, Lowe says that "no purely physical explanation of all of the sub-events of which *E* is the fusion can provide an interesting explanation of this sort" (Lowe 1999, p. 238). With respect to a physical explanation of the movements of the snooker player of the sort Dennett's Martians might give, he says that "such a purely physical explanation makes *E* appear to be a merely coincidental event and a 'fluke', in the sense that it provides us

with no rational expectation that an event of *this kind* would still have occurred even if many of the individual movements of the balls had been rather different" (Lowe 1999, p. 238). In conclusion, Lowe's 3d and 4th arguments establish that mental properties differ from neural properties; however, these arguments say nothing about the issue of the reducibility of the former to the latter.

The crucial issue now is how best to make sense of this result. According to Lowe, only dualism can make sense of abstract causes, such as decisions to raise one's arm, and abstract effects, such as performing an action belonging to the abstract type of raising one's arm. "My opponents are in the grip of an unduly simple concept of causation – one which admits only of the causation of one event by one or more chains of causation. (…) Since this is the only sort of causation recognized by the physical sciences, intentional causation on the NCSD model is bound to be invisible from the perspective of such a science" (Lowe 2006, p. 19; similarly Lowe 2010, p. 461).

Lowe's diagnosis of his opponents' (i.e. physicalists who deny property dualism) conception of causation is certainly oversimplified. First of all, the thesis that the only adequate concept of causation is event causation is not scientific but philosophical, and has been developed by such philosophers as Hume and Davidson rather than by scientists. But secondly and more importantly, there is a conception of causation, in terms of types of events, which provides the means of accounting for the premises of all of Lowe's arguments without abandoning physicalism. An event such as an action can either be conceived as a particular occurrence whose identity is defined by the region of space and the interval of time that it occupies or as an event belonging to a type. The same event belongs in general to many different types. In particular, it belongs to several types that are more or less determinate, or more or less abstract. The event occurring at time t, which is an act of raising my arm, is also an act of raising my left arm, and an event of raising my left arm in two seconds. The event at which, at t_0, I take the decision to raise my arm is both an event of the type of taking that decision and an event that belongs to many neurological types. It is an event of the type characterized by my brain being in a precise neural state N, which characterizes the state of activity of each single neuron in my brain, but also an event of many more abstract types of states of activity of my brain.

The first two of Lowe's four arguments we have considered show that the decision, at t_0, to perform a certain action at t is not identical with any neural event at t_0. The issue of whether a decision is reducible to some underlying neural event remains open. The third and fourth arguments show that, even if decisions have neural properties over and above their

psychological properties, psychological and neural properties have different causal powers. These differences can be interpreted by supposing that psychological and neural types categorize the same event in a more or less abstract way, the neurological category being less abstract than the psychological. Once more, nothing is said about the reducibility or not of the psychological type in terms of the neural type. To say that an event belongs to a neural type or to a psychological type is just another way of saying that it has a neural or a psychological property. The weak property dualism that is established by Lowe's arguments is equivalent to the thesis that the relevant psychological categories are distinct from all neural categories.

The thesis that these different categories correspond to different causal powers, as it is shown by the third and fourth argument, can be made clearer with the help of a powerful formalism that has been elaborated in recent decades, in which causal relations are represented by means of structural equations (Pearl 2000, Spirtes, Glymour, Scheines 2000, Woodward 2003). I can here only present the fundamental ideas of the formalism, following Halpern (2000) and limiting myself to deterministic models with a finite number of dimensions. The construction of a structural equations (SE) model proceeds in three steps.

In the first step, the system under study is represented by a finite set of variables, corresponding to the predicates characterizing the relevant features of the system. There are two sorts of variables: Endogeneous variables are such that their values are determined by other variables within the model, whereas the values of exogeneous variables are determined in a way that is independent of the other variables of the system. The structural equations describe the functional dependence of the endogeneous variables on other (endogeneous and exogeneous) variables in the model. So this first step of the construction of the model consists in determining the set of exogenous variables U, the set of endogenous variables V, and a set of functions R associating with each variable Y a non-empty set R(Y) of values.

The situation of a person p's raising her arm can be modelled with an exogenous variable D0 representing the person p's decision to raise her arm and the endogenous variable R1 representing the raising of p's arm. The fact that D0 takes the value d0+ represents the fact that p takes the decision to raise her arm, whereas D0=d- represents the fact that she doesn't take that decision. Similarly for R1: The fact that R1 takes the value r1+ represents the fact that p raises her arm, whereas R1=r1- represents the fact that she doesn't.

In a second step, the dependence relations among the variables introduced in the first step are represented by a set of structural equations. The simplest model representing the situation of our example consists in the equation R1=D0. It expresses the assumption that the question of whether (R1=r1+) or not (R1=r1-) p raises her arm is perfectly determined by a

single factor, i.e. the value of D0. If D0=d0-, then R1=r1-, and if D0=d0+, then R1=r1+[6]. Our model expresses the hypothesis that the raising of p's arm is a function of p's decision to raise her arm. The fact that the representation of the dependence is functional expresses the supposition that it is deterministic and without exceptions: the same type of event (D0) is always associated with same result (R1)[7].

Instead of representing the causal dependence of R1 on D0 with the structural equation R1=F(D0), it can also be represented with a diagram where nodes represent endogeneous and exogenous variables and arrows between the nodes represent causal relations. In our case, there are only two variables and one causal relation between them:

D0→R1.

The formalism of structural equations is powerful enough to represent Lowe's observation that the decision *D* has the causal and counterfactual profile that makes it adequate to mention in a causal explanation of *R*, whereas the underlying neural state *N* does not. The relevant difference between the functions expressing the dependence of *R* on its neural and psychological causes, *N* and *D*, can be expressed by a mathematical criterion.

R is a function of both its neural cause *N* and of its psychological cause *D*. However, the latter function R=F(D) is *injective*, whereas the former function R=F(N) is not. Saying that R=F(D) is an injective function means that each value of D is associated with a different value of R[8].

r+ = arm raised, r- = arm not raised

d+ = decision taken, d- = decision not taken.

If we use arrows to represent the associations between values of variables related by causal influence, the values of the variables related by an injective function are associated one by one by such arrows.

d+ → r+

d- → r-

---

[6] This model is of course very much oversimplified because it does not represent the many other factors that may influence whether a person raises her arm at a given moment. The decision may be overridden by interference of external or internal factors, so that the arm does not rise although the decision has been taken (D0=d0+ but R1=r1-), and the arm may raise for reasons independent of the decision, so that R1=r1+ although D0=d0-.

[7] In a third step a value is assigned to each of the external variables. In our model, it simply consists in attributing to the exogenous variable D0 one of its two values, d0+ for situations in which the decision is taken and d0- for situations in which it is not taken.

[8] The idea of analyzing specific causation with the help of the concept of an injective function is due to Woodward (2010, p. 305). My account differs from Woodward's in that in his analysis, a function is specific if it is both injective and surjective, whereas I have used a weaker notion that requires only injectivity but not surjectivity. A function Y= f(X) surjective if and only if, for every value y_i of Y there is some value x_j of X such that $y_i=f(x_j)$. The concept of specific causation defined in the text is similar to Yablo's (1992) "proportional" causation, Lewis' (2000) causation as influence, and List & Menzies' (2009) "realization-insensitive" causation.

I suggest calling a dependence relation that can be represented by an injective function a specific causal influence. In that case, each difference in values of the cause value corresponds to a difference in values of the effect variable. By contrast, in a non-specific model of the causal dependence between two variables, it is not the case that each value of the cause variable is associated with a different value of the effect variable. In other terms, there are values ni≠nj of the cause variable that are associated with the same value of the effect variable.

Let us suppose that the causal dependence of R on N is also represented by a function R=F(N), which is however not injective: different values of N are mapped on the same value of R. A diagram where causal relations between variables are represented by arrows would contain both an arrow from D to R, D→R, and an arrow from N to R, N→R. The difference between the specific causal relation from D to R and the non-specific causal relation from N to R is conspicuous at the level of the association of the values of these variables. If the cause variable N has values n_i, and the effect variable R, has values r-, r+, the non-specificity is represented by the fact that different values $n_i \neq n_j$ are associated with each of the values of R, as in figure 3.

The fact that R is an injective function of D whereas R is not an injective function of N is a clear mathematical expression of the fact that the decision D is specific for behaviour of type R, whereas the pattern of neural events N is not specific for R. It is the association of values in a one-one way that explains why the choice of D is more appropriate for a causal explanation of R than the choice of N.

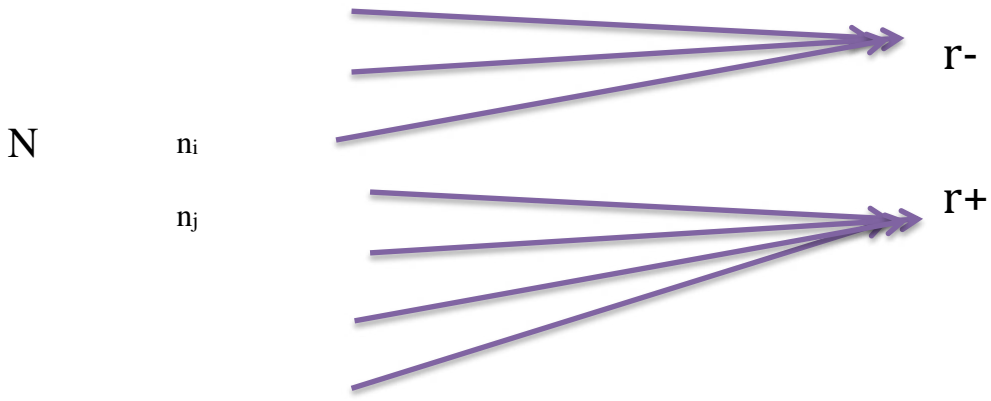

Fig. 3. Schema of the function R(N) that is not injective, i.e. which associate different values ni ≠ nj to the same value of N.

Let us take stock. We have seen that all four of Lowe's arguments analyzed above are compatible with weak property dualism, where differences between properties can be interpreted in terms of differences between causal profiles. Causal profiles of properties can be expressed by structural equations of the variables representing the properties. The first two arguments show that the cause of a given effect belongs in general to more than one category, so that there are several true type level causal judgments about the same token causal process.

At the level of events, there is just one causal relation between the event at which the decision D is taken and the event of the bodily movement B. However, a response to a request for a causal explanation must go beyond the identification of that causal relation. Every such request asks for the identification of two types of events D and B, such that B is functionally dependent on D. Type B is fixed by the explanatory question. I want to know why your bodily movement belongs to a certain type of action, e.g. why you raise your arm, or why you raised your left arm in 2 seconds.

There are in general several types T of events to which the cause belongs and which are such that there is a functional dependence between B (the type chosen to be relevant by the question) and T. In the situations considered in Lowe's arguments, there is a functional dependence both between the fact that I did something of type B and the fact that I took decision D, and between the fact that I did something of type B and the fact that my brain was in neural state N. The premises of the first argument can be interpreted as showing that the same event that causes B belongs to both a mental type D and a neural type N, which have different properties: N can be entangled with neural causes of another action B2 executed in parallel, whereas D cannot be so entangled. The premises of the second argument can be interpreted in this framework as expressing another important difference between the neural and mental types to which the cause of B at time t_0 belongs. The neural type has parts that are independent of each other, in the sense of having no common cause, which makes the effect appear as a coincidence B with respect to its neural cause N, whereas the psychological type does not have any such parts. For this reason, B does not appear as a coincidence insofar as it is caused by the decision D.

According to the third argument, decisions and their neural underpinnings have different causal profiles. Here is a way to represent this fact in a framework that acknowledges (weak) property dualism, without abandoning physicalism or the possibility of reduction.

The difference between the two dependence relations that are highlighted in Lowe's third argument can be explained by the difference in the type of functional dependence: B is

an injective function of D but not of N. The counterfactual scenarios considered by Lowe can be interpreted as scenarios in which the value of the cause variable is modified. This always makes a difference in the case of an injective function, such as the function characterizing the dependence of the action on the decision, but does often make no difference in the case of a non-injective function, such as the function characterizing the dependence of the action on the detailed state of neural activation of my brain N.

The same difference is highlighted in a different way in the fourth argument. The fact that decisions are intentional is a way of expressing the fact that the effect event, as an event belonging to a certain type of action, stands in a relation of specific causal dependence with respect to the decision. There is no specific dependence of events of that type of action on fine-grained type of neural events N.

What is perhaps more surprising is that the framework of representing causal dependence in terms of functional dependence among variables provides a means of making sense of Lowe's thesis that decisions are uncaused (Lowe 2006, p. 18; Lowe 2010, p. 459).

Lowe would certainly allow that there are causes of decisions in the following sense. There are neural events immediately preceding decisions, such that, if they had been very different, the decision wouldn't have been taken. Some major perturbation of the brain, due to drugs or accident, would have prevented the decision. This can be expressed by saying that there are variables X expressing types of neural events occurring before D such that D=F(X). However, no such type X of neural event is specific for D. There is no type X of events such that X-events are difference makers for D-events. Lowe's statement that a decision D is "uncaused" can be interpreted as meaning that there is no variable X such that D is an injective function of X. There is no variable X representing any type of event occurring before D in such a way that differences in the value of X correspond systematically to differences in the value of D, so that it would depend on the value of X whether or not the person takes the decision or not. This interpretation of Lowe's claim that decisions are "uncaused" would make it compatible with physicalism. True, decisions would not be uncaused in the sense of having no causes whatsoever. However, I think the thesis that there is no event on which a decision specifically depends so that nothing makes a difference to which decision is taken, captures the intuition behind Lowe's thesis.

Conclusion

The four arguments for property dualism of E.J. Lowe's we have examined do establish weak property dualism. They show that the property of an event to be a decision differs from its neural property of being constituted by a specific pattern of activity a set of neurons. An event at which a person takes a decision belongs both to various psychological types and to various neural types, each of which corresponds to a different property. However, none of Lowe's arguments establishes *strong* property dualism, according to which mental properties are independent of neural properties, in the sense of not being grounded on them and not being reducible to them.

We have seen that the representation of causal relations in the framework of structural equations can account for the specific differences between neural and mental causes, which are highlighted in some of Lowe's arguments. The psychological (or mental) cause of a bodily movement B constitutive of an action, such as raising one's arm, i.e. the decision to raise the arm, is specific for B. This corresponds to a feature of the function expressing the dependence of the variable B representing the action on the variable D representing the decision. If the function is injective, the dependence is specific, and the cause D makes a difference with respect to the effect B, in the sense that every modification of the cause variable D corresponds to a variation in the effect variable B. We have also seen that it is possible to interpret, in this framework, Lowe's claim that decisions are uncaused. I have suggested that the intuition behind Lowe's claim can be accounted for by the hypothesis that decisions are special in that they do not *specifically* depend on any type of event preceding them. According to this interpretation, it is not literally true that they have no causes at all. Some events such as a stroke would after all prevent all decisions. But decisions are special in having no specific causes[9].

References

- Ashcroft, Neil W. and Mermin, N. David (1976), *Solid State Physics*, Saunders College, Philadelphia, 1988.
- Bliss, Ricki and Trogdon, Kelly (2016), Metaphysical Grounding, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy,* <https://plato.stanford.edu/archives/win2016/entries/grounding/>.

[9] I thank Louis-Etienne Villeneuve, Reinaldo Bernal Velasquez, and my auditors in Warsaw for helpful comments.

- Davidson, Donald (1970), Mental Events, in *Actions and Events*, Oxford: Clarendon Press, 1980.
- Dennett, Daniel C. (1981), True Believers: The intentional strategy and why it works, repr. in D.J. Chalmers (ed.) *Philosophy of Mind*, Oxford University Press 2002, p. 556-568.
- Dennett, Daniel C. (1990), Real Patterns, *Journal of Philosophy* 88, p. 27-51.
- Dretske, Fred I. (1988), *Explaining Behavior. Reasons in a World of Causes*. Cambridge, MA: MIT Press.
- Fodor, Jerry A. (1983), *The Modularity of Mind*, Cambridge: MIT Press.
- Halpern, Joseph Y. (2000), Axiomatizing Causal Reasoning, *Journal of Artificial Intelligence Research* 12, p. 317-337.
- Hume, David (1739), *A Treatise of Human Nature*, Oxford, Clarendon Press, 1888, second edition, L.A. Selby-Bigge and P.H. Nidditch (eds.), 1978.
- Kim, Jaegwon (1998), *Mind in a Physical World*, Cambridge (Mass.), MIT Press.
- Kistler, Max (2005), Lowe's Argument for Dualism from Mental Causation, *Philosophia* 33, p. 319-329.
- Kistler, Max (2006), *Causation and Laws of Nature*, London, Routledge.
- List, Christian & Peter Menzies (2009), Non-reductive physicalism and the limits of the exclusion principle, *Journal of Philosophy* 106 (9), p. 475-502.
- Lowe, E.J. (1993), The Causal Autonomy of the Mental, *Mind* 102, p. 629-644.
- Lowe, E.J. (1996), *Subjects of Experience*, Cambridge, Cambridge University Press.
- Lowe, E.J. (1999), Self, Agency and Mental Causation, *Journal of Consciousness Studies* 6, p. 225-39.
- Lowe, E.J. (2000), Causal Closure Principles and Emergentism, *Philosophy* 75, p. 571-585.
- Lowe, E.J. (2006), Non-Cartesian Substance Dualism and the Problem of Mental Causation, *Erkenntnis* 65, p. 5-23.
- Lowe, E.J. (2010), Substance Dualism: A Non-Cartesian Approach, in R.C. Koons and G. Bealer (eds.), *The Waning of Physicalism*, Oxford: OUP, p. 439-461.
- Lowe, E.J. (2013), Substance Causation, Powers, and human Agency, in S.C. Gibb, E.J. Lowe, and R.D. Ingthorsson (eds.), *Mental Causation and Ontology*, Oxford: Oxford University Press, p. 153-172.
- Menzies, Peter (1988), Against Causal Reductionism, *Mind* 97, p. 551-574.
- Millikan, Ruth G. (1984), *Language, Thought, and other Biological Categories*. Cambridge, MA: MIT Press.

- Needham, Paul (2017), *Macroscopic Metaphysics: Middle-Sized Objects and Longish Processes*. Cham: Springer.
- Pearl, Judea (2000), *Causality: Models, Reasoning, and Inference*, Cambridge University Press.
- Quine, W.V.O (1960), *Word and Object*, Cambridge: MIT Press.
- Rosenberg, Alexander (1985), *The Structure of Biological Science*, Cambridge, Cambridge University Press.
- Searle, John R. (1992), *The Rediscovery of the Mind*, Cambridge, Mass., MIT Press.
- Smyth, Mary M., Collins, Alan F., Morris, Peter E. and Levy, Philip (1994), *Cognition in Action, 2nd Edition*, Hove, Lawrence Erlbaum Associates.
- Spirtes, P., Glymour, C. and Scheines, R. (2000), *Causation, Prediction and Search*, Second edition, Cambridge (Mass.), MIT Press.
- van Riel, Raphael and Van Gulick, Robert (2019), Scientific Reduction, in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy,* <https://plato.stanford.edu/archives/spr2019/entries/scientific-reduction/>.
- Woodward, James (2003), *Making Things Happen*, Oxford University Press.
- Woodward, James (2010), Causation in biology: stability, specificity, and the choice of levels of explanation, *Biology and Philosophy* 25, p. 287-318.
- Yablo, Stephen (1992), Mental Causation, *Philosophical Review* 101, p. 245-280.