WILLIAM JAWORSKI

MENTAL CAUSATION FROM THE TOP-DOWN

ABSTRACT. Dual-attribute theories are alleged to face a problem with mental causation which commits them to either epiphenomenalism or overdetermination – neither of which is attractive. The problem, however, is predicated on assumptions about psychophysical relations that dual-attribute theorists are not obliged to accept. I explore one way they can solve the problem by rejecting those assumptions.

Materialists have been working in earnest to reduce (or eliminate) the mind for over forty years... It’s time to start hedging our bets. Assuming mental–physical property identities are not forthcoming, that the mind cannot be eliminated, and a return to substance dualism is out of the question, it seems we must endorse the following theses:

[Physical Composition] Substances instantiating mental properties either are or are constituted entirely by fundamental–physical entities;
[Property Dualism] There are mental properties, and there are physical properties, and mental and physical properties are distinct.

Call any theory committed to this conjunction a *dual-attribute theory* (DAT). DATs have been much neglected in philosophy of mind, but apart from a general aversion to dualism of any sort, the reasons for rejecting DAT, with one or two exceptions, are ill-defined. My purpose here is to discuss one of the exceptions, and to explain how dual-attribute theorists might address it. The result is a hitherto unexplored solution to the problem of mental causation, and a DAT that resembles in key respects Aristotle’s philosophy of psychology and action.

1. THE PROBLEM OF MENTAL CAUSATION

DATs face a problem with mental causation. The following theses all appear to be true:
(1) Causal closure of the physical domain: Every physical event that has a fully sufficient cause has a fully sufficient physical cause;

(2) Commonsense mental realism: (a) Actions are caused by mental events, and (b) they would not have occurred if their mental causes had not;

(3) Actions are physical events.

Suppose, however, that Madeleine is writing at t. The movements she produces with the pen clearly have a physiological provenance. She is nevertheless writing for a certain reason, to tell Alexander of her weekend at Eleanor’s, say. As a result, there appear to be two explanations for her writing:

(E1) Madeleine is writing at t because she wants to tell Alexander about her weekend at Eleanor’s;

(E2) Madeleine is writing at t because $p_1,...,p_j$,

where $p_1,...,p_j$ is a series of physiological events. What is the relation between rationalizing explanations of action and physical (e.g. physiological) ones? If both provide true, distinct causal explanations, which are not jointly sufficient, we face a problem: Madeleine’s wanting and $p_1,...,p_j$ are both fully sufficient to cause her writing. But then neither by itself is necessary; her action could have occurred without one or the other; it might have occurred, for instance, even if she had not wanted it to, perhaps even if she’d had no mental states at all; and this violates (2b), which entails that Madeleine’s action would not have occurred without its mental causes. Call this the overdetermination problem. There is another. Suppose that in addition to (1)-(3) the following is also true:

[Causal-explanatory exclusion] Every event has only one fully sufficient causal explanation.

In that case, E1 and E2 cannot both be true; we are forced to choose: either $p_1,...,p_j$ do not cause Madeleine’s writing, or else her desire does not. But neither option is attractive; the former implies there are gaps in the physical causal chain; the latter implies epiphenomenalism. Call this the exclusion problem.

Exclusion and overdetermination are two sides of the same coin. The solution I am going to propose can handle both, and for that reason I will speak generically of the problem of mental causation.
The problem is predicated on the following assumptions:

(A1) (a) There are true rationalizing explanations of actions, and
      (b) true physiological explanations of actions;
(A2) (a) Rationalizing explanations of action are causal, and (b) physiological explanations of action are causal;
(A3) The events cited by rationalizing explanations of actions are not identical to the events cited by physiological explanations of them.

The falsity of any one of these is sufficient to dispose of the problem. If rationalizing and physiological explanations of action pick out numerically identical causes, or if explanations of either sort fail to pick out causes at all (either because they are all false or because they are not causal explanations), the problem vanishes. The solution I propose denies (A2). The strategy is not new; the 50s and 60s witnessed a spate of theories denying (A2a). My strategy, however, is to deny the other conjunct. Rationalizing explanations pick out causes of actions, I want to say; it is physiological explanations that do not. The claim is not that physiological explanations cannot pick out causes; doubtless the occurrence of $p_1,...,p_j$ causally explains a host of physical events; it is rather that physiological explanations do not pick out the causes of action; that task falls to rationalizing explanations. I want to suggest that physiological explanations of action are more reasonably construed as explanations of a different sort: they pick out the realizers of actions, not their causes. What exactly a realizer is I will discuss later, but first I want to discuss what’s involved in denying (A2b), and explain why I find the denial of (A1) and (A3) unattractive.

My adherence to (A1) stems from a conviction that eliminativist solutions are out of bounds. On that much, in any event, we must agree before proceeding. Denying (A3), on the other hand, looks promising. DAT requires that mental and physical properties be distinct, but denying (A3) requires only that mental and physical events be identical. A token identity solution therefore recommends itself: identify events without identifying the corresponding properties. This type of solution is not compatible with just any theory of events, however. Suppose, for instance, that events are property-exemplifications in Kim’s (1973, 1976) sense. On this view, an event is a complex entity consisting of an $n$-tuple of objects standing in an $n$-adic relation at a time. Event $e_1$ is identical to event $e_2$, on this view, if and only if the constitutive objects,
relations, and times of $e_1$ and $e_2$ are identical: $x$’s being $P$ at $t = y$’s being $Q$ at $t'$ iff $x = y$, being $P = Q$, and $t = t'$. Since event identity entails property identity on this account, a token identity solution is out of bounds. A solution targeting (A2), by contrast, does not depend on a theory of events in this way. I will argue that dual-attribute theorists can solve the problem of mental causation even if they endorse a Kim theory of events.

Given reasonable assumptions, (A2b) is implied by the conjunction of causal closure and the claim that actions are physical events: if the physical domain is causally closed, and actions are physical events, then actions have physical causes. It seems a reasonable step to postulate explanations citing those causes; so developing a solution that denies (A2b) will require either denying causal closure, or denying that actions are physical events. I want to suggest the latter tack is more promising. To appreciate this, however, one must recognize that (3) conceals an ambiguity.

Suppose the type to which an event belongs, whether mental or physical, is determined by its constitutive property. Thus, $x$’s being $P$ at $t$ is a physical event just in case being $P$ is a physical property. An event’s constitutive property, however, is distinct from the properties it instantiates (Kim 1973, 12). Madeleine’s writing at $t$ might instantiate the property of occurring in her study, but is not constituted by it. The constitutive property of that event is writing (or writing that such-and-such, depending on how exactly acts are individuated). An event might instantiate physical properties, therefore, without itself being a physical event. $x$’s being in pain at $t$, a mental event, is also a physical event if being in pain is identical to a physical property. Irrespective of its physical status, however, that event might still instantiate physical properties, e.g. occurring at the gym. Given the constituting/instantiating distinction, (3) must be understood in the sense of (3*) not (3°):

(3*) Actions are constituted by physical properties.
(3°) Actions instantiate physical properties.

Denying (A2b) requires denying (3*) only. Causal closure requires that physical events, i.e. events constituted by physical properties, have physical causes; it says nothing about non-physical events. If the properties constituting actions are not physical, actions are not subject to causal closure (even if they should happen to instantiate physical properties). As a result, an action having a mental cause need not have a physical cause besides.
To establish the denial of (3*) one must show that the properties constituting actions, i.e. act-properties (or simply acts) are not physical properties. One way of going about this is to find physical properties slated for identity with acts, and to show they don’t pass muster. The most plausible physical candidates, it seems, are physiological properties. But acts and physiological properties do not map onto each other in a straightforward way; a given act may correspond to numerous distinct types of physiological event. The act of protesting the war might consist in pinning a ribbon on one’s lapel (physiologically described), or in setting oneself ablaze with gasoline (physiologically described)–events which, it is fair to say, are physiologically distinct. Likewise, it seems that a particular type of physiological event might underwrite a number of distinct types of action. The same nodding of one’s head (physiologically described) might in different counterfactual circumstances realize an act of greeting a neighbor, of answering in the affirmative, of insulting a colleague, etc. These observations suggest that act-properties are not physiological properties.

Of course, establishing the denial of (3*) would require the inspection of every physical candidate, a task that would surely take us beyond the scope of this paper. There are nevertheless some general reasons for being skeptical about the possibility of finding physical types identical with acts. Consider, for instance, what those physical types would have to look like, especially in the case of so-called “negative acts”: refusing to sign one’s name, abstaining from fatty foods, refraining from voicing a wry quip, invoking one’s Fifth Amendment rights, etc. These appear to be genuine acts. They can be done intentionally, willfully; one can try to do them, fail in one’s attempt, merit praise or blame as a result, etc. Their performance, however, does not consist in the occurrence of any particular physical events. To be identical with negative act-types, physical types would have to carve out regions of space-time broad enough to accommodate the relevant contexts. Those contexts, however, involve complex relations among agents and their social environments, so the physical types in question would have to be individuated broadly enough to encompass relations constitutive of the life of, perhaps, an entire society or culture. It seems extremely unlikely there are physical types up to the task. To recast a remark of Fodor’s (1974): nothing but brute enumeration would convince us there were
physical types coextensive with act contexts, and even if there were, there would seem to be no chance of the coextensions being lawlike, much less expressions of identity.

The foregoing considerations are obviously not decisive against (3*); they nevertheless suggest grounds for thinking act- and physical properties distinct. If they are, it follows that actions need not have physical causes as required by causal closure; one is free to deny that actions have causes constituted by physical properties. This suggests a strategy for dealing with the problem of mental causation. We can illustrate it as follows:

\[
\begin{align*}
\text{Madeleine’s wanting} & \rightarrow \text{Madeleine’s writing that q} \\
\text{Alexander to know that q} & \leftarrow R \\
\text{… → } p_j & \rightarrow p_k \rightarrow \text{…} \\
\text{→ } p_l & \rightarrow p_m \rightarrow \text{…} \\
\end{align*}
\]

Kim has sometimes referred to this as a “dual-explanandum” or “two explananda” strategy.\(^{13}\) Past advocates have included Melden (1961), Malcolm (1968), and Dretske (1988, 1989). The arrows represent causal relations; the Rs, what I will refer to as “realization” relations. Madeleine’s action has a mental cause, and no physical causes. It and its mental cause are nevertheless realized by physical events \((p_j, p_k, \text{etc.})\) that do have physical causes. The ellipses appearing on the lower level of the diagram represent unbroken chains of causal relations running from physical events occurring earlier in time (i.e. to the left) to ones occurring later. The plurality of physical events relative to the action and its cause is meant to indicate that events in the “lower-level” domain need not correlate one-to-one with events in the “higher-level” domain.

According to the proposed solution, actions and mental events are type- and token-distinct from their contemporaneous physiological realizers (taken either individually or in conjunction). Intuitively, the events realizing Madeleine’s action include the hand movements responsible for pushing her pen, muscle contractions in her arm, neuronal firings in her brain and spinal cord, etc. They are physiological events necessary in some strong sense for the occurrence of Madeleine’s writing, but nevertheless not identical to it. The same holds *mutatis mutandis* for the realizers of mental events. The resulting picture is compatible with (1), (2), and (3*).
MENTAL CAUSATION FROM THE TOP-DOWN

This is an outline of the strategy I plan to use. The basic points, again, are these: Actions and physical events are type- and token-distinct. Given the view of events adopted here, the result is that actions are not subject to causal closure of the physical domain; they need not have physical causes, but could have mental causes only. This does not mean their mental causes do not instantiate physical properties; given certain assumptions, they clearly might; what it means is simply that their constitutive properties are not physical (something required by property dualism in any case). Despite that, however, the solution claims mental events and actions are physically realized. On this score, it seems reasonable to endorse something like the following:

[Physical Realization Thesis] (a) Necessarily, actions and mental events occur only if they are realized, and (b) necessarily, they are realized physically.

(Note: the modality of (a) and (b) is probably not the same). What it means for one event to realize another I will discuss momentarily; but first, I want to put some flesh on the proposed solution.

3. A SOLUTION TO THE PROBLEM OF MENTAL CAUSATION

Let us agree that an explanation, \( e_1 \) because \( e_2 \), has three components: the explanans, \( e_2 \), the explanandum, \( e_1 \), and the “because,” which expresses an explanatorily relevant relation between the two. Let us also agree to a *metaphysically realist* account of explanation according to which the “because” in an explanation expresses an objective relation among entities in the world, e.g. events (Kim 1987). Thus, causal explanations express causal relations: a statement ‘\( e_1 \) because \( e_2 \)’ provides a true causal explanation just in case \( e_2 \) causes \( e_1 \). The solution I am proposing trades on the assumption that causation is not the only explanatorily relevant relation; in particular, the relation expressed in a physiological explanation of action is not one of causation, but of realization. This is not to say physiological explanations do not sometimes express causal relations, only that when they do, careful examination gives us little reason to suppose the intended explananda are actions.

The proposed solution hinges on the following claims:

(4) (a) The explananda of rationalizing explanations are actions, and (b) the “because” occurring in those explanations express causal relations;
(5) Either (a) the explananda of physiological explanations are not actions, or (b) the “because”s occurring in those explanations express realization relations.

Premise (5) implies the denial of (A2b). If the “because”s in physiological explanations are not causal, or if their explananda are not actions, those explanations and rationalizing ones do not pick out competing causes. They cannot overdetermine actions, and neither need be excluded. Consider again E1 and E2:

(E1) Madeleine is writing at t because she wants to tell Alexander about her weekend at Eleanor’s;
(E2) Madeleine is writing at t because $p_1, ..., p_j$.

According to the proposed solution, E1 picks out the cause of Madeleine’s writing; E2 does something else: Either its “because” expresses a realization relation instead of a causal one, or its intended explanandum is not really Madeleine’s writing. In the latter case, E2 represents a derelict attempt at saying something else, e.g.

(E2’) Madeleine’s hand is moving at t because $p_1, ..., p_j$

(where the hand movements $\neq$ Madeleine’s writing). Either way, E2 does not cite a cause that could in any way compete with the mental cause cited in E1.

The intuition driving this solution is not new. In the *Phaedo*, Plato’s Socrates expresses a similar idea when recounting his experience reading Anaxagoras:

...My wondrous hopes were swept away, my friend, when I proceeded to read, and saw the man neither appealing to mind, nor citing any of the causes responsible for the ordering of things, but instead citing air, and aether, and water, and many other absurdities as causes. To me it seemed exactly the same as someone saying that Socrates does everything he does with mind, and then in undertaking to state the causes of each thing I do were to say that I am sitting here now because, first, my body is composed of bones and sinews, and the bones are hard and have joints separating them, while the sinews for their part contract and relax, and cover the bones along with the flesh and skin that contains them, and that because the bones move freely in their joints, the contracting and relaxing of the sinews somehow enables me to bend my limbs now, and this is the cause of my sitting here in a bent position... But to call such things causes is most absurd. If someone were to say that without having bones, and sinews, and such, I would not be able to do what I believe best, that would be true. But to say that I do what I do *because* of these, and therein act with mind, but not on account of choosing what I believe best – *that* would be an extremely careless way of speaking. (98c–99b).14
According to Socrates, the problem with Anaxagoras’ theory is that it imputes to physiological events a causal role in the production of actions. But physiological explanations of action are not causal; to call them such, according to Socrates, would be Pickwickian at best. In a similar vein, the solution I am proposing purports that physiological explanations are either not causal, or else their explananda are not actions at all, but events of a different type described in a careless way. Whether particular physiological explanations satisfy this disjunct is, of course, an empirical matter. What thesis (5) supplies, however, is a strategy for dealing with them. Consider how this might go.

3.1. Case 1: Non-causal physiological explanations

Suppose the intended explanandum of E2 is indeed an action. According to the proposed solution, this is a case in which the “because” in E2 expresses something like (6) instead of (7):

(6) Madeleine’s writing at t is realized by \( p_1, \ldots, p_j \); 
(7) Madeleine’s writing at t is caused by \( p_1, \ldots, p_j \).

The events \( p_1, \ldots, p_j \) are the realizers of Madeleine’s writing, not its causes. Contexts in which an action’s realizers interest us are legion. The following quote from a textbook on the psychology of reading represents just one example:

When speaking, the [neural] circuitry is as follows: (1) The underlying pattern of an utterance arises in Wernicke’s area, (2) The pattern is transferred along the arcuate fasciculus to Broca’s area which evokes a program for vocalization, and (3) The program is relayed to the facial areas of the motor control gyrus which activates the appropriate tongue-lip-larynx muscles of speech. When reading, the visual pattern of print is received in the visual cortex, is transmitted to the angular gyrus which applies a transformation that elicits the auditory form of the word in Wernicke’s area. When reading orally, the circuitry is the opposite: Information is passed from auditory reception to Wernicke’s and back to the angular gyrus. (Robeck and Wallace 1990, 168)

Other examples are of a more clinical nature: Madeleine underwent a complete hemispherectomy at the age of three, the result of which was a nervous system organized rather differently from the average person’s. What, someone might wonder, is going on physiologically when she writes? Similarly, suppose Madeleine’s seizures are controllable only by lobectomy. Because the location of the lesion imperils her ability to write, surgeons are careful to determine first
which cortical structures are responsible for realizing that ability. In such contexts, the realizers of actions, not their causes, are what interest us.

3.2. Case 2: Non-action-directed physiological explanations

Of course, what E2 purports to explain might not be an action at all; it might instead represent a mistake on the part of a speaker who utters E2 but intends it as an explanation of something else. Suppose, for instance, that careful study of the speaker’s intention reveals E2’ as a more felicitous expression of that intention than E2. The “because” in E2’ may express a causal relation, but what it causally explains is not an action. Perhaps the hand movements are the physiological events that realize Madeleine’s writing, in which case the explanation may pick out the physiological causes of those movements.

But if E2’ is the accurate expression of the speaker’s intention, E2 represents a linguistic or conceptual mistake. According to the proposed solution, such mistakes are likely to occur given the intimate nature of the realization relation. The latter makes it easy to suppose the action is the target of physiological explanation when in fact it is the realizer. Often, in such cases, the description or name of an action functions as a convenient label for a series of events more accurately described in purely physiological terms. Lack of verbal caution then lends itself to mistakes. For instance, let “controlled hand movements” designate a type of very complex physiological event that frequently realizes a human act of writing. Given the frequency of the association, and the intimacy of the realization relation, circumstances may lend themselves to subtle confusions of a sort manifest in the following:

What causes controlled hand movements, e.g. what causes Madeleine to write?
Madeleine appears unable to regulate lateral movements of her hand. Watch the way she writes... Why is she doing that?

In these contexts a request for information has been made that apparently concerns Madeleine’s writing. In both cases, however, the real focus, the real explanandum, is not her action, but the physiological events realizing it. E2 might be uttered in such contexts with an eye to explaining the occurrence of those events, but it would miss the mark. Careful examination would reveal something like E2’ as the
more accurate utterance, and a physiological event as the real explanandum.

These examples are just a smattering; a fully worked out solution would obviously have to cover more since there are many ways in which confusions of the foregoing sort can arise. Once the realizers of actions are mistakenly taken for actions themselves, for instance, the physiological causes of those realizers are easily confused for the causes of action. My goal here has simply been to illustrate how the strategy might work, and in that respect, I hope the foregoing examples have sufficed.

Before continuing, I want to clarify what the proposed solution is by saying a word about what it is not. Given its property dualism, it is obviously not a type identity solution; nor, given a Kim view of events, is it a token identity solution, for on that view token identity implies type identity. It is clearly not an overdeterminist solution; that much is clear from its endorsement of (2b), the claim that an action would not have occurred if its mental causes had not; but neither is it a form of epiphenomenalism, token or type (McLaughlin 1989). Token epiphenomenalism, which claims mental events cannot enter into causal relations, is obviously incompatible with (2a), the claim that mental events cause actions, while type epiphenomenalism, which allows mental events to enter into causal relations, but denies they do so on account of their mental properties, is ruled out by (2a) in conjunction with the plausible assumption that events enter into causal relations on account of their constitutive properties. Finally, the solution is not a form of interactionism. The latter claims that actions are physical events having exclusively mental causes; the proposed solution, however, denies both that actions are physical events (in the sense of (3*)), and that physical events could fail to have physical causes.

Can we say more, in positive terms, about what the proposed solution is? A solution along these lines is compatible with several views of psychophysical relations. In what follows I will sketch one of them.

4. DATS: TOP-DOWN VERSUS BOTTOM-UP

I’ve claimed there’s an important sense in which actions are not physical events, and are therefore not subject to causal closure of the physical domain. What this proposal effectively does is incorporate actions into the mental domain in a manner reminiscent of one of
Davidson’s (1970, 110) remarks: “intentional actions are clearly included in the mental domain along with thoughts, hopes, and regrets.” Once we understand actions as mental phenomena, the question is how the entire domain of thought and action relate to physical events, physiological ones in particular. I’ve attached the label “realization” to that relation, but I’ve so far said nothing about it.

The use of “realization” as a label for the psychophysical relation, was ushered into philosophy of mind by Putnam (1960, 1967) as a way of characterizing the relation between computational procedures and the physical states of the systems that execute them. The two are clearly not identical, he said, for the same procedure might be carried out in systems physically very different. There is nevertheless a sense in which the physical states of a computer are the procedure as it is being carried out; they implement or realize it. That relation, Putnam suggested, also characterized the relation between mental and physical states: the physical states of persons realize their mental states the way physical states of computers realize their programming. “Realization” and its cognates have since become buzzwords in the Philosophy of Mind. I want to consider one way of understanding the realization relation, and consider how dual-attribute theorists might make use of it.

In order for a two explananda solution to the problem of mental causation to work, the realization relation must not be a species of causation. I will nevertheless assume that realization is like causation in being an irreflexive, asymmetrical relation between events (we needn’t decide here as to its transitivity). What distinguishes it from causation is that its relata are necessarily synchronic and co-locational, i.e. necessarily, if $e_1$ realizes $e_2$, those events occur at the same time and place. Causal relations, by contrast, are paradigmatically diachronic, and action at a distance seems at least logically possible.

I want to suggest there are two shapes a DAT can take depending on the model of realization its advocates endorse. One model claims that the occurrence of a realizer determines the occurrence of the mental event it realizes (Putnam 1970; LePore and Loewer 1989; Block 1990; Kim 1992; Melnyk 2003); a second model denies this. DATs committed to the first model approach mentality “from the bottom up,” so to speak. They begin by considering which physical events occur in a world, and go on to consider how thought and action fit onto that physical base. I will refer to these as bottom-up DATs. Top-down DATs approach matters from the other end. They
begin by considering which mental events and actions occur in a world, and only then consider the physical conditions that make that world possible.\textsuperscript{16}

Kim (1993b, 2001) has argued that bottom-up DATs cannot solve the problem of mental causation. Briefly: suppose causation is a type of determination relation: causes determine the occurrence of their effects. According to bottom-up DATs, realization is also a determination relation.\textsuperscript{17} It’s not difficult to see why this would pose a problem: \textit{overdetermination need not be specifically causal to violate commonsense mental realism}. If an action’s mental cause and its realizer both determine its occurrence, if each would have been sufficient by itself to produce it, then neither by itself is necessary; the action might have occurred without its mental cause, and this violates (2b), the claim that actions would not occur without their mental causes.

There may be ways friends of bottom-up DATs can respond to this challenge. It might also be worthwhile, however, to explore a top-down alternative.

5. THE REALIZATION RELATION

For the remainder of the paper, I want to sketch out what a top-down DAT might look like. Although there may be several ways of making a top-down approach work, I want to consider one that owes its inspiration to Aristotle.

On Aristotle’s view, an organism’s actions are caused by its desire for something presented to it in thought, perception, or imagination.\textsuperscript{18} Those actions, as well as their psychological causes, nevertheless occur “in” or “through” changes in the organism’s matter;\textsuperscript{19} they are enmattered formulae (logoi enhuloi). Consequently, any account of them must combine material and formal factors according to a scheme like the following: “a movement of such-and-such a body (or part or power of it) \textit{[caused] by this for the sake of this}” (403a26–27). Aristotle’s example is the state of anger. The latter might be defined as a boiling of blood around the heart caused by pain for the sake of retaliation. For me to be angry is for my matter to undergo a change, but it is not simply that: to be in a state of anger, my blood must boil \textit{in the right mental and social context}, so to speak. My state of mind must be caused by pain, and must be directed toward retaliation, so that being angry is not identical to having boiling blood; they are not “one in being,” as he would say.
Having boiling blood, or undergoing a similar change in one’s matter, is nevertheless necessary for my state of anger to exist, and there is a sense in which having it in the right context is anger. The “is” here resembles what I will refer to as the “is” of realization.

Because Aristotle understood the relation between psychological and physiological states as an instance of the relation between form and matter, I will refer to the account inspired by it as a hylomorphic account of realization. Unlike other accounts of realization that have appeared in the literature, the account does not attempt an analysis of realization, but takes it as a primitive. Friends of the hylomorphic account must therefore rely on examples to elucidate it. The clearest of these involve actions: Madeleine’s hand movements (physiologically described) realize her writing; Eleanor’s arm waving (physiologically described) realizes her hailing a cab. The hand movements and arm waving do not cause the actions; they are the actions, where the “are” is the “are” of realization. Henceforth, I will attach the subscript “R” to forms of the verb “to be” when the intended sense is that of realization.

On the hylomorphic account, physiological changes occurring in the right contexts are\_R actions and mental events. But to say a certain physiological event is\_R Eleanor’s cab-hailing is not to say the two events are identical. Realization is a contingent relation. A different set of physiological events such as a shrill whistle (physiologically described) could have been\_R Eleanor’s cab-hailing. Conversely, the arm waving (physiologically described) could have been\_R a different action, such as bidding farewell. The same is true of mental events: c-fiber firing might in fact be\_R pain, but z-fiber firing could be\_R pain as well; and the physiological events that are\_R in fact Alexander’s belief that H\_2O is colorless, might have been\_R a belief that XYZ is. The “is” of realization is not the “is” of identity.

Unlike accounts of realization endorsed by bottom-up DATs, the hylomorphic account does not require that realizers be sufficient for what they realize. It is compatible with the view depicted in the figure below, according to which the action’s sole sufficient condition is its

\[
m \quad \xrightarrow{R \ \ R} \quad \text{The action} \quad \xrightarrow{p \ \ \ \ p'}
\]
mental cause, m. As a result, DATs committed to the hylomorphic account do not face a problem with overdetermination the way bottom-up accounts do.

But if the hylomorphic account does not require that realizers be sufficient for what they realize, neither does it preclude strong dependence relations between them. It is reasonable to suppose, for instance, that there are constraints on the types of physical events capable of realizing thoughts and actions. For example, it seems manual writing can be realized only by physical events that produce marks on a surface. Worlds in which the laws of nature render surfaces impervious to marking are therefore worlds in which manual writing cannot occur.20

Aristotle discussed such constraints under the heading of *hypothetical necessity*. If there is to be a saw, he said, it must be made of a suitably hard material, e.g. iron. The goal of wood-cutting cannot be accomplished unless the potential wood-cutting tool is made of the right stuff (*Physics* 199b34ff.). The same, he said, is true of psychology. Thought, perception, and the like can be accomplished only if a potential perceiver is made of the appropriate materials. For instance, visual perception, he argues, can occur only in something transparent (419a11), and that means an organ made of either air or water, the only two transparent materials in his physics.21 Exponents of the hylomorphic account may insist, like Aristotle, that physical events meet certain constraints to operate as realizers: not only are physical events necessary for mentality (that much is entailed by the Physical Realization Thesis), but certain *types* of physical events may be necessary to realize certain types of mental ones as well.

6. MENTAL CAUSATION FROM THE TOP-DOWN

A moment ago I remarked that according to the hylomorphic account physiological changes occurring in the right contexts areR actions and mental events. But what are the right contexts? The temptation here is to understand them as magical devices that somehow transform ordinary physiological occurrences into thoughts and actions. Friends of top-down DAT insist, however, that this is exactly the *wrong* way to look at things, a way infected by the bottom-up mentality. According to top-down DATs, the *right contexts are precisely those in which the thoughts and actions occur*. Friends of top-down DAT approach mental causation by determining first what the mental context is, i.e. what thoughts and actions occur in a world,
and only then go about determining how those thoughts and actions are realized. This prioritizing of the mental domain is precisely what "going top-down" is all about: it means approaching mental causation (et al.) in a way that takes thought and action as primary, and only secondarily considers how physical events make them possible; it means viewing physical events not as supplying conditions that determine thought and action, but as supplying conditions that allow thought and action to be.\(^\text{22}\)

Contexts in which thought and action occur are precisely ones in which the physical is \(r\) the mental. It is only when things break down, when physical events fail to satisfy the conditions necessary for thought and action, that the mental and physical come apart, and physical events cease to be \(r\) mental. To engage in thought and action, certain (perhaps very specific) physical conditions must be in place, but sometimes, alas, they are not. Thought requires coherent neural patterns of a certain sort; hence, certain neural events are \(r\) not thoughts; they are (not "are\(^r\)") instead disruptions in the normal operation of mind in the world (e.g. "seizures"). Likewise, visual perception requires the right sort of causal contact between visual organs and reflected light. Physical circumstances in which this contact is inhibited are \(r\) not instances of perception. The same is true of "overt" action. Limb movements produced by the pathological firing of motor neurons are \(r\) not actions, but situations in which structures responsible for realizing actions have failed to meet the conditions necessary for it.

The top-down approach obviously raises a number of questions about how top-down advocates understand psychophysical dependence relations. I obviously cannot paint a detailed picture here, but I can offer, perhaps, a quick sketch of how they might look if viewed from the top-down.

By all empirical accounts, physical events were occurring long before there were minds; the physical enjoys what we might call a \textit{temporal priority} over the mental. Given the Physical Realization Thesis, moreover, it also enjoys an \textit{ontological priority} in the following sense: physical events can exist without mental events, but not vice-versa. There is a different sense, however, in which the mental enjoys an ontological priority over the physical, a sense somewhat difficult to specify, but central to the top-down approach. It comes out as a rejection of a certain procedure for generating philosophical problems.

Advocates of the bottom-up approach characteristically pose questions with the following form: "Given that the physical is
thus-and-so, how is it possible for the mental to be like this?” – where the illocutionary force of the question is to suggest it can’t be like that. Suppose, for instance, that at time t in world w, Madeleine wants to tell Alexander of her weekend at Eleanor’s, and this causes her to write “I had a wonderful time at Eleanor’s.” But, a bottom-up advocate might urge, given that the physical domain is causally closed, and assuming the movements of the atoms constituting Madeleine’s arm in w have causes, it follows that they must have physical causes, and given that causation is a species of nomic sufficiency, it follows that each atomic movement is caused by another physical event completely sufficient for it, and that by another such physical event, and so on. But in that case, how is it possible for Madeleine’s desire to cause her writing? If each physical event was fully sufficient for the next, how could the mental be necessary in the way required by commonsense mental realism?

Top-down advocates insist this represents the wrong way of approaching mentality, one beholden to a prejudice about the nature of the physical. Consider instead, they say, that thoughts and actions do in fact occur. Why not tack down the theoretical rug in that corner, rather than in a corner that conceives of the physical as something that could preempt the causal efficacy of the mental? The knowledge we have of the physical domain is compatible with a different conception of the physical, one that doesn’t pose a preemptive threat. On that conception, the physical is that which makes the mental possible.

Given that Madeleine’s desire and writing occur in w, the Physical Realization Thesis entails that some physical occurrences must be in those events. It might also be the case that Madeleine’s desire can cause her writing only if those occurrences are links in a continuous physical causal chain. None of this, however, requires that those physical events should determine the mental ones in the way bottom-up advocates insist they must. When it comes to mental causation, top-down advocates begin by determining which mental events and actions occur in a world, and only then consider what physical events make that world possible.

Which physical events those are in any given case is to large extent an empirical matter. Cognitive science, neuroscience, and the like have (or will have) a lot to say about how thought and action are or must be realized. Given the information they provide (or hopefully will provide), one might be able to infer that, say, engaging in writing requires that a person be constructed of components like this or like that, that during writing these be thus-and-so, or perhaps like
such-and-such, and that consequently, a potential writer must be constituted by these tissues and molecules, or by those, etc. When it comes to empirical matters, then, the methodological recommendations of top-down DAT are not significantly different from those of, say, homuncular functionalism (Lycan 1987, Ch. 4). Where it departs from the latter is in rejecting the claim that thought and action might be determined by conditions specified in non-mental, non-actional terms; physical events merely realize mental ones in the foregoing sense of supplying conditions necessary for their instantiation. The dependence this entails is, admittedly, profound: without physical events of the appropriate sort, mental events and actions would not occur, but this in no way diminishes the priority enjoyed by thought and action when they do occur. The world cannot run without a physical engine, but thought and action do the driving.

Finally, I wish to consider a sense in which the physical is dependent upon the mental. Certain distributions of matter and energy would not be in place had certain thoughts and actions not occurred. As the case of Berkeley’s famous naked man shows, for instance, the atoms that comprise your clothing would not be where they are had you not dressed yourself, and according to commonsense mental realism, you would not have dressed yourself if you hadn’t mental cause to. The positions of those atoms depend in a certain way on mental events that occurred earlier today (presumably); nearby worlds in which you hadn’t a desire to dress are worlds in which you didn’t; those are worlds in which your act of dressing didn’t occur, and was therefore not realized; and so in those worlds, the relevant atoms remained in your closet.23

7. CONCLUSION

The foregoing remarks are programmatic but also, I hope, suggestive. Bottom-up advocates insist there is an urgency to certain questions in philosophical anthropology. I hope these remarks raise questions about the extent to which that perceived urgency is predicated on a conception of the physical that is by no means obligatory.

Top-down DAT, the hylomorphic account of realization, and the solution to the problem of mental causation they support, are a hedge in uncertain times. My aim here has not been to defend them against every possible objection, but simply to advance them as options worth further consideration.24
Many thanks to Robert Audi, Marian David, Jaegwon Kim, Michael Kremer, Mike Loux, and Leopold Stubenberg for helpful comments on an earlier draft of this paper. Many thanks also to two anonymous referees for Erkenntnis, and to audience members at the University of Notre Dame where a version of this paper was read in January 2002.

NOTES

1 Anglo-American representatives of DAT include Strawson (1959), and Searle (1995).
2 In what follows I am assuming a nomic sufficiency account of causation (see Section 4) – hence the need for a distinction between the counterfactual condition expressed by (2b) and the causal condition expressed by (2a).
3 Note that to generate the problem, the weaker claim that only some actions are physical events would suffice. It would simply limit the problem’s scope.
4 As Kim (1990, 1992, 1998) has emphasized, this is a question any respectable theory of mental causation must answer.
5 See, for instance, Anscombe (1957) and Melden (1961). The classic defense of the claim that rationalizing explanations are causal is Davidson (1963).
6 For approaches to mental causation that reject a Kim theory of events see Davidson (1970), Macdonald and Macdonald (1986, 1991), and Yablo (1992).
7 I will follow the convention of using “action” to refer to events, and “acts” or “act-properties” to refer to event types. We should also note that a property-exemplification view of events commits us to a fine-grained account of act individuation (Goldman 1970).
10 The strategy here is to place actions within the mental domain. The claim that actions are part of the mental domain has been suggested by others. Davidson (1970), for instance, claims that “intentional actions are clearly included in the mental domain along with thoughts, hopes, and regrets” (110); and Armstrong (1970) suggests Behaviorists were right “in thinking that our notion of a mind and of individual mental states is logically tied to behavior” (195; emphasis original).
11 Jennifer Hornsby (1980) suggests that even negative actions can be identified with physiological changes of certain sorts. She quotes Sherrington (p. 13) to the effect that neural inhibition is as much an event as neural excitation. The problem is that neural inhibition does not have the right sort of grain for the purposes she has in mind. Everything we do involves neural inhibition at some level. It’s simply out of the question to suppose that the concept of neural inhibition might do yeoman’s service for that of negative action.
12 The tacit assumption of many reductionists is that reduction in the social sciences will follow on the heels of reduction in psychology. But these considerations suggest exactly the opposite: actions are fundamentally social; reduction of human behavior (mind and action) requires as a first step the reduction of social relations.
Consequently, some advocates of supervenience-based theories have called realizers “causes,” albeit material causes. We are obviously using “cause” in a different, narrower sense. According to the Aristotelian, “cause” in the sense used throughout this paper it designates only one type of causal relation, a type resembling the Aristotelian notion of an efficient cause. Consequently, when we say that realizers are not the causes of actions, the Aristotelian might express the idea by saying realizers are not the efficient causes of actions; i.e. material causes are not efficient causes.

A third approach distinct from both top-down and bottom-up DATs denies that psychophysical dependence relations are asymmetrical. Parallelism is the thesis that mental and physical events do not interact causally, but are nevertheless correlated in interesting counterfactual-supporting ways. It has been traditionally associated with substance dualism according to which mental properties are instantiated by non-physical persons or minds. Mental events, while capable of causing other mental events, are not capable of causing physical events despite being related in counterfactual-supporting ways to the instantiation of physical properties by bodies. But this traditional understanding of parallelism is by no means compulsory; mental and physical events might occur in parallel causal lines even if persons are physical entities such as organisms. Parallelism is therefore compatible with dual-attribute theory, and we might refer to a dual-attribute variety of parallelism among these lines as neoparallelism to distinguish it from the traditional alternative.

It is difficult to know what to say about supervenience-based theories here. Their advocates typically want supervenience to provide determination, but supervenience has not seemed up to the task. What would have to be added to it to yield determination is debatable; one possibility, however, is that the instantiation of the supervenient properties must somehow explain the instantiation of the supervenient ones. Cf. LePore and Loewer’s (1989) analysis of realization.

E.g., De Anima 406b24–25, 433a10–31, 413b12; De Motu 700b17, 701a29–34, 701b1.

E.g., Phys. 194a13–14; De Anima 403a16–24, 408b7–10, 412b4–25, 413a3–5, 414a19–21, 423b30, 424a25–26, 30–31, 433a1, 433b18; De Sensu 436a7–10, 436b2–7; De Motu 701b2ff.; Metaphys. 1026a2–3, 1036b27ff. For some excellent discussion of these texts see Nussbaum (1978, 1984, 1986), and Nussbaum and Putnam (1992).

This echoes some recent remarks of Block (1997): “In Walt Disney movies, teacups think and talk, but in the real world, anything that can do those things needs more structure than a teacup... laws of nature impose constraints on ways of making something that satisfies a certain description. There may be ways of making such a thing, but not just any old structure will do. It is easy to be mesmerized by the vast variety of different possible realizations of a simple computational structure... But the vast variety might be cut down to very few when the function involved is mental... I’ve already mentioned one trivial constraint: a thinker requires a structure different from that of a teacup. Here is a guess as to a more general and slightly less trivial constraint: a thinking thing cannot be composed entirely of a liquid or a gas”
(120–121). Although Block endorses a different account of realization, the idea is one friends of the hylomorphic account can adopt: there are constraints on the types of physical events capable of realizing mental events and actions.

21 In fact, it turns out that because air is difficult to contain, visual perception can be accomplished only by animals with eyes made of water (438a13ff.).

22 Is the same type of account in order for chemical, biological, neurological, and other kinds? Answering this would go beyond the scope of this paper. The account presented here is committed to psychophysical property dualism; it says nothing about bio- or chemo- or neuro-physical property dualism, and is compatible with the claim that biological or chemical or neurological properties are identical to physical properties narrowly construed.

23 Worlds in which those atoms didn’t remain in your closet, we may suppose, are not nearby.

REFERENCES


27–56.  
362–385. 
Art, Mind and Religion, University of Pittsburgh Press. Reprinted as ‘The Nature 
Cambridge University Press. 
Robeck, M. C. and R. R. Wallace: 1990, The Psychology of Reading: An Interdis- 
ciplinary Approach 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ.  
Cambridge, MA. 
London. 
Strawson, P. F.: 1959, Individuals: An Essay in Descriptive Metaphysics, Methuen, 
London. 

Philosophy Department  
Fordham University  
113 W. 60th Street  
New York, NY  
10023-7484  
USA  
E-mail: jaworski@fordham.edu  

Manuscript submitted 4 May 2005  
Final version received 22 February 2006