

# **Reason, Metaphysics, and Mind**

NEW ESSAYS ON THE PHILOSOPHY  
OF ALVIN PLANTINGA

Edited by Kelly James Clark and Michael Rea

**OXFORD**  
UNIVERSITY PRESS

# OXFORD

UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further  
Oxford University's objective of excellence  
in research, scholarship, and education.

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi  
Kuala Lumpur Madrid Melbourne Mexico City Nairobi  
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece  
Guatemala Hungary Italy Japan Poland Portugal Singapore  
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2012 Oxford University Press

Published by Oxford University Press, Inc.  
198 Madison Avenue, New York, New York 10016

www.oup.com

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means, electronic,  
mechanical, photocopying, recording, or otherwise,  
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

Reason, metaphysics, and mind : new essays on the philosophy of Alvin Plantinga / edited by Kelly James  
Clark and Michael Rea.

p. cm.

ISBN 978-0-19-976686-4 (alk. paper)

1. Plantinga, Alvin. I. Plantinga, Alvin. II. Clark, Kelly James, 1956-  
III. Rea, Michael C. (Michael Cannon), 1968- IV. Title.

B945.P554R43 2012

191—dc22 2011009250

1 3 5 7 9 8 6 4 2

Printed in the United States of America  
on acid-free paper

## Causation and the Mental

Peter van Inwagen

I have some rather extreme ideas about ontology—and when I say this, I’m not alluding to the ideas about tables and chairs and organisms that the phrase “van Inwagen’s extreme ideas about ontology” would no doubt suggest to many philosophers. I’m alluding rather to certain ideas I have that belong to the most abstract (and the most abstruse) part of ontology—the part that pertains to the concepts of substance and attribute and the relations between them.

I have some very odd ideas about causation—notice that I distinguish the extreme from the very odd in the realm of ideas—and some very odd ideas about the relation between the mental and the physical. (Or perhaps I should say “about the traditional opposition between the mental and the physical,” since the phrase “the relation between the mental and the physical” suggests something having to do with causation, and the odd ideas I’m alluding to are not ideas about the way the mental and the physical are *causally* related.)

What I want to do in this essay is to try to bring my extreme ideas about ontology and my odd ideas about causation and my odd ideas about the mental and the physical together to see what emerges—to try to see whether my ideas about the abstract and the concrete, my ideas about causation, and my ideas about the mental/physical opposition have, as it were, a vector sum. I am particularly interested in the question whether this sum, this issue—whatever one wants to call the result of putting ideas about disparate subject matters together—has any implications in the matter of mental causation, any implications about how to answer traditional philosophical questions about the causation of the physical by the mental, and the causation of the mental by the mental, and the causation of the mental by the physical. (In the end, it will transpire that I do have some rather odd ideas about how the mental and the physical are causally related. But these ideas will figure in this essay as conclusions, not as premises.)

## I. My Extreme Ideas about Ontology

The world, I say, divides into abstract and concrete objects: everything is either abstract or concrete and nothing is both.

I must concede at the outset that I don't know how to define either of these terms. I think it's *probably* right to say that an object is concrete if and only if it can enter into causal relations and that an object is abstract if and only if it cannot enter into causal relations. (As you can probably guess from my opening remarks, I'll presently have a good deal to say about what I mean by causal relations. For the present, I'll simply remark that I don't regard the fact that the law of universal gravitation is an inverse-square law as implying that the number 2 enters into causal relations with gravitating bodies—nor do I regard the fact that the number 100 is the measure of boiling point of water in the Celsius temperature scale as implying that *that* number can enter into causal relations with the water in your teakettle.)

Nevertheless, I don't think that these true biconditionals—these *necessarily* true biconditionals—constitute real definitions of “abstract” and “concrete.” In my view, those two biconditionals, considered as definitions, are of little more value than these two biconditionals:

$x$  is a word if and only if  $x$  has a spelling

(Compare:  $x$  is a representative of the category “the basic unit of speech.”)

$x$  is the number 2 if and only if  $x$  is the even prime

(Compare:  $x$  is the successor of the successor of the cardinal number of the empty set.)

One does feel that to treat any of these biconditionals as a definition would be to treat a superficial feature of a concept as if it were a fundamental feature of that concept.

There are, therefore, abstract objects, concrete objects, nothing that is anything other than abstract or concrete, and nothing that is both. But if there are abstract things and concrete things, what *sorts* of abstract/concrete things are there?

In my view, the only concrete objects are *substances* and the only abstract objects are relations—that is, “relations-in-intension.” (I include in the category of relation, one-place or unary relations—that is properties or attributes or qualities—and zero-place relations or propositions.)

In addition to zero-place and unary relations, there are the items that are more usually called relations: binary, ternary, and so on, relations—*proper* relations, so to call them. And I would place other things in the category “proper relation” as well: “variably polyadic relations,” relations expressed by sentences in which plural variables are free, sentences like “the  $x$ s are carrying a beam,” “the  $x$ s and the  $y$ s are conspiring” and “ $x$  is conspiring with the  $y$ s to persuade  $z$  to betray the  $w$ s.”

So much for relations. What about substances? I can't really tell you what substances are. That is to say, I don't know how to tell you which concrete things are substances and which are not. My reason for this is that I can't see how anything could

manage to be a concrete thing but avoid being a substance. Anyone who thinks that this can be managed seems to think something like this: some concrete things more fully realize the idea of “being a real thing” than other concrete things do. The concrete things that most fully realize this idea are the substances. Here are some things that many people think *exist* and think are concrete rather than abstract, and which (they say) despite their existing and their being concrete do not manage to be fully real: artifacts, severed limbs, sticks and stones, holes, cracks (e.g., in a vase), surfaces, waves, reflections, and shadows. There would be some tendency, on the parts of some philosophers, to say that artifacts, severed limbs, and sticks and stones were incomplete or defective substances, and that holes and cracks and waves and reflections and shadows were not even that—that they were mere *modes* of substance.

I can say only that I don’t really understand any of this. I cannot grasp the idea of one thing’s being more real than another—much less the idea of something’s not being a real thing at all. (I don’t deny that “real” is a very useful adjective, but I don’t think its purpose is to mark out a boundary between the class of things it applies to and the class of things it does not apply to.) As far as I can see, if there *were* artifacts and severed limbs and waves and the rest, they would be as real as anything could be—and would therefore be substances. Now I in fact *don’t* think that there are any such things as these, but that’s another story.

All right. There are the concrete things, the substances, and there are also the abstract things, the propositions, properties, and proper relations. How are these two classes of things, the abstract things and the concrete things, related to each other?

Well, let’s look at a simple case—a case that will perhaps serve further to explain how I look at abstract things.

Let’s consider someone’s sky-blue scarf. (Of course I don’t think that there are any scarves, but I expect most of my readers do, and I don’t object to tailoring my examples to my audience. I do think that there are sky-blue things—various exotic birds, say—and it doesn’t much matter what my example is.) Here we have a substance and an attribute. I first note that the attribute is a universal—it is had by the scarf, but it is had by lots of other things as well.

And all the properties and relations I shall speak of are universals—with the possible exception of haecceities and certain other properties that in some sense “involve” particular concrete things—properties like Platonity, the property of being Plato. I suppose that, strictly speaking, haecceities are not universals, since they can’t be shared, but if they are not universals, neither are they what philosophers who believe in them have variously denominated as tropes or individual accidents or particularized properties. And it is these items, whatever they may be called, that I mean to distance myself from. In my view, although I concede that the phrase “the blueness of Jill’s scarf” might denote something, its referent might very well be identical with the referent of “the blueness of Julia’s scarf”—if those two scarves were of exactly the same shade of blue. The idea that a garment might have a color that was essentially peculiar to it—because it, the color, was in some recon-dite sense a *constituent* of the garment—is opaque to me. No one would say—or

would they? I hope not, but I've been caught out in matters like this more than once—that Jill's scarf had a width or a length or a texture that was all its own, and could not possibly be the width or length or texture of some other object. Why then suppose that an object can have an incommunicable color, an incommunicable blueness or greenness?

Nor do I suppose that *universals* are in any sense, however recondite, constituents of concrete things. (This would seem to be equivalent to saying that properties do not “inhere in” the things that have them.) In my view, the relation that relates the scarf and the color sky-blue—call it instantiation, exemplifying, having, or what you will—is as abstract and bloodless as the relation (counting or numbering) that relates Jill's scarf and Julia's scarf, on the one hand, and the number two, on the other. The properties or attributes are, so to speak, up there in the Platonic heaven with the numbers and other mathematical objects,<sup>1</sup> and the substances, the concrete things, are not. Thus, the property being sky-blue is no more a constituent of a sky-blue scarf than the number two is a constituent of a pair of scarves.

These properties, these inhabitants of the ontological empyrean are as abstract and Platonic as anyone could imagine: they cannot be *seen*, for example, not even if, like the property of being sky-blue, they are properties that imply the property of being visible. (If properties could be seen, then an abstract thing could enter into a causal relation, for “seeing” is certainly a causal relation.) Consider for example, the color sky-blue. (The color sky-blue, I suppose, is the same object as the property of being sky-blue. At any rate, I can't see what else it could be.) You can't see it. You can see things that have it, of course—scarves, for example—, and you can see *that* they have it, but you can't see *it*. (Cf: you can see pairs of things and see *that* they are pairs, but you can't see the number two.)

The interlocutor speaks: “But what about the experience you have when you look at a cloudless sky on a fine summer's day. Then you see that color but (since there is no firmament) nothing that *has* it. It might be that there were no sky-blue objects at all—no scarves or parrots. Still, look at the sky and you'll see sky-blue, you'll see that *color*.”

I must meet the interlocutor's speech with a flat denial. I say that when you look at an empty patch of sky on a fine day, you see *nothing*. (Note that “see nothing” is ambiguous. In one sense, it means “have no visual experiences.” When I say that you see nothing when you look at the sky on a fine day, I don't of course mean that you have no visual experiences. I mean, rather, that nothing—no thing—stands in the relation “is seen by” to you.) If there were no sky-blue scarves, parrots, and so on, *nothing* would be sky-blue. For *what* would be sky-blue? Not the sky—for there is no sky. Not a reflection of the sky in a pond, for there are no reflections. Not some quale, for there are no qualia. There would, in that case, be nothing “there”

---

<sup>1</sup> If my ontology is correct, of course, then, if there are such things as mathematical objects, they must *be* attributes or some other kind of relation. The number 2, for example, may be something like the logical sum of all irreflexive binary relations.

but you, looking upwards, and, as Chisholm liked to say, sensing sky-blue—*a* sensory state with no object.<sup>2</sup>

The color sky-blue, I would say, is an abstract thing, and is, as abstract things tend to be, necessarily existent. Since there are obviously possible worlds in which there is nothing sky-blue, it follows that this color can exist even if nothing has it—which, I suppose, is not surprising if the “having” relation is as abstract and external as the numbering relation. (A number can exist even if there are not enough concrete things for it to number. No doubt most numbers, even most finite numbers, do just that.) Universals are thus *universalia ante res* and not *in rebus*.

Properties, moreover, are not located where their instances are—and are thus not “multiply located”—and they are not “wholly present” where their instances are. Rather they have no kind of presence or location at all. One might as well say that numbers are present when things are present in the appropriate number as say that properties are present in those places at which things have them—one might as well say that whenever two or three are gathered together, the number two, and, it may be, the number three, are with them.

Finally, properties abound. There are not only such properties as are presented to our senses as belonging to the objects we sense<sup>3</sup>—properties such as roundness and whiteness and warmth—but there are also such properties as being (either [warm and not round] or [round and not white]) and being the second marine biologist to be married to a president of the United States in the twenty-third century.

I *almost* want to say this: that to every one-place open sentence (with a precise meaning) there corresponds a property—the property expressed by that sentence. Well, Russell showed that that can't be right, but I'd like to go as far in that direction as possible. (And of course I do not mean to imply that there are only such properties as can be expressed in some language: most sets of real numbers can't be singled out in any language, and for every set of real numbers there is the property of belonging to *that* set.)

<sup>2</sup> If you want to know what “sensory states” are, they are simply properties. The sensory state “sensing sky-blue” is the property of (being a perceiver who is) sensing sky-blue. Many of the central problems of the philosophy of perception are summed up in the question, What is the relation between the property “sensing sky-blue” and the property “being sky-blue”?

<sup>3</sup> In a noncausal sense of “presented.” If the reader supposes that “presentation” is essentially a causal concept, I would cite this case: if I look at two coins lying on a table, I know that there are two coins before me, and this knowledge is noninferential. It must therefore be that, if there is such a thing as the number 2, it—and no other number—is in some sense presented to me in the experience I have when I look at the tabletop; and it is presented to me *as* the number of the coins that are before me on the table. And yet it enters into no causal relation with me: the coins affect me, but the number of coins does not affect me. (If there is such a thing as the mereological sum of the two coins, *it* no doubt affects me. But the number of its maximally connected parts does not affect me.) It is true that the effects on me of the tabletop and the objects lying on it would be different if the number of coins among those objects was different; nonetheless, the number of coins does not affect me. For every statement made in this note about the number 2 and the experience of looking at a pair of coins, an exactly parallel statement can be made about the color sky-blue and the experience of looking at a sky-blue scarf. Thus, the color sky-blue can be “noncausally presented” to me in experience in a way that very closely parallels the way in which the number 2 can be noncausally presented to me in experience.

To return to concrete objects: if what I have said is correct, then concrete things are what David Armstrong has called blobs—they are without ontological structure. Their only constituents are their parts, their parts in the strict and mereological sense. Since the only concrete objects are substances, the only proper constituents—in any sense of “constituent”—of any concrete thing are smaller substances.

## 2. My Very Odd Ideas about the Mental and the Physical

These odd ideas tend to be consequences of my extreme ontological ideas. Let me try to explain how certain very abstract metaphysical ideas can have consequences for the philosophy of mind.

Let’s say that a mental property is a property that entails either thought or sensation—a property such that, necessarily, whatever has it is either thinking or (inclusive) sensing.<sup>4</sup> For example, thinking about Vienna, understanding Brouwer’s proof of his fixed-point theorem, being in pain, and sensing sky-blue are mental properties.

Philosophers, and particularly philosophers of mind, like to talk about mental states. What could these mental states be—according to those who accept my extreme ontological ideas?

Well, what but certain mental properties? What could my mental state at  $t$  be if not the conjunction of all the mental properties I have at  $t$ ? What could mental states be if *not* properties? They’re certainly not substances, are they? Substances like the persons or beasts whose mental states they are?

“But that can’t be right,” says the Interlocutor, “because if it were right, then mental states would exist *necessarily*. They’d exist *no matter what*. They’d exist in possible worlds in which nothing is sapient and nothing is sentient. And they’d exist at all times. Consider the mental state of some English soldier at some moment during the battle of Culloden—a complex mixture of fear, excitement, calculation, and all manner of desires and beliefs—a desire to avoid being killed, for example, and a belief that he had been born early in the eighteenth century. If mental states are, as you suppose, *universalia ante res*, that momentary mental state exists now just as surely as it existed in 1745; and it already existed when dinosaurs roamed the earth.”

---

<sup>4</sup>Since impossible properties entail every property (a fact that has the consequence that, by the strict terms of the definition, the property “is both spherical and cubical” is a mental property), some might wish to say instead that a mental property is a *possible* property that entails thought or sensation. The alternative definition will, of course, entail that there are no impossible mental properties—that, for example, the impossible property “being unaware that one is in excruciating pain” is not a mental property. Either definition will have the consequence that “knowing that Obama is the President” and “being a six-foot-tall man who enjoys music” are mental properties. I am willing to accept this consequence.

I have no reply to this other than to say that I don't know what a "state," mental or physical, could possibly be if it were not an attribute. And I don't see how an attribute can possibly avoid being a necessarily existent thing. (Suppose that there is a possible world  $w$  in which a certain attribute does not exist. The attribute "being cubical," say. If that attribute does not exist in  $w$ , then it is hard to see how *propositions* about cubes—the proposition that there are cubes, the proposition that it is possible for there to be cubes—could exist in  $w$ . And if it is possible in  $w$  for there to be cubes, it's hard to see how it could fail to be the case that the proposition that it is possible for there to be cubes could be anything other than true in  $w$ . It follows that if the attribute "being cubical" does not exist in  $w$ , then it is not possible in  $w$  for there to be cubes. And from *that* it follows that the actual world, which is well endowed with cubes, is not a possible world in  $w$ . That is, the accessibility relation is not symmetrical and this very situation in which we find ourselves might have been not only nonactual but impossible. I find it a lot easier to believe that attributes are necessarily existent than I do to believe that this actual state of affairs might have been impossible.)

My extreme ideas about ontology also imply either the falsity or the mere vacuous truth of the so-called identity thesis—the thesis that every mental event is a physical event. They have this implication because they imply that there are no events, no events of any description, either mental or physical. I must point out that the thesis that there are no events is obviously not the same thesis as the thesis that substances never gain or lose properties or never begin or cease to stand in certain relations. I grant the substances and the properties and the relations, but I see no reason to affirm the existence of items denoted by phrases like "the acquisition of the property hunger by the substance Socrates" or "the substances Socrates and Xanthippe coming to stand in the relation 'marriage.'" I have recently read the draft of a book by a very famous philosopher that contains the following argument: When a cold poker becomes hot, that is a change, and, therefore, changes exist. (The count-noun "change" is, so I am willing to grant, synonymous with "event.") This argument, however, is formally invalid (even if its conclusion is true). It is formally invalid because its premise contains a pronoun, the demonstrative pronoun "that," which has no antecedent. Its formal invalidity is precisely analogous to the formal invalidity of "This poker is hot, and that's a property. Therefore, there are properties." (That argument has a true conclusion but is nonetheless invalid.)

The thesis that there are, speaking strictly and philosophically, no events, obviously has consequences for the philosophy of causation, since many treatments of causation involve quantification over events, but I'll put off talking about that matter till later.

Let us now turn to the topic of Cartesian or Platonic dualism—substance dualism. I regard that thesis as false, but not because it's in conflict with my extreme ontological ideas or my odd ideas about the mental/physical opposition. Cartesian dualism is false, I say, but perfectly intelligible. I'm less happy about the intelligibility of "property dualism." How can this thesis be stated? Perhaps like this?

There are physical and nonphysical properties; mental properties are among the nonphysical ones.

But this proposal raises a difficult question: What are physical and nonphysical properties? We've seen what mental properties are—but what are physical properties, and, more to the point, what are *nonphysical* properties?

Nonphysical properties cannot be understood as properties that are not physical, properties that are not physical things or physical objects, for on that reading of “nonphysical property,” all properties are nonphysical properties.

Are nonphysical properties, then, properties that *entail* the property of being a nonphysical thing, that is, properties that can belong only to nonphysical things? (That would be a definition parallel to our definition of “mental property”: mental properties are properties that can belong only to sapient or sentient things.) Well, certainly not according to those philosophers who profess and call themselves property dualists, for one of the core theses of property dualism is that physical things not only can have but do have nonphysical properties: some physical things have mental properties, and mental properties, according to property dualism, are nonphysical properties.

Might a nonphysical property be a property that does not entail the property of being a physical thing but is consistent with that property? This is a less obviously objectionable proposal, but it has the consequence that the property “thinking about Vienna” is a nonphysical property only if it is possible for there to be a nonphysical thing that thinks about Vienna. If this proposal is accepted, therefore, property dualism will entail the metaphysical possibility of substance dualism (or of idealism), and that is an unwanted consequence of the definition: property dualists will not be happy about being committed to the thesis that it's metaphysically possible for there to be immaterial or nonphysical thinkers. (It's not that they are committed to denying that thesis, but they'd certainly prefer not to have to affirm it as a logical consequence of their theory.)

I have been unable to find in print a definition of “nonphysical property” that seems to me both to be intelligible and likely actually to capture what property dualists mean by the phrase. Therefore, at one point, I adopted the simple expedient of *asking* selected property dualists what they meant by “nonphysical property.” Here's what I learned when I subjected David Chalmers to an extended line of Socratic questioning:

A non-physical property is a property of some physical things such that the pattern of its instantiation among physical things does not supervene on the totality of the distribution of matter and radiation in space-time.<sup>5</sup>

---

<sup>5</sup> Possibly this was intended only as a sufficient condition for a property's being nonphysical. Possibly Chalmers did not intend his definition to imply that, as a matter of metaphysical necessity, a nonphysical property can belong only to a physical thing.

Property dualism can therefore be framed in these words:

In any pair of possible worlds in which matter and radiation are distributed in space-time in exactly the same way, every physical being—and, in particular, every living organism—in one of those worlds, will have a “counterpart” in that other world; and there will be pairs of “same distribution” worlds such that an organism in one of them and its counterpart in the other will differ in their mental properties. Take you, for example, you as you are in the actual world. There are worlds in which matter and radiation are distributed just as they are in the actual world and in which your counterpart has different mental properties from yours. Indeed there are “same distribution” worlds in which you have a counterpart who is a “zombie”—a creature whose behavior and physiological and anatomical structure are the same as yours but who has no mental properties at all, a creature that neither thinks nor feels.

(I should remark that my use of the word “counterpart” is not meant to bring David Lewis’s modal ontology to mind. If  $w_1$  and  $w_2$  are two “same distribution” worlds, and if  $x$  exists in  $w_1$  and  $y$  exists in  $w_2$  and is the counterpart of  $x$  in  $w_2$ , then  $y$  may or may not be  $x$ ; our statement of property dualism is noncommittal on that question.)

This is a thesis I can understand. I’m not sure why anyone would want to call it a form of *dualism*, but I won’t go into that question. In any case, whatever the thesis is called, I reject it. This rejection, like my rejection of substance dualism, is not a consequence of my extreme ontological ideas—although it probably does reflect the fact that Chalmers and I have very different ideas about modal epistemology. I simply think it’s false: it seems quite evident to me that if I am (as I suppose myself to be) composed entirely of quarks and electrons, then my intrinsic properties supervene on the distribution of matter and radiation in space-time (and necessarily so). More generally: If God’s creation is entirely physical, then, once he’s ordained a distribution of matter and radiation in space-time, there’s nothing more for him to do (*qua* creator, at least). If he has ordained the actual distribution of matter and radiation in space-time, he has thereby caused *every* proposition about created things to be true or to be false.<sup>6</sup> The truth of the following propositions, for example,

There are over five hundred thousand species of beetles.

It’s easier for a German to learn Dutch than it is for an Italian.

---

<sup>6</sup>Here I assume, in constructing a theological “intuition-pump” that is intended to support a non-theological metaphysical thesis (the thesis that if substance dualism is false, then the pattern of the instantiation of mental properties in space-time supervenes on the distribution of matter and radiation in space-time) that God ordains the distribution of matter and radiation in space-time. I hasten to assure the reader that I do not believe this: in my view, many aspects of the distribution of matter and radiation in space-time are due to chance.

Most mathematicians do not regard either mathematical logic or statistics as real mathematics.

Bratislava and Vienna are closer to each other than any two other capital cities.

Some “ancestral Africans” (some human beings all of whose human ancestors were born in Africa) are more closely related to most Swedes than they are to some other ancestral Africans.<sup>7</sup>

is a metaphysically necessary consequence of the distribution of matter and radiation in space-time that he has ordained. To recapitulate: two important metaphysical theses that I reject, theses concerning the relation between the mental and the physical (substance dualism and—so called—property dualism), are both consistent with my odd ideas about the relation between the mental and the physical.

### 3. My Very Odd Ideas about Causation

These ideas can be summarized in just a few words:

Causal relations exist (and many of them are instantiated). Causal explanations exist (and some are sometimes correct). Causation does not exist.

But perhaps this summary does not wear its sense on its sleeve. Perhaps these words require some sort of commentary or gloss. I will try to provide it.

(1) Causal relations exist but causation does not exist.

Causal relations, as I see matters, are relations that hold, not between events, but between substances. They are relations that are expressed by verbs like “push,” “press,” “kick,” “kiss,” and so on. Even if you took the disjunction of them all—“influences”? “affects”? “acts on”? “has an effect on”?—it would still be a relation that held between and only between substances. This, then, is what I mean by saying that causation does not exist: there is no relation whose properties in any way resemble the properties that philosophers ascribe to causation.

To a certain extent—I admit—the wording of my thesis (“Causation does not exist”) was chosen for dramatic effect. We might distinguish causation the *relation* from causation the *phenomenon*. The phenomenon of causation consists simply in the fact that things act on other things: shoppers carry parcels, sunlight warms

---

<sup>7</sup>Three of my example sentences by which I identify the propositions contain proper nouns. If you pressed me, I'd probably be willing to say that what really supervened on the distribution of matter in space-time would be the truth of the propositions expressed by sentences obtained from these sentences by replacing the proper nouns they contain with qualitative descriptions having the same referents. (For example, replace “Africa” with a description along the lines of “the continent having such and such a size and shape and such and such a geological history on a planet with the following intrinsic and extrinsic features...”).

stones, falling water turns waterwheels, children annoy their parents. But, I contend, causation-the-phenomenon is not identical with, is not reducible to, does not supervene on the holding or not holding of a relation called “causation” between the members of each pair of events (indeed, its reality is consistent with there being such things as events at all).

I should say, too, that my thesis does not imply that if someone, engaged in the ordinary business of life, says “Her death was caused by a traffic accident,” then that person says something false. It does imply that the truth of that person’s assertion does not entail that the words “her death” denote an item that stands in a relation expressed by “*x* was caused by *y*” to an item that belongs to the extension of “*x* is a traffic accident.” Nor, I maintain, does, a piece of discourse like, “Mrs. White, Detective O’Malley and I are just a little unclear about the order in which things happened. We’d like you to describe carefully the sequence of events that led up to your husband’s death.” presuppose that, in addition to substances, attributes, and relations, there are objects that are the acquisitions of attributes by substances or the “comings-to-stand” in relations by pluralities of substances.<sup>8</sup> When Mrs. White says, “Well, my husband came through the door, and then he tripped over the dog and hit his head on the corner of the kitchen table,” her statement does not imply—at any rate, it need not be taken to imply—that there are objects that are the referents of phrases like “my husband’s coming through the door” and “my husband’s tripping over the dog.”

And why do I say that there is no such relation as causation? Well, one reason, of course, is that my ontology implies that there are no events, and thus no objects of the sort that that relation requires as its terms. But if there were good accounts of the phenomenon of causation that identified it with the pattern of instantiation of a relation that holds only between events (and which, as part and parcel of this task, presented us with a logically serviceable account of the ontology of events), I’d certainly want to modify my ontology to include events. But I have not been impressed by the attempts of a vast army of very able philosophers either to “define causation”—which seem to lead only to ever-more-elaborate epicycles—or to provide an ontology of events.

I want to suggest a way to understand our causal discourse that does not presuppose that there is such a relation as causation. I propose to understand our causal discourse in terms of *causal explanations*. For I say that

- (2) Causal explanations exist.

That is to say, people explain things, and causal relations play a central and essential role in some of these explanations (particularly in explanations of contingent states of affairs).

---

<sup>8</sup>I use “object” as the most general count noun: everything is an object. Those who like to oppose the terms “object” and “event” may wish to use some other term as the most general count noun: “item,” perhaps, or “thing.”

Sometimes our causal explanations are “achronic,” like the following explanation (essentially Aristotle’s) of why the Earth is a ball: if it *weren’t* a ball, the force of gravity would immediately pull it into that shape. Newton’s explanation of why the planets move in elliptical orbits with the Sun at one focus is also an achronic causal explanation. Achronic explanations provide answers to “Why?” questions: “Why is the earth a ball?”; “Why do the planets move in elliptical orbits?”—or even “Why does the car make that funny whuh-whuh-whuh noise when it goes over forty?” (Of course, the car’s making that funny noise is no doubt a very temporary state of affairs, even in relation to the life of the car. I call an explanation like “A strut is out of alignment” “achronic” to call attention to the fact that the state of affairs cited in the *explanans* obtains simultaneously with the state of affairs that is the *explanandum* and, as it were, underlies it.) Other causal explanations, however, take the form of stories or histories or narratives. Such explanations are generally more easily thought of as answers to questions that begin with “how” than questions that begin with “why.” Questions, for example, like “How did Winifred die?” or “How did the lion escape from its cage?” or “How did the strut that you say is out of alignment *get* out of alignment?” It is causal explanations of this latter sort—causal narratives I shall call them—that I am primarily interested in in the present essay. Causal narratives are answers to questions of this general sort:

The so-and-so now has the property F (or the so-and-sos now stand in the relation R). How did that happen? How did things get to be that way?

That is to say, someone observes that *things have changed*: a certain object now has a certain property or a certain relation now holds among certain objects, and that person asks why things are *now* this way when they *didn’t use to be*.

And typically—almost always—the answer to such questions will consist in a description of how those current properties of, and those relations that now hold among, various objects are a consequence of how those objects (and perhaps other objects) have acted on one another. In some of the simpler cases, cases involving objects like billiard balls, the answer will consist in a description of how things have moved and have been moved, on how they have pushed one another and have been pushed by one another.

Causal narratives contain and depend essentially on causal verbs, verbs like “push,” “pull,” “strike,” and “turn,” “annoy,” “comfort,” and “kill.” I want to make a few remarks about the logic of these verbs. We may distinguish causal verbs, which are typically transitive verbs,<sup>9</sup> from “verbs of change”—verbs like “move,” “break,” “become warmer,” and “grow angry” (all of them intransitive). (Note that

---

<sup>9</sup>The intransitive verbs “eat” and “write” are causal verbs—being essentially abbreviations of “eat something” and “write something.” (See note 12.) If “commit suicide” is a verb, it is an intransitive causal verb.

in English and many other natural languages, the “same” verb may be transitive or intransitive, depending on context: Aristotle’s hand moved; Aristotle’s hand moved the staff.) We can think of verbs as predicates. For example, “**1** moves” or “**1** moves **2**.”<sup>10</sup> (This has the advantage of clearly distinguishing transitive from intransitive verbs and, therefore, of introducing a clear syntactical difference between causal verbs and verbs of change.)

I introduce a predicate-operator “C.” This operator takes two predicates and makes a single predicate, a causal verb. For example, the expression

C **1** strikes **2**; **2** moves.

is a causal verb. It may be read as “**1** strikes **2**, and, as a result, **2** moves” or “**1** strikes **2**, and, in consequence, **2** moves” or (rather more informally, since this predicate does not contain “**2** moves”) “**1** strikes **2** thereby causing **2** to move.”<sup>11</sup>

Other examples are: “C **1** will shout; **1** will warn **2**” (“**1** will shout and thereby warn **2**”) and “C **1** strikes **2**; **1** destroys **2**” (“**1** strikes **2** thereby destroying it”) and “C **1** strikes **2**; **1** destroys **1**” (“**1** strikes **2** thereby destroying itself”); “C **1** moved; **1** began to roll”<sup>12</sup> (“**1** moved, with the consequence that it began to roll”).

The purpose of “C” is, of course, to increase the available stock of causal verbs. If we had not already had the causal verb “move” at our disposal, we could have generated an equivalent causal verb from the “all-purpose” causal verb “act on” by applying “C” to that verb and the verb of change “move”:

C **1** acts on **2**; **2** moves.

<sup>10</sup> The boldface numerals contained in these “predicates” are a device that allows us to use reference to predicates to make general statements about open expressions (expressions containing free variables) without displaying any particular variables. For example, “**1** moves **2**” stands in for any open sentence in which the same variable replaces a numeral at each occurrence of that numeral: “*x* moves *y*”; “*y* moves *x*”; “*z* moves *z*.”...

<sup>11</sup> Predicates are of two kinds: *perfect* and *imperfect*. Perfect one-place predicates contain any number of occurrences of the boldface numeral **1** and contain no other boldface numerals; perfect two-place predicates contain any number of occurrences of both **1** and **2** and contain no other boldface numerals—and so on. Any predicates that are not perfect are imperfect. Examples of imperfect predicates are “**2** moves” and “**1** moves **3**”—predicates with “missing” numerals. Imperfect predicates can occur in our causal discourse only as components of perfect predicates. An expression formed by prefixing the operator C to two predicates (both perfect, one perfect and one imperfect, neither perfect) is well formed only if it is perfect. For example, the expression “C **1** strikes **2**; **2** moves” is a perfect two-place predicate (and hence is well formed)—formed from a perfect and an imperfect predicate—because it contains occurrences of both **1** and **2** and contains no other boldface numerals. In contrast, “C **1** strikes **3**; **3** moves” and “C **2** strikes **3**; **3** moves” are imperfect and not well formed.

<sup>12</sup> An intransitive causal verb (see note 9). Any predicate formed by prefixing the operator C to two predicates each of which contains **1** and no other boldface numeral will be an intransitive causal verb—although of a sort not found in English. (Not found as a one-word “dictionary-entry” verb. The English verb phrase “**1** moved, with the consequence that **1** began to roll” is an intransitive causal verb *phrase*.) Ordinary-language intransitive causal verbs are generally produced by—in effect—quantifying into transitive causal verbs (the variable occupying the direct-object position). For example, the English intransitive causal verb “eat” or “**1** eats” is equivalent to “ $\exists x$  **1** eats *x*.”

But, of course, we already have the transitive “move.” Here is a verb that has no equivalent in ordinary speech:

C **1** acts on **2**; **3** moves.

This verb means something like “act on a thing, thereby causing a thing (not necessarily *that* thing) to move.”

Our predicates, are, of course, *predicates*. That is, the boldface numerals that occur in them represent “sites” at which a term (a variable or a denoting phrase) can occur. We can thus say or write

$\exists x C x$  struck the ball; the ball moved,

which may be read “Something struck the ball, thereby causing it to move.”

I contend that the only “causal” vocabulary that a causal narrative need contain is a stock of causal verbs sufficient for our purposes and the operator “C.” (And if we have the operator “C” in our vocabulary, the single causal verb “act on” will comprise a “sufficient stock” of causal verbs for any purpose.) And I contend, in presenting a causal narrative, one need quantify only over objects of the following two kinds: (i) concrete objects, objects capable of acting on other objects (and perhaps on themselves) and being acted on by other objects—that is, substances; (ii) abstract objects—that is, properties and relations.<sup>13</sup> In sum: in presenting a causal narrative, one need refer to neither the relation philosophers call “causation” nor any of its supposed relata, events.

#### 4. Bringing It All Together

Finally, I want to bring this all together—everything I’ve said about ontology, the mental, and causation—by considering the relation between two kinds of causal narratives:

Causal narratives or histories (largely imaginary) involving only the vocabulary of physics

Causal narratives or histories of the everyday sort, narratives involving the vocabulary we use in everyday life, including our everyday mental vocabulary

Let us begin by considering a narrative that is permeated by this everyday mental vocabulary. Suppose that a friend asks Alice this question: “I see that Tom has a black eye. How did that happen?” And suppose that Alice responds with the

---

<sup>13</sup> A person presenting a causal narrative will almost certainly have to quantify—*apparently* to quantify—over “times” or “occasions” or something of that sort (and, quite possibly, over “places” as well). One problem that faces my very sparse ontology is this: how to understand apparent quantification over times (and, it may be, places) as being, in the final analysis, quantification over substances, properties, and relations—and nothing else. I will not consider this problem in the present essay.

following causal narrative—a narrative that provides an answer to this “How did it happen?” question and in that sense constitutes an explanation of the fact that Tom has a black eye:

Well, you know what a jealous husband Fred is. At the party last night, Fred confronted Tom and demanded to know whether Tom was having an affair with his wife, and Tom confessed that he was. That enraged Fred and he punched Tom in the face.

This causal narrative is certainly “permeated by our everyday mental vocabulary.” (In addition to obvious pieces of mental language like “jealous” and “enraged,” there are words like “confront” and “demand” and “confess” that can be applied only to sapient beings.)

Contrast this “mental” narrative with what we might call “God’s physical narrative”—a narrative that describes the behavior of the individual elementary particles that collectively compose Tom and Fred and some significant portion of their environment—over a certain stretch of time.

Of course God’s narrative wouldn’t strictly speaking constitute an explanation of Tom’s black eye or of Fred’s having struck Tom; it would rather constitute an explanation of how a certain set of elementary particles came to be arraigned in a certain way at a certain time. Nevertheless, the arrangements of those particles at various points in God’s narrative were such that the states of affairs we mere human beings call “Fred’s striking Tom in the face” and “Tom’s having a black eye” supervened on those arrangements. That is to say, God’s narrative explains a lot *more* than Fred’s striking Tom and Tom’s having a black eye, but in some sense it explains those things among many other things—all the things that supervene on the truth of the narrative.

Now let’s leave these two explanatory narratives aside for the moment and consider two other explanatory narratives. We will first consider an “everyday” physical explanation—an explanation conducted in the language of everyday life that involves nothing more than changes in and interactions among inanimate objects. A child asks: “Daddy, when you put the key into the keyhole and turn it, the door stops being locked. What makes that happen?” And Father responds to this question with a causal narrative:

Well, sweetheart, here’s what happens when you turn the key in the keyhole. There are little teeth and slots on the key and the key fits into a thing inside the lock that also has little teeth and slots. The teeth and slots on this key just match those on the thing inside the lock, so it’s possible to turn the key; when the door is locked, there’s a little bolt that sticks out of the lock into the doorjamb, and turning the key pushes the bolt out of the jamb and back inside the lock—that’s what we call “unlocking the door”—and then we can push the door open.

Contrast this narrative with “God’s narrative”—a story of the evolution of an enormous assemblage of elementary particles, a story that ends with a description of a

vastly intricate distribution of particles that “includes” the door’s being unlocked: that is, the truth of the proposition that the door is unlocked supervenes on that distribution.

Now let us compare these two pairs of narratives—on the one hand, the human story of how Tom’s eye came to be black and the divine story of the evolution of a system of elementary particles into a state on which Tom’s having a black eye supervenes, and, on the other, the human story of how turning the key leads to the door’s being unlocked and the divine story of the evolution of a system of elementary particles through a state on which the key’s turning supervenes into a state on which the door’s being locked supervenes.

The “human” narrative that explains how Tom got a black eye and the “human” narrative that explains how turning a key leads to an unlocked door, and the two God’s-eye stories of the evolution of systems of elementary particles, are all four of them explanatory narratives, causal narratives. All are stories whose only characters are substances and attributes and relations; in each case, the successive “chapters” of the story are episodes of substances acquiring or losing certain attributes or coming to stand in certain relations and ceasing to stand in certain relations.

Let’s consider Father’s explanation of how turning a key unlocks a door. I can see no reason to think that his explanation isn’t correct. That is, I see no reason to doubt either of the following statements:

- (a) When Father says, “Well, sweetheart, here’s what happens when you turn the key in the keyhole,” that is a true statement about the story he is about to tell the child. Of course, that story doesn’t include *everything* that goes on inside the key and the lock and the door and the bolt and the jamb. It’s not God’s narrative. God’s narrative includes (or it does if current physics is more or less right) lots of statements about the exchange of photons by charged particles, and Father’s narrative includes nothing about photons or charged particles. Nevertheless, Father speaks the truth about the narrative he is about to present when he says, “Here’s what happens when you turn the key in the keyhole.”
- (b) If Father’s preface to his narrative is a true statement, then his causal narrative does count as an answer to his child’s request for an explanation. And that is what a *causal explanation* of *x*’s having come to be *F* is: A correct answer to the question “How did *x* get to be *F*?” in which causal verbs “fit” “turn,” and “push” play a central and essential role. (And in which my operator “*C*,” or everyday expressions that do much the same work—“thereby causing,” and “and, as a consequence”—play an essential role.)

No one would suppose, I think, that Father’s statement “Here’s what happens when you turn the key in the keyhole” is falsified or vitiated or in any way undermined by the fact that God can give an unimaginably intricate explanation of a vastly complex state of affairs that in a certain sense “includes” the door’s being unlocked after the key has been turned.

Is there any reason to think that the fact that God can present a causal narrative—a narrative couched entirely in terms of the interactions of elementary particles, one that in this same vague sense includes Fred’s striking Tom and Tom’s consequent black eye—in any way falsifies or undermines or vitiates the following explanation?

Well, you know what a jealous husband Fred is. At the party last night, Fred confronted Tom and demanded to know whether Tom was having an affair with his wife, and Tom confessed that he was. That enraged Fred and he punched Tom in the face.

(Or that it falsifies or vitiates or undermines the claim that that little narrative is a correct answer to the question, “How does Tom come to have a black eye?”) This explanation contains “mental” language. Father’s explanation of the unlocked door contains no mental language. Does that make a difference?—a difference that would have the consequence that the latter explanation was correct but the former incorrect?

Well, there are arguments that might suggest something of that kind, arguments associated with the work of Jaegwon Kim.<sup>14</sup> (These arguments are thought by many, at the very least, to pose a genuine and difficult philosophical problem.) Here is one argument of this sort:

If mental states supervene on physical states, and if the physical states of the things that make up the world at a given time cause all subsequent physical events, it cannot be that such mental states as some things may have or be in at a certain time cause subsequent physical events.

Or, if this is thought by some to fall short of being a convincing argument, they may still find its “interrogative counterpart” to be a troubling philosophical question:

If mental states supervene on physical states, and if the physical states of the things that make up the world at a given time cause all subsequent physical events, how can it be that such mental states as some things may have or be in at a certain time cause subsequent physical events?

For example: how can it be that a sudden access of pain causes me to wince; or that my desiring to vote for the measure and my belief that raising my right hand will be a vote for the measure jointly cause my hand to rise, if my having the pain at  $t$  or having the desire and the belief at  $t$  supervene on the distribution of matter and radiation at  $t$ , and that that distribution is causally sufficient for my wincing or my raising my hand shortly after  $t$ ?

---

<sup>14</sup>Kim has presented this argument in many places. A representative statement can be found in his *Physicalism, or Something Near Enough* (Princeton, NJ: Princeton University Press, 2005), 39. See also his essay, “Causation and Mental Causation,” in *Contemporary Debates in Philosophy of Mind*, ed. Brian McLaughlin and Jonathan Cohen (Oxford, UK: Blackwell, 2007), 227–42.

Suppose we agree—with those who pose this question—that the distribution of the instantiation of mental properties (properties that imply either or both of the properties “being a thing that thinks” and “being a thing that feels”) supervenes on the distribution of matter and radiation in space-time. But suppose we *also* affirm the following three theses: that mental states and physical states are abstract objects and thus have no causal powers whatever; that there are no events, either mental or physical; that while there are causal relations and causal explanations, there is no such relation as causation.

The interlocutor speaks: “But how can you possibly say that a physical state like ‘being red hot’ has no causal powers? If a poker is in that state, it obviously has the power to heat or burn things with which it comes into contact.”

And I reply: Yes, but the obvious truth of what you have said doesn’t imply that “being red hot” has the power to heat or burn things. “Being red hot” doesn’t *have* causal powers—it *is* a causal power. It is, as you have said: the poker that *has* the power to heat and burn things. Just as changes of position are motionless and shapes are shapeless, powers are powerless.

If these three theses are true—that mental and physical states have no causal powers, that there are no events, that there is no such relation as causation—then it’s hard to see how a Kim-style argument for epiphenomenalism (so to describe its conclusion) can even get started. It seems that the conclusion of a Kim-style argument for epiphenomenalism must be equivalent either to “Physical states can be causes of physical states and mental states cannot” or “Physical events can be causes of physical events and mental events cannot.”

It further seems that those who affirm these three theses are free to contend that explanatory narratives involving the language of everyday life (the language of everyday life involves mental vocabulary but also includes many words and phrases that apply to nonsentient and nonsapient objects, words and phrases like “turn,” “push,” “lock,” “bolt,” and “fit exactly into”) are by no means falsified or undermined or vitiated by the existence and correctness of God’s physical explanations—that is, explanations couched entirely in terms of the interactions among those fundamental physical entities on whose features and arrangement the truth values of the propositions expressed by sentences couched in the language of everyday life supervene. Mental language does not apply to elementary particles, but neither do words and phrases like “turn,” “push,” “lock,” “bolt,” and “fit exactly into.”<sup>15</sup> If the correctness of explanations that contain vocabulary of the latter sort (and no mental vocabulary) can supervene on the distribution of matter and radiation in space-time, why should it be that the correctness of explanations that involve mental vocabulary cannot also supervene on the distribution of matter and radiation in space-time?

---

<sup>15</sup>“Push” is a possible exception. If you think that one electron can “push on” another (in virtue of their both being negatively charged) in the same sense of “push on” as that in which a human being can “push on” a jammed door, I think you’re wrong, but I won’t argue with you; the other items in my list will suffice for my purposes.

I do not pretend to have shown that (or even to have presented an argument for the conclusion that) the correctness of explanations containing mental vocabulary *can* supervene on the distribution of matter and radiation in space-time. My purpose is only to call into question an argument for the conclusion that this *cannot* be—or for the conclusion that a substantive philosophical problem confronts those who say that this can be. I would point out, however, that it does seem to be the default position that explanations involving mental vocabulary are sometimes correct. And this default position entails the following conditional: *if* all truths supervene on the distribution of matter and radiation in space-time, then the correctness of some “mental” explanations supervenes on the distribution of matter and radiation in space-time.

## Response to Peter Van Inwagen, “Causation and the Mental”

Robin Collins

In responding to Peter, I will work back from the end of his paper toward the beginning. Peter concludes his paper by claiming that

If these three theses are true—that mental and physical states have no causal powers, that there are no events, that there is no such relation as causation—then it’s hard to see how a Kim-style argument for epiphenomenalism (so to describe its conclusion) can even get started.

The problem I have with Peter’s statement is that I do not see how Kim-style argument depends on these theses, and, thus, why the denial of these theses weakens such an argument. Let me elaborate by presenting one way in which a person could hold that there is a relation of causation in which a Kim-style argument fails. In particular, one could adopt a *nomological subsumption* account of causation. Under such an account, one could define a state C as constituting a complete cause of E if and only if: (1) C is temporally prior to E; and (2) there is a law such that if C occurs, then E will occur. Further, C and E could be either mental or physical states. Given that these two conditions are satisfied, C and E will have a corresponding relation between them: namely, the relation of being such that one of them is temporally prior to the other and that there is a law requiring that, if the state that is temporally prior exists, so does the temporally posterior state. One could then define a “causal relation” as this relation.

Does this notion of cause run into a Kim-style? Not necessarily. To see why, consider the regularist view of the laws of nature, in which metaphysically laws merely express universal regularities in the world. In physics, such a law would be expressed by an equation – say Newton’s law of gravity,  $F = Gm_1m_2/r^2$ .