

Philosophy Department

Articles

University of Connecticut

Year 2013

Mental Causation, Invariance, and
Teleofunctional Content

Crawford L. Elder

University of Connecticut - Storrs, crawford.elder@uconn.edu

Crawford L. Elder; Philosophy Department, Unit 2054; University of Connecticut; Storrs, CT 06269-2054; U.S.A. Forthcoming 2013 in a special issue of *The Monist* on the topic “constitution and composition”. Made available by kind permission of the editors. This manuscript is copyrighted: **Copyright 2009, Crawford L. Elder.**

Mental Causation, Invariance, and Teleofunctional Content

Abstract. Can we vindicate the common sense conviction that a person’s beliefs and desires are causally efficacious—that they shape that person’s behavior? Three obstacles stand in the way. First, there are no exceptionless laws linking beliefs and desires to behavioral outcomes; second, the microphysical (or neurochemical) states of affairs on which beliefs and desires supervene have as strong a claim as beliefs and desires themselves to causing behavioral outcomes; third, even if we could surmount the threat of causal exclusion, we would still face “the problem of mental quausation” for beliefs and desires—we would have to show that beliefs and desires cause outcomes in virtue of having the contents that they do. I argue that the “manipulationist” account of causation not only side-steps the first obstacle, but also shows that there is no competing causation at the level of microparticles or neurochemistry. I then briefly suggest that “the problem of mental quausation”, as applied to beliefs and desires, is confused.

Are a person’s mental states, in particular a person’s beliefs and desires, causally efficacious—do they shape that person’s behavior? Common sense is deeply convinced that the answer is Yes. Jerry Fodor voices this conviction when he remarks, after citing some examples of ways that mental events apparently shape behavior, “if none of this is literally true, then practically everything I believe about anything is false and it’s the end of the world” (Fodor 1989, p. 77). Yet persons are physical entities, made up entirely of the microparticles of physics, and this leads many philosophers to *doubt* whether it is literally true that mental states are causally efficacious.

There are two ways to express the idea that persons are entirely physical entities. One is that any person is constituted by an aggregate of microparticles; the other, that any person is composed of many, many microparticles. Constitution is always a one-one relation; composition is typically many-one. The difference between these two expressions may seem trivial. But some philosophers—notably Lynne Rudder Baker (2007)—have put forth views that suggest that *constitution by an aggregate* may make it easier to defend the mental causation in which common sense believes, than does *composition by many microparticles*.

This paper undertakes to vindicate the common sense belief in mental causation. But it does so on exactly the opposite basis. It argues that while *composition by many microparticles* initially appears to defeat the idea that mental states are causally efficacious, it actually entails that that idea faces no serious challenge. A key step in this entailment is that we have no reason to suppose that the many microparticles within a person *act as an aggregate*: that is, we have no reason to think that *the aggregate of microparticles* has any causal powers at all. Since “the constitution view” holds that the aggregate of microparticles does have causal powers, it actually *obstructs* the vindication of mental causation.

But how different the situation at first appears to be! For constitution is a phenomenon by which importantly different entities may occupy exactly the same volume of space, even across the whole span of their existences. Two such entities may differ in their modal profiles, for starters. Consider the well-worn example of the statue made entirely of clay. The lump of clay can survive being flattened, while the statue cannot. Yet the relation of constitution forestalls any mystery about how these distinct objects can manage to fit into the same space: the lump *constitutes* the statue, and constitution not only permits but requires exact co-location. Why, then, mightn’t the relation of constitution enable a person, whose mental states produce behavioral outcomes via mental causation, to be exactly co-located with a purely physical entity, whose physical states produce, via purely physical causation, physical outcomes which “compose into” or subvene exactly those behavioral outcomes? That is one of the questions which this paper will indirectly address.

I

What this paper will directly address is the question whether the causal efficacy of mental states (and in particular beliefs and desires) can be vindicated; the question of how well the constitution view assists that vindication will be considered only *pari passu*. To frame the discussion adequately, we must begin by noting that the vindication of mental causation is generally thought to confront three main obstacles. The first is the idea that any true causing must instantiate some exceptionless law (Davidson

1970). For it seems clear that the closest we can come to laws that link beliefs and desires to behavioral outcomes are mere generalizations that must be hedged by *ceteris paribus* clauses. The second obstacle comes from the thought that a person's mental states supervene upon particular neurochemical states characterizing his brain—and perhaps ultimately supervene upon extremely complex states involving microparticles within his brain. The subvening states of affairs, at either level, may seem to have at least as strong a claim to being the causes of those behavioral outcomes that the person's beliefs and desires seem to cause, as do those beliefs and desires themselves. Since it is hard to believe in systematic causal overdetermination of those behavioral outcomes, the subvening states of affairs thus appear to exclude the putative causation by the beliefs and desires. And even if we can dispel the threat of causal exclusion, we still face a third obstacle. Even if beliefs and desires can somehow be said to cause (aspects of) a person's behavior, there is still a question whether they can be said to do so *quâ* beliefs and desires—in virtue of having the particular contents they do. This third obstacle is an aspect of “the problem of mental quausation” (Horgan 1989; cf. McLaughlin 1993).

The defender of mental causation might confront these obstacles head-on, and that is what this paper will undertake to do. Alternatively, the defender might enlist “the constitution view” to argue that these obstacles are more fictitious than real—that they appear only if we force the kind of causation that runs from people's mental states to people's behavior onto a Procrustean bed suited only to the causation that microparticles exercise (or, perhaps, microparticles and likewise biological tissues). The defender who enlists the constitution view might, for example, say that while it is appropriate to restrict genuine causing to exceptionless causing—causing that can be subsumed under some exceptionless law—when dealing with causation by microparticles, or aggregates of microparticles, or elements of organisms, it does not follow that *mental states* genuinely cause behaviors only when *they* do so in accordance with exceptionless law. Such a philosopher might also contend that the worry about overdetermination is inappropriate. We can grant, this philosopher might say, that any individual behavioral outcome brought about by a person's mental states supervenes on some vastly complex state of affairs involving microparticles, and can allow that this microphysical state of affairs was caused by an equally complex

state of the aggregate of microparticles that constituted the person when the mental states obtained—a complex state on which those mental states themselves supervened. It still doesn't follow that the *patterns* of causation that lead from mental states to behavioral outcomes can be reduced, on a type-type basis, to *patterns* by which complex states of such an aggregate lead to complex events subvening behavioral outcomes, and neither does it follow that the mental-to-behavioral patterns capture causings that are any less real than the causings captured by the microparticle patterns. Finally, this philosopher might say that even though content-bearing states do not figure in the causal inputs proper to the microparticles (or aggregates of microparticles or biological elements), content-bearing states *could* perfectly well figure in the causal inputs proper to people as conscious agents. The premise that would motivate these charges of Procrustean misunderstanding would, precisely, be that the relation between any person and the co-located aggregate of microparticles—and, for that matter, the co-located biological organism—is *constitution*. Constitution is a form of “unity without identity” (Baker 2007, Ch's 2 and 6). The co-located entities between which it holds can be importantly different from one another. They can differ in their modal profiles. They likewise can differ in their ways of causing. The so-called obstacles to believing in the efficacy of mental states arise only from treating the person as if it *is*, rather than *is constituted by*, either the aggregate or the organism.

But to the philosopher who takes the standard obstacles to mental causation seriously, such a reply may sound more like a bare assertion that mental causation is genuine, than like an argument to that effect. This reply offers no *argument*, for example, against the claim that all genuine causing is causing in accordance with some exceptionless law. If that claim is indeed true, what follows is not that the regularities that subsume causings-of-behavior by mental states are less than exceptionless. What follows is instead that the genuine causation of human behaviors can be traced only by using the vocabulary of microphysics—or possibly the vocabulary of biochemistry—and that mental states can qualify as causes only if they can be said to *be* complex states of affairs involving microparticles or biochemical elements. Further, if content-bearing states do not figure in the causal inputs characteristic of a microparticle system, or a biochemical system, then what will follow is that they do not figure among the causes of

human behavior. And even the response offered to the worry about overdetermination appears inconclusive. The response does not speak to the point that in any individual instance where mental states in a person have supposedly caused some behavior by that person, the person's microparticles will have had to "go along for the ride". Not one bit of human behavior can occur without microparticles in that person having moved and altered in particular ways. Yet microparticles within a person, like microparticles outside the person, are constrained in their every movement and every alteration by the laws of microphysics—microphysics, as it is commonly thought, is closed and complete. How then can it happen that microparticles, which individually are just "minding their microphysics", manage collectively to subvene the patterns that lead from particular types of mental state to particular types of behavior? One has to suspect that in any individual instance in which this happens, *what it is* for the microparticles to subvene such a pattern just *is* for them individually to be "minding their microphysics". But the worry about overdetermination then remains in place. For any individual case in which a mental state putatively causes behavior, there will be two accounts of the causation that is taking place. One will run from complex microphysical causes to the complex microparticle outcome that subvenes the behavior; the other, from mental states to behavior. But only the former account will invoke exceptionless laws. The threat of causal exclusion will remain.

This paper will therefore tackle the three obstacles to mental causation head-on. It will not invoke "the constitution view" as a premise, although in the following sections it will offer an argument that helps one element of that view. It will lay greater emphasis on the first two obstacles, and will then briefly sketch a reply to the third. It begins by arguing that, to judge by recent work by Woodward and Hitchcock, many genuine cases of causation instantiate only some explanatory invariance—where an "invariance" is a generalization that need not be exceptionless (Woodward and Hitchcock 2003a, 2003b; Woodward 2003; Hitchcock 2001a, 2001b). (Here is where help is provided to "the constitution view".) It next argues—in a way not anticipated by Woodward or Hitchcock—that there is typically *no* invariance, of the sort crucial for a claim of causation, linking a complex microparticle state or event in the brain to those behavioral outcomes that common sense credits beliefs and desires with producing. So

the same reasoning that defends mental causation against the first of our obstacles defeats the second obstacle as well: it shows that at the level of the microparticles, there is no causation that threatens causal exclusion at all. Nor will it help the causal exclusionist to move up one level, to neurochemical states in the brain: here too no invariance links the would-be excluding states to behavioral outcomes, except insofar as we take those states as subvening causally efficacious beliefs and desires.

Finally, this paper argues briefly that “the problem of mental causation”, for beliefs and desires, rests upon a confusion. It supposes that what it is for beliefs and desires to have their characteristic contents is one thing; that their causing—or putative causing—of behavioral outcomes is another; and that we can coherently ask whether it is in virtue of the first thing that the second thing obtains. On the view which I will sketch—a view put forth over many years by Ruth Millikan (Millikan 1984, 1993, 2000)—*what it is* for beliefs and desires to have their contents just *is* for them to cause, across “historically normal” circumstances, particular shapings of behavior. The having of content *includes* causal shaping of behavior, and so one cannot properly ask whether it *underlies* shaping of behavior. Millikan’s view pictures beliefs and desires as states that naturally-selected programs or devices in the brain assume, by virtue of their design. Thus beliefs and desires are themselves biological products. And biological products are par excellence entities that produce outcomes not in accordance with exceptionless laws, but only in accordance with explanatory invariances.

II

A line of thought that has been influential in philosophy of mind ever since Davidson’s paper “Mental Events” (1970) maintains that every true causing must instantiate some exceptionless law. This line of thought furnishes the first of my three obstacles to vindicating mental causation. For it does seem clear that any so-called law linking beliefs and desires to behavioral outcomes would have to be hedged by *ceteris paribus* clauses, clauses that have the effect of saying “except when this ‘law’ fails to hold”. But recent work by Hitchcock and Woodward has put in question the general view on causation that this

line of thought defends. The application of that general view to beliefs and desires seems particularly questionable if we suppose—as I shall be assuming—that beliefs and desires are biological products. For biological products—the traits and devices that natural selection fashions—seem to be characterized by a kind of causality that the insistence on exceptionless laws would rule out. Biological products are therefore the first place we should look, if we want to see reason for suspecting that there is something wrong with the “exceptionless laws” view on causation.

Evolutionary biology claims to identify and study traits and devices that are characterized not just by a recurrent morphology but also by a shared history of natural selection. Tokens of these traits proliferate across the generations of a lineage, the thinking runs, as a causal consequence of effects that ancestor tokens produced in earlier generations—effects that increased the reproductive fitness of their host organisms. Evolutionary biologists commonly take this sort of shared history to be a necessary condition for two individual traits to be tokens of the same type: the same trait occurs in two different organisms, evolutionary biologists commonly judge, only if the two tokens share not just a common morphology, but also stem from the same sort of selectional history. Thus evolutionary biologists commonly take it as essential, to the individual traits and devices belonging to a common type, that they are replicated from ancestor tokens that produced a particular fitness-enhancing effect.

But the crucial point is that these ancestor tokens need not have invariably—exceptionlessly—produced that beneficial effect. Natural selection does not insist on perfection. All that is needed is that the ancestor tokens have *often enough* produced the beneficial effect—often enough to give the genes that code for them selectional advantage over their alleles. Tokens of a given biological trait or device are, essentially, such as *often enough* to cause a particular kind of effect. They almost never are such as *exceptionlessly* to cause that effect.

How might this image of biological design be reconciled with the thesis that all genuine causing is exceptionless causing—causation that can be subsumed under some exceptionless law? Davidson’s own way of handling the apparent causal efficacy of mental events provides the first and most salient suggestion. Exceptionless laws do obtain at the level of fundamental microphysics, Davidson held, and

so one need only view any individual mental event as just *being* some highly complex microphysical event, and the way is cleared for seeing how that mental event might produce a behavioral outcome in accordance with some exceptionless law—provided, of course, one likewise views the behavioral outcome itself as just *being* some further, equally complex, microphysical event. In parallel fashion, one could argue that whenever a particular biological trait or device produces the sort of outcome, for which its biological lineage was naturally selected, what really has happened is that some massively complex event involving a vast plurality of microparticles has, in accordance with the laws of microphysics, produced some further and equally complex event involving microparticles.

But this response incurs a cost. It suggests that the taxonomy proper to evolutionary biology does not carve events in the world along causally significant lines. For if the instances of a given type of biological device are viewed as enormous aggregations of microparticles, instances of the *same* biological type turn out to be instances of highly *divergent* (and highly complex) microphysical types. One preserves the idea that instances of a given biological type often enough produce a characteristic sort of outcome, but preserves it by representing the different instances of that outcome as being wildly diverse outcomes of a microphysical character. Indeed one reduces away the central biological idea as to what causes instances of the given biological trait or device to get replicated in generation after generation after generation. The central biological idea is that ancestor devices in such a lineage often enough caused the same sort of outcome, that this common causal history itself caused the replication of the lineage, and that this common causal history is part of what holds current instances of the lineage together as a kind. If real causation is causation at the level of microphysics, not one element of this idea remains in place. Evolutionary biology, we must rather judge, studies a strangely sustained illusion: really, highly divergent aggregations of microparticles, located at different points in time, cause divergent microphysical outcomes, but sustain at the level of biological description the appearance that the same sort of causing has—often enough, and not invariably—not only occurred at these different points in time, but has caused fresh instances of such causing.

There is a second and less radical way in which one might try to reconcile the standard image of biological design with the idea that all true causing fits under exceptionless laws. One might argue that there is after all exceptionless causing at the level of biological description itself. The idea would be that, for any (type of) biological trait or device, there are “historically normal circumstances”, of a particular qualitatively specifiable type, in which tokens of the type *invariably* produce the characteristic effect. These would be just those circumstances of some one particular biologically-specifiable type that obtained on the all the propitious historical occasions, on which ancestor tokens actually accomplished the trick that caused them to be replicated into the next generation. Really, however, nothing in actual biological practice warrants the supposition that there are such infallibly supportive circumstances: a device might perfectly well get selected for even if, under the very best of circumstances, it performs only *often enough* the trick that gave it selectional advantage over whatever devices were coded for by the relevant alleles.

Evolutionary biology, in sum, rests on the idea that natural selection is shaped by histories of causation that obtain *often enough but not invariably*. Before we dismiss this idea, we should ask very seriously whether there might not be a rival account of causation—one on which true causing does not require exceptionless laws. And in fact there is one very plausible rival, put forth over recent years by Woodward and Hitchcock (Woodward and Hitchcock 2003a, 2003b; Woodward 2003; Hitchcock 2001a, 2001b). In the next section I will sketch this rival account of causation, and will show that it is compatible with the idea that beliefs and desires cause shapings of the host’s behavior. In the section after that I shall use the same account of causation to confront the specter of overdetermination. The argument there will allow that whenever an individual belief or desire gets tokened in a person’s brain, a vast plurality of microparticles in that person’s brain comes to instantiate some vastly complex microphysical property, but will show that, on the Woodward/Hitchcock account of causation, there is no clear content to the claim that the occurrence of this complex microphysical property causes *any* shaping of the host’s behavior. The section following will ask whether overdetermination gets entailed by a

different route—whether overdetermination follows from the efficacy of physical properties instantiated at a level above that of the microparticles, namely that of neurons and their parts.

III

The account of causation put forth by Woodward and Hitchcock takes its cue from what events actually get treated as causes in actual sciences. Sciences as diverse as economics, medicine, evolutionary biology, mechanics, and agronomy take themselves to have found a causal connection between two types of events just when they have identified an *invariance* between events of the two types. An invariance is a relation such that to variations on a putative cause there correspond, or there would correspond, variations on the putative effect. There is a variable that subsumes the putative cause—typically a multi-valued variable—and alterations on the value of that variable produce answering alterations in the variable that subsumes the putative effect. (“Produce” is of course a causal locution, but Woodward and Hitchcock do not purport to offer a reduction of causation to non-causal phenomena.) To put it metaphorically, by intervening on the cause variable one can manipulate the effect variable. But the manipulation need not always work: the function from values of the cause variable to values of the effect variable need not be exceptionless (Woodward and Hitchcock 2003a, Woodward 2003 pp. 65-70). In particular, the dependence of the effect variable on the cause variable may itself depend on other factors in the causal system, factors whose character is not fixed by the value taken by the cause variable: such other factors will be values of “exogenously determined” variables, and for some settings of these further variables, the dependence of the effect variable upon the cause variable may break down. So long as there is *some* range of settings of the exogenously determined variables such that interventions on the cause variable will reproducibly and reliably be accompanied by commensurate variations of the effect variable, there is a true case of causation.

Are beliefs and desires tied by invariances to particular shapings of behavior? I now argue that the answer may well be Yes—and that to see this, one must consider systems in which a fair number of

exogenously determined variables have just the right settings. My argument rests on two fairly plausible assumptions: that folk psychology is roughly right about the roles that beliefs and desires play in the brain's operations, and that beliefs and desires are physical states produced by naturally-selected programs or devices in the brain. The detailed ramifications of these twin assumptions are much explored in Ruth Millikan's work, and little explored elsewhere. But the basic picture is pretty much common coin—it is not specifically Millikanian.

The idea that beliefs and desires are the product of naturally-selected devices entails that beliefs and desires have their own proper functions (Millikan 1984, Ch. 2, and 1993, Ch. 1). The proper function of beliefs is different from the proper function of desires; so I will discuss these in turn. Desires—as folk psychology seems to suggest—have it as their proper function to target various goals for action. If the program that tokens desires was fashioned by natural selection, we can be sure that at least many of the goals that desires historically have targeted have been ones the attainment of which was useful or beneficial for the host. But there may also have been quite a few ill-formed desires, targeting goals the attainment of which would have been useless or outright harmful for their host. There have no doubt been many unrealistic desires—ones that targeted goals that their host could never be in a position to attain or even pursue—and many unactionable desires, desires such that, as things happened to work out, their host was never actually in a position to act on them. Even those desires that targeted useful goals which their hosts actually could and did pursue are not likely to have brought about right away the useful results of goal-attainment. Most desires are probably stored for a fair length of time before their host receives the information that here, now, there is a situation in which the targeted goal can really be attained. So it is not realistic to say that any individual desire *all by itself* brings about the attainment of some goal that is beneficial to its host. The right circumstances have to be in place, and have to be represented as being in place by the host, before this can happen. (Those are the “exogenously determined variables”, as I shall say presently.) But we can be sure that if the desire-fashioning program is the result of natural selection at all, desires do often enough come to be accompanied by inner

registrings that the right circumstances indeed are in place, and do then bring about actions that result in beneficial outcomes.

These inner registrings are beliefs. The proper function of beliefs is to represent particular states of affairs as obtaining in the world, states of affairs that may prove relevant to the enactment of those desires that the host comes to form. If the program that fashions beliefs is itself the product of natural selection, we can be sure that at least often, the states of affairs that historically have been represented *as* obtaining really *have* obtained—the beliefs have been *true*—and have sooner or later been crucial to the attainment of the goals that the host’s desires came to target. More fully, we can be sure that often, beliefs have attuned or adapted the host’s goal-pursuing behaviors to really-obtaining environing circumstances, in just such a way that those goal-pursuing behaviors succeeded. But here too we can assume that natural selection left much room for error and misfiring. The belief-fashioning program may have fashioned a fair number of false beliefs. It may also have fashioned irrelevant beliefs—ones that would never have been relevant to any desires that the host would, in the normal course of things, come to form—and many non-useful beliefs, beliefs that, as things happened to work out, represented states of affairs that never came to be relevant to enactment of the host’s actual desires. Even those beliefs that accurately represented states of affairs, to which the host at one point had to attune his goal-pursuing behaviors in order for those behaviors to succeed, are not likely to have brought about their useful attunement right away. Most beliefs are probably stored for a fair length of time before they can play a role in successfully aiming or adjusting the host’s pursuit of his goals. So it is not right to say that any belief can *all by itself* cause such attunement-to-circumstances of goal-directed behaviors that the host undertakes. The host has to have form desires to which the represented states of affairs are, in the circumstances which the host occupies at a particular time, relevant. (Again, these are the “exogenously determined variables”.) But we can be sure that if the program that fashions beliefs was indeed a product of natural selection at all, historical beliefs often enough were true, came to be relevant to actually-present desires, and actually attuned the pursuit of those desires in ways that intersected with the states of affairs that they represented so as to result in successful pursuit.

In this way, it seems, we can credit particular desires and particular beliefs with causing particular shapings of behavior. Even so, I have not yet squarely faced the question about invariances. Is it really true that—when the exogenously determined variables are set in just the right ways—variants on a given belief would have gotten accompanied by corresponding variants on particular attunements of behavior? Or that—again fixing the variables in the right way—variants on a particular desire would have engendered pursuit of a correspondingly variant goal? But the answer does seem to be Yes. We are supposing that natural selection somehow hit upon a device that could produce all manner of useful-goal-targeting states, even ones the actual enactment of which would depend upon the host's eventual acquisition of all manner of new beliefs. Similarly, we are supposing that natural selection hit upon a device that could produce all manner of beliefs, including ones that could attune and guide the host's enactment of all manner of desires that the host might eventually form, but did not yet have at the time the beliefs were produced. In order for the first device to be able to produce so rich an array of desires, and for the latter device to be able to produce so rich an array of beliefs, there would have to be some *system* by which the devices fashion new desires and new beliefs (Millikan 1993, pp. 80-82, 89-91, 98-99). That is, the particular states that are the host's goal-targeting states—its desires—would have to display a certain compositionality; so too would the states that are the host's potentially-action-attuning states, its beliefs. Put differently, the devices that produce desires and beliefs would have to have systematic ways of varying the desires they produce, and the beliefs they produce, so as to target variant goals and represent variant states of affairs. For it is unrealistic to suppose that the belief-producing device, for example, could have been rigged up to produce non-composite and unrelated brain states for each surrounding circumstance that it is the job of beliefs to represent; there are just too many such states; and so, *mutatis mutandis*, for the desire-producing device. In the case of either device, the only design that is simple enough for nature to have been able to hit upon it, while still being powerful enough to do the job that we are supposing it to do, would be one that allows systematic variations and similarities among the beliefs and desires that are produced.

Thus it is plausible to assume that, when the belief-producing device was functioning in the ways that won the favor of natural selection, there would have been variants on any actually produced belief that would have led to variant attunements of behavior—provided, that is, we set the exogenously-determined variables so that the variant beliefs would all have been relevant to one or another actually-present desire. It is plausible that, when the desire-producing device was functioning in the ways that won the favor of natural selection, systematic variants on a given desire would have led to actions that achieved variant goals—provided we set the exogenous variable for beliefs so that beliefs are present that could attune the variant actions in ways that lead to success.

IV

Suppose that whenever a particular belief gets tokened in a person's neural hardware, a shadow event occurs at the level of the microparticles in the person's brain. Billions of microparticles, between themselves, come to instantiate some massively complex structural property, describable only by some massive conjunctive predicate, each conjunct of which spotlights a property or interaction found in an individual microparticle or a small group of microparticles. The relation between the tokening of the belief, and the occurrence of this complex microparticle event, is not simply simultaneity and spatial coincidence, let us further suppose. Let us say that the tokening of the belief (call it *b*, for “belief”) supervenes, at least weakly, on the complex microparticle event (call it *c*, for “cause”—though the very point of the argument here will be that we have no grounds for thinking that *c* *really* causes *anything*). Then the question opens up whether *c* might perhaps overdetermine any effect that we credit *b* with causing. Suppose *b* does indeed eventually cause some attunement in the host's behavior: suppose that the state of affairs that *b* represents does at some point become relevant to the way the host can go about pursuing one or another of his desires. This attunement will have a description at the level of muscle contractions and nerve impulses, and perhaps also at the level of inferences and practical planning (and let us call the event so described *a*, for “attunement”). But that attunement, so described, will itself be

accompanied by another shadow event at the level of other microparticles. Let us further suppose that a supervenes on this other complex microparticle event (which we will call e , for “effect”). The question opened up is this. Does c cause e —and by causing e , does c thereby cause a ? If so, the shaping of the host's behavior which b brings about, namely a , will be causally overdetermined.

From the viewpoint of this paper, this question amounts to asking: is there a variable that subsumes c , and a variable that subsumes e , such that there is an invariance joining values of the one variable to values of the other? The contention that b causes a is thus far safe, at least until overdetermination rears its ugly head and gets transformed into causal exclusion. For as we noted in the previous section, we should expect that the design of the belief-producing device ensures that variants on b would—when all goes according to design—be accompanied by variants on a . Is there an invariance that links a similarly multi-valued variable that subsumes c to a multi-valued variable that subsumes e ? Is there even an invariance linking a *single*-valued variable subsuming c to a single-valued variable subsuming e ? In this section I argue that there is no clear case for answering Yes to either question. Before I begin the argument, let me point out that it will be crucial for us to consider c *strictly* as a complex event involving billions of microparticles, and not also as an event upon which b supervenes—and likewise to consider e strictly as a complex microparticle event, regardless of whether it subvenes a . For the suggestion that we will be trying out is that c causes e in a way that simply *bypasses* the apparent causation that runs from b to a . Only if this suggestion is vindicated can it emerge that c causes a in a way that renders b 's apparent efficacy redundant—a way that therefore *excludes* b 's apparent causal efficacy. Maintaining this focus is harder than it may appear to be, as I shall point out in the last paragraph of the *following* section, section V.

There is an invariance between two multi-valued variables, to use the metaphor favored by Woodward and Hitchcock, just in case intervening on the value of the putative cause variable would produce answering variations in the value of the putative effect variable. The precise function that takes values of the cause variable over into values of the effect variable can vary. It can be continuous, so that every small alteration in the cause variable produces an equally small alteration in the effect variable. It

can be discontinuous, so that before any alteration occurs in the effect variable one must run through a number of small variations in the cause variable (Woodward 2003, p. 66). It might even be non-linear. But this much seems clear: if there is to be any function at all, by which alterations in the effect variable are tied to alterations in the cause variable, there must be a fact of the matter as to which variations in the cause variable are smaller and which are greater. More precisely, there must be an ordering of possible alterations in the cause variable, along which greater and greater alterations make for values of the effect variable that are more and more different from the value with which we began.

The event *c* which, we are now suspecting, may causally overdetermine the attunement-to-circumstance (viz., *a*) of the host's behavior consists in the instantiation of a massively complex structural property at the level of microparticles. Can we really think of a variable that will subsume *c*, such that different values of this variable will amount to variants on *c* that line up in an ordering of greater and greater difference from *c* itself? Let us first address a prior question: do we even understand what it would be for complex microparticle events that are variants on *c* to differ from *c* to differing, but commensurable, degrees—to differ from *c* itself *more* and *less*? What makes this question seem formidable is that the property that *c* instantiates is so complex, being specifiable only by a conjunctive predicate that involves thousands and thousands of conjuncts. It seems to follow that the complex microparticle properties, which *variants* on *c* would instantiate, could differ from the property which *c* instantiates along thousands and thousands of different dimensions of difference. Now there can indeed be properties that are variants on one another, and that differ from one another on several different dimensions, for which there is nevertheless a clear-cut phenomenon of “more and less different”. Colors differ from one another in saturation, hue, and brightness, and yet the difference between any one color and another is comparable to—commensurable with—the difference between the first color and a second other (at least provided we are willing to regard the differences as distances within the Munsell color solid). But that is because the different dimensions on which colors differ from one another are mutually independent, and the relation between them is merely additive.

In contrast, when one looks to the dimensions on which c differs from alternative complex microparticle events, one finds that the dimensions are not independent of one another. A particular distance on one dimension of difference, for two such complex microparticle events, can amount to a massive difference or a trivial difference between the overall events themselves, depending on the distance between the two events on *other* dimensions of difference. This claim will be illustrated in a moment. The upshot is that while any one such complex microparticle event may lie at different distances from each of two others, on any dimension that appears to permit a comparison of degrees of difference from the first event, those different distances never in and of themselves amount to true cases of “more and less different” relating the overall events concerned. Indeed those different distances do not even make stable and fixed *contributions* to a phenomenon of greater and lesser difference. So while there are countless (literally!) variants on c , which differ from c itself in countless different ways, there is in general no phenomenon of “more and less” that orders those differences. Our intuitions may, it is true, seem to suggest that we can think of *some* variants on c that differ from c itself far less than many, many other variants. We will return to this point shortly, and will see why we perhaps should not trust these intuitions. But even if we give evidential weight to these intuitions, and rule that there are *some* definite cases of “_____ is less different from c than _____”, among complex microparticle events that are variants on c , it will remain unbelievable that there is any *general* ordering of relative differences from c , such that any two *randomly-selected* variants on c will occupy determinate positions in that ordering.

Consider, then, eight possible variants on c .

Event $c1$: it instantiates a structural property, the description of which differs from the description of c 's structural property with respect to 1000 conjuncts belonging in the latter description.

Event $c2$: instantiates a structural property which differs with respect to 5000 conjuncts.

Event $c3$: with respect to every conjunct where its description differs from c 's description, a microparticle bears a markedly different charge or energy-state from that of the microparticle described in the corresponding conjunct of c 's description.

Event *c4*: with respect to every conjunct where its description differs from *c*'s description, a microparticle bears only a slightly different charge from that of the microparticle described in the corresponding conjunct of *c*'s description.

Event *c5*: its description requires only 1100 conjuncts—that is, *c5* is a *less complex* microparticle event than *c*—but differs at 1000 of those conjuncts from *c*'s description.

Event *c6*: its description is the same, conjunct-for-conjunct, as *c*'s, and requires no more conjuncts than *c*'s description does, but its conjuncts are satisfied at regions far more spread out from one another, spatially, than are *c*'s conjuncts.

Event *c7*: its description incorporates every conjunct in *c*'s description, but requires 18000 more conjuncts as well—*c7* is a far *more complex* microparticle event than *c*.

Event *c8*: thirty-six conjuncts within its description depict the microparticles which they describe as lying within a volume which common sense would suppose to be occupied by a dog.

Event *c9*: three hundred conjuncts depict the microparticles which they describe as lying within a volume which common sense would suppose to be occupied by a pond.

Event *c1* seems, at first blush, to differ from *c* itself less than *c2* does. But if *c1* also differs from *c* in the way that *c5* differs from *c*, *c1* may be more different from *c* than *c2* is. This will especially be true if *c1* further differs from *c* in the way that *c3* does. But if *c2* likewise differs from *c* in the way that *c3* does, *c1* may after all be less different from *c* than *c2* is. Of course, that will all depend on whether *c2* further differs from *c* in the way that *c6* does. And what about *c7*—is it more different from *c* than *c2* is, if *c2* itself differs from *c* in the way that *c3* does? Does the difference between *c8* and *c9* amount to different degrees of difference from *c*? The point that these questions illustrate is that there is often systematic interference, and sometimes systematic irrelevance, among the various dimensions along which variants on *c* can differ from *c* itself. The dimensions do not relate additively to one another. The consequence is that relative distances from *c*, along any dimension that affords an ordering of “more and less different” for complex microparticle events that are variants on *c*, will not in general *amount*, in and of themselves, to degrees of difference for those complex microparticle events themselves from *c* itself.

But then in general, for *pairs* of variants on *c* that differ *differently* from *c* on any one dimension of difference—that can, in respect of that dimension, be said to be “more and less” different from *c*

itself—there will simply be no fact of the matter as to which member of that pair is *as a whole* more different from *c*, and which less. But then even if we could think of a multi-valued variable that subsumes *c* and all its complex microphysical variants, there would be no prospect of claiming that values of that variable that are *more and more different* from the value found at *c* bring about *corresponding* variants on *a* that are more and more different from *a* itself. In the relevant sense, *there is no such phenomenon* as “more and more” different from *c*—no neat ordering of the microphysical variants on *c* in terms of relative difference from *c* itself. So there is no prospect of an invariance that links a multi-valued variable subsuming *c* to a multi-valued variable subsuming *a*—or even to a multi-valued variable subsuming *e*.

But if it has no clear sense to ask how matters would have gone if complex micro-events *more and more different* from *c* had occurred in *c*'s place, it still might make perfectly good sense to ask how matters would have gone if *c* had not occurred at all. Philosophers inclined to think that microparticles found where *b* occurs between them cause *e*, and thus overdetermine *a*, might therefore suppose that there is an invariance between a *two-valued* variable *C* that subsumes *c*—the values of *C* being {*c* occurs, *c* does not occur}—and a two-valued variable *E* (namely {*e* occurs, *e* does not occur}) that subsumes *e*. So let us ask: what really *would* have occurred, if *c* had not occurred? Most philosophers will agree that to ask this question is to ask about the *closest possible worlds* to the world in which *c* does occur, i.e. closest to the actual world. These will be worlds in which some complex microparticle event very much like *c*, and differing from *c* just slightly, occurs in place of *c* itself. Our intuitions may, as we noted several paragraphs above, suggest that we can fairly readily identify at least some such variants on *c*. The recipe we will follow, if we heed these intuitions, will be to envision a variant event that is “just like *c* on every dimension but one”, and that involves some minimal divergence from *c* on the remaining dimension. Thus we might suppose that a variant that is “otherwise just like *c*” except that one of the microparticles, spotlighted by one of the conjuncts in the specification of *c*, is placed at a slightly greater spatial remove from its nearby neighbors than *c* depicts it as being. But this overall variant event may be *less* like *c*, in respect of its causal powers, than a variant that differs from *c* on *two* dimensions—a variant

that features stronger charges or higher energy states obtaining in that spatially-removed microparticle and in its neighbors. In general, as we have seen, there are countless ways in which a small divergence from c along one dimension of difference can be made to amount to no overall difference from c , or to sizeable overall difference from c , by divergences (or samenesses) that obtain along other dimensions. We must be very hesitant, therefore, to claim that we can identify *any* variants on c that differ “only minimally” from c itself. But even if there objectively are such variants, we must remind ourselves that if there are any, there are many. For possible variants on c diverge from c on thousands and thousands of different dimensions of difference. Even if we trust our intuitions to have identified three or four candidates for the title “minimally different variant on c ”, there will be thousands and thousands of candidates that we have not thought of. And then the crucial question is this: if we take c together with these thousands and thousands of other variant events, is it plausible that c alone was causally *sufficient* to produce e —in other words, that c itself was not just causally sufficient for e , but causally necessary? If c had not occurred, are we really sure that e would not have gotten produced anyway? It seems out of the question that we should really have warrant for answering Yes to these questions. But then we have no warrant for believing in a variance linking even single-valued variables that subsume c and e . We have no reason for thinking, therefore, that c causes e . Consequently, we have no reason for thinking that c causes a . The claim of b to have caused a remains unrivalled.

A note on the argument of this section. One might worry that the argument “proves too much”. For if the phenomenon of “more and less different from” is ill-defined for the variables that might subsume the microparticles involved in c , it is ill-defined in many other cases as well. The complex microphysical event c attracted our attention because, we supposed, it subvented the occurrence of mental state b . But similarly complex microphysical events appear to subvene all manner of macro-level states involving macroscopic objects: such events appear to subvene a baseball’s flying towards a window, for example, or a car horn’s blowing loudly. Thus the argument of this section may seem, if sound, to show that there is no real worry about causal exclusion for even such events as these, involving *non-conscious* macroscopic objects. To this misgiving I plead guilty as charged. I think the argument of this section is

sound, and *does* show that microparticles are in general powerless to rival, or steal away, the putative causal efficacy of familiar macro-objects. I am nevertheless content to focus, in this paper, on threats to the apparent causal efficacy of *mental* states—in particular, of beliefs and desires. For in the particular case of beliefs and desires, as I have argued, we have reason, derived from plausible surmises about the course of natural selection, for supposing that they are tied by robust invariances to shapings of behavior. Mental causation is a case in which macro-causation appears to be genuine, precisely while microparticle causation appears to be spurious.

Yet one further note, on the relation of this section to the constitution view. That view holds that the aggregate of microparticles found within a person constitutes an organism; that the organism constitutes the person himself; and that different patterns of causation characterize each of these three co-located objects. The argument of this section concludes that we have no warrant for supposing that the aggregate of microparticles, as such, causes anything. If Alexander’s Dictum is correct—if “to be is to have causal powers”—then we have no warrant for supposing that there even *exists* any such thing as the aggregate of microparticles found within the person. That may seem to violate the principle of Universal Mereological Composition. Again, I plead guilty as charged. That principle is false (Elder 2008).

V

Whenever a particular belief or a particular desire gets tokened in a person’s neural hardware, an event occurs at the level of the neurons in the person’s brain. Certain axons and dendrites, say, pass into certain new states. Therein lies a remaining threat to the vindication offered by this paper of mental causation.

This paper has assumed that beliefs and desires are both brain states and biological products. That is, it has assumed that individual beliefs and desires are brain states tokened by devices fashioned by natural selection. It follows that beliefs and desires have proper functions, which often enough they actually perform. Desires have it as their proper function to cause behaviors which—when steered by

beliefs in ways assumed by the design of the desire-forming device—result in the attainment of goals useful to the host organism. Beliefs have it as their proper function to attune the host's enactment of one or another desire, in such a way that some environing state of affairs gets enlisted in making the host's behavior, thus attuned, turn out to be successful. Beliefs and desires, on this way of looking at them, are like other devices and traits that evolutionary biology studies. They are essentially the products of a certain history of natural selection. More precisely, they are essentially the products *of* products—the belief-producing and desire-producing devices themselves—of a certain history of natural selection.

But precisely where beliefs and desires, so conceived, occur in the brain, there also occur brain states that are *only* brain states, and not *also* products of natural selection. These are the neurochemical excitation-states of particular axons and dendrites, and the distribution and movement of neurotransmitter molecules within those axons and dendrites, upon which those individual beliefs and desires supervene—the neurochemical states and events that *realize* or *compose into* beliefs and desires. These states and events are *not* essentially the products (or the products of products) of histories of natural selection. Their entire natures can be specified without any mention of history. If *they* are sufficient to cause whatever shapings of behavior we might want to credit beliefs and desires with causing, then we still face a worry that those shapings of behavior are causally overdetermined. Indeed we might then worry that the causal credentials of these neurochemical states and events *exclude* any claim of causal efficacy that we might want to make on behalf of beliefs and desires.

In other words, beliefs and desires on the one hand, and the neurochemical states and events upon which they supervene on the other, seem to lodge competing claims to causal efficacy—to shaping the host's behaviors. The competitors are both states (and events) that occur in the brain—*brain* states and events. But they are brain states (and events) discerned from different points of view. One point of view, we could say, is that of the *neuroanatomist*. Here one pictures the realm of the neurons as populated by structures or mechanisms, both simple and complex, that were fashioned by natural selection. One then sees individual neurons, or groupings of neurons, as having designed ways of functioning. One's judgements as to how long “the same” neuron-level event is going on, or over how wide an area it occurs,

will be structured by one's understanding of what the stages are in these designed ways of functioning. From this point of view, the events that occur at the neuron-level will be tokens of historically-replicated types. The other point of view is that of the *neurochemist*. Here one focuses on molecules, and on interactions among molecules, that are salient strictly because of the biochemical properties that they now possess and produce. One's judgements as to how long "the same" neuron-level event is going on, or over how wide an area it occurs, will not trade (even tacitly) on one's suppositions about how natural selection has structured the brain: that is, even if it is true that neurochemically different configurations, occurring at different times and places in the brain, all realize a common stage in some designed way of functioning, one will nevertheless rule that these are *different* neuron-level events. History will be irrelevant to one's taxonomy of events. One will not assert that neuron-level events have proper functions.

Here is an illustration of the contrast between these points of view. Linguists and evolutionary biologists commonly surmise that there resides within the brain something that can properly be called "the language organ". This is a surmise which the neuroanatomist might either confirm or disconfirm. But even if the surmise is true, "the language organ" is likely to appear, from the standpoint of the neurochemist, as a disparate plurality of separate neurons, neuron-parts, and neurotransmitter molecules. Indeed it may appear as a *shifting* disparate plurality, one that incorporates numerically different neuronal and sub-neuronal elements on different episodes of its operations.

Just as with language, so too—we must be prepared to discover—with beliefs and desires. From the standpoint of the neuroanatomist, it may well turn out that there resides within the brain something that can properly be called "the belief-tokening region" (or "the desire-tokening region"). Even if so, we must be prepared to learn that from the standpoint of the neurochemist, this unitary "region" is a hodgepodge of disparate and spatially separated neurons, neuron parts, and neurotransmitter molecules. Indeed it may be a *shifting* hodgepodge, incorporating numerically different parts on different occasions. From the standpoint of the neuroanatomist, the tokening of an individual belief in this region will count as a single (if not simple) event. But from the standpoint of the neurochemist, we must be prepared to learn,

the tokening of a belief will count as a plurality of independent and diverse events, which may not even occur at the same time as one another, and which will assuredly occur in different subjects of change—in different, indeed scattered and diverse, neuron-parts and neurotransmitter molecules and neurons.

For this paper, the remaining worry about overdetermination is this. If a belief that is present in the brain—present, as we may suppose, in “the belief-tokening region”—comes to cause some attunement in the host’s behavior, can it equally be said that what the neurochemist sees in that “region” causes that same attunement? What the neurochemist sees, just where some individual belief is realized in the brain, is a disparate plurality of scattered, perhaps non-simultaneous, small neurochemical states and events involving separate neurons, neuron-parts, and neurotransmitter molecules—or so we are supposing may well be the case. Our remaining worry is therefore this: can such a plurality of scattered small events be credited with causing a single outcome—namely the same behavioral attunement as we suppose the belief to cause?

To address this question, we must face a more general one. This is the question of when, whether, and why it may be legitimate to think of many events as collectively causing a single outcome. Many philosophers speak as if this question has an obvious answer. They speak as if it is obvious that if each element of such a “many” causes a single small outcome, it automatically follows that the “many” itself causes the collection of those small outcomes. But on the face of it, this simple inference seems merely to be a case of “the fallacy of composition” (Elder 2000). In any case, the idea that any true causing must instantiate some explanatory invariance entails that this simple inference is too quick. That is what the argument of the preceding section in effect showed. It showed that to defend the claim that many small events, each involving a separate small object, between them cause some single outcome, we must defend the claim that the instantiation of a single complex structural property, by that plurality of small objects, is linked by an explanatory invariance to that single large outcome (or to something single and large that upon which that outcome supervenes).

Just so here. The threat of overdetermination collapses unless the neurochemist can argue that what happens from *his* perspective, when (as we are supposing) an individual belief causes an attunement

of behavior, falls under an explanatory invariance. The neurochemist must argue that the many neurons, neuron-parts, and neurotransmitter molecules between them instantiate some complex structural property, such that the right sort of invariance holds—such that variants on that structural property, more and more different from the actually-instantiated complex property, would be accompanied by behavioral attunements more and more different from the attunement actually produced (or would be accompanied by large-scale events on which such variant attunements would supervene).

But again we must ask, as we did in the previous section: is the phenomenon of “more and more different from” well defined, in this context? If we view the neurotransmitter molecules, neuron-parts, and neurons on which the putatively-efficacious belief supervenes in the abstract—putting out of our mind the thought *that* they subvene this belief—the contrasting structural properties, which would qualify as variants on the structural property actually instantiated, seem to contrast along many different dimensions. The structural property actually instantiated, let us say, is the one specified by this predicate: “incorporates a neurotransmitter of type a at location l_1 moving across a synapse at location l_2 ; and incorporates an axon of type b and location l_4 spiking at such-and-such a frequency; and incorporates a pyramidal cell at location l_5 rigged to attune its outputs to six other pyramidal cells, at locations....”, and so on, through 23 conjuncts. And now let us consider some variants on this structural property, with which it would contrast. One incorporates 22 conjuncts that are the same as conjuncts in the actually-instantiated property, and 70 more conjuncts, of which this is a typical example: “...and incorporates a third baseman who is batting .267”. Another incorporates the same conjuncts as the first 18 in the actually-instantiated property, but no others. Another incorporates pretty much the same conjuncts as the original 23, and no others, but specifies that each conjunct is satisfied at a location 500 miles or greater from the location at which any other conjunct is satisfied. My question: which of these variant structural properties is more like the structural property actually instantiated, and which less?

“But this is unfair!”, I expect the neurochemist to reply. “These are ridiculous, made-up structural properties—they aren’t really relevant alternatives to the structural property that is instantiated in the brain, from my point of view, where the actual belief is present!”. Precisely. But ask *why* these are

not relevant alternatives. The answer must be: “these aren’t alternative structural properties which the many neurons, neurons-parts, and neurotransmitter molecules might between them instantiate, while that same brain is still doing its usual business!”. But the only explication of “doing its usual business” that will help the neurochemist here is this: “while still composing into beliefs that are variants on the one that actually causes the actual attunement of behavior”. That is, the neurochemist has to impose some discipline on which dimensions of difference, from the actually-instantiated structural property, will count as relevant. And the only way to impose this discipline is to keep constant the assumption that some belief or other is present, that is efficacious in causing the actual behavioral attunement or some variant on it.

In sum: the neurochemist *may*, in a way, be able to assemble a competing cause of the behavioral outcome, at the level of the small objects he recognizes, that will be sufficient to cause the behavioral attunement—but it will qualify *as* sufficient to cause this behavioral outcome only by composing into the belief which, we are supposing, causes that behavioral attunement. If this is causal overdetermination, it is entirely non-worrying causal overdetermination. (For a closely similar form of non-worrying causal overdetermination, see claim “(5)” in Merricks 2001, pp. 100-101.)

A parallel point applies to the argument of the previous section, the section that considered *microphysical* overdetermination of a given shaping of behavior. That argument asked whether a genuine invariance might link *c*, the enormously complex event involving billions of microparticles, with an equally complex microphysical event *e* that subvened the given shaping of behavior. That argument contended that variants on *c* diverge from *c* on so many different, sprawling dimensions, that the very phenomenon of “more and less different from *c*” was undefined. But here friends of the microparticles might be tempted to make much the same *sort* of protest as I have attributed to friends of the neurochemist’s standpoint. “These so-called variants on *c*,” friends of the microparticles might protest, “feature ridiculous, made-up structural properties! Surely the *relevant* alternatives to *c* are ones that involve just as many microparticles as *c* itself does, not more or fewer, and involve *only* microparticles, and only microparticles that are just as close to one another in space and time as are the microparticles in

c itself!” But ask *why* these might be the only relevant alternatives—just what the rationale would be for confining divergence from *c* to just these dimensions. The answer will have to be: “because only so will we be looking at microphysical events, involving billions of microparticles, that might subvene the occurrence of a mental event!” (cf. Sturgeon 1998, §6). But then we will have been treating *c* as essentially a subvener of *b*. Yet the whole point was to ask whether *c* might cause *e* (and thereby *a*) in a way that simply *bypasses* the apparent causation that runs from *b* to *a*—in a way that has nothing to do with *c*’s subvening *b*. If we save the claim that an invariance really does tie *c* to *e*, but does so *only insofar as c* subvenes the mental event *b*, we will not have attributed to *c* a causal efficacy that *bypasses b*’s claim to causal efficacy.

VI

Now a brief comment on “the problem of mental quausation”.

When individual beliefs perform in the way they did perform, on the occasions when the device that produces them was winning the favor of natural selection, they attune the host’s pursuit of its desires to the presence in the world of particular states of affairs. That is, an individual belief—when operating in the way that got the belief-making device to be selected for—initiates, steers, and modulates the host’s enactment of one desire or another, in such a way that some environing state of affairs gets enlisted in making the host’s behavior, thus steered, turn out to be successful. More briefly: the individual belief causes the host’s goal-pursuing behavior to be caused to be successful by some particular environing state of affairs.

Now what Ruth Millikan has argued is that this two-step causal connection to an environing state of affairs is *what it is* for the individual belief to have that state of affairs as its semantic value, its content (1993, Ch’s 3 and 4). If that is correct, a belief’s having the content that it does is a function of its being such as to cause—when all proceeds according to design—the behavioral outcome that it does cause.

Having content presupposes efficacy in shaping behavior. It is therefore incoherent to ask whether or not it is *in virtue of* having its content, that an individual belief has efficacy in shaping behavior.

When individual desires operate in the way they did perform, on the occasions when the device that produces them was winning the favor of natural selection, they target the host's behaviors towards goals which, thanks to steering of those behaviors by the host's beliefs, actually get attained. That is, an individual desire—when operating in the way that got the desire-producing device to be selected for—causes the attainment of some goal. Millikan (1993, Ch's 3 and 4) has argued that this belief-steered causing of a particular outcome is *what it is* for the desire to have that outcome as its content—to be a desire *that* that outcome obtain. If this is correct, then “the problem of mental causation” is as confused in the case of desires as it is in the case of beliefs. A desire's having the content it does *is* its being such as to cause—when all proceeds according to design—a particular outcome. Content presupposes causal efficacy in shaping behavior. One cannot coherently ask whether it *underlies* or *yields* causal efficacy in shaping behavior.

References

- Baker, Lynne Rudder 2007: *The Metaphysics of Everyday Life*. Cambridge: Cambridge Univ. Press.
- Davidson, Donald 1970: "Mental Events". In Foster, Lawrence and Swanson, J. W., eds., *Experience and Theory*. Amherst: Univ. of Massachusetts Press.
- Elder, Crawford 2008: "Against Universal Mereological Composition". *dialectica*, 62, pp. 433-54.
- _____ 2000: "Physicalism and the Fallacy of Composition". *Philosophical Quarterly*, 50, pp. 332-43.
- Fodor, Jerry 1989: "Making Mind Matter More". *Philosophical Topics*, 17, pp. 59-79.
- Hitchcock, Christopher 2001a: "The Intransitivity of Causation Revealed in Equations and Graphs". *Journal of Philosophy*, 98, pp. 273-99.
- _____ 2001b: "A Tale of Two Effects". *Philosophical Review*, 110, pp. 361-96.
- Horgan, Terence 1989: "Mental Quausation". *Philosophical Perspectives*, vol. 3 (*Philosophy of Mind and Action Theory*), pp. 47-76.
- Merricks, Trenton 2001: *Objects and Persons*. Oxford: Oxford Univ. Press.
- McLaughlin, Brian 1993: "On Davidson's Response to the Charge of Epiphenomenalism". In Heil, John, and Mele, Alfred, eds., *Mental Causation*. Oxford: Oxford Univ. Press.
- Millikan, Ruth 1984: *Language, Thought, and Other Biological Categories*. Cambridge, Massachusetts: M.I.T. Press.
- _____ 1993: *White Queen Psychology and Other Essays for Alice*. Cambridge, Massachusetts: M.I.T. Press.
- _____ 2000: *On Clear and Confused Ideas*. Cambridge: Cambridge Univ. Press.
- Sturgeon, Scott 1998: "Physicalism and Overdetermination". *Mind*, 107, pp. 411-32.
- Woodward, James 2003: *Making Things Happen*. Oxford: Oxford Univ. Press.
- Woodward, James and Hitchcock, Christopher 2003a: "Explanatory Generalizations, Part I: A Counterfactual Account". *Noûs*, 37, pp. 1-24.

Woodward, James and Hitchcock, Christopher 2003b: “Explanatory Generalizations, Part II: Plumbing Explanatory Depth”. *Noûs*, 37, pp. 181-99.