University of Chicago Interlibrary Loan
OCLC: CGU / RLG:ILCG / DOCLINE: ILUJCL

ILLiad TN: 1880302

Borrower: NUI

Lending String:
*CGU,IAY,MHY,AAA,COA,ITD,NXW,UUE,FHM,MYG,OHS,COO
In Process: 20151007

Patron:
Reference #: 4566902

Journal Title: Consciousness & emotion.

Volume: 4
Issue: 1
Month/Year: 2003
Pages: 43–61

Article Author: Carrie Figdor

Article Title: Can mental representations be triggering causes?

Imprint: Amsterdam ; Philadelphia : John Benjamins Pub. Co., c2000-c2003.

Notes:

ILL Number: 155261163

Call Number: 10/7/2015 11:47:10 AM
Need By:
Not Wanted After: 11/06/2015
Received Via:

Call #: BF309.C667 c.1 v.3-4 2002-03

Location: ASR / JRLASR

OCLC
CIC
Charge
Maxcost: 35IFM
Category: Exempt
Billing Notes:

Shipping Address:
Interlibrary Loan
University of Iowa Libraries
125 W Washington St
Iowa City, Iowa 52242
United States

Fax: 319-335-5830
Ariel:
Email: lib-ill@uiowa.edu

ODYSSEY REQUEST

☐ **SCAN for PLATES** (Pattern 2)

Email / Ariel
Mail

SENT_____

Please report all Ariel transmission problems within 48 hours of receipt

Can mental representations be triggering causes?

Carrie Figdor
Rutgers University Newark

Fred Dretske's (1988) account of the causal role of intentional mental states was widely criticized for missing the target: he explained why a type of intentional state causes the type of bodily motion it does rather than some other type, when what we wanted was an account of how the intentional properties of these states play a causal role in each singular causal relation with a token bodily motion. I argue that the non-reductive metaphysics that Dretske defends for his account of behavior can be extended to the case of intentional states, and that this extension provides a way to show how intentional properties can play the causal role that we wanted explained.

Keywords: behavior, content, intentional state, mental representation, non-reductive process

Introduction

Remember Clyde? He's the beer-swilling couch potato in Fred Dretske's (1988) account of how the contents of our thoughts explain our behavior. Clyde went to the fridge to get a beer because he wanted a cold beer and he believed that there was cold beer in the fridge. Clyde's adventure was the subject of an extensive analysis by Dretske of how the fact that these mental states were *about beer* could make a causal difference to his behavior, given that a complete account in physical terms could be given of how he heaved himself off the couch and shuffled into the kitchen. But many philosophers found Dretske's account untenable. One of the main criticisms was that Dretske missed the target; he explained something that wasn't what we really want to know.[1] What we wanted was an account of how the contents of his thoughts caused him to act, and what we got was an account of why his contents caused that action and not some other.

Consciousness & Emotion 4:1 (2003), 43–61.
ISSN 1566–5836 / E-ISSN 1569–9706

I think, however, that Dretske's programme deserves another close look. I believe that an extension of his account provides us with a plausible framework for explaining how beliefs, desires and other intentional mental states can be causally implicated in producing behavior in virtue of their intentional content. That framework is not his own account of intentional causal efficacy, as his critics are correct to argue; rather, it is one that can be derived from his account of behavior.

In this article I will sketch this alternative solution to the problem of intentional causal efficacy, identifying but leaving unbridged an important gap in the account. The resulting framework is not Dretske's, nor is it necessarily anything that he would endorse; but it is a natural outgrowth of an analysis he has already given us, and it may deserve further attention in its own right. I will proceed as follows. First I will describe the metaphysics behind the view of behavior that is dominant in the philosophy of mind and show how it motivates Dretske to provide an alternative. I will then provide a reading of Dretske's alternative view of behavior, and suggest how the metaphysics he suggests for behavior can be extended to the case of mental representations. With this extension in place, I suggest, we possess a framework for addressing Dretske's critics and explaining exactly what we wanted to know: how the fact that Clyde's thoughts were about beer caused him to get off the couch in virtue of their intentional content.

Action, bodily motion, and "behavior": The account that Dretske rejects

When Clyde got off the couch, he performed what we usually call an intentional action; no one forced him to get up, and his getting up was not merely a series of bodily motions. Clyde may be a couch potato, but he isn't a puppet. These bland observations hide a tangle of philosophical issues, but the one of interest here is to understand the relation between intentional action and bodily motion — between Clyde's getting off the couch and his body moving around just so.

Action theorists spend much of their careers trying to explain this relation in the course of debating what an action is. As Mele (1991) notes, philosophers have defended many different models of action. An action might be a bodily event whose cause was a mental event; a bodily event of this sort in combination with certain of its products; a composite of a mental cause and bodily event; a composite of this sort plus certain of its products; a mental event alone; or none of the above. The five specific proposals just cited involve several distinct relations between an action and a bodily motion. For example, in the first, the relation is identity: the action is a bodily motion, albeit one with a

particular kind of cause in virtue of which it sustains an action description. In the third, the relation is part-whole: an action is partly composed of a bodily motion. In the fifth, the relation is contingent and causal: an action is a trying or a willing, and a bodily motion may perhaps be caused by this action. The relations in the second and fourth may perhaps be assimilated to the first and third respectively, but these proposals and the fifth will not matter in this context.

Given the interest of philosophers of mind in psychological explanations of behavior, and of explanations of intentional action in particular, we might expect that each of these proposals about action would be closely scrutinized. After all, if each is a different view of what action is, then each is a different view of the explanandum of interest. What we find instead is a longstanding, lackadaisical, even negligent, approach to the whole issue. It is endemic in the philosophy of mind literature to find discussions of psychological or intentional explanation that introduce the problem using standard examples of intentional action, substitute the term "behavior" for "action" and proceed to discuss the etiology of "behavior" as if the first view of action, which identifies it with bodily motion, were uncontroversial. For example, Block (1986) defends the role of narrow content in explaining behavior this way:

> [T]o ascribe a narrow meaning is to ascribe a syndrome of causes and effects, including, in some cases, behavioral effects (or at least impulses in motor-output neurons). The reason my twin and I both jump [into our respective swimming pools] is that we have representations with conceptual roles that have, as part of their syndrome of effects, jumping behavior. The reason that wide meaning is not as relevant to the explanation of behavior as is narrow meaning is that differences in wide meaning that do not involve differences in narrow meaning ... do not cause behavioral differences.[2]

"Jumping into a swimming pool" begins, at least for rhetorical purposes, as an ordinary intentional action that our friend Clyde would never contemplate doing, and morphs into motor-neuron activity. The term "behavior" — or hybrids like "behavioral output" and "bodily action" — blur any possible distinction between action and bodily motion.[3]

One reason for this neglect is the temptation from a computational psychological perspective to identify action with bodily motion under the terminological guise of "output". To do so seems to be a harmless extension of the popular computer model of the mind. But the fact that an analogy can be extended is no argument that the extension is legitimate. The input-output model itself has come under attack in recent years, and the identity of action with output is surely not something that can be taken for granted.[4]

The more important reason, however, seems to be the background assumption of a metaphysics of events exemplified by Davidson (1970).[5] I do not mean Davidson's analysis of the logical form of action sentences or his criteria for individuating events. I mean the extension of his version of non-reductive materialism for mental events to the case of intentional actions. His view of mental events combines token-token event identity with the denial of type-type identity, such that the tokens to which mental terms refer (or which mental properties classify) are events that satisfy various descriptions, mental and physical. Since actions are (on his view) events, the extension of this metaphysics to actions is natural: the relation between action and bodily motion is one of token-token identity between individual events (generically, "behavior") that can be alternately described as actions or bodily motions. As noted above, this is roughly the first of the proposals cited above about what action is. If the term "behavior" can be legitimately substituted for "action" — that is, if the term is *neutral* in its reference to either bodily motion or action, rather than *ambiguous* — it is because this metaphysics is presupposed.

This assumption of a "Davidsonian" metaphysics of action (as I will call it) is not entirely neutral with regard to the question of intentional causal efficacy, as Dretske was entirely aware. Since all claims of causal efficacy are relative to a given effect, *any* property can seem causally inert if an inappropriate effect is assumed. To borrow an example from Block (1990), although thermal and electrical conductivity nomologically covary, the electrical conductivity of a heated metal rod that is rigged to a bomb is causally inefficacious — *relative to* the bomb's exploding. But if we rig the same rod into an electrical circuit with a light bulb, the thermal conductivity is causally inert — *relative to* the light bulb's turning on.[6] Similarly, the meaning of the soprano's song, sung at high C, won't explain why the glass breaks, but the song's being sung at high C won't explain why the tenor staggers off the stage in mock despair.

In particular, a metaphysics that denies that actions are ontologically distinct from bodily motions imposes a significant handicap on the defender of intentional causal efficacy. For given the assumption that the effect, if any, of mental phenomena is an event which may be described as action but which at the token level is a bodily motion, and given that we can provide a complete physical explanation of every token bodily motion, then a causal explanation of action that involves mental causes can easily seem redundant or not genuinely causal.[7] The "Davidsonian" ontology sketched above does not determine the outcome of this debate, but it does prejudice our intuitions about causal relevance. Dretske's alternative metaphysics is motivated precisely by this difficulty.

Dretske's alternative view of "behavior"

Dretske discusses the relation between action and bodily motion in the first one-third of his book *Explaining Behavior*. Only afterwards does he defend a view of how contents causally explain action. He justifies this way of proceeding for the reason just mentioned: if behavior — of which intentional action is considered a species — is confused (that is, identified) with bodily motion or output, then it is easy to see how mental phenomena can fail to be causally explanatory. In his words,

> [I]f one uses this model [in which behavior = output] to think about human and animal behavior, great harm can be done. One can easily be misled into thinking that the cause of behavior is necessarily the cause of output. And once this confusion is in place, one will have no choice but to identify causal explanations of why we *do* the things we do with causal explanations of why our body *moves* the way it does. One will, in other words, have succeeded in confusing psychological explanations of behavior with neurobiological explanations of motor activity. Reasons — our *thinking* this and *wanting* that — will have been robbed of an explanatory job to do.[8]

The "Davidsonian" metaphysics, in which an action-event is token identical to a bodily movement, is an example of this conflation. Dretske's account of behavior is an attempt to avoid it.

Dretske's model blends a notion of behavior taken from various biological and behavioral sciences with Irving Thalberg's component model of action.[9] On Thalberg's view, an action has a bodily motion, as well as a psychological entity, as components. In the behavioral sciences, behavior — whether of plants or animals — is internally caused change. A rat's moving its paw is something the rat does — an internally caused change — whereas a movement of a rat's paw is something I can do.[10] Dretske's synthesis is a model that extends the Thalbergian view so that it applies to all behavior, including intentional action. On his view, behavior is a composite process that starts with an internal cause — in humans, a compound of belief and desire — and has a bodily motion as its product. Adopting Dretske's symbols, if "C" designates this internal cause and "M" a bodily motion, then behavior is (identical to) the process C's causing M, or [C → M]. It is not identified with the product M, or (more precisely) with the product M's occurring.

One important metaphysical feature in Dretske's account is that it substitutes a part-whole relation between behavior and bodily motion for an identity relation. On Dretske's view, to identify an action with a bodily motion at the

token level is to commit a sort of fallacy of division, whereby one confuses a part with the whole. While the Dretskean view still allows for different descriptions of behavioral tokens, it is incompatible with the "Davidsonian" view of the nature of the token. A bodily motion can be either a part of a token behavior or the whole of it; a token action is either the process of C's causing M, or the event of M's occurring.[11] The latter of course, is a "Davidsonian" or "product" view of action.

A second, more critical, metaphysical feature in Dretske's process view is the fact that behavioral processes are non-reductive in the following sense. Even if we grant that each action is a token process [C → M], it remains an open question as to whether or not this process is nothing but the sum of the token events that comprise it (assuming, as most people do, that processes are composed of events). If it is, the result is a reductive or mereological view in which a process is identical to the sum of its parts. If it is not, then a behavioral process is not reducible to its parts; it is something more than their sum. Dretske chooses the second possibility, in effect positing a relation of constitution, but not identity, between a behavioral process and its parts.

The non-reductive view of processes is directly related to Dretske's motivation for his alternative account of behavior. If a process is nothing but the sum of its component events, there is no deep ontological difference between a process and its components; and if there is no deep ontological difference, then it is open to a "Davidsonian" to analyze a process to its components, identify the behavior with the final product M, and then simply add as an explanation that M must have the right kind of internal causal history in order to be truly describable as a piece of behavior. However, this decomposition would abet the very confusion that Dretske is eager to avoid. The non-reductive view of processes reflects his insistence that we not conflate behavior with output and that explaining the one is not explaining the other.[12]

Dretske provides two arguments for a non-reductive view. His first argument is by analogy, when he describes a process as a "complex entity" in the way that a marriage is a complex entity, such that it is "a mistake to identify *a marriage* with the pair of people who stand in such a [marriage] relation to each other". He writes:

> A process, as I am now using the term, is not simply a temporally extended entity, a mere succession or sequence of events. ... Even the simplest movements have temporally distinct phases — an infinite number of them, if Zeno is correct. Processes are something else. Or something *more*. A process is the bringing about, the causing, of a terminal condition, state or object — what I

> have so far called, and will continue to call, its product. ... A marriage is a more complex entity than a pair of people *who* stand to each other in the marital relation. It is their *standing* to each other in this relation, their *being married.* And so it is with a process. A process isn't a sequence of events *which* stand in certain causal relations to one another. It is their *standing* in these relations to one another — one event (or two or more events) producing or bringing about another.[13]

Tokens of C's occurring and M's occurring and a causal relation between them are all components of the behavioral process C's causing M, which is, in Dretske's terms, "something more" than the sum of the parts of this causal sequence.

He supplements the analogy with an argument that builds on the idea that a process brings about a product. The suggestion that processes are individuated as *purposive* causal sequences, and not just causal sequences *simpliciter,* may reflect the idea in behavioral science that behavior is *environmentally responsive* internally caused movement.[14] The products in terms of which behavioral processes are defined are typically environmentally determined: a gazelle's fleeing from a lion is a behavior described in terms of a goal or purpose relative to its immediate environment. Dretske argues that a process is *defined by* its product — that the occurrence of the product is what makes the process the process it is.[15] Striking out a batter begins with the arm movements involved in pitching and ends with the batter's missing his third swing at a ball in the strike zone. The occurrence of the result defines the process of which it is the result. There is no striking out of the batter until the batter has been struck out.

He illustrates the point in his discussion of learning and, in particular, the phenomenon of response generalization, in which what is learned in a given trial is properly described in terms of a goal rather than in terms of making particular movements. In one study he cites, human subjects had their arms strapped to a board with the palm facing downwards, and were conditioned to withdraw their fingers from a shock-emitting electrode at the sound of a buzzer. After they were conditioned, their arms were strapped with the palm facing upwards. When the buzzer sounded, they *withdrew their fingers* as conditioned even though the movements that partly constituted this action were quite different. In Dretske's terms, what the subjects learned to do was

> a process defined by its result (getting the finger away from the electrode) rather than by any particular way of producing that result. Though N was always brought about via M during learning, what the subject learned was to produce N, not to produce it via M. This is why reasons, when they help explain behavior, explain why N is produced, not why it is produced in the way it is.[16]

In this example, the bodily motion components of the action tokens differed without there being any difference in the type of action being performed. Although the bodily motion types differed, the behavioral types did not. Moreover, this behavioral type would not have differed even if the arms had *only* been in the palm-downward position — that is, if there had been only one bodily motion type. If the behavioral type defines the process, each token process is what it is because it satisfies that type.

It seems to follow that for any given token process, even if its component bodily motion had differed by a few muscle fiber contractions, its identity would not change; it would remain the same token that it is, individuated by its behavioral type. This suggests that token behaviors are *essentially* individuated by their behavioral-types, where these process-types are defined by a product-type that is not identical to a bodily motion type. A token behavior counts as a token of a given process-type iff it yields a token of that product-type, even if the physical makeup of each token product is never exactly the same twice. This would mean that each *token* behavior is not identical to its components — that constitution is not identity for behavioral processes. For example, if token finger-withdrawing B at time t had involved component motion M′ rather than the M it actually did, where M and M′ differ by a few muscle fibers, B would have been the same token finger-withdrawing. Its identity would persist through changes in physical composition.

Whether Dretske himself actually intended his non-reductive account to include the denial of token-token identity (and not just type-type identity) for behavioral and physiological processes is not clear. But this reading is consistent with what he does say and, just as importantly, with his original motivation. For it would not be of much use in avoiding the conflation of action with bodily motion if behavioral and physical process types differed yet each behavioral token were identical to a physical process token. To switch from a view in which action is identified with a bodily event to one in which action is a process identified with a bodily process would not make much sense given this motivation. So while I will not further defend this reading of Dretske here, it is one that Dretske might endorse. Moreover, a non-reductive view of this sort poses a challenge to the "Davidsonian" view of action that Dretske opposes. This challenge can be illustrated by Mele's (1991) criticism of Dretske.[17]

A non-reductive "product" view?

Mele, who defends a "product" view of action, argues that Dretske's process view of behavior is not essential to his defense of intentional causal efficacy — that Dretske's account of intentional causal efficacy works (if it works at all) when combined with the "product" view. Although Dretske's account of intentional causes will be discussed in the next section, for the moment it suffices to say that it explains behavior [C's causing M] by saying that C was *recruited for* that purpose because of its content; if C had had a different content, this type of behavioral process would not have come about. C might have been hooked up with N, or with nothing at all. Mele's point is that the very same recruitment story can be told even if we identify action with the product M. We can say that M has the action description it does because it is caused by a C that was recruited as the cause of M in virtue of its content. So, Mele concludes, the fact that Dretske adopts a process view of action isn't truly essential to his defense of intentional causal efficacy.

Mele's criticism is correct in the sense that a recruitment story can be told to explain action whether or not action is a process or a product. However, it is incorrect if the claim is that this is all a "product" theorist needs to do to take advantage of Dretske's defense of intentional causes. This is because the recruitment story is of no use as a defense of intentional causal efficacy on *either* view of action *without* the non-reductive metaphysics that Dretske's process view involves. For if the action is identified with the product M, the obvious response — the one that motivates Dretske's account of behavior — is that we already have a complete physical explanation of M qua movement. Mele's suggestion simply illustrates how the critical part of Dretske's non-reductive process account of behavior is not that it is a process, but that it is non-reductive: a process is constituted in, but not identical to, the sum of its parts. In fact, Dretske's embrace of a process view may be due in part to the fact that non-reducibility may seem easier to defend in the case of processes than in the case of single events. The point remains that the recruitment story works (if it works at all) for a *non-reductive* view of action, whether action is a process or a product.

This presents a challenge to those like Mele who would combine a "product" view of action with Dretske's account of intentional causes: for given the motivation for Dretske's non-reductive account, a "product" theorist is also motivated to deny that the relation between a token action and a token bodily motion is identity.

Such a view has already been defended in the case of the relation between

mental events and neurophysiological events, most recently by Pereboom (2002).[18] Take the case of a token of C (or C's occurring). By hypothesis C is "internal" in terms of its physical realization or "whereabouts".[19] But if we take seriously the widely held view that a mental representation's content is essential to it, then we may want to deny that it is identical to some neurophysiological token. For if the neurophysiological elements of the neural token had differed — by a few molecules, say — we would want to say that the token of C would be the same token, with the same content.

Mele's proposal and Dretske's motivation provide reason to defend a relation of constitution, not identity, between actions and bodily movements even on the "product" view. Such a relation is plausible for two reasons. First, we already know that M is *not* individuated by its physiological constituents. As Dretske notes, we can perform tokens of the same type of bodily motion many times without any two of those motions being the same at the physiological level.[20] Bodily motions *supervene* on physiology in that a difference in bodily motion entails a physiological difference, and sameness of physiology entails sameness in bodily motion. But this does not entail that each token bodily motion could not differ physiologically in some way without remaining the same token bodily motion. If such changes make no difference to its identity, the identity of the token is not determined by its physiology.

Second, it is plausible to think that the bodily motions that count as actions on the "product" view are individuated by their relations to the intentional state in virtue of which they count as actions. For what distinguishes bodily motions that are actions from those that are not is this causal link; it would follow that this link is essential to them, and that the precise physiology of any token M is not. The action would be constituted by the event M, which is essentially individuated not by its physiology but by its relation to the intention that is its cause.

Before I proceed, I will summarize the argument so far. I have shown how Dretske argues for a process view of behavior, in which bodily motions are components of behavioral processes. This is incompatible with a "Davidsonian" or "product" view in which behavior is a bodily motion event with a special kind of cause. I also argued that the critical element in Dretske's account of behavior is that it is non-reductive: processes are more than just the sum of their parts. This view is motivated by Dretske's desire to avoid conflating action with bodily motion, a conflation that can prejudice the case against intentional causal efficacy. As such, it challenges those like Mele who prefer a "product" view to defend a non-reductive version that respects Dretske's motivation.

What I will argue in the remainder of this paper is that the non-reductive

process ontology that Dretske defends for behavior can be extended to the case of intentional causes, and that by doing so the intentional properties of mental representations can be triggering causes of behavior. This is the causal role for intentional content that critics found wanting in Dretske's account of intentional causes.

Mental representations as non-reductive processes

The primary example that Dretske uses in his defense of intentional causal efficacy is the "behavior" of a thermostat, which is analogous to more complex animal and human behavior in ways that illustrate the central elements of his account. Using this example, a quick summary of his position is as follows.

On his view, two distinct causal explanations can be given as to why a given behavior occurred. Suppose the thermostat opened the garage door. Each explanation of the thermostat's "behavior" supports a distinct true counterfactual. (a) If the engineer hadn't wired the thermostat to the garage door, the thermostat would not have opened the garage door. (b) If the temperature had not dropped, the thermostat would not have opened the garage door. The first counterfactual relies on an explanation that picks out what Dretske calls the *structuring* cause. This is the cause of the [C → M] process-type or behavior-type that is selectively instantiated on occasion, normally as an effect of environmental stimuli. The second counterfactual relies on an explanation that picks out what Dretske calls the *triggering* cause. This is the environmental stimulus that sets off a [C → M] process token.

According to Dretske, contents play a causal explanatory role as structuring causes, not triggering causes. States of type C are recruited in learning to cause products of type M, yielding behaviors of type [C → M], *because* states of type C have the content they do. In the example, C corresponds to the bending of the thermostat's bimetallic, which represents the temperature. It is "recruited" by the engineer to open the garage door because it has this representational content. (We may suppose, in this case, that the engineer wants to avoid getting out of his car to open the garage door in cold weather, and that otherwise he doesn't mind doing so.)

As noted, one widespread criticism of this account was that it doesn't tell us what we wanted to know. As Block (1990) puts it, "we can tell the whole mechanistic story about this causal process without saying anything about how the mechanisms that subserve it arose." In other words, what we get is a role for content as

structuring causes of behavioral types, but what we want is a role for content as triggering causes of behavioral tokens. We want an account of how Clyde's thinking about beer causes his getting up to get beer, not why his thoughts about beer cause beer-getting behavior rather than pretzel-getting behavior.

However, now there are two degrees of freedom in Dretske's account: on the behavior side, processes or products; on the causal side, triggering or structuring causes. Dretske himself, after making these distinctions, combines the process view on the behavior side with structuring causes on the cause side. Mele's criticism shows that these variables are independent, since he argued that structuring causes can be combined with the product view of behavior. What I want to suggest is yet a third possibility: combining the process view on the behavior side with triggering causes on the cause side, where the intentional causes are also non-reductive processes.

Using Dretske's symbols, the idea is that the process [C's causing M] is a composite of *two* non-reductive processes, one causing the other, each of which is designated by a name or description of the product whose occurrence makes the process the process it is. Representing is a process that has a mental representation C as its product; the term "C" refers to this process. Similarly, bodily motion is a process that has a final movement M as its product; the term "M" refers to this process. This requires a modification in what we call [C's causing M], but this is not a deep problem; for convenience, Dretske himself often uses "M" to refer to behavior. In any case, there is nothing unusual in the claim that processes can trigger other processes. If so, then in principle intentional causes *can* trigger behavior: Clyde's beer-representing process, which is so defined because it has a beer-representation as a product, causes his getting off the couch to get beer. Defending this claim requires defending the suggested view of representations as non-reductive processes and showing how this difference allows for token representations to be causes in virtue of their intentional contents. I'll address the first issue here and turn to the second issue in the next section.

Although no thermostat has mental states, the thermostat example can illustrate the proposal. Suppose that a thermostat is set at 65°F; in a 65°F environment, the "goal" of the thermostat is to register that temperature. Its registering that temperature in those conditions is an environmentally sensitive process which involves responding to the environment so as to end up representing 65°F. The thermostat's bending of the bimetallic strip in response to environmental stimuli is this representational process, a process defined by the "goal" or product of registering 65°F. When everything goes normally —

when I'm not poking the bimetallic with a pencil, for example — the "product" (the bimetallic's being bent at the right arc) "represents" 65°F in virtue of its relation to the environment.

This view is consistent with Dretske's argument that in general we distinguish processes and events because the times of events and of processes are not the same.[21] The thermostat's representing 65°F (the bimetallic's bending to the right arc) occurs over a time interval, of which its reaching the right arc is a product. It follows, however, that if its bending is a dynamic process, then this product may be an interval of that process and not a particular event. It may not make sense to identify a particular event in a continuum as "the" representation of 65°. It is more plausible to say that there is an interval at the end of this process any point during which it may be said to have a "mental representation" of 65°F.

In the human case, we already know that human cognitive processing occurs over time rather than at a time.[22] That it takes time is presupposed in cognitive research that employs the "subtractive" method whereby (for example) the time it takes for a person to discriminate between two lights is measured by subtracting the time it takes to respond to a single light task from the time it takes to respond to one of two lights in a second task. An example from Posner and Raichle (1994) can help to illustrate how mental representation is plausibly viewed as a process.

In these experiments, subjects positioned before a computer screen displaying an unmarked grid are asked to form visual images, say of the letter "F", and to verify whether a mark on the grid falls on or off the image. Immediately after the mark is presented, subjects are slow to make this verification. After a short delay, verification is quick for marks that lie on the upright stroke of the "F" but remain slow for marks on the cross bar. It seems, then, that the letter image is generated in much the same way as a subject might write it.[23] This generating-an-"F" process is defined by its goal, the generation of an "F" image; there is no generating-an-"F" process unless and until an "F" image is produced. Tokens of generating-an-"F" can include "F" images that differ somewhat, but these will all count as generating-an-"F" tokens because the images all count as "F" images. And even though the cell assemblies that subserve any token of this process need never be precisely the same, the process-type does not differ if the product does not differ.[24] Consequently, it seems plausible to claim that if a token generating-an-"F" had been subserved by a cell assembly containing, for example, a neuron with a shorter dendrite than it actually has, it would have been the same token process. Alternatively, a token generating-an-"F" might

determine a set or equivalence class of several functionally equivalent neural aggregates that differ by a few neurons but which can subserve the token cognitive process equally well. It would be arbitrary to identify the token generating-an-"F" with any one member of the set. Either way, constitution will not be identity for representational processes defined by their products.

This non-reductive result conforms with intuition: for example, when I bend the thermostat's bimetallic with a pencil, although the same physically described series of events might occur, there is no representing going on, no representation results, and the subsequent opening of the garage door does not count as thermostat behavior. Similarly, if I stimulate Clyde's brain with electric probes and his body moves off the couch as a result, we do not ascribe beer-thoughts to him, nor do we credit him with performing an intentional action. These judgments are consistent with a view in which representations and behaviors are not defined by their physical constitution.

There is, of course, the obvious rejoinder to the non-reductive view: if the underlying physical series of events are the same, from stimulus to mental representation and from mental representation to bodily motion, what does the intentional stuff matter? Even with all this fancy talk of representing as a dynamic process and its products as intervals, it's all just physical, the objection goes. The objection holds that there is no difference between the thermostat's representing the temperature or my poking the bimetallic with a pencil, or between Clyde's deciding that he wants a beer or my stimulating his brain with electrodes. More precisely, it holds that in each of these pairs of cases there is no *essential* difference: we are merely given different descriptions of what in each case is essentially nothing but physical activity. If the underlying physical activity doesn't differ, there is no difference. By the same reasoning, the causal relations involved in these processes are also just physical; the intentional properties don't really do anything. In short, if the process's underlying physical activity differs, then its causal powers differ. It is this last claim that I must deny if mental representations are triggering causes in virtue of their intentional properties.

Mental representations as triggering causes

Before proceeding, I will review the account so far. Suppose subject A's generating an "F" now (a token of C) causes her pressing the key now (a token of M). By Dretske's account, the process-type [C's causing M] is explained by saying that states of type C are recruited as the cause of states of type M *because* they are C's; their content is the structuring cause of [C's causing M]. I extended this

account to argue that C and M are themselves processes, one cognitive, one behavioral, such that C's causing M is a case of one process's causing another. I also argued that Dretske's non-reductive view of processes implies the denial of token-token identity for processes: tokens of C-processes (generating-an-"F") and of M-processes (key-pressings) are constituted in but not identical to physical processes. This is because processes are defined by their products, which may be defined in terms of some goal relative to the environment.

I then raised the obvious objection: at the token level these are all just physical processes. It should be clear, however, that the objection begs the question as to whether the physical makup of these events or processes *also* determine their identity. It simply asserts that a token mental representation just is a neural process which can be intentionally described. There is no question that physical events or processes underlie, implement or constitute Clyde's representing and his behaving. But the burden of my argument above, following Dretske, was to show that, *pace* the objector, the identity of representation or action tokens is not so fragile that any physical difference yields a different representation or action.

A parallel question arises with reference to causal powers. The objection asserts that a token mental representation's causal powers just are the causal powers of its neural correlate. To show how this can be false, I'll begin with an analogy, derived from Block (1990). A sleeping pill has the power to cause sleep in humans. It has the mix of active and inactive ingredients that it has in virtue of the fact that it is supposed to have this power. Each molecule of active ingredient has its own causal powers. But only the pill has the causal power to cause sleep in humans, a causal power defined in terms of its desired product. This is the causal power is what we are asking about when we ask whether the pill's causal power is identical to the causal powers of its individual ingredients. And it seems clear that the loss or gain of a few molecules of active ingredient will *not* affect the causal power ascribed to the pill, for it will still have the power to cause sleep in humans despite these changes. So the causal powers ascribed to the pill should not be identified with those of its physical makeup, since its physical makeup can change and the pill's causal power will not.

In the cognitive case, the causal powers of individual neurons are not in dispute. The issue is whether the causal power ascribed to a representation just is that of the cell assemblies that subserve it. But if the cell assemblies subserving subject A's generating an "F" now can undergo certain physical changes without there being a difference in the causal powers ascribed to the mental representation, then the representation's causal powers are not identical to those of its

physical constitution. This seems to be the case. The identity of a token representation's causal power is not so fragile that a difference in the length of a dendrite, for example, entails a difference in its causal power. Alternatively, if there are several candidate cell assemblies that can subserve a token cognitive process, it would be arbitrary to identify its causal power with that of any one of the candidates. In either case, the representation's causal powers are constituted in but not identical to the causal power of the aggregate neural processing that subserves it.

We end up with a story that goes like this. Clyde's believing that there's beer in the fridge and his desire for beer cause him to go fetch a beer. His believing that there's beer in the fridge is a bit of cognitive processing defined in terms of its product, a belief that there's beer in the fridge. This token cognitive processing is constituted in but not identical to some bit of neural processing. It causes a token beer-fetching in virtue of its being about beer because it is a token of a cognitive process-type that causes tokens of beer-fetching processes. *That* lawlike regularity, between beer-thinkings and beer-fetchings, simply follows from Dretske's structuring cause story, which no one has denied. What I've added is that his believing that there's beer in the fridge is constituted in but not identical to a bit of neural processing, and that its causal powers are similarly constituted in but not identical to those of its correlated neural processing. The objection fails because it claims that the token (and its causal powers) is nothing but its neural processing (and its causal powers), and this is what I have denied.

I should add that the refusal to identify cognitive or behavioral processes and their causal powers with physical processes and causal powers does not entail a denial of the causal closure of the physical, the claim that every event that has a cause has a physical cause. This may be understood as claiming that every event has a cause that is either a simple physical event or is constituted by physical events. This reading leaves open whether the identities of tokens of complex events, series of events or processes composed of events are determined by their physical constitution or not. Someone who denies that constitution is identity for cognitive or behavioral processes and their underlying physical processes can assert this principle of physical causal closure without begging the question against herself.

If this sort of defense of intentional causal efficacy is possible, then Dretske's error may just have been that of not drawing out the potential of his process metaphysics in addressing the problem of intentional causes. His critics may have been disappointed with the account of intentional causes that he actually gave. But I hope to have shown that there still remains the account he might have given, drawn from his non-reductive account of behavior.

Acknowledgements

I wish to thank the editors of this journal, Profs. Nakita Newton and Ralph Ellis, for their helpful comments and encouragement. Thanks also to Prof. Pete Mandik and the audience at the William Paterson University of New Jersey philosophy department colloquium, in November 2002, where I had the opportunity to present an earlier version of this paper.

Notes

1. For example, see Block (1990), Cummins (1988) and Mele (1991).

2. See Block (1986:668). A more subtle example is found in Pettit (1986:35–36): "The following two propositions are irresistible: (a) Every action is at least partly constituted by a change in the context-independent character of the agent, usually a manifest change describable as a piece of behaviour. (b) Every such change is caused by a narrow state of the agent, a state requiring nothing of the environment. The first thesis is entirely uncontroversial, since all that it rules out is action at a distance of the sort that would involve no narrow change in the agent. The idea of such action is empirically outlandish and not even clearly coherent." In this quote, the first token of "action" refers to intentional action, while the second token refers to a bodily movement, a physical action. The first proposition — that every action contains a bodily movement as a constituent — is compatible with (e.g.) the component view of action, defended by Thalberg and Dretske. This proposition is entirely independent of the second proposition, that bodily movements have internal physical causes, which is accepted by everyone. Accepting the component view of action has nothing to do with ruling out physical action at a distance.

3. Examples that have become part of the standard vocabulary in philosophy of mind include Stich (1983) "autonomous behavior", Block (1990) "behavioral output" and Kim (1982) "bodily action".

4. Two prominent critics of the input-output model, which casts behavior as output, include Hornsby (1986) and Hurley (1998a).

5. More generally, such views are sometimes called "product" views of behavior. Mele (op. cit.) defends a product view.

6. See Figdor (2002) for an extended defense of this point.

7. The epiphenomenalism problem — the view that intentional properties don't play genuine causal roles in explanations of behavior — has been raised against Davidson's view directly, but versions of it that apply more generally appear in Block (1990) and Kim (1993) among many others.

8. Dretske (1988:36). The point is made throughout Ch. 2.

9. Thalberg (1977).

10. Dretske (op. cit), 13.

11. The claim that behavior has an internal cause does not entail the component view. A Davidsonian can identify behavior with bodily motion at the token level and distinguish them by saying that only bodily motions that have the right kind of cause can be the subject of true behavioral descriptions. For the purposes of this paper, we can suppose that the component view of action is supported by at least some good arguments and that it is a credible position for Dretske to take.

12. It is also consistent with the view of critics that have described his account as a "dual explanandum" strategy, such as Kim (1993) and Jacob (1997).

13. Dretske (op. cit., 34–5).

14. Dretske of course emphasizes this, but see also Millikan (1993).

15. I paraphrase Dretske's description on p. 19.

16. Dretske (op. cit., 135).

17. See Mele (op. cit.). I thank an anonymous referee for this reference to Mele's criticism.

18. Pereboom and Kornblith (1991) and Pereboom (2002) defended a similar metaphysics for mental events, in which token-token identity is denied on the basis that constitution is not identity. The account I provide below is inspired by their account and by Johnston (1991).

19. Dretske, (op. cit., 38).

20. Dretske (op. cit., 131–2) notes that fixed motor patterns, where physiology is identical, are relatively rare. Also, bodily motions should not be confused with so-called basic actions, such as moving my finger, which are still actions.

21. See Dretske (op. cit., 15–18).

22. See Posner and Raichle (1994) for detail on the experiments discussed in this paragraph.

23. Petersen and Fiez (1993) note that even just the processing of single words involves many elemental operations that occur in distributed functional areas of the brain. Each of these operations is a process that occurs over time.

24. Paul Churchland (2002) calls the brain "an internal Heraclitean river" in which "the current context into which each sensory input arrives is never exactly the same twice, not twice in a lifetime".

References

Baker, L.R. (2000). *Persons and Bodies* (New York: Cambridge University Press).

Block, N. (1990). Can the mind change the world?, in Boolos, G., ed., *Meaning and Method: Essays in Honor of Hilary Putnam* (Cambridge: Cambridge University Press), 137–170.

Churchland, Paul (2002). Outer space and inner space: The new epistemology, *Proceedings and Addresses of The American Philosophical Association* 76 (2) (November), 25–48.

Cummins, R. (1988) Comments on Dretske, in R. Grimm & D. Merrill, eds., *Contents of Thought* (Tucson: University of Arizona Press), 44–55.

Davidson, D. (1970). Mental events. Reprinted in *Essays on Actions and Events* (New York: Oxford University Press, 1980), 207–225.

Dretske, F. (1988) *Explaining Behavior* (Cambridge: The MIT Press).
Figdor, C. (2002). What the computer model does not entail (unpublished MS).
Hornsby, J. (1986) Physicalist thinking and conceptions of behaviour. In Pettit & McDowell, 95–115.
Hurley, S. (1998a) *Consciousness in Action* (Cambridge: Harvard University Press).
Hurley, S. (1998b) Vehicles, contents, conceptual structure, and externalism, *Analysis* vol. 58, no. 1 (January), 1–6.
Jacob, P. (1997) *What Minds Can Do* (Cambridge: Cambridge University Press).
Johnston, M. (1992) Constitution is not identity, *Mind* vol. 101 no. 401 (January 1992), 89–105.
Kim, J. (1976). Events as property exemplifications, in M. Brand and D. Walton, eds., *Action Theory* (Dordrecht: D. Reidel Publishing Co.), 159–177, reprinted in *Supervenience and Mind.*
Kim, J. (1982) Psychophysical supervenience, *Philosophical Studies* vol. 41, 51–70, reprinted in *Supervenience and Mind.*
Kim, J. (1993) *Supervenience and Mind: Selected Philosophical Essays* (New York: Cambridge University Press.
Mele, A. (1991). Dretske's intricate behavior. *Philosophical Papers,* vol. XX, no.1, 1–10.
Millikan, R. (1993) *White Queen Psychology and Other Essays for Alice* (Cambridge: The MIT Press), 135–170.
Pereboom, D. (2002). Robust nonreductive materialism. *The Journal of Philosophy* 99 (10) (October), 499–531.
Pereboom, D. & Kornblith, H. (1991) The metaphysics of irreducibility, *Philosophical Studies* vol. 63 no. 2 (August), 125–145.
Petersen, S. & Fiez, J. (1993) The processing of single words studied with positron emission tomography. *Annual Review of Neuroscience* 16: 509–30.
Pettit, P. (1986) Broad-minded explanation and psychology, in Pettit, P. & McDowell,J. 17–58.
Pettit, P. & McDowell, J., eds. (1986) *Subject, Thought and Context* (New York: Oxford University Press).
Posner, M. & M. Raichle (1994). *Images of Mind.* (New York: Scientific American Library).
Steward, H. (1997) *The Ontology of Mind: Events, Processes and States* (New York: Oxford University Press).
Stich, S. (1983). *From Folk Psychology to Cognitive Science.* (Cambridge: The MIT Press).
Thalberg, Irving (1977) *Perception, Emotion and Action: A Component Approach* (New Haven: Yale University Press).