Information, Belief and Causal Role

Paul Skokowski
McDonnell-Pew Centre for Cognitive Neuroscience
Oxford University
South Parks Road
Oxford, OX1 3UD
England

Introduction

In an important series of books and papers, Fred Dretske has developed a theory of content to explain how information can be put to work in a system in which internal states indicate outside conditions.[1] The theory formulates an account of belief which combines information-based semantics with a naturalistic approach to learning. This combination allows for an explanation of the fundamental ability of beliefs to misrepresent, or have false content. The theory accounts for the difference between information and representation both causally and in terms of content.

Dretske's theory of belief ultimately depends on the concept of the promotion, also called the selection or recruitment, of an internal indicator type of state to a representational type.[2] Promotion is what allows learning to solve the "Design Problem" of getting a system to do M when and only when an external condition F exists. Indeed, Dretske tells us that an indicator is "... *selected* as a cause of M because of what it indicates about F. Unless this is done, the Design Problem cannot be solved. Learning cannot take place."[3] Promotion, therefore, is a cornerstone of the entire program. But the concept of promotion offered by this theory is flawed by a failure to individuate internal state types by the only tools available: content and causal role. This failure leads to a breakdown of the concept of promotion with the consequence, as we have just seen, that learning cannot take place.

Work on this paper was made possible by a McDonnell-Pew Fellowship from Oxford University. I would like to thank John Perry for critical comments and sage guidance on earlier versions of this paper, and Martin Davies for a particularly helpful piece of advice.

[1]See, for example, Dretske (1986), (1988a), (1988b), (1989), (1990a), (1990b), and (1991).

[2]Dretske (1991, pp. 214-216) is explicit that it is *types* of states that get promoted. Not individual states. Also note that promotion, recruitment and selection of types are equivalent denotations in his analysis.

[3]Dretske (1988, p. 101).

This is a problem. Without learning belief -- and in particular the ability of beliefs to misrepresent -- can no longer be accounted for. Frankly, I do not wish to undermine the theory of content Dretske proposes. I am convinced that he is on the right track. Therefore I propose that more explanation is needed, and so I view this paper as pinpointing a problem that can be solved with modifications. Indeed, the solutions sketched at the end are consistent with the spirit and framework of Dretske's theory of content. Nevertheless the problems examined here will need to be overcome in order to retain a viable model of the acquisition of representational states through promotion and learning.

I will begin by giving a short introduction to the key points of the theory of promotion and learning. This will include discussions of the causal relationships that internal states enter into with respect to external conditions and subsequent movements. I will then give a simple example to illustrate the causal problems with the theory of promotion, and consider counter-arguments to this example. I will introduce the notions of innocent and efficacious properties, and show that Dretske is committed to efficacious properties when identifying internal state types. Finally, I will discuss how the problems raised by this example are due to inconsistent requirements for the causal roles of types of internal states involved in promotion.

In conclusion two compatible approaches will be offered for resolving the problem. The concept of promotion on offer involves tracking *types* of information carrying states through the learning phase. One approach for resolving the problem with promotion is to track concrete objects instead, while another is to hold that a new state must be installed which then plays the right causal role.

1. Background

Central to the account of promotion and learning are the concepts of *indication*, *representation* and *function of indication*, which help describe how beliefs acquire content and get hooked up to motion through *causal* encounters with the environment. I will therefore begin by discussing these concepts before turning to the details of promotion and learning which rely on them.

For Dretske an internal state of an agent indicates an external condition because of a causal regularity: a certain type of external condition F will cause a type of internal

state B. The trees in front of me cause a state inside my head about trees. Generally, when a state *indicates* some condition, this means the state carries *information* about that condition. Information is different from representational content in this important way: information is content that is true, whereas representational content may be false.[4] That is, the content of a representational state may *mis*represent its surroundings. Beliefs are a paradigm example of representational states. I may believe the Giants are the best team in baseball, while in fact the A's are the best team in baseball. My belief is a state carrying a content, but the content happens to be false. It is a representational state which is managing to *mi*-represent the Giants.

Further, the representational content of a system is that content which the system has the *function* of indicating. A system may indicate, that is, carry information about, many different things; but its representational content is limited to that content which it has the function to indicate. A plane's altimeter carries information about, and therefore indicates, both pressure outside the airplane *and* altitude. Nevertheless the dial only *represents* altitude. That is the function of the altimeter. It has the function of indicating altitude; it does not have the function of indicating pressure.

Beliefs, in this theory, are internal states that can cause motion. My belief that a tree is in front of me causes me to change direction during my stroll. Beliefs are arrived at from the *promotion* or *recruitment* of internal indicators through *learning*. Before learning, a type of internal state *indicates* an external condition of some type, and after learning, my internal state type *represents* that type of condition. Internal states thus acquire the ability to *mis*represent due to the learning process, and become full-fledged beliefs. Using Dretske's notation, after learning, state B gets its hands on the steering wheel by becoming a cause of motions M. It becomes an executive state capable of guiding behavior. Learning is a promotion or recruitment of B to cause M in conditions F. Because of these facts, the content of B *explains* the agent's behavior, M.

A brief word on conative states and external contingencies. Usually it takes both a belief state *and* a desire state to cause a motion. I will assume a desire state for a goal exists from now on, and concentrate on beliefs only, for simplicity. This doesn't affect the analysis that follows, since promotion concentrates on representational, not conative, states.[5] I will also assume that the external contingencies required for the promotion of information states to beliefs remain more or less constant throughout the learning

[4]Dretske (1981), (1988), and Skokowski (1994).
[5]This simplification follows the development in ch. 4 of Dretske (1988).

process. Note that individual learning results in an *internal* change in an agent rather than a change in some external contingency. I can hit a slice backhand now because I have learned to do so; nothing has changed in the ball, racquet, court, opponent, or other external circumstance that is responsible for the action. (Even my desire to use slices as a strategy needn't change throughout the learning process.) As we've just seen and will soon see again, the changes from learning are exhibited in the internal states of the agent.

The internal indicator states denoted by B are the same type of state before and after learning.[6] Types of states are what are promoted or recruited. Note that individual, particular, states cannot be promoted or recruited in any interesting way. Particular states occur at a time and so cannot re-occur at a later time as the *same* particular. But the same *type* of state can re-occur. I turned left yesterday and again today when an oak was in front of me. But yesterday my belief was caused by the oak in my yard, while today my belief was caused by the oak in the park. The very same particular belief did not cause my movements on both days, though both beliefs were of the same type. So when Dretske talks of promoting or recruiting a state, he means a *type* of state is so promoted or recruited; he does not mean the very same particular state re-occurs.[7]

We could, with Kim, call an individual occurrence of an internal state B an event.[8] Similarly for a particular external condition F or a motion M. I will follow Dretske in calling such individual occurrences *states*. Dretske and Kim also use physical properties to determine a state's type (they refer to the physical property N of a token state as the *type* this token falls under). Think of a state as an object exemplifying a property or properties at a time. Then, given a state with a physical property ***A***, I will follow Dretske and Kim and say that the state is of *type **A***. Thus it is the properties of states which determine their type.[9]

Outside conditions cause internal states; types of internal states get promoted to new roles as beliefs; beliefs cause motions. Causality is a crucial thread which ties this theory of content together, and we need to be clear about the causal roles of internal states throughout the learning process.

---

[6]Dretske (1991, pp. 214-216) is explicit that it is *types* of states that get promoted, and that type stays unchanged as a consequence of promotion. Also see Section 2 immediately below.
[7]Dretske (1991), pp. 214-216.
[8]Kim's (1973) gives a clear discussion of events and causal relations between types of events.
[9]This usage follows Dretske (1988), (1991), Kim (1973), (1991), and Goldman (1970).

Note that pairs of individual states don't constitute nomic dependencies. Saying that an external condition F causes an internal state B just denotes a particular occurrence. The causal relationship, rather, is a dependency we capture by saying that things of type ***F*** are regularly followed by things of type ***B***.[10] It is a relationship between types of objects, not between particular objects at particular times. Though individual occurrences don't constitute nomic dependencies, they may be particular instances of such a dependency (at a time). The lawful character of particular causal occurrences is grounded in a relationship between *types* of states. That is why we continue to use the word *cause* for individual occurrences. For the purposes of this paper, I will understand causal relationships to be strict, and nomological in character.[11]

2. Promotion and Learning

We are now in a position to understand how a type of internal state gets promoted to new executive capacities by changes to its causal properties through learning. These changes are found when "... individual learning is occurring, places where internal states *acquire* control duties or *change* their effect on motor output ...."[12] So a type of internal state gets harnessed to output by acquiring a new physical property, one which enables it to cause motions appropriate under the outside conditions it is indicating. The state type thereby gets recruited (promoted) for an executive role. This process is illustrated by considering how the type of internal indicator C gets its hand on the steering wheel:

> C is recruited as a cause of M because of what it indicates about F, the conditions on which the success of M depends. Learning of this sort is a way of shaping a structure's causal properties in accordance with its indicator properties. C is, so to speak, selected as a cause of M because of what it indicates about F.... learning of this sort must recruit indicators of F as causes of M.[13]

Internal indicators that are promoted are the same type of state before and after learning. Before learning, these states indicate conditions F, and after learning, they cause motions M (and hence comes to represent F). As Dretske says of a state that has been promoted:

---

[10]Causal correlations are between types (properties), not tokens. See Dretske(1991, pp. 214-215) and Kim(1973)

[11]Probabilistic accounts of action are definitely interesting, but they do not figure in Dretske's analysis. Hence, I will not consider them in this paper.

[12]See Dretske (1988, p. 95).

[13]See Dretske (1988, p. 101).

> ...when it represents (say) F, it does so because it is a token of type B, and this type, in virtue of its indicator (informational) properties (its correlation with F) was recruited for control duties (as a cause, say, of movements of type M) because it was an indicator of F. By saying that it -- this structure *type* -- was recruited as a cause of M because of what it indicated about F, I simply mean that as learning progresses later tokens of B...cause M....[14]

So this state type B started as an indicator, and later acquired an executive capacity and began to cause motions. The model of an indicator state type being promoted is reinforced elsewhere:

> B is the type of physical condition whose correlation with condition (type) F makes tokens of B indicate (carry the information that) F (when they do so) and whose relationship with M (established through learning) makes tokens of B (when circumstances -- motivational and otherwise -- are right) cause M.[15]

The internal state type B has two important constitutive properties: the semantic property of indicating external conditions *and* the physical property of causing movements. This conjunction of properties of the internal state captures the executive nature of beliefs, and so I will refer to this conjunction as the *executive principle*. In addition, we have seen that the internal indicator state exhibiting these constitutive properties is of the same type before and after learning. I will call this notion the *promotion hypothesis*. We will return to evaluate the executive principle and the promotion hypothesis after examining a specific example of promotion.

3. The Car Axle

Consider the axle of a certain car. Suppose, for simplicity, that this car has only one gear working. By virtue of a direct connection to the car's wheels and connection to the engine via a clutch, the axle carries information. Due to these connections, the axle is a good indicator of the car's speed and of the number of rpm's of the car's engine. There is a nomic relationship between the vehicle's speed and the speed of rotation of the axle, and between the engine's rpm's and this rotation. These nomic relationships are between types of states: rotation and velocity on the one hand, and rotation and rpm's on the other. The spinning axle, therefore, is a state carrying information about vehicle speed and engine turnover.

---

[14]See Dretske (1991, p. 215). I have substituted my notation of types (B and F) for his type notation.
[15]See Dretske (1991, p. 214, 215). I have substituted my notation of types (B and F) for his type notation.

Now, the types under which these indicator states fall can acquire *functions* of indicating.[16] Note that this is not the same as indication. Types of internal states which have a function of indicating have been selected to indicate what they do. For mechanical systems this requires the intervention of an intelligent agent, while for intelligent agents themselves, acquisition of function occurs through learning. Internal states which have a function of indicating are *executive*, that is, these states help control output in the system of which they are a part. In addition these states have the ability to misrepresent. Our axle can acquire the function to indicate speed by being hooked up via a flywheel and spring to a speedometer. This is the axle's analogue of learning. By being hooked up to a speedometer, the spinning axle is an internal state of a type that has acquired the function of indicating the vehicle's speed. Note that it is the spinning axle and not the speedometer which is playing the analogue role of a belief state here.[17] Internal executive states of a system cause outputs: belief states of an agent cause motions; spinning axle states of our automobile cause speedometer readings.

We have, then, that the spinning axle is of a type that has the function of indicating the car's speed. The type has this function due to a physical hookup with the speedometer. A spinning axle hooked up in this way therefore exemplifies two properties: the property of spinning, and the property of being hooked up to the speedometer. After this hookup, then, a spinning axle occurring causes a motion of the speedometer. Note that in virtue of the hookup, the spinning axle *represents* the vehicle's speed. In Dretske's notation, this would normally be written B(F), that is, the state B represents conditions F; B carries the content that F. The moving axle represents the car's velocity.

Note that B(F) is a relation of representation, rather than pure indication, which gives a fine-grainedness to the content of tokens B. The state of the axle spinning is now about the vehicle's speed rather than about the vehicle's speed *and* engine rpm's. As Dretske points out,

> The specificity of functions to particular properties, even when these properties are related in ways (e.g., by logical or nomological relations) that

[16]Dretske (1988).
[17]This distinction is made explicit in Dretske (1988, p. 105), when indicator functions of thermostats are discussed. The movement of the bi-metallic strip plays a "purely cognitive" role within the thermostat. As such it is an internal state which causes outputs. The subsequent states it causes, such as the shutting off of the furnace, or the movement of the temperature readout dial, are outputs, or motions, resulting from this 'cognitive' state, and are not to be confused with the cognitive state itself. In my example, the axle plays the "purely cognitive" role of an internal state.

> prevent one's being indicated without the other being indicated, is easy to illustrate with assigned functions, functions we give to instruments and detectors. . . . We can make something into a voltmeter (something having the function of indicating voltage differences) without thereby giving it the function of indicating the amount of current flowing even if, because of constant resistance, these two quantities covary in some lawful way.[18]

The spinning axle has been assigned the function of indicating car speed and not engine rpm's. It therefore represents conditions of type speed and not of type rpm.

Suppose we unhook the speedometer from the axle and hook up a tachometer. It is now clear what happens. By being hooked up to a tachometer, the axle acquires the function of indicating the engine's rate of turnover. The motions of the tachometer are a causal consequence of this hookup with the axle. A spinning axle hooked up in this way therefore exemplifies two properties: the property of spinning, and the property of being hooked up to the tachometer. After this hookup, then, a spinning axle occurring causes a motion of the tachometer. Similarly to what we noted for the speedometer, in virtue of the hookup, the spinning axle state now *represents* the motor's rpm's. Again using the above notation, this would be written B(G), that is, the state B represents conditions G; B carries the content that G. Thus the axle has been assigned the function of indicating engine rpm's and not car speed. It therefore represents conditions of type rpm and not of type speed.

Let's pause now, and take stock of the situation. First, we have seen from the previous section that pre-learning indicator states are of the same type as post-learning belief states. Second, we have seen two analogues of a learning situation. In both of these examples, types of states get functions of indicating. In the first case, spinning axle states are of a type that acquire the function of indicating conditions of vehicular speed, and in the second case, spinning axle states are of a type that acquire the function of indicating states of engine rpm. We have then, that the pre-learning spinning axle state is of the same type as the post-learning state which has the function of indicating vehicular speed. We also have that the pre-learning spinning axle state is of the same type as the post-learning state which has the function of indicating engine rpm. So the post-learning state indicating speed is of the same type as the state indicating rpm's.

But this conclusion cannot be correct. Compare instances of the two tokens along with their representational contents: the first is a representation that F, whereas the

[18]See Dretske (1988, p. 76).

second is a representation that G. It is the nature of representations to have fine-grained contents. In this case the contents are not only fine-grained, but they also happen to differ. Different functions of indicating lead to different contents; it is this very fine-grainedness of functionality that leads to the fine-grained differences in contents.[19] Representational relations such as beliefs are different from informational relations. A belief that the engine is revving at 4000 rpm is different from a belief that the car is going 47 mph, simply by virtue of differences in content, even though the car may be going 47 mph at the same time it is revving at 4000 rpm and there may be a nomic relation connecting that speed with that rpm. If beliefs are to be type-individuated by their contents, then we have no choice but to say that beliefs that differ in content differ in type.

But that is not all. The conclusion is wrong for another reason: the two states cause different types of motions. We know that one spinning axle state causes a motion of the speedometer, and that another state causes a motion of the tachometer. In this way, the two output instruments will generally differ not only numerically but also structurally. For example, the speedometer may be an analogue mechanical pointer display, whereas the tachometer may be an digital electronic display. States of these output devices have radically different constituent physical properties, and hence are of different types. The two post-learning representational states (beliefs) cause different types of motions, and they do this because the causal relationships in effect are between different types of states.

Hence the two types of beliefs cause different types of motions. This is not surprising since it is by hooking an indicator state up to a type of movement that eliminates its indeterminacy of function; that is, it is this very causal connection that confers a function of indicating upon the type of state in question, thereby making states of that type into representations with a fine-grained content.[20] Thus beliefs that cause one type of motion are a different type from beliefs that cause another type of motion. If beliefs are to be type-individuated by their causal consequences, then we have no choice but to say that beliefs that differ in their consequences differ in type.

[19]See Dretske (1988, p. 76, 77).
[20]See Dretske (1988, p. 70 and Chapter 4).

4. An Objection

One might object to the above by claiming that the sort of thing described in the car axle example just doesn't happen. One function of indicating isn't lost by a system when another is gained. And since there is no loss of function, the same type of indicator state will remain in effect even after promotion and learning. And what is accomplished by this is that the representational state that results is a single one with the function of indicating *both* conditions. Thus, the true analogue to belief in agents isn't the one laid out in the speedometer/tachometer example above. Rather, it is one where the speedometer connection is left intact when the tachometer connection is made. That is, the state to consider is not either connection in isolation, but rather the complex state of both devices being hooked simultaneously. This latter state now has the function to indicate both vehicular speed and engine rpm's.

A reply to this is that the car axle has just given an example of successive and different types of belief states arising from the same type of indicator state. This sort of thing can indeed happen, at least for automobiles. Of course I can disconnect the tachometer when I connect the speedometer. Tachometers may bore me. My axle therefore does lose the function to indicate speed when it acquires the function to indicate rpm's.

The same change in function of indication can occur in agents. Consider John Major's internal Maggie-Detector, which initially indicates that Thatcher is both a potential ally *and* a serious threat. After Major learns to be a smooth politician, his Maggie-Detector acquires the function of indicating that Maggie is a potential ally, and causes appropriate motions such as giving speeches supporting her. But after Maggie criticizes Major on his performance as Prime Minister, Major has gone through a different learning situation. Now whenever his Maggie-Detector lights up, he has a *new* belief that Maggie is a serious threat, which produces appropriate motions such as giving speeches denouncing her views as outdated and irrelevant. This is a case of two distinct types of beliefs being promoted from the same indicator state, and in the process, one function of indicating is lost when another is gained. This results in different types of internal indicator states before and after learning.

The same thing happens for car axles. Different selection processes, or learning situations, were executed by the designer in order to 'recruit' the two different

connections. I hooked up the speedometer because I required a representation of the speed. I hooked up the tachometer because I required a representation of engine turnover. I did not choose that particular tachometer connection because I wanted a representation of speed *and* rpm's. I wanted a representation of rpm's. Remember the fine-grainedness of representational content. This has its origin in the assignment of function to a type of indicator state. Representations of conditions of one type may be established independently of representations of conditions of another type, as was the case for our automobile. The second connection was added in order to obtain a particular representational content. It was not added in order to obtain a dual representation. After all, I can consider the content represented by one dial in isolation from the other. I don't have to consider both together just because there are two connections.

A second reply is that the executive nature of these representational states ensures that they make the system of which they are a part behave differently in different situations. The different states have different remote effects. In this way the movements of agents are akin to the readings on the car's two dials, the speedometer and tachometer: different movements (readings) issue from different beliefs (hookups with the axle). I learn that yucca plants are spiny from painful experience. I learn from my botany class that yuccas are succulents. Assume that there is a nomic relation between yucca plants being in front of me, and a certain yucca-style neural firing in my visual cortex when I fixate on one. Then the neural firing type indicates yuccas in front of me, in virtue of the regularity. Believing (or perceiving) spininess makes me behave in certain ways, and believing (or perceiving) succulence makes me behave in other, distinct, ways. This is so because we move for *reasons*, in order to achieve goals.[21] We behave differently with respect to objects depending on which of their properties best allows us to achieve those goals. Spininess has nothing to do with my digging up a yucca to achieve the goal to bring home a succulent plant. Succulence has nothing to do with running around the yucca to achieve my goal of avoiding getting impaled during the cross-country race. Because of this difference in output effects, beliefs with differing contents remain of distinct types even when they share the common origin of being *learned* from a single type of indicator state.

Finally, note that it is by having belief states hooked to motion that the types to which these states belong acquire their function of indicating. This resolves the

[21]See Davidson (1980), and Dretske (1988) and (1989).

indeterminacy of function indicator state types otherwise have.[22] Our axle originally indicates both ***F***'s and ***G***'s. According to the above prescription, by hooking up the speedometer we remove this indeterminacy. But then, if we accept that by hooking up the tachometer we give the axle the function to indicate ***F***'s *and* ***G***'s, then we give the axle back this very indeterminacy of function we set out to overcome in the first place by giving it representational abilities.

5. Learning Difficulties

The problems raised by the car axle example point to an inconsistency between the executive principle and the promotion hypothesis. Recall that the promotion hypothesis says that internal indicator states are the same type before and after learning. The executive principle says that internal state types indicate external conditions *and* cause certain types of motions. But causal relations between states are nomic relations: if states of one type *cause* states of another type, then they always do so. Before learning, internal indicator states do *not* cause motions of a given type. After learning and recruitment, the *same* type of state *does* cause motions of that type. Hence, internal indicator states are *not* of the same type before and after learning. One type does not cause motion; the other does. We thus have that internal indicator states both *are and are not of the same type* before and after learning. This conclusion can't be right. Hence, the executive principle and the promotion hypothesis are inconsistent.

This argument shows that the notion of promotion or recruitment is in trouble if the executive principle and the promotion hypothesis are both retained. Since internal (belief) states which indicate external conditions are taken to be of the same type both before and after learning, then whenever such a state occurs, a motion follows -- even *before* learning. This brings into question what role recruitment could *ever* have, since internal indicators will *always* cause the appropriate motion under conditions F. What needs to be learned?

A central reason for these problems is that the constitutive properties of internal indicator states are real physical properties with causal consequences. Instances of these states presumably don't cause a motion before learning (they must be *recruited* to do this), therefore they don't have those crucial physical properties required for them to

[22] Ibid.

cause motions. Thus, before indicator states acquire causal efficacy they are of a different type than those which occur after learning, for the reason that the former lack the constitutive physical properties required for causing motions.

But now here is another problem. According to this model, after learning, indicator states are not only of a different *causal* type than before learning, but these states are no longer pure indicators: promotion has resulted in belief states. Beliefs are fallible, and so they can not be *strict* indicators since their contents may be false. And this means that they can no longer satisfy the causal relationship that has so far held for strict indicators, viz., that external conditions of a given type will cause internal indicator states of a certain type. After all, it can no longer be true for representational states, as opposed to informational states, that such a law holds for them. My long lost brother now stands in front of me, but I don't believe my eyes. This is the nature of belief. Hence internal states will also differ in type with respect to content: before learning, these states carried informational content; after learning, they carry fallible content.

These dilemmas show there are problems with the theory of content we have been considering. It appears that learning cannot promote types of states from not being causes into being causes of motion, for, as a consequence of learning, internal states begin to play a different causal role -- and that means either that one type of state disappears when another appears, or that a new type is created which is different from the first type. Perhaps, then, what learning does is replace one type of indicator with another type of indicator. But this isn't quite right either, for the latter type is a representational type, not an indicator type, according to the theory.

6. Innocent and Efficacious Properties

It could be replied that there is a level at which the states do not change type after promotion. Rachel has a can of red playdough. When she takes it out of its container, it is cylindrical. After crafting by the toddler it becomes a (nearly) perfect sphere. Throughout, however, it has remained red. These shapes, which are physical properties, have different causal consequences: one type (cylindrical) can roll only along one axis, while the other type (spherical) rolls along any axis. They were also caused by different causal (nomic) processes: the cylindrical shape was caused by a container, while the spherical shape was caused by pressing between cupped hands.

We can think of Rachel as taking the red cylinder and *promoting* or *selecting* it to a red sphere. Being cylindrical and being spherical are two different properties that enter into different causal relations. Before promotion the red cylinder only rolled along one axis. After promotion the red sphere could roll along any axis, and could be used in new ways; it had a new property. But it was the red object that was promoted. The object is of type RED throughout its acquiring different physical properties. Surely this is the sense in which Dretske means an object remains the same type before and after promotion and learning.

Let me call this the *innocent* interpretation of type identification, and call a property such as RED an *innocent property*. I call RED innocent because this property does not enter into the causal relations which signify instances of promotion. RED isn't caused by a container or by a hand as cylindricality and sphericity are, and RED doesn't enter into the degrees of freedom for the motions of cylinders and spheres. Let me call *efficacious properties* those properties that can enter into causal relations -- in this example, properties such as shapes. Likewise, types will be innocent or efficacious, depending on the properties which determine the types. Therefore by holding that internal states do not change types upon promotion, Dretske must be referring to innocent properties and types in his theory of learning.

But this theory of learning does not choose innocent properties and types for the internal states promoted during learning. It explicitly chooses properties which indicate, and therefore are *caused by*, outside conditions, and other executive properties that *cause* motions. Recall that, before learning an internal state type ***B*** indicates (and hence is caused by) external conditions of type ***F***, but doesn't cause motions of type ***M***; and that after learning, the same type ***B*** represents ***F***, and now causes motions ***M***. Properties such as ***B*** are not innocent properties; they are efficacious properties: properties that enter into causal relations. And again, since ***B*** has different contents and different causal consequences before and after learning, it cannot remain of the same type before and after learning. Efficacious properties are individuated by causal relations; hence states with efficacious properties that enter into different nomic relations are of different types. Innocent properties do not figure in determining the types of states promoted during learning.

7. Conclusions

We learned from the car axle example that different types of representational states may indeed issue from a single type of indicator state, and these new states are typed according to their various causal roles. The conclusion is that these fine-grained beliefs are distinct in content and causal role from the original internal indicator state. We have also seen that Dretske's theory types internal states by their efficacious properties, and not their innocent properties. Thus the promotion hypothesis must be rejected, since we have seen that indicator states must indeed change type as a consequence of promotion and learning.

A more general conclusion, however, is that mental states need to be typed strictly according to their constitutive properties. Goldman has made this claim for acts, as has Kim for events. Both Goldman and Kim are strict about types in this way because without being careful about constitutive properties, their acts/events will have untoward causal consequences. It is important for Goldman to distinguish between the acts John's pulling the trigger and John's killing Smith because the two acts have different causal consequences: John's pulling the trigger causes the gun to fire, whereas John's killing Smith does not.[23] Kim's strict typing of events by their constitutive properties is also designed to avoid such difficulties.[24]

The moral for the theory of content we have been considering is the same: one needs to be careful about the typing of mental states. *Causation* is the crucial link throughout the entire analysis: it underlies the pickup and utilization of information, and is responsible for ensuing behavior. That is why the concept of promotion flaws Dretske's theory of learning by not adequately distinguishing the causal roles of indicator and representational states.

There are three related lessons to be learned from this analysis. First, pre-learning indicator states *differ in type* from post-learning belief states which succeed them. Second, these two types of states differ both in virtue of their contents, which they acquire due to causal relations with outside conditions, and in virtue of their abilities to cause further, succeeding, states (movements). Thus it is causal relationships, *not instances of promotion*, that are the final arbiters when determining types of mental

[23]See Goldman (1970, p. 2).
[24]Kim (1970, p. 227).

states. By examining the causal relationships that mental states enter into with respect to outside conditions and subsequent movements, we may determine the types of these states. Third, because promotion cannot work in the way proposed, Dretske's theory of representation and learning is weakened. But I believe there are modifications available to restore the theory of representational content.

I see two compatible approaches towards formulating a concept of promotion which can work within Dretske's theory of representation. The first approach is to hold that it is concrete objects, rather than *types* of states, which get promoted to new causal roles.[25] These concrete objects will reside inside the head, and must exemplify the very physical *properties* which determine their mental type (indicator vs. representational) and their causal role. Thus the concrete object in question remains the same, and so may be tracked through instances of promotion, but will acquire different properties over time. This approach accepts that the type of state carrying representational content differs from the original indicator state type.

The second approach is to hold that the internal state type remains the same, while, through learning, a connecting state is formed or installed in such a way that when the indicator state is activated, then the connecting state causes an output motion to proceed as required by the theory. The problematic burden of causing motion that the original theory required of the indicator state is therefore taken on by the new, connecting state. This is I think the most promising avenue, in part because it has (some) biological plausibility. Let me explain. Suppose we accept, by way of an example, that we have some sort of chair-indicator which fires in our brain when we visually fixate on a chair. The causal process might look something like this: we look at a chair, an upside down image of it is projected onto the retina, the retinal cells fire in a pattern characteristic of chairs, and this causes, by way of the optic nerve, the Lateral Geniculate Nucleus, and other connections, a characteristic pattern of activation (chair indicator B) in certain neurons in the visual cortex. So much for the hardwiring we normally inherit at birth. Assume this process occurs both for infants, who don't yet have beliefs about chairs, and for adults, who do.

The biological plausibility arises through neural changes which are a direct effect of learning, such as Long Term Potentiation, or LTP. LTP appears to be a chemical

[25]For automobiles, the car axle may serve as the internal concrete object, whereas for agents, a bundle of neurons may do the trick.

process which allows changes to our neural substrate as a consequence of learning.[26] In particular, learning can induce, via mechanisms of LTP, connections between sensory neurons (which indicate outside conditions) and motor neurons (which cause motions).[27] View learning then as installing a new matrix of connections, call it W, between sensory and motor neurons.[28] W becomes a permanent part of the structure of our brain. Now when our internal indicator B goes off, W is in place, and the two states together, B and W, cause a motion M. Notice that the causal burden has been taken off of B, allowing it to be promoted in the sense Dretske's theory would like. Before learning B didn't cause anything. Learning then installs W. B and W together now cause M. B hasn't obtained a new causal property because if you take away the separate state W after learning (chemically or surgically alter the synaptic connections) then B alone won't cause anything. But it *will* continue to be a chair indicator.

Here, then, is a way that a type of state can be promoted into a representational state. The difference is that learning installs a new state W to enable the promotion of B. B still becomes executive, since after learning, when B occurs, M will ensue. B also has the capacity for misrepresentation that Dretske's theory requires: if B fires randomly, or by accident after learning, it now will cause motions M, even when F is not around. This is a form of misrepresentation that is not possible for a pure indicator, because indicators cannot cause inappropriate motions when they indicate, as they are causally inert.

Both of these approaches, I believe, show some promise in accounting for the changes in content and causal role that Dretske's theory requires for representational states. The latter one, however, can maintain the indicator and representational properties of a single type that is promoted, while avoiding the difficulties of giving that same type a new causal role.

[26]Thompson (1986), and Cotman and Lynch (1989).
[27]Greenough, W., Larson, J., and Withers, G. (1985).
[28]Skokowski (1992)

REFERENCES

Armstrong, D.M. 1978: *A Theory of Universals*. Cambridge: Cambridge University Press.

Cotman, C. and Lynch, G. 1989: The Neurobiology of Learning and Memory, *Cognition*, Vol 33.

Davidson, D. 1980, "Actions, Reasons and Causes", in Donald Davidson, *Essays on Actions and Events*, New York: Oxford University Press.

Dretske, F. 1981: *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F. 1986: "Misrepresentation", in Bogdan, R. (ed.), *Belief*. Oxford: Oxford University Press.

Dretske, F. 1988a: "The Explanatory Role of Content". *Contents of Thought: Proceedings of the 1985 Oberlin Colloquium in Philosophy*. Tucson, AZ: University of Arizona Press.

Dretske, F. 1988b: *Explaining Behavior*. Cambridge, MA: MIT Press.

Dretske, F. 1989: "Reasons and Causes", in Tomberlin, J. (ed.), *Philosophical Perspectives,* vol. 3, *Philosophy of Mind and Action Theory*. Atascadero, CA: Ridgeview Publishing.

Dretske, F. 1990a: "Does Meaning Matter", in Villanueva, R. (ed.), *Information, Semantics and Epistemology*. Oxford: Basil Blackwell.

Dretske, F. 1990b: "Putting Information to Work", in Hanson, R. (ed.), *Information, Language, and Cognition*. Vancouver, B.C.: University of British Columbia Press.

Dretske, F. 1991: "Replies to Critics", in McLaughlin, B. (ed.), *Dretske and his Critics*. Cambridge, MA: Basil Blackwell.

Goldman, A. 1970: *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice-Hall.

Greenough, W., Larson, J., and Withers, G. 1986: Effects of Unilateral and Bilateral Training in a Reaching Task on Dendritic Branching of Neurons in the Rat Motor-Sensory Forelimb Cortex, *Behavioral Neural Biology*, Vol. 44, pp. 301-314.

Kim, J. 1973: "Causation, Nomic Subsumption, and the Concept of Event". *Journal of Philosophy*, 70, No. 8, pp. 217-236.

Skokowski, P. 1994: "Can Computers Carry Content Inexplicitly?" *Minds and Machines,* 4, No. 3, pp. 333-344.

Skokowski, P. 1992: *From Neural Networks to Human Agents.* Ph.D. Dissertation, Stanford University.

Thompson, R.F. 1986, The Neurobiology of Learning and Memory, *Science*, Vol. 233, p. 941-947.