

MINDS, MACHINES, AND MONEY:
WHAT REALLY EXPLAINS BEHAVIOR

According to a certain philosophical picture of the way mind and body are related, the mind is to intentional action what money is to the behavior of a vending machine. Just as coins are in (or get deposited in) vending machines, beliefs, desires, and intentions are in us. Just as the right coins deposited in the machine cause the machine to behave in a certain way — to yield its contents: cokes, cigarettes, or candy, as the case may be — so the right mental entities occurring in us cause us to perform various actions. Furthermore, just as what makes money money is not its intrinsic character — shape, size and density of the coins, for example — but certain extrinsic or relational facts about these coins (the fact that they possess monetary value) so too what makes a belief a belief is not its intrinsic neurobiological character, but, rather, certain extrinsic facts about it — the fact that it has a certain meaning or content, the fact that it has certain *intentional* properties.

But if we take this analogy seriously, it suggests that beliefs, *qua* beliefs, are as irrelevant to animal behavior as is money, *qua* money, to the behavior of vending machines. Since it is facts about the shape and size of coins, not facts about their monetary value, that explain why coins cause a machine to yield its contents, the analogy, if we take it seriously — and a good many philosophers do — compels us to conclude that it is the intrinsic features of beliefs, their neurobiological properties, not their extrinsic properties, their meaning or content, that explains why we do what we do. We thus seem driven to the conclusion that what we believe is causally irrelevant to what we do.

I do not think we are *driven* to this conclusion, although, I admit, some people seem willing to drive there. It is the purpose of this essay to say why this conclusion is not forced on us. Let me begin, then, by enlarging the analogy.¹

¹ After writing this paper I came across (Allen 1995) in which a similar analogy is developed to reach a similar conclusion.

1. MONETARY-MACHINE INTERACTIONS

The United States Government does its best to make the (legal) monetary value of objects supervene on the intrinsic properties of the objects that have that value. In the case of paper money, special watermarks, high quality paper, intaglio printing, and security strips (visible only with transmitted light to frustrate photocopying) make successful counterfeiting difficult and increasingly rare. There are no (at least not many) non-\$20 bills that look and feel exactly like real \$20 bills. The same is true, of course, for other denominations and coins. This is no accident. The entire system of monetary exchange depends on it. For understandable reasons, then, the U.S. Treasury Department is dedicated to maintaining the strictest supervenience.

The monetary value of an object is a relational property of that object. It has to do with its *history* — was it produced in a mint or in someone's basement? — and the economic practices of the community in which it exists — are such objects generally accepted as a medium of exchange in the community? Since the usefulness of money depends on its easy identification, governments make every effort to see to it that this extrinsic property of money supervenes on the observable (intrinsic) properties of money — size, shape, markings, weight, and so on. If two objects are observationally indiscernible, if they *look* and *feel* the same, then (if the government is doing its job) they are indiscernible with respect to monetary value. If it *looks* like a \$20 bill, it *is* a \$20 bill. Another way of expressing this² is to say that (as long as counterfeiting is kept in check) monetary values (V) are necessarily realized (usually multiply realized) in an object's intrinsic properties (S), and each value of S has the same value of V. This corresponds to what Kim calls *weak supervenience*.³

² Corresponding to Kim's second formulation of weak supervenience (1984a, p. 64). Though citations are to individual articles, all page references to Kim are to (Kim 1993b) in which the individual essays are collected.

³ At least it is a form of *local* weak supervenience — local to a given nation or economic unit. Though it would complicate monetary exchanges, there is no reason two countries might not assign the same (type of) object different monetary values. If this happened, then, even without counterfeiting, there would be local (i.e., national), but no global (international), supervenience. In speaking of monetary value supervening on the intrinsic properties of an object, I should, therefore, be understood as referring to a given country or economic unit.

As a result of this (normally) widespread supervenience and the correlation associated with it, we can (and regularly *do*) use the fact that something is money to predict and “explain” (more about the scare quotes in a moment) the effects money has in transactions of various sorts. Why did the cashier give me \$8 in change? Because lunch cost \$12 and I gave her \$20. Why didn’t the vending machine give me the candy I selected? Because I only deposited \$.55 and the candy bars cost \$.65.

Are these familiar explanations really correct? Is the fact that I gave the cashier \$20 really the (or part of the) explanation of why she gave me \$8 change? Is the monetary value of the paper I gave her a causally relevant property? The coins I deposited in the vending machine are only worth \$.55, but is this fact relevant to why the machine did not give me a candy bar? Is the value, the legal worth, of these coins a causally relevant fact about them? I know we talk this way. I know that everyday explanations of such results are replete with references to monetary value, but is this extrinsic property the causally relevant property?

It is important to understand that these are questions about the causal relevance of an object’s properties (its being worth \$20), not the causal efficacy of the objects (the \$20 bills) that have these properties. These are, in other words, questions about what *explains* the result, not what *causes* it. Giving the cashier an object with a monetary value of \$20 *caused* her to give me \$8 change. About that there is no argument. The question we are asking, though, is not whether a \$20 bill is a causally effective *object*, but whether its being a \$20 bill explains its effectiveness. Is the value of the paper I give her a fact about the paper that explains the result of giving her the paper? What if I, instead, give her a piece of paper that looks and feels exactly like a real \$20 bill? Would the result be different if we suppose the bill was perfect counterfeit? No, of course not. If she can’t tell the difference, how could it be? Well, if we really believe this, as I assume we all do, then why say that the cashier gave me \$8 change *because* I gave her \$20. Giving her \$20 is the cause, but that it was \$20 is not the explanation, of her giving me \$8 change. The correct explanation is that I gave her a piece of paper that looked and felt (to her) like a \$20 bill. The causally effective properties, those that explain why the effect occurs, are the intrinsic, the observable, properties of the paper on which its being \$20

supervenies, the properties you and I, cashiers and machines, use to *tell* whether it is \$20.

I am not, mind you, recommending that we change explanatory practice. Though I am convinced that its being money is (in most imaginable cases) totally irrelevant to the results obtained, I will go right on explaining the results of monetary transactions in terms of the money exchanged. Though we predict the behavior of vending machines by mentioning the value of the money we put in them (“You have to deposit \$.75 to get a coke.”) we all know that it isn’t the value of the money that explains the result. It is the shape, size, weight, and (for machines that take bills) visible marks of the objects we put in them that explains why machines behave the way they do. An object with the same S and a different V (a slug) would produce the same behavior. Vending machines (not to mention store clerks) are equipped to detect the shape, size, and density, but surely not the economic history of the objects they receive. We nonetheless pretend to explain machine behavior by mentioning the historical-social properties (\$.75) of the internal objects (coins) that cause behavior. We ignore the intrinsic properties that are causally relevant. We ignore them because, often enough, we don’t even know what they are. Nonetheless, given the facts of supervenience, we know that, normally, inserting \$.75 will get you a Coke even if we don’t know which properties of the \$.75 are responsible for this effect (is density relevant?). V is, after all, multiply realizable in S. We can use a variety of different coins, of different shapes and sizes, to make \$.75. The machine will give us a coke, it will behave in the same way, if we insert quarters, dimes and a nickel; or seven dimes and a nickel; or fifteen nickels. As long as the coins add up to \$.75 we get the same result. So it is simpler and much more convenient in our explanations of machine behavior to mention the extrinsic V all the different S’s have in common even though we know it is S, not V, that explains the result. Convenience explains the explanatory pretense.

This, incidentally, is why I am suspicious of philosophical appeals to our ordinary explanatory practice, or to the explanatory practices in the special sciences, to support accounts of what causally explains what (see, for example, Burge 1986, 1989, 1993, 1995; Baker 1995). Our explanatory practice is often governed by practical convenience and, sometimes, theoretical ignorance. I know, for example, that we commonsensically invoke beliefs and desires to explain human (and

sometimes animal) behavior. That, I am willing to concede, is the accepted practice. Even in cognitive psychology and computer science (presumably special sciences) there are a variety of intentional ideas (e.g., data structures, information, representation) that regularly appear in causal explanations. But *saying* that *x*'s having *P* causally explains *x*'s *Q*-ing, when *P* is a relational or — even worse — an intentional property of *x*, doesn't make it so. Even if *everyone* says it. If I trusted explanatory practice this blindly, I would have to conclude that the monetary value of objects explains their effect on vending machines. It will take more than our explanatory practice to convince me of this.

It may be thought that I am constructing a false dichotomy, that the two explanations of a cashier's or a vending machine's behavior — one in terms of the intrinsic *S* properties, the other in terms of extrinsic *V* properties — do not (as I have been assuming) really compete. They aren't mutually exclusive. They can *both* be correct. The explanation in terms of a coin's intrinsic properties is a *proximal* explanation of its effect on the vending machine while the explanation in terms of monetary value is a more *remote* explanation of this same result. It is like explaining a behavioral deficit (stuttering, say) by describing the brain damage that produces the stutter (explanation by intrinsic properties of the stutterer) or by mentioning the incident — being dropped on his head as an infant — that causally explains this brain damage (an explanation by extrinsic — i.e., historical — properties). The first is a proximal, the second a remote, explanation of the stuttering. Similarly, if we think of the fact that the paper I give the cashier has a monetary value of \$20 — that it has the kind of history and use that makes it \$20 — as the causal explanation of its having the observable properties it now has, then social-historical *V* properties causally explain intrinsic *S* properties and, thus, explain (in a more remote way) whatever the *S*-properties causally explain — why, for example, the cashier gave me \$8 change for my \$20.

This objection, though it gets at something interesting about the connection between extrinsic and intrinsic properties in explanations of this sort, is not, as it stands, correct. The facts that give coins and bank notes their value (the *V*-facts) do not causally explain why these objects have the size, shape, and markings they have (the *S*-facts). The reason why \$20 bills have Andrew Jackson's picture on them while \$5 bills have Abe Lincoln's picture, the reason they have *these* particular

observable features⁴, is not because these bills have the value they have. It has to do, rather, with the various decisions and policies of administrators in the U.S. Treasury Department. The pictures on US coins and bank notes might well have been different. If everybody (including the government) agreed, we could, in fact, *make* \$20 bills (the bills that are *now* worth \$20) into bills worth \$5, and *vice versa*.

Nonetheless, though I think the objection mistaken, it raises an interesting possibility, the possibility that the explanatory efficacy of an object's extrinsic properties lies in the complex causal relations between an object's extrinsic properties and its intrinsic nature. I will return to this point later in order to explore this possibility. Pending deeper investigation, though, I assume that the output of people and vending machines in monetary exchanges is not to be explained, not even remotely, by the extrinsic value of the money that produces that output. The causal efficacy of money is not explained by its being money.

When externally individuated properties (like V) supervene on intrinsic properties (S), and the supervenient property is *multiply* realized in S (thereby making it practically convenient to express generalizations in terms of V rather than S) talk of the supervenient properties begins to dominate explanatory contexts and one finds little or no mention of S.⁵ Imagine trying to explain why Clyde got a Coke, not by saying he deposited the required \$.75, but by describing the S-properties that were actually causally relevant. If we happen to be ignorant of exactly which coins Clyde deposited in the machine, the explanation would, of necessity, be radically disjunctive: 15 coins of this sort; or 2 coins of this sort and 7 coins of that sort; and so on and so on. Nobody gives *those* kinds of explanation. What does this show? Nothing. Or, perhaps, only that we are lazy or ignorant.

Despite this undeniable tendency in explanatory practice to drift to the most conveniently expressible generalizations, V-generalizations are not the sort that will support explanations. Predictions, yes, but not

⁴ I understand, of course, that having X's picture on Y is an extrinsic, not an intrinsic, property of Y, but I think my point (about differences in observable markings) is clear enough without going into these fussy details.

⁵ We advert to S only when the V (and design) stance fails — when, for example, there is a breakdown or malfunction in the machine: e.g., the machine didn't give us a candy bar because the coin was bent. See (Dennett 1987) for the same point about the intentional stance.

explanations. In more careful moments — when, for instance, we are doing metaphysics — we realize that it is the object's being S, not its being V, that explains its impact on the system in which it exists. Our explanatory practice does not respect metaphysical scruples. There is no reason it should. In giving and receiving explanations we are not doing metaphysics. We take explanatory shortcuts. We leave the metaphysics for later. Or never.

That concludes my example. Its intended purpose, of course, is as an analogy with mind-body interactions. What I hope to do is to draw some useful lessons from the analogy. For convenience, I will refer to causal interactions between coins and vending machines as *monetary-machine* interactions. If we stopped here, if we looked no deeper, then, despite common explanatory practice, we would have to conclude that, with respect to machines, money was epiphenomenal. That is, the fact that money is money does not *explain* the effects of money on machines (or people, for that matter). It is the S-facts that do all the explanatory work. What explains machine behavior are not “broad” facts about the value of internal coins, but “narrow” facts about their size, shape, and density. Once we have the analogy fully in place, though, I will return to the analysis of these interactions and take a deeper look. What I hope to show is that, contrary to what I have just been arguing, there is a sense in which monetary facts about money *is* causally relevant to machine behavior.⁶ The form of this relevance is, I think, suggestive about the way the mind is causally relevant to human and animal behavior.

2. THE ANALOGY

There is a prevalent view in the philosophy of mind that the propositional attitudes (including belief) are something like internal coins. What you believe (intend, desire, conclude, regret, etc.) is an extrinsic property of the internal belief (intention, etc.) in the same way the value of coins is extrinsic to the coins in a machine. For a materialist (who is not an eliminativist) a belief (some brain state, say) has intrinsic (neurobiological) properties, but it also has a content or meaning (=

⁶ To machine behavior, *not* to machine output. This distinction between output and behavior is a distinction that figures importantly in my account of the way reasons explain behavior in *Explaining Behavior* (1988). Here I merely note the distinction. I return to it later.

what it is one believes) and this is determined, in part at least, by the relations this internal state bears to external affairs. The relational individuation of belief is why the same belief can occur in much different heads; what makes it *that* belief is not the brain state that realizes it (this can be quite different), but the way that brain state is related to the rest of the world. Putnam's (1975) and Burge's (1979) examples have convinced many that, in this respect, beliefs are "internal money."

I will call this view the Standard Theory. I call it the Standard Theory not just because it is widely accepted, but because, if you are a materialist⁷ it is hard to see how something like this view could fail to be true. Beliefs (just like coins) have to be inside the system whose behavior they causally explain. How else could they cause the behavior they are said to explain? Nonetheless, what gives these internal states their content (just like what gives coins their value) is not inside the head. The representational (intentional) character of a belief, what makes it a belief about football rather than philosophy, is a matter of how that internal state is related to external affairs. Surely there is nothing *in the brain* that makes one neurological event about football and another about philosophy. What a person believes — its meaning or representational content — must, it seems, be extrinsic to the believer.

It is the same with words. Words and sentences are printed in books, but what makes some words about football and others about philosophy is not in a book.

3. THE PROBLEM: EPIPHENOMENALISM

The Standard Theory is commonly thought to have the kind of epiphenomenal implications we uncovered in examining monetary-machine interactions. Though the content of a belief — *what* one believes — is routinely mentioned in explanations of behavior (just like the value of coins is mentioned in explanations of machine behavior), this content is, according to Standard Theory, as irrelevant to what we do as is the value of coins to what a machine does. If you want to know what makes vending machines dispense cokes and candy bars, look to the intrinsic properties of the internal causes — the shape, size, and weight of the internal coins that trigger its responses. For the same reason, if you want to know what makes people do the things they do, look not

⁷ And a realist (i.e., not an eliminativist) about the mind.

to the relational properties of belief (those that constitute *what* we believe) but to the intrinsic (i.e., neurobiological) properties of the belief. Look to the “shape” and “size” — i.e., the syntax — of these internal “coins,” not their semantics.

This is a form of epiphenomenalism because although beliefs, on this view, turn out to be causally active (just as the coins deposited in vending machines are causally active), the properties of the internal cause that make it mental, the extrinsic properties that give it content (and thus make it into a belief) — are not relevant to the causal efficacy of the belief. Thus, the Standard View, while denying neither the reality nor causal efficacy of the mental, leaves little or no room for understanding the causal efficacy of the mental *qua* mental. Beliefs, *qua* beliefs, have as much effect on the behavior of persons as do quarters, *qua* quarters, on the behavior of vending machines.

4. SOLUTIONS.

Standard theorists are aware of this problem, of course, and they have adopted a variety of different strategies to neutralize its impact. Some (e.g., Campbell 1970; Stich 1978, 1983) simply accept the implication and try to live with it. Others (e.g., Burge 1986, 1989, 1993, 1995; Baker 1995) insist that it should be actual explanatory practice, not *a priori* metaphysical principles, that determines what is a causally relevant property. So if, in ordinary causal explanations of behavior, we invoke what is believed to explain what is done, then what is believed — content — *is* causally relevant to behavior and that is an end to the matter — metaphysical principles to the contrary be hanged. Still others (e.g., Fodor 1987, 1991), concede the irrelevance of extrinsic or broad content and look for a satisfactory substitute — an intrinsic content, narrow content. Or, like Davidson (1980), one takes comfort in the fact that beliefs are causes and refuses to worry about what it is about them that explains their effects (on Davidson’s theory it turns out to be the intrinsic physical properties of the belief — the ones that figure in strict laws). It is hard to see why some of these strategies (e.g., Fodor’s and Davidson’s) for vindicating the *explanatory* role of belief are not so much ways of solving the problem as (like the first) gritty ways of learning to live with (and talking around) it.

In a series of insightful articles, Jaegwon Kim (1984a, 1984b, 1987, 1989, 1990, 1991, 1993a) has explored the idea that mental causation is a form of supervenient causation (I denote supervenient

causation by “causation_s”). One macro event (increasing temperature of a fixed volume of gas, for instance) causes_s another macro event (an increase in the gas’s pressure) in virtue of the fact that both macro properties — temperature and pressure — supervene on causally related micro-states of the gas. Kim offers this as a model for the way mental states cause behavior: beliefs causally_s explain behavior by supervening on “micro” states of the believer (neurophysiological states) that cause bodily movements. If mental causation is really causation_s, if the mental really supervenes on the physical states of the body, then, he says, “mental causation does take place, but it is reducible to, or explainable by, the causal processes taking place at a more basic physical level.” (1984, p.107) If we assume that mental states supervene on biological states of the brain, then, Kim suggests (1984, p. 107), this (i.e., causation_s) would redeem the causal powers we attribute to mental states. Mental properties (the content of the propositional attitudes) would be as efficacious on this account as is temperature, pressure, heat, and a variety of other physical macro properties that derive their efficacy from the micro events on which they supervene.

This account of mental causation is plausible in the case of those mental states that supervene (or are thought by some philosophers to supervene⁸) on the intrinsic (biological) constitution of a person: e.g., pains, itches, tingles, sensations, and feelings. This, no doubt, is why Kim chose pain and the sensation of fear (1984, p. 106) to illustrate the theory.⁹ For such mental states — call them *phenomenal* states — there is a strong intuition (most philosophers seem to have it) that physically indiscernible individuals *must* be in the same phenomenal state. Even if physical twins (as a result of much different histories) might be having different beliefs — they must (or so the intuition dictates) be having the same sensations (pains, etc.). If this is, indeed, so, then there is no particular obstacle to supposing that phenomenal states

⁸ Not all philosophers think this. Some (including myself — see Dretske 1995) have a representational view of sensations that identifies experienced qualities (qualia) with representational properties. Thus, just like beliefs, the mental properties of sensations turn out to be extrinsic or relational properties of internal states: see Harman (1990), Lycan (1987, 1996), Tye (1994, 1995).

⁹ This is confirmed by his doubts a few pages later (p. 107) about whether the account of supervenient causation will work for *intentional* states — states (like belief) that have a propositional content.

derive their causal efficacy from the physical states on which they supervene. For phenomenal mental states might — who knows? — *strongly* supervene¹⁰ on the intrinsic physical properties of an individual, and it is *strong* supervenience (of the macro on the micro) that Kim requires (1989, p. 283; 1984, p. 104, 106) to support causal_s relations between macro events. Only if the macro supervenes on the micro in the way temperature supervenes on molecular motion or being water supervenes on being H₂O is it plausible to attribute the causal efficacy of the micro to the macro.

But however plausible supervenient causation may be as an account of the way *phenomenal* states bring about their effects, it does not seem to be available as an account of the way intentional states bring about their effects.¹¹ For intentional states, according to Standard Theory, do not, like phenomenal mental states, strongly supervene on the intrinsic biological properties of the person that occupies these states.¹² Intentional states, unlike phenomenal states, are relational states of an individual, and it is difficult to see how such relational properties could strongly supervene on an object's non-relational, its intrinsic, properties. Our monetary example illustrates the way, under ideal circumstances (no counterfeiting), a relational property might *weakly* supervene on intrinsic properties: as a matter of fact, thanks to the government's efforts, every piece of paper that has a particular set of intrinsic properties is a genuine \$20 bill. But weak supervenience is clearly not enough for supervenient causation. Even though the value of money supervenes on its shape and size, its being money does not

¹⁰ Unlike weak supervenience, strong supervenience requires that anything with a given base property *necessarily* has the property that supervenes on it. This requires, as Kim notes, some kind of nomological *dependence* between the supervenient property and those properties on which it supervenes. If water (in this world) is H₂O then, if being water strongly supervenes on being H₂O, then nothing *can* be H₂O without being water. Contrast this with being a \$20 bill. If pieces of paper of this size and shape (in this world) are \$20 bills, this does *not* imply that, things of that size and shape *in other possible worlds* are also worth \$20 (i.e., are *necessarily* worth \$20).

¹¹ Despite Kim's suggestion (1991, pp. 303ff) of supervenient causation as a possible replacement for my own theory (see Dretske 1988). My own theory is explicitly about the explanatory status of *intentional* (not phenomenal) states, and, as noted earlier, Kim himself seems dubious about whether supervenient causation applies to extrinsic mental states (beliefs and desires).

¹² Kim seems to agree with this; see 1987, p. 87.

share in the glory of causing what its having that shape and size causes. Temperature shares (in a derivative, but nonetheless real enough way) in the glory of causing whatever the events on which it supervenes causes, but the monetary value of a quarter does not share — not even derivatively — in the glory of causing what the coin's shape and size cause — e.g., an elliptical shadow in obliquely falling light. The value (being extrinsic) and the physical appearance (intrinsic) remain distinct attributes of the coin with different causal powers. To get supervenient causation we need strong supervenience, but what could it mean to suppose that the monetary value of a piece of paper or the value of a quarter was *necessarily* tied up with its having a particular shape, size and set of marks? This, it seems, could only be the case if the monetary value of the paper, its being a genuine \$20 bill, was not in fact relational at all but, rather, reducible to the paper's having just that set of intrinsic properties.¹³ This, though, is precisely what Standard Theory denies.

I do not think, therefore, that supervenient causation is a viable account of the causal powers of extrinsic mental states.¹⁴ If what I believe is a genuine relational property of me, then it might, in some local way¹⁵, weakly supervene on my intrinsic physical properties, but I do not see how it can display the kind of *dependence* on my intrinsic physical properties that would tempt us to say that *it* explains whatever the physical states on which it supervenes explains.

5. A BETTER SOLUTION.

We have, however, neglected an important aspect of the causal relations at work in both monetary-machine and mind-body cases. In the monetary-machine interaction, for instance, there is the fact that the

¹³ Kim (1989, p. 283) makes exactly this point — the point, namely, that strong supervenience — the kind necessary for supervenient causation — only occurs when there is a possibility of reduction of the macro properties to the micro. That is the basis of his argument that *non-reductive* materialists should derive no comfort from supervenient causation as a way to give the mental some causal punch in the material world.

¹⁴ Despite his suggestion (1991) that supervenient causation be considered a 'modified' version of my own theory (of belief and desire), I suspect Kim would agree with this.

¹⁵ Kim stresses the need to localize the supervenience (the supervenience base for your thoughts may not be the same as mine) in Kim 1991.

machines on which coins have a causal impact were designed and manufactured to be sensitive to objects having those intrinsic properties (S) on which monetary value supervenes, and they were made that way precisely *because* V supervenes on S. Business being what it is, machines that dispense commodities like cigarettes, food, and drink would not be designed to yield their contents to objects having S *unless* objects having S had V. Remove the fact of supervenience (as a result of widespread counterfeiting, say) and S-objects will soon lose their causal power. They will no longer produce the effects they now produce. They will lose their causal power because machines will no longer be built to respond to objects having S. The causal efficacy of intrinsic S (on machines — not to mention people) depends on the supervenience of extrinsic V on S. Let V supervene on a different set of properties, T, and T-objects will, quickly enough, assume the causal powers of S-objects.

This additional dimension to the causal story does not show that a vending machine's output is explained by the monetary value of the coins deposited in it. No, the cokes come rolling down the chute not because an object with a certain value is deposited in the machine, but because an object with a certain size and shape is. Nonetheless, if what we want to explain is not why a coke came sliding down the chute (the shape and size of the coins deposited will explain that), but why objects having the size and shape of nickels, dimes and quarters *cause* cokes to come rolling down the chute, why objects of that sort have effects of this sort, the answer lies, in part at least, in the fact that there is a reliable (enough) correlation between objects having that size and shape and their having a certain monetary value. It lies, in other words, in the fact that there is a supervenience (*weak* supervenience) of V on S. The value doesn't explain why the cokes come out, but it does explain why coins — objects of that size and shape — cause cokes to come out.

When we turn to the mind-body case, this dimension of the causal story is suggestive. If we think of ourselves as “vending machines” whose internal causal structure is designed, shaped and modified not, as with vending machines, by engineers, but, in the first instance, by evolution and, in the second, by learning, then we can say that although it is the “size” and “shape” (the syntax, as it were) of the internal causes that makes the body move the way it does (just as it is the size and shape of the coins that releases the cokes) it is, or may be, the fact that a certain extrinsic property supervenes on that neurological “size”

and “shape” that explains why internal events having these intrinsic properties have the effect on the body that they have. What explains why a certain neurological event in the visual cortex of a chicken — an event caused by the shadow of an overhead hawk — causes the chicken to cower and hide, is the fact that such neurological events have a significant (to chickens) extrinsic property — the property of normally being caused by predatory hawks. It is, or may be, possession of this extrinsic property — what the internal events *indicate* about external affairs — that explains why objects having those intrinsic properties cause what they do.

There is but a short step from here to the conclusion that it is the extrinsic, not the intrinsic, properties of internal events that causally explain behavior. All that is needed to execute this step is the premise that behavior is not the bodily movements that internal events cause, but the *causing* of these movements by internal events. All that is required, that is, is an appropriate distinction between the behavior that beliefs explain and the bodily movements that (in part) constitute that behavior. For if moving your arms and legs (behavior) is not the same as the movements of the arms and legs, but it is, rather, some internal event causing the arms and legs to move, then although the intrinsic properties of our internal “coins” will explain (via activation of muscles) the movements of our arms and legs, the extrinsic properties, properties having to do with what external conditions these internal events are correlated with, will explain why we move them.

This is not the place to amplify this account. I tried to do this in (Dretske 1988). The only point I want to make here is that the account I gave there of how reasons explain behavior depends on a correlation between the extrinsic (informational) and the intrinsic (biological) properties of reasons. It depends on weak supervenience of the extrinsic on the intrinsic. Without that supervenience, reasons cannot get their hand on the steering wheel. This is not because the extrinsic causally_s explains the movements of the body. No. That would require strong supervenience, and the relational properties underlying mental content do not strongly supervene on neurobiological properties any more than the value of coins strongly supervenes on their size and shape. It is rather because supervenience — *weak* supervenience — explains why the internal events that cause the body to move cause it to move the way it does. If I am right about behavior, that is exactly what we want beliefs to explain — viz., behavior, what a person *does*.

So to our series of opening questions, the answers are as follows: Yes, beliefs stand to human behavior in something like the way money stands to vending machine behavior. Does this show that what we believe is causally irrelevant to what we do. No, it does not show this anymore than it shows that the fact that nickels, dimes and quarters have monetary value is irrelevant to the behavior of vending machines. The fact that these coins have monetary value, the fact that they are a widely accepted medium of exchange, explains why the machines (are built to) dispense their contents when objects of this sort are placed in them. In this sense, the fact that these coins have monetary value explains why machines behave the way they do when the coins are in them. The same is true of belief: the extrinsic properties of these beliefs — *what it is* we believe — explains why we behave the way we do when these beliefs occur in us.

Department of Philosophy
Stanford University

REFERENCES

- Allen, C. (1995), "It Isn't What You Think: A New Idea About Intentional Causation", *Nous* 29:1, pp. 115-126.
- Baker, L. (1995), *Explaining Attitudes: A Practical Approach to the Mind*. Cambridge: Cambridge University Press.
- Burge, T. (1979), "Individualism and the mental", *Studies in Metaphysics (Midwest Studies in Philosophy 4)*, Peter A. French, Theodore E. Uehling, J., and Howard K. Wettstein, eds. Minneapolis: University of Minnesota Press, pp. 73-122.
- Burge, T. (1986), "Individualism and Psychology", *The Philosophical Review*, XCV, No. 1, pp. 3-45.
- Burge, T. (1989), "Individuation and Causation in Psychology", *The Pacific Philosophical Quarterly* 70, pp. 303-322.
- Burge, T. (1993), "Mind-Body Causation and Explanatory Practice", in *Mental Causation*, John Heil and Alfred Mele, eds. Clarendon Press, Oxford: pp. 97-120.
- Burge, T. (1995), "Reply: Intentional Properties and Causation", in *Philosophy of Psychology: Debates on Psychological Explanation*, Cynthia and Graham Macdonald, eds. Oxford: Blackwell, pp. 226-235.

- Campbell, K. (1970), *Body and Mind* (New York).
- Crane, T. ed. (1992), *The Contents of Experience*. Cambridge: Cambridge University Press.
- Dennett, D. (1987), *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- Dretske, F. (1988), *Explaining Behavior. Reasons in a World of Causes*. Cambridge, MA: MIT Press/A Bradford Book.
- Dretske, F. (1995), *Naturalizing the Mind*. Cambridge, MA: MIT Press/A Bradford Book.
- Fodor, J. (1987), *Psychosemantics*. Cambridge, Ma: MIT Press
- Harman, G. (1990), "The Intrinsic Quality of Experience", In Tomberlin (1990).
- Kim, J. (1982), "Psychophysical Supervenience", *Philosophical Studies* 41, pp. 51-70. Reprinted in Kim (1993b).
- Kim, J. (1984a), "Concepts of Supervenience", *Philosophy and Phenomenological Research*, 45, pp. 153-176. Reprinted in Kim (1993b).
- Kim, J. (1984b), "Epiphenomenal and Supervenient Causation", *Midwest Studies in Philosophy*, 9, pp. 257-70. Reprinted in Kim (1993b).
- Kim, J. (1987), "'Strong' and 'Global' Supervenience Revisited", *Philosophy and Phenomenological Research* 48, pp. 315-326. Reprinted in Kim (1993b).
- Kim, J. (1989), "The Myth of Non-reductive Materialism", *Proceedings and Addresses of the American Philosophical Association* 63, pp. 31-47. Reprinted in Kim (1993b).
- Kim, J. (1990), "Supervenience as a Philosophical Concept", *Metaphilosophy* 21, pp. 1-27. Reprinted in Kim (1993b).
- Kim, J. (1991), "Dretske on How Reasons Explain Behavior", in *Dretske and His Critics*, ed. Brian McLaughlin (Oxford: Basil Blackwell), pp. 52-72. Reprinted in Kim (1993b).
- Kim, J. (1993a), "The Nonreductivist's Troubles with Mental Causation", in *Mental Causation*, eds. John Heil and Alfred Mele. (Oxford: Oxford University Press), pp. 189-210. Reprinted in Kim (1993b).
- Kim, J. (1993b), *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- Lycan, W. G. (1987), *Consciousness*, Cambridge, Mass.: MIT Press/A Bradford Book.

- Lycan, W. G. (1990), "What is the 'Subjectivity' of the mental?", In Tomberlin (1990).
- Putnam, H. (1975), "The Meaning of "Meaning"", in *Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science*, vol 7, K. Gunderson, ed. Minneapolis: University of Minnesota Press.
- Stich, S. (1978), "Autonomous Psychology and the Belief-Desire Thesis", *The Monist* 61, pp. 573-91.
- Stich, S. (1983), *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press/A Bradford Book.
- Tomberlin, J. E., ed. (1990), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind, 1990*, Atascadero, CA: Ridgeview Publishing Company.
- Tye, M. (1992), "Visual qualia and visual content", in Crane 1992, pp. 158-176.
- Tye, M. (1994), "Qualia, content, and the inverted spectrum", *Nous*.
- Tye, M. (1995), *Ten Problems of Consciousness*. Cambridge, MA: MIT Press/A Bradford Book.

WHAT CAN THE SEMANTIC PROPERTIES OF
INNATE REPRESENTATIONS EXPLAIN?

I

Some of the things I do, I do for no reason: I inhale oxygen, I vomit, I cough, I hiccup, I perspire, I snore, and so on and so forth, if and when I do, for no reason. Other things I do, I do for reasons and some of my reasons for doing them are the contents of my beliefs and desires, i.e., my propositional attitudes. So I recently went to the travel agent closest to the place where I live because I wanted to reserve my flight from Paris to Amsterdam, I believed I could do so by going to a travel agent and I wanted to go as close to my place as possible.

This distinction may strike you as it strikes me as quite intuitive. It is the distinction between intentional voluntary behavior — the kind of things I do when I have reasons for doing them — and non-intentional instinctive behavior — the kind of things I do when I don't have reasons for doing them. Going to my travel agent was under the control of my beliefs and desires. In this sense, it is rational behavior. Salivating, blushing or shivering are not. Assuming that an individual's propositional attitudes are mental representations, his or her intentional voluntary behavior is behavior explainable by the contents of his or her representations. And the explanation I have in mind is causal explanation. An individual's non-intentional non-voluntary behavior on the other hand is not to be explained by what he or she wants or believes: it is not to be explained by any of the individual's mental representations. That much I will assume is uncontroversial.

Not all of an individual's mental representations, however, are propositional attitudes. I will assume that an individual's propositional attitudes are paradigmatic conceptual representations of properties and states of affairs in his or her environment. Like many non-human animals though, humans also entertain non-conceptual, sensory (or sensuous) representations of properties or states of affairs in their environments. Many philosophers have provided examples of mental sensory representations with non-conceptual content, particularly visual sensory representations. So for example, there certainly is something it's like to enjoy the visual experience of a square surface: the space is