

International Phenomenological Society

Knowledge in Action

Author(s): John Gibbons

Reviewed work(s):

Source: *Philosophy and Phenomenological Research*, Vol. 62, No. 3 (May, 2001), pp. 579-600

Published by: [International Phenomenological Society](#)

Stable URL: <http://www.jstor.org/stable/2653537>

Accessed: 23/03/2012 12:38

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



International Phenomenological Society is collaborating with JSTOR to digitize, preserve and extend access to *Philosophy and Phenomenological Research*.

<http://www.jstor.org>

Knowledge in Action

JOHN GIBBONS

New York University

This paper argues that the role of knowledge in the explanation and production of intentional action is as indispensable as the roles of belief and desire. If we are interested in explaining intentional actions rather than intentions or attempts, we need to make reference to more than the agent's beliefs and desires. It is easy to see how the truth of your beliefs, or perhaps, facts about a setting will be involved in the explanation of an action. If you believe you can stop your car by pressing a pedal, then, if your belief is true, you will stop. If it is false, you will not. By considering cases of unintentional actions, actions involving luck and cases of deviant causal chains, I show why knowledge is required. By looking at the notion of causal relevance, I argue that the connection between knowledge and action is causal and not merely conceptual.

“What knowledge adds to belief is not psychologically relevant.”¹

—Stephen Stich

A natural view about psychological or intentional explanation has it that whenever you do something intentionally, there is an explanation of that sort for your action under its intentional description. If Alice stops the car and thereby inadvertently spills the coffee on the dashboard, we can explain some features of the action but not others in terms of her mental states. She stopped because she saw that the light was red. This psychological explanation gives part of her reason for stopping, and a slightly more complete explanation might give more of the reason: she wanted to stop if the light was red in order to avoid an accident. In general, given the connection between doing something intentionally and doing it for a reason, there ought to be a reason-giving or psychological explanation for whatever you do intentionally. But, since Alice did not spill the coffee for a reason, there may be no psychological explanation of the action under that description.

Reason-giving or rationalizing explanations of intentional action serve two masters. On the one hand, they appear to be a species of causal explanation, and all the rules for causal explanation seem to apply. If I explain a fire in terms of a short circuit, I need not make explicit reference to any back-

¹ Stephen Stich, “Autonomous Psychology and the Belief-Desire Thesis,” *The Monist* 61 (1978) p. 574.

ground conditions, but I invite my audience to infer that conditions were normal, for whatever counts as normal in that context. If I explain Alice's stopping in terms of desire to avoid an accident, you assume that conditions are normal. Perhaps we can say something in general about what counts as normal conditions for reason-giving explanations by looking at what is presupposed in ordinary explanations of this kind.

On the other hand, reason-giving explanations justify or make sense of the relevant behavior from the agent's perspective. You might think there is a conflict here. Since reasons are conceptually connected to actions they cannot be causally connected to actions. We shall look at an updated version of this argument. But you might think that like the case of belief, the justification of action depends, at least in part, on how it was caused. Of course, not just any old causation will do. You need the notion of causally related in the appropriate way. If reason-giving explanations serve one master by serving the other, if they play their justifying role by pointing to the appropriate kind of causation, then an understanding of this appropriate way, a normally presupposed background condition, will help us understand not only the nature of intentional action but also the nature of intentional or reason-giving explanations.

An explanation of Alice's stopping in terms of a belief-desire pair is, in some sense, a more complete explanation of the action than one in terms of a desire alone. When we think about giving more and more complete explanations of intentional behavior, perhaps making explicit presupposed background conditions, it is clear that sooner or later, we will have to refer to more than just non-factive mental states like belief, desire, emotion, and intention. The more "world involving" the description of the action, the more obvious this is. We explain Alex's conquering the known world in terms of his deep-seated need to impress his philosophy teacher. But there must be more to it than that, even if we add to the explicitly mentioned pro-attitude the belief that by conquering he could impress. Unless we're only explaining attempts, there is always more to intentional action than beliefs and desires.

Part of what is involved in the more complete explanation of intentional action is knowledge. Alice knew that she could stop the car just by pressing a pedal. And Alex, though it may have taken him a while, discovered a way of conquering the world. We do not usually make explicit reference to knowledge in our ordinary explanations of action, just as we rarely make reference to both a belief and a desire and almost never to an intention. But, I will argue, the role of knowledge in the explanation and production of intentional action is as indispensable as the roles of belief and desire.

I think this means that knowledge is a kind of mental state. I shall only say a few words about this at the end, but this is part of what is at issue. Consider two individuals in different possible worlds with the same beliefs, desires, and intentions. Though they both believe that *p*, only one of them

knows that p. Perhaps one of them lives in a world in which it is false that p, or the justification for the relevant belief is insufficient for knowledge. Is this difference enough to show that the two individuals are in different kinds of mental states? I say that it is. Now this may seem like a somewhat non-standard view or a modest linguistic proposal to alter our use of the expression “mental state.” But the claim that knowledge is psychologically relevant is required by the natural view of psychological explanation, and our conception of psychological states is largely determined by the kinds of explanation in which they figure. So the psychological relevance of knowledge is already implicit in our ordinary understanding.²

No Way Out

A restriction of psychological explanantia to beliefs, desires, and intentions leads naturally to similar restriction to attempts as the only psychological explananda.³ If you want to predict whether or not Alice will stop the car, you need to know more than just what she intends. You need to know, for example, whether or not the brakes are working. If we assume that people typically succeed in their endeavors, a fairly dubious assumption in many cases, then we can use our knowledge of her mental states to predict that Alice will stop the car. But even given the prediction, it seems that our account of Alice’s stopping the car will be seriously incomplete if we ignore the automotive facts. The functioning of the brakes is as crucial to bringing it about that Alice stopped the car as are her beliefs and desires.

Given the intuitively plausible assumption that the automotive facts are not psychological facts along with the seemingly innocuous claim that a psychological account of Alice’s A-ing can make reference only to psychological facts, we seem stuck with the conclusion that there can be no complete psychological account of Alice’s stopping the car. We can understand the notion of completeness in terms of the provision of a *ceteris paribus* nomologically sufficient condition, or a provision of all of the factors causally relevant to the production of the effect. Since the automotive factors are causally relevant to Alice’s stopping the car, and the purely psychological account leaves them out, the account is incomplete.

If you are tempted to retreat inward at this point, be prepared for a long walk. We might try letting the psychologist explain the “more basic” action of depressing the brake pedal. But *depressing the brake pedal* has exactly the same problems as *stopping the car*. If the pedal is stuck in place, Alice will

² For very different arguments for a similar conclusion see Timothy Williamson, “Is Knowing a State of Mind?” *Mind* 104 (1995), and John McDowell, “Knowledge and the Internal,” *Philosophy and Phenomenological Research* 55 (1995).

³ William Alston argues for the restriction to attempts in “Conceptual Prolegomena to a Psychological Theory of Intentional Action,” in Brown, ed., *Philosophy of Psychology* (New York: Harper & Row, 1974), 71–101.

not depress it, regardless of her intentions. Even if we assume universal functioning of brake pedals, our account of her depressing it will be incomplete without reference to the automotive facts. The argument is general enough to deny the psychological relevance of all behavioral types individuated in terms of your effect on the environment.

The next step on the retreat inward is to give up on environmental effect types, and their intentional counterparts, and turn to bodily movement types. We might think that the most obvious problem with a psychological theory that restricted itself to bodily movements is that it would miss important generalizations. A folk psychologist may try to explain the success of an ad campaign by saying something like this. "Everybody buys the stuff because it makes them feel like an individual." Since buying is not only multiply realizable in terms of, but does not even supervene on bodily movements, a psychological theory restricted in this way would be unable to express, let alone explain, general facts about shopping behavior.

To insist that there are such generalizations within the purview of psychology is simply to insist that environmental effect types are psychologically relevant. The problem with bodily movement types from our perspective is that they face the same difficulties as environmental effect types. If the brake pedal is frozen in place, this purely automotive fact will affect the trajectory of Alice's leg. If someone surreptitiously glues the bottom of Alice's sneaker to the floor of the car, she will simply not move her leg in the specified manner in these circumstances.⁴

The final step on the inward retreat, of course, is to say that the primary explananda of purely psychological explanations are attempts. Regardless of the presence of glue, functioning brakes or brake pedals, Alice will *try* to stop the car. William Alston concludes that the primary explananda are intentions.⁵ Perhaps some may prefer acts of will. But the differences between these options need not concern us. All three options have the following two features. Willing to A, like trying to A and intending to A, entails having a representation of A-ing but does not entail that you succeed in A-ing.

The nature of psychological explananda is rarely explicitly discussed in the literature, but Stephen Stich in *From Folk Psychology To Cognitive Science*,⁶ is an exception. Since he thinks that psychological explanations can only make reference to properties that supervene on current, internal physical properties of organisms, Stich concludes that what gets explained

⁴ It is tempting to suppose that the "*ceteris paribus*" clause in the *ceteris paribus* laws will cover proper bodily functioning. This is why my examples point to obstacles rather than paralysis. The former show that the importance of an accurate representation of your environment is the same regardless of whether you are talking about the intended location of your car or the intended location of your arm.

⁵ Alston, p. 95.

⁶ Stich, *From Folk Psychology to Cognitive Science*, (Cambridge, Mass.: MIT Press, 1983).

must be restricted to behavior under an autonomous description. An *autonomous behavioral description* is a description of a way of acting such that if you would act in that way in a given setting, so would your replica, i.e., something that shares all of your current, internal, physical properties.⁷ While “performing one’s millionth weld” is not an autonomous description, “performing a weld” presumably is autonomous.⁸

Suppose we do restrict attention to autonomous behavioral descriptions. Can we explain someone’s performing a weld if we make reference only to properties and relations that supervene on the internal, physical properties of the welder at the time of the welding? Wilbur performs a weld. His replica, with a clogged acetylene torch, does not. Can we explain Wilbur’s welding completely in terms of properties he shares with someone who fails to perform a weld? Looking at the definition of an autonomous behavioral description, it seems that they apply partly in virtue of internal properties and partly in virtue of facts about the setting, facts that clearly do not supervene on the current, internal, physical properties of agents. Autonomous behavioral descriptions include a proper subset of descriptions of action that are defined in terms of your effects on the environment and share all of the same difficulties.

Anything But the Truth

The above considerations are based on the assumption that the truth of your beliefs is psychologically irrelevant. If the presence or absence of a feature of your mental states is systematically connected to different kinds of behavior, the feature is psychologically relevant. Wanting water is systematically connected to different kinds of behavior than wanting milk. Is wanting water systematically connected to different kinds of behavior from wanting twin-water?⁹ Well, that all depends on how you individuate behavior. If you’re talking about trying to get water, getting water, or intentionally getting water, they seem to lead to different sorts of behavior. This is a familiar argument from the days of narrow content.¹⁰ We’re now asking whether the

⁷ *Folk Psychology*, p. 167.

⁸ One difficulty for this view is determining what counts as a setting or the circumstances. If it is part of the setting that you have performed 999,999 welds, then “performing one’s millionth weld” is autonomous. If we restrict attention to current features of the external world in determining the setting, then no description of behavior that makes reference to artifacts, like “stopping a car” will count as autonomous. Something counts as an artifact of a certain sort only if it has a certain sort of history. It must have been produced for a certain purpose or at least be used for that purpose.

⁹ For twin-water, see Hilary Putnam, “The Meaning of ‘Meaning,’” in *Mind, Language, and Reality* (Cambridge: Cambridge University Press, 1975), pp. 215–71 and Tyler Burge, “Individualism and the Mental,” *Midwest Studies in Philosophy* vol. 4, eds. French et al. (Minneapolis: Minnesota University Press, 1977), pp. 73–121.

¹⁰ See for example, Tyler Burge, “Individualism and Psychology,” *Philosophical Review* 95 (1986), pp. 3–45, Jerry Fodor, *Psychosemantics* (Cambridge, Mass.: MIT Press, 1987),

truth values, rather than the contents of our beliefs make a systematic difference to the kinds of behavior we engage in. If you think that truth makes no difference, this entails that you will act the same way regardless of whether your beliefs are true or false. So the idea that truth is irrelevant brings with it some notion of acting in the same way, i.e., psychologically relevant behavioral types.

If the belief in question is that you can stop the car by pressing the pedal, then clearly, if your belief is true, you will stop the car. If your belief is false, you will not stop the car. If truth is psychologically irrelevant, so are intentional actions. If the belief in question is that your sneaker is not glued to the car or that there is nothing in the way of your leg, then the bodily movements you produce depend in obvious ways on the truth or falsity of your beliefs. If we assume that truth is irrelevant, the argument for restricting psychology to attempts is quite straightforward.

Why should we believe that truth is psychologically irrelevant? It is not *obvious* that Alice's stopping at a red light and her rolling into an intersection count as the same type of behavior from the psychologist's perspective. So it is unlikely that a belief in the psychological irrelevance of intentional action types figures as a premise in an argument for the irrelevance of truth. Presumably, it is the other way around. Perhaps the idea is something like this. Suppose that Alice stops the car because the light is red. Now we imagine a possible situation in which Alice, or her psychological duplicate with the same beliefs, desires, intentions, and so on is in a car with faulty brakes. Alice stops the car, but, through no fault of her own, her duplicate does not stop the car. The idea is that, given the psychological similarity, there must be some respect in which the two actions are the same. Furthermore, this is the only kind of respect that can interest the psychologist, since, intuitively, the behaviors differ for purely automotive reasons.

The duplication argument offers no support for the irrelevance of truth, since the argument presupposes the irrelevance. When we imagine a psychological duplicate of Alice, we duplicate certain features of her without duplicating others. We keep fixed Alice's beliefs, desires, and intentions without keeping fixed the truth values of her beliefs. This procedure tells us something about the psychological relevance of certain behavioral types only on the assumption that those truth values are psychologically irrelevant.

Of course, the truth of some of your beliefs is irrelevant to some of your behavior. The truth of your belief about the capital of Burkina Faso is, in many contexts, irrelevant to what you will have for lunch. With equal rele-

especially Chapter 2, and "A Modal Argument for Narrow Content," *Journal of Philosophy* 88 (1991), pp. 5–26. If Fodor's modal argument is sound, this is an *a priori* argument against restricting psychological explanations to attempts. Unfortunately, the argument is a version of the Logical Connection Argument run at the level of types, a version of which we'll look at later.

vance, we might point out that the truth of your belief that by A-ing you will B is irrelevant to whether or not you will A. Of course, the truth of that belief is relevant to whether or not you will B, and the truth of other beliefs is relevant to whether you will A, unless A-ing is an attempt type. So the typical examples presented in favor of the irrelevance of truth (you will still step on the brake pedal even if that will not in fact stop the car) simply ignore an equally large set of examples in favor of the relevance of truth (if you're wrong about the brake pedal, you will not stop the car).

Assuming that truth is psychologically irrelevant leads to the view that psychological explanations must be restricted to attempts. But there are two ways of understanding the restriction. It is consistent with one reading of this claim that there are explanations of someone's intentional action, under that description, possibly in terms of propositional attitudes like perception, knowledge, and desire. It is just that these explanations should not be called "psychological." This may sound like a purely linguistic dispute, but I shall return to this question at the end of the paper when I ask whether knowledge is a kind of mental state. For now, I'm interested in the nature of these explanations. Since these explanations, whatever we call them, must make reference to more than beliefs and desires, what else is involved, facts about the setting, the truth of the relevant beliefs, or knowledge? I shall argue for the third option.

On a stronger understanding of the claim that psychology is restricted to attempts, this is an expression of the idea that intentional action kinds cannot figure in any serious scientific explanations at all. Since psychology does explain attempts, the problem with these kinds is not that they're too mentalistic. They share that feature with attempts. The apparent problem with these kinds is that they involve interaction with the environment. So what goes for stopping the car and paying the rent also goes for feeding, fleeing, fighting, and all the rest. Maybe, in the case of animals without the conceptual resources to do these things intentionally, a merely biological explanation is possible. But once people start doing these things on purpose, they necessarily disappear from the perspective of science. I do not know what to say about such a view or even if anyone believes it. But I suppose that if I can show, at least in outline, how such explanations would work, I have defended their logical possibility. If such explanations are possible, we're back to the first way of understanding the claim that psychology is restricted to attempts.

The Truth Will Set You Free

When we give an ordinary explanation of intentional action, we may only explicitly mention a desire. But the explanation only works in the context of a number of implicit assumptions. Given the action and the desire, you do not need to make reference to the belief. Sometimes giving a more complete

explanation is simply a matter of making explicit the assumptions that were already implicitly involved in the original explanation. A rough and ready test for when this is going on is fairly simple. If you can figure it out from the armchair, it was probably already involved in the explanation. If you explain Alice's stopping the car in terms of her desire to avoid an accident, you do not need to add the fact that her beliefs about how to stop were true. But that does not mean that the assumption about truth was not already implicitly involved in the original explanation.

Any explanation of an intentional action, rather than an attempt, depends on the truth of the relevant beliefs in this way. Consider a set of means-ends beliefs, or beliefs that by A-ing you will B. Restricting attention to beliefs of this sort that play a role in the production or guidance of behavior, it seems that the truth of just these beliefs is relevant to what types of action you will perform. A belief, no matter how externally individuated, can always be a mistake. And when you act on a false belief, you will not, except by accident, do what you intend. If your psychological explanations make reference to what you believe as well as the truth of those beliefs, this amounts to the claim that truth is psychologically relevant.

As long as we focus on truth, the following may seem like an alternative. Rather than talking about a belief being true, we should simply talk about the facts that make the belief true. Or maybe psychological laws and explanations make reference both to non-factive mental states and to settings or contexts. If a setting is just the set of facts that make your means-ends beliefs true, this sounds to me like another way of putting the view that truth is psychologically relevant without coming out and saying it.

If you do not think of settings simply as whatever it takes to make the relevant beliefs true, you're faced with the following methodological difficulty. Suppose psychological laws have something like the following form:

(PL) If you are in mental state M and setting S, you will perform action A.

Discovering such a law would require filling in for the "M," the "S," and the "A." Suppose we replace "A" with the description "performing a weld" and "M" with your favorite description of whatever internal states are sufficient, or at least close to sufficient for trying to perform a weld. What do you need to know to know what goes in for "S"? You need to know how to weld. You need to know what can go wrong and what must go right in order for the weld to be successful. Considering the variety of intentional action descriptions, including "stopping a car," "fixing a lock," and "splitting an atom," the study of psychology becomes the study of practically everything.

One of the virtues of making reference to truth is its generality. If you say that someone's beliefs about how to fix a lock were true, you do not need to know how to fix a lock to explain the behavior. If you say that Alice's beliefs about how to stop the car are true, you need not make explicit reference to any of the particular facts that would keep her from performing the action by making her belief false. The failure of the brakes, the presence of glue on her sneaker, the end of the world as we know it, or any of a number of other conditions could be incompatible with Alice's stopping the car. The only thing all of these conditions have in common is that they are incompatible with the truth of Alice's means-end beliefs. If we say that the beliefs are true, we rule out all of these conditions without having to list them.

Knowledge is Power

The real problem with an explanation of action in terms of attempts and the facts is that belief-desire pairs plus the facts that make the beliefs true do not add up to intentional action. Explanations of action do not just rely on the attribution of true belief. They rely on attributions of knowledge. The natural view says that we can give a psychological explanation of your A-ing whenever you intentionally A. But you can act on a true belief about how to A and still A by accident. You cannot give the same explanation of one person's intentionally A-ing that you give of another person's unintentionally A-ing. Consider a case of a deviant causal chain.¹¹ A climber wants to rid himself of the danger of holding a rope attached to another climber and believes that letting go will rid him of the danger. The belief and desire so unnerve him that he unintentionally lets go. In this situation, there is something clearly wrong with the following explanation: he let go of the rope because he wanted to rid himself of the danger. "Because" is stronger than the inverse of "caused." While the climber's letting go of the rope may not have been an action (it was not caused in the right way by beliefs and desires), the description of the behavior still applies. If asked why he let go of the rope, he will not say, "I did not do it" but rather "because I got nervous." The same notion of causally related in the appropriate way, whatever that is, is involved in the idea of behavior being intentional and in the "because" of explanation. So what you get, either implicitly or explicitly, in an explanation of action lets you know that the action was intentional. Just assuming that their beliefs were true will not let you know the action was intentional. You need to assume that they knew what they were doing.

Examples of lucky action and examples of deviant causal chains both show the importance of knowledge to intentional action. For luck, consider the following three lottery stories. In the first, Cindy buys a lottery ticket,

¹¹ Donald Davidson, "Freedom to Act," p. 79 in *Essays on Actions and Events*, (Oxford: Oxford University Press, 1980).

knowing her chances are a million to one, and she wins. In this case, with a fair lottery, Cindy's winning is too accidental, or too lucky to count as intentional. In the second case, Cindy mistakenly believes someone rigged the lottery in her favor. She believes, on the basis of this, that if she buys a ticket, she will win. She buys the ticket and wins. So her belief about winning is true. She even has a justified true belief. But her winning is just as accidental and just as lucky as it was in the first case. So her winning is not intentional. What is missing? It looks like knowledge. In the third case, Cindy knows the lottery is rigged and knows that if she buys a ticket she will win. In this case, Cindy intentionally wins the lottery.

Examples of deviant causal chains suggest that there is a gap between doing something you intend to do and doing something intentionally. In some of these cases, the causal deviance occurs after you initiate the action but before you get what you want. For example, you're driving down the road and see an accident up ahead. You intend to stop the car by pressing on the break pedal and turning the wheel hard to the right. This intention causes you to move your leg. Unfortunately, you hit the gas pedal instead of the break, and this, along with your turning the wheel, causes you to crash into the guardrail. You stopped the car. You did what you intended and tried to do, but you did not do so intentionally. The intention even caused you to do what you intended, but it did not cause it in the right way. The problem, of course, is to say what that right way is.

There are two fairly intuitive things you can say about this story. On the one hand, you got what you wanted, but you did not do it in the way that you intended to do it. On the other hand, your getting what you wanted was just an accident. Your stopping the car was not really under your control. We can say both of these things about Cindy's second lottery story where she mistakenly believes that the lottery is rigged in her favor. Whether or not this is a case of a deviant causal chain, I do think that we should treat Cindy's case the same way we treat the car accident case, and both of our fairly intuitive judgments allow us to do so. Cindy got what she wanted, but she did not do it in the way she intended. She intended to win by having the lottery rigged in her favor, but this is not what happens. Also, her winning was just an accident. It was not, despite what she believed, under her control.

Here is one way of making our first fairly intuitive judgment a little more precise.¹² Think of the content of an intention as an action plan. Action plans can be more or less complete. You may intend to go to the movies some

¹² For versions of this strategy see Myles Brand, *Intending and Acting* (Cambridge, Mass.: MIT Press, 1984), Gilbert Harman, "Practical Reasoning," *Review of Metaphysics* 79 (1976), pp. 431–63, and Alfred Mele, *Springs of Action* (Oxford: Oxford University Press, 1992), especially chapter 11. For more on action plans see Michael Bratman, *Intention, Plans, and Practical Reason* (Cambridge, Mass.: Harvard University Press, 1987).

time this month without any idea about what movie, when, or where. Or you may intend to go to the movies tonight, and your plan is to go to the theater around the corner, see what is playing, and buy a ticket for something that looks interesting and starts fairly soon. Or you might have the entire evening planned out: first buy the tickets, then go to dinner at your favorite restaurant, then walk diagonally across the park, etc. These are all action plans of varying degrees of specificity. What you believe determines, at least in part, how specific or complete your action plan is. If you do not yet have a view about how you will get to the movies, that part of the plan is left open.

If we use our first intuitive judgment to understand why your stopping the car and Cindy's winning the lottery are not intentional, we need to look at the whole action plan rather than focusing only on the final goal. Neither of these actions are intentional because the facts do not fit the content of the intention. The general strategy is to say that if the facts do fit the content of the intention, if things go according to plan, then this kind of causal deviance will be ruled out.¹³ But here is the problem, and here is how our first intuitive judgment differs from the second. No matter how complete an ordinary person's action plan may be, the facts will always be more fine-grained than the plan. So you intend that A will lead to B and B to C. Since the specificity of the content of this attitude is limited by your beliefs and conceptual resources, there will come a point where you just do not know how A leads to B. Once we get to this point, that is as specific as the plan can get. Still, there will be some causal process or another leading from A to B, and if we imagine a deviant causal chain here, it will just be an accident that you get what you want. But the facts will fit the content of the intention. So a fit between the facts and the content is not sufficient to rule out this kind of causal deviance. Perhaps an example will help.

Bobby intends to kill his uncle by planting a bomb in his house and then, after moving a safe distance away, pressing the large red button on the remote control device. He does not know much about how these things work. He thinks that pressing the button will cause the bomb to detonate but has no idea about the details of this process. His belief is true and, we can suppose, justified. But here is what happens. A satellite, launched by the National Security Agency and designed to prevent bombings of just this kind, intercepts Bobby's transmission; this causes the satellite to send a warning to the intended victim; but, because of an unfortunate choice of frequency, this causes the bomb to detonate. Bobby killed his uncle and caused the bomb to detonate, but he did not do either of these things intentionally.

We cannot explain this case of causal deviance by saying that though Bobby got what he wanted, he did not do it the way he intended. If Bobby had

¹³ Of course the intention must cause the action. The notion of fit between the facts and the content is used to understand causing *in the right way*.

a view about what would happen after he pressed the button, it probably would not involve NSA satellites. Under those circumstances, we could explain why Bobby's action was unintentional by saying that things did not go according to plan. Assuming that the content of an intention is an action plan, this strategy explains causal deviance in terms of a failure of fit between the facts and the content of the intention.

In the case as described, Bobby's ignorance about the details means that his plan is quite simple: press the button and blow up the house. In this respect, Bobby is like many of us who call people on the phone, play CDs, and stop cars without having any idea about how these things work. The simplicity of the plan means that we cannot use the common strategy. In this case, things did go according to plan. What Bobby lacks is knowledge. Bobby believes that he can blow up the house by pressing the button. He is right, but this is not a case of knowledge. One thing makes him justified—information about the reliability of the device—while something very different makes his belief true. This is like the original Gettier examples.¹⁴ Information about Nogot justifies the belief, but the facts about Havit make it true. Now we can imagine a case where the trick with the satellite becomes the standard way of blowing people up so we have the right connection between the justifier and the truth maker. In that case the belief counts as knowledge, and the action is intentional.

Knowing how is more than just being able. You are able to do something if you have a plan of action that will work. Bobby's action plan works, but it is just an accident that it does. To the extent that *knowing that* is something like having a nonaccidentally true belief, *knowing how* is something like having a nonaccidentally effective action plan. I do not intend either of these as a serious analysis, but the kinds of things that can go wrong with a true belief that keep you from *knowing that* can also go wrong with action plans and keep you from *knowing how*. And when things do go wrong in this way, and you act on that plan, even if you do what you intended, you will not do so intentionally.

Our notion of intentional action involves our notion of control. In a fair lottery, you have no control over the outcome. Bobby is not in control of his bomb. But in ordinary circumstances, you do have control over where your car goes and when it stops. This is why cars are such useful devices. To say that someone has control over a chunk of the world or some feature of that chunk is to attribute to them a worldly mental state. Control requires awareness, so it clearly involves the mental. But this mental state is far more worldly than a mere externally individuated belief. It is not sufficient for being in control of the car that you were causally related in the appropriate way to cars at some time or that you interact with the automotive experts.

¹⁴ Edmund Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23 (1963), pp. 121–23.

There has to be a car. In this respect, control resembles knowledge and perception with their specific entailments about the external world.

Of course, you can be pushed into thinking that in the strict and philosophical sense, we really only have control over what goes on in our own minds. All you can do is try. The considerations that push in this direction are essentially the same as those that push toward Cartesian scepticism. No matter how recently you've had the brakes checked, there is always the possibility that they will fail. If the mere logical possibility of failure is inconsistent with the strict and philosophical sense of control, then we are talking about complete control. It is one philosophical mistake to assimilate control to complete control, a mistake with an obvious analogue in epistemology. It is another to suppose we have complete control over our own minds. Perhaps for some, weakness of the will is an interesting theoretical possibility. But for many of us, it is a daily occurrence. And it is often easier to make a cup of coffee than it is to stop thinking about what happened last night. Talk of intentional action presupposes a certain degree of control on the part of the agent. Control, like perception, requires the right kind of connection between the agent and the facts. An essential ingredient in this kind of connection is knowledge.

Causal Relevance

Typical explanations of intentional action only explicitly make reference to a belief or a desire. But the real explanation makes reference to both. To the extent that the real explanation is determined both by what is implicit and explicit, explanations of intentional behavior make reference to truth and knowledge. If your means-end belief is false, you will not do what you want. If your means-end belief does not constitute knowledge, you will not do it intentionally. The role knowledge plays in the explanation of behavior is as indispensable as the roles of belief and desire.

To say that knowledge is as indispensable as belief and desire is not quite to say that it is indispensable. When you know that *p*, this fact about you will be determined by some set of specific facts about your beliefs, experiences, and your relation to the world. Perhaps a serious scientific explanation should give up on the general facts, like knowing that *p*, and restrict attention to the specific facts upon which they depend. To use the jargon, serious explanations make reference to subvening properties not supervening properties.¹⁵

Of course, if you believe that *p*, this will also be determined by some set of specific facts. If subveners are always better explainers, belief is in danger of elimination. But there is an at least apparent difficulty with the view that

¹⁵ See Jaegwon Kim, "Mechanism, Purpose, and Explanatory Exclusion" in *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993).

subveners are always better explainers. Explanations by reference to the general facts will cover cases that, in some sense, ought to be treated alike. This difficulty may not be insurmountable. Since I am only arguing for equal rights for knowledge, I am only concerned to show that the considerations that cause trouble for an elimination of belief from our explanatory scheme make a similar elimination of knowledge equally problematic.

So I think knowledge plays a causal role in the production of action. But consider the following alternative. It is not knowing that *p* that plays a causal role. The belief that *p*, the justification for the belief, the fact that *p*, and whatever other facts are needed for the belief to constitute knowledge, all of these play a causal role in action. Your knowledge that *p* is an epiphenomenon that, in some sense, results from these more basic, causally relevant features. Once you've made reference to these more basic features, there is no further work for knowledge to do.

There is a problem with the view that mental facts and external facts always figure independently in psychological explanations of behavior. When the general facts are the same in two cases but the specific facts differ, only an explanation in terms of the general facts will capture the relevant similarity. On the view I'm opposing, facts about beliefs and desires are causally relevant with respect to intentional behavior. And facts about settings or the external world are causally relevant. Any talk about knowledge or true belief, on this view, is at best a mere shorthand description of these more basic causally relevant facts. Truth or knowledge is not an independent causal power. To show that there are cases where truth or knowledge is the causally relevant feature, we need to understand the notion of causal relevance presupposed. To keep the issues focused, consider the following pair of examples.

Marcia stopped at a red light because she wanted to avoid getting a ticket and thought that stopping would help her avoid getting one. Greg also stopped at a red light because he wanted to avoid a ticket and thought that that would help him avoid one. Greg and Marcia act in the same way and do so for the same reasons. Of course, they do not only act in one way. Marcia is riding her cousin's Harley and stops at the red light by squeezing a lever with her right hand. She also, without realizing it, casts a shadow in my direction. Greg is driving the family station wagon, and he stops by pressing a pedal with his foot. The two events are similar in important respects and different in important respects.

According to the natural view of psychological explanation with which we began, the important respects in which behaviors may be similar and different are the intentional respects. If we type behavior purely in terms of the end results, Marcia's stopping at a red light is psychologically on a par with her casting a shadow in my direction. But we can explain the similarity between Greg and Marcia's behavior, their both stopping at red lights, in terms of similarities in their mental states. We cannot, however, explain the differ-

ence, her casting a shadow but his failure to do so, in terms of some difference in their mental states.

So we know what the psychologically relevant effect types are: types of the sort *intentionally A-ing*. Our present concern is with the question of what sorts of properties are causally relevant with respect to effects of this sort. On our picture of explanation, there will often be causally relevant features that do not figure explicitly in ordinary explanations. We can explain the fire in terms of the short circuit without mentioning the presence of oxygen. But we may still believe that the presence of oxygen was causally relevant to the production of the effect. If there had not been any oxygen, there would not have been a fire. The truth of such counterfactuals is usually taken as a test for causal relevance, where causal relevance is supposed to be a relation between a property of the cause and a property of the effect. Unfortunately, a precise analysis of this relation has so far eluded our grasp.¹⁶ There are even those who are sceptical about the existence of such a relation.

Without some notion of causal relevance, a relation between properties and not merely events, it is very difficult to understand what Mill's methods are supposed to be methods for. To apply the method of difference, you construct or consider a situation in which you have A, B, and C, and you check the effects. Then you consider a similar situation in which you have B and C without A and check the effects. What could "A," "B," and "C" possibly refer to in these instructions? If the instructions are possible to follow, they cannot refer to concrete, particular, nonrepeatable events. The expressions refer to properties, and the test determines the causal relevance of repeatable features.

In addition to scientific practice, there are familiar examples that show the intuitive difference between causally relevant and causally irrelevant features. With your indulgence, I will add my personal favorite to the list. When you play pool in a bar rather than a pool hall, you have to pay for each game. With one exception, when you pocket a ball, the ball remains inside the pool table until you put in more money. The exception, of course, is the white ball. Certain regularities and corresponding counterfactuals hold for a wide range of pool tables of various sizes, constructions, and designs. If the ball that goes in is white, it will come back out; if it is not, it will not.

We can think of the regularities and counterfactuals in terms of the color of the ball. This, by itself does not convince us that color is causally relevant. If there is a color video camera inside the pool table that plays an

¹⁶ See, for example, LePore and Loewer, "Mind Matters," *Journal of Philosophy* 84 (1987), pp. 630–41; Fodor, "Making Mind Matter More," *Philosophical Topics* 17, (1989), pp. 59–79; and Block, "Can the Mind Change the World?" in *Meaning and Method: Essays in Honor of Hilary Putnam* edited by George Boolos (Cambridge: Cambridge University Press, 1990).

important role in the release of the white ball, then color is causally relevant. But we do not think that this is how it works. How does it work?

Suppose that there is some spring loaded mechanism inside the table, and when the cue ball rolls over it, it is heavy enough or massive enough to trigger the mechanism, while the other balls are not. How do we find out whether the causally relevant feature is mass or weight? Are there any facts that would make it the case one way or the other? Well, if you can consider or construct a situation in which the mass is the same but the weight is different, and if, in that situation, the relevant effects are different, this would suggest that weight rather than mass is the causally relevant property. But of course we can consider such a situation: take the pool table to the moon. On the moon, none of the balls are heavy enough to trigger the mechanism. So you get a relevantly different effect. The cue ball is not treated differently from the others. So it is the weight of the cue ball, rather than its mass that is causally relevant. This looks like a case where a relational property rather than some intrinsic correlate of that property is “doing the causal work.”

So we have a conception, though not an analysis, of this relation of causal relevance. And we want to know what features of Marcia are causally relevant with respect to her intentionally stopping at a red light. We know that claims about truth and knowledge pass the counterfactual test. If she did not know about the brake pedal or if her belief about the brake pedal had been false, she would not have intentionally stopped at the red light. But there is more to causal relevance than counterfactuals. Though I think the real causal power is knowledge, to make things easier, let us begin with truth.

On the view that mental facts and external facts figure independently in psychological explanations of behavior (i.e., on the view that the antecedents of psychological laws say something like “If you are in mental state M and setting S...”), we do not point to the single fact that Marcia’s belief is true. We point to the fact that Marcia believes that p and the further fact that p.¹⁷ So on this view, a slightly fuller explanation of Marcia’s stopping would include facts like these:

- (1) Marcia wanted to stop at a red light.
- (2) Marcia believes that if she squeezes the lever she will stop.
- (3) If Marcia squeezes the lever she will stop.

Of course there will be other facts involved in a complete explanation, but they will all be facts of the same three sorts: facts about Marcia’s desires,

¹⁷ The inspiration for this view comes from discussions of the deflationary conception of truth. See, for example, Paul Horwich, *Truth* (Oxford: Basil Blackwell, 1990).

facts about her beliefs, and facts about the world. There is no need to mention further facts that involve the notion of truth.

I say you have to talk about the truth of Marcia's belief about how to stop. They say you only have to talk about Marcia's belief that by squeezing the lever she will stop, and the fact that by squeezing the lever she will stop. There is supposed to be a difference here? Yes. Our conception of causal relevance is intimately connected to Mill's methods. You use the methods to find out about causal relevance. Mill's methods are about looking at a wide variety of situations that are similar and different in certain important respects and seeing in what other ways those situations are similar and different. Our notion of a causally relevant property is a property of a cause that is responsible for a property of an effect in a wide range of situations. To determine the causal relevance of a feature, you need to look at more than just one case.

So let us look at another case. Greg, you remember, stopped at a red light. He, like Marcia, stopped in order to avoid traffic tickets. There are some interesting similarities between the two cases. Both causes include desires to avoid tickets, and both effects are of the type, *intentionally stopping at a red light*. Are there any further similarities? It seems so. They both stop partly because they both have true beliefs about how to stop, and they both know how to stop. This sort of similarity in the causes is missed by any explanation that makes reference to any particular belief about how to stop, since they have different beliefs about how to stop. He believes you stop by pressing a pedal, and she believes you stop by squeezing a lever. Any explanation that makes reference to facts about pedals will only cover one case and not the other.

This is not an isolated instance. Any explanation of intentional action will presuppose knowledge and control on the part of the agent. If you're tempted to reject the request for an explanation that covers both of these cases, you should do so whenever there is a difference in how two people do the same thing. Aside from science fiction stories, any time two people do the same thing for the same reasons there will always be some difference in how they do it. If Greg and Marcia are in different model cars, there will probably be some difference in how they stop. She needs to press more firmly on the brake pedal or move her leg in a slightly different direction. Give up on stopping at a red light, something you can do in a car or on a motorcycle, and you begin the same inward retreat we have already seen. Once we remember that knowledge rather than true belief plays the causal role, on the view that restricts attention to specific facts, any difference in the justification for the means-end beliefs requires a difference in the explanations.

I am not arguing for reform. Explanations of intentional action already rely on implicit assumptions about the agent's knowledge. I'm not suggesting that we change anything, only that we face the facts. If you want any

psychological explanation to cover more than one actual case, rather than restricting attention to intrinsic duplicates in identical settings, something that only happens in philosophical stories, you need to keep talking about knowledge. Of course, there could also be other explanations that treat the two cases differently. Perhaps a neurological/mechanical explanation will satisfy our desire for the details. Or maybe we need to go all the way down to physics for that. But to the extent that generality is also a virtue in explanation, the psychological, and hence epistemological explanation is necessary for that.

Many people have the intuition that if two individuals are the same on the inside, they must be the same in all psychological or mental respects. This intuition is not restricted to philosophers. This intuition may influence the kinds of questions some psychologists ask and the kinds of theories that interest them. But the fact that scientists share this intuition does not elevate it to the status of a scientific fact. Now I think we should, in the end, reject this intuition. This does not mean that any of the theories based on the intuition are false. Assuming that the distinction between the inside and the outside makes sense, this may be an important distinction in nature to be studied by empirical methods. The question is whether this is the only important distinction in the neighborhood.

I have tried to call this intuition into question by presenting a clash of intuitions. Most of us think that the distinction between intentional and unintentional action is a real distinction in nature. If one person does something on purpose and another does the same thing but just gets lucky or does it by accident, we think it is more likely that the first person will do that sort of thing again in the future. "Luck" and "accident" are, almost by definition, non-projectable. To call it lucky or accidental is to say that you should not count on it.

To treat intentional action as a natural kind is to take seriously questions about what instances of that kind have in common as well as what the causes and effects of those instances have in common. If you restrict attention to what is inside, you will not find this. To find it, you need to look at the connection between the inside and the outside, the connection embodied in knowledge, perception, and control. If you do take this connection seriously, you need to relax your grip on the importance of the distinction between the inner and the outer. You can, presumably, hold onto the internalist intuition come what may, and so reject altogether the importance of the distinction between intentional and unintentional action. Whether scientists do this or philosophers do this, it looks like rejecting one set of intuitions in favor of another. I recommend the other option. The difference between the two views as I see it, is the difference between a restrictive and a non-restrictive view. Though I think there is more to your mind than what is inside your head, I do not think that what is inside is irrelevant.

Conceptual Connections

Since intentional action requires control, and control is a clearly relational notion, something like being in the right relation to the facts, we need a clearly relational notion like knowledge to explain intentional action. But maybe the connection between knowledge and intentional action is merely conceptual and not causal. There is some sort of conceptual connection between knowing how to A and intentionally A-ing. In the case of reasoning, one hopes that there is a conceptual connection between the contents of the reasons and the content of the conclusion. But we do not conclude that reasoning is not a causal process. Not every conceptual connection is inconsistent with a causal connection. So we need to look more closely at the kinds of causal and conceptual connections involved.

Given my commitment to causal relevance, I cannot take the familiar Davidsonian line on the logical connection argument. According to Davidson,¹⁸ only tokens are causally related; only types are conceptually related; and a conceptual connection between types is consistent with a causal connection between tokens. Since I think that causal relevance is a kind of causal relation involving types, I cannot rely on this to block the move from the presence of a conceptual connection to the absence of a causal connection. I think the type *knowing how to A* is both causally and conceptually connected to the type *intentionally A-ing*. How could this be?

First of all, wanting to A, intending to A, or if you like belief-desire pairs, wanting to B and believing that you can B by A-ing are all as conceptually connected to intentionally A-ing as knowing how to A. In this respect, knowledge is on a par with beliefs, desires, and intentions. To the extent that I'm only arguing for parity, this should be enough. But to see the parity, let us look at two specific kinds of conceptual connection between kinds of mental states and the kinds of action they cause.

The most obvious conceptual connection between the mental kinds and the action kinds is a similarity or identity in their intentional properties. Looking for and wanting water, like seeing that and knowing that there is water in front of you, are all directed at water in some sense. The kind of intentionality characteristic of belief and desire also applies to knowledge, perception, and action. In terms of a conceptual connection between the intentional properties of the causes and effects, intentional action is on a par with reasoning. If reasoning itself is an action or if the conclusion of practical reasoning is an action, this may be more than an analogy. But the analogy will do. If you come to believe that q on the basis of your beliefs that p and that if p then q, we have the same kind of conceptual connection between contents here that we have in the case of action. Even in cases that you might

¹⁸ Donald Davidson, "Actions, Reasons, and Causes," in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980).

not call reasoning, for example, a case where an appearance that *p* causes a belief that *p*, we have the same pattern. And when you say that you believe it because it looks that way to you, there is a strong presumption that that “because” is causal.

There may be another kind of conceptual connection between knowledge and action. It may be that a logically necessary condition for intentionally A-ing is that you know how to A. It may also be a logically necessary condition on intentionally A-ing that you have some pro-attitude toward A-ing. You need not see it as an end in itself or as a straightforward means to an end. It may be an inevitable consequence of something you want to do. So whatever else you have against it, it has at least that much going for it. I do not think this stronger conceptual connection entails that desires are causally irrelevant with respect to action.

We need to distinguish the argument from conceptual connections from a more general argument from uninformativeness. In most contexts, an explanation of someone’s A-ing simply in terms of their knowing how to A will not be very informative. But informativeness is not a very good indicator of causal relevance. In most contexts, an explanation of a fire simply in terms of the presence of oxygen will not be very informative. Of course, you can imagine contexts in which such an explanation is informative. The same goes for knowledge. Why did Jan, rather than Peter, finish the puzzle first? Because she knew how to do it. The same goes for intentions. Telling you that he did it because he intended to might not tell you nothing, but it does not tell you much. This is why knowledge and intention are typically implicit in the explanation of behavior. It is not that they play no causal role. It is that the causal role they play is so common and pervasive that for the practical purposes of explanation, they are relegated to the realm of background conditions. But I assume that the distinction between causes and background conditions is pragmatic, not metaphysical.

On the other hand, if the knowledge attributed is that you can A by B-ing, this is not a logically necessary condition for intentionally A-ing, since there may be other possible ways of A-ing. It is also, in many contexts, more informative, since it tells you how the agent went about A-ing. Again, we have the analogy with intentions. We usually say that you intend to go to the movies, and that is the end of the description. But at least in many cases, the content of the intention will include a plan specifying, perhaps, what movie, where and when, and a way of getting there. Perhaps in some cases, a more complete description of the intention will give you the information that interests you. But the intention, like the knowledge embedded in it if it is to be successful, does play a causal role in the production and guidance of the behavior, whether anyone is interested or not.

Philosophers of mind tend to think of the explanation of action in terms of belief-desire pairs. Since they require both a belief and a desire, they need

to dig just slightly below the surface of explicit explanations. Action theorists tend to think of the production and guidance of action in terms of action plans, the contents of intentions, and this requires a little more digging. But whoever we are, if we're interested in intentional action rather than attempts, we need to dig all the way down to knowledge.

Knowledge and the Mental

Knowledge plays a systematic role in the explanation and production of that paradigm of mentality, intentional action. So knowledge is a kind of mental state. But what difference does it make if we call knowledge mental? Why is this not a merely terminological dispute? The question of whether or not knowledge is a mental state raises the question of what it is to be mental. If there are philosophical questions that are not merely verbal, I think this is one of them.

One way to start thinking about what it is to be mental is to begin with some plausible examples and some plausible general principles and go from there. As for examples, I like the list embedded in the first sentence of "Mental Events." Davidson's list is "perceivings, rememberings, decisions, and actions."¹⁹ Of course, I would like this list since three out of four are factive, but there is at least some initial plausibility to including the items on this list. Of course, we need to add to this things like beliefs, desires, itches, and tickles. As for principles, some are more relevant to our concerns than others. A plausible first guess might be that something is mental only if it has qualitative or intentional properties. But this does not distinguish knowledge and perception from their non-factive cousins.

A more relevant, though I think more problematic principle is epistemic. You can state this in terms of infallibility or indubitability, but let us stick with the weaker, more common version in terms of introspection. M is a kind of mental state only if you can tell through introspection whether or not you're in M. This is what is funny about calling knowledge a kind of mental state. It violates principles like this. There need not be any introspectable difference between knowledge and mere belief. But if your list of examples also includes unconscious mental states, whether of the Freudian or cognitive-science variety, this should call into question your commitment to epistemic principles like this. If you generally distinguish questions about our epistemic access to a thing from questions about what the thing is, you should be suspicious of epistemic principles about what it is to be mental. Unconscious mental states count as mental states because they play a systematic role in the explanation and production of behavior, and they do so

¹⁹ Davidson, "Mental Events," p. 207, in *Essays on Actions and Events* (Oxford: Oxford University Press, 1980).

in virtue of their content. Knowledge counts as a kind of mental state for the same reason.

There are, finally, our notions of the internal and the external. In discussions of wide content, the notion of the internal is fairly literal. A mental state is internal when it supervenes on what is literally inside your head. But the significance of the distinction between what does and what does not supervene on the neurological remains unclear. Not all neurological facts are mental facts, and the causal relevance of relational properties is quite common. In epistemological discussions about internalism and externalism about justification, the notion of the internal is understood in terms of introspective accessibility.²⁰ You may think there are specific reasons for thinking that justification must be introspectively accessible. But you need a further argument to show that such access is part of what it is to be mental.

A question that plagued the Cartesian dualists was where and how the mental met the physical. As any token identity theorist will tell you, the mental is physical all the way through, so we need not worry about peculiar causal interactions at the edges of the mental, i.e., perception and action. This option was unavailable to the Cartesians since they thought of “mental” and “physical” as contraries. Numbers may be neither, but nothing could be both. As long as we think of the inner and the outer as contraries, with the purely psychological on one side and the external world on the other, we’re faced with the same sort of question about what happens when they meet. Of course mental events are located inside the head, but they’re individuated in terms of their relation to the world.²¹

²⁰ See William Alston, “Internalism and Externalism in Epistemology,” in *Epistemic Justification* (Ithaca: Cornell University Press, 1989), and Alvin Goldman, *Epistemology and Cognition* (Cambridge, Mass.: Harvard University Press, 1986).

²¹ I would like to thank Ned Block, Paul Boghossian, Brie Gertler, Jaegwon Kim, Thomas Nagel, Stephen Schiffer, Sigrun Svavarsdottir, and Peter Unger for helpful comments on earlier drafts of this paper.