THE JOURNAL OF PHILOSOPHY

VOLUME LXXII, NO. 1, JANUARY 16, 1975

BECAUSE HE THOUGHT HE HAD INSULTED HIM *

YOU are talking with your friend Donald when a wild-eyed red-faced man rushes into the room, delivers a right to Donald's nose and stalks out. Donald explains, "That was Leo. He's like that; he thought I had insulted him, and he attacks anyone whom he believes to have insulted him."

Donald has explained Leo's actions by appealing to a general principle: Leo attacks anyone he thinks has insulted him. Principles like this occur in many of our everyday psychological explanations; I think they are central to our everyday understanding of mind. In this paper I shall begin by discussing some linguistic issues about the sentences that express explanatory principles such as this. These issues derive from the fact that such principles involve what we have learned is an extremely problematic construction: a quantifier that governs a psychological expression.[1] Then I shall derive some conclusions about the conception of mind and action that underlies the use of such principles. My program is to develop analyses of quantification into several kinds of psychological context, and to argue that such quantification serves a function which, given the way we take people to operate, is essential to many kinds of psychological explanation.

* Ancestors of this paper were read at colloquia at Temple and Simon Fraser Universities. I have benefited from these discussions and also from conversations on related issues with Philip Hanson and Louis Loeb.

[1] The classical arguments are in Quine's "Quantifiers and Propositional Attitudes," this JOURNAL, LIII, 5 (March 1, 1956): 177–187, reprinted in his *The Ways of Paradox* (New York: Random House, 1966). My project involves combining the problems whose importance Quine first saw with an attitude toward explanation whose classical source is Carl Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," in Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965). For a valuable discussion of the application of Hempel's account to psychological explanation, see Paul Churchland, "The Logical Character of Action Explanations," *Philosophical Review*, LXXIX, 2 (April 1970): 214–236.

Let *P* be the principle: Leo attacks anyone whom he believes to have insulted him. The 'anyone' in *P* cannot mean that all instances of "If Leo believes that . . . has insulted Leo, then Leo attacks . . ." are true. For Leo believes that people have insulted him who are too remote in time or space for him to attack. Or, to get to more fundamental problems, suppose that Leo knows that there is a shortest among those who have insulted him, although he cannot identify him further.[2] Suppose that David is in fact the shortest person to have insulted him. Then we would have to predict that Leo attacks David. But this is an unwelcome prediction, since Leo may have no idea that David has insulted him.

A principle like *P* entails only those of its instances in which the belief got by filling the . . . place with a name or description of an object is *about* the object named or described. Discussions of what it takes for a belief to be about an object usually focus on three characteristics. The belief must involve a true or approximately true specification of the object; it must be informative; and it must be specific—it must be about *that* object rather than any other. Truth is not enough, as the example of the shortest insulter shows. Nor is it enough to exclude uninformative beliefs, such as the belief that the shortest person to have insulted Leo has insulted Leo. For suppose that Leo has a hallucination as of a man walking through the doorway before him and making an insulting gesture. Leo leaps forward and succeeds in attacking a man who happens to be walking through the doorway and who by pure coincidence has all the visual properties of the man in the hallucination. Leo's action should not be explained by *P* alone, for Leo's belief that there is a man in the doorway is not about the man who actually was his victim.[3] The point can also be made with examples in the manner of Grice,[4] in which Leo shoots a man behind a mirror who resembles the man whose image actually is in the mirror.

One way of spelling out the three requirements, David Kaplan's, is to argue that some names and descriptions are special. For 'Leo believes that *x* has insulted Leo' to be true of an object, some sentence 'Leo believes that *D* has insulted Leo' must be true, where '*D*' is a name or description that applies to that object and is both *vivid*

[2] The example is essentially David Kaplan's. See his "Quantifying In," in Donald Davidson and Jaakko Hintikka, eds., *Words and Objections* (Boston: Reidel, 1969). A related point is made by Hintikka in sec. 6.3 of *Knowledge and Belief* (Ithaca, N.Y.: Cornell, 1962).

[3] But *P* may be *part* of an explanation of such an action. See below.

[4] That is, inspired by the examples in Paul Grice, "The Causal Theory of Perception," *Proceedings of the Aristotelian Society*, supp. vol. xxxv (1961).

and *of* it for Leo at that time. ('Vivid' and 'of' are technical terms of Kaplan's theory, having a slight resemblance to their English counterparts: in essence, the description must mean a lot to Leo, and must have a geneology tracing back to the victim.) But this does not seem to solve the puzzle. For *P* can be used to explain why Leo attacks someone whom he glimpses briefly in a crowd, buttoning up his raincoat in a way that Leo construes as insulting. What would be the *vivid* and *of* name here? Perhaps it is something like Leo's mental image of the man. But it is hard to see how a dispensation allowing casual mental images into the privileged class can exclude *any* name or description. And in fact it does seem that, given any name related however casually to an agent and an object, one can imagine situations in which a belief that we would express by using that name can be taken as an instance of a principle, along the lines of *P,* which explains the agent's action. More on this later. There are examples that tend in the opposite direction too; for suppose that Leo does believe that *D* has insulted him, where '*D*' satisfies all the conditions that one is inclined to impose on a special name, and then *D* approaches Leo disguised as a dancing bear. We don't want to predict that Leo will punch *D* on his furry nose.

There is a moral here. If we explain actions by appealing to general principles, then we must quantify into psychological contexts to state these principles. Since we explain a variety of everyday actions on the basis of a variety of everyday beliefs, desires, and the like, the conditions that determine the applicability of such principles cannot be too restrictive. On the other hand they cannot be too permissive either, as is shown by the shortest enemy, the faraway enemy, and the dancing bear.

The dancing-bear example points to some helpful platitudes. Leo won't attack when he doesn't know that anyone in attacking range has asked for it, however much someone who happens to be in attacking range has asked for it. To get attacked by Leo (for insulting him), first one must be someone he thinks has insulted him, then one has to be situated so that he can attack one, and finally he has to know that one, thus situated, is a person he thinks has insulted him. Part of what *P* asserts, then, is that, if Leo believes there is someone who has insulted him and who now possesses a certain characteristic, his awareness of which will enable him to get about attacking that person, and if there actually is someone possessing that characteristic at that moment, then Leo will attack that person. The relevant characteristic will usually concern the spatial location of the insulter-victim, but need not; any character-

istic which in Leo's possession will allow him to attack the person having it will do.

I think this is a roughly correct account of the truth and informativeness requirements, and can be generalized to apply to all principles like *P*, which quantify into an agent's beliefs while connecting them with his actions. Postponing for a moment discussion of the general case until I have discussed the specificity requirement, and also postponing discussion of quantified psychological sentences that do not refer to particular actions, we may use what we have obtained so far to account for some features of psychological explanation.

First, the congruence of particular fact explained and general explanatory principle may depend upon the nature of the action involved. A belief that hooks into an object in a way that allows one principle to apply to it may not involve the right sort of description for another principle, referring to another kind of action, to apply. For example, *D*, who escaped Leo's ire by wearing a bear costume, may be the object of other of Leo's actions. If a principle such as 'If Leo thinks that someone has insulted him, then he will sue him' is true, then *D*'s disguise will not save him—as long, that is, as *D* possesses some property, such as a name and an address, Leo's or his lawyer's knowledge of which will enable them to start a lawsuit. The belief, '*D*, who can be subpoenaed at 111 West 111th St., has insulted me', fits a principle about lawsuits, and does not fit a principle about assaults. In fact even a description like 'the shortest insulter' could be the description under which someone is the object of Leo's energies. For example, suppose Leo has a console on which are buttons corresponding to all of Leo's insulters, ordered by height; when he pushes a button an assassin is put on the track of the corresponding person. (Leo doesn't know the names and addresses, but the computer does.) The relevant psychological principle might be, 'If Leo thinks that someone has insulted him, then he will try to cause him harm'. What matters about a description for supporting quantification in psychological principles is not any intrinsic feature of the description or the agent's grasp of it, but rather the nature of the action being explained and the associability in the agent's reasoning of the description with true information, about whatever the description describes, which will enable him to perform the action.[5]

[5] Other arguments for related conclusions are found in Ernest Sosa, "Propositional Attitudes *De Dicto* and *De Re*," this JOURNAL, LXVII, 21 (Nov. 5, 1970): 883–896, and in sec. VII of Roderick Chisholm, "The Logic of Knowing" *ibid.*, LX, 25 (Dec. 5, 1963): 773–795.

Second, we can explain why one often does not qualify a psychological principle like *P* with clauses such as 'if he gets a chance' or 'if nothing stops him'. For many such qualifications are built into the meaning of the quantifier. We can also explain why such principles seem not to be injured by many apparent counterexamples. For, e.g., Leo can believe that *D* has insulted him and fail to attack *D*, and the conclusion to be drawn not be that principle *P* is false, but rather that under the circumstances '*D*' was not a quantifiable description.

We must now discuss specificity. Examples such as that of a belief which, though caused by a hallucination, is by chance true of someone, and that of a belief produced by a mirror image that happens to fit someone whose image it is not, show that we must ensure that a principle like *P* cover only cases in which the agent's belief is about, and in a certain sense due to, some particular object. Some sort of causal condition seems to be needed. I think that the right causal condition is implicit in what I have already said, and needs only to be made explicit. To do so I must say more about how one's actions depend on one's beliefs.

To carry out an action directed at something, one normally needs more information about the object than one needs to initiate it; one has to keep track of the object while acting. When, for example, Leo attacks a supposed insulter, in order to get started he has to know the approximate location of his victim, and then to carry out the assault he has to keep track of the victim's location well enough to get near to him and then lay hands on him. He has to have information that will *guide* his action. Guiding action requires more than initiating it, since in the interval during which the action is being performed the object may change—for example move—so that new information must be delivered. A source of information that can guide actions directed at an object requires some causal commerce between object and agent; just the right causal commerce, I think, to serve as our condition. I suggest that a belief refers to a particular object, in the context of a psychological explanation, when the belief is sustained by processes that normally can guide actions of the kind being explained to objects situated as the object is—if such processes were not operating the agent would not continue to hold the belief. (I might instead have required that the belief be sustained by processes that register changes in the object's situation in corresponding changes in the agent's beliefs; but then I would have to specify which changes in the object's

situation are relevant—and the only answer is: those which must be registered in guiding actions of that kind.)

The explanatory force of the condition is that it requires the agent not only to have the information needed to get about the job, but also to have the capacity to get enough information to complete it. We can therefore see a reason why our quantifications require specificity; specificity improves the likelihood of psychological principles involving them being true. For if Leo connects with the man who happens to fit his hallucination it is purely by luck.

Putting together truth, informativeness, and specificity, we get the following. The truth of 'If a believes that something is Q, then he will do A to it' requires that a do A to any thing x such that (a) a possesses information which is true of x and which will allow a to get about doing A to it, (b) a believes that there is some thing (body) of which this information holds and of which Q is true, and (c) the information is sustained by a process that normally can guide A-ings and similar actions to their ends.

This account still makes the force of a quantifier in psychological principles of the sort I have discussed depend on the type of action named in their consequents. Yet it does help us to understand simple quantifications such as 'There is an x such that a believes that x is Q' ('a believes of something that it is Q'), which do not refer to any particular kind of action.

Two distinct phenomena seem to underlie such idioms. In the first place, one can have implicit reference to an action. If an object is such that it satisfies the conditions for being an object of someone's belief at a given time relative to an action that he is capable of performing at that time, then it is the object of that belief at that time *simpliciter,* and the unembellished quantification is true. In the second place the beliefs of an agent may be included in the preparations that lead to the actions of someone else, or of a group. For example, an astronomer has been told by her colleagues about a comet that exhibits certain characteristics. She deduces a prediction about the behavior of the comet. It seems natural and right to say that she believes of that comet that it will behave in that way (there is something which she believes to be a comet that will behave thus), although she does not know its location or anything else that would allow her to act on it. The reason that her belief is about the comet is that actions directed at the comet (e.g., observing it) could be performed by other people using the astronomer's

conclusions, and the process of group practical reasoning that leads to the action would be sufficiently like that of a single person's deliberations that the same general patterns of explanation apply. (To act on the comet, the group must have true information about it, and must be able to get information to guide as well as to initiate action.) Thus it can be that sometimes one's belief is about an object although one's only true information about it is that certain other people have beliefs about it.

The principles I have examined so far apply in only a small part of the range of situations in which we can explain what people do. The most important limitation of principles such as *P* is that they require the action in question to be specified in terms of the successful bringing about of a state of affairs involving a particular object. We are rarely so explicit; we more often describe the action in such terms as 'trying to . . .' or 'trying not to . . .' or 'avoiding . . .' or 'doing something to . . .'. It might seem that the majority of psychological principles are of a sort to which my analysis does not apply.

One class of principles unlike *P* is important because it includes principles that can be used together with principles like *P* to explain what an agent does on account of false or accidentally true information. These are principles like 'If someone would have performed a certain action on any object of which he believed *Q*, then if he believes that there is something such that *Q* he will try to perform that same action'. Or, less cumbersomely and less accurately, 'People try to do what they would do if their beliefs were satisfied'. Such a principle, conjoined with a principle like *P*, will allow one to predict what the agent will try to do, say, when hallucinating. Then sometimes further information about the particular agent and the particular situation will allow one to predict what the agent will succeed in doing. The very general principles I have just stated are, I think, principles that we appeal to repeatedly. We rarely state them. For one thing, they are hard to state, and for another we all know that we all believe something like them. They differ from principles such as *P* in that they avoid describing the agent's psychological state in terms of his beliefs about actual objects, by describing the consequent action in intensional terms, as attempts to produce certain effects.

Trying to bring about *s* amounts to doing something that one believes is likely to effect *s*. Doing it intentionally, that is; one hasn't tried to do something, whatever one's intentions, if one acts by

accident. If a psychological law asserts that under certain circumstances an agent will try to do something, it therefore asserts that under these circumstances there is an action which the agent will intentionally perform and of which he believes that it will tend to certain consequences. So laws involving 'try' do not avoid quantification into psychological contexts. They conceal such quantifiers, which range over actions rather than objects. I shall speak of *action quantifiers* and *object quantifiers*.

Action quantifiers are almost as hard to interpret as object quantifiers, when they govern psychological contexts. They require truth, and informativeness, but not specificity. Examples to show the need for truth and informativeness may be got by embedding a false or uninformative description of an object in a description of an action. For example, someone wants to get a door to open and believes that one of the buttons before him operates the door. He surely thinks that pushing the button that will open the door will lead to the door's opening, but, since he does not know which button fits the bill, he does not know which action fits the bill. If he stabs at random and by chance hits the right button, then he hasn't done something that he thought would get the door opened. Another kind of example is got by direct construction of uninformative descriptions for actions. Suppose that someone believes that there is a quickest way to get the door open, but he has no idea what it is . . .

It is hard to find examples that show the need for a specificity condition on action quantifiers like that on object quantifiers. And on reflection this is not surprising, for one can have no causal interaction with an action not yet (and possibly never) performed. The fact is that in psychological explanations actions just are referred to in a less specific way than objects are. For the object of the belief or desire that eventuates in action is the *type* of action of which the eventual performance is an instance. The statement, '*a* will perform an action of which he believes *Q*' is tricky; even besides the problematic quantifier there is a hidden shift of level. For it means "there is a type of action of which *a* believes *Q*, an instance (or token) of which *a* will perform." [6]

Putting these observations together, we get the following: an agent believes *Q* of an action if he believes that there is an action

[6] As Davidson says, "it makes no sense to demand that my want be directed at an action performed at any one moment or done in some unique manner. Any one of an indefinitely large number of actions would satisfy the want, and can be considered equally eligible as its object." "Actions, Reasons, and Causes," this JOURNAL, LX, 23 (Nov. 7, 1963): 685–700.

of a certain description which is Q and if there is an action of which the description gives true information that will allow the agent to begin to perform an action of that type. (The description must describe the action as taking place in a situation that the agent could recognize from the description well enough to get about acting.)

But although quantification over actions is not as specific a business as quantifications over objects, we do sometimes need devices that apply something rather like specificity conditions to actions. The concept of an intentional action seems to be one such device. For, just as a belief refers to an object (modulo a kind of action) if it is connected to it by means that are responsive to certain changes in the object, so an action is intentional if it is connected with beliefs and desires by means that are responsive to certain changes in the beliefs and desires. Let me try to clarify this with an example.

Leo thinks that the only way to injure his insulter is to point his pistol north-northeast and pull the trigger. He is getting ready to do this when, just as he gets the pistol pointed in the right direction, his hand trembles from the excitement of revenge and the weapon accidentally goes off.[7] This example is like the hallucination and mirror examples, which showed the need for a specificity requirement on object quantification, in that it involves a belief that is causally related to something (in this case an action) but in a degenerate way, which prevents the belief being about *that* thing. Perhaps Leo wanted to do something like what he did, but he hadn't intended to do that, then. It is unlike those examples in two ways. The causal chain runs in the opposite direction: from belief to action rather than from object to belief. And the analogous quantification 'There is an action which Leo thinks will injure his insulter and (an instance of) which he performs' is true, rather than false; Leo had thought that that type of action would injure his insulter, he just didn't intend to perform it that way. In fact, a quantified principle like "If Leo thinks that someone has insulted him, then he will do something he thinks will injure that person, but he'll be so excited that by mistake he'll act prematurely," might be true, and might explain his action.

When an agent acts intentionally, the beliefs and desires that constitute his intention are about the type of action he performs, and his performance is of a particular instance of the type.

[7] This example is modeled on an example which Davidson uses to show that an action may be caused by appropriate beliefs and desires and not be intentional. See his "Freedom to Act," in Ted Honderich, ed., *Essays on Freedom of Action* (Boston: Routledge & Kegan Paul, 1973).

He normally does not have beliefs about the instance before he performs it; at least not beliefs that support quantification, for one's intentions almost never specify an action so exactly that no action besides the particular action one performs will satisfy them. Yet there does seem to be a sense in which the intention is directed at the particular action that results in it. At any rate, I find a strong inclination to describe cases such as that in the last paragraph, in which an intended action is performed unintentionally, as cases in which the intention, though an intention to perform an action of a type of which the performance is an instance, is not directed at that very performance. I think that this purely verbal intuition can be backed up with an analysis, which is plausible in some interesting cases.

Intentional action is action that is guided by information to which it is responsive. At any rate, this responsiveness seems to be what is missing in examples such as that of Leo's premature shooting. For if Leo's information (true or false) about the location of his supposed insulter had changed at the last moment, his time and direction of firing would still have been as they were. The excitement that led to his pulling the trigger was independent of changes in the information that partially initiated it. If Leo's excitement can be modulated by his beliefs, so that when he believes his insulter has moved it decreases, rising to a trigger-pulling level only when he is sure the insulter is in the direction the pistol is pointed, then his shooting is intentional, and his excitement a medium rather than a diversion of his will.

I don't think we should try to generalize these observations into a definition of intentional action. For they leave too much undetermined: which are the beliefs, changes in which must produce changes in the action; which changes are relevant; and how sensitive must the action be to them? However, a vague principle along these lines does seem to mark intentional action: if an action is intentional under a certain description, then the agent has information relevant to performances of actions of that description, to specific features of which the action is sensitive—other information would have produced another action. If this is right and if what I said in the earlier sections of this paper is right, then we can see why intentional actions are the main focus of informal psychological explanation. For, inasmuch as we explain by appealing to general principles, we make use of laws involving quantification into psychological contexts. Many of these involve object quantifiers, and the specificity conditions on these ensure that the properties of

agents which account for their actions are not simply particular complexes of beliefs and desires but changing patterns of beliefs and desires that reflect changing information about the objects of action. Reliable correlations between such changing patterns of beliefs and desires and actions are most likely to be brought about by psychological mechanisms that reverse the mirroring process and make the particular action performed reflect changes in the beliefs and desires that occasion it.

The general function I have ascribed to quantifiers in psychological explanation may be described as follows. A commonplace model of action makes action depend upon relatively long-term beliefs and desires and other similar states (lasting for upwards of a few minutes, say) which bring about action through practical reasoning and the use of a fluctuating mass of information about the agent's environment. We know very little about the reasoning that transmutes our long-term states into our actions; we know even less about the processes by which information about the environment is gathered, collated, and used to guide action. Common sense embodies very little knowledge about these latter processes, and, as a result, everyday principles of psychological explanation usually refer explicitly to states, such as beliefs and desires, and to actions, but not to processes, such as practical reasoning or the management of information we get from our senses and elsewhere. However, action cannot occur without these processes, and so we need ways of asserting that they are working as they normally do without having to *describe* their normal workings.[8] Quantifiers that govern psychological contexts, the idioms of intentional action, and verbs such as 'try' and 'avoid,' which are quantifiers over intentional actions, do this.[9]

ADAM MORTON

Princeton University

[8] The examples one finds (and makes) in philosophical discussions of quantification and intentional action seem curiously dominated by search and violence. Though it is a far-fetched and unsupported idea, I would speculate that this is no accident, for the psychological mechanisms that lie behind the applicability of these concepts were developed during a long history of hunting, flight, and murder.

[9] For a similar assertion about some modal idioms, see sec. 4 of my "The Possible in the Actual," *Noûs*, VII, 4 (November 1973): 394–407.