

REASON AND CAUSATION IN DAVIDSON'S THEORY OF ACTION EXPLANATION*

CARLOS J. MOYA
University of Valencia

Davidson's famous 1963 paper "Actions, Reasons, and Causes" contains, *in nuce*, the main lines of Davidson's philosophy of action and mind. It also contains the seeds of some major problems of Davidson's thought in these fields. It will be instructive to go back to this paper in order to uncover the basic structural lines and notions which underlie Davidson's influential theory, as well as the roots of the difficulties that affect it. We might thereby, from the advantageous perspective provided by the work of a philosophical giant, try to improve our vision of the intricate landscape of human agency and mind. I shall defend, following Davidson, that rationalization or reasons expla-

* Research for this paper has been funded by the DGES of the Spanish Ministry of Education and Culture, as part of the project PB96-1091-C03-02. I am also grateful to the Instituto de Investigaciones Filosóficas of the UNAM, and especially to its Director, Dra. Olbeth Hansberg, for inviting me to take part in the XVI International Symposium of Philosophy, this time dedicated to the work of a philosopher I highly admire, Donald Davidson. This paper is a modified version of the one I presented to this Symposium. My thanks also for his reactions and comments to Donald Davidson himself (who, I guess, will probably still disagree with some aspects of my interpretation of his work in the present version), as well as to John McDowell, Akeel Bilgrami, Marcia Cavell, Carol Rovane and the audience of the Symposium, which constituted a certainly memorable event.

nation is a species of causal explanation, but I will be contending, against Davidson's approach, that causality is best viewed, in this kind of explanation, as an integral aspect of justification itself, and not as an independent, additional condition.

Davidson's main target in the aforementioned paper was the thesis that reasons explanations of actions were not causal explanations, and that reasons were not causes of actions. Champions of this thesis used to defend it by pointing to the characteristic logical or conceptual structure of reasons explanation. A remarkable feature of reasons explanation is that the concept of the explanandum is also used in a fully explicit statement of the explanans. Since there is a conceptual connection between explanans and explanandum, the explanans (the reason) cannot be a cause of the explanandum (the action), for, according to the orthodox Humean view of causation, cause and effect must be conceptually independent. This conceptual connection, moreover, is an essential ingredient of the characteristic justificatory role of this sort of explanation. The description under which the action is explained by the reason has to appear in the description of the desire or the belief that jointly constitute the reason if the action is to be justified, under that description, by citing that reason. Reasons explanation allows us to see the action an agent has performed as anticipated by the agent and envisaged by him as a means to an end he positively values. That is why the action description also appears in the description of the reasons. Now this justificatory role of reasons explanation is also appealed to in the course of arguing against the causal character of this kind of explanation and against the causal character of reasons: since reasons justify actions and causes do not justify effects, reasons cannot be causes of actions. Davidson reports this line of argument as follows:

Noting that nonteleological causal explanations do not display the element of justification provided by reasons, some philosophers have concluded that the concept of cause that applies elsewhere cannot apply to the relation between reasons and actions, and that the pattern of justification provides, in the case of reasons, the required explanation (Davidson 1982, p. 9).

Davidson can afford to concede to his opponents both that rationalizations, or reasons explanation, include this element of justification and that part of this justificatory aspect is the existence of a conceptual connection between the descriptions of the reason and the action. He denies, however, that the non-causal character of rationalizations can be validly inferred from these two points. Justification is certainly a central feature of rationalizations, but this feature can be conceived as the *differentia specifica* of rationalizations inside the general category of causal explanations. Justification adds to the common requirement of causality as a “differentiating property” (ibid.) of rationalizations. This differentiating property is spelled out in Davidson’s first condition on primary reasons:

C1. *R* is a primary reason why an agent performed the action *A* under the description *d* only if *R* consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property (Davidson 1982, p. 5).

It is worth pointing out that, in later articles, Davidson himself will find the statement of this condition inappropriate and propose an alternative to it; however, I will not enter into this change, important though it is, since it is not essential to my line of argument in this paper. We can see that C1 includes a conceptual connection between explanans and explanandum: the description under which the action is to be explained occurs also in the content of the

agent's belief. Only in this way can we see the performing of an action of the corresponding kind, to use Churchland's expression, as "reasonable in the light of" the agent's mental attitudes. To a certain extent, Davidson's remarks can be seen as both an endorsement and a clearer statement of some central presuppositions of non-causal theories of action explanation. Davidson shares with these theories the view that the justification relation between reasons and actions is essentially logical or conceptual, not causal: it is a matter of having the right logical relations between descriptions of reasons and actions and does not include the idea of a causal influence of reasons on actions. I shall try to countenance this claim below. My contention will be that this concession to his opponents, this non-causal view of the justification relation, is an important source of some major problems that Davidson's theory of action and mind has to confront, namely the problem of wayward causal chains and the problem of epiphenomenalism of mental properties.

Davidson's distinctive position is that rationalization cannot be reduced to justification, in the sense of C1 or similar conditions, for an agent can have reasons and perform actions that conform to the pattern of C1 without those reasons' being the reasons why he acted that way: "... A person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it" (ibid.). If this is the case, then the reasons do not explain the action, even if an action of the corresponding kind is reasonable in the light of those reasons. Hempel, for example, objected to William Dray that from the fact that a way of acting is reasonable or appropriate for someone, given his reasons, it does not follow that he will act that way (Hempel 1966). In reducing rationalization to the reasonableness or justification relation, understood in a sense close to condition C1, non-causal theories are unable to

distinguish the case we are considering from the case in which the agent acts *for* those reasons. Davidson's point is that the most natural way of drawing this distinction is to say that in the latter case, unlike the former, the reasons caused the action. Hence his second condition for primary reasons runs as follows: "C2. A primary reason for an action is its cause" (Davidson 1982, p. 12).

In his defence of C2, Davidson insists that a conceptual or logical connection between certain ways of describing or classifying reasons and actions, linked to the justification aspect of reasons explanation, is compatible with the existence of a causal relation between reasons and actions as particular events. Since condition C2, the causal requirement, is common to all true causal explanations, I think it is fair to interpret Davidson as holding that the *differentia specifica* of rationalizations or rational explanations of actions, namely the feature of justification, is captured exclusively by condition C1, which is what distinguishes rationalizations from other kinds of causal explanations. This means that justification and causation are seen, in Davidson's theory, as fully independent conditions on reasons explanation. This interpretation might be challenged by identifying justification and rational explanation, so that justification included also condition C2. But I do not think this move is open to Davidson, for he thinks that all causal links are covered by strict laws, which are physical laws. Now it is clear that physical descriptions of reasons and actions, under which they would be covered by physical laws, do not possess any special justification capacity. On the other hand, the intentional descriptions characteristic of justification do not play any role in backing the existence of a causal link between the particular reasons and actions that are so described. So, in Davidson's theory, the kinds under which reasons and actions are classified for purposes of justification are not the kinds under which

they are, or would be, classified for purposes of causal explanation. If this interpretation is correct, as I think it is, then Davidson's concept of justification does not include the idea of a causal link between what justifies and what is justified. The causal link is conceived as a condition on rationalizations which adds to the justification condition and is independent of it. In this way, the causal requirement on rationalizations is seen as fully independent of the justification requirement.

This independence between justification and causation, this non-causal view of justification, which Davidson inherits from non-causal theorists, gives rise to the aforementioned problems of wayward causal chains and epiphenomenalism, as I will try to show in what follows. To sum up my position, I contend that Davidson is right in holding that rationalizations are causal explanations, but wrong in holding that rationalizations consist in something more than justification. Non-causal theorists, in turn, are right in holding that rationalizations consist in justifications, but wrong in holding that justification is a non-causal explanation.

The structure of cases of wayward causal chains is the following. An agent has a desire to p and a belief that by performing an action of kind A he will satisfy his desire; moreover, he performs an action of kind A and the causal antecedents of his performance include the indicated desire and belief; nevertheless, we cannot say that this action is rationalized or rationally explained by the agent's belief and desire. A possible conclusion to be drawn from these cases is that conditions C1 and C2 do not fully capture the concept of reasons explanation, so that these conditions are only necessary, but not sufficient, to have that kind of explanation. This is, as I see it, Davidson's position. The usual move of requiring that reasons cause actions 'in the right way' is handwaiving unless the right way can be

specified in non-circular terms, that is, in terms that do not include or presuppose the very idea to be elucidated, namely the idea of reasons explanation or rationalization, but this non-circular specification is something that, according to Davidson himself, cannot be given: “What I despair of spelling out is the way in which attitudes must cause actions if they are to rationalize the action” (Davidson 1982, p. 79). Besides, this common move, in leaving untouched the general structure of the analysis of rationalizations, conceals the deeper roots of the problem, which, in my view, have to do with that general structure, rather than with its details. In cases of waywardness the action is certainly not rationalized, i.e. not rationally explained, by the agent’s reasons; but, if we reflect on these cases, we can see that the action is not even *justified* by those reasons in that particular causal context. The action performed by the agent is of a kind that could be justified by the kind of reasons the agent has. But the particular performance of that action is not justified by those reasons. This suggests that justification, or at least the sort of justification at hand in reasons explanations of actions, is not merely a general relation holding between kinds or descriptions of reasons and actions, but rather a relation that holds between particular reasons and actions in particular contexts, although partly in virtue of their falling under certain descriptions or kinds. If justification were just a general relation between kinds or descriptions, it could not be affected by the particular form taken by a causal relation which holds between particular events, provided that these events still have the right true descriptions. But it is, as cases of waywardness show. My view is that, in the context of reasons explanations of actions, causation and justification are not independent and disparate relations between reasons and actions. Rather, justification includes, as a necessary condition, a causal relation between what justifies and what is

justified. That is, in order for an action to be *justified* by an agent's reasons, these reasons have to cause the action, they have to be the reasons why he acted as he did. In this way, reasons explanation is causal explanation, but it amounts to no more than justification, as long as justification is held to involve a causation requirement. On this perspective, cases of waywardness are excluded in that in these cases the performed action is not justified by the agent's reasons. We may accept that these reasons cause the action, but they do so independently of the relation of justification: the action is not, in these cases, so to speak, the conclusion the agent draws from, *and because of*, the relationship between the content of his attitudes and the appropriate action description. Instead, cases of waywardness are endemic in a perspective in which the causal relation is conceived as running parallel and independent of justification, understood as a non-causal, general relation between descriptions of reasons and actions.

Let me now try to indicate, somewhat tentatively, what I think is included in the idea of an action being justified by a reason, and how the conceptual and the causal components mesh together. I shall use the analogy with the activity of deducing conclusions from premisses. There are, in this case, abstract logical relations between the propositions expressed by the premisses and the conclusion. However, considering the content of the premisses, seeing what is logically implied by them and drawing this as a conclusion are particular psychological events, distinct existences, which can be causally linked. Nevertheless, if the drawing of the conclusion is to be rationally justified, the causal link has to be grounded in the abstract, logical relations between the contents of the premisses and the conclusion, that is, a particular drawing of a conclusion is rationally justified as long as the agent draws the conclusion because of (his perceiving) the logical relationships between the contents

of the premisses and of the conclusion itself. Applying this to ordinary physical actions we have: there are conceptual relations between the contents of the reasons and the appropriate description of the action, but the reasons and the action are, as Davidson himself insists, particular (psychological) events which can be causally linked. However, the particular action is justified by those reasons as long as the agent performs that action because of (his perceiving) the conceptual relations between the contents of his reasons and the appropriate description of his action. The mere existence of these conceptual relations between reasons and action does not amount to justification unless the action is performed because of these relations. And, when the action is so justified, it is also rationally explained. Causation is an aspect, a necessary condition of justification itself, not a relation independent of it and justification is, then, partly a causal relation.

Viewing justification as a non-causal relation and conceiving of causation as independent of justification itself is also at the root of the problem of epiphenomenalism of content and mental properties. If the line of causation between reasons and actions runs parallel to the justification relation and rests on different grounds, as happens in Davidson's theory, there is always the possibility that attitude contents remain causally inert. On the contrary, if justification is already a causal relation, if we accept that the justifying role of a reason includes its causal influence on what is justified, then we can question the causal efficacy of content only at the price of questioning justification itself, and with it the idea of reasons explanation. If the content of an agent's reasons is not causally efficacious in his performing of the corresponding action, this is not justified, and hence not rationally explained, by those reasons. Viewing justification as a causal-explanatory relation runs against any form of irrealism about the mental. For an action to be rationally

justified it is not sufficient that it can be given a plausible interpretation in terms of reasons and intentions. For intentional realists such as Dennett and Churchland, the existence of rational action depends on the possibility of rational interpretation of behaviour, on the adoption of a specific perspective on behaviour which Dennett has called “the intentional stance”. There is no commitment to the idea that the reasons alluded to in this interpretation are in fact causally responsible for the agent’s behaviour. However, on the causal view of justification I am defending, this position becomes incoherent: we cannot have rationally justified behaviour if we deny the causal efficacy of reasons, for this efficacy is a necessary condition of justification. On this perspective, the existence of rational action is incompatible with epiphenomenalism of content, so one cannot have both epiphenomenalism and rational action.

As for Davidson, his view of justification as a non-causal relation between descriptions of reasons and actions is congenial with an interpretative view of rational action, where mental attitudes are attributed in a holistic manner, addressed to maximizing rationality of agents. On this perspective, several empirically equivalent, but mutually incompatible, interpretative schemes can be used in order to make sense of an agent’s behaviour, several mutually incompatible sets of mental attitudes can be attributed to him that fit the empirical behavioural data. On the other hand, however, his early thesis that reasons are causes of actions leads him towards a more realist view of the mental. On this realist perspective, an interpretative scheme can be false, even if it fits the data, if the reasons attributed in it to the agent did not in fact cause the agent’s behaviour. This realist approach to the mental, clearly present in Davidson’s theory, is however undermined by the fact that the justificatory links between contents and intentional descriptions of actions are seen as independent of the causation relation,

so that interpretation and rational explanation of behaviour can run in fact free from realist causal constraints, giving rise to a tendency to irrealism and to epiphenomenalism. Intentional realism, I contend, implies a rejection of the independence between the conditions of justification and causation on reasons explanation of action.

Given the centrality in our argument of the view that justification is not independent of causation, but includes it as a necessary condition, I shall finish this paper with some considerations in defence of this view.

Our first consideration has already been suggested above. It has to do with the problem of wayward causal chains. Cases of waywardness show, among other things, that the justifying relation between reasons and actions can be undermined or destroyed by the way the reasons cause the action, which should not happen if justification were just a matter of logical or conceptual relations between descriptions of reasons and actions and did not include a causation link.

Our second consideration has to do with Davidson's distinction between singular causal statements and causal explanations. The former are extensional, while the latter involve logical relations between laws and descriptions of events and are therefore intensional. This distinction is used by Davidson, among other purposes, to clarify some confusions involved in the logical connection argument and to resist its conclusion. If we take this distinction into account, we can see that the existence of logical or conceptual relations between descriptions of two events is compatible with the existence of a causal relation between the events themselves. So, even if there is a logical relation between descriptions of reasons and actions, reasons may still be causes of actions. But I think that rationalizations do not clearly fall under any of these two categories. Davidson, it seems, must hold that rationalizations are singular causal

statements, for the intentional descriptions appearing in them do not fall under strict laws. But this assumption faces difficulties, for the descriptions of reasons and actions play a crucial role in making intelligible the causal link between them, a role that is played by laws in explanations of physical events. In this sense, rationalizations get closer to causal explanations, but they fall apart from them owing to their lack of strict laws covering reasons and actions under intentional descriptions. Moreover, descriptions of reasons and actions in true rationalizations not only affect the justification relation, but also the causation relation: the descriptions of beliefs and desires under which the action is justified are also the descriptions under which those beliefs and desires lead agents to act. This is part of what I mean in saying that causation is not independent of justification, but an internal aspect of it and, by the way, it is also the natural reading of Davidson's condition C2, according to which a primary reason for an action is its cause. Rationalizations cross-cut, then, Davidson's distinction between singular causal statements and causal explanations. This oblique position of rationalizations can be accounted for if we accept that they consist in justifications, which in turn include both the assumption of a causal relation between particular reasons and actions and a statement of the general relations between descriptions or kinds that form the intelligibility basis of that causal relation. Rationalizations are not mere statements of a singular causal relation waiting for more perspicuous, physical or physiological descriptions which get them closer to causal explanations; rather they include precisely the descriptions that are relevant to understand the causal link reported in them.

Our third consideration is based on certain natural analogies between action theory, on the one hand, and epistemology and ethics, on the other. In epistemology, it is not uncommon to assume that a belief is not justified by oth-

er beliefs unless it is caused by them. This assumption, I think, should also be made in action theory: an action is not justified by a reason unless caused by it (see Audi 1990, p. 232). This leads us to ethics. It seems pretty clear to me that we do not take an action to be morally *justified* by a reason just because there is a coherence relation between the reason and the action: we also require the agent to be caused to act by the reason offered. To see this, think of the following example. John is an upright civil servant, who thinks that public services should not be distorted in favour of private interests. John discovers that Peter, a colleague who competes with him for a higher post, has altered the order of a file in order to give preeminence to the application of a relative. Consequently, John decides to report this fact to his superiors and does so. Before his other colleagues he offers, as a moral justification of his act, his desire to keep public services clear of private interests. Let us suppose, however, that what actually led him to denounce Peter was the opportunity this gave him to get rid of a competitor. Even if the moral reason he gives to his colleagues was a reason he had to denounce Peter and is coherent with his action, it was not the reason why he did it. Davidson would say that John's action is not *explained* by the reason he gives owing to the lack of the causal condition. I want to say something stronger: John's action is not even morally *justified* by that reason precisely because this was not the reason that caused his action. If this is our intuitive judgment about this example, then this shows that the causal conception of justification I am putting forward is not merely stipulative, but reflects the concept of justification we actually have and employ in our moral assessments of certain actions. Again, it is plausible to hold that this applies also to rationalizations in general: being coherent with a certain reason is not enough for an action to be rationally justified by that reason. We also

require, at least, that the action be caused by the reason. Our concept of justification is not merely formal, but presupposes the causal efficacy of reasons.

If this is on the right lines, Davidson's important thesis of the causal character of rationalizations is not rejected, but defended on slightly different grounds and vindicated against some usual objections. There is no denying, however, that this defence runs against other, maybe no less harder objections. But their treatment, however, as one usually says at this point, must be left for another occasion.

REFERENCES

- Audi, R., 1990, "An Internalist Conception of Rational Action", *Philosophical Perspectives*, 4, pp. 227, 245.
- Davidson, D., 1982, *Essays on Actions and Events*, Clarendon Press, Oxford.
- Hempel, C.G., 1966, "Explanation in Science and in History", in W.H. Dray (ed.), *Philosophical Analysis and History*, Harper & Row, New York, pp. 95-126.

Recibido: 17 de febrero de 1998

RESUMEN

En la concepción de Davidson, las explicaciones de la acción en términos de razones incluyen dos aspectos o condiciones independientes entre sí: una condición de racionalidad o justificación racional y una condición causal. La satisfacción de la primera depende de relaciones lógicas apropiadas entre las descripciones de la razón y de la acción. La segunda exige únicamente la existencia de un vínculo causal entre razón y acción. Es esta independencia entre las dos condiciones la que, en nuestra opinión, genera en la teoría davidsoniana los problemas de las cadenas causales desviadas y del epifenomenismo de las propiedades mentales. Frente a esta independencia entre racionalidad y causalidad, sostenemos que la justificación racional de una acción incluye ya, como condición necesaria, la eficacia causal de la razón, y no requiere, pues, una condición causal independiente de la propia justificación.