

Reasons and first-person authority

Book or Report Section

Published Version

Schroeder, S. ORCID: <https://orcid.org/0000-0002-4480-6458>
(2017) Reasons and first-person authority. In: Padilla Gálvez, J. and Gaffal, M. (eds.) Intentionality and Action. de Gruyter, pp. 123-138. ISBN 9783110559095 Available at <https://centaur.reading.ac.uk/75535/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Publisher: de Gruyter

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Reasons and First-Person Authority

1 Introduction

Typically, an explanation of an event cites the causes that made it happen. According to what has been called the ‘standard view’ in the philosophy of action,¹ intentional human actions are just a subclass of events and the agent’s reasons are the causes of their actions. In the 1930s this widespread causalism was disputed by Ludwig Wittgenstein, whose alternative view that reason-giving explanations should *not* be seen as a type of causal explanations, held sway in the 1960s. It was defended by Elizabeth Anscombe,² Peter Winch,³ and A. I. Melden,⁴ but was then forcefully challenged by Donald Davidson, in his influential paper ‘Actions, Reasons, and Causes’,⁵ which by the 1970s had re-established causalism about reasons as the standard view. The standard causalist view, however, has generally been acknowledged to be troubled by a major problem: the problem of deviant causal chains.⁶ Davidson himself had confessed not to be able to resolve this problem,⁷ yet many of his followers remained more optimistic.

A more recent causalist view, propounded for example by Carlos Moya⁸ and John Hyman,⁹ is that the flaw both in Davidson’s version of causalism and in the anti-causalism of Wittgenstein and his followers was their commitment to a Humean account of causation, as a law-governed relation between events. On a more flexible construal, Moya and Hyman argue, causalism can be defended against its critics. — In this paper, I shall, after a few preliminary clarifications, re-consider the deviant causal chain problem, with those more flexible causalist ideas in mind, and then try to give a defence of Wittgenstein’s key anti-causalist idea: first-person authority about reasons.

1 Cf. Sandis 2009, 2.

2 Anscombe 1957.

3 Winch 1958.

4 Melden 1961.

5 Davidson 1963.

6 Moya 1990, 114–28.

7 Davidson 1973, 79.

8 Moya 2014, 200–4.

9 Hyman 2015, ch. 5.

2 Reasons and causes

To begin with, let us note that reasons are not causes: the two concepts are categorially different. Although a thought may well be cited both as a reason and as a cause, this is due to a well-known ambiguity: by a thought we can mean either the occurrence of a thought (the mental event of thinking of something at a given moment) or the contents of the thought (what we think). Thus, the mental event of my thinking or remembering that I've run out of cash may be said to *cause* me to set off in search of a cash machine, whereas the *reason* I offer to justify my behaviour is the content of my thought: that I need some cash.¹⁰ It can occasionally happen that my reason for acting is not *what* I think, but the fact that a certain thought occurs to me. This is nicely illustrated by a pair of examples given by Hyman:¹¹

(1) Roger fled to Brazil because he believed that he was being pursued by the Security Services.

(2) Roger saw his doctor because he believed that he was being pursued by the Security Services.

In the second case, on a natural interpretation, Roger's action was not motivated by the contents of his belief, but by its occurrence, which he took to be pathological. However, even in such a case reason and cause are not identical. The occurrence of such irrational thoughts caused Roger to seek medical advice, but the reason for his action was not the (datable) *occurrence* of the thought, but the (non-datable) *fact* that it occurred.¹² The fact that a football match lasts 90 minutes does not itself last 90 minutes, and if a stone hit the window it is not *the fact that a stone hit the window* that hit the window (and caused it to break). Similarly, *my thirst* may have caused me to drink a whole pint, but the reason I did so was *that I was thirsty*: a fact, not a desire.

This categorial difference between reasons and causes does of course not rule out a causalist construal of reason-giving explanations. But instead of saying that reasons are causes, a causalist should propose, more carefully, that a reason-giving explanation is correct if and only if a corresponding causal claim is correct. For example, on the causalist view, the following two statements are equivalent (subject perhaps to some minor qualifications):

¹⁰ Wittgenstein *PLP*, 123; Gnassounou 2014, 58.

¹¹ Hyman 2001, 182.

¹² Gnassounou 2014, 59.

- (3) I went in search of a cash machine because I had run out of cash.
 (4) The thought that I had run out of cash caused me to go in search of a cash machine.

3 The problem of deviant causal chains

However, the causalist view faces the problem of deviant causal chains. Two versions of this problem need to be distinguished.

(i) First, in the well-known examples given by Roderick Chisholm¹³ and Davidson,¹⁴ the resulting action although caused by the right belief and desire is not even an intentional action. Rather, the agent's thoughts make him nervous and thus cause him to bring about the envisaged action unintentionally (e.g., careless driving makes him inadvertently and coincidentally run over the very person he is thinking of murdering). This kind of case serves to highlight that the existence of a suitable desire-belief pair is not sufficient for the intentional action in question to be carried out. I may want to do it and know how to do it — but not be able to bring myself to do it. Therefore, causalists proceeded to insist that the mental cause of an intentional action must be an *intention* to carry out that action.¹⁵

However, this modification does not overcome the deviant causal chain problem. Assuming for the moment that an intention is indeed a mental occurrence suitable for causing an action (perhaps envisaging it as a kind of decision to act),¹⁶ it may well lead to the action being carried out unintentionally. In Chisholm's example of the homicidal driver, we can imagine that the man was already determined and on the way to committing the crime when his eagerness to carry it out made him kill his victim accidentally. It is evidently not enough for the intention to figure among the causes if in the end it doesn't inform the action itself. As Moya puts it, in his critical discussion of Irving Thalberg's causalist theory: it is not enough that as a result of an intention an agent fulfils that intention; 'the agent has to be conscious of fulfilling his intention'¹⁷. That is indeed the crucial point: for an action to be intentional the agent has to be aware of carrying it out. But this awareness is a feature of the action itself and

¹³ Chisholm 1966, 29–30.

¹⁴ Davidson 1973, 79.

¹⁵ Brand 1984, 31–5; Thalberg 1984, 256; Stout 2010.

¹⁶ For reasons to think otherwise, see Anscombe 1983; Rundle 1997, 197–206; and Clot-Goudard 2015, 68–76.

¹⁷ Moya 1990, 122.

not a causal factor preceding it. It is certainly possible that a plan or intention precedes the action, but that is neither a sufficient nor a necessary condition of intentionality. It is quite enough for the intention to be there in the act, as an ability I have of explaining what I'm doing. The causal preliminaries cancel out as conceptually irrelevant.

(ii) Another version of the deviant causal chain problem occurs when the intentionality of the action is not in question, but we want to ascertain the reason for which the person acted. This is the point where Davidson presents his only positive argument for a causalist construal of reasons. Where a person performs a certain action and is aware of more than one reason to do so, we need a criterion by which to decide which of the reasons did in fact motivate the person to act. For Davidson, the only plausible criterion to single out the (actual) *reason why somebody* \square -*ed* from (potential) *reasons to* \square is a causal link between reason and action.¹⁸ But, as Davidson had to acknowledge, it is easy to think of cases in which an action is caused by the thought of a reason that was not the one for which the person acted. In a fairly common scenario, considering reason *A* may lead to the discovery of reason *B* which then motivates me to act. The thought of reason *A* would have caused the action, although it wasn't the reason for which I acted. An obvious response would be to demand that the required causal link between the thought of the reason and the action must not be indirect, but proximate.¹⁹ But that doesn't help, for we can easily imagine that the proximate consideration that triggers the action provides a mere reminder or pretext, not the reason for which the person acts. For example, hearing a moving account of the economic difficulties of the local bicycle shop leads me to buy a bicycle there; yet I may have been planning to make this purchase anyway for selfish reasons and now I have just been reminded of this intention. Perhaps I do indeed like the idea of supporting a local shop, but it would never have been a reason for me to buy a bike had I not wanted to do so anyway.

Others have suggested that the (thought of the) reason must be required to cause the action 'in a way that is characteristic of' acting for that reason,²⁰ or in the way in which a disposition causes its manifestations.²¹ But it is hard to see how that could provide us with a criterion for deciding, in a given case, whether somebody acted for reason *A* or for reason *B*. One suspects that what this kind of uninformative restriction really amounts to is that we do not in fact rely on a

¹⁸ Davidson 1963, 9.

¹⁹ Thalberg 1984, 257.

²⁰ Stout 2010, 162.

²¹ Hyman 2015, 116–127.

causal analysis for identifying agential reasons. Rather, when we think we know independently that somebody acted for reason *A* and not for reason *B*, we say post hoc that only (the thought of) reason *A* caused the action ‘in the right way’. In other words, causalism, unhelpfully qualified, becomes a mere postulate that doesn’t do any work.

The truth is that even if somebody did specify what the ‘right’ causal pathway was supposed to be, we know far too little about the causal pathways of our thoughts and desires for such a causally construed concept of an agential reason to be usefully applicable. The insufficiency of our knowledge of the causal impact of thoughts and desires tends to be disguised by the fact that we focus on a few examples where the causal pathways, and the strength or weakness of particular desires are simply assumed to be obvious. Yet psychological causes are not like inner billiard balls whose paths can be followed by introspection. This may to some extent be the case with mental occurrences: I can, at least sometimes, fairly reliably report a causal succession of ideas (— how a given sentence reminds me of my sister — which memory in turn brings up a mental image of my school — which then reminds me of the book on architecture I just bought — which then makes me pick up that book). But dispositional phenomena, such as beliefs and inclinations, are not mental occurrences and so their effects cannot be read off their temporal position.

Acknowledging our ignorance in this area, Alvin Goldman suggested that an ‘explanation of how wants and beliefs lead to intentional acts would require extensive neurophysiological information’²². But for one thing, it is far from clear that in a case where the agent is aware of various potential reasons neurophysiological data could neatly decide which considerations were the principal cause of his action. For all we know, even thoughts that we don’t regard as relevant may play an important role in the complicated causal network of our psyche and its underlying physiology. For another thing, our concern is not with a possible future concept of a reason, defined in the light of future scientific findings and possibilities, but with a concept that has played a central role in our discourse for centuries: for as long as people have been articulate enough to inquire about and explain their behaviour to each other. Clearly, relying on criteria that remain either unspecified or that we are in fact unable to apply, causalism fails to offer a plausible analysis of this everyday concept.

²² Goldman 1970, 62.

4 Reasons and first-person authority

At the core of Wittgenstein's distinction between reasons and causes lies the idea that the grammar of the former is characterised by first-person authority:

...a person does not generally know the *causes* of his activities. ...he will frequently be mistaken in specifying the cause. Strangely enough, he cannot be mistaken in specifying his *reason*. ... That is, we *call* the reason that which he gives as his reason. The cause of an action is established by observation, namely *hypothetically*, i.e., in such a way that further experiences can confirm it or contradict it.²³

John Hyman sees in this (and a similar passage in *BB* 15) only the claim that causal explanations are *conjectures*, whereas we are immediately certain of our reasons,— to which he rightly objects that causal statements can be certain and immediate too.²⁴ In fact, that is something Wittgenstein himself remarked and illustrated by the example of being startled by suddenly seeing a light: in which case one does not *hypothesise* on the basis of repeated observations about the cause of one's bodily movement, but is immediately aware of that cause.²⁵ So it is indeed inaccurate to say that 'one can only *conjecture*' the cause (as Wittgenstein does in *BB* 15), but that is not his main point.

The crucial contrast is not between immediate knowledge and mere conjecture based on repeated observation, but between, on the one hand, an authority conferred by our concepts (hence not really a matter of knowledge at all) and, on the other hand, any kind of empirical finding, be it immediate perception or a tentative conjecture. In the latter case there is an independent state of affairs of which I come to know, be it step by step through repeated observation or immediately. Either way, my cognition is fallible. Error may be highly unlikely or even practically impossible, but extraordinary circumstances that would prove my judgement erroneous are at least conceivable. Thus, it is at least imaginable that the car I see right in front of me in broad daylight proves to be an ingeniously constructed hologram. Similarly, when I suddenly see a light or a face and startle, it is conceivable (however far-fetched and unlikely) that my startling reflex was triggered by some electrodes in my brain quite independently of what I happened to see at that moment.

²³ Wittgenstein *VW*, 109–11.

²⁴ Hyman 2015, 118.

²⁵ Wittgenstein *CE*, 408.

By contrast, first-person authority does not underwrite reports about independent occurrences, but marks certain kinds of sincere avowals of a person's impression as constitutive of a certain concept. Thus, when I can sincerely complain that something feels painful to me, then — by definition or by grammar — it is painful. This concept of feeling does not allow any room to a distinction between appearance and reality; that is, error in this case is logically impossible. (Not because I know so reliably whether I'm in pain, but simply because this is not really a case of knowledge at all.²⁶)

Beliefs and preferences provide other examples of first-person authority. Whereas my *considered* opinions and *long-term* preferences may require some working out (researching and weighing of evidence, testing how different things agree with me), there are also concepts of current belief and spontaneous preference that capture how things appear to me and what I am inclined to choose at a given moment. There, again, error is logically excluded. If I can sincerely say that I don't think Arsenal will win the Champions League and that I'd like to have a cup of tea now, then that is *ipso facto* what I believe and what I want at that moment.

Similarly, according to Wittgenstein, there is a concept of a person's reason that is simply constituted, by that person's possible sincere avowal of his reason. Not because we take people to be so reliable at finding out what their reasons are, but because a reason (in this sense), unlike a cause, is not an independent occurrence at all: it is simply defined to be what an agent can honestly say in response to the question: 'Why are you doing it?'. Thus, Wittgenstein's account offers a straightforward solution to the problem that stymied Davidson: the problem of identifying the reason for which somebody acted.

To see more clearly how first-person authority is built into our concept of an agential reason, consider the concept of an intentional action. Intentional actions are a subclass of conscious voluntary actions. They are 'actions that are chosen either as ends in themselves or as means to some other ends'.²⁷ The important thing to note is that to be conscious of an intentional action — to know what it is one is doing — is to be conscious or aware of its intentional characteristics, which make it the action it is. In the case of an intentional action, to be aware what one is doing must include being aware that one is doing it as an end in itself or as a means to some other end. In other words, it conceptually includes an awareness of one's reasons, i.e. first-person authority.

²⁶ Wittgenstein *PI*, § 246; Schroeder 2006, 208–10.

²⁷ Kenny 1989, 45.

Why, for example, am I pushing a plastic card into a slot and pressing a few buttons? My reason for doing so (— I'm making a payment, buying a bicycle, with which I'd like to go on weekend bike rides —) is an essential characteristic of my very action. Instead of 'Why are you pressing those buttons?' one might just as well ask: 'What are you doing there pressing those buttons?'. My reason is not just describing the preliminaries that led to the action (as causalism suggests), it characterises the action itself. If I weren't able to explain my action's intentional features (by either giving a reason or, less likely, saying that I thought pressing buttons was worthwhile in itself), then it just wouldn't qualify as an intentional action. It might just be a kind of fidgeting.

Note that first-person authority in giving one's reasons — as all kinds of first-person authority — is strictly speaking only a matter of present tense. That as an intentional agent I can authoritatively say what I am engaged in doing does obviously not guarantee the truth of my (sincere) descriptions of my past actions. However, with regard to one's actions in the recent past where, given ordinary human memory capacities, one can be trusted to remember the relevant circumstances of the situation, first-person authority is also granted retrospectively. That is because, as long as a person's views and preferences haven't changed, we can assume that the justification he offers now is probably not much different from the one he would have offered at the time of the action.

Giving an explanation of one's recent action is to perform another action, often manifesting the same want and belief as the original action. To explain one's verbal utterance is normally a matter of paraphrasing or supplementing it. 'It's a disgrace!' —?— 'I mean: it's very unfair that he got sacked after all he has done for the company.' In this case the philosophical question as to the virtual infallibility of an agent's explanation of his action is easily answered. Action and explanation are on the same footing. In both cases the person speaks his mind; and there is no puzzle about our ability to say repeatedly what we think about a subject, being a little more explicit the second time.

There are similar pairs of actions in the non-verbal sphere. Suppose you see Fanny Robin throwing a snowball against a wall, and ask yourself: 'What on earth did she do that for?' Then you see her throwing another snowball against a first-floor window, and you understand that that is what she was trying to do the first time: hit the window to attract someone's attention. The second action explains the first by pursuing the same purpose more expediently, and thus more evidently.

Now consider the case where being asked for her reason after the first throw she explains: 'I want to hit that window to attract Sergeant Troy's attention'. Her ability correctly to give her reasons ceases to look puzzling if we see this kind of

case as a variation upon the preceding ones. Action and explanation are simply two actions manifesting the same intention (e.g. to attract someone's attention) and the same belief (e.g. that it can be done by throwing a snowball against a window). Given that she can speak, her ability to say for what reason she acted is not more surprising than her ability to repeat the action.

Here we have an illustration of how giving one's reason is not a matter of describing what happened in one's mind. One only has to express truthfully what one wants and believes (or knows) in the matter. Of course, things are a little different if one's wants or beliefs are no longer the same. Once Fanny Robin has succeeded in hitting the window and thus attracted Sergeant Troy's attention, she will no longer have a want to do so. Hence, explaining her own behaviour will no longer be merely a matter of expressing a *current* want and belief. But even then no more is required of her than a sincere expression of her inclinations, only in a conditional form: 'If I hadn't yet succeeded in hitting the window, what would I want to do?' Again, the answer that will explain her action need not draw on any mental events she might remember; she only has to express truthfully her present inclinations to say what under such circumstances she *would* want to achieve and believe to be achievable by the behaviour in question.

However, when it comes to explaining our behaviour months or years ago we are obviously fallible. I may not remember what my wants, interests and plans were a few months ago, so that the reasons I now attribute to myself retrospectively may not be the ones I had at the time (as others remembering my earlier explanations may know). And we are also fallible with respect to the history and development of our long-term intentions and reasons. Past intentions that led to an action over a longer period of time need not be part of the intentional characteristics of the action itself. — Suppose pity with the plight of our local bicycle shop owner gives me the idea of supporting him by buying a bike, even though I never had any thoughts of using a bicycle. But once I decided on buying a bike from him, I naturally started thinking what I could do with it. The idea of weekend bike rides arose and pleased me so much that by the time of the purchase I regard it as the only reason for my action. It may be that I've forgotten the original motive of compassion, but it may also be that I regard it as a change of mind: two months ago I might have bought a bike in order to support the shop, but now I don't feel particularly concerned about the shopkeeper anymore and give only my weekend plans as my reason. The original reason figures in the causal history of my action, but is not the reason for which I acted.

5 The ‘real’ reason

There are various situations in which we are not satisfied with the reasons given by an agent for their action, situations in which we are likely to dig deeper for (what we may be inclined to call) the ‘real’ reason. Does that contradict Wittgenstein’s account of first-person authority about reasons? — Five kinds of cases can be distinguished:

5.1 People can be dishonest about their reasons

It often happens that one suspects a person’s stated reason not to be their real reason, simply because they may not have been entirely honest. Such suspicions are typically based on a perceived inconsistency between the reason stated and the views and preferences the person manifested by their behaviour on other occasions. For example, a ruthlessly selfish businessman’s donations to charity may be suspected to be given merely for prudential reasons, to improve his image, even if he solemnly declares that he only wanted to alleviate suffering. That is straightforward and in no way in conflict with the view that reasons are what agents can *truthfully* present as their reasons.

But is it not possible that even a *sincere* statement of one’s reasons is contradicted by one’s behaviour? Here two kinds of cases need to be distinguished. First, an intentional action and the reasons it involves may, of course, be out of character. Human characters are rarely entirely homogenous. Even a money-grabbing egoist may have moments of altruistic generosity. Secondly, we could try to imagine a more serious inconsistency. Suppose somebody is honestly convinced to act for a reason, a belief, which his surrounding behaviour shows him not to have. For example, Jones turns down the music at 11 pm in order not to disturb the neighbours — although his knowledge that the neighbours are away is shown by the fact that he regularly enters their flat to feed the cat and water the flowers during their absence. Even that would not be particularly uncommon, as the knowledge may have momentarily slipped his mind, so that when turning down the music he was indeed sincerely concerned not to disturb the neighbours. If, however, we assume that the two contradictory beliefs are being manifested more or less simultaneously, the ascription of either becomes problematic. The person’s behaviour, verbal or non-verbal, would just be too confused for our concepts of belief and reason to get a firm foothold. A certain consistency in our verbal and non-verbal expressions of beliefs has to be pre-

supposed for it to be possible to attribute reasons to people, just as it is a precondition of our attributing beliefs to them.

5.2 People can be unclear about their reasons

I may feel unable to give a clear answer to the question why I bought a bike. I may be aware of a number of reasons — considerations speaking in favour of it that I find plausible — without being prepared to single out one of them as decisive. Then it just has to be accepted that there is no single reason why I did it.

5.3 Agential reasons can be (patently) erroneous

As Jonathan Dancy puts it: reasons are non-factive.²⁸ The reason for which I acted may be that p (e.g., that the bicycle I bought was of good quality) — even if it is not in fact the case that p (— I was wrong in thinking that the bicycle was of good quality). In such a case, where the agent is mistaken, or even deluded, our interest may well shift from a reason-giving explanation to a causal explanation. Yet it would be a conflation of categories to say that, in such a case, the unknown cause was the agent's 'real reason'.

For an example of such a conflation, consider the following psychological experiment:²⁹

...passers-by were invited to evaluate items of clothing — four different nightgowns in one study (378 subjects) and four identical pairs of nylon stockings in the other (52 subjects). Subjects were asked to say which article of clothing was the best quality and, when they announced a choice, were asked why they had chosen the article they had. There was a pronounced left-to-right effect, such that the rightmost object in the array was heavily chosen. For the stockings, the effect was quite large, with the right-most stockings being preferred over the leftmost by a factor of almost four to one. When asked about the reasons for their choices, no subject ever mentioned spontaneously the position of the article in the array.³⁰

The psychologists Richard Nisbett and Timothy Wilson interpret this to show that people tend to be ignorant of the real reasons of their actions.³¹ But, as Con-

²⁸ Dancy 2000, 134.

²⁹ I owe this example to Constantine Sandis 2015.

³⁰ Nisbett and Wilson 1977, 233.

³¹ Nisbett and Wilson 1977, 249.

stantine Sandis has pointed out,³² Nisbett and Wilson just fail to distinguish two different concepts: that of an agent's reason — what the agent is prepared to offer as a justification of his action — and that of the causal factors influencing an action.

The conflation is partly due to the fact that the word 'reason' can also be used to denote a cause (rather than an agential reason). Thus it is perfectly acceptable to say:

(5) The reason that so many subjects chose those nylon stockings is that they were placed in the rightmost position.

Just as one can plausibly say:

(6) The reason that he proposed to her that night was that he was drunk.

Here, instead of 'the reason' one could also say: 'what explains the fact'. By contrast, an agential reason would be indicated by prefacing the word with a possessive pronoun (*their* reason, *his* reason). Yet it would take a very odd character to make the following statement true:

(6a) His reason for proposing to her that night was that he was drunk.

It is much more probable that the man's reason for proposing to the lady in question was that he thought her a desirable wife, while the drunkenness just allowed him to overcome his usual shyness; or alternatively, it may have clouded his judgement. In the first case, the drunkenness caused him to act on his reason; in the second case, the drunkenness induced him to have a new (and probably ill-judged) reason (namely, to find the lady suitable as a wife).

Similarly, consumer psychology cunningly identifies causal factors, such as location or packaging, that make people believe certain products to be of superior quality. The reason for their choice, then, is what they mistakenly believe — not the marketing trick that caused them to have such a belief. Again, the following would describe a rather odd scenario:

(5a) Miss Jones's reason for choosing those nylon stockings was that they were placed on the extreme right.

³² Sandis 2015, § 3.

We would be entitled to ask her why she thought that it was preferable to take stockings from the right side of the shelf. (Is she superstitious? Or has she heard of a convention that that is where shopkeepers tend to put the best products?)

In short, what psychologists, such as Nisbett and Wilson, show is not that people are mistaken about their reasons, but simply that they are often ignorant of the causes of their having, and acting on, certain reasons. That what they believe and take as their reason is false — that they have been duped into believing it by factors unknown to them — doesn't mean that it isn't their real reason. It only means that in such a case the reason-giving explanation itself stands in need of a (further) explanation, which is likely to be a causal explanation. For given the falsity of the reason (that in fact the preferred stockings were *not* of better quality), we want to know why the agent held such a false belief.

5.4 The action to be explained was not in fact voluntary.

The most extreme cases of delusional reasons ('confabulation') are those that involve post-hypnotic action. Wittgenstein considers the following case:³³

The experimental subject, who has under hypnosis been given a particular task to execute, e.g. to put up his umbrella, does precisely this — but has no inkling why he does what he does; asked to account for it, he may well invent a motive — 'I wanted to check that the umbrella is in good shape', and believe in it perfectly sincerely; all the same he is mistaken.³⁴

However, what according to Wittgenstein the experimental subject is mistaken about is the *cause* of his action, not his reason.³⁵ The action was not caused by a preceding thought that the umbrella might be defective, yet we can still accept his stated reason as the best justification of his behaviour he can think of. Of course, under the circumstances, we are not much interested in such a justification, since we know that his behaviour wasn't voluntary. What he gave us may still be regarded as *his reason* — but what he did was not *his intentional action*, nothing that could be explained by the agent's reason. What happened here was, in fact, the hypnotist's doing, and so we might be more interested in the hypnotist's reason for it.

³³ It should be noted that in this passage Wittgenstein uses the word '*Motiv*,' translated as 'motive', for 'what somebody specifies as a reason for his action' Wittgenstein *VW*, 424–5.

³⁴ Wittgenstein *VW*, 424–5.

³⁵ Wittgenstein *VW*, 424 f.

(v) Agential reasons may be found insufficient to justify the action (even in light of the agent's views and preferences).

There are certain cases of self-deception where the agential reason given is not false or deluded, but does not appear sufficiently strong to justify the action.

Of course somebody's reasons can appear feeble to us simply because we don't share their views and preferences. I should never regard the concert of a Heavy Metal Band in London as a good reason for going there, but I should not for that matter question others' explanation that that was their reason for going to London taking into account that they may have weird tastes in music. But there are other cases where a stated reason appears insufficiently weighty even by the agent's own ideas and standards.

I may, for example, sincerely give as a reason for my telling someone off his lack of politeness. Yet a careful observer who knows me sufficiently well realizes that the occasion was fairly trivial and that I would never have lost my temper so much, had it not been for the fact that I felt jealous of the person I told off, perhaps without being clear about it myself. Here, again, a reason sincerely offered by the agent is not to be contradicted. The *reason* for my action was indeed the one I gave (always assuming my sincerity, which might require that I reflect carefully on my relevant views and beliefs). However, since my behaviour was less under my control than I believed, my behaviour could not be fully accounted for by my reason. Part of it (the inappropriate vehemence of my outburst) could only be explained causally: by my being carried away by an emotion I hadn't yet fully taken stock of.

The domain of reason-giving explanations is reasonable behaviour. Hence, to the extent to which an action is unreasonable we will look for a *causal* explanation, to supplement, or in some cases even to replace the reason-giving explanation.

6 Conclusion

As argued, following Wittgenstein's ideas, it is a mistake to think that reason-giving explanations are just agents' attempts at causal explanations. The concept of an agential reason is crucially different from that of a cause (— just as a person's avowal of how much they value something must not be misconstrued as their attempted estimate of the thing's objective monetary value). Nevertheless, many reasons can also be said to identify causes. I stopped because the traffic lights turned red. That is my reason, but it is equally correct to say that the red lights caused me to stop. Again, if I develop a desire to own an *Opel*

Insignia this desire may lead (cause) me to buy such a car; yet I may also give as the reason for my purchase that I desired to own such a car.

The upshot of Wittgenstein's account is not that reasons cannot refer to causal factors, but merely that the *concept* of a reason is not the *concept* of a cause. Due to this conceptual difference, there is considerable overlap, but no complete correspondence between reason-giving and causal explanations of an action. That is, for one thing, because not all reasons refer to independently existing psychological occurrences or states.³⁶ The second fact responsible for divergences between the two concepts is the one discussed in this paper, namely that reasons, unlike causes, are covered by first-person authority. It means that although I may be infallible in giving certain facts or circumstances as a reason for my action, I do not enjoy the same infallibility with respect to the causes of my behaviour. Although normally when I stop in front of a red traffic light, the red light provides both the reason and the cause for my stopping, it is conceivable that my stopping had different causes: perhaps I hallucinated the traffic lights or it was something else (say, hypnosis) that caused me to brake. In such a case I would still be correct to give the red lights as my reason for stopping.

But first-person authority is relevant not only in cases of misperception or error. It also allows the agent to single out circumstances as decisive from a justificatory point of view when they play only a minor role, or perhaps no role at all, from a causal perspective; while, on the other hand, an observed circumstance that was causally crucial, getting the action under way, may rightly *not* be regarded by the agent as his reason for acting.³⁷

7 References

- ANSCOMBE, G.E.M., 1957: *Intention*. Cambridge/Mass., Harvard University Press, 1963.
- ANSCOMBE, G.E.M., 1983: 'The Causation of Action', in: M. Geach & L. Gormally (eds), *Human Life, Action, and Ethics*. Exeter, Imprint Academic.
- BRAND, M., 1984, *Intending and Acting*. Cambridge/Mass., MIT Press.
- CHISHOLM, R.M., 1966, 'Freedom and Action', in: K. Lehrer (ed.), *Freedom and Determinism*. New York, Random House, 11–44.
- CLOT-GOUDARD, R., 2015, *L'explication ordinaire des actions humaines*. Montreuil-sous-Bois, Ithaque.
- DANCY, J., 2000, *Practical Reality*. Oxford: Oxford University Press.

³⁶ Rundle 1997, 12, 168; Schroeder 2001, 159–63.

³⁷ I am grateful to Casey Doyle for his comments on this paper.

- DAVIDSON, D., 1963, 'Actions, Reasons, and Causes', in: *Essays on Actions and Events*, Oxford, Clarendon Press, 1980, 3–19.
- DAVIDSON, D., 1973, 'Freedom to Act', in: *Essays on Actions and Events*. Oxford, Clarendon Press, 1980, 63–82.
- GNASSOUNOU, B., 2014, 'Intention et action', in: R. Clot-Goudard (ed.), *L'explication de l'action: Analyses contemporaines*. Paris, Vrin, 57–78.
- GOLDMAN, A.I., 1970, *A Theory of Human Action*. Princeton, Princeton University Press.
- HYMAN, J., 2001, 'Knowledge and Self-knowledge', in: S. Schroeder (ed.), *Wittgenstein and Contemporary Philosophy of Mind*. Basingstoke, Palgrave; 171–193.
- HYMAN, J., 2015, *Action, Knowledge, & Will*. Oxford, Oxford University Press.
- KENNY, A., 1989, *The Metaphysics of Mind*. Oxford, Clarendon Press.
- MELDEN, A.I., 1961, *Free Action*. London, Routledge & Kegan Paul.
- MOYA, C. J., 1990, *The Philosophy of Action*. Cambridge, Polity Press.
- MOYA, C. J., 2014, 'Les aspects causaux de l'explication de l'action', in: R. Clot-Goudard (ed.), *L'explication de l'action : Analyses contemporaines*. Paris, Vrin, 187–206.
- NISBETT, R.E. and T.D. WILSON, 1977, 'Telling More Than We Can Know: Verbal Reports on Mental Processes', *Psychological Review*, Vol. 84, No. 3, 231–259.
- RUNDLE, B., 1997, *Mind in Action*. Oxford, Clarendon Press.
- SANDIS, C. (ed.), 2009, *New Essays on the Explication of Action*. London, Palgrave Macmillan.
- SANDIS, C., 2015, 'Verbal Reports and "Real" Reasons: Confabulation and Conflation', in: *Ethical Theory and Moral Practice*, 18:2, 267–280.
- SCHROEDER, S., 2001, 'Are Reasons Causes? A Wittgensteinian Response to Davidson', in: S. Schroeder (ed.), *Wittgenstein and Contemporary Philosophy of Mind*. Basingstoke: Palgrave, 150–170.
- SCHROEDER, S., 2006, *Wittgenstein: The Way Out of the Fly-Bottle*. Cambridge, Polity.
- STOUT, R., 2010, 'Deviant Causal Chains', in: T. O'Connor & C. Sandis (eds), *A Companion to the Philosophy of Action*. Chichester, Wiley-Blackwell, 159–165.
- THALBERG, I., 1984, 'Do our intentions cause our intentional actions?', *American Phil. Quarterly* 21, 249–260.
- WINCH, P., 1958, *The Idea of a Social Science*. London, Routledge & Kegan Paul.
- WITTGENSTEIN, L., 1958, *The Blue and Brown Books* (BB). Oxford, Blackwell.
- WITTGENSTEIN, L., 1965, *The Principles of Linguistic Philosophy* (PLP), by F. Waismann, based on Wittgenstein's dictations, edited by R. Harré. London, Macmillan.
- WITTGENSTEIN, L., 1993, 'Cause and Effect: Intuitive Awareness' (CE), in: *Philosophical Occasions 1912-1951*, eds: J. Klagge & A. Nordmann. Indianapolis, Hackett, 370–426.
- WITTGENSTEIN, L., 2003, *The Voices of Wittgenstein. The Vienna Circle* (VW), by Ludwig Wittgenstein and Friedrich Waismann, edited by Gordon Baker. London, Routledge.
- WITTGENSTEIN, L., 2009, *Philosophical Investigations* (PI), Fourth Edition. Edited by P. M. S. Hacker and J. Schulte. Translated by G. E. M. Anscombe, P. M. S. Hacker and J. Schulte. Oxford, Wiley-Blackwell.