

Action Explanation and the Unconscious

Edward Harcourt

Any question about action explanations that appeal to the unconscious raises the prior question of what's meant by unconsciousness, and by unconscious mentality. Though a full treatment of the latter questions lies beyond the scope of this entry, some comment on them is needed in order to avoid taking for granted conventional philosophical uses of 'unconscious' as introducing a well-demarcated topic or set of topics. As regards action explanations that appeal to the unconscious themselves (henceforth for brevity 'unconscious action explanations'), the questions addressed here are twofold. First, to what extent are unconscious action explanations of a distinctive kind or kinds; in particular, how if at all do they differ from the kinds of action explanations supplied by conscious factors, and from the kinds of explanation, conscious or unconscious, of doings of ours which are not actions? Secondly, how closely is the phenomenon of irrationality in action connected with the availability of an unconscious explanation?

Many things people do are explained by things about themselves that they don't know – that is, by things about themselves of which they are in a non-technical sense unconscious. A man knocks over a glass thanks to a Parkinsonian twitch without knowing he has the disease; a woman vomits because she is pregnant, though she doesn't know it yet. But these doings of ours do not qualify as actions. Irritably criticizing a friend, however, *is* an action, and might be explained by the

fact that my blood sugar is low, which I don't know. But when philosophers speak of 'unconscious explanations of action' this is often not the kind of case they have in mind. For some, this is because the explanation is not 'Freudian', that is, because (whatever else there is to say about Freudian explanations) it does not refer to the kinds of unconscious mental processes distinctively discussed by Freud. But even among some philosophers who are sceptical about the validity of explanations of *that* kind, there's an unwillingness to treat the explanation in the blood sugar case as an example of an unconscious explanation of action, thanks to an assumption that the term 'unconscious' is properly applied only to *mental* states. Since Descartes defined the mental as the conscious (in a certain specialized sense of that term), and his definition has been highly influential, there is a polemical interest in affirming that there are phenomena (and thus potential explanations of action) that are both unconscious and mental. But now compare 'He irritably criticized his friend because his blood sugar was low' and 'He irritably criticized his friend because he was hungry'. Hunger is a state one can easily be unaware of (for example, because one is so busy being irritated by one's friend) and, at least to the philosopher's ear, it is mental. But is there any reason of principle for distinguishing the kind of explanation involved in the two cases? Arguably not. Moreover the kind of explanation of action offered in both these cases seems the same as the following unconscious mental explanation of a non-action: 'She was tearful because she was suffering from depression'. (I'm assuming a case where, though the subject may consciously feel miserable, (a) she does not know she is depressed because she has not yet had the diagnosis, but (b) the state the diagnosis ascribes, when she gets it, is a mental state.) Indeed the kind of

explanation seems the same as the following unconscious *non*-mental explanation of a non-action: 'She was tearful because she was pregnant (without knowing it)'. From cases of these kinds it may appear that the class of unconscious mental explanations of actions is carved arbitrarily from a much larger class, the class of explanations of things we do which cite facts about ourselves of which we are unaware. (The class is even larger if we do not restrict it to facts about ourselves; cp. 'She was irritable because of the Mistral', whose blowing up she hadn't noticed because she was so absorbed by the object of her irritation.) Since it is very hard to give a tidy non-stipulative definition of the mental, there is all the more reason to expect not to find well-motivated distinctions within the class of facts about ourselves of which we're unaware when it comes to explaining behaviour.

What the above cases have in common, however, is that the explanations all explain without providing a reason: that is why it is so easy to be indifferent as to whether the behaviour explained is a non-action (doings of ours for which there are never reasons) or an action, and as to whether the *explanans* is mental or not (because though some mental *explanantia* are rationalizing, those in the examples are not). But there does seem to be a subclass of explanations of actions which readily leap to mind when 'unconscious explanations of action' are mentioned, and which are of special interest. A neat way to introduce this subclass is as the subclass of cases where the *explanans* is a state of a kind apt to provide a rationalizing explanation (whether or not it does so in the particular case), and which is unconscious. Since propositional attitudes are widely thought to be the only states of this kind, we have a rationale for singling out

explanations which invoke unconscious propositional attitudes for special attention. It should be noted, however, that this neat way of introducing the interesting subclass depends on another issue that lies beyond the scope of this entry, the issue as to what kinds of thing are reasons for action. According to an alternative view on this issue, any old facts can be reasons ('Why did you do that?' 'The weather', 'the look on your face'). So there is no kind of *explanans* which is rationalizing or nonrationalizing *per se*. Moreover, the role unconsciousness plays in action explanation cannot reliably be put down to the unconsciousness of an explaining mental state (since there need not be one), and will have to rest instead on something like the agent's (conditional) inability to give the explanation.

However, factoring this controversy into what follows would make for excessive complication, so I am going to set the more liberal view of reasons on one side. And in any case, whether or not mental states that are apt to rationalize actions are the only things that are apt to rationalize them, advocates of the more liberal view of reasons need not deny that there can be unconscious desires, beliefs and so on. So it is worth asking what further features, if any, such states have in common. There is at least one further feature and which they don't share with non-mental causes either of actions or of non-actions, or even (as the depression example above might be taken to show) with every mental cause of either. For though depression is a mental state, ascriptions of it do not display first/third person asymmetry, either in the strong sense that we are authoritative about whether we have it (we aren't), or in the weak sense that self-ascriptions of it have a special status in confirming other-ascriptions of the state to oneself: I may

just come to accept I am depressed by accepting the expert diagnoses of those who have evidence that I am. In the case of propositional attitudes, by contrast, my self-ascriptions do have a special status in confirming the ascription of the state to me (though of course I can be insincere, self-deceived and so on). By the same token it is doubtful whether I could count as *consciously* (e.g.) intending something if my *only* reasons for thinking I did were indirect (for example that others who had been behaving in the same way as me had all accepted that they had the intention in question). It may thus be said that there are (at least) two senses in which a state of mine may be unconscious, one where self-ascription has a special status in relation to the ascribability of the state to me, the other where it does not. We can call these characteristics avowability and non-avowability respectively. It's the apparent fact that only mental states are avowable in the sense explained that further fuels the common assumption, noted earlier, that 'unconscious' *means* 'unconscious and mental'.

Before leaving the subject of how unconsciousness is to be defined, it's appropriate to mention one more distinction commonly applied to unconscious mental states, sometimes termed the 'procedural' versus the 'dynamic' unconscious. According to this distinction, states that are procedurally unconscious are unconscious (roughly) because they don't need to be conscious (for example, my belief that there is a hole in the road which explains my swerving to avoid it, though when asked about it I say 'what hole?'); dynamically unconscious states are unconscious (roughly) because they need not to be conscious (for example any state that is unconscious as a result of repression). But things are more complicated than this distinction allows. Consider a case of

what we might call simple lack of self-knowledge, as when I am upset about something but haven't realized it yet. At a given time I am not disposed to *assert* that I am upset (despite behaving as if I am), so my state is, at that time, unconscious. But when it is put to me that I am upset, I agree readily and find it helpful to be able to put my state into words, so prior to that the state wasn't repressed. So was it procedurally rather than dynamically unconscious? In the example, being upset is an avowable state in the sense previously explained: my ready agreement has a special status in confirming the suggestion that I am indeed upset. But this is not the case with my belief about the hole in the road: my sole reason for accepting that I had the belief may be that there was a hole, I was in a position to form the belief that there was, and I swerved. ('I suppose I must have believed there was one.') One response to this would be to deny that the state I was in was an unconscious mental state at all. To see why the latter seems too radical, contrast yet another case. There is evidence that male preference for female faces tracks female pupil size, though subjects report no awareness of differences of pupil size (Rey, 1998, p. 523). Here again, then, subjects would accept that they were aware of differences of pupil size, if they accepted it at all, solely on the basis of indirect evidence. The pressure to deny that this awareness is an unconscious mental state is strong: the supposed unconscious state doesn't explain in the way that state would if it were conscious, since conscious awareness of pupil size would have no bearing on preference. In the light of this example, consider again the 'hole' case. Here, the state I was in *does* explain my swerving in the way a conscious belief ('Oh look, there's a hole') would have done, i.e. thanks to the rational relation it bears to the way things were, my wants and so on. That is why it is going too far to deny that

it is a mental state. But there is still the interesting fact that, like pupil-size awareness, the state is non-avowable. So in place of the two-way procedural/dynamic distinction, we appear to need a three-way distinction between those unconscious mental states that are neither repressed nor avowable, those that are avowable but not repressed, and those that are avowable and repressed.

With these clarifications out of the way, I want to focus on the interesting subclass I picked out before, namely the subclass of unconscious explanations of action that appeal to mental states that are both avowable and apt to rationalize the actions they explain.

One question that immediately arises in this connection is whether such states do in fact rationalize the actions they explain when the states are unconscious. According to one line of thought, there is no difference in the kind of explanation provided by propositional attitudes when they are unconscious from the kind they provide when conscious. In some cases, at least, this seems right. Suppose that love-rivalry with another suitor rationalizes my monopolizing a woman's attention (I love her, I don't want her to fall for him, so I form the intention of keeping them apart). My behaviour would, it seems, be equally rational relative to my own state of mind and the lie of the land if I did not know I was in love, and instead offered a self-deceiving explanation for my monopolizing behaviour (e.g. that if I didn't set out the argument of my very long book all in one go, she wouldn't be able to form a proper opinion of it). This is not to say that this is a good wooing tactic, but it's what I would have done (rightly or wrongly) even if I had not been self-deceived. Indeed there are cases where my desires are more

likely to be fulfilled if motive and intention are unconscious than if they're conscious. Thus suppose I am in love with someone and want to become close to them but am so painfully shy that were I to be conscious that this is what I wanted, I would become inhibited, and so fail. By self-deceivingly describing my conversations with the object of my desire as 'routine discussions with a colleague' or 'just passing the time of day', I may become close to them before I realize what I'm doing, and so succeed. Or, to take a case reported by Freud, a Jewish convert to Christianity went on a family holiday to friends who did not know his background. When his hostess began to make anti-Semitic remarks, he felt that he ought to declare it, but at the same time was afraid of making a scene, so he said nothing – or perhaps 'said nothing'. For, though he meant to say 'run along to the garden, you boys [*Ihr Jungen*]' to his sons, what came out was 'run along to the garden, you Jews [*Ihr Juden*]' (Freud 2002, pp. 87-8). On one reading of the episode, the slip was motivated by an unconscious desire on the speaker's part to communicate his Jewish origins. What is more, it is the desire which, had the speaker not akratically yielded to his fear of making a scene, he would have acted on. Both examples remind us that our rationality may be compromised in more than one way at a time (shyness and self-deception; weak will and (I guess) a further specialized form of incontinence with respect to utterance), and that one defect can correct for another.

From this it does not follow, however, that explanations of action that appeal to unconscious propositional attitudes are *always* rationalizing explanations. In a recent Preface to his *The Unconscious*, Alasdair MacIntyre emphasizes the distinction between acting for reasons and 'acting *as if* unconsciously guided by

reasons'. In the latter case, 'the motives that control [one's] behaviour preclude one from acting as a practical reasoner does', since it's constitutive of the notion of a reason that one be capable of asking whether it's a *good* reason for one's action. But where our motivations 'give expression to infantile wishes and early traumas' we are unable to do this (MacIntyre 2004, pp. 25-6). Insofar as this is so, the exercise of our rational powers is inhibited or subverted, so there are cases where, though the *explanantia* are of a kind that is apt to provide reasons, the explanation itself displays a defect in our rationality. Interestingly, however, this defect is not, on MacIntyre's account, due to the unconsciousness of the desires, emotions etc. *per se*: they explain action in the way 'an attack of panic terror' does, panic terror being a state in which we are also not capable of evaluating our reasons for acting, but where the explaining state is fully conscious. Here we may compare Davidson on akrasia: 'the standard case of akrasia is the one where the agent knows what he is doing, and why, and knows that it is not for the best, and knows why' (Davidson 1982, p. 304). On this type of account, then, the connection between the impaired rationality of the action and the unconsciousness of the states appealed to in explaining it is very loose. Actions can apparently be fully rational but unconsciously explained, and they can be irrational for the same sorts of reason whether the explaining states are unconscious or not.

Another type of explanation which cites unconscious states that are apt to rationalize, but don't in the particular case, is exemplified by some other Freudian cases of 'symptomatic error': 'Frau F. said, of her first lesson in a language course ...: "It's really interesting, the tutor is a nice young Englishman.

In the very first lesson he indicated to me *durch die Bluse* [through my blouse]” – and corrects herself: “*durch die Blume* [lit, ‘through the flower but colloquially ‘with veiled hints’] – that he would rather give me private tuition” (Freud 2002, p. 78). Though desires are states of a kind apt to provide rationalizing explanations, there is no unconscious communicative intention on Frau F.’s part in this case: her attraction to the tutor simply betrays itself involuntarily in her utterance, much as emotions betray themselves in gestures and facial expressions. Here, as with the essentially non-rationalizing causes of action cited earlier, we find a pattern of unconscious explanation of action that’s common to the explanation of non-actions, and indeed to the explanation of non-actions by *conscious* states (I can express amusement by smiling involuntarily, but know perfectly well that I am smiling because I am amused).

Equally indebted to psychoanalysis, however, is a different model of the unconscious explanation of action (Gardner, 1993) that insists on its difference from rationalizing explanation; discerns patterns of explanation common to actions (e.g. obsessional behaviour) and non-actions (e.g. dreams); draws on the notion of expression and of expressive behaviour; but tries to draw the connection between the irrationality of the *explananda* and the unconsciousness of the *explanantia* much more tightly than in any of the models reviewed so far. Does this model compete for logical space with the others? No: though Gardner claims that if the correct explanation of the action is ‘unconscious rationality’ (as in the monopolizing suitor case), we have the operations not of the unconscious but of the ‘preconscious’, the disagreement here is surely over the word ‘unconscious’ rather than over the viability of the other models of explanation.

As an example of this, most exigent, model of unconscious explanation, consider Freud's case of a girl's obsessive bedtime ritual, which involved arranging her bedding in a particular way (Freud 1952, pp. 224-7). Notice first of all that the behaviour consists of a series of intentional actions (e.g. arranging the bolster so it does not touch the end of the bedstead). It's also irrational (like the man in Anscombe's *Intention* filling a saucer with mud 'for no reason'), since the agent's reasons for so acting give out immediately, and yet the action - unlike whistling or running one's fingers through one's hair - does not remotely make sense *without* a further reason. It's natural, then, to look for further unconscious reasons of the kinds canvassed before: beliefs, desires and so on which, though unconscious, either fully rationalize the action or at least display the action as a case of action *as if* for reasons, though in fact a case of impaired rationality because of the agent's incapacity to evaluate them. Indeed reasons of this kind are ready to hand: the girl's ritual (which prevented her parents from going to bed till it was over) might be seen as manifesting her unconscious intention to prevent her parents from having sexual intercourse, thereby averting an outcome she (unconsciously) very much does not want, namely a sibling who would be her rival. Gardner concedes that rationalizing factors (including perhaps pleasure in the ritual) may help to perpetuate the obsessional behaviour and make it hard to shift. However, its primary explanation, though dependent on the content of unconscious states and so not merely mechanical, is not rationalizing but depends on (i) unconscious phantasy, (ii) expression, and (iii) symbolism. The girl is motivated by a wish-fulfilling phantasy of undoing her mother's pregnancy, which expresses itself in intentional action in the way anger may express itself in smashing a glass, but symbolically, via the details of the

ritual (e.g. fluffing the eiderdown in such a way that it makes a hump, then smoothing it again = symbolically undoing pregnancy). The ritual is thus irrational not only because of the inadequacy of the reasons the girl herself is able to offer, but also because the mental states that explain it are themselves irrational (though representational, they are 'not subject to the requirements of rationality that govern belief, desire and preference' (Gardner 1993, p. 116)). It's not that the girl unconsciously wants her mother to cease to be pregnant and believes, unconsciously and falsely, that wishing for this state of affairs will bring it about (which would make her wish instrumentally rational relative to her false belief): it's rather in the nature of phantasy to function as if wishing something makes it so. But states of this kind are also, in their nature, minimally accessible to consciousness (Gardner 1993, p. 89): thus the special irrationality of the action and the unconscious character of its explanation don't just go together by accident.

This psychoanalytic account of irrational action and its unconscious explanation draws attention to a feature of explanation by intentional states overlooked by the previous accounts of rationalizing explanation, namely that even when explaining actions by appeal to *conscious* emotions 'makes sense' of them (e.g. smashing the glass in anger), it need not do so by representing the action as practically rational relative to the agent's desires, beliefs etc.: the action may simply express the subject's state. However, there is room for controversy about the view of unconscious mental states in play on this account, both with respect to the states' content and with respect to their alleged inaccessibility to consciousness. On a psychoanalytical model of mental structure emphasized by

Bion, Winnicott and others, ego- or self-formation is favoured by the availability of a mirroring or containing other, who articulates the growing child's thoughts and experiences to it in a way it can grasp and make its own; from this pattern of nurture the child gradually acquires the capacity to 'contain' its own thoughts and experiences (Shuttleworth 2002, p. 36). Perhaps because they are traumatic, or because of the absence of a containing other, or both, these contents may however make themselves felt in other ways, for example, through symptom-formation, 'concrete thinking' (Dubinsky 1997, p. 8) or 'acting out'. When this is so irrationality results, since the agent's incapacity to articulate the contents to him- or herself debars them from interacting in the right way with the rest of his or her beliefs, desires and so on. But there is nothing about the nature of the states themselves that precludes them from interacting with the subject's other mental states in the right way. Insofar as it aims to help the subject 'contain' these contents – or, in Jonathan Lear's words, converts 'physical throwings-up into verbal throwings-up [of meanings]' (Lear 1999) – psychoanalysis sees itself as restoring or enhancing the subject's rational powers.

EDWARD HARCOURT

References

Davidson, D. (1982). Paradoxes of irrationality. In R. Wollheim and J. Hopkins (Eds.), *Philosophical Essays on Freud*. Cambridge: CUP.

- Dubinsky, A. (1997). Theoretical overview'. In M. Rustin et al. (Eds.), *Psychotic States in Children*. London: Duckworth.
- Freud, S. (1952). *Introductory Lectures on Psycho-analysis*. (J. Riviere, Trans.). London: George Allen & Unwin.
- Freud, S. (2002). *The Psychopathology of Everyday Life*. (Anthea Bell, Trans.). London: Penguin.
- Gardner, S. (1993). *Irrationality and the Philosophy of Psychoanalysis*. Cambridge: CUP.
- Lear, Jonathan (1999), 'Love and Its Place in Nature', www.pschoanalysis.org.uk/lear.htm
- MacIntyre, A.C. (2004). *The Unconscious* (revised edn.). London: Routledge.
- Rey, Georges (1998). Unconscious Mental States. In E.J. Craig (ed.), *The Routledge Encyclopedia of Philosophy*, vol. IX. London: Routledge.
- Shuttleworth, K. (2002). Psychoanalytic Theory and Infant Development'. In L. Miller et al. (Eds.), *Closely Observed Infants*. London: Duckworth.