

## CHAPTER 26

- Reading 26.1 **Spence, S.A. (1996a). Free will in the light of neuropsychiatry. *Philosophy Psychiatry, & Psychology*, 3: 75-90 (Extracts pp. 78-81, 83-5).**
- Reading 26.2 **Ryle, G. (1963). *The Concept of Mind*. London: Penguin, (Extract pp. 62-6).**
- Reading 26.3 **From Davidson, D. (1980). Actions, reasons and causes. In *Essays on Actions and Events*. Oxford: Oxford University Press, pp. 3-197. (Extract pp. 11-17).**
- Reading 26.4 **Davidson, D. (1982). Paradoxes of irrationality. In *Philosophical Essays on Freud* (ed. J. Hopkins and R. Wollheim). Cambridge: Cambridge University Press, pp. 289-305. (Extract pp. 296-8).**
- Reading 26.5 **Tanney, J. (1995). Why reasons may not be causes. *Mind and Language*, 10: 105-128 (Extracts pp. 108-110, 111-2, 113)**
- Reading 26.6 **Strawson, P.F. (1974). Freedom and resentment. In *Freedom and Resentment, and Other Essays*. London: Methuen, pp. 1-25 (Extracts pp. 6-9, 10-11).**

**Reading 26.1****EXERCISE 1**

From: Spence, S.A. (1996a). Free will in the light of neuropsychiatry. *Philosophy Psychiatry, & Psychology*, 3: 75-90 (Extracts pp. 78-81, 83-5).

**Extract 1****The Experience of Free Will: The Normal Case**

Under normal conditions many acts may be performed without undue attention or conscious processing being devoted to them. A common example is driving a car while talking or listening to music. Under these conditions behavior is clearly 'automatic' to some extent and may be accompanied by very little subjective volition. Indeed it is not uncommon to reach the end of such a process without any conscious awareness of the preceding interval (e.g., the details of a familiar journey).

When activity is complex or novel it receives greater attention, and is more likely to be remembered (Frith 1992). Under these circumstances the sensory aspects of the activity are more likely to be attended to, and awareness of one's own volition increasingly apparent (e.g., the first faltering steps of learning to drive or to play a musical instrument). In this normal situation it can be difficult to distinguish the sensory experience of freedom from the associated experiential belief in the same, but as stated before, there exists a distinction between motor and sensory perceptual phenomena e.g., the sequence of finger movements on a keyboard, and the thoughts (beliefs) which address these phenomena.

When anxious an individual may become acutely aware of his own actions, interrupting the flow of normally fluent activities e.g., walking and speech. In such a condition the experience (perception) of voluntary activity is heightened, but there is not necessarily any change in the experiential belief in the freedom of a specific act.

In normal individuals engaged in creative activities there is frequently a phenomenological account which describes altered volition. Such cases have been documented elsewhere by Boden (1992) as evidence for non-conscious processing producing novel computations in response to creative problems. These solutions subsequently access consciousness. The following descriptions taken from different sources speak of an alteration in the phenomenology of volition.

A1. 'Sometimes when I'm singing there's no sensation in my throat. It feels as if some outside influence has taken over. The air is coming through me—as if I'm being sung like a flute.' Sheila Chandra (quoted in 'The Wire', May 1994)

A2. 'I just play what comes into my mind.' Thelonious Monk (quoted in *Jazz* at Ronnie Scott's, 1979)

A3. 'So . . . I did one of those things you do and never find out why, even though something in you knows you could never have done anything else.' William Gibson (1993)

A4. 'When you blurt something out and haven't censored yourself, your first instinct is to cover up because it doesn't make sense, rather than just letting it be there. It's a mistake to censor those kind of irrational utterances. They are often the deepest things you say.' David Byrne (*The Wire*, 1994)

A5. 'How can I tell what I think until I see what I say?' E. M. Forster (1960)

A6. 'It was late before the two guests left and Russell was alone with Lady Ottoline. They sat talking over the fire until four in the morning. Russell, recording the event a few days later, wrote, 'I did not know I loved you till I heard myself telling you so—for one instant I thought 'Good God, what have I said?' and then I knew it was the truth.'" Clark (1975) quoted by Dennett, 1991, 246

A7. 'I was sitting in a chair in the Patent Office at Berne when all of a sudden a thought occurred to me: 'If a person falls freely he will not feel his own weight.' I was startled. This simple thought made a deep impression on me. It impelled me toward a theory of gravitation.' Albert Einstein, quoted in Highfield and Carter (1993)

In these accounts from normal individuals there are circumstances described in which the subjective experience of voluntary action is diminished or lost, and experiential belief is altered. They comprise altered motor performance e.g., singing, playing the piano, speaking, and also, in the case of Einstein, insightful thought (or 'inner speech'). None of the individuals attribute the phenomenon to an outside force, although, in the case of A1 there is clearly a metaphorical allusion to such passivity ('as if I'm being sung like a flute'). However, despite the obvious fact that they are all physically responsible for their respective acts they do not experience as freely willed action the impulses which produced these acts.

Thus, in 'normals' it appears possible for the experience of free will to be heightened, diminished or lost, and for the experiential belief in free will (pertaining to a specific act) to be altered (but nevertheless retained). Any allusion to outside forces remains metaphorical and does not reach delusional intensity.

**Pathological Cases**

A variety of pathologies underlying neuropsychiatric illnesses may disturb 'willed' action. From these 'natural experiments' it may be possible to infer aspects of normal functional organization. In some cases the pathology may be so devastating as to leave little prospect of willed movement, e.g., the limbs affected by dense hemiparesis, or the 'off' state of Parkinson's disease. Similarly in severe psychiatric illness the disturbance may be so profound as to offer little prospect of teasing apart aspects of volition e.g., the depressed individual with both psychomotor retardation and nihilistic delusions: 'I am dead; I have no feelings and no will' (Sims 1988, 235).

In such a case we may clearly hypothesize that mechanisms of willed action and their representation in the subject's consciousness are disturbed. But is this disturbance one of higher processing, the abstract belief in freedom, or is it perhaps 'lower,' related in

some part to an experience of impeded action, a failure to initiate volition at some level?

In an attempt to dissect volition some specific examples will now be addressed.

## 1. The Alien Hand Syndrome

The 'alien hand' syndrome is a rare but potentially informative example of a pathological process affecting willed movement. Initially the term was used to describe the inability of the affected subject to identify their (affected) limb (Brion and Jedynak 1972). However, in later descriptions the emphasis shifted to the subject's inability to control the latter (Gasquoine 1993). The limb is experienced as being autonomous. The syndrome is usually found in association with lesions of the medial frontal lobe (contralateral to the affected limb) and anterior corpus callosum (Gasquoine 1993). (However, it is common for the pathology e.g., cerebral infarction or tumor to encroach on neighboring structures.)

The striking phenomenology encountered has been reported by a number of authors:

B1. '[A]t one point it was noted that the patient had picked up a pencil and had begun scribbling with the [affected] right hand. When her attention was directed to this activity, she reacted with dismay, immediately withdrew the pencil, and pulled the right hand to her side using the [normal] left hand. She then indicated that she had not herself initiated the original action of the right arm. She often reacted with dismay and frustration at her inability to prevent these unintended movements of the right arm. She experienced a feeling of dissociation from the actions of the right arm, stating on several occasions that "it will not do what I want it to do."' (63 year old right handed female with left medial frontal infarction; Goldberg et al 1981)

B2. 'He viewed these (autonomous) actions as unwanted, unintentional, and uncontrollable. He described the [affected] right hand as 'the bad one, it has a mind of its own,' and that it was 'always trying to get into the act.' A very specific complaint, which was chronic and persisted for a year after the infarction, was that the patient felt the right hand anticipated future actions and performed movements prior to the patient actually intending them. The [normal] left hand did not show this tendency?' (68 year old right handed male with left medial frontal and callosal infarction; Feinberg et al 1992)

B3. 'Astonishment and frustration were expressed regarding her inability to voluntarily release her grasp of objects in the [affected] left hand. She would attempt to restrain the unwanted movements of the left hand by keeping her hands folded together or by gripping an object in the left hand. The left hand would perform activities with objects around her in a compulsive way that was not associated with a feeling of voluntary control. . . . Her left hand was noted to wander around in the 'alien mode' when not restrained. . . . Another incident occurred when the [normal] right hand picked up a bowl of hot soup and the left

hand threw it to the ground. On another occasion, her left hand began to remove a cigarette from her mouth as she was about to light it with the right hand. She stated that the left hand 'was trying to keep me from smoking.'" (53 year old right handed female with right medial frontal lobe and corpus callosum infarction; Goldberg and Bloom, 1990)

B4. 'She often spoke to her own [affected] left hand, asking it to perform some movement, but the hand 'did only what it wanted to.' In bimanual tasks (e.g., lighting a match) she tried to use only her [normal] right hand but was regularly impeded by her left hand's groping. At that time she experienced the lack of ownership of her left hand and could not indicate her own among others' hands..' (65 year old right handed female with right medial frontal infarction extending to anterior cingulate gyrus, supplementary motor area and anterior fibres of corpus callosum; Trojano et al 1993)

B5. '[S]he noticed that her [affected] left arm began to move 'on its own.' The left hand would tenaciously grope for and grasp any nearby object, pick and pull at her clothes, and even grasp her throat during sleep. These autonomous left arm and hand movements tended to occur in bouts usually signaled when her hand did 'silly things' such as scratching her knee. She slept with the arm tied to prevent nocturnal misbehaviour. She never denied that her left arm and hand belonged to her, although she did refer to her limb as if it were an autonomous entity.' (39 year old right handed female with trauma to medial frontal white matter bilaterally, corpus callosum, right basal ganglia, internal capsule and thalamus; Banks et al 1989)

B6. 'Sometimes the [affected] left hand would pick up a cup and move it into the path of the [unaffected] right hand so as to interfere with eating. While playing checkers on one occasion, the left hand made a move that he did not wish to make, and he corrected the move with the right hand; however, the left hand, to the patient's frustration, repeated the false move. On other occasions, he turned the pages of a book with one hand while the other one tried to close it; he shaved with the right hand while the left one unzipped his jacket; he tried to soap a washcloth while the left hand kept putting the soap back in the dish; and he tried to open a closet with the right hand while the left one closed it.' (40 year old right handed male with ruptured anterior communicating aneurysm, and subsequent resection of right frontal gyrus rectus; Banks et al 1989)

As may be seen from these accounts the alien hand is experienced as behaving in an autonomous and purposeful way. Its interference with normal activity is such as to provoke patients to developing strategies to restrain or 'distract' the hand. The patients acknowledge that the limb belongs to them but that its actions are not under their control. Yet they do not attribute its activity to outside forces (cf. passivity, below). Their appreciation of their predicament suggests that the subjects retain a sense of their own will, in that they form action plans (which fail). There is a circumscribed inability to impose that will upon one domain of their body/action, i.e., the affected limb.

Thus it is clear that the experience of free will may be absent with regard to part of a subject's anatomy and repertoire of voluntary behavior (limb movement). This is particularly noticeable in those examples where these activities occur during sleep (e.g., B5). Experiential belief relating to the action of the 'alien' limb is clearly disturbed. Once the aberrant activity of the limb is noticed the patient tries to exert conscious control but cannot. The limb is experienced then as having a 'mind of its own.' This belief is not delusional since it retains an 'as if' quality, and is clearly supported by the subjectively and objectively unusual behavior of the limb.

The subject loses the experience of free will (with respect to the affected limb), and develops an accommodating abnormal experiential belief that the limb is autonomous.

That these subjects have sustained lesions of their medial frontal lobe and anterior corpus callosum is of interest, given the role of these areas in the genesis of willed action (see 'Temporal Aspects of Action', below) and information transfer between the hemispheres. We will return to the issue of the timing of initiation of activity, but for the present it is interesting to note example B2 in which the subject experienced his right hand as anticipating his intentions. In this case the autonomy of the affected limb is further extended (to anticipating the will of the subject).

What does the alien hand tell us of free will? It suggests that the sense of free will accruing to action may have a neurological component which is discrete from (but related to) the neurological correlates of action itself; in other words that the initiation of action and the 'labelling' of that action as 'intentional' ('willed' or 'belonging to the subject') are potentially separate.

## Extract 2

### The Temporal Aspects of Action

Does a meaningful concept of free will require that we may, at least on some occasions, consciously choose a certain course of action? Must choice and decision be conscious? If so, what temporal relationship must pertain between conscious awareness of intent and the initiation of action?

As alluded to in the introduction, there is an assumption embedded within the concept of free will that conscious choice may occur (or appear to occur) on at least some occasions. For this to be a real choice, however, it is here argued that the choice made, e.g., a decision to act, would of necessity be prior to the initiation of the specified act. Initiation would 'need' to follow decision. By decision here we refer to the conscious 'making' of a decision to act. Is this sequence of events a biological possibility?

For the sake of example I will assume that observable motor acts provide a model for all acts, mental or physical. Although this is an assumption I hope to justify it in due course.

Voluntary finger movement in the human is preceded by a well documented pattern of cerebral electrical activity measured from surface electrodes (Kornhuber and Deecke 1965; Deecke, Scheid

and Kornhuber 1969; Libet et al 1982). There are 3 potentials identifiable using electroencephalography (EEG):

1. Readiness Potential (RP): a slowly increasing, negative potential occurring 850 msec prior to voluntary movement (as recorded by EMG). The RP is bilateral, symmetrical, and seen over pre- and post-central regions, being maximal at the vertex.
2. Pre-motion Positivity (PMP): a symmetrical positive potential seen 86 msec prior to voluntary movement.
3. Surface negative motor potential: maximal over the contralateral, pre-central hand area, 56 msec prior to voluntary movement (Deecke, Scheid and Kornhuber 1969).

Thus the early stages of the initiation of a motor act precede its objective execution by approximately 850 msec. At what point does the subject become subjectively aware of the 'decision' to move? If the voluntary decision is itself initiating the movement, then one would intuitively suggest that the former precede the latter in time. (One would then need to explain how a 'thought' might 'produce' such an action.) If, however, the neural substrates of action precede the subjective decision to act, then clearly the latter cannot be causal to the former (at least in terms of our current understanding of causality), i.e. conscious 'decisions' do not cause acts.

Libet et al. (1983) addressed the above issue by requiring subjects to note when they decided to act in a movement paradigm. Subjects observed an oscilloscope screen on which a spatial clock was represented. While their EEG and EMG data were recorded, the subjects noted the 'time' according to the oscilloscope clock at which they 'decided' to move. Their RPs began approximately 350 milliseconds prior to their subjective 'decision.' Thus Libet et al. conclude that volitional activity is initiated non-consciously (Libet et al., 1983). [See fig. 1]

If the findings of Libet and co-workers are replicated, then we must conclude that 'decisions' to act arise prior to our conscious awareness of them. Thus our 'decision' or 'freedom' is illusory (if by these terms we mean conscious phenomena). It occurs after activity has been initiated. Libet et al. (1983) wish to retain free will, however, when they suggest that the subject may still have the power of veto over arising intentions. Thus an act may be planned, and initiated out of consciousness, but might still be rejected when entering the latter. Could this be so? I will suggest that in the light of Libet's own work this is unlikely to yield 'true' freedom.

When the somatosensory cortex is directly stimulated, at liminal intensity, stimulus train durations of about 500ms are necessary to elicit any reportable conscious sensation. Stimuli below that intensity or train duration produce no sensation at all (Libet, 1993 for review). Similarly an afferent sensory stimulus may be altered (by direct cerebral stimulation) for up to 500ms prior to its subjective appearance' (Libet 1993). Libet suggests that a period of 'neuronal adequacy' of approximately 400–500 ms is required for neuronal activity to yield a conscious sensation.

For the purposes of this discussion, it does not matter what duration in time, or magnitude of stimulus, is required for such a sensation to 'arise.' The important point is that a finite period of neuronal activity precedes conscious awareness of the latter. Neuronal events precede subjective phenomena and hence phenomenology.

Thus, to suggest as Libet does that free will is still sustainable, as a power of veto, is flawed. It is flawed because the 'decision' (in consciousness) to act or not to act is itself the 'result' of preceding neural activity.

The decision does not make itself. It is a product or correlate of preceding neuronal events.

Similar conclusions have been reached by Velmans in his analysis of the literature concerning human information processing (Velmans 1991). The weight of the evidence derived from cognitive neuroscience points to sophisticated analysis of both incoming sensory information and ongoing volitional activity 'taking place' out of consciousness. '[O]ne becomes aware of a stimulus only after one has analysed and selected it, and aware of one's own response only after one has executed it' (Velmans 1991, 666).

## Reading 26.2

## EXERCISE 3

From: Ryle, G. (1963). *The Concept of Mind*. London: Penguin, (Extract pp. 62–6).

Volitions have been postulated as special acts, or operations, 'in the mind'; by means of which a mind gets its ideas translated into facts. I think of some state of affairs which I wish to come into existence in the physical world, but, as my thinking and wishing are unexecutive, they require the mediation of a further executive mental process. So I perform a volition which somehow puts my muscles into action. Only when a bodily movement has issued from such a volition can I merit praise or blame for what my hand or tongue has done.

It will be clear why I reject this story. It is just an inevitable extension of the myth of the ghost in the machine. It assumes that there are mental states and processes enjoying one sort of existence, and bodily states and processes enjoying another. An occurrence on the one stage is never numerically identical with an occurrence on the other. So, to say that a person pulled the trigger intentionally is to express at least a conjunctive proposition, asserting the occurrence of one act on the physical stage and another on the mental stage; and, according to most versions of the myth, it is to express a causal proposition, asserting that the bodily act of pulling the trigger was the effect of a mental act of willing to pull the trigger.

According to the theory, the workings of the body are motions of matter in space. The causes of these motions must then be *either* other motions of matter in space *or*, in the privileged case of human beings, thrusts of another kind. In some ways which must forever remain a mystery, mental thrusts, which are not movements of matter in space, can cause muscles to contract. To describe a man as intentionally pulling the trigger is to state that such a mental thrust did cause the contraction of the muscles of his finger. So the language of 'volitions' is the language of the para-mechanical theory of the mind. If a theorist speaks without qualms of 'volitions', or 'acts of will', no further evidence is needed to show that he swallows whole the dogma that a mind is a secondary field of special causes. It can be predicted that he will correspondingly speak of bodily actions as 'expressions' of mental processes. He is likely also to speak glibly of 'experiences', a plural noun commonly used to denote the postulated non-physical episodes which constitute the shadow-drama on the ghostly boards of the mental stage.

The first objection to the doctrine that overt actions, to which we ascribe intelligence-predicates, are results of counterpart hidden operations of willing is this. Despite the fact that theorists have, since the Stoics and Saint Augustine, recommended us to describe our conduct in this way, no one, save to endorse the theory, ever describes his own conduct, or that of his acquaintances, in the recommended idioms. No one ever says such things as that at 10 a.m. he was occupied in willing this or that, or that he performed five quick and easy volitions and two slow

and difficult volitions between midday and lunch-time. An accused person may admit or deny that he did something, or that he did it on purpose, but he never admits or denies having willed. Nor do the judge and jury require to be satisfied by evidence, which in the nature of the case could never be adduced, that a volition preceded the pulling of the trigger. Novelists describe the actions, remarks, gestures, and grimaces, the daydreams, deliberations, qualms, and embarrassments of their characters; but they never mention their volitions. They would not know what to say about them.

By what sorts of predicates should they be described? Can they be sudden or gradual, strong or weak, difficult or easy, enjoyable or disagreeable? Can they be accelerated, decelerated, interrupted, or suspended? Can people be efficient or inefficient at them? Can we take lessons in executing them? Are they fatiguing or distracting? Can I do two or seven of them synchronously? Can I remember executing them? Can I execute them, while thinking of other things, or while dreaming? Can they become habitual? Can I forget how to do them? Can I mistakenly believe that I have executed one, when I have not, or that I have not executed one, when I have? At which moment was the boy going through a volition to take the high dive? When he set foot on the ladder? When he took his first deep breath? When he counted off 'One, two, three—Go', but did not go? Very, very shortly before he sprang? What would his own answer be to those questions?

Champions of the doctrine maintain, of course, that the enactment of volitions is asserted by implication, whenever an overt act is described as intentional, voluntary, culpable, or meritorious; they assert too that any person is not merely able but bound to know that he is willing when he is doing so, since volitions are defined as a species of conscious process. So if ordinary men and women fail to mention their volitions in their descriptions of their own behaviour, this must be due to their being untrained in the dictions appropriate to the description of their inner, as distinct from their overt, behaviour. However, when a champion of the doctrine is himself asked how long ago he executed his last volition, or how many acts of will he executes in, say, reciting 'Little Miss Muffet' backwards, he is apt to confess to finding difficulties in giving the answer, though these difficulties should not, according to his own theory, exist.

If ordinary men never report the occurrence of these acts, for all that, according to the theory, they should be encountered vastly more frequently than headaches, or feelings of boredom; if ordinary vocabulary has no non-academic names for them; if we do not know how to settle simple questions about their frequency, duration or strength, then it is fair to conclude that their existence is not asserted on empirical grounds. The fact that Plato and Aristotle never mentioned them in their frequent and elaborate discussions of the nature of the soul and the springs of conduct is due not to any perverse neglect by them of notorious ingredients of daily life but to the historical circumstance that they were not acquainted with a special hypothesis

the acceptance of which rests not on the discovery, but on the postulation, of these ghostly thrusts.

The second objection is this. It is admitted that one person can never witness the volitions of another; he can only infer from an observed overt action to the volition from which it resulted, and then only if he has any good reason to believe that the overt action was a voluntary action, and not a reflex or habitual action, or one resulting from some external cause. It follows that no judge, schoolmaster, or parent ever knows that the actions which he judges merit praise or blame; for he cannot do better than guess that the action was willed. Even a confession by the agent, if such confessions were ever made, that he had executed a volition before his hand did the deed would not settle the question. The pronouncement of the confession is only another overt muscular action. The curious conclusion results that though volitions were called in to explain our appraisals of actions this explanation is just what they fail to provide. If we had no other antecedent grounds for applying appraisal-concepts to the actions of others, we should have no reasons at all for inferring from that actions to the volitions alleged to give rise to them.

Nor could it be maintained that the agent himself can know that any overt action of his own is the effect of a given volition. Supposing, what is not the case, that he could know for certain, either from the alleged direct deliverances of consciousness, or from the alleged direct findings of introspection, that he had executed an act of will to pull the trigger just before he had pulled it, this would not prove that the pulling was the effect of that willing. The connexion between volitions and movements is allowed to be mysterious, so, for all he knows, his volition may have had some other movement as its effect and the pulling of the trigger may have had some other event for its cause.

Thirdly, it would be improper to burke the point that the connexion between volition and movement is admitted to be a mystery. It is a mystery not of the unsolved but soluble type, like the problem of the cause of cancer, but of quite another type. The episodes supposed to constitute the careers of minds

are assumed to have one sort of existence, while those constituting the careers of bodies have another sort; and no bridge-status is allowed. Transactions between minds and bodies involve links where no links can be. That there should be any causal transactions between minds and matter conflicts with one part, that there should be none conflicts with another part of the theory. Minds, as the whole legend describes them, are what must exist if there is to be a causal explanation of the intelligent behaviour of human bodies; and minds, as the legend describes them, live on a floor of existence defined as being outside the causal system to which bodies belong.

Fourthly, although the prime function of volitions, the task for the performance of which they were postulated, is to originate bodily movements, the argument, such as it is, for their existence entails that some mental happenings also must result from acts of will. Volitions were postulated to be that which makes actions voluntary, resolute, meritorious, and wicked. But predicates of these sorts are ascribed not only to bodily movements but also to operations which, according to the theory, are mental and not physical operations. A thinker may ratiocinate resolutely, or imagine wickedly; he may try to compose a limerick and he may meritoriously concentrate on his algebra. Some mental processes then can, according to the theory, issue from volitions. So what of volitions themselves? Are they voluntary or involuntary acts of mind? Clearly either answer leads to absurdities. If I cannot help willing to pull the trigger, it would be absurd to describe my pulling it as 'voluntary'. But if my volition to pull the trigger is voluntary, in the sense assumed by the theory, then it must issue from a prior volition and that from another *ad infinitum*. It has been suggested, to avoid this difficulty, that volitions cannot be described as either voluntary or involuntary. 'Volition' is a term of the wrong type to accept either predicate. If so, it would seem to follow that it is also of the wrong type to accept such predicates as 'virtuous' and 'wicked', 'good' and 'bad', a conclusion which might embarrass those moralists who use volitions as the sheet-anchor of their systems.

## Reading 26.3

## EXERCISE 4

From: Davidson, D. (1980). *Actions, reasons and causes*. In *Essays on Actions and Events*. Oxford: Oxford University Press, pp. 3–19. (Extract pp. 11–17).

## IV

In order to turn the first ‘and’ to ‘because’ in ‘He exercised *and* he wanted to reduce and thought exercise would do it’, we must, as the basic move,<sup>1</sup> augment condition C1 with:

C2. A primary reason for an action is its cause.

The considerations in favour of C2 are by now, I hope, obvious; in the remainder of this paper I wish to defend C2 against various lines of attack and, in the process, to clarify the notion of causal explanation involved.

A. The first line of attack is this. Primary reasons consist of attitudes and beliefs, which are states or dispositions, not events; therefore they cannot be causes.

It is easy to reply that states, dispositions, and conditions are frequently named as the causes of events: the bridge collapsed because of a structural defect; the plane crashed on takeoff because the air temperature was abnormally high; the plate broke because it had a crack. This reply does not, however, meet a closely related point. Mention of a causal condition for an event gives a cause only on the assumption that there was also a preceding event. But what is the preceding event that causes an action?

In many cases it is not difficult at all to find events very closely associated with the primary reason. States and dispositions are not events, but the onslaught of a state or disposition is. A desire to hurt your feelings may spring up at the moment you anger me; I may start wanting to eat a melon just when I see one; and beliefs may begin at the moment we notice, perceive, learn, or remember something. Those who have argued that there are no mental events to qualify as causes of actions have often missed the obvious because they have insisted that a mental event be observed or noticed (rather than an observing or a noticing) or that it be like a stab, a qualm, a prick or a quiver, a mysterious prod of conscience or act of the will. Melden, in discussing the driver who signals a turn by raising his arm, challenges those who want to explain actions causally to identify ‘an event which is common and peculiar to all such cases’ (87), perhaps a motive or an intention, anyway ‘some particular feeling or experience’ (95). But of course there is a mental event; at some moment the driver noticed (or thought he noticed) his turn coming up, and that is the moment he signalled. During any continuing activity, like driving, or elaborate performance, like swimming the Hellespont, there are more or less fixed purposes, standards, desires, and habits that give direction and form to the entire enterprise, and

there is the continuing input of information about what we are doing, about changes in the environment, in terms of which we regulate and adjust our actions. To dignify a driver’s awareness that his turn has come by calling it an experience, or even a feeling, is no doubt exaggerated, but whether it deserves a name or not, it had better be the reason why he raises his arm. In this case, and typically, there may not be anything we would call a motive, but if we mention such a general purpose as wanting to get to one’s destination safely, it is clear that the motive is not an event. The intention with which the driver raises his arm is also not an event, for it is no thing at all, neither event, attitude, disposition, nor object. Finally, Melden asks the causal theorist to find an event that is common and peculiar to all cases where a man intentionally raises his arm, and this, it must be admitted, cannot be produced. But then neither can a common and unique cause of bridge failures, plane crashes, or plate breakings be produced.

The signalling driver can answer the question, ‘Why did you raise your arm when you did?’ and from the answer we learn the event that caused the action. But can an actor always answer such a question? Sometimes the answer will mention a mental event that does not give a reason: ‘Finally I made up my mind.’ However, there also seem to be cases of intentional action where we cannot explain at all why we acted when we did. In such cases, explanation in terms of primary reasons parallels the explanation of the collapse of the bridge from a structural defect: we are ignorant of the event or sequence of events that led up to (caused) the collapse, but we are sure there was such an event or sequence of events.

B. According to Melden, a cause must be ‘logically distinct from the alleged effect’ (52); but a reason for an action is not logically distinct from the action; therefore, reasons are not causes of actions.<sup>2</sup>

One possible form of this argument has already been suggested. Since a reason makes an action intelligible by redescribing it, we do not have two events, but only one under different descriptions. Causal relations, however, demand distinct events.

Someone might be tempted into the mistake of thinking that my flipping of the switch caused my turning on of the light (in fact it caused the light to go on). But it does not follow that it is a mistake to take. ‘My reason for flipping the switch was that I wanted to turn on the light’ as entailing, in part, ‘I flipped the switch, and this action is further describable as having been caused by wanting to turn on the light’. To describe an event in terms of its cause is not to confuse the event with its cause, nor does explanation by redescription exclude causal explanation.

The example serves also to refute the claim that we cannot describe the action without using words that link it to the alleged cause. Here the action is to be explained under the description: ‘my flipping the switch’, and the alleged cause is ‘my wanting to turn on the light’. What relevant logical relation is supposed to

<sup>1</sup> I say ‘as the basic move’ to cancel any suggestion that C1 and C2 are jointly *sufficient* to define the relation of reasons to the actions they explain. For discussion of this point, see the Introduction and Essay 4.

<sup>2</sup> This argument can be found in one or more versions, in Kenny, Hampshire, and Melden, as well as in P. Winch, *The Idea of a Social Science*, and R. S. Peters, *The Concept of Motivation*. In one of its forms, the argument was of course inspired by Ryle’s treatment of motives in *The Concept of Mind*.

hold between these phrases? It seems more plausible to urge a logical link between 'my turning on the light' and 'my wanting to turn on the light', but even here the link turns out, on inspection, to be grammatical rather than logical.

In any case there is something very odd in the idea that causal relations are empirical rather than logical. What can this mean? Surely not that every true causal statement is empirical. For suppose 'A caused B' is true. Then the cause of B = A; so substituting, we have 'The cause of B caused B', which is analytic. The truth of a causal statement depends on *what* events are described; its status as analytic or synthetic depends on *how* the events are described. Still, it may be maintained that a reason rationalizes an action only when the descriptions are appropriately fixed, and the appropriate descriptions are not logically independent.

Suppose that to say a man wanted to turn on the light *meant* that he would perform any action he believed would accomplish his end. Then the statement of his primary reason for flipping the switch would entail that he flipped the switch—'straightway he acts' as Aristotle says. In this case there would certainly be a logical connection between reason and action, the same sort of connection as that between, 'It's water-soluble and was placed in water' and 'It dissolved'. Since the implication runs from description of cause to description of effect but not conversely, naming the cause still gives information. And, though the point is often overlooked, 'Placing it in water caused it to dissolve' does not entail 'It's water-soluble'; so the latter has additional explanatory force. Nevertheless, the explanation would be far more interesting if, in place of solubility, with its obvious definitional connection with the event to be explained, we could refer to some property, say a particular crystalline structure, whose connection with dissolution in water was known only through experiment. Now it is clear why primary reasons like desires and wants do not explain actions in the relatively trivial way solubility explains dissolvings. Solubility, we are assuming, is a pure disposition property: it is defined in terms of a single test. But desires cannot be defined in terms of the actions they may rationalize, even though the relation between desire and action is not simply empirical; there are other, equally essential criteria for desires—their expression in feelings and in actions that they do not rationalize, for example. The person who has a desire (or want or belief) does not normally need criteria at all—he generally knows, even in the absence of any clues available to others, what he wants, desires, and believes. These logical features of primary reasons show that it is not just lack of ingenuity that keeps us from defining them as dispositions to act for these reasons.

C. According to Hume, 'we may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second'. But, Hart and Honoré claim, 'The statement that one person did something because, for example, another threatened him, carries no implication or covert assertion that if the circumstances were repeated the same action would follow' (52). Hart and Honoré allow that Hume is right in saying that ordinary singular causal statements imply generalizations, but wrong for this very reason in supposing

that motives and desires are ordinary causes of actions. In brief, laws are involved essentially in ordinary causal explanations, but not in rationalizations.

It is common to try to meet this argument by suggesting that we do have rough laws connecting reasons and actions, and these can, in theory, be improved. True, threatened people do not always respond in the same way; but we may distinguish between threats and also between agents, in terms of their beliefs and attitudes.

The suggestion is delusive, however, because generalizations connecting reasons and actions are not—and cannot be sharpened into—the kind of law on the basis of which accurate predictions can reliably be made. If we reflect on the way in which reasons determine choice, decision, and behaviour, it is easy to see why this is so. What emerges, in the *ex post facto* atmosphere of explanation and justification as *the* reason frequently was, to the agent at the time of action, one consideration among many, *a* reason. Any serious theory for predicting action on the basis of reasons must find a way of evaluating the relative force of various desires and beliefs in the matrix of decision; it cannot take as its starting point the refinement of what is to be expected from a single desire. The practical syllogism exhausts its role in displaying an action as falling under one reason; so it cannot be subtilized into a reconstruction of practical reasoning, which involves the weighing of competing reasons. The practical syllogism provides a model neither for a predictive science of action nor for a normative account of evaluative reasoning.

Ignorance of competent predictive laws does not inhibit valid causal explanation, or few causal explanations could be made. I am certain the window broke because it was struck by a rock—I saw it all happen; but I am not (is anyone?) in command of laws on the basis of which I can predict what blows will break which windows. A generalization like, 'Windows are fragile, and fragile things tend to break when struck hard enough, other conditions being right' is not a predictive law in the rough—the predictive law, if we had it, would be quantitative and would use very different concepts. The generalization, like our generalizations about behaviour, serves a different function: it provides evidence for the existence of a causal law covering the case at hand.<sup>3</sup>

We are usually far more certain of a singular causal connection than we are of any causal law governing the case; does this show that Hume was wrong in claiming that singular causal statements entail laws? Not necessarily, for Hume's claim, as quoted above, is ambiguous. It may mean that 'A caused B' entails some particular law involving the predicates used in the descriptions 'A' and 'B', or it may mean that 'A caused B' entails that there exists a causal law instantiated by some true descriptions of A and B.<sup>4</sup> Obviously,

<sup>3</sup> Essays 11, 12, and 13 discuss the issues of this paragraph and the one before it.

<sup>4</sup> We could roughly characterize the analysis of singular causal statements hinted at here as follows: 'A caused B' is true if and only if there are descriptions of A and B such that the sentence obtained by putting these descriptions for 'A' and 'B' in 'A caused B' follows from a true causal law. This analysis is saved from triviality by the fact that not all true generalizations are causal laws; causal laws are distinguished (though of course this is no analysis) by the fact that they are inductively confirmed by their instances and by the fact that they support counterfactual and subjunctive singular causal statements. There is more on causality in Essay 7.

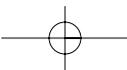
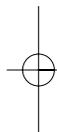
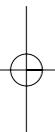


both versions of Hume's doctrine give a sense to the claim that singular causal statements entail laws, and both sustain the view that causal explanations 'involve laws'. But the second version is far weaker, in that no particular law is entailed by a singular causal claim, and a singular causal claim can be defended, if it needs defence, without defending any law. Only the second version of Hume's doctrine can be made to fit with most causal explanations; it suits rationalizations equally well.

The most primitive explanation of an event gives its cause; more elaborate explanations may tell more of the story, or defend the singular causal claim by producing a relevant law or by giving reasons for believing such exists. But it is an error to think no explanation has been given until a law has been produced. Linked with these errors is the idea that singular causal statements necessarily indicate, by the concepts they employ, the concepts that will occur in the entailed law. Suppose a hurricane, which is reported on page 5 of Tuesday's *Times*, causes a catastrophe, which is reported on page 13 of Wednesday's *Tribune*. Then the event

reported on page 5 of Tuesday's *Times* caused the event reported on page 13 of Wednesday's *Tribune*. Should we look for a law relating events of these *kinds*? It is only slightly less ridiculous to look for a law relating hurricanes and catastrophes. The laws needed to predict the catastrophe with precision would, of course, have no use for concepts like hurricane and catastrophe. The trouble with predicting the weather is that the descriptions under which events interest us—'a cool, cloudy day with rain in the afternoon'—have only remote connections with the concepts employed by the more precise known laws.

The laws whose existence is required if reasons are causes of actions do not, we may be sure, deal in the concepts in which rationalizations must deal. If the causes of a class of events (actions) fall in a certain class (reasons) and there is a law to back each singular causal statement, it does not follow that there is any law connecting events classified as reasons with events classified as actions—the classifications may even be neurological, chemical, or physical.



## Reading 26.4

### EXERCISE 5

From: Davidson, D. (1982). Paradoxes of irrationality. In *Philosophical Essays on Freud* (ed. J. Hopkins and R. Wollheim). Cambridge: Cambridge University Press, pp. 289–305. (Extract pp. 296–8).

What requires explaining is the action of an agent who, having weighed up the reasons on both sides, and having judged that the preponderance of reasons is on one side, then acts against this judgment. We should not say he has no reason for his action, since he has reasons both for and against. It is because he has a reason for what he does that we can give the intention with which he acted. And like all intentional actions, his action can be explained, by referring to the beliefs and desires that caused it and gave it point.

But although the agent has a reason for doing what he did, he had better reasons, by his own reckoning, for acting otherwise. What needs explaining is not why the agent acted as he did, but why he *didn't* act otherwise, given his judgment that all things considered it would be better.

A person who appreciates the fact that he has good reasons both for and against an action should not be thought to be entertaining a contradiction. It follows that moral principles, or the judgments that correspond to desires, cannot be expressed by sentences like 'It is wrong to lie' or 'It is good to give pleasure'. Not, that is, if these sentences are taken in the natural way to express universal statements like 'All lies are wrong' or 'All acts that give pleasure are good'. For one and the same act may be a lie and an act that gives pleasure, and so be both wrong and good. On many moral theories, this is a contradiction. Or to take an even simpler case, if it is right to keep promises and wrong to break them, then someone who through no fault of his own has made incompatible promises will do something wrong if he does something right.

The solution to this puzzle about the logic of practical reasoning is to recognize that evaluative principles are not correctly stated in the form 'It is wrong to lie'. For not all lies are wrong; there are cases when one ought to lie for the sake of some more important consideration. The fact that an action is a lie, or the breaking of a promise, or a consumer of time is a count against the action, to be weighed along with other reasons for the action. Every action we perform, or consider performing, has something to be said for it and something against; but we speak of conflict only when the pros and cons are weighty and close to being in balance. Simple deduction can tell me that if I wish to keep promise *A* I must be in Addis Ababa on a certain date, and if I wish to keep promise *B* I must be in Bora Bora at that same time; but logic cannot tell me which to do.

Since logic cannot tell me which to do, it is unclear in what respect either action would be irrational. Nor is the irrationality evident if we add that I judge that all things considered I ought to

keep promise *A*, and yet I keep promise *B*. For the first judgment is merely conditional: in the light of all my evidence, I ought to do *A*; and this cannot contradict the unconditional judgment that I ought to do *B*. Pure internal inconsistency enters only if I also hold—as in fact I do—that I ought to act on my own best judgment, what I judge best or obligatory, everything considered.

A purely formal description of what is irrational in an akratic act is, then, that the agent goes against his own second-order principle that he ought to act on what he holds to be best, everything considered. It is only when we can describe his action in just this way that there is a puzzle about explaining it. If the agent does not have the principle that he ought to act on what he holds to be best, everything considered, then though his action may be irrational from *our* point of view, it need not be irrational from his point of view—at least not in a way that poses a problem for explanation. For to explain his behaviour we need only say that his desire to do what he held to be best, all things considered, was not as strong as his desire to do something else.

But someone who knowingly and intentionally acts contrary to his own principle; how can we explain that? The explanation must, it is evident, contain some feature that goes beyond the Plato Principle; otherwise the action is perfectly rational. On the other hand, the explanation must retain the core of the Plato Principle; otherwise the action is not intentional. An account like this seems to satisfy both requirements: there is, we have agreed, a normal reason explanation for an akratic action. Thus the man who returns to the park to replace the branch has a reason: to remove a danger. But in doing this he ignores his principle of acting on what he thinks is best, all things considered. And there is no denying that he has a motive for ignoring his principle, namely that he wants, perhaps very strongly, to return the branch to its original position. Let us say this motive does explain the fact that he fails to act on his principle. This is the point at which irrationality enters. For the desire to replace the branch has entered into the decision to do it twice over. *First* it was a consideration in favour of replacing the branch, a consideration that, in the agent's opinion, was less important than the reasons against returning to the park. The agent then held that everything considered he ought not to return to the park. Given his principle that one ought to act on such a conclusion, the rational thing for him to do was, of course, not to return to the park. Irrationality entered when his desire to return made him ignore or override his principle. For though his motive for ignoring his principle was a reason for ignoring the principle, it was not a reason against the principle itself, and so when it entered in this second way, it was irrelevant as a reason, to the principle and to the action. The irrationality depends on the distinction between a reason for having, or acting on, a principle, and a reason for the principle.

Another, and simpler, example will make the point clear. Suppose a young man very much wishes he had a well-turned calf and this leads him to believe he has a well-turned calf. He has a normal reason for wanting to have this belief—it gives him pleasure. But if the entire explanation of his holding the belief is that he

wanted to believe it, then his holding the belief is irrational. For the wish to have a belief is not evidence for the truth of the belief, nor does it give it rational support in any other way. What his wish to have this belief makes rational is that this proposition should be true: He believes that he has a well-turned calf. This does not rationalize his believing: I have a well-turned calf. This is a case of wishful thinking, which is a model for the simplest kind of irrationality. Simple as it is, however, the model has a complexity which is obscured by the ambiguity of the phrase 'reason for believing'.

In some cases of irrationality it is unlikely, and perhaps impossible, for the agent to be fully aware of all that is going on in his mind. If someone 'forgets' that today is Thursday because he does not want to keep a disagreeable social commitment, it is perhaps ruled out that he should be aware of this. But in many cases there is no logical difficulty in supposing the agent

knows what is going on. The young man may know he believes he has a well-turned calf only because he wants to believe it, just as the man who returns to the park to replace the branch may realize both the absurdity and the explanation of his action.

In standard reason explanations, as we have seen, not only do the propositional contents of various beliefs and desires bear appropriate logical relations to one another and to the contents of the belief, attitude or intention they help explain; the actual states of belief and desire cause the explained state or event. In the case of irrationality, the causal relation remains, while the logical relation is missing or distorted. In the cases of irrationality we have been discussing, there is a mental cause that is not a reason for what it causes. So in wishful thinking, a desire causes a belief. But the judgment that a state of affairs is, or would be, desirable, is not a reason to believe that it exists.

## Reading 26.5

### EXERCISE 6

From: Tanney, J. (1995). Why reasons may not be causes. *Mind and Language*, 10: 105–128 (Extracts pp. 108–110, 111–2, 113)

### Extract 1

#### 2.1 External Reasons

Having heard from a prophet that he was destined to murder his father and marry his mother, Oedipus attempted to escape his fate by leaving the village in which he lived. As he was walking along, he was accosted by a number of men who were travelling by carriage. They got into a brawl, and Oedipus killed them; in particular, he killed an old man who hit him over the head with his staff. This man was Laius, the King, Oedipus's father. What reason did Oedipus have to kill this man? His motive was self-defence, and we can easily translate this into a rationalizing reason: he believed the old man was about to kill him and he believed that unless he killed the old man first he would die, and he wanted to live. What is the good reason Davidson refers to that Oedipus 'has' but which isn't 'his in acting'? Perhaps it is that Oedipus's father, who also heard the prophecy, ordered his son's murder when Oedipus was an infant, and Oedipus might want revenge; perhaps it is that if Laius is out of the way, Oedipus will be able to take his father's place in his mother's bed. In any case, Oedipus doesn't know that the old man in the carriage is Laius the King, or his father, or the husband of Jocasta, or the husband of his mother. So neither the vengeful motives nor the 'Oedipal' motives could be construed as one of the reasons *for which* Oedipus acted. But this doesn't show the insufficiency of primary reasons, since neither of these *are* primary reasons as Davidson defines them. That is, neither of these motives could be construed as constituents of reasons that could be attributed to Oedipus to explain his action: whether or not he has the vengeful or the Oedipal desires, he has not got the requisite beliefs to combine with these desires to rationalize the action of killing his father. He merely has reasons that rationalize killing the old man in the carriage.

So far, thinking of Oedipus doesn't really help. But maybe we can try to reconstruct the example. What we need is a sense of 'reason' such that a reason is properly attributable to the agent and yet still not be the reason *for which* the agent acts.

#### 2.2 Overridden Reasons

So let's change the story a little. Let's suppose that Oedipus *does* know that the man in the carriage is his father and let's suppose that he both wants to kill his father because he ordered his murder, and he wants to kill him so that he can marry his mother, and he wants to preserve his own life from harm. Of course there ought to be nothing wrong with saying he acted for *all* of the reasons. If someone tells you I bought my house because it was affordable, it has a big garden, and a beautiful view, it would be strong to demand *the* reason *for which* I acted. In certain cases,

the more reasons we give, the more explicable the event; not the other way around. But it is easy enough to complicate the story to make sense of the idea that Oedipus has all of the above reasons for acting, yet only acts because of one of them. Suppose that although Oedipus wants revenge, and would like to take his father's place in his mother's bed, he nonetheless has strong moral inhibitions against killing a parent for any reason, let alone to satisfy vengeful or incestuous desires. Strong enough, so that he would resist the temptation to kill his father for either of these reasons. But suppose, too, that Oedipus has an even stronger desire to live, and in the end, this outweighs his moral inhibition against killing his father. So he kills him. In this sense, perhaps we could argue that it would be inappropriate to cite the vengeful and incestuous reasons to explain Oedipus's action. So, perhaps in this case although Oedipus has good reasons for killing his father, they aren't the ones *for which* Oedipus acts.

Here is an example that shows that citing a primary reason isn't sufficient to (rationally) explain an action. But the *more* that we might add is precisely what I did add in describing the story: a more complex justificatory machinery which would allow us to attribute not only beliefs and desires (or primary reasons), but *competing* primary reasons, values, and weighted judgments as well.<sup>1</sup> But nothing yet has been argued about the necessity of positing a *causal* relation between reasons and actions: we just need to introduce judgments, weights, and values into the 'anaemic' analysis of reasons. But these judgments needn't necessitate talk of mental processes; they may simply be, as reasons and intentions are for Davidson, part of a more complete analysis of the concept of acting for reasons.

Citing a primary reason might not be sufficient to (rationally) explain an action even when it isn't overridden by competing value judgments. It may simply be that one primary reason 'carries more weight' than the others. It may be true, for example, that my house has a big garden and a beautiful view and it was within my price range, but the reason I bought it was because of the garden; I shouldn't have bought it had it not been for this. And how are we to understand this unless we speak of this reason as being 'causally efficacious'?<sup>2</sup> But we can accommodate this example by allowing that certain reasons, values, or desires have more weight than others. But again, to assign weights to desires, etc., is merely, according to the story I shall tell, to complicate the justificatory machinery. It doesn't imply the existence of mental processes. (pp. 108–110)

### Extract 2

#### 2.4 Buridan's Ass

It might be responded that there is still a lacuna in explanation left by pure rationalizations.<sup>3</sup> Suppose, as suggested, our primary

<sup>1</sup> This suggestion is made in Davidson, 1970 and in subsequent articles. But by then the thesis that reasons are causes was already in place, so that the addition of principles, weighted judgments, etc., was *added* to a reasons-as-causes model and not substituted for it.

<sup>2</sup> This was suggested by Jim Hopkins.

<sup>3</sup> This objection was suggested by Christopher Peacocke.

reason account is supplemented by attributions of values and weighted judgments. Now suppose Oedipus finds himself in a 'Buridan's Ass' type situation: he finds himself at the crossroads after the brawl and has to choose which way to go. Suppose that each of the three paths leads to the same destination and requires travelling the same distance. Suppose that Oedipus knows this and although he has reasons for choosing any one of the paths (he wants to get to where it leads) he has no reason for choosing one path over another; that is, no amount of additional *deliberation* will deliver a stronger judgment in favour of taking one path over the other. Now suppose Oedipus takes the path on the right. It might be thought that his having reasons for taking the right path doesn't provide an explanation of his action since his reasons for taking one of the other paths are just as strong. If we want an explanation why he took the right path instead of one of the others, then (it might be argued) rationalizing explanation needs to be supplemented with a causal explanation. Oedipus's taking the right path instead of one of the others would be explicable if his reasons for taking the right path *caused* his action.

But surely the moral of the tale of Buridan's ass is that if the ass were a rationalist, and he only acted on his preferences, then he'd starve to death; or if Oedipus only acted on his preferences then he'd never get to his destination. So much the worse for a *rationalism* that insisted on the contrary. The moral is that rationalizing explanations have a limit, and that when faced with a choice for which there are equally strong reasons, an individual is eventually forced to *pick*, not choose.<sup>4</sup> That is, although Oedipus has reasons for taking the right path his taking the right path *instead of another* cannot be rationalized. (pp. 111–112)

<sup>4</sup> By 'pick' I mean take a course of option more or less arbitrarily, and by 'choose' I mean to take a course of option determined by one's preferences. See Ullman-Margalit and Morgenbesser, 1977.

### Extract 3

#### 2.5 Weakness of Will

Suppose Oedipus is weak-willed and acts contrary to his own best judgment.<sup>5</sup> Doesn't this show that primary reasons, together with weighted judgments aren't sufficient to (rationally) explain Oedipus's action? For the reasons which were weighted higher according to Oedipus were not, by hypothesis, the reasons for which he acted. Hence we need to introduce causation into the analysis in order to provide an account of rational action. What was irrational about the akratic action was that Oedipus's best judgment didn't cause his action; whereas in rational action Oedipus's best judgment does cause the action.

This issue is complicated, and there are a number of ways to tackle the challenge posed above. Davidson's own account of akrasia is rather too complicated to go into here, but in the end it doesn't support the suggestion that introducing causation is necessary (nor does he take it to); on the contrary, I think it is certain aspects of his causal account of reasons that renders irrationality paradoxical for him.<sup>6</sup> In any case, I think that this example collapses into those already considered. Let's accept that Oedipus assessed his situation rationally and was 'overcome' by weakness or a recalcitrant desire. What does this amount to other than the fact that (1) Oedipus's better judgment doesn't (rationally) explain his (akratic) act; and (2) Oedipus's other reasons (which were not deemed his best) do rationalize his akratic act? Why must the reasons that rationalize his action also be causes? (p. 113)

<sup>5</sup> This example is due to Nancy Cartwright.

<sup>6</sup> I address these points in Tanney, 1995; see also Note 25, below.

## Reading 26.6

### EXERCISE 7

From: Strawson, P.F. (1974). *Freedom and resentment*. In *Freedom and Resentment, and Other Essays*. London: Methuen, pp. 1–25 (Extracts pp. 6–9, 10–11).

### Extract 1

#### IV

It is one thing to ask about the general causes of these reactive attitudes I have alluded to; it is another to ask about the variations to which they are subject, the particular conditions in which they do or do not seem natural or reasonable or appropriate; and it is a third thing to ask what it would be like, what it is like, not to suffer them. I am not much concerned with the first question; but I am with the second; and perhaps even more with the third.

Let us consider, then, occasions for resentment: situations in which one person is offended or injured by the action of another and in which—in the absence of special considerations—the offended person might naturally or normally be expected to feel resentment. Then let us consider what sorts of special considerations might be expected to modify or mollify this feeling or remove it altogether. It needs no saying now how multifarious these considerations are. But, for my purpose, I think they can be roughly divided into two kinds. To the first group belong all those which might give occasion for the employment of such expressions as ‘He didn’t mean to’, ‘He hadn’t realized’, ‘He didn’t know’; and also all those which might give occasion for the use of the phrase ‘He couldn’t help it’, when this is supported by such phrases as ‘He was pushed’, ‘He had to do it’, ‘It was the only way’, ‘They left him no alternative’, etc. Obviously these various pleas, and the kinds of situations in which they would be appropriate, differ from each other in striking and important ways. But for my present purpose they have something still more important in common. None of them invites us to suspend towards the agent, either at the time of his action or in general, our ordinary reactive attitudes. They do not invite us to view the *agent* as one in respect of whom these attitudes are in any way inappropriate. They invite us to view the *injury* as one in respect of which a particular one of these attitudes is inappropriate. They do not invite us to see the *agent* as other than a fully responsible agent. They invite us to see the *injury* as one for which he was not fully, or at all, responsible. They do not suggest that the agent is in any way an inappropriate object of that kind of demand for goodwill or regard which is reflected in our ordinary reactive attitudes. They suggest instead that the fact of injury was not in this case incompatible with that demand’s being fulfilled, that the fact of injury was quite consistent with the agent’s attitude and

intentions being just what we demand they should be.<sup>1</sup> The agent was just ignorant of the injury he was causing, or had lost his balance through being pushed or had reluctantly to cause the injury for reasons which acceptably override his reluctance. The offering of such pleas by the agent and their acceptance by the sufferer is something in no way opposed to, or outside the context of, ordinary inter-personal relationships and the manifestation of ordinary reactive attitudes. Since things go wrong and situations are complicated, it is an essential and integral element in the transactions which are the life of these relationships.

The second group of considerations is very different. I shall take them in two subgroups of which the first is far less important than the second. In connection with the first subgroup we may think of such statements as ‘He wasn’t himself’, ‘He has been under very great strain recently’, ‘He was acting under post-hypnotic suggestion’; in connection with the second, we may think of ‘He’s only a child’, ‘He’s a hopeless schizophrenic’, ‘His mind has been systematically perverted’, ‘That’s purely compulsive behaviour on his part’. Such pleas as these do, as pleas of my first general group do not, invite us to suspend our ordinary reactive attitudes towards the agent, either at the time of his action or all the time. They do not invite us to see the agent’s action in a way consistent with the full retention of ordinary inter-personal attitudes and merely inconsistent with one particular attitude. They invite us to view the agent himself in a different light from the light in which we should normally view one who has acted as he has acted. I shall not linger over the first subgroup of cases. Though they perhaps raise, in the short term, questions akin to those raised, in the long term, by the second subgroup, we may dismiss them without considering those questions by taking that admirably suggestive phrase, ‘He wasn’t himself’, with the seriousness that—for all its being logically comic—it deserves. We shall not feel resentment against the man he is for the action done by the man he is not; or at least we shall feel less. We normally have to deal with him under normal stresses; so we shall not feel towards him, when he acts as he does, under abnormal stresses, as we should have felt towards him had he acted as he did under normal stresses.

The second and more important subgroup of cases allows that the circumstances were normal, but presents the agent as psychologically abnormal—or as morally undeveloped. The agent was himself; but he is warped or deranged, neurotic or just a child. When we see someone in such a light as this, all our reactive attitudes tend to be profoundly modified. I must deal here in crude dichotomies and ignore the ever-interesting and ever-illuminating varieties of case. What I want to contrast is the attitude (or range of attitudes) of involvement or participation in a human relationship, on the one hand, and what might be called the objective attitude (or range of attitudes) to another human being, on the other. (pp. 6–9)

<sup>1</sup> Perhaps not in every case *just* what we demand they should be, but in any case *not* just what we demand they should not be. For my present purpose these differences do not matter.

## Extract 2

What I have called the participant reactive attitudes are essentially natural human reactions to the good or ill will or indifference of others towards us, as displayed in *their* attitudes and actions. The question we have to ask is: What effect would, or should, the acceptance of the truth of a general thesis of determinism have upon these reactive attitudes? More specifically, would, or should, the acceptance of the truth of the thesis lead to the decay or the repudiation of all such attitudes? Would, or should, it mean the end of gratitude, resentment, and forgiveness; of all reciprocated adult loves; of all the essentially *personal* antagonisms?

But how can I answer, or even pose, this question without knowing *exactly* what the thesis of determinism is? Well, there is one thing we do know: that if there is a coherent thesis of determinism, then there must be a sense of 'determined' such that, if that thesis is true, then all behaviour whatever is determined in that sense. Remembering this, we can consider at least what possibilities lie formally open; and then perhaps we shall see that the question can be answered *without* knowing exactly what the thesis of determinism is. We can consider what possibilities lie open because we have already before us an account of the ways in which particular reactive attitudes, or reactive attitudes in general, may be, and, sometimes, we judge, should be, inhibited. Thus I considered earlier a group of considerations which tend to inhibit, and, we judge, should inhibit, resentment, in particular cases of an agent causing an injury, without inhibiting reactive attitudes in general towards that agent. Obviously this group of considerations cannot strictly bear upon our question; for that question concerns reactive attitudes in general. But resentment has a particular interest; so it is worth adding that it has never been claimed as a consequence of the truth of determinism that one or another of *these* considerations was operative in every case of an injury being caused by an agent; that it would follow from the truth of determinism that anyone who caused an injury *either* was quite simply ignorant of causing it *or* had acceptably overriding reasons for acquiescing reluctantly in causing it *or* . . . , etc. The prevalence

of this happy state of affairs would not be a consequence of the reign of universal determinism, but of the reign of universal goodwill. We cannot, then, find here the possibility of an affirmative answer to our question, even for the particular case of resentment.

Next, I remarked that the participant attitude, and the personal reactive attitudes in general, tend to give place, and it is judged by the civilized should give place, to objective attitudes, just in so far as the agent is seen as excluded from ordinary adult human relationships by deep-rooted psychological abnormality—or simply by being a child. But it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition.

Now this dismissal might seem altogether too facile; and so, in a sense, it is. But whatever is too quickly dismissed in this dismissal is allowed for in the only possible form of affirmative answer that remains. We can sometimes, and in part, I have remarked, look on the normal (those we rate as 'normal') in the objective way in which we have learned to look on certain classified cases of abnormality. And our question reduces to this: could, or should, the acceptance of the determinist thesis lead us always to look on everyone exclusively in this way? For this is the only condition worth considering under which the acceptance of determinism could lead to the decay or repudiation of participant reactive attitudes.

It does not seem to be self-contradictory to suppose that this might happen. So I suppose we must say that it is not absolutely inconceivable that it should happen. But I am strongly inclined to think that it is, for us as we are, practically inconceivable. The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question. (pp. 10–11)