

- mad, for example. Second, Borges never insists that 'the' Zahir is actually one thing, as opposed to many.
9. My discussion of Wolff here draws heavily from Velleman 2002.
10. Uriah Kriegel (p.c.) suggests another non-objectivist theory of essences, on which essentiality (a property of properties) is response-dependent. Roughly, a property F is essential to x iff: if x lost F, we would (under normal conditions) intuit that x has gone out of existence. The objection below (to Lewis) is also an objection to Kriegel's idea.

Bibliography

- Borges, J. (1999), 'The Zahir,' in his *Collected Fictions* (Penguin), pp. 237–49.
- Chisholm, R. (1967), 'Identity through Possible Worlds: Some Questions,' *Notis* 1.1, 1–8.
- (1976), *Person and Object* (London: Allen and Unwin).
- Cooper, J. M. (ed.) (1997), *Plato: Complete Works* (Indianapolis: Hackett Publishing).
- Fine, K. (1994), 'Essence and Modality,' *Philosophical Perspectives* 8, 1–16.
- Hazlett, A. (2006), 'Disassembly and Destruction,' *The Monist* 89.3, 418–33.
- Kafka, F. (2000), *The Metamorphosis and Other Stories* (New York: Penguin).
- Lewis, D. (1983), *Philosophical Papers*, vol. 1 (Oxford: Oxford University Press).
- (1986), *Philosophical Papers*, vol. 2 (Oxford: Oxford University Press).
- Locke, J. (1975), *An Essay Concerning Human Understanding* (Oxford: Oxford University Press).
- Markosian, N. (1998), 'Brutal Composition,' *Philosophical Studies* 92 (3), 211–49.
- McKeon, R. (ed.) (1941), *The Basic Works of Aristotle* (New York: Random House).
- Nielsen, L. O. (2003), 'Peter Auriol,' in *Blackwell Companion to Philosophy in the Middle Ages* (Blackwell), pp. 494–503.
- Robertson, T. (1998), 'Possibilities and the Arguments for Origin Essentialism,' *Mind* 107, 729–50.
- Salmon, N. (1989), 'The Logic of What Might Have Been,' *Philosophical Review* 98, 3–34.
- (2005), *Reference and Essence*, 2nd edn. (New York: Prometheus).
- Yablo, S. (1987), 'Identity, Essence, and Indiscernibility,' *Journal of Philosophy* 84 (6), 293–314.
- van Inwagen, P. (1990), *Material Beings* (Ithaca: Cornell University Press).
- Velleman, J. D. (2002), 'Identification and Identity,' in Buss and Overton (eds.), *Contours of Agency* (Boston: MIT Press), pp. 91–123.
- Wolff, R. P. (1976), 'There's Nobody Here but Us Persons,' in Gould and Wartofsky (eds.), *Women and Philosophy: Toward a Theory of Liberation* (G.P. Putnam's Sons), pp. 128–44.

5

Bringing Things About

Neal Judisch

The origin of action – its efficient, not its final cause – is choice, and that of choice is desire and reasoning with a view to an end.

—Aristotle, *Nicomachean Ethics*

The stick moves the stone and is moved by the hand, which is again moved by the man; in the man, however, we have reached a mover that is not so in virtue of being moved by something else.

—Aristotle, *Physics*

1 Natural agency

In this essay, I hope to dissolve a problem for naturalistic theories of human action. The problem I aim to dissolve is generated when two independently plausible theses concerning human action are combined: on the one hand, it is plausible that action consists in sequences of suitably related events – desires and beliefs give rise to mental events such as choices, or states such as intentions, which choices or intentions subsequently cause the agent's body to move in ways aimed at satisfying her goals. On the other hand, actions are distinct from 'mere happenings' in that they are brought about by the agents whose actions they are: actions are things agents *do*, not things that merely occur to or within their bodies.

To put the latter point another way, when a person acts she is not simply the locus of a series of reflexes, whether mental or physical or both, but is (as Taylor indicates) an initiator of action herself. 'In describing anything as an act there must be an essential reference to an agent as the performer or author of that act, not merely in order to know whose act it is, but in order even to know that it is an act at all.... Another perfectly

natural way of expressing this notion ... is to say that, in acting, I make something happen, I cause it, or bring it about' (1966, pp. 109, 111). This manner of description is naturally conjoined with the metaphysical intuition that actions are distinct from mere happenings-involving agents, at least in part, because actions are brought about by agents in a way that differs in kind from the mechanical outworking of causally related event patterns within them. For the mere unfolding of events within a person, however complex, does not obviously add up to an agent's making anything happen, and from this it follows (according to Chisholm) that 'at least one of the events that is involved in any act is caused, not by any other event, but by the agent, the man' (1966, p. 29). Taken literally, these sentiments express the contention that any person who acts must enjoy the possession of a unique capacity: the exercise of which cannot be reduced to any sequence of event-causes: necessary to the performance of any action is a direct causal contribution emanating from the agent herself, and nothing *about* the agent – whether mental or physical – is causally sufficient for it.

Such 'agent-causation' theories are not without their difficulties, and the majority of action theorists are not prepared to accept them. A widely popular alternative thus seeks to specify the necessary and sufficient conditions for action without recourse to an 'agent-causal' power of the sort envisaged here. On this approach, human action is constituted by the right kinds of events being related in the right kinds of ways, which does not require the aetiology of any act to include a special variety of causation occurring nowhere else in nature.¹ Theories operating within these parameters are often called 'causal theories' of action, since they attempt to analyse action in causal terms whilst making use of event-causal relations only within their analyses. They are also called 'naturalistic theories' of action. But it is important to recognize that these approaches are 'naturalistic' only in the sense that their theorising is confined by the strictures imposed by an exhaustively event-causal ontology. 'Naturalism' so understood is therefore not equivalent to ontological materialism: there is nothing inconsistent about rejecting materialism in favour of strongly dualistic theories of mind, for example, and nevertheless accepting that action is analysable in terms of appropriately related mental and physical events. The difficulties for naturalistic theories of action should therefore be distinguished from whatever problems besetting materialist theories of mind (or persons) specifically there may be. Thus when N. M. L. Nathan says, strikingly, that 'materialism is false if anyone ever performs an action' (1975, p. 501), he should be understood as registering the conviction

that materialism cannot allow 'agent-causal' powers, and that, since powers of this sort are required for action, materialism is not consistent with action. If, however, a given anti-materialist metaphysics is no more hospitable to 'agent-causal' powers, then such an anti-materialism would likewise be false if anyone ever performs an action, and for the same reason. 'Materialism,' in this instance, isn't the primary culprit.

Similar remarks apply to the issue of mental causation. One of the ironies of the last few decades has been the emergence of the problem of mental causation as a difficulty for materialist theories of mind, not just for overtly dualistic ones. Yet although it is clear that any naturalistic theory of action is committed to the reality of mental causation (inasmuch as it relies upon mental causes in its analysis), an acceptable solution to the latter problem will not solve the problem of interest here. For our problem concerns, not how mental events could be causes, but how, even when they are, such events occurring within an agent could amount to the agent's making anything happen.²

Closer to home, but still distinct from our difficulty, is a challenge to naturalistic theories that is often supposed to be the chief hurdle they must overcome. This is the problem of causal deviance, and it challenges us to explain exactly *how*, or in what *way*, mental states must cause bodily movements in cases of intentional action. The concern is that an agent may form an intention to perform some action A in order to bring about a state of affairs S, and that very intention may in fact cause the agent to A and S to obtain as a result, but the causal trajectory between the intention and A may be such as to preclude the agent from exercising control over what she does, rendering her behaviour unintentional. The intention in question might produce in the agent a state of nervousness, for example, which causes her temporarily to lose control over her behaviour but which, fortuitously, also causes her body to move in just the way she had intended it to. And in cases like this the agent either does not act intentionally, or, if all actions are intentional, falls altogether to act.

Examples of deviance tend to vitiate the hope that sufficient conditions for intentional action might be framed without appeal to a form of agency that cannot be analysed by way of the interplay between an agent's desires, intentions and motions; and to that extent they too pose a problem for naturalistic action theories.³ Yet even granting that this difficulty may by itself be thought to undermine naturalism about action, this isn't the only thing that might jeopardize a naturalistic orientation. For a successful solution to the problem of deviance may specify the 'right way' for mental events to cause bodily

movements without thereby illuminating how these rightly-related happenings amount to a doing. Simply put, the deviance problem is essentially concerned with explaining how mental and physical events are related when a person (intentionally) acts, and our problem is concerned with how action can consist of 'happenings' and nothing more, quite apart from the precise manner of the relation between the happenings.

A final clarification. The question before us is naturally of concern to philosophers interested in the metaphysics of free will, insofar as there can be no free agency unless there is such a thing as agency 'period.' But it is not a problem afflicting one theory of human freedom only. Specifically, although philosophers inclined towards 'agent-causation' in action theory may also be inclined towards libertarianism about free will, the challenge of explaining how actions are distinct from mere happenings is neither generated by libertarian theories of freedom nor uniquely applicable to them. It is a problem, in other words, that the libertarian faces qua action theorist, not qua libertarian, for it applies in the first instance to agency simpliciter and to free agency only secondarily. (I shall return to this point and address its significance in the final section.)

Our puzzle, then, although easily conflated with the foregoing topics, is distinct from and rather more general than them. Irrespective of a person's stance on the mind/body relation or the freedom of the will, it confronts anyone who wishes to locate instances of agency within nature, or to explain what goes on in a world of events when someone acts: 'Of course action differs from other behaviour in that the agent brings it about, but the problem is how to accommodate such bringing about within a naturalist ontology' (Bishop 1989, p. 69).

2 Can agency be reduced

Broadly speaking there are three available naturalistic strategies for accommodating the intuition that actions are assignable to agents, as being in a special sense their causes. The first involves identifying agents with the complexes of states and events, or some subset of them, which are productive of action. The second is similar, but takes on a distinctively functionalist cast. It attempts to specify the characteristic causal role played by agents in cases of action, and then to identify a state or disposition causally relevant to behaviour that satisfies the agent's functional description. And the final strategy tries first to solve the problem of causal deviance and then to argue that because no

further (reductive) analysis of action is possible, no further analysis is necessary. I take them each in turn.

One interesting version of the first strategy has been formulated by Ekstrom (1993), whose 'coherentist' theory of the self is targeted at explaining how an action may be said to derive from an agent as its 'source.' To forestall confusion it must be acknowledged that Ekstrom's proposal is geared specifically towards explicating the conditions for 'self-determination' in a sense that confers *autonomy*, and a mildly refined version of it is put into service of the (naturalistic) libertarian account of free will developed in her (2000). I have argued that the problem of 'bringing things about' is not equivalent to worries about autonomy or free will, and that it ought not be tied to any particular conception of them. But although Ekstrom's overarching goal is to specify the conditions for autonomous action, it should be noted that the uniquely libertarian aspects of her theory can with propriety be dissociated from those aspects of it that are directly related to the more fundamental problem we have isolated, and it seems to me that her analysis of self-determination may offer the resources for a promising first run at its solution. So I shall examine only the elements of her theory relevant to our question in what follows.

Note first that Ekstrom's account is 'reductive' in a twofold sense, since 'agent-causation' is in her view 'an ontologically and conceptually reducible notion' (2000, p.114). It is *ontologically* reducible because an agent's causing an action, or causing an intention to act, consists in events (involving agents) causing her behaviour, or prior deliberative events causing her intentions to act. And it is *conceptually* reducible because all we have in mind when we say that an agent causes something, according to Ekstrom, is that some event involving the agent causes the something in question to occur:

Surely we do have experience of people doing things, *bringing about* events and states of affairs.... But in speaking of ourselves as *being* the cause of these events... what we really mean is that certain events caused other events.... Speaking of an object or an agent as the cause of some event is just shorthand for the more specific event-causal explanation upon which the agent causal explanation supervenes. (p. 94)

Readers familiar with the mental causation literature will be alive to the concern that a 'supervening' state (or explanation) may sometimes be eclipsed by the underlying causal process upon which the

state (or explanation) supervenes. For given the causal sufficiency of the underlying process, and the corresponding sufficiency of the causal explanations couched in their terms, it is tempting to dismiss the 'supervening' states and explanations as superfluous – unless, of course, the 'supervening explanation' is demonstrably just a restatement of the lower-lying one.⁴ The conceptual reducibility of agent-causation is therefore not peripheral to her project: to see that the 'agent causal explanation supervenes' in a way that secures the agent's contribution to action, a conceptual reduction of the 'agent' or the 'agent-cause' is required.

And this is precisely what she seeks to provide. Under her analysis, an agent is identifiable with a 'character-system,' which is comprised of preferences, beliefs, and desires, along with a capacity to fashion and refashion the character via critical evaluation. Of particular significance are the agent's *preferences*, understood as 'specially processed desire[s]' that have been 'formed by a process of critical evaluation with respect to one's conception of the good' (p. 106). A preference is thus a second-order desire that some first-order desire of the agent be effective in action, namely, whichever first-order desire coheres with her evaluation of what is good to want, and what sorts of desire she reflectively believes ought to move her to act.⁵ It is the preferences in particular which underwrite the thought that the agent herself is the source of her behaviour when she acts upon them, because the preferences she endorses are most expressive of her character, or the sort of person she is: 'Since a "preference" represents what an agent wants as the outcome of her reflection on what is good... when an agent acts on a decisively formed preference... she is involved in the action – she is its source' (p. 107). Given the reducibility of agents to these character-systems, then, we can specify that 'an intention is agent caused just in case it results by a normal causal process from a preference for acting as specified in the content of the intention, where the preference itself is the output of an uncoerced exercise of the agent's evaluative faculty, the inputs into which (various considerations) cause... the decisive formation of the preference' (p. 114).⁶

Part of what lends appeal to Ekstrom's proposal is the attractive idea that when an action is attributable to the agent who performs it, her conduct should be reflective of who she is. And insofar as her preferences and acceptances form a comparatively resilient and coherent system, which is supported by her values or entrenched convictions, there is reason to think that her character is identifiable with this causally relevant set of psychological states.⁷ Yet if the agent can likewise

be reduced to her character, then we can affirm that the agent herself brings her behaviour about whenever her character-system plays a causal role in generating it. The question is whether this series of reductions goes through.

There is cause for concern that at least one of these reductions fails. It is not, I think, excessively naive to protest that the beliefs, wants, and even the enduring values of a person, however closely connected with his 'character' they may be, remain for all that characteristics of the person – things had or possessed by him – and therefore at any rate not *conceptually* equivalent to the person who possesses them. To be sure, Ekstrom does not intend to deliver a theory of 'persons,' so much as a theory of 'agents' or the 'self.' And it may be that, on her theory, the person should be seen as causing his actions 'qua' agent, as opposed to 'qua' person or enduring object. Nevertheless, the intuition underlying this protest, I take it, is that it must be the person – 'the man,' to borrow from Chisholm – who is causally responsible for what he does, and not simply the complex of traits or beliefs and desires 'characteristic' of him. To the extent that 'the man' isn't conceptually reducible to some subset of properties or states he exemplifies, to that extent the man resists reductive analysis – and so therefore does the man-as-cause.

Perhaps this anti-reductive intuition may be shaped into a specific objection. It seems that a conceptual reduction of the agent will fail, if what an agent does when he weighs his reasons and preferences, forms intentions to act in their light, and then acts accordingly, is distinct from what his reasons and intentions themselves accomplish when he acts. Notice how natural it is to say (as indeed Ekstrom says above) that an agent 'acts on' his preferences, after having 'formed' them by way of his evaluation of the good. These locutions reflect the fact that, at least in idea, 'intentional causation is in certain important respects unlike billiard-ball causation. Both are cases of causation, but in the case of desires and intentions, in the case of normal voluntary actions, once the causes are present they still do not compel the agent to act; the agent has to *act on* his reasons or on his intention' (Searle 2001, p. 231). Thus despite the ineliminable causal contribution his mental states undoubtedly provide, the agent's characteristic role is intuitively to mediate between his reason states and his intentions, by forming his intentions in light of his wants and beliefs, and again, to mediate between his intentions and his behaviour by executing his intentions in act. Yet his reasons cannot 'mediate' between themselves and his intentions, nor can his intentions sensibly be said to 'act upon' themselves.

This is not yet to decide that the ontological reduction of agency cannot be matched with a corresponding conceptual reduction of the agent's causal contribution to action. What it does suggest, according to Velleman, is that involvement of the kinds of mental states that have so far been articulated fails to capture the agent's distinctive involvement in generating his conduct, because what the agent does is to intervene between reasons and intention, and between intention and bodily movements, in each case guided by the one to produce the other.... When reasons are described as directly causing an intention, and the intention as directly causing movements, not only has the agent been cut out of the story but so has any psychological item that might play his role' (1992, p. 463). Since the agent's characteristic role is to mediate between the occurrences to which standard naturalistic theories appeal, the agent will fail to participate in his actions unless he – or his agential role – can be reduced to some state or event within him, which performs the work we suppose him to do. A second attempt at the conceptual reduction of agent-causation may thus remain in the offing, if we can add to the mix of action-generating states a state or disposition that is functionally identical to the agent, in the sense that it plays the 'intermediary' role ordinarily ascribed to him.

According to Velleman we should view the agent's definitive role as being 'that of a single party prepared to reflect on, and take sides with, potential determinants of behaviour at any level in the hierarchy of attitudes' (p. 477), so as to ensure that the motives which provide the best reasons for acting 'prevail over those whose rational force is weaker' (p. 478). The psychological state that fills this bill, he argues, cannot be identical to any of the reasons or motives that may potentially come up for evaluation, but must be a distinct state *by* which the agent evaluates his motives: a state with sufficient motivational force to 'throw its weight' behind what the agent considers are his best reasons for acting, by selecting and actuating the superior motivations 'in the agent's name.' What might this state be? It 'can only be a motive that drives practical thought itself' (p. 477), a desire to act in accordance with reasons, 'to do what makes sense, or what's intelligible to [the agent], in the sense that [he] could explain it' (p. 478). 'Only such a motive would occupy the agent's functional role, and only its contribution to his behaviour would constitute his own contribution' (p. 477).

The form of reductive strategy driving Velleman's procedure is I think unquestionably the most promising one. The specifics of his reductive analysis, however, suggest that the problem he solves is not the problem with which he began; it is not the problem of 'bringing things about.' As

a general matter, one might worry that it makes little sense to identify a desire as performing the work ordinary parlance attributes to agents since, as Mele points out, even granting that a person's deliberations and decisions are motivated by something akin to the desire figuring centrally in this account, the desire itself 'cannot play [the agent's] role because no desire can deliberate and decide' (2003, p. 225). Whatever it is that is caused to evaluate motives by the desire to behave rationally, in other words, may not itself be reductively identifiable with the desire that moves the agent to engage in his evaluative efforts, and it is 'engagements' of precisely this sort that the critics of naturalism contend are irreducibly actional. But there is a more telling indication that Velleman's solution misses its mark. Notice that akratic or weak-willed actions are not such that the agent's disposition to ensure the victory of his rationally superior motivations contributes decisively to their occurrence, nor indeed is any action that the agent does not 'throw his weight behind' in Velleman's sense. Yet akratic acts are most certainly actions agents perform, even if they are not paradigms of rational, self-controlled agency.⁸

The central difficulty with Velleman's solution can best be appreciated if we borrow a bit of his terminology. What began as an attempt at 'finding an agent at work amid the workings of mind' turned quickly into a theory of 'full-blooded' human action, or human agency '*par excellence*,' within which the basic activity of the agent remains unanalysed and in effect presupposed. This tendency to slide from the relatively more fundamental phenomenon of 'doing' to a special form of agency, wherein the agent may be considered more 'fully involved,' is similarly in evidence when he remarks that 'full-blooded human action occurs only when the subject's capacity to make things happen is exercised to its fullest extent' (2000, p. 4), where the extent of his involvement is to be appraised along normative lines. But our problem is not primarily normative. It emerges most clearly when we consider the analogies Velleman formulates in an effort to convey that human agency is of a piece with other varieties of 'behaviour' that fit less problematically within event-causal schemes. Thus a person counts as the 'initiator' of his actions in the same sense that a person counts as a digester of food or a fighter of infections: namely, by virtue of the fact that some proper part of the person contains a food-digesting system, an infection-fighting system, and an 'action-initiating system' – a system which itself performs the functions in virtue of which he qualifies as an agent' (1992, p. 475–6). But these analogies invite the very misgivings that induce scepticism concerning naturalism in the first place; for the goal of a naturalistic theory

should be to 'earn the right to make jokes about primitive agent-causation' (p. 469), precisely by explaining how naturalism can accommodate the distinction between actions on the one hand and mere sub-actional physiological processes (such as digestion) on the other. To presuppose the distinction can be accommodated, and then proceed to differentiate agency simpliciter from action 'par excellence' is one thing; to accommodate the distinction is something else.

When reductive analyses persistently fail, it is natural to revisit the item targeted for reduction with an eye towards greater clarification of just what this phenomenon is supposed to be, and why it creates such persistent trouble. It's reasonable, too, to ask whether the item in question either needs no more reduction than whatever degree it has already received, or whether for that matter it ever really needed any reduction to speak of at all. Suppose for example we succeed in answering the argument from deviance, by establishing criteria the satisfaction of which entails that, in any naturally possible world, a person whose behaviour is caused in *that way* is a person who intentionally acts. What more about agency requires reduction, if intentional action has already been reduced?

John Bishop (1989) asks and answers this question by arguing that once the challenge of causal deviance has been addressed it is no longer reasonable to entertain doubts about the compatibility of naturalism and agency. His own analysis of (basic) intentional action, he allows, may not apply to any possible worlds containing subjects with irreducible agent-causal powers, but he claims that the conditions he sets forth do apply 'to all possible worlds that have the same kind of ontology and causal order that science understands the actual world to have.' In judging the merits of his theory, then, what the reader has to do is 'to ask whether, within a natural scientific ontology, you could conceive of just those conditions it specifies applying without being prepared to attribute a basic intentional action to the agent, or conversely.' And if we're ready to concede that the absence of intentional action is inconceivable under such circumstances, we must a fortiori be prepared to admit that agency as such is compatible with naturalism: 'Postulating that agency involves a nonnatural factor (agent-causation) may be quite sensible while we lack uniform correlates for actions; but it becomes plain silly once we can understand how such correlates may be obtained' (p. 179). Whatever additional factors may be fuelling 'agent-causation syndrome' – or the urge to view agency as nonnatural, irreducible – 'dwindle by comparison' with the challenge of causal deviance, says Bishop (p. 180), and may at this point permissibly be ignored.

I have voiced my agreement that specifying normality conditions on the event-causal relations involved in cases of action is insufficient to allay fears that something essential to agency – viz., its 'initiation' by the agent – has been overlooked. At the same time, Bishop's position is perhaps closer to my own than are the reductionist approaches we have already examined. Bishop does not think any 'more' conceptual reduction of agency is required than what he (let us grant) has already achieved, and here I am prepared to agree. Where we part ways, however, is on the reasonableness of persistent scepticism concerning natural agency. For if there remains something about agency that cannot be reduced, then we want an explanation for this failure of reducibility. In particular, we want an explanation for this failure that is consistent with the naturalism which was thought initially to require it, for supposing a person is 'on the fence' about ontological naturalism, she may well view the irreducibility of (some features of) agency as providing reason to reject the naturalism assumed in Bishop's account and his theory of action along with it.

The point is not merely a dialectical one. Here I tip my hat to Sydney Shoemaker, whose remarks in another connection I find equally applicable to our situation: 'To a large extent, the mind-body problem, including the problem of personal identity, arises because of considerations that create the appearance that no naturalistic account could be true; and I think that solving the problem has got to consist in large part in dispelling that appearance (while acknowledging the facts that give rise to it)' (1984, p. 71). What's missing from Bishop's presentation is not necessarily some aspect of the metaphysics of action that has escaped his analytic attention, but an explanation for the appearance that 'natural agency' somehow leaves out its agent, and why agent-causation syndrome remains hard to shake.

3 The phenomenology and conceptual irreducibility of agency

I propose we attempt to dispel the appearance that naturalism could not be true by acknowledging that when we conceive of action as a 'natural' phenomenon we no longer conceive it as centred upon the agent as its source. Hornsby is no doubt correct when she says that 'Our conception of a person as an agent is a conception of something with a causal power... to initiate series of events containing some we want. An action is the exercise of such a power, and a person's actions are the events at the start of those series she initiates' (1993, p. 164). But the question is how such a conception can be maintained alongside a

very distinct conception, according to which the agent's 'initiation' of any act is a matter of particular states within her being caused by prior events. Viewed far enough 'from the outside,' as Nagel rightly notes, nothing of agency looks to remain within this causal network, for both the agent and her putative initiating powers are inevitably engulfed in the relentless tide of occurrences: 'Something peculiar happens when we view action from an objective or external standpoint. Some of its most important features seem to vanish under the objective gaze. Actions seem no longer assignable to individual agents as sources, but become instead components of the flux of events in the world of which the agent is a part' (1986, p. 110).

Once more, the problem cannot with justice be laid on the doorstep of 'materialism' per se. Stipulating the presence of a robustly efficacious non-physical mentality does not alter the fact that '*my doing* of an act – or the doing of an act by someone else – seems to disappear when we think of the world objectively ... Even if we add sensations, perceptions, and feelings we don't get action, or doing – there is only what happens' (p. 111).⁹ The anxiety about agency to which this alienating vision leads thus stems fundamentally from our being situated within a purely event-causal nexus; a 'nexus' from which we seem unable to coherently imagine escaping, but within which we likewise cannot conceive of ourselves as genuine springs of activity.

The conceptual disorientation that attends viewing our world 'objectively' is not unfamiliar. Many of philosophy's 'hard problems' involve something recognizably similar – normativity, meaning, and conscious experience are prime examples of phenomena that, each in their own way, prove uneasy to cognize in terms of the 'natural' facts thought to 'give rise' to them. Between these things and the natural world is a notorious 'explanatory gap.' And what lies behind the gap isn't necessarily a dearth of pertinent empirical knowledge – of a sort that might be rectified by conservative extension of the stock of empirical concepts we currently possess – but rather an intuitive failure of conceptual fit.

The conditions that invite this conceptual displacement may vary from case to case. What is naturalistically recalcitrant about normative or teleological phenomena (e.g.) probably isn't what makes phenomenal consciousness difficult to naturalistically explain. But although agency certainly involves normative and telic dimensions, I do not see that these features are responsible for the problem of 'bringing things about.' Rather, I claim, the conceptual residue of agency, which remains even after the deviance problem is pushed past, derives from something

like what plausibly generates the 'explanatory gap' between phenomenal consciousness and our physical makeup.

On the approach I have in mind, the celebrated gap between conscious states and physical ones is conceptual, not ontological. 'Dualism syndrome' (as we may call it) urges us to infer the metaphysical irreducibility of phenomenal qualia from their conceptual irreducibility. But this urge can itself be diagnosed, according to this approach, without undermining the naturalism it presses against. It derives in effect from a sense/reference confusion: from reflecting on our conscious experience – itself a natural phenomenon – from the internal point of view, under a mode of presentation that is very unlike modes presenting it as a natural phenomenon. More specifically, the 'phenomenal concepts' we deploy when we attend introspectively to the 'feel' of our conscious states are not reducible to the third-person concepts of, say, brains and central nervous systems; nor do the functional roles these sets of concepts play in our thinking coincide. The gap between these items might thus be categorized as unreal, a 'cognitive illusion' to which we succumb when we fail to appreciate the special character of phenomenal concepts and the way these concepts function within our cognitive hierarchy.¹⁰

How precisely phenomenal concepts should be understood is a matter of dispute.¹¹ But this question need not delay us. What appears undeniable is that phenomenal concepts, however exactly they should be construed, are not reducible to third-person, purely 'natural' ones, that there is no a priori connection between 'what it's like' for a subject of conscious experience and the physical (functional, representational) states with which his conscious experience is related. Yet whether or not this recognition safeguards materialism in the mind/body case, it may be extended to explain why agency seems to disappear under 'the objective gaze.'¹²

I suggest, accordingly, that it is because the phenomenology of agency is absent from third-person conceptualizations that action theories developed from this perspective will inevitably appear to ignore something significant.¹³ And in a sense they do. What they leave undressed is how things *seem* to us when we act; they ignore what Carl Ginet (1990) has called the 'actish phenomenal quality.' Of course, a theory of action need not include a description of what it's like to act, because it needn't be judged incomplete as a metaphysical theory if it fails to include such a description. Our problem may therefore be contrasted with the problem of consciousness, because in this case the phenomenal feel of conscious experience is itself the thing that needs

explanation; and the action theorist, for her part, is simply concerned with action. But it does not follow that the phenomenology of agency is irrelevant to an assessment of naturalism about action. For it may be that what is apparently missing from the naturalistic account is a particular *phenomenal* conception of ourselves as sources of our conduct – one that we come by from the internal point of view – and which cannot in the nature of the case be conceptually reduced to any item given in a third-personal analysis.

Here I'm indebted to Horgan et al., who identify 'three especially central elements of the phenomenology of doing: (i) the aspect of self-as-source, (ii) the aspect of purposiveness, and (iii) the aspect of voluntariness' (2003, p. 323). Not surprisingly, the phenomenology they describe is so pervasive as to recede into the background. It may however be surfaced by way of contrast with defective cases. You can imagine being the subject of a deviantly caused action: you can appreciate the experiential difference between involuntarily shaking and rapidly moving your hand, or losing your grip and releasing some item on purpose. And what it's like to trip is very different from what it's like to pretend to be tripping. In typical cases of doing, the phenomenology represents a subject (to herself) as an embodied agent whose behaviour is generated, guided, voluntarily controlled from within – it 'tells us,' so to say, that things are working properly, or that we're at the helm of the ship.

If the phenomenology of doing is in this way intentional, then it has veridicality conditions; it can accurately reflect and it can misrepresent how things are. And it is initially reasonable to believe that the aspect of self-as-source, which Horgan et al. dub the phenomenology of 'immanent generation,' satisfies its veridicality conditions only if some component of any act is in Chisholm's sense 'immanently caused' – produced immediately by the agent, and not by any mere event at all. Indeed, this intuition is I think what's principally responsible for the feeling that naturalism fails to account for 'bringing things about' and gives us mere happenings instead.

Compare Timothy O'Connor, who remarks that a preference for agent-causation over causal theories of (free) action:

may be bolstered by a simple appeal to how things seem to us when we act. It is not, after all, simply to provide a theoretical underpinning for our belief in moral responsibility that the agency theory is invoked. First and foremost... the agency theory is appealing because it captures the way we experience our own activity. It does not seem

to me... that I am caused to act by the reasons which favor doing so; it seems to be the case, rather, that I produce my decision *in view of* those reasons.... Such experiences could, of course, be wholly illusory, but do we not properly assume, in the absence of strong countervailing reasons, that things are pretty much the way they appear to us? (1995, pp. 196–7)

But this intuition may not be as trustworthy as it looks. Certainly it must be admitted that action is not phenomenologically represented as a sequence of events wherein, say, the subject experiences an occurrent desire to *A*, and then idly watches as her body moves about in an *A*-ish way. Nor does the experience of action present it as consisting of bodily motions *being caused* by pertinent mental states – granting that causal processes are sometimes perceived as causal processes, and not simply as strings of contiguous occurrences. Experiences of either sort, Horgan et al. (p. 328) rightly point out, would be unfamiliar and indeed alienating. Still, this does not imply that the phenomenology of doing is non-veridical if, in fact, an agent's 'bringing things about' consists in suitably related psychological and physical events.

Consider here that most 'agent-causation' theories locate the agent-caused event at the start of an extended action, but not as being operative throughout it. Thus according to Chisholm (1964) what the agent causes is a cerebral event of some kind, subsequent to which standard event-causal processes (constituting the action proper) unfold as a consequence of the cerebral event produced by him. Or again, on O'Connor's (1995) approach what the agent causes is an 'action-triggering intention,' which itself causes the events comprising (the rest of) his action and persists throughout the series, so as to guide the agent's behaviour to completion. But the phenomenology does not present things this way: It does not present the agent as 'immanently' discharging some force (e.g.) and then allowing a series of 'transcendent' causes to take over and perform or complete the act on his behalf. Rather, the *sense* of immanent generation remains throughout the duration of an act, and this despite the fact the agent-cause has according to these theories already done its work: I seem directly to bring about my hand-waving at least as much as I seem to bring about whatever nameless cerebral event is supposed to make my hand wave, and I seem to be guiding my behaviour at least as much as my intentions seem to be doing the guiding for me. Yet if *its not seeming like* our actions are purely event-caused provides evidence that they are not, then it likewise provides evidence that whatever remains of our acts once we agent-cause

their beginnings is not accomplished by any event-causes – whether cerebral or intentional – either. At no point, when we act, does it seem as though events within us are doing the actional work for us. Yet all sides agree that at least some degree of event-causation is nevertheless operative in any instance of agency.

This recognition underscores a significant point about the phenomenology of doing, especially as it relates to its veridicality conditions. That action is *not represented as being* event-caused is not to say that it is represented *as being not* event-caused.¹⁴ If the latter were true then the phenomenology of doing would misrepresent its nature, assuming a naturalist stance. (Indeed, it would similarly undermine any theory specifying that action is in large measure constituted by event-causal processes; and so far as I know all agent-causation theories do.) But since the former is true, the phenomenology itself does not unambiguously support the familiar intuitions behind the thought that ‘bringing things about’ cannot consist in event-causation alone. And, intuitively, this isn’t what we should expect the phenomenology to tell us in any case. Its representational function plausibly consists in providing ‘feedback’ to subjects: information that enables us to monitor, supervise and guide the direction of our activity, while alerting us to potential lapses in control and allowing for adjustments to behaviour as needed. But it is hard to see why the usefulness of such a feedback system should depend upon there being an irreducible agent-cause, or why its proper functioning would be compromised if agency consists in event-causes after all. Thus whereas we can affirm that the phenomenology of agency would be non-veridical (assuming naturalism) if it represented action *as being not* event-caused, we cannot infer the falsity of naturalism on the basis that action is *not represented as being* event-caused. For in this latter case the truth-value of naturalism about action is undetermined by the phenomenological evidence; and to say that the phenomenology doesn’t disclose the presence (or absence) of event-causation is not equivalent to affirming that the phenomenology is non-veridical if event-causation is present in action after all.

The approach I’ve put forward is clearly only a sketch, and as such it stands in need of development along several fronts. But it seems to me both promising and worthy of development. In any event, a sustained consideration of the phenomenology of agency as it relates to conceptual analyses of action is certainly overdue, and what analysis remains to be done from the third-person perspective will I think most fruitfully be done in interaction with it.

4 Natural free agency

In closing I want to briefly indicate how the approach sketched here impinges upon the topic of free will, particularly as concerns the viability of ‘causal indeterminist’ theories of libertarian freedom. These theories are libertarian inasmuch as they deny the compatibility of freedom and determinism while affirming the reality of freedom; and they are ‘naturalistic’ because they seek to construct a theory of free agency by employing only those resources available to naturalistic theorists of action.¹⁵

Since these approaches are naturalistic, they are impervious to the complaint that libertarianism requires the inclusion of naturalistic-agent-causal powers and the like – and they are for that reason better placed to avoid a common accusation lodged against libertarian views. But according to libertarian critics of causal indeterminism, this gain in strategic posture is purchased at the cost of the theory’s ability to deliver a genuinely libertarian account of free will. The worry is that an agency consisting in probabilistic event-causes results in ‘basically a compatibilist strategy’ for understanding human freedom, as Hasker (1999, p. 97) puts it. It fails to sufficiently enhance agential control, because if actions result from indeterministic event-causal processes they do not originate with agents themselves.

The assimilation of ‘origination’ with ‘enhanced control’ is widespread, and understandably so. Intuitively, unless an agent directly brings about one action as opposed to any other action he could have performed in those circumstances, he lacks the kind of control over his behaviour required for free will: and it seems he cannot bring his actions about in this manner without an agent-causal capacity. While I lack the space fully to defend this claim here, however, I contend that this assimilation of ‘origination’ and ‘control’ obfuscates two distinct issues. An agent ‘controls’ his behaviour in the most basic sense when he acts intentionally, or when he guides his behaviour in accordance with his reasons for acting so as to achieve the end(s) specified in the content of his intentions. The claim that such control can be captured in purely event-causal processes, irrespective of whether they operate deterministically or probabilistically, is something that contemporary agent-causalists (in contrast to some of their forebears) typically concede. This suggests that the payoff of introducing an agent-cause consists in establishing a strict or literal sense in which agents ‘originate’ their actions, but not in explaining how an agent’s rational or voluntarily control may be more fully ‘enhanced’ by it.

What is striking about the stance of contemporary proponents of agent-causation is that, in opposition to Taylor, Nathan and (the early) Chisholm, they hold that an agent-causal capacity is needed only for *free* agency, and not for agency *as such*: thus 'agent causation should be seen as required for acting with free will, but not for acting,' according to Clarke (1993, p. 192).¹⁶ But making this concession (so I claim) undercuts the need for an irreducible agent-cause at all, even in the case of free will. For if we concede that naturalistic theories of action do not turn 'actions' into 'mere happenings,' and are consistent with agents rationally and purposively bringing about what they do – and isn't this what it means to say that agent-causation is not needed for agency? – then it is unclear why the causal-indeterminist should be obliged to add an agent-cause to explain how agents 'originate' their actions in the indeterministic setting, or what it is such theorists are missing if they don't. If an agent-cause is needed at all, I suggest, it is needed for agency simpliciter – so as to secure the distinction between 'mere happenings' and 'bringing things about' – and it is needed for free agency only derivatively. Thus if it is not needed for agency simpliciter, then it is not required to explain how agents freely 'bring about' their behaviour either – granting the obvious, that the conditions for free agency may *otherwise* outstrip the conditions for agency itself. The debate about causal indeterminism therefore cannot be resolved in isolation from this more fundamental issue we have been considering.

I have argued that dissatisfaction with naturalistic theories stems from the appearance that the agent plays no role in originating 'his conduct,' that this appearance may be diagnosed as being generated by the conceptual gap between first and third person perspectives on action; that this gap is itself explainable with reference to the phenomenal concepts we deploy when we consider agency from the interior perspective; and that the existence of this conceptual gap does not entail that 'bringing things about' cannot consist in suitably related event-causes. The dissatisfaction with naturalistic theories of libertarian freedom, I claim, should be understood and dealt with in a similar fashion. For the problem of 'bringing things about' is not created by causal-indeterminist theories specifically; and if it is answerable naturalistically at all it is answerable naturalistically for them, too.

Notes

1. E. J. Lowe (2008) has argued that agent-causation is continuous with causation throughout the rest of nature, because causation is a relation between

substances that produce changes in others. The proposal is interesting and worth a serious look, but I shall not evaluate it here.

2. Thus Melden: 'It is futile to attempt to explain conduct through the causal efficacy of desire – all *that* can explain is further happenings, not actions performed by agents.... There is no place in this picture... even for the conduct that was to be explained' (1961, pp. 128–9).
3. See Bishop (1983).
4. See Kim (1989).
5. See Ekstrom (1993, pp. 603–5).
6. I've excluded from this definition Ekstrom's libertarian specification that the inputs to the evaluative faculty 'cause but do not determine' the decisively formed preference.
7. See Ekstrom (1993, pp. 606–10).
8. Cf. Mele (2003, p. 220).
9. It is important to note that 'adding' such mental items to the mix here does not imply that what it's like to act has been successfully captured from the 'objective' standpoint. The claim is rather that affirming the existence and efficacy of such items does not help us to see how agency as we experience it from the internal perspective is compatible with event-causation (mental or physical) alone.
10. The phrase 'cognitive illusion' is Michael Tye's (2000).
11. See the essays in Alter and Walter (2006).
12. To clarify: it is consistent with my proposal that phenomenal concepts may not help explain consciousness naturalistically, for it's possible that the experience of agency is metaphysically irreducible *as a phenomenal state* even if the phenomenon it represents is naturalistically explicable.
13. A person can of course enjoy phenomenological experiences of agency while she is theorising about agency from the third-person standpoint, but it does not follow from this that the phenomenal content is itself contained within the third-person theoretic conceptualization of action.
14. Cf. Horgan et al. (p. 335).
15. See Kane's (1996) approach.
16. See for discussion and references O'Connor (2000, pp. 49–55) and Kane (1996, pp. 120–3).

Bibliography

- Alter, T., and Sven W., (eds) (2006), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (New York: Oxford University Press).
- Bishop, J. (1989) *Natural Agency: An Essay on the Causal Theory of Action* (Cambridge: Cambridge University Press).
- (1983), 'Agent-Causation,' *Mind* 92, 61–79.
- Chisholm, R. (1966), 'Freedom and Action,' in Lehrer (ed.), *Freedom and Determinism* (New York: Random House).
- (1964), *Human Freedom and the Self*, Lindley Lecture (Lawrence, Kans.).
- Clarke, R. (1993), 'Toward a Credible Agent-Causal Account of Free Will,' *Noûs* 27, 191–203.

- Ekstrom, L. W. (1993), 'A Coherence Theory of Autonomy,' *Philosophy and Phenomenological Research* 53, 599–616.
- (2000), *Free Will: A Philosophical Study* (Boulder, CO: Westview Press).
- Ginet, C. (1990), *On Action* (Cambridge: Cambridge University Press).
- Hasker, W. (1999), *The Emergent Self* (Ithaca, NY: Cornell University Press).
- Horgan, T., Tienson, J., and Graham, G. (2003), 'The Phenomenology of First-Person Agency,' in Walter and Heckmann (eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action* (Imprint Academic).
- Hornsbj, J. (1993), 'Agency and Causal Explanation,' in Heil and Mele (eds.), *Mental Causation* (Oxford: Oxford University Press).
- Kane, R. (1996), *The Significance of Free Will* (New York: Oxford University Press).
- Kim, J. (1989), 'Mechanism, Purpose, and Explanatory Exclusion,' *Philosophical Perspectives* 3, 77–108.
- Lowe, E. J. (2008), *Personal Agency: The Metaphysics of Mind and Action* (Oxford: Oxford University Press).
- Melden, A. (1961), *Free Action* (London: Routledge and Kegan Paul).
- Mele, A. (2003), *Motivation and Agency* (Oxford: Oxford University Press).
- Nagel, T. (1986), *The View from Nowhere* (New York: Oxford University Press).
- Nathan, N. M. L. (1975), 'Materialism and Action,' *Philosophy and Phenomenological Research* 35, 501–11.
- O'Connor, T. (1995), 'Agent-Causation' in O'Connor (ed.), *Agents, Causes, and Events* (New York: Oxford University Press).
- (2000), *Persons and Causes: The Metaphysics of Free Will* (New York: Oxford University Press).
- Sartre, J. (2001), *Rationality in Action* (Cambridge, MA: MIT Press).
- Shoemaker, S. and Swinburne, R. (1984), *Personal Identity* (Oxford: Blackwell).
- Taylor, R. (1966), *Agent and Purpose* (Englewood Cliffs, NJ: Prentice Hall).
- Tye, M. (2000), *Consciousness, Color, and Content* (Cambridge, MA: MIT Press).
- Velleman, J. D. (1992), 'What Happens When Someone Acts?' *Mind* 101, 461–81.
- (2000), *The Possibility of Practical Reason* (Oxford: Oxford University Press).

6

Interpretation: Its Scope and Limits

Uriah Kriegel

According to interpretivism, all there is to having an intentional property is being best interpreted as having it. I present a regress-or-circularity argument against this. In Section 1, I elucidate interpretivism, and in Section 2, I present the argument against it.

1 Interpretivism

It is sometimes said that intentionality is a relation, though a special one, one that does not require the existence of all relata. In a pair of previous papers (Kriegel 2007, 2008a), I have argued that this is non-sense, akin to positing a special monadic property that can be instantiated even in the absence of something that instantiates it. Instead, I have argued for an adverbialist account of intentionality, according to which intentionality is a certain kind of intrinsic modification – a non-relational property of some mental states. What both adverbial and relational metaphysics of intentionality share is a certain objectivism about intentionality: the idea that there are objective facts of the matter about the correct assay of intentional properties. This assumption can be rejected, however. It is rejected, most notably, by interpretivist approaches to intentionality. It is therefore my goal, in this chapter, to argue against the viability of such interpretivism.

The term 'interpretivism' is used in a variety of different ways throughout the humanities. Here I will be concerned with the view of intentionality and content associated with Dennett and Davidson when they say such things as 'all there is to really and truly believing that p (for any proposition p) is being an intentional system for which p occurs as a belief in the best (most predictive) interpretation' (Dennett 1981, p. 72) and 'what a fully informed interpreter could know about what a speaker