Prolegomena 9 (2) 2010: 199-242

Frankfurt on Second-Order Desires and the Concept of a Person

CHRISTOPHER NORRIS

School of English, Communication, and Philosophy, Cardiff University, Humanities Building, Colum Drive, Cardiff, CF10 3EU, Wales
norrisc@Cardiff.ac.uk

ORIGINAL SCIENTIFIC ARTICLE / RECEIVED: 04–04–09 ACCEPTED: 01–10–09

ABSTRACT: In this article I look at some the issues, problems and (it seems to me) self-imposed dilemmas that emerge from Harry Frankfurt's well-known essay 'Freedom of the Will and the Concept of a Person'. That essay has exerted a widespread influence on subsequent thinking in ethics and philosophy of mind, especially through its central idea of 'second-order' desires and volitions. Frankfurt's approach promises a third-way solution to certain longstanding issues – chiefly those of free-will *versus* determinism and the mind/body problem – that have up to now resisted the best efforts of philosophical deliverance or therapy. It looks very much like the kind of answer that would avoid the 'high priori road' of any Kantian or suchlike metaphysical approach by adopting a broadly naturalized conception of human moral agency while not going so far down the path toward wholesale ethical naturalism as to lose the benefits (of personhood, choice, self-knowledge, and at any rate relative autonomy) that come with the Kantian conception. However I suggest that this appearance is deceptive and that Frankfurt's way of addressing these issues – especially his leading idea of second-order desires and volitions – lies open to a long-familiar range of objections from both a naturalist (anti-Kantian) and a strong autonomist (anti-naturalist) quarter. More specifically, I show that his notion of moral will as possessing a multiplex structure whereby higher-order volitions can reject or countermand the promptings of unregenerate first-order desire is one that must inherently give rise to various problems of a logical, metaphysical, and – most importantly in this context – ethical character. I conclude that a thoroughgoing naturalism is the only response that can meet the kinds of challenge increasingly mounted from various scientific quarters, notably those of neurophysiology and cognitive psychology.

KEY WORDS: Desires, Frankfurt, mind/body problem, naturalism, will.

I

Harry Frankfurt's 1971 essay 'Freedom of the Will and the Concept of a Person' is one of the best known recent contributions to philosophy of mind, moral philosophy, and – in the broad and I dare say proper if not currently received sense – cognitive psychology.[1] Indeed its main line of argument concerning the intimate link between our concept of human personhood and the distinctively human possession of 'second-order' desires or volitions has probably exerted as wide an influence and been subject to as many detailed re-workings as any other in the past half-century.[2] Nor is this at all surprising given Frankfurt's lucid and graceful yet pointedly analytic style, his regular choice of telling examples, and – more than anything – his having here tackled a complex of themes that not only go very much to the heart of recent philosophical debates but are apt to strike most reflective human beings from time to time.

Thus his essay seeks to articulate just what it is that makes such beings 'human' according to a certain criterion – or self-understanding – of genuine personhood that sets them apart both from non-human animals and also from those among their conspecifics who fall short of humanity in that regard. Moreover it does so in an intellectual climate marked by the increasingly rapid accumulation and diffusion of new advances in just the areas (e.g., neurophysiology, genetics, and the more naturalistic varieties of cognitive science) that many people – philosophers and others – regard as a real or potential threat to any such conception. That is, those disciplines are viewed as laying siege to the basic attributes that are taken to mark human beings out as truly human since capable of actions, choices, decisions, or commitments which involve both an unconstrained exercise of will and the condition that such will be exercised in accord with the agent's best idea of what makes for good conduct or a life well lived by her own and other persons' ethical lights. Where Frankfurt's approach has proved so attractive is through the promise of offering a viable alternative to naturalism in its hard-line physicalist and (presumptively) determinist mode while not going along with any downright anti-naturalist or dualist creed that would fly in the face of so much presently accredited scientific knowledge. Thus it offers a way of thinking about these issues which avoids any too direct invocation of Kantian ethical precepts or imperatives yet still leaves room – *via* the doctrine of 'second-order' desires – for a kind of self-transcendence through the bringing to bear of higher, more reflective or worthy volitions on those that belong to our unregenerate

[1] Frankfurt (2003). See also Frankfurt (1988).

[2] For further discussion from a range of viewpoints, see van Inwagen (1983); Kane (2004); O'Connor (2004); G. Strawson (1986); Watson (2003).

or animal natures. I should add that Frankfurt is not overly attached to this latter sort of talk, whether in its Kantian or its more traditional Christian-moralizing mode. All the same his essay falls in with such ideas at least to the extent of making it a requirement for the status of personhood (as distinct from mere membership of the human species) that the beings in question should possess that power of self-critical and self-evaluative thought which enables them to pass their first-order wants in review and, where appropriate, form the desire to cultivate a different set of wants or rank their existing set differently.

It seems to me that Frankfurt's essay cannot in the end make good on its promise to vindicate the claims of human autonomy or free-will through this notion of our pre-reflective desires being subject to a higher-level process of quasi-juridical assessment as to their fitness or otherwise when measured against our best conceptions of the human good. For one thing, and most problematically, it fails to address the standard objection to all such ideas of progressive ascent from one to another level of increasing descriptive, explanatory, or justificatory power in whatever specific context of debate. Thus the point is often made – most famously with regard to old-style logical positivist attempts to draw a firm line between first-order 'material-language' statements and second-order (metalinguistic) statements about those statements – that this procedure not only opens the way to a potential vicious regress but fails to establish any adequate formal (as opposed to pragmatic) grounds for supposing that such a line can be drawn with the required degree of conceptual or logico-semantic precision.[3] There is no need here to rehearse the manifold ways in which this difficulty first came to light and thereafter tended to re-surface with all the more disturbing or disruptive force in the wake of various claims to have resolved it on viable, that is to say, regress-blocking and clearly specified terms.[4] That account would be something like a potted history of the mainstream analytical enterprise from its early stage of programmatic assurance, following Frege and Russell, that such distinctions could reliably be drawn to its wholesale demolition – as the orthodox version goes – in Quine's 'Two Dogmas of Empiricism', his famous attack on every last vestige of the analytic/synthetic dualism or the distinction between empirically verifiable 'matters of fact' and logically self-evident 'truths of reason'.[5]

Of course this particular chapter of developments has to do primarily with issues in epistemology and philosophy of logic and language, rather

[3] See for instance Tarski (1956); Carnap (1969).

[4] For a good sampling of views, see Ayer (1959).

[5] Quine (1961); also – for a more detailed account of these developments – Norris (2004) and (2006).

than with issues in ethics and philosophy of mind such as occupy the main focus of interest in Frankfurt's essay. Nevertheless, as I have said, they do have a bearing on his claim that the possession of second-order desires or volitions is what makes for – indeed, what properly constitutes – the character of genuine human personhood as distinct from those various creaturely traits that mark human beings as just another, albeit highly evolved and psychologically complex kind of animal. Frankfurt himself concedes that there might be difficulties in this direction if it is allowed that 'a person may have . . . desires and volitions of a higher order than the second' (p. 91). For of course this idea brings along with it the possibility that the series thus envisaged might have no rationally assignable stopping-point and threaten not merely a notional regress – a problem for logically-minded philosophers – but a dissolution of authentic personhood in just the sense of that term that Frankfurt is seeking to establish. After all, if 'there is no theoretical limit to the length of the series of desires of higher and higher orders', then it would seem that 'nothing but common sense and, perhaps, a saving fatigue prevents an individual from obsessively refusing to identify himself with any of his desires until he forms a desire of the next higher order' (p. 91). However, he thinks, there is no need to entertain such extravagant worries since the stopping-point may always occur – and the threatened regress thus come to a merciful halt – at the stage of some decisive personal commitment which either exerts such a powerful grip on any higher-order volitions that they cannot but fall into line or else renders their existence altogether superfluous. In such cases it is a matter of indifference 'whether we explain this by saying that this commitment implicitly generates an endless series of confirming desires of higher orders, or by saying that the commitment is tantamount to a dissolution of the pointedness of all questions concerning higher orders of desire' (p. 92).

Nevertheless these solutions have a somewhat makeshift or stopgap air, as if to ward off the regress objection by declaring it simply irrelevant for practical purposes or for anyone who knows 'from inside' what it is to have arrived at such a moment of decision. Indeed they are reminiscent of patch-up attempts in other areas of thought like Russell's Theory of Types, i.e., his rule against self-predicative expressions or those that crossed the line between different orders of statement and thereby threatened serious trouble for the enterprise of set-theoretical thought and the attempt to place mathematics on a wholly consistent logical basis. Only thus, he believed, was it possible to ward off the kinds of damage that would otherwise be wrought by the paradoxes of self-reference or the potential for mischief contained in such expressions as 'the set of all sets that are not

members of themselves'.[6] If this all seems pretty remote from the context of Frankfurt's argument – since concerned with issues of a purely formal or logico-conceptual as opposed to a subjective, experiential and ethico-evaluative nature – then one might pause to consider how often Frankfurt appeals to the language of first- and second-order desires or volitions, a language that cannot but raise certain questions of that same Russellian sort. After all, it is his leading contention that the criterion of genuine personhood is precisely the possession – *ex hypothesi* denied to non-human animals – of a capacity for freely-willed discriminative choice amongst the various first-order desires that are taken to constitute a separate realm of altogether 'lower' since unreflective or instinctually driven needs, cravings, instincts, appetites, and so forth. Then again, they might be construed as belonging to a somewhat more thoughtful stage of reflective ascent but only in so far as this is held subject to a further process of evaluation brought to bear from the next-higher level of deliberative judgment which thereby ensures – crucially for Frankfurt's case – that there remains the necessary leeway for free-will and responsibly exercised choice.

So Russell-type problems do have a bearing here despite his above-cited confident claim that the regress objection can successfully be met just by seeing that it loses any force or point once we grasp that some higher-order volitions have a purposive strength and depth of commitment that renders them proof against any such regress-based or paradox-mongering objection. According to Frankfurt, it is possible to place a definite, non-arbitrary limit on the series of thoughts-about-thoughts-about-thoughts or desires-to-desire-to-desire (etc.) simply by grasping the salient fact that some of our decisions – those with this identity-shaping power – are such as to render that objection otiose, along with the related puzzle about self-predication that is apt to arise if one asks how any clear and subjectively relevant (i.e., other than purely logical) distinction can be drawn between those various levels. Thus '[w]hen a person identifies himself *decisively* with one of his first-order desires, this commitment "resounds" throughout the potentially endless array of higher orders' (p. 91). However this leaves open the obvious question as to what room is left for the joint desiderata of freedom and responsibility – the constituents of full human personhood on Frankfurt's account – if that regress can blocked only by appeal to such a locus of self-determined and self-determining choice. For it then seems that this concept of a halt to the 'endless array' must also entail certain clearly-marked restrictions on the scope of that freedom if indeed it is to allow for any well-defined idea of personal responsibility. That is, there has to come a stage where the notion of choice as somehow both

[6] See for instance Russell (1930); also Russell (1994).

determined and determining – the former in so far as it pertains to lower-order desires, the latter in so far as those desires are subject to reflective evaluation – ceases to operate as the means of conserving a space for autonomous thought and becomes something more like the means of fixing a limit to the otherwise vertiginous range of possibilities thus opened up.

My point is that Frankfurt's use of the distinction between first- and second-order desires or volitions is one that inevitably leads to problems since it confronts thinking with a choice between (1) the kinds of potentially endless regress that ironists in the German romantic tradition found utterly engrossing but which possess fewer charms for modern analytic philosophers, and (2) the kind of regress-blocking move which decrees that such complexities must have an end and that this involves the determination to make certain desires or volitions absolutely and completely one's own.[7] Yet it is then very hard to avoid the question implicitly posed by that seeming quirk of 'ordinary language' which enables us to say, as if without conceptual strain, that we are *determined to do* whatever we freely choose, or that our choice is a matter of *determination* to pursue some given life-plan, project, or mode of conduct itself undertaken – so the presumption goes – of our own spontaneous or unconstrained (hence responsibly exercised) will. It seems to me that Frankfurt's first-order/second-order way of framing these issues leads him into various, more or less nuanced or qualified statements of the thesis which all give rise to this unresolved – maybe unresolvable – conflict between the two senses of 'determine'. Thus on the one hand it signifies 'cause to think, feel, or act in a certain way through certain crucially *determinative* factors of a physical, environmental, social, or cultural sort', while on the other it carries the contrary or at any rate strongly opposed sense: 'decide, resolve, or *determine* to proceed in accord with one's best judgment or freedom to choose the most rational, desirable, advisable, or ethically responsible option'. Frankfurt takes it that metaphysical debates about free-will *versus* determinism had much better be recast as debates about the relationship between first- and second-order desires, that is, the extent which the former may be influenced, guided, enhanced, reformed, or even subject to outright veto by the power of adjudicative challenge or review exercised by the latter. But he nowhere explains how that power of higher-level endorsement or rejection can itself be legitimized – thought of as rightfully binding – if its edicts have the kind of self-validating force that comes of their regress-blocking role and their function in defining what shall ultimately count as the motivating interest (or complex of interests) with which a person is so deeply identified that it becomes the very ground or criterion of personhood in their particular case.

[7] On this chapter in German post-Kantian philosophy, see Simpson (1988).

II

This raises the question as to how there could exist that margin of freedom that Frankfurt considers indispensable if we are to make any sense of likewise indispensable notions such as those of will, decision, commitment, dedication, resolve, reflective or self-critical judgment, and responsibly exercised choice. For it would seem to place a drastic limit on any such freedom – rather, to exclude its very possibility – both in so far as lower-order desires are thought of as properly subject to being overruled without right of appeal and also in so far as higher-order volitions are themselves determined in both senses of the term. That is, they seem to figure on Frankfurt's account as products of human will – of the determination to follow through on some chosen course of action or mode of conduct – but a will that is constrained (along with all those lower-order desires that fall within its jurisdiction) by the fact of its so completely pervading the person's motivational mindset as to leave no room for deviation from the path marked out or determined in advance. In this respect Frankfurt's way of framing the issue can be seen as raising all the same problems as Kant's tortuous attempts to explain how the moral will can be conceived as an autonomous source of laws for its own proper guidance, or how the edicts, maxims, and imperatives of practical reason can be thought of both as issuing from and as bearing upon the subject in its self-legislative role.[8] Thus the word 'subject' in Kantian moral discourse is one that splits, like the word 'determine', into two quite distinct and indeed contradictory senses, with the first strongly linked to notions of the subject as locus of autonomy, agency, and the will to realise intentions formed through an exercise of independent thought while the second connotes the state of being subject – passively obedient – to laws handed down by some higher authority.

That this problem goes deep into the structure of Kant's entire critical project can be seen from the fact that it crops up again in relation to his epistemological arguments in the First *Critique*. More specifically, it concerns the gap that opens up between phenomenal (or sensuous) intuitions and concepts of understanding, a gap that is not so much closed or bridged as pasted over by Kant's notoriously vague talk of 'judgment' and 'imagination' as mediating faculties.[9] In the present context what it helps to bring out is the extent to which discourse on these topics since Kant has been marked – not to say hamstrung or hobbled – by a strain of deep-laid dualist thinking that constantly re-surfaces to vex any project aimed toward the

[8] Kant (1976).
[9] Kant (1998).

final overcoming of all such 'metaphysical' residues.[10] With Frankfurt, it appears in the frequent signs of conceptual strain around statements to the general effect that, for all practical (including legitimate philosophical) purposes, the issue of free-will *versus* determinism comes down to the question whether subjects or agents are capable – unlike non-human animals – to have the sort of will or will themselves to cultivate the sorts of first-order desire that they wish to have *qua* reflective individuals or persons in the plenary sense of that term. Thus '[j]ust as the freedom of an agent's action has to do with whether it is the action he wants to perform, so the question about the freedom of his will has to do with whether it is the will he wants to have' (p. 90). However it is not at all clear how this shift to a second-order level of analysis does anything to ease the conceptual strain involved in understanding how freedom, autonomy or genuine volition can have any place in this notion of a will that is itself the outcome of a 'wanted' (hence willed) commitment, one that may either – at risk of vicious regress – be referred to some yet higher stage of wanting and willing or else taken as a terminal point and therefore as simply not open to reflective or critical-evaluative thought. According to Frankfurt this problem simply doesn't arise since '[i]t is in securing conformity of his will to his second-order volitions . . . that a person exercises freedom of the will' (p. 90). But again it is hard to see that there is room for freedom of will where such freedom is conceived as bringing will into conformity with second-order volitions, or that those volitions can themselves be conceived as somehow introducing the necessary space for freedom to regain its foothold.

Of course it will be said that there are two sorts of 'will' in question here, the first- and second-order sorts, and that Frankfurt's whole purpose in writing the essay was to point up the crucial distinction between them precisely as a means of explaining how genuine personhood, autonomy and freedom differ from the mere unfettered pursuit of first-order wants and desires. However this claim is more easily stated as a matter of abstract principle than convincingly borne out as a matter of finding room for those higher-level modes of volition, along with that vital margin of freedom in the absence of which there could be no making sense of the higher/lower distinction except as a merely technical device with no further moral or significant human implications. The main trouble lies in grasping what it could mean for somebody – let us say a rational, reflective, second-order-volition-forming individual who meets all Frankfurt's

[10] For further discussion and extended commentary on various recent approaches to these problems with Kant's doctrine of the faculties, see Norris, 'Kant disfigured: ethics, deconstruction, and the textual sublime', in Norris (1993: 182–256) and Norris (2000).

specified criteria for genuine personhood – to *freely wish* that her first-order desires should arrange themselves in this or that order of priority, or *freely prefer* that one such desire should take second place to another (conflicting) desire on well-considered moral, prudential, long-term beneficial, or other such grounds. After all, to the extent that those wishes are indeed well-considered and subject to reflective assessment of the kind that, in his view, constitutes the *sine qua non* of authentic second-order will-formation they are *for just that reason* not 'freely' arrived at – singled out for adoption from amongst some range of competing interests or conflicting priorities – as if through a rationally underdetermined or under-motivated act of choice. Rather they must be taken to result from a more-or-less lengthy and complex process of rational deliberation whereby the subject achieves, or seeks to achieve, that measure of reflective equilibrium that would allow him or her to draw the appropriate conclusion and, where called for, translate it into action of a likewise appropriate sort. To suppose otherwise is to endorse the doctrine of 'doxastic voluntarism' according to which we can, often do and indeed often should choose or decide what's best to believe on the basis of preference or inclination as opposed to the basis of empirical evidence or rational-demonstrative warrant.[11] Among further objections to this idea – one espoused by latter-day pragmatists or 'strong'-descriptivists like Richard Rorty – is the standing temptation it offers to indulge a strain of Mary-Poppins-like wishful thinking as well as those other, more dangerous or sinister kinds of self-deception that Bertrand Russell anatomised in his response to William James concerning the latter's essay 'The Will to Believe'.[12] Also there is the case – of major import for the history of growing resistance to various forms of religious and political persecution – that since beliefs are properly arrived at through a process of persuasion by the best evidence, arguments or reasons to hand and therefore cannot (or should not) come down to a matter of wish-driven preference or choice therefore it is grossly unjust to punish people on account of their heterodox beliefs.

So there are problems with Frankfurt's central idea that one can keep open that vital space for the exercise of human freedom by making it a criterion of personhood that persons – as distinct from non-human animals or sub-personal human beings – should be capable of having second-order volitions and moreover of identifying with them in so deep or self-constitutive a way as to block the threatened regress from stage to stage through endless orders of reflection. What is chiefly problematic is just that equa-

[11] See Norris, 'Ethics, Autonomy, and the Grounds of Belief', in Norris (2006: 130–154) and Norris (2005).

[12] James (1907) and (1909); Russell (1999); also – for another vigorous counter-statement to the Jamesian pragmatist approach – Clifford (1999).

tion between regress-blocking potential and fixity of purpose – or 'determination' in both senses of the word – as that which distinguishes genuine (mainly second-order) volition from mere first-order desire. For this inescapably poses the question with regard to doxastic voluntarism, that is, the question as to whether we can make any sense of the idea that beliefs are volitional, i.e., that they result from some exercise of will or decision to pursue one or other of the options amongst our preferentially ranked array of first-order wants or desires. That Frankfurt subscribes to some such idea – that in his view it is what makes room for freedom and sets persons apart from sub-personal (whether human or non-human animal) creatures – is strongly implied by numerous statements in the course of his essay. Thus, for instance, 'it is having second-order volitions, and not having second-order desires generally, that I regard as essential to being a person' (p. 86). Here he seems to be drawing a three-sided distinction between (1) non-human animals with first-order desires plain and simple, (2) human persons who have both first and second-order desires along with second-order volitions, and (3) that intermediate and in some way defective class of human beings – Frankfurt rather quaintly calls them 'wantons' – who are so much at the mercy of their first-order promptings, instincts, impulses, lusts, or whatever that they fall short of fully-fledged personhood. Of the latter he writes that they 'are not persons because, whether or not they have desires of the second order, they have no second-order volitions' (p. 86).

Actually there is some looseness of phrasing or terminological slippage at points in the course of Frankfurt's essay since earlier on he can be found positing a bipartite distinction between those 'many animals' that 'appear to have the capacity for . . . "desires of the first order", which are simply desires to do or not to do one thing or another', and the human animal who uniquely 'appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires' (p. 83). As the argument develops so Frankfurt comes to lay greater emphasis on the further – in his view cardinal – distinction between, on the one hand, persons who possess the reflective *and* the volitional capacity to turn their second-order desires into a fully motivating power of will and, on the other, 'wantons' in whom those second-order desires are too weak, confused, or indecisive to have any such effect. What this shows, I think, is his increasing sense of the above-mentioned problems with his basic first-order/second-order dualism. To begin with there is the problem about infinite regress to which one, albeit debatable solution is the citing of volitions (rather than desires) as the ultimate or buck-stopping locus for ascriptions of genuine second-order will. Then there is the problem as to why second-order desires should be thought of as any more free – or

any less subject to various heteronomous drives and compulsions – than first-order desires. That the term 'heteronomous' imposes itself here as the best, most philosophically pointed or relevant term in this context is again a sure sign of just how deeply Frankfurt's thinking is caught up in the formal structure and, resulting from that, the various internal conflicts and antinomies of Kantian moral philosophy. Just as Kant conceives the 'autonomy' of practical reason in terms of its being held to account by a moral law of which it is somehow both author and compliant or non-compliant subject, so likewise Frankfurt appears to conceive volition as that which can somehow be called upon to reconcile the realms of desire (whether first- or second-order) and will (here cast in a jointly legislative, executive, and juridical role). This is indeed what he regards as the difference between 'will' in its everyday or normal philosophic usage as applied to human agents and 'will' in the distinctive sense that it acquires in the course of his own argument. As Frankfurt puts it, the notion of the will 'is not the notion of something that merely inclines an agent in some degree to act in a certain way', but is rather 'the notion of an *effective* desire – one that moves (or will or would move) a person all the way to action' (p. 84).

If we are here pointedly encouraged to think of 'persons' rather than 'agents' the reason is doubtless that Frankfurt sees – or suspects that readers will be apt to see – too close a relation between talk of agents or acts and the idea that first-order desires may carried right through to the point of direct implementation without any need for complicating detours through the realm of second-order reflection upon them. One is reminded of Hamlet's self-reproachful brooding on the contrast between his own state of chronic vacillation – 'sicklied o'er with the pale cast of thought' – and the purely impulsive, unthinking willingness of a soldier like Fortinbras to take any risk for the sake of some perhaps worthless or doomed venture.[13] Of course Frankfurt's point is just the opposite of Hamlet's: that in the absence of second-order reflection on the promptings of first-order impulse we should lack what most conspicuously sets human persons apart from both non-human animals and 'wantons', namely the capacity for just that sort of self-aware, self-critical, and self-evaluative thought that may sometimes – as perhaps in Hamlet's case – exert an overly inhibiting effect on the power of decisive or resolute action. Thus, for Frankfurt, it consists in the reflective step back from a desire-driven will that all too readily translates into action, or in the pause for thought required by any such refusal to simply go along with the force of first-order instincts, impulses, or appetites. More than that: it consists in the difference between the kind

[13] William Shakespeare, *Hamlet*, Act 3, Scene 1, line 85.

of motivational force exerted by those first-order promptings even when subject to the rule of some purposive coordinating will and volitions of a higher, reflective-evaluative kind such that the will is held in check – or prevented from putting its purposes too quickly into practical effect – by a critical tribunal (call it the voice of conscience or of better judgment) which may or may not turn out to exercise its power of veto.

Here of course we are very much in Kantian country and obliged to fall back, as if necessarily, on legal or juridical idioms and metaphors. Such talk becomes well-nigh unavoidable if one thinks, like Frankfurt, constantly in terms of 'higher' and 'lower' faculties of mind and of these as possessing more or less warrant, authority, or power of command according to their place on a scale that is conceived in strongly hierarchical terms. Moreover, whether taken metaphorically or at face value, those terms connote something like an ordering of moral worth that runs from 'mere' (non-human animal or human sub-personal) desire, *via* the various intermediate levels of purposive will-formation, to the fully achieved capacity for weighing different possibilities and deciding between them on the basis of reflective and autonomous judgment rather than direct impulse or drive. My point is not so much that this smuggles in certain ideological values under cover of a 'purely' philosophical approach but rather that it shows how pervasive is the influence on Frankfurt's thought of those Kantian ideas about the nature, scope, and proper function of the faculties, along with their due degrees of role-specific subordination one to another. No doubt Kant's doctrine of the faculties is one that he would regard, like most philosophers nowadays, as metaphysically over-mortgaged and hence as meriting serious interest only in a suitably naturalised, de-transcendentalized, or scaled-down (e.g., Strawsonian) descriptivist form.[14] However it is not hard to make out the lineaments of that same Kantian doctrine in Frankfurt's leading premise that the truly definitive mark of genuine human personhood is the capacity to attain that measure of critical-evaluative detachment from our first-order wants which permits the formation of a second-order will superior to and exempt from the impulsions of rationally unconstrained or unreflective desire. For on this account there is no escaping the idea of an ascent through successive, increasingly complex and more fully human orders of conscious and reflective thought, or again – perhaps more to the point – a descent from that level of achieved autonomous personhood through stages on the downward path to a level of non- or pre-human animal drive. What is at any rate clear from numerous passages in Frankfurt's essay is the fact of

[14] See P. F. Strawson (1966).

his subscribing to a uni-directional or top-down conception of that which fixes the essential difference between human and other sentient beings.

Frankfurt's essay was published in 1971 at a time when that distinction had not yet been challenged with anything like the range of arguments more recently brought against it by proponents of a radically de-anthropomorphised approach to these issues, among them Peter Singer.[15] All the same there is a certain brisk assertiveness about his remark that 'we do in fact assume . . . that no member of another species is a person', or that 'one essential difference between persons and other creatures is to be found in the structure of a person's will', or again – more strikingly – that '[i]t does violence to our language to endorse the application of the term "person" to those numerous creatures which do have both psychological and material properties but which are manifestly not persons in any normal sense of the word' (pp. 81–2). These sentences all occur within the first two pages of Frankfurt's essay and take rise from his opening critique of P. F. Strawson's conception of the person (in his book *Individuals*) as just that type of entity to which can be ascribed both predicates that specify physical or corporeal characteristics and predicates that specify states of mind, intentions, beliefs, psychological conditions, and so forth.[16] What Frankfurt objects to about this conception is the fact that it seemingly admits to the category 'person' a variety of other (non-human) animal beings in relation to which, by his own lights, we should adopt a quite different attitude and not 'do violence' to language by sinking – or even occasionally blurring – the relevant line of demarcation. Thus it is safe to suppose that Frankfurt would have raised a strong and principled voice of dissent had he come across Singer's proposal that certain non-human animals – among them but not exclusively the higher primates – should properly be accorded the basic rights that are taken to go with the possession of personhood on any unprejudiced understanding of the term, that is, quite apart from the largely irrelevant (as Singer sees it) issue of biological species-membership.[17] For Frankfurt, conversely, 'no animal other than man … appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires' (p. 83).

III

I think it is liable to strike the reader of these and other passages in Frankfurt's essay that his insistence on the point has an air of protesting too

[15] See P. F. Strawson (1966); also (1959).
[16] P. F. Strawson (1959).
[17] Singer (1990); also Singer (1985); Regan and Singer (1976).

much, or of fending off doubts on this score – doubts concerning the security of just that prescriptive line of demarcation – that have been steadily gathering in strength since the time (actually well before Darwin) that science started to point toward an outlook of thoroughgoing naturalism with regard to human beings and their status *vis-à-vis* other animals.[18] Here I might instance some marvellously deft observations in Agamben's *The Open* concerning this pressure on received (whether religious or secular-humanist) ideas of human exceptionalism and the various ways in which thinkers of differing doctrinal adherence managed to evade or accommodate the challenge of an emergent evolutionary-naturalist worldview.[19] It seems to me that Frankfurt's essay is a late offshoot of the same dilemma confronted by those who have refused to accept that worldview at its full and, as they see it, humanly degrading or ethically debilitating force while none the less declining to seek refuge in the alternative appeal to a realm of anti-naturalist (e.g., religious or Kantian) values that are taken to transcend any such grossly reductive physicalist approach.

That we are here addressing matters of a plainly metaphysical if not directly theological import is nowhere more evident than in a passage where Frankfurt stakes out the ground – or the region of conscious and self-conscious or reflective being – which belongs exclusively to human persons, or to those capable of attaining genuine personhood.

> The concept of a person is not only, then, the concept of a type of entity that has both first-order desires and volitions of the second order. It can also be construed as the concept of a type of entity for whom the freedom of the will may be a problem. This concept excludes all wantons, both infrahuman and human, since they fail to satisfy an essential condition for the enjoyment of freedom of the will. And it excludes those suprahuman beings, if any, whose wills are necessarily free. (p. 89)

This passage can be heard to echo any number of theologically inspired disquisitions concerning the proper human place – more precisely: the range of niches available to humans of various types – on the 'great chain of being' that was taken to run all the way from God, *via* the hierarchy of angels, to human beings, non-human animals, and thence on down through the sundry grades of vegetable or wholly inanimate mineral life.[20] Moreover it seems to share their concern with defining what constitutes the human norm in contradistinction on the one hand to that which falls outside and below the norm and on the other hand to that which so far

[18] See especially Agamben (2004); also Empson (1951).

[19] Agamben (2004).

[20] For a classic study of this doctrine in its various historical forms, see Lovejoy (1936).

transcends it as simply to preclude all the conflicts and complexities of human existence.

Not that Frankfurt is in any way committed to a theocentric or religiously motivated outlook. After all, he is talking here about a certain well-defined range of issues in just those closely related areas – philosophy of mind, philosophical psychology, ethics, or metaphysics of the will from a broadly naturalised viewpoint – that need not (and nowadays for the most part do not) involve any overt or covert appeal to a validating ground beyond the compass of unaided human reason. Still there is there is a definite line of descent that runs from that old theocentric conception, through Kant's doctrine of the faculties and the complex, strictly ordered system of relationships between them, to Frankfurt's likewise strongly hierarchical conception of first- and second-order desires or volitions. Moreover this affinity is underlined by his idea that we most closely resemble non-human animals when we allow first-order desires to govern our conduct without the kinds of constraint and guidance afforded by reflective second-order volition, while conversely we attain to authentic personhood by engaging in just such higher-level processes of thought, self-criticism, and effective (i.e., practical even if difficult and sometimes temporally drawn out) will-formation. At any rate his use of 'wanton' – a censorious mode of description, even if deployed with a certain tongue-in-cheek archaizing tone – cannot but suggest that we take our intellectual-moral bearings from a scalar conception of the various orders of being with the human (unlike the animal or the divine) occupying a certain zone of the scale that is always potentially the site of a struggle waged between conflicting forms or forces of desire. Thus the two classes of 'wantons' and 'persons' are mutually exclusive if taken in the full, unqualified sense of each term since it is the having of second-order volitions, not just second-order desires, that qualifies a human being for personhood while the wanton is defined precisely as lacking any such capacity for moral character-shaping on the basis of rational self-evaluation combined with critical-reflective judgement and – crucially – a will to act upon any verdict thereby arrived at. What typifies a wanton is the fact that 'he does not care about his will', that '[h]is desires move him to do certain things, without its being true of him either that he wants to be moved by those desires or that he prefers to be moved by other desires' (pp. 86–7). What distinguishes persons, on the contrary, is that they make some desires truly and properly their own through a process of deliberative second-order thought that enables them to select and to will just those amongst the range presently competing for notice which qualify as worthy of adoption on all relevant, i.e., rationally and ethically acceptable grounds.

Here again Frankfurt reveals the Kantian lineage of his central concepts by insisting – as against any too rationalist or over-intellectualized account of these matters – that it is the exercise of will or capacity for such exercise that singles out some and not other human beings as eligible candidates for personhood rather than the exercise of rational thought if conceived in isolation from the active will. On his account 'a wanton may possess and employ rational faculties of a high order', since '[n]othing in the concept of a wanton implies that he cannot reason or that he cannot deliberate concerning how to do what he wants to do' (p. 87). What is missing from such deliberations when undertaken by the 'rational wanton', so Frankfurt maintains, is precisely that person-constitutive element of will. That is to say, it is the uniquely (though not universally) human capacity to take up a critical-reflective distance from our first-order desires and yet – where deemed appropriate – to adopt, endorse, or decisively confirm them as measuring up to our own best standards of moral or veridical warrant. So rationality, even 'of a high order', is not enough to justify the ascription of personhood in so far as '[w]hat distinguishes the rational wanton from other rational agents is that he is not concerned with the desirability of his desires themselves', and thus 'ignores the question of what his will is to be' (p. 87). Though Frankfurt doesn't say so I suppose one could conclude that the upshot of such conduct – if pushed to a pathological extreme – is the Marquis de Sade's distinctly crazed but weirdly 'rational' imagining of the multiform perversities to which certain appetites or cravings can be carried through a rigorous working-out of their various possible or logically conceivable permutations.[21] Thus the rational wanton is inexorably driven by first-order desires and by the drive to maximize their means of fulfilment through application of rational decision-procedures yet is totally incapable of conducting any enquiry as to whether they are justifiable, that is, whether they can stand up to reflective or critical-evaluative scrutiny from a higher (second-order) viewpoint. 'Not only does he [the rational wanton – here for once the gender-non-specific masculine pronoun seems altogether appropriate] pursue whatever course of action he is most strongly inclined to pursue, but he does not care which of his inclinations is the strongest' (p. 87). Yet it also needs stressing, on Frankfurt's account, that rationality is a necessary if not sufficient condition for personhood rather than mere human-being or membership of the species *homo sapiens*. After all, 'it is only in virtue of his rational capacities that a person is capable of becoming critically aware of his own will and of forming volitions of the second order', in which case '[t]he structure of a person's will presupposes . . . that he is a rational being' (p. 87).

[21] See for instance Barthes (1977); Lacan (1989) and Zizek (1998).

It seems to me that there are deep-laid conceptual tensions in Frankfurt's argument here and that these have to do with the underlying conflict between his basically Kantian conception of will as a matter of rational constraint upon the unregenerate promptings of desire and his semi-naturalised conception of human beings – even fully-fledged 'persons' – as inescapably subject to just such promptings, however qualified or held in check by the countervailing agency of second-order desires and/or volitions. For although this idea of a dualism between lower and higher levels of human being is one with distinctly Kantian antecedents it departs very markedly from Kant in adopting the term 'desire' to describe not only the former but also the latter, i.e., that level of second-order reflective, self-critical, and sometimes desire-thwarting evaluative thought wherein human will is most authentically manifest. For Kant, such a usage would certainly have betrayed an attachment to some grossly inadequate conception of ethical values, a naturalistic conception that reduced them to merely 'pathological' products of human will in its lowest, desire-driven or appetitive mode.[22] To this extent he can clearly be seen to inherit a Christian-influenced tradition of thought which stresses the unregenerate nature of our post-lapsarian state and imposes a pitiless divorce between moral law as enjoined upon the subject by the autonomous, self-legislative power of practical reason and that other barely human (indeed near-animal) realm of 'inclination' where instinct and desire hold sway. At its most extreme, in the writings of Saint Augustine, this tradition has given rise to some strange and at times fairly comic theological-ethical contortions, as in Augustine's notion that before the Fall male erections were fully subject to control by the conscious and deliberative rational will rather than (as now) seeming to possess a perverse will of their own. (For the record: Augustine compares that erstwhile happy state to the capacity of certain people with unusual musculature to wiggle their ears with extraordinary skill or break wind to impressively sustained and musical effect.[23]) That a kindred way of thinking has managed to exert so powerful a grip across such a range of otherwise diverse disciplines, schools and periods of thought is doubtless a result of its fitting so well with Plato's famous conception of the human soul as a charioteer pulled aloft by one horse toward the heaven of contemplative reason but dragged down by the other to the realm of ignoble sensuous appetite or instinctual desire.[24] What Kant gives us in his practical philosophy – conceived as the cornerstone to his overall system or 'architectonic' – is essentially a doctrine of the faculties

[22] See Note 8, above.
[23] St. Augustine (2001).
[24] Plato, *Phaedrus*, 246a–254e.

that seeks to make humanly intelligible sense of this starkly Manichaean vision.[25]

It is therefore not surprising, given all the well-known problems with Kant's remorselessly dualist view, that Frankfurt – like many philosophers nowadays – rejects any rigid application of the doctrine that opposes the dictates of rational (autonomous) will to the promptings of un-self-controlled (heteronomous) desire. One need not altogether go along with Jacques Lacan's scandal-provoking claim in 'Kant avec Sade' in order to see how very odd is the idea that a subject should be thought of as enjoying his or her greatest scope for autonomy of moral conscience or ethical choice at just the point where their inclinations are most strictly and remorselessly held in check, albeit at the bidding of a rational will that is somehow an expression of their own more principled or higher self.[26] Nor need one subscribe to an overly sanguine view of human nature in order to think that there is something wrong with Kant's Augustinian idea of human instinct, desire, and inclination as intrinsically corrupt or as always exerting a pernicious effect when allowed to interfere with the deliberative workings of 'autonomous' practical reason. Frankfurt is decidedly at one with the majority of present-day ethical naturalists in wishing to break with that hugely problematic legacy and abandon the idea of a sharp dichotomy between instinct, inclination, and desire on the one hand and reason, principle, volition, and will on the other. Hence, to repeat, his use of the same word 'desire' to describe both first- and second-order motivations. This suggests that the two orders are not so much locked in a struggle of absolute opposites or mutually exclusive drives and principles but engaged in a contest where they have at least enough in common for the issue and its outcome to possess real import at a humanly meaningful rather than a quasi-theological level. Moreover, Frankfurt's qualified naturalist sympathies show up in his account of how second-order desires relate to second-order volitions, that is, through an exercise of rationally motivated choice that determines just which of those maybe conflicting or as yet unclearly prioritized desires shall become identified with the subject's active will.

Still there are definite limits to Frankfurt's naturalism and also to his scope for throwing off that Platonic-Augustinian-Kantian legacy, given the marked persistence in his thought of other dualist themes which effectively replicate the same structure of assumptions. Chief among them, since the most crucially load-bearing, is his distinction between first- and

[25] See Notes 8 and 9, above; also Deleuze (1984).

[26] For a scholarly, perceptive and unusually angled approach to these themes, see Harpham (1987); also Norris, 'Kant disfigured: ethics, deconstruction, and the textual sublime' in Norris (1993: 182–256).

second-order desires, along with that between second-order desires and volitions. Thus volitions are taken – in distinctly Kantian mode – to fall within the category of 'desires' in so far as they belong to that range of motivating interests with which the subject identifies as a matter of active and personal commitment, but to stand outside (and potentially against) lower-order desires in so far as these latter embody the promptings of mere sensuous, instinctual, or creaturely gratification. This distinction is pressed home with maximum force in Frankfurt's comparison between two kinds of drug-addict, the one perpetually at war with himself over his constant struggle (and repeated failure) to break the habit which he knows to be destroying his life, while the other is subject to no such agonies of self-division and self-reproach since he has entirely given in to the habit. Of course this omits the case of the addict who has successfully managed to kick the habit, an omission that reflects Frankfurt's main interest in the workings of ambivalent, divided, or conflictual rather than straightforwardly efficacious moral will. The second kind of addict is a type-case of the 'wanton', one who 'does not prefer that one first-order desire rather than the other should constitute his will' (p. 88). Indeed, as Frankfurt puts it in an oddly ambiguous sentence, 'the wanton addict may be an animal, and thus incapable of being concerned about his will . . . . In any event he is, in respect of his wanton lack of concern, no different from an animal' (ibid).

This passage leaves one in doubt – whether through deliberate or accidental looseness of phrasing – as to just what is meant by the non-human animal reference. That is, one is hard put to decide between the reading 'this is what happens in the case of mere non-human animals', and the reading 'this kind of blank indifference with regard to the relative value or worthiness of various motives and desires is what reduces human beings to the level of mere sub-human animality'. At any rate it is clear that Frankfurt adopts something very like the Kantian view of desire as inherently a product or expression of whatever in our unregenerate natures inclines us to act against the dictates of practical reason, that is, the maxims of virtuous conduct laid down by our own autonomous or self-legislative better selves. Hence the contrast he draws with the first type of drug-addict, one who sincerely hates his habit and despises himself for his failure to kick it yet is unable to find the requisite determination or strength of will. More precisely: his self-loathing and his genuine wish to pursue the alternative course are just not strong enough to subdue or vanquish the overwhelming first-order compulsion that impels him to continue in the bad way. In short, '[t]hese desires are too powerful for him to withstand, and invariably, in the end, they conquer him . . . [h]e is an unwilling addict, helplessly violated by his own desires' (p. 87).

As I have said, this is what locates Frankfurt's essay squarely within that mainstream tradition of Western ethical thought which identifies morality with some kind of check upon our natural, instinctual, or (above all) our merely 'animal' inclinations. Most often – and most notably in Kant – it has taken the form of a self-denying, self-thwarting, self-abnegating drive with its likeliest source in the kinds of ascetic imperative enjoined with varying degrees of rigour by religious (especially Christian) doctrines of salvation through mortification of the flesh.[27] What is so striking about Frankfurt's albeit more moderate, less Manichaean rendition of this theme is the extent to which it replicates the Kantian structure of argument along with all the conflicts, aporias, antinomies, and other such symptoms of conceptual strain to which that argument gave rise. Moreover they are here expressed in terms of a veritable psychomachia, a drama acted out between contending desires, inclinations, impulses, motives, volitions, or exertions of will that often reads like an allegory with personified virtues and vices in the mode of *Pilgrim's Progress* or a medieval morality play. Such is the passage cited above where Frankfurt describes the unwilling addict as 'helplessly violated by his own desires', and also the following which I shall quote at length since it captures very well the way that this updated, semi-naturalized version of Kant tends to create its own dramatic plot-line with a suitably varied (if somewhat typecast) list of characters to move the action along.

> The unwilling addict identifies himself . . . through the formation of a second-order volition, with one rather than with the other of his conflicting first-order desires. He makes one of them more truly his own and, in so doing, he withdraws himself from the other. It is in virtue of this identification and withdrawal, accomplished through the formation of a second-order volition, that the unwilling addict may meaningfully make the analytically puzzling statements that the force moving him to take the drug is a force other than his own, and that it is not of his own free will but rather against his will that this force moves him to take it. (p. 88)

What is so curious about this passage is that it treats those statements not only as 'meaningful', i.e., as making good sense to the unwilling addict and to anyone who wishes to grasp his tragic predicament but also as possessing a strong claim to philosophical validity since, after all, they can be seen to articulate Frankfurt's own considered views on the matter. Thus we are here presented with a story of mental desires, compulsions, conflicts, dilemmas, identifications, withdrawals, alliances, forces and counter-forces that seem to be envisaged almost after the manner of those ancient theories of consciousness or mind that posited the existence

[27] See Notes 21 and 26, above.

of homunculi, that is, Russian-doll like miniature beings who or which served to 'explain' the otherwise mysterious workings of our various sensory, cognitive, and intellectual faculties.

Of course I am not suggesting that Frankfurt subscribes to anything like that doctrine in its primitive or mythical form. Rather I am pointing to the fact that he tells this story – the story of the unwilling addict – very much in the style of a novelist who identifies so closely with the thoughts, feelings and self-interpretations of a given fictional character as to take theirs as the privileged narrative viewpoint in relation to which all events and other characters will ultimately fall into place. This is why, rhetorically and thematically speaking, it belongs very much to that older tradition where the novel first emerges from the morality tale or play and where there is not, as yet, any room to exploit those added possibilities of complex understanding that come of a more detached, ironic, or sceptically-inclined relation between author and character.[28] No doubt there is a sense in which the unwilling addict is here being put in his place for having manifestly failed to achieve that degree of practical wisdom that would allow him to act upon his better judgement by making it his own as a matter of effective will, that is, by identifying closely enough with his habit-kicking as against his habit-reinforcing desires. All the same it is clear from the various descriptions of the ongoing conflict or turmoil of motives that make up the unwilling addict's mental life that his is the chief point of reference both for Frankfurt's understanding of personhood in general and for what we are to think of his counterpart, the wanton, in whom no such struggle takes place and who therefore – it is argued – falls below the threshold for admission to the class of persons as distinct from mere human beings. Thus, in contrast to the above-cited passage concerning the unwilling addict and his ventures into self-analysis, there is Frankfurt's description of the wanton who may well suffer the same kind of first-order conflict between desire to continue with the habit and desire to give up but who 'does not prefer that one of his conflicting desires should be paramount over the other' (p. 88). In such cases, '[i]t would be misleading to say that he is neutral as to the conflict between his desires, since this would suggest that he regards them as equally acceptable'. On the contrary: what is lacking in the wanton is precisely that capacity to stand back and to assess, judge or critically evaluate his own first-order desires that would be requisite even to a subject who found himself unable to plump for one or the other of some sharply conflicting pair since the reasons and motives for each seemed to him so evenly balanced. As concerns the wanton, however, '[s]ince he has no identity apart from his

[28] For a classic treatment of this episode in literary history, see Watt (1957).

first-order desires, it is true neither that he prefers one to the other nor that he prefers not to take sides' (p. 88).

IV

Hence perhaps Frankfurt's ambivalent remarks, as previously noted, with regard to what he thinks of as the hard-to-draw line between human and non-human (animal) wantons, since both strike him as manifesting the same incapacity for critical, reflective, second-order, revisionary, or self-evaluative thought. Just how hard-to-draw, on his account, comes out in a passage where the human wanton is described in terms that would seem to fit him more for placement in a zoo – or in one of those medieval bestiaries depicting human vices in animal form – than for the company of humankind. Thus:

> [t]he wanton addict cannot or does not care which of his conflicting first-order desires wins out. His lack of concern is not due to his inability to find a convincing basis for preference. It is due either to his lack of the capacity for reflection or to his mindless indifference to the enterprise of evaluating his own desires and motives. There is only one issue in the struggle to which his first-order conflict may lead: whether the one or the other of his conflicting desires is the stronger . . . . [I]t makes no difference *to him* whether his craving or his aversion gets the upper hand. He has no stake in the conflict between them and so, unlike the unwilling addict, he can neither win nor lose the struggle in which he is engaged. (p. 89)

This is a remarkable passage in several ways, among them its sheer determination to enforce the person/wanton dualism and its consequent refusal – somewhat against the logical grain of the argument here – to concede that such distinctions must at best be a matter of degree or of locating various points on a scale not only as between different individuals but also with respect to any given (not to say 'self-same') individual from one to another time or context. Where it raises logical problems is by presenting this as in some sense a struggle 'in which he [the wanton addict] is engaged' while at the same time working hard to persuade us that in truth there can be no such engagement in the wanton's case since he is altogether lacking in just those capacities which, if he possessed them, would earn his admission to the class of fully-fledged persons. In short, '[w]hen a *person* acts, the desire by which he is moved is either the will he wants or a will he wants to be without', whereas '[w]hen a *wanton* acts, it is neither' (p. 89). Again there is something odd – logically as well as metaphysically perplexing – about the idea of *wanting* or *not wanting* to have (i.e., to have one's desires, thoughts and actions directed by) some particular exercise of will, or again, about this notion of the will to engage in or desist from

this or that course of action as itself somehow subject to the prompting of our 'wants', preferences, or inclinations. For it is surely just the point of Frankfurt's cardinal distinction between first- and second-order desires that the latter should be thought of as strictly – indeed, by very definition – super-ordinate since this is what leads to their adoption in the form of definite commitments – or volitions – and hence their taking-up into the subject's will through an act of decisive identification with those that have proved most deserving of his or her allegiance.

So it is hard to make sense of his idea that the will must in turn be considered subject to 'wants', whether positive or negative, and therefore – it would seem – be thought to exert its authority only on condition of approval by the range of desires that predominate in some given subject at some given time in response to some given situation. Most likely my use of the word 'subject' in two different senses in the course of that last sentence will come across either as clumsy phrasing or as a kind of archly philosophical pun that strains the limits of semantic and conceptual decorum. However that sense of strain is endemic to all thinking (at least, all mainstream philosophical thinking) about the nature, scope and limits of human epistemic and ethical responsibility, that is, the extent to which human beings can properly or intelligibly be held accountable for arriving at certain beliefs or convictions and putting them into effect through precept or practice.[29] This is partly, as I have said, a question of the logical regress that opens up whenever one seeks – in the manner first established by Plato and carried to its high point of elaboration by Kant – to specify a pecking-order of the faculties. It is reinforced or brought home in epistemic and ethical terms by the way that such thinking makes it hard to avoid that problematical idea of the subject as at once the very locus of free-will, personhood, or moral autonomy and the locus of subjection to precepts or dictates of which it is somehow (paradoxically) both author and passive recipient.

There is no escape here – in Frankfurt's naturalized or de-transcendentalized version of the Kantian doctrine – from the problem as to how one can save any plausible or non-contradictory account of what is supposed to constitute the essential difference between humans (or persons) as autonomous beings and non-human animals (along with wantons) as falling entirely under the rule of heteronomous compulsions or desires. Beyond that, it is a matter of the problems faced by any attempt to achieve a workable *modus vivendi* between the defence of free-will, autonomy or Kantian practical reason conceived as distinctively human attributes and a

[29] See for instance Code (1987); DePaul and Zagzebski (2002); Fairweather and Zagzebski (2001); Zagzebski (1996).

moderately naturalized approach that would take on board the more salient aspects of present-day, scientifically informed thinking and yet preserve space for a sufficiently robust conception of those same attributes. On most such accounts this would be a more expansive or accommodating space than is allowed for by genial physicalists like Dennett in his book *Elbow-Room: The Varieties of Free Will Worth Having*.[30] Still it would be nowhere near as large as that claimed by upholders of the strong-autonomist line who maintain – on various philosophic grounds – that the physical sciences (especially neurophysiology) are in the grip of a massive category-mistake when they suppose that the problem of consciousness or the issue of mind/brain identity could ever be clarified, let alone resolved, by any possible advance in the scope or reach of scientific explanation.[31] What emerges most strikingly from the passages cited above, as from so many recent efforts in a moderating vein, is the fact that when thinkers strive to steer a path between contrary temptations or opposed sources of error they will often end up impaled on both horns of a dilemma that they have not so much resolved as temporarily managed to ignore.

I have suggested that these problems all bear witness to a deep-laid, culturally widespread and philosophically recurrent desire to fix a distance between the human and the non-human animal, or again (as Frankfurt would have it) between full personhood and whatever in our natures must be thought to fall short of that status through passive or unthinking enslavement to the rule of first-order motives and desires. His essay thus belongs to that multiform and venerable *genre* which Agamben has memorably traced in his book *The Open: Human and Animal*, namely the long series of attempts – in philosophy, theology, anthropology, ethnology, psychology, linguistics, and other fields – to carve out a unique (and uniquely distinguished) niche for *homo sapiens* despite, or because of, the growing sense that this distinction was under threat from developments spawned by some of those same disciplines.[32] What emerges from Agamben's survey is the way that thinkers were forced back upon ever more resourceful and inventive but also ever more elaborate, tenuous, wire-drawn or conceptually extravagant modes of argumentation so as to keep that niche open or hold that distinction in place. Not that Frankfurt's essay could fairly be described in any such terms, argued as it is in a clear-headed analytic style and with meticulous regard for the scope and limits of philosophic reasoning *vis-à-vis* other (e.g., psychological or natural-scientific) claims in this area. Still it is clear from the outset that this will be another contribution

[30] Dennett (1984).

[31] See for instance Foster (1991); also – from a different, more qualified dualist standpoint – Popper and Eccles (1998).

[32] See Note 18, above.

to the same legacy of thought, one that seeks to re-draw the human/non-human (or person/sub-personal human) line in a more naturalistic way, but which continues to insist that its drawing is prerequisite to any grasp of what it means to possess and exert that capacity for freely-willed autonomous choice that sets the former categorically apart from the latter in each case.

That is to say, like other recent ventures in this semi-naturalized and quasi-Kantian vein – among them John McDowell's *Mind and World* – the approach is very apt to hold out in one hand what it promptly takes back with the other.[33] Thus we seem to be vouchsafed renewed access to a range of possibilities for staving off the threat of hard-line naturalism and determinism, possibilities that had once (not so long ago) been thought of as beyond the pale since irredeemably tainted by association with Kantian transcendental idealism and its various (presumptively otiose) metaphysical commitments. In McDowell this has to do with refurbishing Kant's 'spontaneity'/'receptivity' dualism, taken to denote not a sharp and thus inherently trouble-making distinction like that between 'concept' and 'intuition' but rather a pair of mutually dependent or inseparably inter-involved capacities whose jointly active/passive synthesis is such as to resolve all the bad antinomies that have plagued philosophy from Kant on down. Chief among them are those of subject and object, free-will and determinism, and of course mind and world, all of which – so McDowell firmly believes – can be laid to rest through the benign ministration of a Kantianism suitably filtered and revised (with the customary help from Wittgenstein) so as to coax it down from the transcendental-metaphysical heights and restore it to a properly naturalized sphere of communal practices or life-forms while at the same time deploying its normative resources to protect against any too strongly reductionist strain of philosophic naturalism. In Frankfurt, it takes the form of a first-order/second-order distinction as applied to human wants, desires, or volitions which preserves something – a structural analogue at least – of Kant's categorical distinction between the strictly 'pathological' realm of subjectively driven (especially sensuous) inclinations and the sovereign dictates of autonomous moral law. Here again, as with McDowell's Kant, that dichotomy is treated to a moderately naturalizing gloss whereby to eliminate its hapless since delusive reliance on a realm of transcendentally grounded moral precepts yet also to retain something of its role as a means of suggestively marking – rather than rigidly enforcing – the two distinct orders of desire.

---

[33] McDowell (1994); also Norris, 'McDowell on Kant: redrawing the bounds of sense' and 'The Limits of Naturalism: further thoughts on McDowell's *Mind and World*', in Norris (2000: 172–96 and 197–230).

Thus if Frankfurt, again like McDowell, takes a less extreme view than Kant concerning the difference between persons and non-human animals then this also applies, as might be expected, to their respective views of the difference between those first-order 'lower' desires that pertain most directly to the nature of our instincts, drives, or sensuous appetites and those second-order 'higher' volitions that require the application of critical-evaluative thought and the exercise of moral will. That is to say, Frankfurt is sufficiently of his own cultural time and place to make some allowance for the various ways in which a science-dominated naturalistic worldview has raised obstacles to any such confident beating of the moral bounds, or at any rate made it far more difficult – philosophically or ethically speaking – to defend a full-scale anti-naturalist approach along strict Kantian lines. What has mostly taken the place of such efforts is just the kind of moderately naturalized account that finds expression in Frankfurt's essay, even if it appeared some decade before the wider movement in that direction to which McDowell's book bears witness. It is for this latter reason, I think, that Frankfurt writes with a vigour and a sense of breaking new philosophical ground that is very rarely apparent in more recent work on this and related themes. All the same it is worth looking more closely at some of the conceptual tensions or unresolved conflicts of aim and priority present in his thought. On the one hand these have their ultimate source in the Kantian antinomies of pure and practical reason, the largely unacknowledged impact of which on analytic philosophy in its various departments or sub-divisions is a story that I have told in some detail elsewhere.[34] On the other they also result from the encounter between that residual Kantian way of thinking and a naturalistic imperative which, even in its moderate forms, has put such ideas under growing philosophical strain. Indeed that strain has if anything been ratcheted up – rather than (as the received wisdom would have it) effectively dissolved or conjured away – through the sense that there must be some middle-ground position capable of easing these conceptual cramps by showing them up as nothing more than products of an old and nowadays discredited metaphysical worldview.[35]

The effect of such therapeutic endeavours is very often to conceal from their own practitioners the extent to which they are still – and now less wittingly – in the grip of those same compulsive dualisms that are thought to have been left safely behind with the transition to a new, metaphysically unencumbered and sufficiently (though not overly) naturalized or 'de-transcendentalized' approach. It is here, I would suggest, that

[34] See Notes 5 and 10, above.

[35] I put this case more fully in Norris (2002).

Frankfurt's essay has its special relevance and interest with regard to the way that these debates have gone since it appeared almost four decades ago. For there is much to be learned about philosophy's perceived situation *vis-à-vis* the natural sciences and its need to stake out some distinctive territorial claim in the fact that the essay has been so frequently cited and that his notion of second-order desires or volitions has become such a staple of attempts to vindicate the idea of free-will as a marker of human (more precisely: of personal) status and identity. Along with that goes the precept of an absolute or principled distinction between human persons and those other non- or sub-human animals whose lack of this second-order capacity for self-critical, reflective or evaluative thought about their own first-order desires is what marks them as belonging to a lower rank on the scale of sentient being. It seems to me that there are large problems with Frankfurt's attempt to make good this case, and moreover that those problems have a lot to do with his and his readers' wish to maintain that decisive margin of the truly, irreducibly, or properly human that would allow us to keep determinism at bay and hang on to our sense of moral privilege despite and against sundry present-day (mainly natural-scientific) threats and encroachments.

Indeed those problems come through with unignorable force in the last few paragraphs of his essay where Frankfurt takes stock of what he hopes to have achieved and enters a number of surprisingly large caveats in that regard. These caveats appear to be prompted by his desire to place the maximum possible distance between his own views on the free-will/determinism issue and Roderick Chisholm's concept of agent causation, namely the idea that 'human freedom entails an absence of causal determination', that 'whenever a person performs a free action . . . it's a miracle', and moreover that 'a free agent has "a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved"' (p. 93).[36] Frankfurt is confident in rejecting such notions as a mere recrudescence of religiously-inspired and philosophically as well as scientifically discredited ideas about human beings as uniquely placed in the divine order of things by virtue of their absolute exemption from the otherwise universal laws of deterministic cause and effect. Besides – as he very reasonably comments – 'why, in any case, should anyone *care* whether he can interrupt the natural order of causes in the way that Chisholm describes?' (ibid). Yet it is clear that Frankfurt does have concerns of this kind and, moreover, that his whole line of argument here is very largely driven by the desire, motive, volition, or will to carve out just such a space of freedom for the exercise of human autonomy as against

[36] See especially Chisholm (1977); also Chisholm (1976).

all manner of perceived encroachments from the natural-scientific quarter. My point in listing those various candidates for the role of carver-out is that they each have a definite place in Frankfurt's account of how human beings – or persons – manage to buck the scientific (determinist) trend and also that they each mark a certain point on the scale that runs from a naturalistic to a broadly Kantian or autonomist conception of human agency. What is most striking about these last few paragraphs is the extent to which Frankfurt backs away from any strong autonomist claim while none the less continuing to fend off the combined (at least as he perceives them) threats of naturalism and hard-line physicalist determinism.

Thus the paradoxes at this stage come thick and fast, to the point where it is remarkably hard to discern just what is meant by the various terms that make up Frankfurt's philosophical lexicon. 'Whatever his will', we are told, 'the will of the person whose will is free could have been otherwise; he could have done otherwise than to constitute his will as he did' (p. 94). But in that case clearly there must be some superordinate constituting agency – some 'want', to use his own expression – that is so placed within the overall structure or economy of motivating drives as to rank higher than will in terms of its directive or governing power. This interpretation seems to be supported by Frankfurt's going on to say that 'the assumption that a person is morally responsible for what he has done does not entail that the person was in a position to have whatever will he wanted' (ibid). So far as I can see the only way to make logical sense of this claim is to take it as involving the huge concession – huge, that is, for anyone who wishes to defend the principles of human free-will or autonomy – that we can only will what we 'want' to will, or again (in negative terms) that our will to do or to refrain from doing this or that is itself constrained and potentially subject to veto by the force of some other, more powerful desire that overcomes any reasons, precepts, or guiding principles that the will may muster on its own behalf. All the same, 'the assumption that a person is morally responsible for what he has done does not entail that the person was in a position to have whatever will he wanted' (ibid). Here again it is hard to know what Frankfurt means since this sentence seems to push the spiral to a higher stage where not only is the person's will subject to wants that may ultimately thwart or frustrate it but those wants are themselves subject in turn to various possible 'positions' – or scope-restrictive predicaments – in which the person concerned is unable (and moreover, by the logic of the case, unwilling) to act upon them.

Thus one is puzzled to grasp what can possibly distinguish the willing from the unwilling drug addict as Frankfurt describes their respective and, he thinks, decisively different cases. After all, they are both stymied – prevented from acting in their own best interests, everything considered – by

a crucial failure of will even if that failure, according to Frankfurt, pertains to different levels or stages of will-formation. On his account it occurs either (as concerns the wanton) through a sheer lack of second-order desires with sufficient strength or tenacity of purpose to become second-order volitions or else (as concerns the unwilling addict) through the fact that those second-order desires fall short of what is required to transform themselves into the kind of full-strength, self-defining, or properly personal volition that could constitute a subject's chief motivation for behaving in this or that way. Still it might be asked – from a more naturalistic or less residually Kantian standpoint – just why we should accept this *a priori* dualism between, on the one hand, conflicts of motive or allegiance that belong to the presumptively lower sphere of first-order desires and, on the other, conflicts of a more elevated sort which, even where they issue in a failure to adopt the preferable path, none the less bear witness to a higher capacity for suffering such deep-laid dilemmas. For it will then look more like a rearguard attempt to shore up the Kantian-autonomist defences against the threat of a consistent naturalism that would find no philosophically legitimate room – that is to say, no other than face-saving or self-image-protective reason – for any such resort to *a priori* notions of fully achieved as distinct from merely 'wanton' and to that extent quasi-animal states of human being.

Things get even murkier when Frankfurt remarks, as if by way of clarification, that '[t]he willing addict's will is not free, for his desire to take the drug will be effective regardless of whether or not he wants this desire to constitute his will' (p. 94). But if the issue of his wanting or not wanting to remain in thrall to the addiction is indeed a matter of indifference in this sort of case – if the fact of his carrying on with the habit is enough to place the two possibilities metaphysically, ethically, or psychologically on a par – then it seems to contradict Frankfurt's thesis concerning the crucial distinction between persons, including motivationally torn or self-divided persons, and that other class of beings (wantons, non-human animals, humans falling short of personhood) whose defective status he has explained precisely in terms of that difference which he here lets go with remarkable ease. '[W]hen he takes the drug', Frankfurt remarks, 'he takes it freely and of his own free will.' (p. 94) This claim is very odd – involving what seems a sizable affront to received philosophical ideas as well as to the dominant narco-physio-psychological conception of addictive behaviour – and looks even odder when Frankfurt goes on to say that he is inclined 'to understand the situation as involving the overdetermination of his first-order desire to take the drug', and that '[t]his desire is an effective desire because he is physiologically addicted'. By now it would an appear that the willing and the unwilling addicts are in truth – as

per the sentence quoted at the start of this paragraph – distinguished only by a nominal shift from one to another distribution of emphases between the three terms 'desire', 'want', and 'will'.

Moreover, paradoxically enough, Frankfurt can then proceed on the basis of just this stipulative re-jigging of semantic bounds to take what amounts to a strong voluntarist line on addiction – or on 'wantonness' of any kind in so far as it issues in bad, irresponsible, recidivist, or self-destructive behaviour – since, after all, his argument has now worked itself around to a point where the Kantian structure of assumptions has pretty much collapsed under pressure from the naturalizing drive that is also a prominent aspect of Frankfurt's thinking. This is why he eventually declares in favour of what seems in ethical as well as in social, political, and legal terms quite a hard-line doctrine of accountability that would – if carried into policy and practice – leave precious little room for pleas of reduced responsibility, mitigating circumstance, or social/cultural deprivation as grounds for special-case treatment. Thus, if the addict's desire is effective on account of his physiological addiction, nevertheless 'it is his effective desire also because he wants it to be' (p. 95). More specifically, '[h]is will is outside his control, but, by his second-order desire that his desire for the drug should be effective, he has made this will his own. Given that it is therefore not only because of his addiction that his desire for the drug is effective, he may be morally responsible for taking the drug' (ibid). In which case, quite simply, there is no getting the confirmed addict off the moral hook since in the end – as the upshot of all these semantic shifts – it is the addict *ipse*, whatever his degree of psycho-physiological enslavement, who must turn out to be 'willing' not only in the negative sense of passively enduring his addiction but also in the positive sense of actively wanting or desiring to continue with the habit. For we have now reached the stage in Frankfurt's dialectic of the faculties where there would seem as much reason to assert 'Human beings are responsible for everything they do' as to assert 'Human beings are responsible for none of their actions'. And we have reached that stage precisely on account of his essaying a normative conception of personhood that is distinctly Kantian in its basic approach to issues of autonomy, free-will, and motivation but which backs away from any full-scale acceptance of the metaphysical doctrines whereby Kant sought to uphold that conception. What results is a semi-naturalized account of the relations between desire, want, volition, and will that ends up by shying away from the determinist conclusion and thus embracing a full-strength voluntarist or autonomist outlook according to which, even if someone's will is 'outside his control', nevertheless 'by his second-order desire … he has made this will his own'.

V

This is, to say the least, an extraordinary conclusion and one that should I think give us pause in reflecting on the kinds of hidden liability to which philosophy is subject when it tries to bring off that particular trick of retaining those elements of Kantian thought that suit its argumentative purposes while emphatically rejecting the rest of Kant's transcendental-metaphysical system.[37] The consequence of this is a curious hybrid doctrine which produces what might very well be thought the worst of both worlds: a theory that combines the unyielding rigour of a deontological ethic (i.e., a doctrine of autonomous or self-legislative practical reason whereby the person is held absolutely responsible for each and every act) with a partially naturalised view of human desires, wants, and volitions which effectively excludes any possible room for the exercise of such autonomy by placing will in a subordinate role to the promptings of want and desire. Such is the upshot of Frankfurt's argument – borne out by the various passages cited above – even though it goes drastically against what he takes to be the main gist of his essay, that is, the vindication of a strong conception of human choice and responsibility in firm opposition to determinist or hard-line naturalistic approaches. I would guess that this conflict of aim with outcome is one chief reason for the odd statement, in his closing paragraph, that his understanding of free-will 'appears to be neutral with regard to the problem of determinism', despite the 'innocuous appearance of paradox' in the statement – one that Frankfurt seems happy to endorse – that 'it is determined, ineluctably and by forces beyond their control, that certain people have free wills and that others do not' (p. 95).

I think this is not so much a paradox, or, if so, not so much an 'innocuous' paradox but rather something more like a *reductio ad absurdum* of Frankfurt's whole line of argument, leading as it does to a flatly contradictory statement at least on his own expressly voluntarist principles. The same applies to his subsequent claim that '[t]here is no incoherence in the proposition that some agency other than a person's own is responsible (even *morally* responsible) for the fact that he enjoys or fails to enjoy freedom of will' (p. 95). Here again one has to say that there is a very basic and, given his own assertions elsewhere, a self-contradictory tension between this and Frankfurt's emphatic commitment to the idea of personhood as crucially consisting in the power of second-order volitions to monitor, check, subdue, deflect, modify or simply overrule the promptings of first-order desire. 'Perhaps', as he writes in a final twist to this sequence of paradoxical musings, 'it is also conceivable . . . for states of

[37] See Notes 15, 16 and 33, above.

affairs to come about in a way other than by chance or as the outcome of a sequence of natural causes' (ibid.). It is hard to conceive what this might mean – what the alternative possibility might be – if not some version of the Chisholm-style appeal to agent-causation that Frankfurt had roundly rejected just a couple of pages before. I can think of no text that more strikingly demonstrates the kinds of dilemma that philosophers are forced into when they attempt to bring off this kind of balancing-act. Thus a great many of them are still very much in denial – and casting around for some such half-way plausible compromise solution – when it comes to confronting the massive challenge to traditional (including received philosophic) modes of thought represented by advances in the natural-scientific domain.

Indeed it is no exaggeration to say that philosophy's self-image and sense of its right to exist as a self-respecting discipline of thought is closely bound up with something very like Frankfurt's conception of autonomous personhood as involving the exercise of second-order, reflective thought and judgment. Nor is this at all surprising, given what philosophers typically take to be their proper sphere of interest, if not – any longer – their uniquely privileged realm of special expertise. What his essay brings out with particular force is the extent to which, in this respect at least, mainstream philosophy has tended to articulate the sorts of attitude that many people take toward developments in neurophysiology and other branches of the natural sciences which threaten – or which might well be taken to threaten – the belief in human autonomy. That is to say, it gives carefully worked-out arguments for a view of the relation between desire, wants, volitions, and will (along with all their internal divisions) that in its own way strikingly replicates the sorts of ambivalence that many scientifically-informed people are liable to feel when trying to square their everyday conceptions of what is distinctive about human personhood with the kinds of thinking that increasingly appear to dominate the view from those scientific quarters. It seems to me a fair inference from the rate and direction of various advances in the relevant fields of research – neurophysiology, cognitive psychology, and related areas – that naturalism in its strong rather than in any qualified or hedged-about guise is the only outlook that makes sense in a way that is consistent with those advances and not prone to self-destruct on the kinds of paradox or flat contradiction that emerge in the course of Frankfurt's essay.

No doubt it is the case for various reasons – moral, legal, political, social, and cultural – that the ideas of autonomy and free-will are deeply built into our elective self-image as human beings, even if (as naturalists are apt to argue) this is chiefly on account of their conducing to a more

stable, less conflict-ridden mode of social co-existence and hence a better prospect for species flourishing by way of natural selection. Philosophy has long been a bastion of such thinking with its sundry suggestions as to where the crucial difference is supposed to reside, from Plato's 'other place' (*topos ouranos*) of ideal forms to Descartes' realm of 'clear and distinct ideas', Kant's domain of synthetic *a priori* intuitions and concepts as distinct from the deliverances of pure reason, and Husserl's notion of transcendental phenomenology as aimed toward a region of pure eidetic essences.[38] Thus, at least until the middle years of the last century, the majority of philosophers were firmly opposed to any naturalistic approach that would threaten the values of autonomy and selfhood conceived as intrinsically beyond the furthest reach of any physical or causal-explanatory account. Of course there were notable exceptions, beginning with the ancient Greek atomists and carrying on through a widely-spaced yet continuous and indeed – given the strong (sometimes lethal) kinds of disincentive ranged against them – impressively persistent line of materialist or radically anti-dualist thinkers.[39] Nowadays this situation has changed to the point where those who reject the mind/brain identity thesis are forced very much onto the back foot or may feel themselves compelled to endorse some version of the dualist argument on intuitive or purported *a priori* grounds. This typically involves either the appeal to qualia ('what it's like' to perceive colours, listen to an oboe, suffer toothache, and so forth) or else – in Searle's somewhat harder-headed but still residually dualist version of the case – the presumptive truth that minds *just are* self-evidently marked by a special quality or attribute (that of intentionality) that could not possibly be realized by means of any inorganic, e.g., silicon-based hardware or support system.[40] The first line of argument, like so much present-day philosophy, trades on a purely linguistic point about differing modes of talk or conceptualization and the problem – maybe the impossibility, though this is again a linguistic-conceptual matter – of translating the mental-experiential-phenomenological into the physical or neuro-biological without significant remainder. The second has the severe disadvantage, albeit often masked by the over-confidence that comes of *a priori* conviction, of simply taking for granted – as if it were self-evident – the absurdity involved in supposing the idea of 'artificial' (i.e., non-hu-

[38] On the waning of this aprioristic way of thinking about issues in epistemology and philosophy of mind, see Coffa (1991).

[39] See for instance Rosenthal (2000).

[40] For a representative range of views, see Carruthers (2003); Gray (2004); Levine (2002); McGinn (1999); Robinson (2004); Searle (1983), (1992), (2002); Shoemaker (2007); Smith and Jokic (2003).

manly-embodied) intelligence to be anything other than a category-mistake of the most blatant kind.[41]

Cartesian dualism finds a kind of muted, shame-faced or last-ditch expression in Colin McGinn's 'mysterian' thesis that even though minds may be brain-dependent or brain-identical in some ultimate sense it is a sense that we'll never be able to fathom because we are just not bright enough to figure it out.[42] One striking feature of this case is that it totally reverses the Kantian idea of a noumenal domain – that of the 'thing-in-itself' – that is capable of being conceived in the abstract through a speculative stretch of pure reason and must be so conceived if we are to render our knowledge and experience metaphysically intelligible but which cannot, on pain of creating insoluble dilemmas, be ascribed to the remit of knowledge whereby sensuous intuitions are 'brought under' concepts of understanding.[43] For McGinn, conversely, what lies beyond the range of human conceptual grasp is that ultimate material (i.e., neuro-physiological) nexus where goings-on in the brain are somehow – through a process which remains, to us, utterly and forever incomprehensible – transmuted into goings-on in the mind. It strikes me after some fairly extensive reading-around in the literature that a good proportion of it – that which strives to discover some *via media* between the claims of hard-line reductive physicalism and an appeal to the phenomenological self-evidence of 'what it's like' to undergo this or that kind of sensory or subjective experience – amounts to a just series of further variations on the theme first sounded by Frankfurt in his 1971 essay. For it does little more than reiterate the notion of phenomenological ascent from level to level of a consciousness presumed *a priori* capable of achieving such ascent (since defined precisely through and by that capacity) whilst none the less conceived as being in some sense identical with its physical hardware in order to head off any imputation of lingering dualism or subjectivism. Hence the frequent talk nowadays of 'emergent' or 'supervenient' properties, conceived as strictly *dependent upon* and even (in some rather mysterious sense) *identical with* their physical support-system or means of instantiation yet by no means simply *reducible to* it in the manner proposed by hard-line physicalists or central-state materialists.[44]

I think it is fair to say that 'consciousness', along with its other more specialized cognate terms such as 'intentionality' and 'qualia', has become an *explanandum* that all too often serves as its own *explanans*. In

[41] See especially entries for Searle, Note 40, above.
[42] See McGinn (1999).
[43] Kant (1998).
[44] See especially Kim (1993); also Kim (2002); Rowlands (1995).

other words it is a name for whatever is both supposed to require some account beyond reach of our present-best (maybe best-attainable) physicalist understanding and taken to constitute precisely the reason, self-evident to consciousness itself, why such understanding must fall so manifestly short. Moreover this involves just the kind of purely notional ascent – along with the attendant risk of vicious regress or mere circularity – that can likewise be seen in the earlier attempts of logical positivists and empiricists to construct a model of 'material' (or empirical) as distinct from higher-order 'formal' (or logical) languages.[45] That programme had two main objectives: to hold a firm line between evidence and theory, in conjunction with that between context of discovery and context of justification, and also – as I have said – to head off any problems of self-reference like those which Russell famously encountered and purported to resolve with his Theory of Types.[46] That the same sorts of problem tend to arise with efforts to conserve some unique, privileged, or distinctive role for human consciousness *vis-à-vis* what non-human animals are thought to enjoy in the way of mental life or 'internal states' is, I think, one sure indication that such efforts are misconceived.

This applies all the more when they are placed in the service of other, more ethically loaded claims for the qualitative difference between human and non-human animal modes of sentient experience and hence – so the argument often runs – the error of supposing that there exist any rules, codes, or obligations that should regulate 'our' treatment of 'them' in a properly responsible way. Here again such thinking finds a precedent in Kant's idea that the error in question is the sort that typically arises when people mistake the promptings of 'mere' sentiment or creaturely sympathy for the demands of strict moral duty or the maxims and imperatives of practical reason.[47] It is this Kantian conception that is invoked – whether explicitly or not – by those who would draw a categorical line between human and non-human modes of awareness or sentient existence. Thus when philosophers mount a case along these lines it typically involves some further dichotomy – as for instance between conscious and self-conscious states or self-conscious states and those that can or could find expression in articulate or propositional form – whereby to shore up that supposedly vital distinction.[48] Yet this begs the question not only as regards the privilege attached to language or speech by thinkers from Aristotle down and open to challenge on the grounds, once again, of inherent circularity but

[45] See Notes 3 and 4, above.
[46] See Russell (1908); also Russell (1959).
[47] See Note 8, above.
[48] See various entries under Note 40, above.

also as regards its frequent deployment as a means of reinforcing traditional attitudes concerning the intrinsic superiority of human over non-human creatures. Such attitudes – whether grounded in religious or secular versions of the exceptionalist doctrine – are always liable to work out as a pretext for humans to treat other animals pretty much as they see fit.

To this extent the main issue is still what it was for Descartes and his contemporaries, namely the question as to whether those other animals are like or decisively unlike 'us' in the most salient physical, perceptual, psychological, social, and (at least arguably) ethical ways. Descartes' answer – that non-human animals were cunningly contrived machines and that merely to raise such a question was both absurd and impious – is one that would find few takers nowadays, at least in anything like so extreme or unqualified a form. On the other hand there are still adherents to a basically Cartesian outlook who would seek to maintain the crucial distinction, albeit (most often) with obligatory gestures toward a scientifically respectable, quasi-naturalistic approach that would draw the line firmly at any avowal of Cartesian substance-dualism. Indeed current versions of property-dualism – for a long time the favoured fallback position in debates of this sort – are more than likely to come hedged around with all manner of caveats to the general effect that any dualist (or mentalist) talk therein to be found is quite compatible with some version of the mind-brain identity thesis, perhaps *via* an appeal to Davidsonian 'anomalous monism'.[49] All the same it is not hard to make out the lineaments of that same old Cartesian position in many recent, philosophically *au courant* efforts to carve out some well-defined niche for what's thought distinctively or uniquely human about the human animal. Thus, for instance, the traditional view looms large when Peter Carruthers sets out to explain how the really important difference is that between the kind of self-awareness that non-human animals (or some of them) perhaps have and the kind of expressible, i.e., articulate, conceptually mediated and therefore speech-apt second-order consciousness enjoyed by human beings and – so he argues – by human beings alone.[50]

It is clear enough what response this would draw from thinkers, such as Peter Singer, whose chief concern is with the effect of such anthropocentric or 'speciesist' attitudes on the wider culture wherein they translate into various practices that are found unacceptable – or downright abhorrent – by those of a contrary persuasion.[51] For them, the relevant point is still best made by Jeremy Bentham's vigorous rejoinder that 'the question

[49] See especially Davidson (1980).
[50] See Carruthers (1992); also Carruthers (2004) and (2005).
[51] See Note 17, above.

is not, Can they reason?' nor, Can they talk? but, Can they suffer?'.[52] After all, it is scarcely deniable even by unabashed upholders of the exceptionalist thesis in its strong form that any appeal to rationality, language, or other such presumptive indices of human uniqueness must be made from (what else?) a human viewpoint which is sure to embody just that range of values and priorities. Moreover, it is here – in the area presently criss-crossed by numerous debates in epistemology, cognitive psychology, linguistics, anthropology, ethnology, and the social sciences – that it becomes most difficult (maybe impossible) to disentangle issues concerning the scope and limits of human *vis-à-vis* non-human animal modes of existence from issues concerning the ethical dimension of 'our' dealings with 'them'. Thus Bentham's point is subject to dispute not only by those, like Carruthers, who take the attributes of language and reason to constitute a definite and privileged mark of the human but also – less explicitly – by those who want to stake out a middle-ground, quasi-naturalist position that would still leave room for some scaled-down version of the same anthropocentric view. It is this approach that Frankfurt's essay exemplifies to most striking effect, along with all the symptomatic problems and stress-points – among them the constant proneness to various forms of circularity and vicious regress – that tend to characterize such projects.

Indeed one might go farther and assert that the problems in question are sure to arise when philosophers allow their naturalist commitments to be more or less qualified or held in check by the conjoint effect of two near-related but distinct motivations. One is the doubtless very deep-laid desire amongst most human beings to find something specific or unique to themselves that sets them decisively apart from other animals. The second is that other, more specialized or intra-disciplinary incentive which leads them (philosophers) to situate precisely this question – that of consciousness *qua* supposed distinguishing feature of humanity – at the centre of their various epistemological, ethical, linguistic, and (in so far as they are nowadays willing to endorse the description) metaphysical concerns. As I have said, this preoccupation extends well beyond the company of those who would defend a strong, even neo-Cartesian doctrine concerning the absolutely privileged status of human conscious, self-conscious, reflective, or ethically autonomous being. It is just as central to the arguments of many who would have no truck with that full-fledged dualist way of thinking and who would see it as the product of an old, scientifically under-informed conception of mind. For them – the majority of present-day philosophers – such ideas are simply unsustainable when confronted with the range of scientific (i.e., neurophysiological and cognitive-psycho-

[52] Bentham (1970: Chpt. 17, para 4, note).

logical) evidence currently on offer. Such evidence is often assumed to create large problems for the neo-Cartesian appeal to a realm of mental, intentional, or phenomenological experience conceived as intrinsically exceeding the limits of any brain-based (no matter how fine-grained) causal-explanatory account. Even so the very persistence of these debates and the fact that philosophers feel constantly obliged to rehearse them – to beat the bounds of physicalism from both sides – is a sure sign that they have not let go of that nagging preoccupation.

VI

What unites these thinkers across some otherwise large divergences of view is the underlying notion that they are able to discuss this topic in a meaningful way only on condition that, as conscious (or self-conscious) creatures, they must possess the kind of understanding that cannot be fully accounted for in downright physicalist or central-state-materialist terms. It would not, I think, be difficult to show that this basically anti-naturalist premise – or residual Cartesian assumption – is surreptitiously at work in a good few of those seemingly hard-headed approaches to the mind-brain issue which lay claim to a physicalism-compatible view while none the less endorsing the received, philosophically sanctioned idea that there is more to consciousness than could ever be cashed out in such reductive terms.

It seems to me that this gets the emphasis precisely back-to-front and that the best way forward is that marked out by those previous major advances in the natural sciences, from Galileo down, that have come about mainly through the willingness to break with anthropocentric or intuitively self-evident habits of thought and to go with the best scientific theories or hypotheses to hand. No doubt there remains a puzzle – one much touted by anti-naturalists and upholders of the various (albeit heavily qualified) dualisms that are nowadays doing the rounds – as to how that outlook can possibly accommodate the sheer variety of human subjective, perceptual, ideational, phenomenal, or other such inherently first-person-indexed, i.e., irreducibly experiential modes of knowledge and awareness.[53] However that puzzle is better thought of by analogy with our still perceiving, again as a matter of intuitive self-evidence, that the sun rotates around the earth – rises in the morning and sinks below the horizon every evening – or with our continuing to 'see' certain well-known instances of visual illusion under their illusory rather than what we know to be their true, dimensionally accurate or (in some cases) geometrically demonstrable properties, ratios,

[53] See Notes 31, 40, 44 and 50, above.

or shapes. Once thought of in this way the 'problem of consciousness' begins to look more like the pseudo-problem of explaining how phenomenal qualities like heat reduce to physical quantities like mean kinetic energy of molecules, or again, how our intuitive feel for 'analogue' (continuously varying) properties like smoothness, softness, roundness, gradualness, and so forth can possibly be registered or represented by discrete or 'digital' patterns of neuronal firing.

Of course if one concludes that these questions are wrongly framed or involve some kind of basic category-mistake – and are hence incapable of finding any other than a vacuous or misconceived answer – then one risks being lumped together with McGinn and the 'new mysterians' as believing that the problem is inherently too deep or complex for us to get our heads around it. In fact this is just the opposite kind of position since it holds that the problem is mainly a result of our clinging to commonsense-intuitive ideas of what it means to be somehow at the focal point of those various sensory, perceptual, epistemic, subjective, affective, or phenomenological states. Much better, I suggest, to take a lesson from various chapters in the history of science which tell how a good many such states – initially those at the sensory-perceptual end of the scale but increasingly those of a high-level cognitive or epistemic character – have been subject to ever more detailed kinds of scientific, i.e., physically specified descriptive and explanatory treatment. Thus the most productive, least mystery-mongering approach is one that views (say) the qualia issue by analogy with cases like the explanation of perceived temperature in terms of mean kinetic energy of molecules plus an incomplete though rapidly expanding knowledge of the sensory and neurophysical processes involved. At least it holds out more promise of advance – of progressive reduction in the range of phenomena considered beyond reach of any such naturalistic account – than the appeal to some quasi-dualist notion of phenomenal experience as either qualitatively *sui generis* or else, in the current mysterian mode, as quite conceivably having some ultimate physical explanation but one so complex as to far surpass our limited powers of scientific or indeed philosophical grasp. In fact this case very often amounts to a point-for-point reversal of the science-first argument which in stead gives pride of place to the standard range of philosophic puzzles about consciousness, qualia, subjectivity, or 'what it's like' to undergo various sorts of phenomenal experience, and which takes their philosophically recalcitrant nature as evidence enough that they are humanly – hence scientifically – forever insoluble.

It strikes me that, despite the intensive focus on these issues during the nearly four decades since Frankfurt's essay was published, the philosophical discussion has moved on rather little while the science has achieved

some very striking advances in the scope and depth of its conceptual as well as its descriptive and explanatory grasp. No doubt there will continue to be many philosophers who reject *a priori* the very idea that advances of that kind could ever resolve such intrinsically hard and distinctively philosophic issues. After all, there are those who reject any notion that the progress of scientific knowledge with regard to microphysical structures, properties, or dispositions should properly be taken to require some adjustment to our sense of whether Locke was justified in denying the possibility of ever advancing from 'nominal' to 'real' essences or definitions, or again, whether Hume was justified in his scepticism concerning the validity of causal explanations. If one can reasonably take the view that things don't move on philosophically in quite the same way that they move on scientifically – and therefore that it not completely off-the-point or just a sign of disciplinary obsolescence when philosophers raise problems from Locke or Hume – still this affords no justification for the idea that an impasse in philosophic thought should be taken to block or to close off the prospect of future scientific advance. This is just another case of speculative reason's proneness to take what Alexander Pope called 'the high priori road' and discover putative grounds for asserting or denying what can only be determined by other, less perfectly self-assured but altogether more reliable investigative means. It is the kind of error that philosophy is especially (perhaps constitutively) prone to by way of reasserting its status, autonomy, or continuing claim to serious attention despite the extent to which other branches of enquiry – pre-eminently the physical sciences – have seemed to make ever greater inroads on its once all-encompassing area of competence or special expertise. All the same it is a notion that philosophy must learn to live without if is not to become either the last redoubt of beliefs that are all the more entrenched and dogmatic for their superannuated character or else, as debunkers like Rorty would have it, at best just another entertaining but otherwise pretty much irrelevant voice in the ongoing cultural conversation.

References

Agamben, G. 2004. *The Open: Man and Animal*, trans. K. Attell (Stanford, CA: Stanford University Press).

St. Augustine. 2001. *De Bono Coniugali, De Sancta Virginitate*, trans. P.G. Walsh (Oxford: Oxford University Press).

Ayer, A. J. (ed.). 1959. *Logical Positivism* (New York: Free Press).

Barthes, R. 1977. *Sade, Fourier, Loyola*, trans. Richard Miller (London: Cape).

Bentham, J. 1970. *Introduction to the Principles of Morals and Legislation*, eds. J. H. Burns and H. L. A. Hart (London: Athlone Press).

Carnap, R. 1969. *The Logical Structure of the World, and Pseudoproblems in Philosophy*, trans. R. George (Berkeley & Los Angeles: University of California Press).

Carruthers, P. 1992. *The Animal Issue: Moral Theory in Practice* (Cambridge: Cambridge University Press).

——. 2003. *Phenomenal Consciousness: A Naturalistic Theory* (Cambridge: Cambridge University Press).

——. 2004. *The Nature of the Mind: An Introduction* (London: Routledge).

——. 2005. *Consciousness: Essays from a Higher-Order Perspective* (Oxford: Oxford University Press).

Chisholm, R. M. 1976. *Person and Object* (La Salle, IL: Open Court).

——. 1977. 'The agent as cause', in M. Brand and D. Walton (eds.), *Action Theory* (Dordrecht: D. Reidel), 199–211.

Clifford, W. K. 1999. 'The ethics of belief', in *The Ethics of Belief and Other Essays* (New York: Prometheus Books).

Code, L. 1987. *Epistemic Responsibility* (Hanover, NH: University Press of New England).

Coffa, J. A. 1991. *The Semantic Tradition from Kant to Carnap: To the Vienna Station* (Cambridge: Cambridge University Press).

Davidson, D. 1980. *Essays on Actions and Events* (Oxford: Clarendon Press).

Deleuze, G. 1984. *Kant's Critical Philosophy: The Doctrine of the Faculties*, trans. H. Tomlinson and B. Habberjam (London: Athlone Press).

Dennett, D. C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting* (Oxford: Oxford University Press).

DePaul, M. and L. Zagzebski (eds.). 2002. *Intellectual Virtue: Perspectives from Ethics and Epistemology* (Oxford: Oxford University Press).

Empson, W. 1951. *The Structure of Complex Words* (London: Chatto & Windus).

Fairweather, A. and L. Zagzebski (eds.). 2001. *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility* (Oxford: Oxford University Press).

Foster, J. 1991. *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind* (London: Routledge).

Frankfurt, H. 1988. *The Importance of What We Care About: Philosophical Essays* (Cambridge: Cambridge University Press).

——. 2003. 'Freedom of the will and the concept of a person', in Watson (2003).

Gray, J. A. 2004. *Consciousness: Creeping up on the Hard Problem* (Oxford: Oxford University Press).

Harpham, G. G. 1987. *The Ascetic Imperative in Culture and Criticism* (Chicago: University of Chicago Press).

James, W. 1907. *Pragmatism: A New Name for Some Old Ways of Thinking* (New York: Longmans).

——. 1909. *The Meaning of Truth* (New York: Longmans).

Kane, R. 2004. *The Significance of Free Will* (Oxford: Oxford University Press).

Kant, I. 1976. *Critique of Practical Reason, and Other Writings in Moral Philosophy*, trans. and ed. L.W. Beck (New York: Garland).

Kant, I. 1998. *Critique of Pure Reason*, trans. and ed. P. Guyer and A.W. Wood (Cambridge: Cambridge University Press).

Kim, J. 1993. *Supervenience and Mind: Selected Philosophical Essays* (Cambridge: Cambridge University Press).

——. (ed.). 2002. *Supervenience* (Dartmouth: Ashgate).

Lacan, J. 1989. 'Kant with Sade', trans. James Swenson, *October*, No. 15 (Winter 1989), 55–104.

Levine, J. 2002. *Purple Haze: The Puzzle of Consciousness* (Oxford: Oxford University Press).

Lovejoy, A. O. 1936. *The Great Chain of Being: A Study in the History of an Idea* (Cambridge, MA: Harvard University Press).

McDowell, J. 1994. *Mind and World* (Cambridge, MA: Harvard University Press).

McGinn, C. 1999. *The Mysterious Flame: Conscious Minds in a Material World* (New York: Basic Books).

Norris, Ch. 1993. *The Truth About Postmodernism* (Oxford: Blackwell).

——. 2000. *Minding the Gap: Epistemology and Philosophy of Science in the Two Traditions* (Amherst, MA: University of Massachusetts Press).

——. 2002. *Truth Matters: Realism, Anti-Realism and Response-Dependence* (Edinburgh: Edinburgh University Press).

——. 2004. *Philosophy of Language and the Challenge to Scientific Realism* (London: Routledge).

—— 2005. 'Choice and belief', *The Philosophers' Magazine*, Issue 29 (2005), 17–23.

——. 2006. *On Truth and Meaning: Language, Logic and the Grounds of Belief* (London: Continuum).

O'Connor, T. 2004. *Persons and Causes: The Metaphysics of Free Will* (Oxford: Oxford University Press).

Plato. 2002. *Phaedrus*, trans. R. Waterfield (Oxford: Oxford University Press).

Popper, K. R. and J. C. Eccles. 1998. *The Self and its Brain* (London: Routledge).

Quine, W. V. 1961. 'Two dogmas of empiricism', in *From a Logical Point of View*, 2nd ed. (Cambridge, MA: Harvard University Press), 20–46.

Regan, T. and P. Singer (eds.). 1976. *Animal Rights and Human Obligations* (Englewood Cliffs: Prentice-Hall).

Robinson, W. S. 2004. *Understanding Phenomenal Consciousness* (Cambridge: Cambridge University Press).

Rosenthal, D. M. (ed.). 2000. *Materialism and the Mind-Body Problem* (Indianapolis: Hackett).

Rowlands, M. 1995. *Supervenience and Materialism* (Aldershot: Avebury).

Russell, B. 1908. 'Mathematical logic as based on the theory of types', *American Journal of Mathematics* 30, 222–62.

——. 1930. *Introduction to Mathematical Philosophy* (London: Allen & Unwin).

——. 1959. *My Philosophical Development* (London: Routledge & Kegan Paul).

——. 1994. *Foundations of Logic, 1903–1905*, ed. A. Urquhart (London: Routledge).

——. 1999. 'William James's conception of truth', in S. Blackburn and K. Simmons (eds.), *Truth* (Oxford: Oxford University Press), 69–82.

Searle, J. R. 1983. *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press).

——. 1992. *The Rediscovery of the Mind* (Cambridge, MA: MIT Press).

——. 2002. *Consciousness and Language* (Cambridge: Cambridge University Press).

Shoemaker, S. 2007. *Physical Realization* (Oxford: Oxford University Press).

Simpson, D. (ed.). 1988. *The Origins of Modern Critical Thought: German Aesthetic and Literary Criticism from Lessing to Hegel* (Cambridge: Cambridge University Press).

Singer, P. 1990. *Animal Liberation: A New Ethics for Our Treatment of Animals*, 2nd ed. (London: Cape).

—— (ed.). 1985. *In Defence of Animals* (Oxford: Blackwell).

Smith, Q. and A. Jokic (eds.). 2003. *Consciousness: New Philosophical Perspectives* (Oxford: Oxford University Press).

Strawson, G. 1986. *Freedom and Belief* (Oxford: Oxford University Press).

Strawson, P. F. 1959. *Individuals: An Essay in Descriptive Metaphysics* (London: Methuen).

——. 1966. *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason* (London: Methuen).

Tarski, A. 1956. ‘The concept of truth in formalised languages’, in *Logic, Semantics and Metamathematics*, trans. J.H. Woodger (Oxford: Oxford University Press), 152–278.

van Inwagen, P. 1983. *An Essay on Free Will* (Oxford: Clarendon Press).

Watson, G. (ed.) 2003. *Free Will* (Oxford: Oxford University Press).

Watt, I. 1957. *The Rise of the Novel: Studies in Defoe, Richardson and Fielding* (London: Chatto & Windus).

Zagzebski, L. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge: Cambridge University Press).

Zizek, S. ‘Kant with (or against) Sade?’, *New Formations*, No. 35 (1998), 93–107.