



Hume's Theory of Motivation-Part 2

Daniel Shaw

Hume Studies Volume XVIII, Number 1 (April, 1992) 19-40.

Your use of the HUME STUDIES archive indicates your acceptance of HUME STUDIES' Terms and Conditions of Use, available at

<http://www.humesociety.org/hs/about/terms.html>.

HUME STUDIES' Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the HUME STUDIES archive only for your personal, non-commercial use. Each copy of any part of a HUME STUDIES transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

For more information on HUME STUDIES contact humestudies-info@humesociety.org

<http://www.humesociety.org/hs/>

Hume's Theory of Motivation — Part 2

Daniel Shaw

Introduction and Summary of Part 1

In an earlier paper of the same title¹ I defended a Humean theory of motivation against rationalist views of B. Stroud and T. Nagel.² In this paper I should like to relate my theory to more recent writings, explain its implications for the topic of moral motivation and provide further support for the main argument of my original paper.

To begin with, a summary of my previous discussion: while two critics of Hume's theory, B. Stroud and T. Nagel, disagree about the role of reason in action (Stroud accepts, Nagel rejects the Humean claim that reason alone can never produce action) they both agree in rejecting Hume's view that motivating desire essentially involves feelings which are identifiable by the agent, in his conscious experience, by introspection, as something extra, something over and above any purely rational considerations he may have for taking action.

As an alternative to this Humean, experiential view of motivating desire, Stroud argues that the concept of motivating desire should be analysed not in terms of any intrinsic property knowable in experience, but rather in terms of its function in leading, along with belief, to action. In Stroud's words: "It might well be that to have a desire for or a propensity towards E is simply to be in a state such that when you come to believe that a certain action will lead to E you are moved to perform that action."³

In a similar spirit⁴ Nagel claims, contrary to Hume, that there do exist at least *some cases* (for example, actions motivated by a person's prudential consideration of his own future interests) in which *reason alone*—that is, the agent's grasp of certain purely rational considerations—is *all* that motivates action. In the context of these purely rationally motivated actions, Nagel argues, our talk of desire does not refer to any *additional* inclination towards, or preferences for, or sentiment about some goal that we have, but is either just another way of saying that the act is motivated, that is, that we are in *some* state which disposes us to do whatever we think will lead to a certain result, or else indicates some structural feature of the reasoning behind our behaviour; for example, indicates the fact that in view of purely rational considerations which confront us, it would be irrational of us not to perform the act in question.

Against any such account of motivating desire which denies the necessity of the experiential component (that is, of Hume's introspectible sentiment), I argued that all such accounts have a false implication. They all falsely imply that some purely rational consideration, a bare belief, could, all on its own (unaccompanied by any actual or potential pro- or con-attitude), motivate a purposive action. With the help of a counter-example I argued⁵ that although it is *logically possible* for bare beliefs of that kind to produce overt behaviour, such pure rationalist 'action' never in fact happens; and, more importantly, if we carefully consider what such 'action' would be like if it did happen, we find it would not count as voluntary, purposive action, in the fullest sense.

So Hume, I concluded, was right to maintain that motivating desire essentially involves introspectible feelings.

However, the standard interpretation of Hume's view about the relation between desires and feelings yields an indefensible theory of motivated action. On the standard reading, desires simply *are* feelings, and the Humean is left with no adequate reply to the obvious objection that, often enough, for example, when acting after calm deliberation, or when performing routine habit-actions, we are not aware of the occurrence of any such feeling immediately prior to or at the time of action. Hume tells us that in all such cases there is—in addition to the operation of reason—a *calm* passion present—a passion so calm as to be readily confused with and mistaken for the determinations of reason. But on the standard reading of 'calm passions' as very faint or imperceptible feelings, this is a lame argument. As Stroud points out,⁶ not only is there no independent justification for the claim that in all such cases there always are such faint or imperceptible feelings,⁷ but also that claim is inconsistent with Hume's incorrigibility thesis—the principle that we cannot be wrong about the contents of our own minds at a given moment.

A reasonable Humean solution to all these difficulties, I argued, is not to give up Hume's essential connection between motivating desires and feelings, but rather to interpret it in a different way. If, instead of simply *identifying* desires with feelings, we understand desires to be *dispositions to experience feelings* (that is, dispositions to experience introspectible pro- or con-attitudes) in connection with thoughts of the object of desire, then 'calm passions' can be interpreted not as faint or imperceptible feelings but as unactualised feeling-dispositions. On this account,⁸ "Person *P*, desires at time *t*, some state of affairs *S*" means roughly, "If, at *t*, *P* were to think about or seriously contemplate the realization of *S* as opposed to the realization of not *S*, he would then experience an introspectible pro-attitude towards the former prospect and/or an introspectible aversion to the latter prospect (that is, he

would then be aware of favouring the former prospect and/or disfavours the latter).⁹ This dispositional fact about the agent, though in certain cases not a fact about a psychological *occurrence*, is nevertheless an *additional* fact about him and his introspectible experience; and it is something over and above any facts concerning purely rational considerations.

On this reading, one which, I argued, receives support from several key passages in the *Treatise* and the *Enquiry*,¹⁰ Hume's overall theory of motivation regains plausibility without abandoning Hume's essential connection between desires and feelings. Stroud's two objections to the calm-passions argument can now be met; for (firstly), on this view, independent evidence of the existence of a calm passion is always available to the agent:¹¹ if he contemplates the object of his desire, he will become directly aware, by introspection, of his pro- or con-attitude towards it;¹² and (secondly) this account of desire and of the calm passions is consistent with a plausible version of the incorrigibility thesis, the version which claims that, if we introspect carefully upon our conscious desires at any moment, we cannot be mistaken about them, or at least, that we ourselves are in a better position to know what we consciously desire at any given moment than anyone else is.

Moral Motivation

My discussion of Stroud's and Nagel's rationalist views made no explicit reference to the topic of moral motivation. That is the main subject of the present paper.¹³ I wish to argue that a distinct¹⁴ desire factor is just as necessary for moral motivation as it is for non-moral motivation. My arguments for this view about the moral case are the same in form as those for the prudential and other non-moral cases.

The arguments are these: the rationalist or cognitivist about moral motivation is committed to the view that purely rational considerations (a purely cognitive state; for example, the state of believing that one ought to do so-and-so) are capable of motivating moral action, without the presence or operativeness of a separately identifiable desire-factor, identifiable by the agent in his conscious experience by introspection as a passion or sentiment, or at least as something which essentially involves introspectible feelings.¹⁵ The way to test this cognitivist claim is to see whether one can construct a conceivable case of morally motivated action in which it is clear that one has stripped away all such non-cognitive motivational elements,¹⁶ both actual and dispositional, and then to see whether what one is left with is intelligible as a case of voluntary, morally motivated action in the fullest sense.

The results of this stripping-down exercise are, I believe, the same as in the case of prudential motivation (that is, my earlier counter-example to Stroud's and Nagel's views).

Let's apply this test to a couple of promising-looking cases for the rationalist/cognitivist about motivation. I choose a couple of cases suggested to me by my colleague, Mr. Ian Fowlie.¹⁷

Case 1: Amanda finds Jasper very attractive (felt desire) and she wants to go out with him, (felt desire) but she knows (purely cognitive state) that Jasper is a member of the National Front. Were she not so infatuated with Jasper, Amanda would surely have experienced feelings of guilt and a feeling of obligation to resist; but as it is, her amorous feelings have wholly extinguished her conscientious ones, and all she *feels* is the urge to give in. Nevertheless, she *believes* that she ought to give Jasper the brush-off (according to the cognitivist, a purely cognitive state) and so, does so (morally motivated action). The cognitivist/rationalist will invite us to view this as a classic case of the combat of reason and passion. Amanda's desires, he will argue, all line up on the side of yielding to this temptation. Reason, however, gives rise in her to the dispassionate belief (according to the cognitivist, a purely cognitive state involving no feelings) that she ought to avoid the fellow, and so she does.¹⁸ That's the way the conflict presents itself (phenomenologically) to Amanda, and that, the cognitivist will argue, is the way it is.

Case 2: I don't want to go, in the worst way, to this afternoon's department meeting. All my desires, the rationalist will argue, line up on the side of not going; but I believe I ought to go (according to the cognitivist, a purely cognitive state) and, for that reason, go. That's the way it seems to me, and that, the cognitivist will argue, is the way it is.

As a student of mine, Christopher Patton, succinctly put it, this rationalist argument breaks down as soon as one tries to unpack the ought-belief. Is it really quite clear that Amanda's *belief* that she ought not have an affair with a National-Fronter has no 'passionate' non-cognitive element? Has she no sentiment of conscientious aversion (feelings of revulsion, repugnance, disapproval, etc.) to the idea of intimately associating herself with bigots, racists and persecutors of innocent minority groups? Has she no feelings whatever about this prospect, either actual or potential?¹⁹ In her feelings, does she care as little about whether she is associated with people of the above

description as she does about whether she ties her left shoe before her right shoe? Why then, from her point of view—which is clearly not the moral point of view—she would do well to find herself an (amoral) psychiatrist and say to him as follows: “I can’t help myself doctor, I don’t *want* to give Jasper the brush-off, nor will my doing so contribute in any way to anything that I do want or help avoid anything that I want to avoid. And I haven’t the least feeling of aversion to the National Front or to any of its policies or activities. I *care* (non-cognitive state) as little about his membership in that organisation as I do about whether or not he belongs to the Automobile Association. It’s just that each time he makes advances to me this obsessional ought-thought turns me into a block of ice. It’s wrecking my love-life! Is there any cure for it?”

Unpack the ought-thought in such a way as to make it as clear as this that it contains no non-cognitive element, no introspectible trace of conscientious aversion or propensity, and reduce the ‘morally motivated action’ to the status of a mechanical compulsion: to something less than voluntary action in the fullest sense. Whether or not morality is strictly speaking judged of, it is most certainly felt.

I shall not discuss the case of the department meeting further than to say that if I were as totally emotionally indifferent to such matters as the possible effects of my non-attendance on the interests of my department, my colleagues, my students, myself, as I am concerning the order in which I do up my shoes in the morning, but still I find myself driven by that irresistible ought-thought to go, I, too, should have my head examined.

These considerations may not amount to a proof of the desire/belief theory: maybe some day a cognitivist will show how some purely cognitive state passes the ‘strip-down’ test; but they do at least shift the onus of proof to the cognitivist side.

For when it comes to detailed consideration of specific cases (such as those discussed above), it is very hard to see how all relevant ought-beliefs can be thoroughly unpacked without any reference to morally relevant desires, but easy to see just which such desires do come into it.

Letting Passions in by the Back Door

In *Moral Vision*,²⁰ David McNaughton rejects the Humean desire-belief theory of moral motivation and argues in favour of the view that morally motivated action may be motivated by belief alone—by a “purely cognitive state.” He then develops qualms (in sec. 7.3, “Cognitivism and Caring”) about an apparent implication of his view—the implication that moral agents may be totally unfeeling—and argues that this alleged implication doesn’t really follow:

Where does the notion of caring fit into the realist picture, given that the realist adopts a cognitive account of moral motivation? Realism might seem to offer us the chilling, and implausible, picture of the virtuous person as someone completely cold and uncaring who simply sees what is right and does it without emotion or concern. Feeling or passion would have no place, except as incidental accompaniments to the entirely unemotional process of recognising the moral facts.

We can now see that this austere portrait is based on a misconception which springs from the false opposition between belief and desire which the cognitivist account of desire seeks to transcend. The crucial mistake, on the cognitivist view, is to fail to realize that a way of seeing a situation may itself be a way of caring or feeling. Again the analogy with aesthetics is helpful here. It is impossible to separate out the musical person's way of hearing Mozart's music from his appreciation of its beauty, and to appreciate its beauty is to care about it. Yet that response is to a quality of the music that, on the realist conception, is there to be experienced in the music.²¹

The analogy with aesthetics could only be "helpful" to the converted. Anyone analytically-minded enough to distinguish desire from belief in moral motivation will be equally inclined to distinguish between perceptual, cognitive, and non-cognitive aspects of aesthetic experience, and to do that for the good reason that these elements occasionally part company in fact as well as in theory. Most people have had the experience of giving their undivided attention to a musical work with which they are well-familiar and which they have appreciated many times in the past, but *not* appreciating it this time: not appreciating it not because one has misperceived or failed to recognise any of its long-familiar aesthetic qualities, but simply because, on this occasion, the music has left one cold—one remains totally unmoved by it (perhaps because one is in a "bad" [emotionally unresponsive] mood at the time). Such cases show, *pace* McNaughton, that it surely is possible to "separate out the musical person's way of hearing Mozart from his appreciation of its beauty."

Apart from this objection to the analogy with aesthetics, McNaughton cannot, without contradiction, hold (i) that moral motivation is a *purely* cognitive state and also claim (ii) "that a way of seeing a situation (purely cognitive state) may also be a way of caring or *feeling* (non-cognitive state)." (This is not to deny that [in some sense] feelings [that is, emotions] have thought-content or even that such

thought-content is [in some sense] conceptually built into them. It is merely to deny that feelings are *wholly reducible* to their thought-components.)

(ii) (at its strongest) implies that ways of seeing are indissoluble compounds of affect and cognition. If that is so, what could possibly be the force of “*purely cognitive*” in (i)? And why call this indissoluble compound theory—this view which claims to “transcend” the distinction between belief and desire—“cognitivism”? Insofar as the proposed transcendence is accomplished through an inseparable union of equal partners, the resulting view has no more claim to the name “cognitivism” than it has to be called “Humean sentimentalism.”

Hume never denied (in fact, explicitly asserted) that, necessarily, reason and sentiment “concur” both in the making of moral judgements and in the motivation of action by them. It is arguable that Hume underestimated or underemphasised both the intimacy of the partnership as well as the variety and importance of the roles of reason. Those differences fall *within* the basic Humean framework of the desire-belief theory. To break out of (transcend) that framework, to construct a distinctively and purely cognitivist anti-Humean view, requires more than just the amalgamation of reason and sentiment or the incorporation of sentiment into reason (via the back door). It requires the *rejection* of sentiment as *inessential* to moral motivation. That view carries the inescapable implication that, although typical moral agents characteristically care deeply/passionately about serious moral issues, they *need not*: their status as moral agents is not in the least compromised or impaired if they *do not*. But that’s the very possibility that McNaughton (along with most rationalists since Kant) rightly fights shy of (in sec. 7.3): the portrait of the moral agent as a totally and unthawably cold fish.

The Authority of Moral Requirements

In sections 3.3 and 7.4 McNaughton rightly observes that moral requirements strike us as having authority over conflicting desires, but (wrongly, in my view) maintains that the desire-belief theory of moral motivation cannot account for this special authority, whereas pure cognitivism can account for it. The idea is the familiar Kantian one that, on the Humean view, moral ‘requirements’ can be ‘got out of’ just by giving up the relevant desire, whereas cognitions (or beliefs) about real requirements are somehow inescapable (or less easily escaped). But why think that it is any harder to turn a blind eye (through self-deception, rationalisation or blinkered vision²²) to ‘real’ cognisable requirements than it is to suppress sentiments of conscience? Cognitions (or beliefs), it will be argued, are at least securely ‘anchored’ in real (or believed) evaluative features of the world ‘out there’, while

'free-floating' (subjective) desires are readily disposed of. But if the moral sentiments (for example, of approval and disapproval, pride, self-esteem, guilt and remorse) derive from essential and universal features of the human mind and of the human condition (as Hume sought to derive them), they can only be got rid of by doing violence to our own natures—a price not worth paying, not even from the viewpoint of (enlightened) self-interest.

Hume derived the moral sentiments from the 'laws' of sympathy and humanity—the natural and nearly universal human tendency to take pleasure in the pleasures of others, to be pained by the pains of others and to experience consequent feelings of approval and disapproval when contemplating actions, motives and traits of character which involve or give rise to pleasure and pain. By an equally natural process (comparable to the mechanisms of veridical sense perception), reason then regulates and corrects these sentiments so as to give to the judgements arising from them the consistency, uniformity and universality that effective moral communication and co-operation requires. Thus rooted in our common humanity, the moral sentiments are neither subjective nor arbitrary, nor can the judgements arising from them be casually set aside by simply forgoing some incidental desire. Hume, of course, acknowledges that, however deeply rooted they may be in normal human nature, the moral sentiments can be uprooted, perverted or prevented from taking root by, for example, the distorting influences of superstition and ideology (much in the way that natural physical development can be stunted by disease). But the psychological deformities of the character who inflicts suffering for its own sake on himself or others, or is ruthlessly indifferent to the sufferings of others, are not of a kind which a normal personality would knowingly and voluntarily take upon himself merely as a convenient escape from the requirements of morality, and the natural abhorrence of such deformities is no more lightly cast aside than are the cognitions of the Kantian.

Hume's account of justice bases our approval of just actions on these same natural sentiments of sympathy and humanity, while explaining the performance of just actions by reference to the equally natural motive of self-interest, enlightened by a rational understanding of the 'artifice' of justice.

I share Barry Stroud's suspicion that "our 'natural' or primitive condition, which is to form the basis of an explanation along Humean lines of the 'origin' of justice, must be understood to be much richer and more complex than [Hume] allows" and that "an elaborate scheme of justice does not serve merely to channel our direct or acquisitive 'avidity' in more mutually beneficial directions, but that a more or less disinterested concern for justice and fairness is a natural outgrowth of

outward looking or socially-orientated emotional needs that human beings inevitably develop."²³ But one need not subscribe to all the details of Hume's account in order to see merit in Hume's general strategy—the strategy of deriving morality from the natural sentiments. I believe that approach can be elaborated and enriched along lines suggested by Stroud (and Rawls,²⁴ to whom Stroud refers) to provide an adequate non-cognitivist account of the authority of moral requirements.

I have not myself worked out the details of such an account, but the points that I would emphasise are these: the natural basis for the authority of the moral requirements has three main aspects: (i) man's nature as an active, purposive, choosing being; (ii) man's nature as a social being—as one amongst others in society; and (iii) man's nature as a being possessing distinctively human needs, powers and capacities. Following Mill,²⁵ I would classify these under the headings: intellect, emotion, imagination.

All three aspects ([i]-[iii] above) are closely inter-related and interdependent, and all three (as well as their inter-relations) are a legacy of man's evolutionary and cultural history. Regarding the first and second: as purposive and social beings, necessarily, individuals enact roles and pursue activities in a social context that are subject to criteria of evaluation (for example, good or bad father, son, colleague, friend, neighbour, citizen, etc.), many of which have a moral aspect and which, for the effective enactment of these roles and activities, must be internalised, and must be motivated, guided and reinforced by internal sanctions, by rewards and punishments of sentiments which also have a moral quality (for example, approval, disapproval, shame, pride, self-esteem, etc.).

Possible roles and activities are far more numerous than any individual could actually enact within the space of a human life. So, necessarily, many of them can be avoided or abandoned (for example, father, neighbour, colleague—one can choose not to have a family, to ignore one's neighbours, or to remain self-employed or unemployed). But the policy of retreating from them all—or just refusing to develop a coherent set of them—as a means of avoiding the requirements of morality, is a recipe for alienation, absurdity, loss of personal integrity and identity, and disfigurement of one's own humanity, in short, for an empty and meaningless life. This is one main (non-cognitivist) source of the authority of moral requirements.

In addition, the social nature of man brings with it a variety of fundamental social needs and desires: for personal love, friendship, companionship, and a sense of belonging to a human community. The various relationships in which these needs and desires find fulfilment have all got a moral dimension, and can be adequately realized and

sustained only by people of a certain moral character. Unless such moral qualities as honesty, loyalty, fairness, compassion, humility, courage, discipline, respect for the rights and freedoms of others, etc. are quite deeply rooted in a person's character, lasting, genuinely fulfilling relationships of love and friendship will, for that person, be quite impossible.²⁶ To the extent that a person fails to develop or undermines these moral qualities in their own character, to the same extent do they cut themselves off from an essential part of their own humanity and of their own self-identity—their need for fulfilling relationships of love, friendship and companionship.

One might wonder whether it was not possible for someone to develop or practise these very moral virtues that I have mentioned, but in a partial and parochial way, practising them only or mainly with respect to some favourite group of people (for example, my family, my friends, my gang, my countrymen, my co-religionists, etc.), thereby achieving deeply meaningful relationships with people within that group, while at the same time acting immorally, unjustly, with callous indifference or opportunistically towards outsiders. Well, no doubt thoroughly unjust people do strike up *some* kind of relationship with their favourites (family, friends and associates). For an argument to the effect that genuinely fulfilling relationships of love and friendship must, necessarily, for all their individuality, also contain an essential element of universality, and are therefore incompatible with indifference, immorality, or injustice towards others see Erich Fromm, *The Art of Loving*.²⁷

Finally (following Mill) I would emphasise the rootedness of the moral sentiments in the higher mental faculties: the distinctively human powers of intellect, imagination and emotion. All sorts of rewarding creative intellectual activities, including the pursuit of knowledge in its many forms, as well as the appreciation of beauty in art, in nature and in everyday life, require these higher levels of thought, imagination and feeling. But, arguably, these powers cannot be developed to high levels without bringing conscience along with them. When we stifle our consciences we stifle these very powers. For example, no one with well developed powers of reason and imagination can fail to identify with the sufferings of the victims of world poverty, or could fail to understand that there are various ways in which he or she could contribute to the relief of at least some of that suffering. But those very powers of reason, feeling and imagination which enable us to appreciate that will trouble our consciences if we fail to do our part. If this is true, then the price of silencing our consciences in morally serious matters such as this is complete or partial loss of those very powers which make us most fully alive as human beings, with the consequent disfigurement of our own humanity.

As a Humean account of the authority of the moral sentiments, these considerations will never satisfy the Kantian rationalist. The requirements of morality, he will argue, are *categorical* demands, universally valid and binding upon the will of every free rational agent, human or otherwise. The above Humean account construes them as a kind of *hypothetical* imperative, albeit a very special kind. On this Humean view, those free rational agents who, as a last resort, are willing to undermine, disfigure or forgo their own humanity, their own human self-identity (Hume's '[dis]ingenuous natures', 'sensible knaves'), can (at this catastrophic cost) escape the requirements of morality and can do so without irrationality. This is not good enough for the rationalist whose categorical imperative can only be violated on pain of irrationality. But why should that same sanction carry more moral authority than the violation—the doing of violence—to one's own humanity? Serious immorality, the kind which goes against *moral requirements*,²⁸ surely has more to do with inhumanity—with man's inhumanity to man, to himself as well as to other men—than with abstract²⁹ irrationality.

Smith's View

In a paper entitled "The Humean Theory of Motivation,"³⁰ Michael Smith offers an account of motivating desires which has striking (though purely coincidental) similarities to my own. In particular, Smith includes an experiential-dispositional element in his account. However, Smith's view differs from both my and Hume's account of desires as introspectible causes of behaviour by also including a behavioural-dispositional component in the analysis of the concept of motivating desire. In this respect, Smith's theory has more in common with modern behaviourist and functionalist accounts than with Hume's, view.

Perhaps the most original feature of Smith's paper is the use he makes of Mark Platts' 'direction of fit' argument³¹ in favour of the desire-belief theory, which, if successful, would dispose of the rationalist/cognitivist theory of motivation forthwith, and without the need for counter-example arguments of the kind I use (against, for example, Stroud's and Nagel's views). My colleague, Mr. Ian Fowlie, has persuaded me that Smith's 'direction of fit' argument does not succeed, or rather that it only works against an implausible version of ethical naturalism: the version that assimilates moral beliefs to empirical beliefs. Briefly, Smith argues that one and the same psychological state (that is, a moral motive) cannot have two different directions of fit towards one and the same state of affairs. Belief fits the world by tending to go out of existence whenever it perceivably fails to match some relevant state of affairs in the world, whereas desire fits

the world by tending to persist in the face of, and aim to remedy, any such perceived mismatch.³² The cognitivist, argues Smith, requires that moral motives relate to evaluative properties in both ways at once, which is logically impossible.

Impossible, replies Fowlie, only on the assumption that moral beliefs fit the world in the same way that empirical beliefs do; that is, by way of a one-way relation to a single property or state of affairs. If that were so, then no doubt it would be a contradiction to suppose that moral beliefs tend both to go out of existence and to remain in existence where such properties or states of affairs fail to match them.³³ But who accepts that view of the relation of moral beliefs to the world? Contemporary cognitivists such as McNaughton argue that moral beliefs have “directions of fit running in both ways”;³⁴ that is, that a moral belief both tends to go out of existence if it fails to match a real moral requirement, and also tends to persist until such requirements have been satisfied by the agent’s morally motivated behaviour. On their view, the moral agent believes that it is a real moral requirement of the situation in which he finds himself that he ought to do so-and-so. That belief of his owes its existence to that requirement and ceases to exist if and when that requirement fails to obtain. But that very same belief also tends to persist until it moves him to act in accordance with that requirement.

For example, the cognitivist may believe that some defensible version of the categorical imperative or the natural law is a real moral requirement which requires the removal of apartheid. On his cognitivist view, his moral belief that apartheid should be done away with owes its existence to the reality of one or other of the above-mentioned moral requirements, in McNaughton’s phrase, “to the way things are, morally.” Had he no other sufficient reasons for accepting it, that belief would go out of existence if, for example, he discovered a serious fallacy in the theory of the categorical imperative, or in natural law theory, or if he discovered that the categorical imperative or natural law didn’t really require the abolition of apartheid. But that very same moral belief of his (that apartheid ought to be done away with) also tends to persist and to continue to motivate him to oppose apartheid until such time as that moral objective is realized.

Nothing in Smith’s ‘direction of fit’ argument rules out this sort of cognitivist theory. For that purpose, my counter-example argument is required.

Further Support for the Counter-Example Argument

In conclusion, I should like to present an additional argument in support of my claim that cases of the kind in my original

HUME'S THEORY OF MOTIVATION—PART 2

counter-example³⁶ (to Stroud's and Nagel's views) really are both logically and empirically possible and really would not count as purposive action in the fullest sense. (In addition, the following argument supports the view that certain empirically possible observations would provide good evidence for thinking that such a case does exist and is rightly described as a case of non-purposive action.) The argument rests on the following three cases:

Case 1: A brain scientist of the future has invented an electrical device, call it Implant 1, which when implanted in the motor cortex of the brain, electrically stimulates neural pathways in such a way as to cause a knee-jerk reflex to take place once every five minutes for as long as it is switched on. For the sake of brain science, one volunteers to have this device installed in one's skull and, sure enough, one's knee jerks once every five minutes for as long as it is switched on. There is no temptation to regard this bodily movement as one's purposive action. One's acquaintance with the facts of the case taken together with one's grasp of the concept of purposive action rules that out.

Case 2: The brain scientist then produces an upgraded version of his original device. Implant 2 so stimulates the motor cortex as to cause a seated person to rise to his feet and stand for a moment and then return to the seated position once every five minutes for as long as it is switched on. The subject of the experiment experiences this event as a purely passive bodily movement bearing no relation to his desires or to his will. (Let us suppose the movement happens without his wanting to rise at that moment, though if he wants to prevent it he can, much in the way one can restrain a mild tendency to cough or blink one's eyelids.) The movement takes place so smoothly that, to an outside observer, it looks like a purposive action.

Again, one volunteers for the experiment and it comes off exactly as the brain scientist said it would. But again, and for the same combination of factual and conceptual reasons, one does not conceive of one's jack-in-the-box performance as one's own purposive action.

Case 3: Our brain scientist has made real progress. Implant 3 is a sophisticated 'teleological' guidance system which connects to and receives information from perceptual and cognitive areas of the brain, information which influences Implant 3's effect upon the motor cortex. The scientist can

program Implant 3 to propel your body through any stretch of 'goal-seeking' behaviour he chooses irrespective of your desires or your will, though activated and influenced by certain of your perceptions and beliefs; that is, guided by sensory-input. He invites you to name some activity you have no interest in engaging in and would never voluntarily undertake. You name assembling, disassembling and reassembling a huge jigsaw picture of some famous tourist attraction. Fine. He programs Implant 3 such that whenever you acquire the belief that you have an opportunity to do that, Implant 3 will drive your body through the relevant bodily movements and will do so so smoothly that any uninitiated observer would assume that you were a keen jigsaw buff. (He can even program Implant 3 to produce appropriate verbal behaviour; for example, "Could it be the Eiffel Tower? I can't wait to see!") You yourself, however, are well aware that this is no purposive activity of yours. You experience the entire sequence of sensory-guided bodily movements, both verbal and non-verbal, as if you were a passenger in your own body, for it answers to no desire, interest, or purpose that you are aware of having. Your bodily behaviour, though more complex and prolonged, is of the same kind as that produced by Implants 1 and 2—non-purposive bodily movement.

The above case satisfies all the requirements of Stroud's functionalist account³⁶ of motivated action; after receiving Implant 3 the above agent is in some state, S, which, in conjunction with certain of his beliefs, accounts for his overt behaviour. Therefore, Stroud's account implies that the above case is a central case of purposive action motivated by the above-mentioned desire. But clearly it is not. It is not because of the missing experiential pro-attitude component.

It might be objected that the case fails to count as purposive action not because of the missing experiential pro-attitude component but because of the queer background; that is, the brain scientist's intervention. But I don't think that this is right; for had the brain scientist modified the subject's brain so as to engender in him an introspectible pro-attitude towards the assembling of the jigsaw pictures, a pro-attitude which then accounted for his activity, neither the agent himself nor anyone else would have had reason for denying that the agent's subsequent pursuit of his new hobby was a case of purposive action motivated by that desire. All we should say in that case, if we knew the background, was that his newly-acquired interest came about through surgical intervention. (There might be some

HUME'S THEORY OF MOTIVATION—PART 2

question about how free such action is, but no question about its status as purposive action.)

Alternatively, if an organic neural network functionally identical to Implant 3 and equally disconnected from the 'agent's' introspectible pro-attitudes grew 'naturally' in someone's brain, given that the case was in all other ways identical to the case of Implant 3, we would not regard it as a central case of purposive action.

The case of Implant 3 will also serve as a counter-example against Nagel's account³⁷ of motivated action—provided that it is made clear that the set of beliefs which are operative along with state S includes some, for example, prudential beliefs. For example, the brain scientist might program Implant 3 to propel the 'agent' through the motions of entering a jigsaw puzzle competition and winning prize money. In this case, Implant 3 will only produce this activity if the agent holds the prudential belief about the prize money. Since he is aware of having *no desire* to compete for prize money in this way (preferring other means of earning a living), he is still a "passenger in his own body," as Implant 3, in conjunction with the aforesaid prudential belief about the prize money, drives him though automatic jigsaw construction.

But since the conditions of Nagel's account of motivated actions are all satisfied, that account falsely implies that his behaviour is a central case of purposive action.

Functionalists, pure rationalists and other philosophers who are sceptical about including an experiential pro-attitudinal component as a necessary condition in the account of desire will probably remain unconvinced by the above arguments. Why, they will protest, should the concept of motive and desire be essentially tied to problematic phenomena of inner experience? Even if the above arguments show that the ordinary concept of desire is so tied, why should a philosophical account of motives give any importance to this alleged connection? Surely the most significant features of motivation pertain to its observable expression in outward behaviour and to those factors, *whatever* they may be, which explain outward behaviour, not to the phenomena of introspection.

One is tempted to answer that if my above arguments do establish that the experiential component is necessary to our ordinary concept of motives and motivated action, that is, that what we are left with when we subtract the experiential component from motivated action is no kind of motivated action at all but rather a peculiar species of automatic bodily movement, then the proposed behaviourist or functionalist account can only be carried out on pain of changing the subject from motivated action to something quite different.

Why not carry out the program anyway? One might as well ask why not reduce an itch to a tendency to scratch (or to the non-experiential causes of scratching 'behaviour')?

In a sense, there is nothing to stop us from such conceptual reforms as long as we realize that what we are left with fails to account for the original subject.

But the above rejoinder may miss the point of the objection. The objector may be seeking a deeper philosophical explanation (than that provided by the argument from counter-examples) of *why it is* that our concepts of motives, desires and motivated action *should* have an experiential component, and of why it is that any attempt to reduce these concepts to wholly non-experiential ingredients *should* be viewed as a misrepresentation and a distortion rather than as a clarifying improvement. If the 'subject' turns out to be infected with a scientifically unmanageable and philosophically problematic experiential component, why not (partially) change the subject?

The challenge, then, is to bring out the *importance* of the experiential component of motivating desires. To do this we would have to trace the connections between the experiential component of desire and other fundamental features of thought and action such as responsibility, happiness, moral evaluation, knowledge of one's own motives and intentions, etc. I shall confine the following remarks to the importance of the experiential component of desire for understanding and explaining motivated action.

As with many reductionist accounts, accounts of desire which represent motivating desires as 'nothing over and above' non-experiential states or conditions undermine and render unintelligible the *explanatory role* of the very phenomenon they purport to account for. If *per impossible* itches were reducible to tendencies to scratch, they could never *explain* them, as it could not be the case, as it undoubtedly is, that one tends to scratch *because* one has an itch. If, as D. M. Armstrong maintains, visual perceptions were nothing over and above acquisitions of perceptual beliefs, it could not be the case that *what explains* my newly acquired belief that there is a red tomato in front of me is the fact that I have just caught sight of one.

So, too, at least part of what explains the fact that certain beliefs plus certain psychological states dispose me to pursue a certain goal is that I have an introspectible pro-attitude towards its realization: that at least is what we commonly believe concerning our conscious desires. (Were some mechanist or epiphenomenalist theory to overthrow that belief, it would only do so by casting doubt on the genuinely voluntary and purposive character of our actions.) But we could not begin to explain our purposive actions in this way if the experiential component did not have a central role in the concept of motivating desires. The

functionalist will protest that we can perfectly well explain teleological behaviour as the effect of some non-experiential (for example, physiological) state. In a sense, such explanations of 'behaviour' are possible. They give us a causal explanation of bodily movements. But do they give us what we normally regard as full, humanly intelligible understanding of motivated action? (What Weber called *Verstehn*—interpretative understanding.)

In an essay entitled "Free Agency,"³⁸ Gary Watson writes: "It would be impossible for a non-erotic being or a person who lacked the appetite for food and drink fully to understand the value most of us attach to sex and dining. Sexual activity must strike the non-erotic being as perfectly grotesque. Or consider an appetite that is in fact 'unnatural' (i.e., acquired): the craving for tobacco. To a person who has never known the enticement of Lady Nicotine, what could be more incomprehensible than the filthy practice of consummating a fine meal by drawing into one's lungs the noxious fumes of a burning weed?"

Surely Watson is right to describe these as cases of incomplete understanding of motivated action. Equally clearly, the failure of understanding is due in these cases to lack of acquaintance with the experiential component of the motivating desires. If the concept of desire did not *have* an essential experiential component, this kind of failure of understanding would itself be unintelligible.

In the kind of cases Watson describes, we may suppose that the puzzled observer is provided with a full account of the agent's belief about sex or smoking—that would not dispel his puzzlement. We may also provide him with a full account of the functional states underlying sexual activity or nicotine addiction—that would not remedy the incompleteness of his understanding. If the subject of his observations were a robot or a chess-playing computer, there would be no gap in his understanding of the 'behaviour' he was observing, for then his understanding would be as complete as the subject admits of. But in that case, the subject would not be humanly intelligible motivated action in the fullest sense of the term.

University of Aberdeen

1. Daniel Skaw, "Hume's Theory of Motivation," *Hume Studies* 15, no. 1 (April 1989): 163-83.
2. See Barry Stroud, *Hume* (London, 1977); and T. Nagel, *The Possibility of Altruism* (Oxford, 1970).
3. Stroud (above, n. 2), 168.
4. Similar apart from the above-mentioned difference about the role of reason in action.

5. For details of this counter-example argument, see Shaw (above, n. 1), 169-77, esp. 171-73.
6. Stroud (above, n. 2), 164-65.
7. I.e., no grounds independent of their alleged effects on behaviour.
8. Of conscious desire. For the extension of this account to cover the case of unconscious desire, see Shaw (above, n. 1), 174.
9. In order to rule out cases in which *P* does not actually desire *S* at *t*, but is merely *disposed* at *t* to acquire a desire for *S*, the following condition must be added to the above account: "Either at *t* or at some time prior to *t*, *P* has (at least) once *actually* experienced an introspectible pro-attitude towards the prospect of *S*."
10. For references to and discussion of these passages see Shaw (above, n. 1), 177-78.
11. Available, though, for as long as the passion remains calm, he does not avail himself of it.
12. This is true for conscious, not unconscious desires. For the somewhat different way in which independent evidence of the existence of unconscious desires (on my account of them) is also available (at least in principle) to the agent, see Shaw (above, n. 1), 174.
13. The subject of all that follows, but for the last section.
14. Distinct from the cognitive component of the moral motive, though, as Hume held, internal to the sincere acceptance of the motivating moral judgement.
15. This characterisation of cognitivism mentions belief but gives no account of the concept of belief. The omission is deliberate. Cognitivists (*qua* cognitivists) differ as to the specific nature of belief, agreeing only about the general claim that belief, in itself, is a purely intellectual and wholly dispassionate state which is alone capable of motivating action in certain cases. By contrast, Humean non-cognitivists maintain that, in addition to belief, something of the nature of a passion (e.g., an introspectible sentiment, feeling, emotion, pro- or con-attitude, etc.) must also be involved in all motivated action, though the involvement may be quite indirect, as in the case of my dispositional analysis of desire and the calm passions. (Non-Humean [i.e., non-sentimentalist] versions of non-cognitivism are also possible. My positive arguments in support of Hume's passion requirement, if sound, would refute these.)

The detailed working out of a defensible Humean account of belief to go along with my account of desire is beyond the scope of this paper and, as far as I can see, is not required for the purpose of my present argument. Briefly and roughly, I think that the most promising line for the Humean to take about belief runs closely

parallel to the present account of desire, representing belief as a disposition to experience an introspectible attitude of assent in connection with the thought of some state of affairs obtaining. On this account, the essential difference between desire and belief would be the intrinsic, introspectively recognisable difference between the dispassionate attitudes of assent and dissent as distinct from the 'passionate' pro- and con-attitudes of favouring and disfavouring. Differences in 'direction of fit' (discussed below) could be represented as differences in the causal relations in which desires and beliefs stand to other psychological states (e.g., perceptual states), as well as to the behaviour which they motivate and to other states of affairs in the world which tend to give rise to them, sustain them, or bring them to an end. I am not convinced that any of these relations must be regarded as internal to the essential nature of desire and/or belief. Clearly, it would be more in keeping with Hume's philosophy of mind and action to treat these relations as causal contingent ones—empirical laws (or generalisations) of human psychology.

16. I.e., anything of the nature of introspectible affect, or feeling, or caring, or introspectible attitudes of favouring, or disfavouring.
17. Fowlie's own view is not—as these examples might suggest—the view criticised below.
18. To paraphrase Kant, "when no longer moved by any inclination she tears herself out of this deadly insensibility and does the action without any inclination for the sense of duty alone." See I. Kant, *The Moral Law*, ed. H. J. Paton (London, 1961), 66 (Kant's original, p. 11). In support of his view that, in cases of this kind, the "sense of duty" must be a purely cognitive state, the cognitivist will point out that, (i) we commonly use the language of belief to describe it ("I believe I ought ... etc."); (ii) we systematically draw cognitive-seeming inferences about it ("If I ought to do X then I ought to do Y"); and (iii) we cannot, in cases of the kind in question, find any non-cognitive motivational factor with which to identify the sense of duty. So what else could it be? My answer is: an unactualised, non-cognitive disposition.
19. Potential feelings are really important here. They are the calm passions, the unactualised feeling dispositions, which could (if she had them) rescue Amanda (not in Kant's, but in Hume's way) from this moral predicament. Though unfelt at the time of action, they do essentially involve feelings and are capable of motivating action, even when unfelt, and even against strongly felt opposing passions. For example (the above example), let us suppose Amanda to be so preoccupied by thoughts of her amorous desires as to give little or no conscious thought to the objects of her conscientious ones. As a

consequence, her conscientious feelings (of obligation, moral aversion, etc.) will remain unactualised, and in the extreme case, wholly repressed (unconscious or subconscious). Nevertheless (on the proposed theory of the calm passions), her (unactualised) *disposition* to have such feelings may still influence her behaviour. For more about how an unactualised disposition can be operative in motivating action, see Shaw (above, n. 1), 175-76. For more about my account of unconscious desire see pp. 174-75.

20. D. McNaughton, *Moral Vision: An Introduction to Ethics* (New York, 1988).
21. *Ibid.*, 113.
22. Freud did not discover self-deception; but the psychoanalytic study of mind and action, whatever its other shortcomings, has certainly taught us a lot about the 'mechanisms of defence'—the many and various ways in which unwelcome reality can be denied. Unpalatable beliefs are no less repressible than are embarrassing desires; and there is no reason to think that the cognitivist's ought-beliefs, as well as the many is-beliefs upon which they would depend, would be any exception. Including as they would all the agent's beliefs about his own moral failings, weaknesses, and vices, they make first class grist for the mills of deception.
23. Stroud (above, n. 2), 217.
24. John Rawls, *A Theory of Justice* (Cambridge, Massachusetts, 1972), chap. 8, 485-96.
25. J. S. Mill, *Utilitarianism*, ed. M. Warnock (Collins: 1982), 258.
26. This view is convincingly argued in Erich Fromm, *The Art of Loving* (London, 1957), esp. chaps. 2, 4.
27. *Ibid.*, chaps. 2, 4.
28. The kind that overrides moral requirements as distinct from non-decisive moral considerations; see McNaughton (above, n. 20), 114-15.
29. The abstract formalism of the Kantian view is well brought out by P. Foot, who writes, "Kant's argument that moral rules have a peculiar and dignified status depends wholly upon his attempt to link moral action with rationality through the mere concept of the form of law and the principle of universalisability, as interpreted by him." See P. Foot, "Morality as a System of Hypothetical Imperatives," *Philosophical Review* 81 (1972): 305.
30. Michael Smith, "The Humean Theory of Motivation," *Mind* 96 (1987): 36-61.
31. M. Platts, *Ways of Meaning* (London, 1979), 256-57.
32. In Smith's words: "the difference between beliefs and desires in terms of direction of fit comes down to a difference between the counterfactual dependence of a belief and a desire that p, on a

perception that not-p: roughly, a belief that p is a state that tends to go out of existence in the presence of a perception that not-p, whereas a desire that p is a state that tends to endure, disposing the subject in that state to bring it about that p. Thus ... attributions of beliefs and desires require that different *kinds* of counterfactuals are true of the subject to whom they are attributed ... this is what a difference in their direction of fit *is*" (Smith [above, n. 29], 54).

33. E.g., if the obligatoriness of an action were like the redness of a post box, then it would be no more possible for my belief that an as-yet-unperformed action ought to be done to persist in the face of my knowledge that it has not yet been done until it motivates me to do the act, than it would be for my belief that an as-yet-unpainted post box is red to persist in the face of my knowledge that it is not yet red until it has motivated me to paint it red.
34. McNaughton (above, n. 20), 110. "[T]he single state in question," McNaughton continues, "can be plausibly held to be a belief even though it has both directions of fit [I]t is a sufficient condition of a state being a belief that it has the direction of fit; this state must fit the world. It does not lose the status of belief if it happens to have another direction of fit as well."
35. Shaw (above, n. 1), 171-73.
36. See the introduction to this paper, paragraph 3.
37. See the introduction to this paper, paragraph 4.
38. G. Watson, "Free Agency," in *Free Will*, ed. G. Watson, Oxford Readings in Philosophy (Oxford, 1982), 103-4.