

Free Agency

Author(s): Gary Watson

Source: *The Journal of Philosophy*, Vol. 72, No. 8 (Apr. 24, 1975), pp. 205-220

Published by: [Journal of Philosophy, Inc.](#)

Stable URL: <http://www.jstor.org/stable/2024703>

Accessed: 25-05-2015 20:43 UTC

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*Journal of Philosophy, Inc.* is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy*.

<http://www.jstor.org>

---

---

# THE JOURNAL OF PHILOSOPHY

VOLUME LXXII, NO. 8, APRIL 24, 1975

---

---

## FREE AGENCY \*

**I**N this essay I discuss a distinction that is crucial to a correct account of free action and to an adequate conception of human motivation and responsibility.

### I

According to one familiar conception of freedom, a person is free to the extent that he is able to do or get what he wants. To circumscribe a person's freedom is to contract the range of things he is able to do. I think that, suitably qualified, this account is correct, and that the chief and most interesting uses of the word 'free' can be explicated in its terms. But this general line has been resisted on a number of different grounds. One of the most important objections—and the one upon which I shall concentrate in this paper—is that this familiar view is too impoverished to handle talk of free actions and free will.

Frequently enough, we say, or are inclined to say, that a person is not in control of his own actions, that he is not a "free agent" with respect to them, even though his behavior is intentional. Possible examples of this sort of action include those which are explained by addictions, manias, and phobias of various sorts. But the concept of free action would seem to be pleonastic on the analysis of freedom in terms of the ability to get what one wants. For if a person does something intentionally, then surely he was able at that time to do it. Hence, on this analysis, he was free to do it. The familiar account would not seem to allow for any further questions, as far as freedom is concerned, about the action. Accordingly, this account would seem to embody a conflation of free action and intentional action.

\* I have profited from discussions with numerous friends, students, colleagues, and other audiences, on the material of this essay; I would like to thank them collectively. However, special thanks are due to Joel Feinberg, Harry Frankfurt, and Thomas Nagel.

Philosophers who have defended some form of compatibilism have usually given this analysis of freedom, with the aim of showing that freedom and responsibility are not really incompatible with determinism. Some critics have rejected compatibilism precisely because of its association with this familiar account of freedom. For instance, Isaiah Berlin asks: if determinism is true,

. . . what reasons can you, in principle, adduce for attributing responsibility or applying moral rules to [people] which you would not think it reasonable to apply in the case of compulsive choosers—kleptomaniacs, dipsomaniacs, and the like? <sup>1</sup>

The idea is that the sense in which actions would be free in a deterministic world allows the actions of “compulsive choosers” to be free. To avoid this consequence, it is often suggested, we must adopt some sort of “contra-causal” view of freedom.

Now, though compatibilists from Hobbes to J. J. C. Smart have given the relevant moral and psychological concepts an exceedingly crude treatment, this crudity is not inherent in compatibilism, nor does it result from the adoption of the conception of freedom in terms of the ability to get what one wants. For the difference between free and unfree actions—as we normally discern it—has nothing at all to do with the truth or falsity of determinism.

In the subsequent pages, I want to develop a distinction between wanting and valuing which will enable the familiar view of freedom to make sense of the notion of an unfree action. The contention will be that, in the case of actions that are unfree, the agent is unable to get what he most wants, *or values*, and this inability is due to his own “motivational system.” In this case the obstruction to the action that he most wants to do is his own will. It is in this respect that the action is unfree: the agent is obstructed in and by the very performance of the action.

I do not conceive my remarks to be a defense of compatibilism. This point of view may be unacceptable for various reasons, some of which call into question the coherence of the concept of responsibility. But these reasons do not include the fact that compatibilism relies upon the conception of freedom in terms of the ability to get what one wants, nor must it conflate free action and intentional action. If compatibilism is to be shown to be wrong, its critics must go deeper.

## II

What must be true of people if there is to be a significant notion of free action? Our talk of free action arises from the apparent fact

<sup>1</sup> *Four Essays on Liberty* (New York: Oxford, 1969), pp. xx–xxi.

that what a person most wants may not be what he is finally moved to get. It follows from this apparent fact that the extent to which one wants something is not determined solely by the *strength* of one's desires (or "motives") as measured by their effectiveness in action. One (perhaps trivial) measure of the strength of the desire or want is that the agent acts upon that desire or want (trivial, since it will be nonexplanatory to say that an agent acted upon that desire because it was the strongest). But, if what one most wants may not be what one most strongly wants, by this measure, then in what sense can it be true that one most wants it?<sup>2</sup>

To answer this question, one might begin by contrasting, at least in a crude way, a humean with a platonic conception of practical reasoning. The ancients distinguished between the rational and the irrational parts of the soul, between Reason and Appetite. Hume employed a superficially similar distinction. It is important to understand, however, that (for Plato at least) the rational part of the soul is not to be identified with what Hume called "Reason" and contradistinguished from the "Passions." On Hume's account, Reason is not a source of motivation, but a faculty of determining what is true and what is false, a faculty concerned solely with "matters of fact" and "relations among ideas." It is completely dumb on the question of what to do. Perhaps Hume could allow Reason this much practical voice: given an initial set of wants and beliefs about what is or is likely to be the case, particular desires are generated in the process. In other words, a humean might allow Reason a crucial role in deliberation. But its essential role would not be to supply motivation—Reason is not that kind of thing—but rather to calculate, within a context of desires and ends, how to fulfill those desires and serve those ends. For Plato, however, the rational part of the soul is not some kind of inference mechanism. It is itself a source of motivation. In general form, the desires of Reason are desires for "the Good."

Perhaps the contrast can be illustrated by some elementary notions from decision theory. On the Bayesian model of deliberation, a preference scale is imposed upon various states of affairs contingent upon courses of action open to the agent. Each state of affairs can be assigned a numerical value (initial value) according to its place on the scale; given this assignment, and the probabilities

<sup>2</sup> I am going to use 'want' and 'desire' in the very inclusive sense now familiar in philosophy, whereby virtually any motivational factor that may figure in the explanation of intentional action is a want; 'desire' will be used mainly in connection with the appetites and passions.

that those states of affairs will obtain if the actions are performed, a final numerical value (expected desirability) can be assigned to the actions themselves. The rational agent performs the action with the highest expected desirability.

In these terms, on the humean picture, Reason is the faculty that computes probabilities and expected desirabilities. Reason is in this sense neutral with respect to actions, for it can operate equally on any given assignment of initial values and probabilities—it has nothing whatsoever to say about the assignment of initial values. On the platonic picture, however, the rational part of the soul itself determines what has *value* and how much, and thus is responsible for the original ranking of alternative states of affairs.

It may appear that the difference between these conceptions is merely a difference as to what is to be called “Reason” or “rational,” and hence is not a substantive difference. In speaking of Reason, Hume has in mind a sharp contrast between what is wanted and what is thought to be the case. What contrast is implicit in the platonic view that the ranking of alternative states of affairs is the task of the rational part of the soul?

The contrast here is not trivial; the difference in classificatory schemes reflects different views of human psychology. For one thing, in saying this (or what is tantamount to this) Plato was calling attention to the fact that it is one thing to think a state of affairs good, worth while, or worthy of promotion, and another simply to desire or want that state of affairs to obtain. Since the notion of value is tied to (cannot be understood independently of) those of the good and worthy, it is one thing to value (think good) a state of affairs and another to desire that it obtain. However, to think a thing good is at the same time to desire it (or its promotion). Reason is thus an original spring of action. It is because valuing is essentially related to thinking or *judging* good that it is appropriate to speak of the wants that are (or perhaps arise from) evaluations as belonging to, or originating in, the rational (that is, *judging*) part of the soul; values provide *reasons* for action. The contrast is with desires, whose objects may not be thought good and which are thus, in a natural sense, blind or irrational. Desires are mute on the question of what is good.<sup>3</sup>

<sup>3</sup> To quote just one of many suggestive passages: “We must . . . observe that within each one of us there are two sorts of ruling or guiding principle that we follow. One is an innate desire for pleasure, the other an acquired judgment that aims at what is best. Sometimes these internal guides are in accord, sometimes at variance; now one gains the mastery, now the other. And when judgment guides us rationally toward what is best, and has the mastery, that mastery is

Now it seems to me that—given the view of freedom as the ability to get what one wants—there can be a problem of free action only if the platonic conception of the soul is (roughly) correct. The doctrine I shall defend is platonic in the sense that it involves a distinction between valuing and desiring which depends upon there being independent sources of motivation. No doubt Plato meant considerably more than this by his parts-of-the-soul doctrine; but he meant at least this. The platonic conception provides an answer to the question I posed earlier (207): in what sense can what one most wants differ from that which is the object of the strongest desire? The answer is that the phrase ‘what one most wants’ may mean either “the object of the strongest desire” or “what one most *values*.” This phrase can be interpreted in terms of strength or in terms of ranking order or preference. The problem of free action arises because what one desires may not be what one values, and what one most values may not be what one is finally moved to get.<sup>4</sup>

The tacit identification of desiring or wanting with valuing is so common<sup>5</sup> that it is necessary to cite some examples of this distinction in order to illustrate how evaluation and desire may diverge. There seem to be two ways in which, in principle, a discrepancy may arise. First, it is possible that what one desires is not *to any degree*

---

called temperance, but when desire drags us irrationally toward pleasure, and has come to rule within us, the name given to that rule is wantonness” (*Phaedrus*, 237e–238e; Hackforth trans.).

For a fascinating discussion of Plato’s parts-of-the-soul doctrine, see Terry Penner’s “Thought and Desire in Plato,” in Gregory Vlastos, ed., *Plato: A Collection of Critical Essays*, vol. II, (New York: Anchor, 1971). As I see it (and here I have been influenced by Penner’s article), the distinction I have attributed to Plato was meant by him to be a solution to the socratic problem of *akrasia*.

I would argue that this distinction, though necessary, is insufficient for the task, because it does not mark the difference between (“mere”) incontinence or weakness of will and psychological compulsion. This difference requires a careful examination of the various things that might be meant in speaking of the strength of a desire.

<sup>4</sup> Here I shall not press the rational/nonrational contrast any further than this, though Plato would have wished to press it further. However, one important and anti-Humean implication of the minimal distinction is this: it is not the case that, if a person desires to do *X*, he therefore has (or even regards himself as having) a reason to do *X*.

<sup>5</sup> For example, I take my remarks to be incompatible with the characterization of value R. B. Perry gives in *General Theory of Value* (Cambridge, Mass.: Harvard, 1950). In ch. v, Perry writes: “This, then, we take to be the original source and constant feature of all value. That which is an object of interest is *eo ipso* invested with value.” And ‘interest’ is characterized in the following way: “. . . liking and disliking, desire and aversion, will and refusal, or seeking and avoiding. It is to this all-pervasive characteristic of the motor-affective life, this *state, act, attitude or disposition of favor* of disfavor, to which we propose to give the name of ‘interest.’”

valued, held to be worth while, or thought good; one assigns *no* value whatever to the object of one's desire. Second, although one may indeed value what is desired, the strength of one's desire may not properly reflect the degree to which one values its object; that is, although the object of a desire is valuable, it may not be deemed the most valuable in the situation and yet one's desire for it may be stronger than the want for what is most valued.

The cases in which one in no way values what one desires are perhaps rare, but surely they exist. Consider the case of a woman who has a sudden urge to drown her bawling child in the bath; or the case of a squash player who, while suffering an ignominious defeat, desires to smash his opponent in the face with the racquet. It is just false that the mother values her child's being drowned or that the player values the injury and suffering of his opponent. But they desire these things nonetheless. They desire them in spite of themselves. It is not that they assign to these actions an initial value which is then outweighed by other considerations. These activities are not even represented by a positive entry, however small, on the initial "desirability matrix."

It may seem from these examples that this first and radical sort of divergence between desiring and valuing occurs only in the case of momentary and inexplicable urges or impulses. Yet I see no conclusive reason why a person could not be similarly estranged from a rather persistent and pervasive desire, and one that is explicable enough. Imagine a man who thinks his sexual inclinations are the work of the devil, that the very fact that he has sexual inclinations bespeaks his corrupt nature. This example is to be contrasted with that of the celibate who decides that the most fulfilling life for him will be one of abstinence. In this latter case, *one* of the things that receive consideration in the process of reaching his all-things-considered judgment is the value of sexual activity. There is something, from his point of view, to be said for sex, but there is more to be said in favor of celibacy. In contrast, the man who is estranged from his sexual inclinations does not acknowledge even a *prima facie* reason for sexual activity; that he is sexually inclined toward certain activities is not even *a* consideration. Another way of illustrating the difference is to say that, for the one man, forgoing sexual relationships constitutes a *loss*, even if negligible compared with the gains of celibacy; whereas from the standpoint of the other person, no loss is sustained at all.

Now, it must be admitted, any desire may provide the basis for a reason insofar as nonsatisfaction of the desire causes suffering and

hinders the pursuit of ends of the agent. But it is important to notice that the reason generated in this way by a desire is a reason for *getting rid* of the desire, and one may get rid of a desire either by satisfying it or by eliminating it in some other manner (by tranquilizers, or cold showers). Hence this kind of reason differs importantly from the reasons based upon the evaluation of the activities or states of affairs in question. For, in the former case, attaining the object of desire is simply a means of eliminating discomfort or agitation, whereas in the latter case that attainment is the end itself. Normally, in the pursuit of the objects of our wants we are not attempting chiefly to relieve ourselves. We aim to satisfy, not just eliminate, desire.

Nevertheless, aside from transitory impulses, it may be that cases wherein nothing at all can be said in favor of the object of one's desire are rare. For it would seem that even the person who conceives his sexual desires to be essentially evil would have to admit that indulgence would be pleasurable, and surely that is something. (Perhaps not even this should be admitted. For indulgence may not yield pleasure at all in a context of anxiety. Furthermore, it is not obvious that pleasure is intrinsically good, independently of the worth of the pleasurable object.) In any case, the second sort of divergence between evaluation and desire remains: it is possible that, in a particular context, what one wants most strongly is not what one most values.

The distinction between valuing and desiring is not, it is crucial to see, a distinction among desires or wants according to their content. That is to say, there is nothing in the specification of the objects of an agent's desires that singles out some wants as based upon that agent's values. The distinction in question has rather to do with the *source* of the want or with its role in the total "system" of the agent's desires and ends. It has to do with why the agent wants what he does.

Obviously, to identify a desire or want simply in terms of its content is not to identify its source(s). It does not follow from my wanting to eat that I am hungry. I may want to eat because I want to be well-nourished; or because I am hungry; or because eating is a pleasant activity. This single desire may have three independent sources. (These sources may not be altogether independent. It may be that eating is pleasurable only because I have appetites for food.) Some specifications of wants or desires—for instance, as cravings—pick out (at least roughly) the source of the motivation.

It is an essential feature of the appetites and the passions that



they engender (or consist in) desires whose existence and persistence are independent of the person's judgment of the good. The appetite of hunger involves a desire to eat which has a source in physical needs and physiological states of the hungry organism. And emotions such as anger and fear partly consist in spontaneous inclinations to do various things—to attack or to flee the object of one's emotion, for example. It is intrinsic to the appetites and passions that appetitive and passionate beings can be motivated in spite of themselves. It is because desires such as these arise independently of the person's judgment and values that the ancients located the emotions and passions in the irrational part of the soul;<sup>6</sup> and it is because of this sort of independence that a conflict between valuing and desiring is possible.<sup>7</sup>

These points may suggest an inordinately dualistic view according to which persons are split into inevitably alien, if not always antagonistic, halves. But this view does not follow from what has been said. As central as it is to human life, it is not often noted that some activities are valued only to the extent that they are objects of the appetites. This means that such activities would never be regarded as valuable constituents of one's life were it not for one's susceptibility to "blind" motivation—motivation independent of one's values. Sexual activity and eating are again examples. We may value the activity of eating to the degree that it provides nourishment. But we may also value it because it is an enjoyable activity, even though its having this status depends upon our appetites for food, our hunger. In the case of sex, in fact, if we were not erotic creatures, certain activities would not only lose their value to us, they might not even be physiologically possible.

These examples indicate, not that there is no distinction between desiring and valuing, but that the value placed upon certain activities depends upon their being the fulfillment of desires that arise and persist independently of what we value. So it is not that, when we value the activity of eating, we think there are reasons to eat no matter what other desires we have; rather, we value eating when food appeals to us; and, likewise, we value sexual relation-

<sup>6</sup> Notice that most emotions differ from passions like lust in that they involve beliefs and some sort of valuation (cf. resentment). This may be the basis for Plato's positing a third part of the soul which is in a way partly rational—viz. *Thumos*.

<sup>7</sup> To be sure, one may attempt to cultivate or eliminate certain appetites and passions, so that the desires that result may be in this way dependent upon one's evaluations. Even so, the resulting desires will be such that they can persist independently of one's values. It is rather like jumping from an airplane.

ships when we are aroused. Here an essential part of the *content* of our evaluation is that the activity in question be motivated by certain appetites. These activities may have value for us only insofar as they are appetitively motivated, even though to have these appetites is not *ipso facto* to value their objects.

Part of what it means to value some activities in this way is this: we judge that to cease to have such appetites is to lose something of worth. The judgment here is not merely that, if someone has these appetites, it is worth while (*ceteris paribus*) for him to indulge them. The judgment is rather that it is of value to have and (having them) to indulge these appetites. The former judgment does not account for the eunuch's loss or sorrow, whereas the latter does. And the latter judgment lies at the bottom of the discomfort one may feel when one envisages a situation in which, say, hunger is consistently eliminated and nourishment provided by insipid capsules.

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 to dining. Sexual activity must strike the non-erotic being as perfectly grotesque. (Perhaps that is why lust is sometimes said to be disgusting and sinful in the eyes of God.) 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?

Thus, the relationship between evaluation and motivation is intricate. With respect to many of our activities, evaluation depends upon the possibility of our being moved to act independently of our judgment. So the distinction I have been pressing—that between desiring and valuing—does not commit one to an inevitable split between Reason and Appetite. Appetitively motivated activities may well constitute for a person the most worth-while aspects of his life.<sup>8</sup> But the distinction does commit us to the possibility of such a split. If there are sources of motivation independent of the agent's values, then it is possible that sometimes he is motivated to do things he does not deem worth doing. This possibility is the basis for the principal problem of free action: a person may be obstructed by his own will.

<sup>8</sup> It is reported that H. G. Wells regarded the most important themes of his life to have been (1) the attainment of a World Society, and (2) sex.

A related possibility that presents considerable problems for the understanding of free agency is this: some desires, when they arise, may “color” or influence what appear to be the agent’s evaluations, but only temporarily. That is, when and only when he has the desire, is he inclined to think or say that what is desired or wanted is worth while or good. This possibility is to be distinguished from another, according to which one thinks it worth while to eat when one is hungry or to engage in sexual activity when one is so inclined. For one may think this even on the occasions when the appetites are silent. The possibility I have in mind is rather that what one is disposed to say or judge is temporarily affected by the presence of the desire in such a way that, both before and after the “onslaught” of the desire, one judges that the desire’s object is worth pursuing (in the circumstances) whether or not one has the desire. In this case one is likely, in a cool moment, to think it a matter for regret that one had been so influenced and to think that one should guard against desires that have this property. In other cases it may not be the desire itself that affects one’s judgment, but the set of conditions in which those desires arise—e.g., the conditions induced by drugs or alcohol. (It is noteworthy that we say: “under the influence of alcohol.”) Perhaps judgments made in such circumstances are often in some sense self-deceptive. In any event, this phenomenon raises problems about the identification of a person’s values.

Despite our examples, it would be mistaken to conclude that the only desires that exhibit an independence of evaluation are appetitive or passionate desires. In Freudian terms, one may be as dissociated from the demands of the super-ego as from those of the id. One may be disinclined to move away from one’s family, the thought of doing so being accompanied by compunction; and yet this disinclination may rest solely upon acculturation rather than upon a current judgment of what one is to do, reflecting perhaps an assessment of one’s “duties” and interests. Or, taking another example, one may have been habituated to think that divorce is to be avoided in all cases, so that the aversion to divorce persists even though one sees no justification for maintaining one’s marriage. In both of these cases, the attitude has its basis solely in acculturation and exists independently of the agent’s judgment. For this reason, acculturated desires are irrational (better: nonrational) in the same sense as appetitive and passionate desires. In fact, despite the inhibitions acquired in the course of a puritan up-bringing, a person may deem the pursuit of sexual pleasure to be worth

while, his judgment siding with the id rather than the super-ego. Acculturated attitudes may seem more akin to evaluation than to appetite in that they are often expressed in evaluative language ("divorce is wicked") and result in feelings of guilt when one's actions are not in conformity with them. But, since conflict is possible here, to want something as a result of acculturation is not thereby to value it, in the sense of 'to value' that we want to capture.

It is not easy to give a nontrivial account of the sense of 'to value' in question. In part, to value something is, in the appropriate circumstances, to want it, and to attribute a want for something to someone is to say that he is disposed to try to get it. So it will not be easy to draw this distinction in behavioral terms. Apparently the difference will have to do with the agent's attitude toward the various things he is disposed to try to get. We might say that an agent's values consist in those principles and ends which he—in a cool and non-self-deceptive moment—articulates as definitive of the good, fulfilling, and defensible life. That most people have articulable "conceptions of the good," coherent life-plans, *systems* of ends, and so on, is of course something of a fiction. Yet we all have more or less long-term aims and normative principles that we are willing to defend. It is such things as these that are to be identified with our values.

*The valuational system* of an agent is that set of considerations which, when combined with his factual beliefs (and probability estimates), yields judgments of the form: the thing for me to do in these circumstances, all things considered, is *a*. To ascribe free agency to a being presupposes it to be a being that makes judgments of this sort. To be this sort of being, one must assign values to alternative states of affairs, that is, rank them in terms of worth.

*The motivational system* of an agent is that set of considerations which move him to action. We identify his motivational system by identifying what motivates him. The possibility of unfree action consists in the fact that an agent's valuational system and motivational system may not completely coincide. Those systems harmonize to the extent that what determines the agent's all-things-considered judgments also determines his actions.

Now, to be sure, since to value is also to want, one's valuational and motivational systems must to a large extent overlap. If, in appropriate circumstances, one were never inclined to action by some alleged evaluation, the claim that that was indeed one's evaluation would be disconfirmed. Thus one's valuational system

must have some (considerable) grip upon one's motivational system. The problem is that there are motivational factors other than valuational ones. The free agent has the capacity to translate his values into action; his actions flow from his evaluational system.

One's evaluational system may be said to constitute one's standpoint, the point of view from which one judges the world. The important feature of one's evaluational system is that one cannot coherently dissociate oneself from it *in its entirety*. For to dissociate oneself from the ends and principles that constitute one's evaluational system is to disclaim or repudiate them, and any ends and principles so disclaimed (self-deception aside) cease to be constitutive of one's valuational system. One can dissociate oneself from one set of ends and principles only from the standpoint of another such set that one does not disclaim. In short, one cannot dissociate oneself from all normative judgments without forfeiting all standpoints and therewith one's identity as an agent.

Of course, it does not follow from the fact that one must assume some standpoint that one must have only one, nor that one's standpoint is completely determinate. There may be ultimate conflicts, irresolvable tensions, and things about which one simply does not know what to do or say. Some of these possibilities point to problems about the unity of the person. Here the extreme case is pathological. I am inclined to think that when the split is severe enough, to have more than one standpoint is to have none.

This distinction between wanting and valuing requires far fuller explication than it has received so far. Perhaps the foregoing remarks have at least shown *that* the distinction exists and is important, and have hinted at its nature. This distinction is important to the adherent of the familiar view—that talk about free action and free agency can be understood in terms of the idea of being able to get what one wants—because it gives sense to the claim that in unfree actions the agents do not get what they really or most want. This distinction gives sense to the contrast between free action and intentional action. Admittedly, further argument is required to show that such unfree agents are *unable* to get what they want; but the initial step toward this end has been taken.

At this point, it will be profitable to consider briefly a doctrine that is in many respects like that which I have been developing. The contrast will, I think, clarify the claims that have been advanced in the preceding pages.

## III

In an important and provocative article,<sup>9</sup> Harry Frankfurt has offered a description of what he takes to be the essential feature of “the concept of a person,” a feature which, he alleges, is also basic to an understanding of “freedom of the will.” This feature is the possession of higher-order volitions as well as first-order desires. Frankfurt construes the notion of a person’s will as “the notion of an *effective* desire—one that moves (or will or would move) a person all the way to action” (8). Someone has a second-order volition, then, when he wants “a certain desire to be his will.” (Frankfurt also considers the case of a second-order desire that is not a second-order volition, where one’s desire is simply to have a certain desire and not to act upon it. For example, a man may be curious to know what it is like to be addicted to drugs; he thus desires to desire heroin, but he may not desire his desire for heroin to be effective, to be his will. In fact, Frankfurt’s actual example is somewhat more special, for here the man’s desire is not simply to have a desire for heroin: he wants to have a desire for heroin which has a certain source, i.e., is addictive. He wants to know what it is like to *crave* heroin.) Someone is a *wanton* if he has no second-order volitions. Finally, “it is only because a person has volitions of the second order that he is capable both of enjoying and of lacking freedom of the will” (14).

Frankfurt’s thesis resembles the platonic view we have been unfolding insofar as it focuses upon “the structure of a person’s will” (6). I want to make a simple point about Frankfurt’s paper: namely that the “structural” feature to which Frankfurt appeals is not the fundamental feature for either free agency or personhood; it is simply insufficient to the task he wants it to perform.

One job that Frankfurt wishes to do with the distinction between lower and higher orders of desire is to give an account of the sense in which some wants may be said to be more truly the agent’s own than others (though in an obvious sense all are wants of the agent) the sense in which the agent “identifies” with one desire rather than another and the sense in which an agent may be unfree with respect to his own “will.” This enterprise is similar to our own. But we can see that the notion of “higher-order volition” is not really the fundamental notion for these purposes, by raising the question: Can’t one be a wanton, so to speak, with respect to one’s second-order desires and volitions?

<sup>9</sup> “Freedom of the Will and the Concept of a Person,” this JOURNAL, LXVIII, 1 (Jan. 14, 1971): 5–20.

In a case of conflict, Frankfurt would have us believe that what it is to identify with some desire rather than another is to have a volition concerning the former which is of higher order than any concerning the latter. That the first desire is given a special status over the second is due to its having an  $n$ -order volition concerning it, whereas the second desire has at most an  $(n - 1)$ -order volition concerning it. But why does one necessarily care about one's higher-order volitions? Since second-order volitions are themselves simply desires, to add them to the context of conflict is just to increase the number of contenders; it is not to give a special place to any of those in contention. The agent may not care which of the second-order desires win out. The same possibility arises at each higher order.

Quite aware of this difficulty, Frankfurt writes:

There is no theoretical limit to the length of the series of desires of higher and higher orders; nothing except 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 (16).

But he insists that

It is possible . . . to terminate such a series of acts [i.e., the formation of ever higher-order volitions] without cutting it off arbitrarily. When a person identifies himself *decisively* with one of his first-order desires, this commitment "resounds" throughout the potentially endless array of higher orders . . . The fact that his second-order volition to be moved by this desire is a decisive one means that there is no room for questions concerning the pertinence of volitions of higher orders . . . The decisiveness of the commitment he has made means that he has decided that no further question about his second-order volition, at any higher order, remains to be asked (16).

But either this reply is lame or it reveals that the notion of a higher-order volition is not the fundamental one. We wanted to know what prevents wantonness with regard to one's higher-order volitions. What gives these volitions any special relation to "one-self"? It is unhelpful to answer that one makes a "decisive commitment," where this just means that an interminable ascent to higher orders is not going to be permitted. This *is* arbitrary.

What this difficulty shows is that the notion of orders of desires or volitions does not do the work that Frankfurt wants it to do. It does not tell us why or how a particular want can have, among all of a person's "desires," the special property of being peculiarly

his "own." There may be something to the notions of acts of identification and of decisive commitment, but these are in any case different notions from that of a second- (or  $n$ -) order desire. And if these are the crucial notions, it is unclear why these acts of identification cannot be themselves of the first order—that is, identification with or commitment to courses of action (rather than with or to desires)—in which case, no ascent is necessary, and the notion of higher-order volitions becomes superfluous or at least secondary.

In fact, I think that such acts of "identification and commitment" (if one goes for this way of speaking) are generally to courses of action, that is, are first-order. Frankfurt's picture of practical judgment seems to be that of an agent with a given set of (first-order) desires concerning which he then forms second-order volitions. But this picture seems to be distorted. As I see it, agents frequently formulate values concerning alternatives they had not hitherto desired. Initially, they do not (or need not usually) ask themselves which of their desires they want to be effective in action; they ask themselves which course of action is most worth pursuing. The initial practical question is about courses of action and not about themselves.

Indeed, practical judgments are connected with "second-order volitions." For the same considerations that constitute one's on-balance reasons for doing some action,  $a$ , are reasons for wanting the "desire" to do  $a$  to be effective in action, and for wanting contrary desires to be ineffective. But in general, evaluations are prior and of the first order. The first-order desires that result from practical judgments generate second-order volitions because they have this special status; they do not have the special status that Frankfurt wants them to have because there is a higher-order desire concerning them.

Therefore, Frankfurt's position resembles the platonic conception in its focus upon the structure of the "soul."<sup>10</sup> But the two views draw their divisions differently; whereas Frankfurt divides the soul into higher and lower orders of desire, the distinction for Plato—and for my thesis—is among independent sources of motivation.<sup>11</sup>

<sup>10</sup> Frankfurt's idea of a wanton, suitably construed, can be put to further illuminating uses in moral psychology. It proves valuable, I think, in discussing the problematic phenomenon of psychopathy or sociopathy.

<sup>11</sup> Some very recent articles employ distinctions, for similar purposes, very like Frankfurt's and my own. See, for example, Richard C. Jeffrey, "Preferences among Preferences," this JOURNAL, LXXI, 13 (July 18, 1974): 377–391. In "Freedom and Desire," *Philosophical Review*, LXXXIII, 1 (January 1974): 32–54, Wright Neely appeals to higher-order desires, apparently unaware of Frankfurt's development of this concept.



## IV

In conclusion, it can now be seen that one worry that blocks the acceptance of the traditional view of freedom—and in turn, of compatibilism—is unfounded. To return to Berlin's question (206, above), it is false that determinism entails that all our actions and choices have the same status as those of "compulsive choosers" such as "kleptomaniacs, dipsomaniacs, and the like." What is distinctive about such compulsive behavior, I would argue, is that the desires and emotions in question are more or less radically independent of the evaluational systems of these agents. The compulsive character of a kleptomaniac's thievery has nothing at all to do with determinism. (His desires to steal may arise quite randomly.) Rather, it is because his desires express themselves independently of his evaluational judgments that we tend to think of his actions as unfree.

The truth, of course, is that God (traditionally conceived) is the only free agent, *sans phrase*. In the case of God, who is omnipotent and omniscient, there can be no disparity between valuational and motivational systems. The dependence of motivation upon evaluation is total, for there is but a single source of motivation: his presumably benign judgment.<sup>12</sup> In the case of the Brutes, as well, motivation has a single source: appetite and (perhaps) passion. The Brutes (or so we normally think) have no evaluational system. But human beings are only more or less free agents, typically less. They are free agents only in some respects. With regard to the appetites and passions, it is plain that in some situations the motivational systems of human beings exhibit an independence from their values which is inconsistent with free agency; that is to say, people are sometimes moved by their appetites and passions in conflict with their practical judgments.<sup>13</sup>

As Nietzsche said (probably with a rather different point in mind): "Man's belly is the reason why man does not easily take himself for a god."<sup>14</sup>

GARY WATSON

University of Pennsylvania

<sup>12</sup> God could not act *akratically*. In this respect, Socrates thought people were distinguishable from such a being only by ignorance and limited power.

<sup>13</sup> This possibility is a definitive feature of appetitive and passionate wants.

<sup>14</sup> *Beyond Good and Evil*, section 141.