4

Identification and Identity

J. David Velleman

When Harry Frankfurt chose a title for the first volume of his essays, he must have been thinking of the direction in which his work was going rather than the direction from which it had come. Retrospect would have led him to the titles of the founding essays in his research program, such as “Freedom of the Will and the Concept of a Person” or “Identification and Externality.” Instead he named the volume after the essay that set the theme for his future work, “The Importance of What We Care About.”[1] In the years since the publication of that volume, Frankfurt has explored many topics suggested by its wonderfully resonant title: how our caring about things makes them important to us; how the process of caring about them is important to us; and how important a matter it is which things we care about.

What I most admire about Frankfurt’s essays on these topics is their candor in reporting one man’s efforts to understand life as he finds it. Reading this work, one has the sense of receiving dispatches from an examined life. Frankfurt’s reflections on caring, in particular, are clearly an expression of what the author cares about, and as such they command a respect that transcends any disagreement.

Disagreement there is bound to be, however, when philosophy cuts so close to the bone. In this paper I am going to disagree with Frankfurt’s view on the last of the topics mentioned above, the importance of which things we care about. Which things we care about is important, according to Frankfurt, because our cares and concerns define our individual essences as persons: what we care about determines who we are. I don’t believe that we have motivational essences of this sort, though I agree that we sometimes seem to have them. I want to look for the source of this misleading appearance.

Frankfurt's New Conception of the Self

Ever since "Freedom of the Will and the Concept of a Person," Frankfurt has sought to draw a distinction among motives as internal or external to the self. The need for this distinction was first suggested to Frankfurt by cases in which an agent lacks autonomy because he is actuated by motives from which he is alienated. These motives seem to assail the agent from without and to compete with him for control of his behavior. Such cases suggest that being autonomous, or self-governed, is a matter of being governed from within—that is, by motives internal to the self. The question is what makes some motives internal in this sense.

Frankfurt's initial answer relied on the concept of identification. He suggested that motives are internal to the self when the subject identifies with them, by reflectively endorsing them as determinants of his behavior. An agent is autonomous, Frankfurt concluded, when he is actuated in ways that he reflectively endorses. This analysis of autonomy elicited a number of objections, which have been the subject of an extensive literature, leading to various revisions on Frankfurt's part.[2] More recently, however, Frankfurt has made a revision that is not obviously prompted by objections: it appears to express a further intuition about the boundaries of the self.

What Frankfurt now says about autonomy is this: "A person acts autonomously only when his volitions derive from the essential character of his will."[3] Frankfurt goes on to explain that inessential characteristics of a person's will are "separable" from it, and in that sense "external" to it, so that their governance of the person's behavior amounts to heteronomy rather than autonomy. Thus, Frankfurt still conceives of autonomy as governance by motives internal to the self, but he has adopted a new criterion of internality. Motives are internal to the self, according to the new criterion, when they are essential to the subject's volitional nature.

Frankfurt disavows what might be perceived as Kantian overtones in this statement. Kant would gladly join Frankfurt in saying that a person is autonomous when his behavior is determined by his essential nature. But what Kant would mean by this statement is that autonomy consists in being determined by practical reason, which places every agent under the same, universal laws.

Frankfurt explicitly rejects this Kantian reading of the relation between autonomy and personal essence. For he is loath to equate the self of self-governance with the anonymous faculty of practical reason:

> [T]his *pure will* is a very peculiar and unlikely place in which to locate an indispensable condition of individual autonomy. After all, its purity consists precisely in the fact that it is wholly untouched by any of the contingent personal features that make people distinctive and that characterize their specific identities. . . . The pure will has no individuality whatsoever. It is identical in everyone, and its volitions are everywhere exactly the same. In other words, the pure will is thoroughly *impersonal*. The commands that it issues are issued by no one in particular.[4]

In Frankfurt's view, the self whose governance constitutes self-governance, or autonomy, must be a thoroughly *personal* self: it must be someone in particular. Frankfurt therefore conceives of personal essences as comprising features of the very sort that Kant purified out of the will—that is, "contingent personal features that make people distinctive and that characterize their specific identities."

Such features, which are contingent to the nature of personhood, can still be essential to an individual person, in Frankfurt's view. "Even though a person's interests are contingent," Frankfurt says, "they can belong to the essential nature of his will."[5] A person can thus have a volitional essence that consists in perfectly idiosyncratic concerns.

What makes contingent interests essential to the nature of a person's will? The answer to this question is best developed in stages, through which an initial intuition is gradually modified.

Frankfurt's initial intuition is that the essence of an agent comprises what is volitionally necessary for him, just as the essence of a triangle comprises what is conceptually necessary for a triangle. Volitional necessity differs from conceptual necessity, however, in that it doesn't constrain how the person can be classified or described: "Volitional necessity constrains the person himself, by limiting the choices he can make."[6] It thus involves the inability to choose some things or to refrain from choosing others.

As Frankfurt goes on to note, however, such an inability may be due to an overwhelming aversion or compulsion of the sort that is alien to the self, constraining the will from without. If an inability is to be constitutive of the self, it ought to constrain the will from within and hence autonomously. Volitional necessity must therefore be a voluntary or

willing inability of the will. And Frankfurt believes that the will can indeed be subject to willing inabilities, such as the subject may express by calling an act unthinkable or saying that he cannot bring himself to perform it. To explain how an inability to will can be voluntary, Frankfurt modifies his initial intuition, by reintroducing the concept of identification. An inability becomes voluntary, Frankfurt explains, when it is due to a motive with which the subject identifies by means of a reflective endorsement.[7] If the motive's effectiveness in constraining his will is due in part to his reflective endorsement of it, then "the constraint is itself imposed by his will."[8]

Yet if the agent could potentially withhold his reflective endorsement from this constraint, then it isn't imposed by his will essentially. Hence more than reflective endorsement is required for the volitional necessities that define the subject's volitional essence. Frankfurt's initial intuition is therefore modified once again, to require not only that the agent endorse the motive constraining his will but that he be unable to help endorsing it. In such a case, the agent has a second-order inability: the inability to will any change in his inability to will. And for reasons already noted, this higher-order inability must itself be of the willing variety, by virtue of receiving a higher level endorsement, and so on. The subject's inability to alter his will thus appears to resound through higher and higher levels of the motivational hierarchy.[9] The subject finds that "it is not only unthinkable for him to perform the action in question; it is also unthinkable for him to form an effective intention to become willing to perform it."[10] Compound inabilities of this sort are what define the subject's essential nature, in Frankfurt's view.

Frankfurt describes these inabilities as "contingent volitional necessities by which the will of the person is as a matter of fact constrained."[11] They are contingent in the sense that they are not logically entailed by the subject's being a person or having a will. But they do have a quasi-conceptual consequence. By constraining the subject's will, they also define his essence as an individual, and so they give rise to a further constraint—namely, that he could not alter them while remaining the same person.

Frankfurt makes this point most clearly as follows:

> Agamemnon at Aulis is destroyed by an inescapable conflict between two equally defining elements in his own nature: his love for his daughter and his love for the army he commands.... When he is forced to sacrifice one of these, he is thereby

forced to betray himself. Rarely, if ever, do tragedies of this sort have sequels. Since the volitional unity of the tragic hero has been irreparably ruptured, there is a sense in which the person he had been no longer exists. Hence, there can be no continuation of *his* story.[12]

The necessitating concerns that make up Agamemnon's essence as a person, as Frankfurt conceives it, cannot be deduced from any of the generic concepts that apply to him, but they are necessary to his individual identity, to his being the particular person who he is. When Agamemnon sacrifices some of these concerns, he becomes a different person, and his former self ceases to exist.

Frankfurt makes the same point in several other ways. For example, he says that if someone is free of any volitional limits, then he has "no essential nature or identity,"[13] and he consequently suffers "a diminution, or even a dissolution, of the reality of the self."[14] A similar dissolution can be caused by boredom, which Frankfurt conceives as a lack of any compelling cares or interests. This state "threatens the extinction of the active self," and our dislike of it can be understood accordingly as an expression of our instinct for self-preservation—"in the sense of sustaining not the *life* of the organism but the *persistence of the self*."[15] Yet a third threat to the self, as conceived by Frankfurt, comes from ambivalence, which guarantees that one or another element of the self will have to be sacrificed when a choice is made.[16] If the self is to have a chance of remaining whole, it must be wholehearted, in the sense of being unequivocal in its essential concerns.

Because these concerns can be sacrificed only at the cost of the self, in Frankfurt's view, they "possess not simply power but authority,"[17] derived from the imperative of self-preservation. The subject has compelling reason not to oppose them, because of the "drastic psychic injuries" that such opposition would entail.[18] What someone cares about is thus important because, by defining who he is, it determines what he must do in order for that person to survive.

These recent developments in Frankfurt's conception of autonomy have, in effect, yoked it to a conception of personal identity, which also involves the boundaries of the self.[19] And Frankfurt's conception of personal identity agrees with the currently prevailing, neo-Lockean conception propounded by Derek Parfit. In particular, Frankfurt's conception of identity agrees with Parfit's in respects that, in turn, distinguish Parfit's

conception from Locke's. Since the differences between Parfit and Locke have not been widely discussed, I want to pause a moment to review them.

Parfit follows Locke in thinking that what makes for the survival of a person is his psychological relation to past and future selves.[20] But Parfit differs from Locke on the nature of the relevant relation, and he thereby makes room for motivational essences of the sort that Frankfurt envisions.

Locke thinks that the psychological relation making for a person's survival is exclusively a relation of memory: one's past selves are those whose experiences one remembers first-personally, and one's future selves are those who will first-personally remember one's experiences. A natural extension of Locke's theory applies this definition recursively, so that one's past and future selves include not only those who are linked to one by memory but also anyone similarly linked to them, and so on. In either version, Locke's theory implies that one may share virtually no motivational characteristics with one's past or future selves. One may in the past have possessed vastly different attitudes and traits of character, so long as one remembers being the person who possessed them (or being someone who remembers being that person, and so on); and one may yet have vastly different attitudes and traits in the future, so long as the person possessing them remembers being oneself (or being someone who remembers being oneself, and so on). Not only can a prince and a cobbler end up inhabiting one another's bodies, according to Locke's theory; they can also end up possessing one another's beliefs, desires, ideals, loves, projects, and so on.

Parfit thinks that some of the latter characteristics are in fact relevant to survival.[21] Part of what makes for one's survival, in Parfit's view, is the persistence of attitudes and traits of character.[22] And for this purpose, some attitudes and traits are more important than others. Characteristics that differentiate one from other people tend to be more important to one's survival, as Parfit conceives it, than those that one shares with everyone else;[23] characteristics that one values in oneself may also be more important than those that one wishes to shed.[24] Parfit's theory thus allows for the possibility that some cluster of desires and intentions might be so distinctive of a person, and so valued by him, that he would be as good as dead without them. Sacrificing these attitudes would be the

end of him, and so they would derive authority from the imperative of self-preservation, just as Frankfurt believes.[25]

The recent developments in Frankfurt's conception of autonomy have thus brought it into harmony with the distinctively un-Lockean strain in Parfit's otherwise Lockean conception of personal identity. Like Parfit, but unlike Locke, Frankfurt believes that one may have to retain particular motives in order to remain oneself. And he believes that motives essential to the self in this sense are the motives whose governance of one's behavior constitutes one's self-governance, or autonomy.

A Critique of Frankfurt's New Conception

Even if I believed that a person had a motivational essence of this kind, I would not infer that his being governed by this essence was what made him autonomous. Being governed by such an essence might amount to authenticity, perhaps, but not autonomy.

To see the difference, consider the paradigm case of inauthenticity, the person who manifests what D. W. Winnicott called a "False Self."[26] This person laughs at what he thinks he is supposed to find amusing, shows concern for what he thinks he is supposed to care about, and in general conforms himself to the demands and expectations of others. The motives that his behavior is designed to simulate are motives that he doesn't genuinely have. And the overriding motive that he does have—namely, to satisfy the expectations of others—is hardly a motive that he cannot help endorsing; on the contrary, he doesn't even acknowledge this motive, much less endorse it. Hence neither the motives that he simulates nor the motive on which he thereby acts belong to his essential nature, as Frankfurt conceives it.

But is this person lacking in self-control, self-governance, or autonomy? To be sure, he has a problem with autonomy, but his problem is one of excess: he is overly self-controlled, overly deliberate; his grip on the reins of his behavior is too tight, not too loose. His failure to be motivated from within his true self makes him inauthentic, but it seems to result from his being all too autonomous.[27]

So if I believed in a person's motivational essence, I still wouldn't identify it as the source of autonomy; but I don't believe in a motivational essence. I am inclined to say of the essential self posited by

Frankfurt what the psychoanalyst Jeffrey Rubin has recently said of the True Self posited by Winnicott:[28]

> The process of searching for one's True Self, regarded as a singular entity waiting to be found, is a quixotic enterprise that may promote self-restriction and self-alienation.... [S]ingular notions like the True Self subjugate selfhood's possibilities by obscuring and limiting its multidimensionality. Facets of self-experience that do not fit into preexisting images of who one *really* is are neglected or not assimilated into one's sense of identity.... Not only is a monolithic sense of self limiting, but psychological health may involve access to, and comfort with, our multidimensionality. From this perspective, a sense of the complexity, multidimensionality, and polyvalency of the self is a developmental milestone and achievement.

With these words in mind, I turn to a critique of Frankfurt's new conception of the self.

I have argued elsewhere against the notion that there must be motivational constancy between a person and his past or future selves. Specifically, I've argued that a person has past and future selves in virtue of psychological connections that give him first-personal access to past and future points-of-view—connections that can be forged by memory and anticipation but not by the retention of motives or traits of character.[29]

I don't deny that some of our concerns are authoritative for us because they are somehow central to our personalities; nor do I deny the temptation to describe these concerns as essential to who we are, integral to ourselves, definitive of our identities, and so on. At the end of this paper, I'll explain how I think we should understand these descriptions. For now, I want to argue only that we shouldn't take them literally by claiming that the defense of our central concerns is a matter of self-preservation, a matter of life or death for the self. Such literalism can easily lead to absurdity.

Consider this passage, which is sometimes cited with approval by critics of impartial morality:

> I am not a person who just happens accidentally and irrelevantly, to be a man, forty years old, the husband of a professor of English literature, the son of two aging and sick parents, the father of two small boys six and four, a comfortably well-off member of the upper middle class, American-Jewish, born and raised in New York. I am *essentially* such a man.[30]

Can a man be essentially forty years old, essentially the father of young children, or essentially the son of elderly parents? One cannot read this

statement without wondering whether the author still believes it, now that he is in his sixties, his parents have passed away, and his children are grown.

Of course, Frankfurt would not include most of these attributes in a person's essential nature. But if we read this enumeration of attributes as an expression of the associated concerns—of the author's love for his wife and children; pity for his ailing parents; pride in his masculinity, his American-Jewish heritage, or his hometown of New York—then we arrive at a personal essence that Frankfurt might recognize. And we can still wonder whether the author would be a different person simply because of having become estranged from his parents or having fallen for someone other than his wife.

According to Frankfurt, such crises may have no sequels, because their protagonist ceases to exist. But surely estrangements and betrayals are precisely what set the stage for a sequel. The Agamemnon legend would lose much of its power if the man who sacrificed his daughter at the beginning didn't survive to be murdered by his wife in the end. If a crisis like the one at Aulis necessarily put an end to its protagonist, we wouldn't just be lacking sequels to particular stories: we wouldn't have the concept of a sequel at all.

In light of how implausible the notion of personal essences can be when applied to particular cases, we have to wonder why it remains so attractive. Frankfurt never offers us a convincing example of someone who ceases to be the same person because of abandoning a project, betraying a commitment, or undergoing some other change of heart; nor does he offer any argument for thinking that motivational changes can have such momentous results. He simply asserts that our projects and commitments are sometimes essential to who we are. We welcome his assertion, but not as something of which he has convinced us; we welcome it as something that we, too, want to say about ourselves. The question is why we want to say it. In the absence of examples or arguments to show that we have motivational essences, what moves us to apply this self-description?

My worry is that believing ourselves to have motivational essences is a case of wishful thinking on our part. We'd like to have motivational essences, and so we're happy to agree when someone says that we do.

Now, a conception of the self cannot be faulted simply for being associated with wishes. Any conception of how we are constituted will yield implications about what it is for us to be well constituted. A conception of the self will thus entail an ideal of the self, to which holders of the conception will naturally aspire. What's crucial, however, is the logical order between conception and aspiration. We are justified in wishing to embody an ideal implicit in our self-conception; but our self-conception should not be tailored to suit our antecedent wishes. My worry is that Frankfurt's conception of the self appeals to us only because its implicit ideal represents us as we wish we could be.

The ideal implicit in Frankfurt's conception of the self is the ideal of wholeheartedness. Frankfurt reasons that if the self is constituted out of irresistible motives, then it had better be constituted out of motives that are in concert rather than conflict, so that it will not be divided against itself. He therefore concludes the well-constituted self is wholehearted rather than ambivalent.

Frankfurt's term "wholeheartedness" does not denote the complete absence of conflicting motives. A person can be wholehearted in Frankfurt's sense while retaining desires that conflict, so long as he has decisively identified with one of the desires and dissociated himself from the other. This process "involves a radical *separation* of the competing desires, one of which is not merely assigned a relatively less favored position but extruded entirely as an outlaw."[31] The motivational conflict is not thereby eliminated. Rather, "the conflict between the *desires* is in this way transformed into a conflict between *one* of them and the *person* who has identified himself with its rival."[32]

Frankfurt compares and contrasts this process with "the self-reparative activities of the body":

> When the body heals itself, it *eliminates* conflicts in which one physical process (say, infection) interferes with others and undermines the homeostasis, or equilibrium, in which health consists. A person who makes up his mind also seeks thereby to overcome or to supersede a condition of inner division and to make himself into an integrated whole. But he may accomplish this without actually eliminating the desires that conflict with those on which he has decided, as long as he dissociates himself from them.[33]

Thus, ambivalence is a disease of the self, to which wholeheartedness stands as the contrasting state of health. What cures the disease, and

restores us to health, is the process of dissociating ourselves from unwelcome desires, a process that expels them from the self without necessarily eliminating them entirely.

This prescription for self-health is undeniably attractive. The question is whether it attracts us by articulating what would in fact be ideal for us, given how we are constituted. I suspect that it attracts us for other reasons.

Frankfurt is not alone in thinking of ambivalence as a disease of the self. One of the most famous discussions of ambivalence casts it literally as an agent of disease—specifically, the mental illness suffered by the patient of Freud's who has come to be known as the Rat Man.[34]

Freud diagnoses the Rat Man's problem as "a splitting of the personality"[35] resulting from "a battle between love and hate [that] was raging in [his] breast"[36] with respect to one and the same person. The Rat Man desperately loved and violently hated his father, and his personality was consequently divided, according to Freud, into distinct loving and hating selves.[37] Freud cites this division to explain the Rat Man's symptoms, which often involved repeatedly doing and undoing an action, or thinking and contradicting a thought.

At first glance, then, Freud seems to agree that ambivalence is a disease of the self, a disease whose cure requires the attainment of wholeheartedness. A second look reveals, however, that the Rat Man's problem was not so much ambivalence as his response to it. What caused the Rat Man's neurosis, according to Freud, was the means by which he sought to cope with the battle between love and hatred within him—namely, by repressing his hatred and acknowledging only his love. This repression is what allowed the two emotions to survive unmixed and hence to continue pulling the patient so violently in opposite directions. Freud concludes, "We may regard the repression of his infantile hatred of his father as the event which brought his whole subsequent career under the dominion of the neurosis."[38]

The Rat Man's strategies of repression were not the ones with which we are familiar from Freud's more accessible writings. Most of the Rat Man's thoughts and feelings, both loving and hostile, were available to his consciousness; he simply disconnected them and reconnected them in such a way as to conceal their true significance.[39] Thus, for example,

he frequently had thoughts of harm befalling his father, but he had disconnected those thoughts from their wishful affect. He insisted that they were merely "trains of thought" rather than hostile wishes.[40] Conversely, his hostile feelings were displaced from their true objects onto others, including his psychoanalyst[41] and himself.[42]

The Rat Man's repression thus consisted in a concerted practice of self-misinterpretation. And what motivated this misinterpretation was precisely the desire to dissociate himself from his own hatred and hostility. Thus, Freud tells us that on the occasion when the Rat Man first divulged the hostile thought that became the centerpiece of his case history (and the source of his analytic moniker), "He broke off his story in order to assure me that these thoughts were entirely foreign and repugnant to him."[43] On another occasion, the Rat Man said "that he would like to speak of a criminal act, whose author he did not recognize as himself, though he quite clearly recollected committing it."[44] His hatred was thus something that he had alienated from himself, so that he no longer regarded its resultant thoughts and actions as his.

We might even say that the Rat Man's hatred had been repressed by being "extruded as an outlaw"—but then we would be quoting Frankfurt rather than Freud.[45] Conversely, Freud's discussion of this case begins with a statement that might easily have been written by Frankfurt. Referring to an erotic wish felt by the Rat Man in childhood, Freud says: "This wish corresponds to the later obsessional or compulsive idea; and if the quality of compulsion was not yet present in the wish, this was because the ego had not yet placed itself in complete opposition to it and did not yet regard it as something foreign to itself."[46] The theory expressed in this statement is that a wish becomes a compulsion when the ego comes to regard it as foreign—which is close to what Frankfurt believes as well.

Unfortunately, this point of agreement between Freud and Frankfurt suggests that the Rat Man suffered, not from the disease of ambivalence, but from something like Frankfurt's cure. What made him ill was his effort to dissociate himself from one of his emotions, which is just what Frankfurt prescribes for cases of ambivalence. The "radical separation of . . . competing desires" recommended by Frankfurt ultimately led to the "splitting of the personality" diagnosed by Freud.[47]

Of course, Frankfurt does not recommend separating desires by repressing some of them. Although Frankfurt's view implies that the Rat Man was right to expel his hatred, it also implies that he was wrong about where to expel it from. He expelled it from his self-awareness or self-understanding; whereas Frankfurt's view implies that he should have consciously rejected it and thereby expelled it from the self.[48]

Yet the suspicion remains that this prescription, though different from what caused the Rat Man's illness, would hardly have been more healthy. Surely, what the Rat Man should have done was to accept his filial hostility as part of himself, to accept *himself* as ambivalent toward his father.[49] The Rat Man's mistake was indeed his attempt to separate competing desires by expelling one of them, not the specific form of expulsion by which he tried to separate them.

We can draw this moral from the Rat Man's story without relying on any distinctively Freudian hypotheses—childhood sexuality, the Oedipus complex, or even repression. Beneath these overlays of psychoanalytic theory is a story intelligible to pure common sense.

To begin with, the species of ambivalence attributed to the Rat Man is familiar to all of us. Almost all of us love our parents, but most of us also retain sources of deep hostility toward them—sore spots that can be inflamed into powerful anger or even hate. Other elements of the Rat Man's history may seem weird or incredible, but the element of filial ambivalence is not extraordinary in the least.

Second, beneath the theoretical apparatus of Freud's account lies a piece of folk wisdom about dealing with mixed emotions. When we are angry with someone we love, the first step toward dealing with our anger is to let it mingle with, and be modified by, our other emotions toward the same person. Isolating our hostility from our other feelings is a way of not dealing with it, of allowing it to remain undigested, a lasting source of inner strife and outer impulsiveness. Of course, new-age common sense about the importance of "processing" problematic emotions is derivative of Freudian theory; but Freudian theory about the return of the repressed is, in turn, derivative of a common sense that is ageless, as Freud himself was the first to point out. When Freud explains why repression brought the Rat Man "under the dominion of the neurosis," his explanation strongly implies, on commonsensical grounds, that any

attempt by the Rat Man to segregate his emotions would have been equally harmful.

A third piece of common sense in this case history is that allowing our emotions to mingle with their opposites is difficult, daunting, even terrifying. The Rat Man chose to regard his hatred as foreign because he was afraid of letting it into his emotional life, even though doing so was his only chance of domesticating it. All of us are like the Rat Man at least to this extent, that we feel threatened by various emotions that would introduce conflict into our lives. We consequently wish that our commitments were not tinged with regret, that our projects were not fraught with doubts, that our loves were not complicated by hate. We wish, in short, that we could be wholehearted.

What has now emerged is that wholeheartedness is an object of wishes that do not necessarily represent a healthy trend in our thought. Our attraction to the idea of being wholehearted is one manifestation of the fears that move us to defend ourselves against our own emotions. Hence our affinity for Frankfurt's ideal may not indicate that he's right about the constitution of the self; it may indicate no more than our own defensiveness.

Conceiving of ourselves to have motivational essences can serve the same defensive purpose. What's threatening about our hostility toward loved ones is that it might efface our love for them and move us to do things that love would never allow us to do.[50] Our love therefore entails a fear of its own obliteration. How comforting it would be to think that our love was indelibly written into our nature, so that we didn't have to protect it from exposure to contrary feelings. Are we attracted to this thought because it is true or because it is comforting?

A similar question can be asked about the very notion of an inner self. When we defend ourselves against unwelcome emotions, we would like to think that we are expelling, excluding, or (in Frankfurt's term) "extruding" them from ourselves. We are trying to neutralize troublesome elements of our psyches, and one way to neutralize troublemakers is to banish them beyond some enforceable boundary. When we picture the inner sanctuary of the self, we are picturing a defensible territory—which is precisely what's needed for successful defenses. Given our wish

for this safe haven, however, our belief in its existence may be another case of wishful thinking.

Indeed, there is also a defensive way of applying Frankfurt's term "identification." I'm sure that Frankfurt didn't intend the term to be applied in this way, but I wonder whether we as readers haven't departed from his intentions. The term "identification" has an ordinary meaning, different from Frankfurt's, whose substitution into our interpretation of Frankfurt's theory would further suit our defensive purposes.

A Digression on Identification

Frankfurt is responsible for bringing the term "identification" into widespread use among contemporary philosophers and for shaping their intuitions about it. Following Frankfurt, philosophers have come to speak of identification as primarily reflective and evaluative—as a process of endorsing some parts or aspects of ourselves.[51] But the term "identification" ordinarily stands for a process that is not, in the first instance, either reflective or evaluative.[52] I want to examine what we ordinarily mean by "identification" and then to consider the possibility that we can be misled by that meaning in our reading of Frankfurt.

In ordinary parlance we are more likely to speak of identifying with other people than with parts of ourselves. The remark "I can identify with that" is a way of saying that we have experienced what someone else is going through and that we empathize with his reaction to it. We speak of identifying with fictional characters or with their actions and reactions in particular scenes.[53] We also describe ourselves as identifying with authority figures and role models in our lives. Identifying is thus something that we do, in the first instance, with people other than ourselves.

In the case of role models, identification involves a positive evaluation: identifying with these people goes hand-in-hand with admiring them and wanting to emulate them. But in other cases identification can be evaluatively neutral or even negative. "I can identify with that" may be our response to someone's self-depreciating tale of ineptitude or weakness. What elicits our identification may be another person's rendition of that which we find most disappointing or embarrassing in ourselves. Fiction

would lack much of its educative force if it couldn't induce us to identify with characters whom we don't admire or wish to emulate. Our identification with these characters may soften our judgment of them, but only because it makes us empathize with them, not because it involves any judgment in their favor.

Perhaps the most extreme example of identification without positive evaluation is the phenomenon that Anna Freud called "identification with the aggressor."[54] When a child plays at being a hungry lion or an angry teacher, he may be identifying with what he most fears, so as to escape from being the target of its aggression. The same mechanism may be at work when an adult directs at others the very sort of criticism to which he feels most vulnerable himself. This person doesn't necessarily admire his critics or want to be like them; he would just prefer being the critic to being the object of criticism.

If identification doesn't necessarily involve a favorable evaluation, then what do instances of identification have in common? The nature of identification can only be obscured, I think, by a moral psychology confined to the categories of belief and desire, or belief and pro-attitude. If identification must sit with either the beliefs or the pro-attitudes, it will sit more comfortably with the latter, in the form of an endorsement or a desire to emulate. Yet to insist on placing identification in the matrix of belief and pro-attitude is to miss the fact that it involves, above all, an exercise of the imagination.[55]

To identify with someone, you have to imagine that you are he, or that he is you. Such an exercise of imagination isn't sufficient for identification: you can imagine that you are Caligula or Lady Macbeth even though you can't identify with them. But identifying with someone can be characterized, I think, as a particular way of imagining that you are he, with particular psychological consequences.

One respect in which identification exceeds merely imagining that you are someone else is that it must be spontaneous. You can deliberately conjure up the thought of being Lady Macbeth, but unless you were spontaneously affected by that thought, you wouldn't be said to identify with her. Another difference is that when you deliberately conjure up the thought of being Lady Macbeth, it occupies the focus of your awareness, while knowledge of your real identity is pushed into the background.

When you identify with someone, the position is reversed: the thought of being that person is in the background—perhaps so far in the background as to be unconscious—while actual identities remain salient.

Finally, deliberately imagining that you are someone else doesn't necessarily affect your realistic attitudes—that is, your attitudes toward the world as you believe it to be rather than as you have imagined it. The activity of imagining that you are Caligula may leave no traces on your thoughts and feelings about the real world, including yourself and the historical figure of Caligula. But when you identify with someone, the thought of being that person, though outside the focus of awareness, somehow colors your attitudes toward yourself or him, toward your individual situations or shared relationship. Your attitudes toward the actual world are modified by your having spontaneously though perhaps unconsciously imagined a world in which you are he.

The most common way to imagine that you are someone else is to imagine being that person, by imagining the world as experienced by him—as seen through his eyes and traveled in his shoes.[56] This sort of imagining is also the most common means of identification. Identifying with someone is usually a matter of having your view of reality colored by a spontaneous image of how things are for him.

This image is of necessity incomplete, in that it doesn't represent every facet of the other person's perspective. Sometimes it can be so incomplete as to represent no more than a single sensation. Watching a sweaty jogger lift a drink to his lips, you may suddenly imagine a cool draught in your own throat; watching a couple pause to embrace on the street, you may spontaneously imagine a warm breath against your own face; seeing someone catch his finger in a car door, you may imagine a shooting pain in your own finger. These brief and fragmentary identifications have only an ephemeral impact on your view of the world, but they exemplify the same psychological process as identifications that are more consequential.[57]

One consequence of imagining things from someone else's perspective is a tendency to empathize with him. Picturing the world as seen through his eyes or heard through his ears, you feel first-personal emotions on his behalf.[58] A further consequence of identification is insight into the other person's thinking and behavior. You are better able to anticipate

the thoughts and actions of someone with whom you identify, because you have imaginatively simulated his situation, either consciously or unconsciously.

Another consequence of identifying with someone is a tendency to behave like him, partly because of empathizing with him but also because of the direct motivational force of the imagination. On the one hand, you tend to do what the other person does or would do because you feel the way he does or would feel. Loving his friends, you tend to favor them; hating his enemies, you tend to oppose them; and so on. On the other hand, you tend to pick up the person's behavioral style—his accent, his idiom, and his body-language as well—as if you were impersonating him.

The latter mechanism is similar to that of deliberate pretending or make-believe. When you played make-believe as a child, you did not just copy the behavior of the character or creature you were pretending to be; you imagined being that character or creature, and your imagination moved you to behave accordingly.[59] Similarly, when you identify with someone, the image of being that person leads to move as if inside his body, to speak as if with his voice, even to think and feel as if with his sensibilities. Your identification with him may thereby become recognizable to observers, if they can detect these echoes of his behavior in yours.

What would happen if we interpreted Frankfurt's term "identification" as referring to this phenomenon, identification as ordinarily understood? Frankfurt's claim that we identify with some of our attitudes would then seem to describe us as imagining ourselves to *be* those attitudes. To identify with a desire or emotion would be to imagine being the desire or emotion. But how could we imagine being one of our own mental states, a proper part of ourselves?[60] How could we identify, in this sense, with something that isn't a whole person?

Come to think of it, there is a famous description of just this process. It goes like this:

> Let us consider this waiter in the café. His movement is quick and forward, a little too precise, a little too rapid. He comes toward the patrons with a step a little too quick. He bends forward a little too eagerly; his voice, his eyes express an interest a little too solicitous for the order of the customer.... All his behavior seems to us a game. He applies himself to chaining his movements as if they were mechanisms, the one regulating the other; his gestures and even his voice seem to

be mechanisms; he gives himself the quickness and pitiless rapidity of things. He is playing, he is amusing himself. But what is he playing? We need not watch long before we can explain it: he is playing *at being* a waiter in a café.[61]

When Sartre says that this waiter is playing at being a waiter, he means that the waiter is playing at being less than a whole person—a waiter-on-tables and nothing more. The waiter imagines that he is nothing but the nexus of motives and skills exercised in his waiting on tables; that he is, not a person choosing to exercise those motives and skills, but a mechanism wholly composed of them; that he is, so to speak, a waiting machine. The waiter thus identifies with a proper part of himself.

This waiter is Sartre's prime specimen of bad faith, which is a mode of defensive thinking. Of course, the defenses diagnosed by Sartre are fueled by a different anxiety from the defenses diagnosed by Freud: they are fueled by a fear of our radical freedom rather than our threatening emotions. But the strategy of defense described by Sartre is available to the latter anxiety as well. If we are afraid of hating our parents, we can imagine being identical with our love for them—parent-lovers and nothing more. We can imagine shrinking to occupy the loving aspect of our personalities, just as the waiter imagines shrinking to occupy his waiterly motives and skills. Once we have imaginatively retreated to within the boundaries of our love, we can hope to keep our hatred at bay.

When Frankfurt describes us as identifying with some of our motives and alienating others, his description rings true, I suspect, because it accurately describes this common defensive fantasy. We do indeed identify with some of our motives, but we thereby engage not in self-definition but self-deception. We identify with some of our motives by imagining ourselves as *being* those motives, to the exclusion of whatever might complicate or conflict with them.

To repeat, I do not believe that Frankfurt uses the term "identification" in this ordinary sense. I think that he initially introduced the term by stipulation, with the intention that it would carry a new, philosophical meaning. "Identification" was meant to describe the psychological process by which a person empowers some of his motives to implicate him in causing behavior, so that whatever they motivate will be attributable to him, as his doing. The attention that Frankfurt has drawn to the process of identification, so defined, has greatly advanced the philosophy of

action over the past twenty years. But I think that Frankfurt's own notion of identification has turned out to involve some assumptions that ought to be reexamined, in light of the conclusions at which he has now arrived.

Frankfurt assumed from the outset that identification works by incorporating motives into something called the self, so that behavior governed by those motives qualifies as self-governed, or autonomous. This assumption is harmless, I think, so long as it leaves open the sense in which the term "self" is being used. The reason some behavior counts as autonomous action, attributable to the subject as his doing, is that it is governed by motives constitutive of something deserving to be called the self in some sense of the word. What has emerged in Frankfurt's recent work, however, is the further assumption that the sense of the word "self" used in an account of autonomy will be the same one that is used in accounts of other phenomena that merit philosophical attention.

Thus, Frankfurt conceives of the self as an inner core or kernel comprising that *in* the person which really *is* the person and whose impact on the world is therefore his. The self so conceived underlies not just autonomy but personal identity as well. It is not just that part of a person whose participation in causing behavior is necessary and sufficient for his participating; it's that part of a person whose existence is necessary and sufficient for his existing. Indeed, it is the former precisely because it is the latter—the source of the person's autonomy because it is the basis of his identity, causing what he can be said to cause by virtue of being what he is.

Similarly, Frankfurt conceives of the self as that to which a person must be true in order to be true to himself, or that which he must not betray lest he be guilty of self-betrayal. Motives constitutive of the self therefore carry a special authority in the subject's practical reasoning. They exercise not only the force of his autonomy but also the claim of his self-worth, and for the same reason—namely, that they constitute what he is.

An Alternative Conception of the Self

I don't believe in the self, so conceived. That is, I don't believe that a person has a proper part that is both the source of his autonomy and

the target of his self-regard because of being the basis of his identity. Expounding my own views is not the purpose of this paper, but I want to state them briefly in order to illustrate that believing in the self is optional.[62]

In my view, "self" is just a word used to express reflexivity—that is, the coincidence of object and subject, either of a verb or of the activity that it represents. ("She accidentally cut herself.") In many philosophical contexts, "self" expresses the reflexivity of representations, especially their notional reflexivity, the property they possess when they represent their object *as* their subject. ("He's always talking about himself.") In this sense, "self" is used to report indirectly a thought or utterance that originally contained a first-person pronoun. We use "self" to report a thought or utterance containing "I" just as we use "present" to report a thought or utterance containing "now."

As a word expressing reflexivity, "self" has various uses in various contexts, including several contexts that are of interest to philosophy. "Self" can express the reflexivity of the control that an autonomous agent exerts over his own behavior; the reflexivity of the memories and anticipations that link a temporally extended person to his past and future; or the reflexivity of any first-personal attitudes that he may hold. Although "self" expresses reflexivity in each of these contexts, there is no single entity to which it refers in all of them. We shouldn't assume, in other words, that there is something called The Self that governs a person's behavior when it is self-governed, persists so long as the person remains himself, and is the object of his self-concept or self-image.

I want to explain briefly how I understand the term "self" in these philosophical contexts. I begin with the context of personal identity.

I think that a person's past or future selves are just the past or future persons whom he can pick out with thoughts that are notionally reflexive, or first-personal.[63] There is no kernel or core whose presence in past or future persons makes them selves of his; there are only the psychological connections that mediate his reflexive references to them, thus enabling him to think of them first-personally. Locke was right to name experiential memory as the psychological medium connecting a person to his past selves, because replaying past experiences is how the person naturally and without contrivance thinks of past individuals as "me." I

would merely add that there are experiential forms of anticipation that can mediate first-personal reference to future persons as "me," thus linking the subject to his future selves.

This conception of selfhood implies that a philosophical theory of the self should have as little substance as a philosophical theory of the present. We can theorize about the reflexive aspect of things, just as we can theorize about their present aspect, but we must avoid reifying the present or the self.[64]

If a person's relation to past and future selves doesn't depend on a shared subset of attributes and attitudes, then it doesn't depend on anything that might be the object of his self-regard. The self for whom the person may have esteem, and with whom he can keep or break faith, is not an inner core of traits and states that he must retain in order to remain himself.

In this latter usage, I think, the term "self" refers—not to the person, or a part of the person, represented reflexively—but to the person's own reflexive representations, which make up his self-image or self-conception.[65] This sense of the word "self" crops up frequently in the field known as self-psychology, where it is often paired with a corresponding sense of the word "identity." When someone suffers an identity crisis, as we call it, what is threatened is not his identity as a person but his conception of himself as a person, which might also be called his sense of identity or his sense of who he is.

In this context, I am happy to say that particular cares and concerns can be definitive of a person's identity or essential to the self.[66] That he has these motives may be a fundamental, organizing principle of a person's self-understanding, without which the rest of his self-image would no longer cohere. If he had to stop thinking of himself as having these motives, he would temporarily lack any coherent conception of himself as a person, and so he might be described as no longer knowing who he was. But the fact that jettisoning the representation of these motives from his current self-image would result, temporarily, in his no longer knowing who he was—this fact doesn't mean that jettisoning the motives themselves would result in his never again being who he is. These motives are essential to his self, or identity, in the sense that refers to his self-conception, which can in time be revised or replaced if his actual motives should change.[67]

This abstract distinction may be clarified by an example. A philosopher recently told me that when he discusses same-sex marriage in his Introduction to Political Philosophy, the fundamentalist Christians in the class find the subject threatening to their identities.[68] Frankfurt might explain this phenomenon by pointing out that the doctrinal commitments of these students involve various volitional necessities, such as an inability to condone homosexuality or even to wish that they could condone it, or a similarly structured inability to question the dictates of scripture. If the students allow these essential aspects of their natures to change, they would bring their current selves to an end—a "drastic psychic injury," in Frankfurt's view.[69] Frankfurt's view thus seems to imply that these students are justified to resist any change of mind on the issue, on grounds of self-preservation.

I would say that a commitment to religious doctrines is essential to these students' identities only in the sense that it is central to their self-conceptions. They think of themselves as Christians first and foremost, and much else that they believe about themselves is based on this premise. If they had to question their faith in the doctrines that they regard as essential to Christianity, they would have to question most of what they currently believe about themselves. Changing their minds on doctrinal matters is therefore threatening to their identities because it threatens to enforce a major revision in their self-conceptions. They would still be themselves after changing their minds, but they would have temporarily lost their grasp of who they are. Some resistance to such radical change may well be justified, but not as much as would be justified for the sake of literal self-preservation.

I have now explained what an aspectual interpretation of "self" implies for discourse about personal identity and self-regard. In discourse about personal identity, "self" refers to those past or future persons whom the subject can denote reflexively, as "me"; in discourse about self-regard, it denotes the subject's reflexive representation, his self-concept or self-image. I turn, finally, to discourse about autonomy, or self-governance. My aspectual interpretation of "self" doesn't require me to deny that a person has a source of autonomy that might be called his essential self: the source of a person's autonomy can be his essential self in an aspectual sense.[70]

Suppose that a person has a part that he is unable to regard nonreflexively, a part on which he cannot attain a truly detached, third-personal perspective. That part of him will be essentially "self" to him, in the sense that it is inalienably "me" from his perspective. Its being his essential self won't mean that it is essential to his identity; only that it always presents a reflexive aspect to his thinking.

Maybe an analogy would help. Consider that spot, right between your eyes, which is at the origin of your visual perspective—the vertex of all the angles that your visual images subtend. That spot is your visual location, or visual standpoint, in the sense that you always see things as projected onto that point. Of course, you can look at yourself in the mirror and refer to the relevant point in space as "over there," in the mirror. But even when you look at it "over there," you are still looking at it *from* that point, and so it remains "back here" as well.

Now, is your visual standpoint an essential part of your visual apparatus? No. Indeed, it isn't a part of your visual apparatus at all. It's just a part of you that always presents a particular aspect to you—the aspect of being visually "here," at the geometric origin of your visual perspective.[71] Surely, we would be making a mistake to regard this point as the origin of your vision in any other sense.

If there is a part of your personality with which you necessarily think about things, then it will be your mental standpoint, always presenting a reflexive aspect to your thought. You will be able to think about this part of your personality as "it," but only from a perspective in which it continues to function as the thinking "I"—just as you can find a reflection of your visual location "over there" only from a perspective in which it is also "back here."

I believe that this phenomenon is what Aristotle had in mind when he said that "each person seems to be his understanding."[72] A person can never conceive of his own conceptual capacity from a purely third-personal perspective, because he can conceive of it only *with* that capacity, and hence from a perspective in which it continues to occupy first-person position. Just as the person cannot attain a visual perspective from which the point between his eyes isn't "here," so he cannot attain a cognitive perspective from which his understanding isn't "I." That's why the person seems to be his understanding.

This Aristotelian observation does not imply that a person's understanding is his essence or the basis of his identity through time. On the contrary, it comports best with the view that his past and future selves are determined aspectually, too, as the past and future persons whom he can think of first-personally, as "me." Their being his past and future selves need have nothing to do with whether they preserve some component of his psyche.

I believe that a person's understanding makes a distinctive contribution to just those behaviors which count as his autonomous actions. Roughly speaking, my view is that autonomous action is behavior motivated in part by the understanding. How the understanding motivates is a question beyond the scope of this paper. What's relevant here is that this part of a person can be the locus of his autonomy, by virtue of being his essential self, but without necessarily constituting his essence or identity as a person. Autonomy can be an aspectual matter, a matter of whether behavior originates in a part of the person that inevitably presents a reflexive aspect to him.[73]

This conception of autonomy remains deeply indebted to Frankfurt. The guiding insight of Frankfurt's work is that a person's capacity to act autonomously rests on his capacity to reflect on aspects of his personality and to feel a special relation to some of them. My aspectual conception of autonomy is little more than a reinterpretation of this insight.

In fact, my aspectual conception of autonomy is just a reinterpretation of a statement from Frankfurt's most recent work on the subject. What Frankfurt now says, in part, is that autonomous behavior is motivated by parts of the subject with which he cannot help identifying.[74] If "identification" is read as a term for first-personal thinking—for thinking of something as "me" or "mine"—then Frankfurt's statement simply becomes the aspectual thesis.

Like much philosophy of action over the past twenty years, then, my view can be expressed as a commentary on Frankfurt—specifically, on what Frankfurt meant, or should have meant, by "identification." I agree with Frankfurt that autonomous action is guided by a part of us with which we cannot help identifying; I disagree mainly with his claim that our identifying with a part of ourselves incorporates it into something called the self.

Acknowledgments

For comments on earlier drafts of this essay, I am grateful to Nomy Arpaly, Linda Wimer Brakel, Michael Bratman, Sarah Buss, Jennifer Church, and Connie Rosati. The essay was presented to a conference on autonomy organized by Joel Anderson and Sigurdur Kristinsson, to the Philosophy Department of Kansas State University, and to the Contemporary Philosophy Workshop at the University of Chicago. Work on the essay was supported by a fellowship from the John Simon Guggenheim Memorial Foundation, together with matching grants from the Department of Philosophy and the College of Literature, Science, and the Arts, University of Michigan.

Notes

1. *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1987).

2. I discuss some of the relevant literature in "What Happens When Someone Acts?" in *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000). For a more recent discussion, and a promising alternative to Frankfurt's view, see Michael Bratman, "Identification, Decision, and Treating as a Reason," in *Faces of Intention* (Cambridge: Cambridge University Press, 1999), 185–206.

3. Harry Frankfurt, "Autonomy, Necessity, and Love," in *Necessity, Volition, and Love* (Cambridge, Cambridge University Press, 1999), 132. Frankfurt first made this claim, though less explicitly, in the paper "Rationality and the Unthinkable," in *The Importance of What We Care About,* 178: "With respect to a person whose will has no fixed determinate character, it seems that the notion of autonomy or self-direction cannot find a grip."

4. Frankfurt, "Autonomy, Necessity, and Love," 132. This remark about Kant's conception of the person as a subject of autonomy bears a striking resemblance to remarks often made about Kant's conception of the person as an object of moral concern. In Kant's view, a person is worthy of moral concern insofar as he is an instance of rational nature, which is the proper object of respect. Many critics reply that concern for someone as an instance of rational nature is concern for no one in particular, since the same generic nature is instantiated in everyone alike. As Robin Dillon puts it: "In Kantian-respecting someone, there is a real sense in which we are not paying attention to *her*, for it makes no difference to how we respect her that she is who she is and not some other individual. Kantian respect is thus not a 'respecter of persons,' in the sense that it does not discriminate or distinguish among persons." Robin Dillon, "Respect and Care: Toward Moral Integration," *Canadian Journal of Philosophy* 22 (1992): 105–132, at 121. See also Elizabeth Spelman, "On Treating Persons as Persons," *Ethics* 88

(1977): 150–161; Robert Paul Wolff, "There's Nobody Here but Us Persons," in *Women and Philosophy; Toward a Theory of Liberation,* eds. Carol C. Gould and Marx W. Wartofsky (New York: G. P. Putnam's Sons, 1976), 128–144; Seyla Benhabib, "The Generalized and the Concrete Other: The Kohlberg-Gilligan Controversy and Moral Theory," in *Women and Moral Theory,* eds. Eva Feder Kitty and Diana T. Meyers (Totowa, NJ: Rowman and Littlefield, 1987), 154–177; Edward Johnson, "Ignoring Persons," in *Respect for Persons,* ed. O. H. Green, *Tulane Studies in Philosophy* 31 (1982): 91–105. I discuss Dillon's statement in "Love as a Moral Emotion," *Ethics* 109 (1999): 338–374.

5. Frankfurt, "Autonomy, Necessity, and Love," 135.

6. Frankfurt, "On the Necessity of Ideals," in *Necessity, Volition, and Love,* 113; see also "Autonomy, Necessity, and Love," 138.

7. Frankfurt, "On the Necessity of Ideals," 111–112; see also "Autonomy, Necessity, and Love," 136–138; and "Rationality and the Unthinkable," 182–183.

8. Frankfurt, "On the Necessity of Ideals," 112.

9. The image is Frankfurt's. See, e.g., "Freedom of the Will and the Concept of a Person," in *The Importance of What We Care About,* 21; "Identification and Wholeheartedness," in *The Importance of What We Care About,* 168.

10. Ibid. See also "Rationality and the Unthinkable," 187. Presumably then, the person must endorse his inability to will any change in his inability to will. Here the threat of a regress reappears.

11. Frankfurt, "Autonomy, Necessity, and Love," 138.

12. Ibid., 139, n. 8.

13. Frankfurt, preface to *The Importance of What We Care About*, vii–ix, at ix. See also "Rationality and the Unthinkable," 188; and "On the Necessity of Ideals."

14. Frankfurt, "Rationality and the Unthinkable," 179.

15. Frankfurt, "On the Usefulness of Final Ends," in *Necessity, Volition, and Love,* 89.

16. Frankfurt, "Autonomy, Necessity, and Love," 139, n. 9.

17. Ibid., 138.

18. Ibid., 139.

19. In "The Necessity of Ideals," 113, Frankfurt says: "The idea that the identity of a thing is to be understood in terms of conditions that are essential for its existence is one of the oldest and most compelling of the philosophical principles that guide our efforts to clarify our thought. To grasp what a thing is, we must grasp its essence—viz., those characteristics without which it is not possible for it to be what it is. Thus, the notions of necessity and identity are intimately related."

20. Parfit denies that the relevant connections constitute a relation of identity, because they can branch in such a way as to connect a person to two distinct selves existing at one time. I will use the term "selves" to denote those person-

stages to whom one bears the survival-relation that matters, whether or not they are stages of a single person.

21. In this discussion I gloss over many other differences between Locke and Parfit. Most important, perhaps, is that Parfit distinguishes survival from strict identity through time. Parfit thinks that survival, unlike identity, admits of degrees and intransitivities.

22. Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), 205–206; see also 301–302.

23. Ibid., 300–301. See also 515, n. 6.

24. Ibid., 299.

25. This consequence of Parfit's view is confirmed by his discussion of "The Nineteenth Century Russian" in *Reasons and Persons,* 327–328. This idealistic nobleman bequeaths his land to the peasants, in a document that can be revoked only by his wife. He then makes his wife promise that she won't revoke the document, even if he later asks her to. Parfit comments, "The young Russian socialist regards his ideals as essential to his present self. He asks his wife to promise to this present self not to act against these ideals. And, on this way of thinking, she can never be released from her commitment. The self to whom she is committed would, in trying to release her, cease to exist."

26. "Ego Distortion in Terms of True and False Self," in *Collected Papers: Through Paediatrics to Psycho-analysis* (London: Tavistock Publications, 1958).

27. One might be tempted to say that the exaggerated self-control of such a person is only an imitation of real self-control, amounting only to pseudo-autonomy. See David Shapiro, *Autonomy and Rigid Character* (New York: Basic Books, 1981), 74–75. But even if this person has only pseudo-autonomy, the reason is not that he is only pretending to be autonomous; it's rather that he's somehow misapplying the self-control that would otherwise count as very real autonomy. His pseudo-autonomy consists in the misdirected self-control by which he holds himself to a pretense; but the pretense is one of authenticity. Autonomy is not what the agent simulates but what he misuses in mounting the simulation.

28. "Does the True Self Really Exist? A Critique of Winnicott's True Self Concept," in *A Psychoanalysis for our Time: Exploring the Blindness of the Seeing "I"* (New York: New York University Press, 1998), 109. Also relevant here is Erikson's discussion of the difference between "wholeness" and "totalism," in *Identity, Youth, and Crisis* (New York: W. W. Norton, 1968), 74–90. Erikson concludes: "To have the courage of one's diversity is a sign of wholeness in individuals and in civilizations" (90). See also John D. W. Andrews, *The Active Self in Psychotherapy: An Integration of Therapeutic Styles* (Boston: Allyn and Bacon, 1991), 7–8, 35; Roderick Anscombe, "The Myth of the True Self," 52 *Psychiatry* 209–217 (1989); Mark Epstein, *Thoughts Without a Thinker: Psychotherapy From a Buddhist Perspective* (New York: Basic Books, 1995), 71–73.

29. "Self to Self," *Philosophical Review* 105 (1996): 39–76, esp. 42, n. 5. I do include long-range intentions among the mental states that connect the self to

itself through time. Hence a person's plans and policies are part of what make him (in Locke's phrase) "self to himself" from one time to another, in my view. But I do not believe that particular plans and policies can play a privileged role simply because of their distinctiveness or importance to the subject.

30. Robert Paul Wolff, "There's Nobody Here But Us Persons," 136–137. This passage is quoted by Johnson, "Ignoring Persons," 97. Another author who thinks that the authority of our concerns can be traced to their place in our identity, and hence to the imperative of self-preservation, is Christine Korsgaard. See *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996). For a critique of Korsgaard on this point, see David Copp, "Korsgaard on Normativity, Identity, and the Ground of Obligation," in *Rationality, Realism, Revision: Proceedings of the Third International Congress of the Society for Analytical Philosophy,* vol. 23 of *Perspektiven der Analytiscen Philosophe*, ed. Julian Nida-Rümelin (Berlin: de Gruyter, 1999), 572–581. I discuss Korsgaard's version of this view in "Willing the Law," to appear in *Practical Conflicts*, ed. Monika Betzler and Peter Baumann (Cambridge: Cambridge University Press).

31. Frankfurt, "Identification and Wholeheartedness," in *The Importance of What We Care About,* 170.

32. Ibid., 172. See also Frankfurt, "The Faintest Passion" in *Necessity, Volition, and Love,* 100: "Wholeheartedness does not require that a person be altogether untroubled by inner opposition to his will. It just requires that, with respect to any such conflict, he himself be fully resolved. This means that he must be resolutely on the side of one of the forces struggling within him and not on the side of any other."

33. Frankfurt, "Identification and Wholeheartedness," 173–174. In "The Faintest Passion," Frankfurt quotes Saint Augustine as calling ambivalence "'a disease of the mind' from which we suffer in punishment for Original Sin" (100). Frankfurt himself calls it here "a disease of the will."

34. Sigmund Freud, "Notes Upon a Case of Obsessional Neurosis," in *The Standard Edition of the Complete Psychological Works of Sigmund Freud,* ed. James Strachey et al. (London: Hogarth Press and the Institute of Psycho-Analysis, 1953–1974), Vol. X, 153–249.

35. Ibid., 177.

36. Ibid., 191; see also 180–183; 237–241.

37. For the image of two selves, see Freud, "Notes Upon a Case," 177.

38. Ibid., 238.

39. Ibid., 196–197; see also 175–176; 231–232.

40. Ibid., 178–180, 222;

41. Ibid., 209.

42. Ibid., 188–189.

43. Ibid., 167.

44. Ibid., 184.

45. Quoted at note 31, above.

46. Ibid., 162–163.

47. These expressions are quoted above at notes 31 and 35, respectively. See Andrews, *The Active Self,* 29.

48. Morris Eagle has argued that expulsion from the self is the ultimate aim of repression, and that expulsion from consciousness is adopted as a means to that aim. "Psychoanalytic Conceptions of the Self," in *The Self: Interdisciplinary Approaches* (New York: Springer-Verlag, 1991), 49–65. Eagle writes: "[T]he logic of Freudian theory, particularly the clinical logic, points to the expulsion of mental contents from *self-organization* as the significant aspect of repression. Or, to put it more fully, it is the *disowning or disavowal* of mental contents, the rejection of these contents as one's own, that is the clinically (and theoretically) significant aspect of repression. According to the view I am suggesting, expulsion from conscious awareness is mainly *a means,* albeit the most frequently employed means, toward the end of rendering certain mental contents as an impersonal 'it,' as not mine, as ego alien. But it is the disowned ego-alien status that is the critical element in dealing with mental contents incompatible with one's self-structure" (55). See note 66, below. See also Morris Eagle, "Psychoanalysis and the Personal," in *Mind, Psychoanalysis, and Science,* ed. P. Clark and C. Wright (Oxford: Blackwell, 1988), 91–111.

49. In other words, the Rat Man needed to attain the "developmental milestone" that Rubin describes as "access to, and comfort with, [his] multidimensionality," or "a sense of the complexity, multidimensionality, and polyvalency of the self" (quoted at note 28, above). Compare here Eagle's interpretation of psychotherapeutic change, as expressed in Freud's formula "where id was, there should ego be": "Freud is defining psychotherapeutic change, not in terms of consciousness and understanding, but in terms of alterations in self-structure. That is, he is saying that real change is marked by an enlargement of the self." Eagle, "Psychoanalytic Conceptions of the Self," 61.

50. See Freud, "Notes Upon a Case," 226–227, 233–236. Freud's view was that we fear the magical fulfillment of our hostile wishes, via the "omnipotence of thoughts."

51. There is at least one passage in which Frankfurt uses the term in a different sense: "A person who cares about something is, as it were, invested in it. By caring about it, he makes himself susceptible to benefits and vulnerable to losses depending upon whether what he cares about flourishes or is diminished. We may say that in this sense he *identifies* himself with what he cares about." Frankfurt, "On the Necessity of Ideals," 111. Here the target of identification is, not one of the subject's own motives, but the external object of those motives.

52. For a related argument, see Nomy Arpaly and Timothy Schroeder, "Alienation and Externality," *Canadian Journal of Philosophy* 29 (1999): 371–387.

53. We also use the term "identification" to describe a person's sense of affiliation with a social group or movement; but this context usually calls for a different construction. What a person usually does with respect to a group or movement is, not simply to identify with it, but rather to identify *himself* with

it—which is a slightly different maneuver. Frankfurt tends to use the constructions "to identify with" and "to identify oneself with" interchangeably.

54. Anna Freud, *The Ego and the Mechanisms of Defense* (London: Hogarth Press, 1937), 109–121; reprinted in *Pivotal Papers on Identification,* ed. George H. Pollock (Madison, CT: International Universities Press, 1993), 105–114. On the variety of motives for identification, see Roy Schafer, *Aspects of Internalization* (New York: International Unversities Press, 1968), chap. 1.

55. See Richard Wollheim, "Imagination and Identification," in *On Art and the Mind* (Cambridge, MA: Harvard University Press, 1974), 73ff.

56. I discuss this claim at length in "Self to Self."

57. In addition to imagining that you are someone else, you can also imagine that he is you, by regarding him externally and thinking "That's me," as you might think in reference to a photograph or a reflection of yourself in a shop window. Can you identify with someone by means of this external sort of imagining?

Imagining an external figure to be oneself may not seem like a compelling way of identifying with him. But consider the universal phenomenon of dreaming about oneself from a perspective outside one's own body, a perspective from which one sees oneself as an external figure. This phenomenon suggests that representing an external figure as oneself is deeply entrenched in the representational idiom of the imagination. And imaginatively representing another person as one represents oneself in dreams ought to be a compelling way of identifying with him.

But how exactly does a dream represent which figure is oneself? As the dream unfolds, one doesn't suddenly focus on one of the characters and think "That's me," as one does when recognizing oneself in a home movie. It's implicit from the outset of the dream that a particular figure is oneself, and so it never gets formulated explicitly in thought. The question is how one's identity among externally represented figures can be implicit in the dream.

Maybe one's identity is implicit in the emotional structure of the dream. Although one sees and hears all of the characters from the outside, one reacts emotionally to the dreamed events from the perspective of one character in particular, fearing whatever threatens him in the dream, resenting whoever insults him, and so on. Indeed, one doesn't so much react emotionally to dreamed events as one dreams them emotionally, to begin with: one fearfully dreams threats to a character and resentfully dreams insults to him. One thus dreams the dream from the emotional perspective of a particular character, and this structural feature of the dream may be what implicitly casts him in the role of one's dream-self.

If so, then dreaming of an external figure as oneself is not very different, after all, from dreaming of being that person. Although the dream doesn't represent things as seen or heard by him, it does represent them as emotionally felt by him, and its so representing them is what casts him in the first-personal role. The emotions experienced in the dream are a partial image of being that person, just as a feeling imagined in the throat can be a partial image of being someone who is seen taking a drink. Thus all identification may involve imagining from within the other's perspective in some respect.

58. On the effects of first-personal imagining, see Wollheim, "Imagination and Identification"; and *The Thread of Life* (Cambridge, MA: Harvard University Press, 1984), chap. 3. I think that empathy itself may be a mode of identification, in which we imagine the world as experienced by someone else emotionally rather than perceptually. My thought here is that the emotions we feel on behalf of the other person play the same role as the sensations that we feel on behalf of the drinking jogger described above. To develop this thought, however, I would need to offer a theory of the emotional imagination, explaining how we can feel imaginary emotions in the way that we can taste an imaginary drink. For some remarks in this direction, see my "On the Aim of Belief," in *The Possibility of Practical Reason* (Oxford: Clarendon Press, 2000), 270, n. 51.

59. I defend this claim at length in "On the Aim of Belief."

60. This question is also raised by Arpaly and Schroeder.

61. Jean-Paul Sartre, *Being and Nothingness; An Essay on Phenomenological Ontology,* trans. by Hazel E. Barnes (New York: Philosophical Library, 1956), 59.

62. The views stated in this section are developed more fully in my "What Happens When Someone Acts?"

63. This paragraph summarizes the thesis of my "Self to Self." Also relevant here is Rom Harré's contrast between the self as a point-of-view and the self as a bundle of personal qualities. *The Singular Self: An Introduction to the Psychology of Personhood* (London: Sage Publications, 1998).

64. Here again see Rubin's critique of Winnicott's "True Self Concept."

65. See, e.g., Frederic J. Levine and Robert Kravis, "Psychoanalytic Theories of the Self: Contrasting Clinical Approaches to the New Narcissism," in *The Book of the Self: Person, Pretext, and Process,* ed. Polly Youaang-Eisendrath and James A. Hall (New York: New York University Press, 1987), 306–330; J. F. Kihlstrom and N. Cantor, "Mental Representations of the Self," in *Advances in Experimental Social Psychology,* ed. L. Berkowitz (New York: Academic Press, 1984), vol. 17, 2–40.

66. This sense of the word also allows us to say that the Rat Man's strategies of repression were strategies of expelling unwelcome emotions from the self. For as I explained above, the Rat Man repressed his hatred by writing it out of his self-conception. (See 194, "Notes Upon a Case," and n. 48, above.)

67. Frankfurt sometimes speaks of a person's "sense" or "grasp" of his identity, or "the clarity with which he comprehends who he is" ("On the Necessity of Ideals," 108–109; see also "Rationality and the Unthinkable," 177). But he doesn't adequately distinguish the person's grasp of his identity from the identity so grasped. He thus slides from the statement that "extensive proliferation of [a person's] options may weaken his grasp of his own identity" ("Rationality and the Unthinkable," 177; see also "On the Necessity of Ideals," 109: "extensive growth in the variety of a person's options may weaken his sense of his identity") to the statement that "an excess of freedom gives rise to a diminution, or even to a dissolution, of the reality of the self" ("Rationality and the Unthinkable," 179;

see also "On the Necessity of Ideals," 110: "With total freedom, there can be no individual identity").

68. Thanks to John Exdell for this useful example.

69. Quoted at note 18, above.

70. I discuss this claim in the introduction to *The Possibility of Practical Reason*.

71. As Dan Dennett demonstrates in his paper "Where Am I?" your visual standpoint can migrate out of your body with the help of prosthetic sense-organs. In *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge, Mass.: MIT Press, 1981), 310–323. I take myself to be making the same point as Dennett—namely, that your visual location is a merely aspectual matter.

72. *Nicomachean Ethics*, 1178a.

73. My claim that selfhood is an aspectual matter does not apply to personhood. What makes someone a person is not merely that he presents a particular aspect to himself; what makes him a person includes various other facts about him, including the fact that he is autonomous. Of course, his autonomy consists in the fact that his behavior is governed partly by his understanding, which inevitably presents a reflexive aspect to him, thus qualifying as his essential self; and being governed by such a self is part of what makes him a person; but the self by which he is governed must not be conflated or confused with the person he thus becomes.

In "Self to Self," I argue for a similar distinction between selfhood and personhood in the discussion of personal identity. A person's past and future selves are those past and future persons who present a particular aspect to him, but they need not be the same person.

74. See the quotations at note 10, above.

REPLY TO J. DAVID VELLEMAN

Harry Frankfurt

1. Some of the apparent conflict between Velleman's ideas and mine is due more, I think, to relatively superficial rhetorical differences than to important philosophical disagreement. This is especially so with regard to his objections to my account of "the self." I shall begin by commenting on those, and then consider certain other issues.

Velleman attributes to me the view that the self is "a singular entity waiting to be found." What I believe awaits being found is not the self. It is the *limits* of the self—the volitional necessities that make it impossible for a person to bring himself to behave in certain ways. I have sometimes referred to these limiting necessities as boundaries of the self, which define its shape. Perhaps this language suggested that, since I said the self has boundaries, I must have been thinking of it as "a singular entity." My figure of speech was not intended, however, to have any ontological implications. The truth is that I am not inclined to construe the self as an "entity" at all.

2. Is it improper for me, then, to speak of "the essential nature" of the self? Velleman chides me for asserting that Agamemnon no longer existed after he ruptured a defining feature of his nature by sacrificing his daughter. He wonders how I can account for the fact that Agamemnon was still around later on when Clytemnestra killed him. What I said about Agamemnon's essential nature, and about his being destroyed, certainly does lead quickly to absurdity if it is taken as Velleman takes it. However, it is not necessary to take it in that way. There is a familiar and respectable usage in which it makes perfectly good sense to speak of the essence of a person, and to say that the continued existence of a person—though not the person's life—depends upon the continuity of his essential nature.

Velleman alleges that I have never offered a convincing example of someone who ceased to be the same person, or who ceased to exist, because of some change of heart. Be that as it may, suitable examples are not so hard to find. Suppose that a hateful bully undergoes a moral conversion, with the result that he genuinely renounces his aggressively

vicious ways. It might correctly be said of him that he is not the same person that he was, meaning (among other things) that he is no longer inclined, and would no longer be able even to bring himself, to do the dreadful things for which everyone once feared and despised him. In virtue of his having reformed, the shape of his will has altered; his volitional limits are different than they were. As we might quite commonly say, he has become a different person. Perhaps there are some people who do not know how he has changed, and who are therefore still wary of him. If so, we might naturally tell them that their anxiety when he is around is misguided because the dangerous brute about whom they are worried no longer exists.

There is of course another sense in which, despite the alteration of his volitional character, the reformed man is identical with who he was prior to his change of heart. He is numerically the same person before and after that transformation. Our customary usage of "the same person" is, however, less univocal than Velleman allows. There is a sense, on which I relied when speaking of Agamemnon, in which people may be defined by their volitional limits and in which the survival of the self therefore requires a certain motivational constancy or continuity. In this sense, of course, survival is not a matter of life and death.[1]

3. In commenting on what I say about ambivalence and wholeheartedness, Velleman discusses the Rat Man, whose repression of his hostility toward his father generated some famous pathology. On my account, as Velleman says, the Rat Man would have been better off if, instead of repressing his hostility and thus secreting it in his unconscious, he had rejected it at the conscious level and had identified himself wholeheartedly with his more benign attitudes. Velleman disagrees; and he suspects that my prescription, "though different from what caused the Rat Man's illness, would hardly have been more healthy." In his view, it would have been better for the Rat Man to be ambivalent than to be wholehearted. "Surely," he says, "what the Rat Man should have done was to accept his filial hostility as part of himself, to accept *himself* as ambivalent toward his father."

The Rat Man should certainly have accepted, rather than hidden from himself, the fact that he had hostile as well as loving feelings toward his father. To say this is hardly the same, however, as saying that he should

have accepted himself as being ambivalent. Ambivalence is not a matter simply of having conflicting feelings. It consists essentially in having a divided will—that is, of being unresolved as to which side of the conflict among one's feelings one is on. Thus it cannot be eliminated by repression, for that only renders the unresolved conflict unconscious.

In order for the Rat Man to be wholeheartedly on the side of his benign attitudes, it would not have been necessary for him to conceal his hostile feelings from himself. Nor would he have had to refrain from making a conscious effort to deal with those feelings in whatever ways might be effective and helpful. It would have required only that, in the struggle between his hatred and his love for his father, he himself come to stand decisively against the hatred and behind the love.

4. Velleman reports Freud's belief that "a wish becomes a compulsion when the ego comes to regard it as foreign," and he says that this "is close to what Frankfurt believes as well." Actually, it is not close to what I believe. When Freud speaks of the ego coming to regard a wish as foreign, he has in mind a situation in which the wish is repressed. When I speak of the self rendering a wish external and regarding it as alien, what I have in mind is a quite different maneuver. This maneuver does not entail repressing the wish or making it unconscious, and it is in no way pathogenic.

After asserting that what made the Rat Man ill was "his effort to dissociate himself from one of his emotions," Velleman adds that this is "just what Frankfurt prescribes for cases of ambivalence." But the Rat Man did not become ill because he tried to achieve dissociation. He became ill because he tried to achieve dissociation *through repression*. Repression is quite unlike the mode of dissociation that I prescribe as an antidote to ambivalence. Resolving ambivalence by taking a decisive stand against certain feelings does not require (or even permit) that a person misrepresent those feelings or that he conceal them from himself.

Velleman acknowledges that I do not recommend repression. But it was repression that caused the Rat Man so much trouble. I am therefore puzzled by his claim that the Rat Man would hardly have been better off if he had unified his will in the way that I do recommend—that is, by identifying with his love for his father, without repressing anything and without neglecting conscious ways of coping with feelings of hatred by

which he did not wish to be driven. I agree that defensive and self-deceptive wishes, which Velleman says may play a role in the desire for wholeheartedness, are inimical to health. However, I see no reason to think that the desire for wholeheartedness naturally depends upon such wishes.

5. Surely it is ambivalence, and not wholeheartedness, that is a disease of the will. Velleman himself makes it clear why this is so. Thus he reports that the Rat Man's symptoms "often involved repeatedly doing and undoing an action, or thinking and contradicting a thought." That sort of self-defeating behavior and thought violates the elementary requirements of rationality. It is not a consequence of repression as such, but of ambivalence. It is a manifestation of the incoherence in which, precisely, the divided will of ambivalence consists. The desire for wholeheartedness is nothing other than a desire to be free of this crippling irrationality.

6. In his discussion of identification, Velleman says that his paradigms are cases in which a person imagines that he is someone else. These, he claims, are instances of "identification as ordinarily understood" (27). The examples he provides, however, are not actually of that kind; they are not cases in which a person imagines that he is someone else. Instead, they are cases in which someone imagines having an experience like the experience another person is having.

What does it mean to identify with another person? When I say that I can (or that I do) identify with someone, I do not mean that I can (or that I do) imagine being that person. I mean just that I can (or do) imagine myself actually responding as that person does to a situation of a certain kind. Suppose that in the aftermath of Waterloo, I tell Napoleon that I can identify with him since I too have suffered catastrophic defeats. I do not mean that I can imagine myself being Napoleon (whatever that might mean) but, at most, that I can imagine what it is like to be in Napoleon's shoes. That is, I know what it is like to suffer a catatstrophic defeat.

It also seems to me inaccurate to suppose, as Velleman evidently does, that when I identify with a desire or with an emotion, I "imagine being the desire or emotion." Identifying with something is, in the relevant sense, not a matter of becoming identical with it. It is, as Velleman

himself puts it, "a psychological process by which a person empowers [authorizes?] some of his motives to implicate him in causing behavior." Velleman believes that this is not what the term ordinarily means, but the usage seems commonplace enough. After all, we do say things like: "You've got me wrong, you don't understand who [what sort of person] I am." In other words, we complain about being identified with desires and motives with which we do not truly identify.

Note

1. This usage, more than the other, lends itself to (perhaps it even encourages) an undesirable mushiness. Although it is well enough established in common speech, perhaps it would after all be better to keep it out of philosophical discourse. I confess that some of the statements in which I have resorted to it do now strike me as a bit overblown.