Motivation by Ideal

Article *in* Philosophical Explorations · May 2002

DOI: 10.1080/10002002058538724

CITATIONS
25

READS
226

1 author:

J. David Velleman
New York University
87 PUBLICATIONS **3,184** CITATIONS

SEE PROFILE

Motivation by Ideal[*]

J. David Velleman

Abstract

I offer an account of how ideals motivate us. My account suggests that although emulating an ideal is often rational, it can lead us to do irrational things.

When philosophers discuss our motive for acting morally, they tend to assume that it serves as one contributor to the broad conflux of motives that jointly determine most of our behavior. Although philosophers recognize the possibility of our being divided into mutually isolated motivational currents of the sort posited, at the extreme, to explain phenomena such as multiple personality, they assume that our moral motive must not be thus divided from our other motives, lest its manifestations in our behavior turn out to be irrational and, at the extreme, insane. Their assumption is that the actions flowing from our moral motive must in fact flow from a unified stream of all our motives, augmented by a moral tributary.

This assumption influences which questions are asked about moral motivation and which answers are considered plausible. The assumption encourages philosophers to ask, for example, how to identify our moral motive among the impulses that pass under the eye of ordinary deliberative reflection, and how that motive can possibly prevail against the impulses that so conspicuously favor immorality.

I am going to argue that the motive behind moral actions can become isolated from our other motives, generating behavior that is irrational in some respects though rational in others. In my view, moral action performed from moral motives can be less than fully rational precisely because of the division in its moti-

* This is the third in a series of four papers on narrative self-conceptions and their role in moral motivation. In the first paper, "The Self as Narrator" (to appear in *Autonomy and the Challenges to Liberalism: New Essays*, ed. Joel Anderson and John Christman), I explore the motivational role of narrative self-conceptions, drawing on Daniel Dennett's notion of the self as a "center of narrative gravity". In the second paper, "Willing the Law" (in *Practical Conflicts*, ed. Monika Betzler and Peter Baumann [Cambridge: Cambridge University Press, forthcoming]), I explore the role of self-conceptions in Kantian "conflicts in the will", drawing on Christine Korsgaard's notion of "practical identities". In a fourth paper, "A Sense of Self" (MS), I explore the role of narrative self-conceptions in Kantian "conflicts in conception", drawing on the work of Thomas Nagel and John Perry on the self.

For initial conversations on the topic of the current paper, I am grateful to Nishi Shah. For comments on an earlier draft, I am indebted to the Departments of Philosophy at the University of Illinois at Chicago, the University of Pittsburgh, and Syracuse University; to members of ORGiE (the Ohio Reading Group in Ethics), including Justin D'Arms, Dan Farrell, Don Hubin, Janice Dowell, David Sobel, Sigrun Svavarsdottir, and special guest Doug Lavin; and to an anonymous referee for this journal.

vation. The reason why moral motivation can become isolated from our other motives, I shall argue, is that it often depends on the force of an ideal; an ideal gains motivational force when we identify with it; and acting out of identification with an ideal is like a game of make-believe, in which we pretend to be that with which we identify. My argument will begin, then, with a consideration of adult make-believe.

For many years, I regularly kicked my wife in the head. We were studying Tae Kwon Do, and we often found ourselves paired together in drills or sparring. There we stood, high-school sweethearts from the sixties, each apparently trying to knock the other's block off.

What is the motivational explanation for such behavior? The motives most obviously actuating me in the circumstances were my desires to enhance my cardiovascular fitness and to have some fun in the process. But surely there would be something odd about saying that I kicked my wife in the head in order to lower my cholesterol or just for fun. Of course, I knew – or, at least, hoped – that my wife would suffer no harm. She was wearing a foam helmet, I was wearing padded footgear, and I didn't strike with all of my strength. You might think, in fact, that I didn't so much kick her in the head as do something else that was only superficially similar, such as tap her on the temple with my toe. Such a tap could indeed have been produced by many motives of mine, including affection. Yet to say that I was trying to deliver a tap would misrepresent the encounter: a pulled punch or kick may feel like a tap to the recipient, but it is in fact quite dissimilar, since it is thrown with full force and "pulled" only in the sense of being aimed to fall short.

What calls for further explanation is not so much the fact that I kicked my wife on these occasions as the spirit in which I did so. For one thing, the effort behind my kicks was disproportionate to the motives that led me to the activity of sparring. The desires and beliefs that militated for kicking my wife may well have been stronger than the desires and beliefs that militated against, but not by enough of a margin to account for the zeal with which I went at her. Shouldn't effort be proportionate to motivation?

Then there is the manner of my kicks, which also seems to require further explanation. One and the same gross movement can evince different motives through subtle differences of posture, timing, muscle tension, and body english. The kicks that I aimed at my wife did not have the inflection of calisthenics or soccer or dance; they had the inflection of combat.

The key to explaining these aspects of my behavior, I think, is that Tae Kwon Do had helped me to solve a familiar motivational problem. The effort that one must expend in order to stay fit tends to require more motivation than can be supplied by one's desire for fitness: that's why so many exercise programs fail. If one wants to stay fit, one needs to find some additional source of motivation to draw on. Some forms of exercise give one access to competitiveness as an additional motive, others to team spirit, a love of nature, or musical inspiration. My additional source of motivation in Tae Kwon Do was aggression, and aggression is what accounted for the energy and inflection of my kicks.

Reflection on this case convinces me that there must be some truth in Freud's theory of the drives. What's true in that theory, I think, is the postulation of highly labile psychic energies, which have only a vague direction in themselves but can be invested in specific activities.[1]

In studying Tae Kwon Do, I discovered that I had a fund of aggression to spend on kicks and punches, whether they were aimed at a leather bag, a handheld target, or a person's head. This aggression is not best characterized in terms of desire and belief. I did not enter the *do jang* wanting to smash something and looking for something to smash: rough contact with medium-sized objects was not something I desired at all. But it was something for which I found a considerable reserve of energy, in the form of aggression; and that aggression could be turned on virtually any solid object, including any person who happened to be my assigned opponent.

I am similarly inclined to believe in a drive corresponding roughly to the Freudian libido. We sometimes describe a person as having a lot of love to give but nowhere to give it. Such a person has a fund of tenderness that could potentially be spent on a lover, a child, a cat, even a garden or a scrapbook. In this case, unspent energy may be experienced as frustration, and so the person may develop a desire for someone or something to love. But such a desire need not develop; and even when it does, it remains distinct from the fund of energy whose disbursement it seeks. The person's desire for someone or something to love is a contingent reaction to his unspent tenderness, not an essential constituent of it.

I realize that talk of psychic energies will strike philosophical readers as intolerably metaphorical. In principle, the metaphor can be eliminated in favor of concepts drawn from propositional-attitude psychology: we can conceive of aggression as a conative attitude whose object is picked out by a mental representation of some kind. But we shall then be forced to conceive of this representation either as radically indeterminate in content or as playing a non-standard role;[2] and the resulting conception of aggression will not lend itself to the kind of formalization that has so endeared propositional-attitude psychology to philosophers ever since Aristotle discovered the practical syllogism.

On the one hand, if we think of aggression as motivating the pursuit of, and being temporarily quelled by, the literal truth of the associated representation, then we shall have to say that the representation is far too vague to be expressed in the concepts with which we consciously reason, or the terms in which we write and speak. There is no finite "that" clause of ordinary language that will suffice to specify the pursuits or satisfactions in which aggression can eventuate. If, on the other hand, we insist on framing a written or spoken "that" clause to express the content

1 Let me emphasize that I am borrowing only some elements of Freudian drive theory. I am not borrowing the model of stimulus reduction, for example, but only the notion of indeterminate motivational forces. Indeed, my conception of their indeterminacy is different from Freud's. Freud described drives as having determinate aims but being readily redirected toward other aims instead. I prefer to think of drives as having only inchoate aims.

2 The substance of this paragraph is borrowed from Linda Brakel's work on primary process. See Brakel (2002).

of aggression, we shall have to concede that what the attitude can motivate someone towards, or be satisfied by, includes not only the literal truth of the clause but also indefinitely many other outcomes related only by analogy, by metaphorical similarity, or by other mental associations of an open-ended variety. Either way, propositional-attitude psychology will not afford the same computational advantages in this case as it does in the case of ordinary beliefs and desires, whose tendencies to motivate and to be satisfied can be summed up in sentences of ordinary language.

Thus, we can accommodate drives within the basic principles of propositional attitude psychology, but only by allowing for a level of mental representation, or a mechanism of motivation, that eludes capture by the explanation schemas characteristic of that theory. The metaphor of psychic energies is a useful reminder that, even if all motives are propositional attitudes in principle, some have motivational possibilities that cannot practically be formalized in spoken or written propositions.

Another idea that I want to borrow from Freud is that drives can take on a specific direction by "leaning" on some other, more specific motive. According to Freud, the infantile libido leans on and takes direction from the motive of hunger, with the result that the nutritive activity of sucking becomes a source of sensual pleasure, and the breast becomes a sexual object. Similarly, I think, aggression can take direction from more specific motives, such as professional ambition or athletic competitiveness. Aggressive energy is then invested in professional or athletic pursuits, which in turn take on an aggressive character.

The spirit of my kicks in Tae Kwon Do can thus be explained by the aggression from which they drew some of their motivation. Yet the explanation can hardly end here. The aggressiveness of my kicks was not like the aggressiveness of my driving, for example, which emerges without my knowledge and even despite my efforts to contain it. The aggressiveness of my kicks was knowing and intentional, because I was engaged in a fight. And yet I had no motives for, and many motives against, literally fighting my opponents. I was behaving aggressively in this case because I was engaged in fictional aggression, and so an explanation of my behavior requires an account of the operative fiction.

A martial art typically relies on a story – indeed, on a story-within-a-story, especially for students in the West. The "inner" story is a story of combat. At the founding of the discipline, this story may have been about combat on the battlefield, but in the modern *do jang* it is often about being attacked on the proverbial street. Some students have actually lived through a version of this story, especially women who seek out the martial arts after surviving rape or domestic abuse. But even these students train under a fiction, insofar as they are not really being attacked by their fellow students.

The "outer" story of a martial art, which is usually a fiction only for beginning students and then only briefly, is that they are devotees of a venerable tradition, transmitted to them by a revered master and shared with others in a spirit of humility and mutual self-restraint. The beginning student acts out this story before it can possibly be true of him, by bowing to his instructor and fellow students, calling them "Sir" and "Ma'am," wearing ritual garments and reciting ritu-

al phrases, all from the first moment of the first class. For some students this story always remains a fiction, in the sense that they are never more than playing at participation in the tradition; but for most it soon becomes a true story, and the phrases of Korean or Japanese that were at first only mouthed come to be sincerely meant.

The inner and outer stories of a martial art are in direct conflict. The ferocity with which one tries to disable or kill an attacker, according to the inner story, is the very opposite of the humble deference that, according to the outer story, one owes to the instructor who may be playing the attacker's role. This conflict is vividly demonstrated when someone is injured in competitive sparring. The competitor responsible for the injury, who a moment ago seemed intent on bloody murder, suddenly kneels with his back to his opponent, in a posture of passivity and penitence, because he has drawn a single drop of blood. The fiction of combat is instantly dispelled, leaving only the outer story of deferential self-restraint.

This scene illustrates two further claims that I want to make about motivation, in addition to my prior claim on behalf of drives. The first of these further claims is that our motives are often manifested in our behavior under the guidance of a story: how we act on them is determined by the story that we are enacting.

The most ambitious version of this claim, which I have defended elsewhere, is that all of our autonomous actions are the enactments of stories, most of which are true but all of which are made up.[3] At any particular time we have motives for taking various actions, and the action we take is usually the one whose story we have in mind to enact. We are therefore in a position to make up the story of our behavior as we go, in the assurance that we'll behave accordingly, provided that we confine ourselves to stories whose enactment could be fueled by motives that we actually have. And the story that we make up is true, not only in that we proceed to enact it, but also in that it represents our action as its own enactment – as the action that we are hereby setting ourselves to take.

My view is that this process depends on a motive that is almost always in the background, and rarely in the foreground, of our autonomous actions: the desire to make sense of what we're doing. This desire moves us to take actions that make sense to us, and the actions that make sense are the ones about which we have a story to tell. Thus, although we ultimately do what is favored by the overall balance of our motives, that balance has often been tipped by the inclusion of our motive for doing things that make sense to us – a motive that is purely formal and does not appear in our conscious story of what we're doing. That story may tell of other motives in light of which the action makes sense to us; but we perform the action not only out of those narrated motives but also out of our motive for making sense, which is enlisted by the availability of the narrative itself.

When a story renders an action intelligible to us, it becomes a *rationale* for the

3 Of course, the "stories" enacted in our autonomous actions are not the stuff of novels: they can be as trivial as the story that goes "My leg itches, so I'm scratching it." For this view of autonomy, see Velleman (1989); Velleman (2000a), Chapters 1, 2, 7, and 9; Velleman (2000b); Velleman (forthcoming); and Velleman (ms). My view of agency bears similarities to: Hollis (1977), Harré (1979), and Anscombe (2000).

action. And when we are thereby led to perform the action, as the intelligible thing to do, we act on the basis of the story in its capacity as *rationale*. In other words, we act for a reason.

Thus, when I entered the *do jang* on a particular evening, I thought of myself as continuing my martial arts training, which I thought of myself as pursuing for the sake of cardiovascular fitness and fun. These thoughts were not just an idle commentary on my behavior; they constituted a story that I was in the process of enacting, with actions that I would not have taken in the absence of a story to tell about them. That I was seeking to continue my training out of desires for fitness and fun – that story was the *rationale* under which I entered the *do jang*. It was my reason for walking in the door.

Yet when I kicked an opponent in Tae Kwon Do, the story I enacted wasn't true, since it was a story of fending off a mortal attack. My behavior was therefore an enactment in the thespian sense – or, if you like, a game of make-believe.

I have argued elsewhere that the term 'make-believe' means "mock-belief," because it refers to a fantasy or imagining that stands in for a belief by playing its motivational role.[4] My examples on that occasion were primarily imaginings that play the role of ordinary instrumental beliefs – such as the belief that I can communicate with someone by speaking to him, which is ordinarily one of the motives behind my verbal behavior. When I address remarks to other drivers on the road, however, or to the referees of a sporting event on television, I am not moved by the belief that I can thereby communicate with them; I'm moved instead by imagining that I can. Because imagining here plays the motivational role of a belief, it qualifies as "mock-belief," and I can be described as making believe. I'm making believe that I can communicate with these people, by acting on a mock-belief to that effect.

In the present context, I want to consider imaginings that substitute for beliefs in a slightly different motivational role. If I really believed myself to be under attack, that belief would serve as a narrative premise under which some courses of action would make sense and others would not, and I would be guided accordingly as I improvised my part in the encounter. Strictly speaking, this belief would be functioning as an instrumental motive, since it would influence me by causing some steps but not others to appear intelligible and hence conducive to making sense of what I do, which is a desired outcome. But this outcome is not an end-in-view – not, that is, an end-in-the-story, something whose pursuit I would enact. It's just something that I want, conduciveness to which makes actions attractive to me. And what it makes attractive to me, in particular, are actions about which I have a story to tell.

I think that fantasy and imagining can play this motivational role as well. When I imagined that I was facing an attack in Tae Kwon Do, I was thereby led to imagine some steps as making sense and others as making none, and I was guided accordingly as I improvised my part in the ensuing fights – make-believe fights, guided by a mock-belief. I then enacted a story that was fictional in every

4 "On the Aim of Belief", in Velleman (2000a), 244 - 81.

sense, since it was not only made up but also untrue.

Part of the story, of course, was that I fought out of a desire to disable or kill my opponent, and in reality I didn't have any such desire to draw on. What I drew on instead, I have argued, is a labile fund of aggression, which leaned in this case, not on any desire to harm my opponent, but on the motivational force lent to the story itself by my inclination to do what made sense in light of that story. I may actually have imagined the felt thrust of aggression to be a desire to harm my opponent, much as I was obliged, in self-defense drills, to imagine wooden batons to be knives. (In that case, my aggression served as a "prop," in the sense defined by Kendall L. Walton.)[5] In reality, however, my aggression's being focused upon my opponent was due to my conceiving of it as a desire to harm him, rather than the other way around. That is, imagining it as a desire to harm my opponent lent intelligibility to the act of kicking, thus giving me a motive for kicking, on which my aggression could lean.

The game of make-believe was thus fueled by two elements – a drive and an imagining – and the game would fail if either element was missing. Some students of the martial arts don't have much aggression to draw on, and they consequently aren't fully equipped to play the game. Merely imagining that they are under attack isn't enough to make them fight, in the absence of a drive that could supply the force of their imagined desires with respect to an attacker. So their threatening yells always sound like peeps, and their blows really are no more than taps. Other students seem to have sufficient aggression but to be inhibited from entering into the requisite make-believe, at least in some circumstances. For example, some men simply can't bring themselves to imagine that they are trying to kill or disable a woman. Though capable of fighting other opponents aggressively, they can't muster the imagining that would bring their aggression to bear on these opponents, and so they merely go through the motions.

Of course, none of us actually tried to kill or disable an opponent. We were restrained by our sense of mutual respect and deference. But I do not think that the motive of deference simply combined with aggression to yield an intermediate vector-sum – a deferential aggression, or aggressive deference, or whatever. To pull a punch is not simply to strike at half-strength, out of some lukewarm mixture of hot and cold motives. This is my second of my further claims about motivation.

In making this claim, I do not mean to reject the principle that a person's behavior flows from the combined force of his motives; I mean only to point out that, because of the motivational force exerted by an agent' self-conception, there are two distinct ways in which his other motives can combine.

One way requires the agent to think of himself as acting on both motives at once and hence to be guided, not only by their combined forces, but also by his conception of how those forces combine. In this case, the agent is consciously engaged in a mixed activity – restrained hostilities, or perhaps hostile self-restraint. The agent's behavior is determined partly by the combined forces of his motives and partly by his conception of what would make sense for him to do in

5 Walton (1993).

light of their combination, his story of how he is acting on both at once.

Another way for motives to combine is for the agent to conceive of himself as acting on only one of them, while the other tacitly modifies this activity. Thus, for example, an agent's desire to avoid bodily harm steers him away from obstacles even when he is single-mindedly engaged in vigorous activity and not consciously exercising caution. What makes for the difference between these two ways of mixing motives is the motivational role of the agent's self-conception, which is not epiphenomenal on his behavior, not just an idle commentary. In one case, the agent deliberately acts on both motives, by enacting a story of both; in the other case, the agent enacts the story of one motive, while this enactment is subject to unheralded modification by the other.

When we think about the mixing of motives, we usually have the former process in mind, because we assume that people are simultaneously aware of the various motives vying for control of their behavior. We may therefore assume that if students of the martial arts are both mutually deferential and mutually hostile, they must conceive of themselves in both terms at the same time. But such a conflicted self-conception would result in sparring that could only be described as half-assed. In fact, students imagine themselves entirely as hostile opponents while they are sparring, but this role is externally constrained by their deferential motives as colleagues.

Consider what happens when a participant in make-believe gets "carried away". Sometimes students do get carried away in sparring, especially new students who haven't yet learned how to manage the conflicting stories that they are supposed to enact. The reason why it's possible to get carried away, I think, is that a participant in make-believe puts his real identity and his real relations to other participants temporarily out of mind. In order to enact his fictional identity and his fictional relations to others, he must devote his mind to the fiction. In doing so, however, he trusts that the motives he has put out of mind will nevertheless hold him back from excesses, or will pull him up short if things get out of hand. His knowledge of who the participants really are, and his inclinations toward those real people, are motives that stand by and supervise, as it were, either by setting boundaries to the game of make-believe, within which they are not in view, or by forcing their way into view and breaking up the game, if it goes too far. The agent gets carried away when this external supervision fails and the game proceeds headlong, without either restraint or interruption.

Getting carried away often leads to irrational action. When someone gets carried away in a philosophical debate, for example, he presses his point at the expense of other people's feelings and his own reputation for collegiality, both of which he cares about, on balance, more than the question under dispute. In some cases, of course, intellectual enthusiasm may have blinded the agent to the undesirable consequences of his behavior; but in others, he sees those consequences yet presses on with the argument regardless.

From the agent's point-of-view, his motives may appear to wax and wane as circumstances change. In the heat of the argument, the prospect of securing his point consumes all of his attention and interest; whereas in a cooler moment, the philosophical point may seem unimportant. But this introspectable change need

not be a change in the agent's desires themselves; it may instead be a change as to which desire is reinforced by the agent's conception of what he is doing. In the heat of the argument, the agent thinks of himself exclusively as pressing his point, and this self-conception provides reinforcement exclusively to his motives for doing so. Even if the agent notices the annoyance of his interlocutor, he doesn't think of it as something that he is currently wants to avoid or to mitigate. Managing his relations with colleagues is not something toward which he thinks of himself as currently motivated, and so his potential motives for that activity are not bolstered by his interest in self-understanding.

These motives are nevertheless present, and as I have suggested, they have two chances to prevent him from getting carried away, corresponding to the two ways in which motives can combine. First, the desire for good relations with colleagues can leave the agent's pursuit of the argument uninterrupted while restraining it from the outside, in the same way as the desire to avoid bodily harm restrains his physical activities even when he isn't deliberately being cautious. And then, if unreflective restraint fails, the agent's desire for good relations with colleagues can obtrude itself on his attention, so that his concentration on the argument is broken and he comes to think of himself, under the circumstances, as having more than one end at stake.

These modes of restraint look quite different, both from the agent's perspective and from the perspective of observers. Some philosophers can throw themselves into an argument without fear of giving offense, because they will be unreflectively restrained from going too far. These philosophers are said to trust themselves in the heat of an argument, where the "selves" they trust are not reflective selves who might be trusted to make the right choice in deliberation but rather motives that can be trusted to restrain them without reflection or deliberation. Other philosophers never fully commit themselves to the point they're trying to make, because they are busy monitoring the expressions of their listeners and interjecting polite qualifications. Because they can't rely on their collegial motives for implicit restraint, they must explicitly adopt self-restraint as an additional activity whenever they get into an argument.

The same contrast applies to participants in the martial arts. If a student can't trust himself in sparring, he must consciously ride two horses at once, both his aggression and his self-restraint. If a student can trust himself, then he can ride his aggression wholeheartedly and count on his self-restraint to run alongside on its own. If the latter strategy fails, the student may be forced to adopt the former – not exactly to switch horses in midstream but to shift part of his weight onto the second horse. And part of what he counted on from his self-restraint, at the outset, was that it would force itself into his activity in this manner if it failed to steer him adequately from the outside.

Both forms of restraint are exemplified in an agent's behavior most of the time. Because moments of true single-mindedness are rare, an agent is often consciously multi-tasking, and yet he is also influenced by additional motives that remain out of view. Bustling down the street on several errands at once, he implicitly trusts himself not to step into potholes or bowl over fellow pedestrians – which is to say, he knows that various latent motives of his will either restrain his conscious pursuits or interrupt them if tacit restraint should fail.

An agent's self-conception thus separates his motives into two groups. One group comprises motives that the agent manifests in the process of consciously enacting them; the other comprises motives that manifest themselves primarily by externally modifying such enactments. The former are the motivational horses that the agent is riding, as I have put it, and the latter are relegated to the role of hemming him in or cutting him off as necessary.

The process becomes further complicated if the agent imagines himself to have motives that he doesn't actually have. The agent may be moved to enact this imaginative self-conception, especially if he has motivational resources that can mimic the force of the imagined motives, such as aggression that can be focussed onto a particular person by being conceived as a desire to kill or disable him. The agent's actual motives are then divided into those on which he is acting under a mistaken or imaginary guise, and those which are relegated to hemming in or cutting off that game of make-believe.

An extreme form of this motivational division may account for various dissociative phenomena, such as multiple personality disorder (or dissociative identity disorder, as it is now called). What seem like distinct personalities may in fact be distinct self-conceptions enacted by the agent at different times. The self-conceptions involved in DID would have to differ in various respects from ordinary self-conceptions, including those involved in make-believe. They would have to be full-blown delusions – that is, conscious fantasies not recognized as such by the agent – and they would have to resist external restraint or interruption to the point that the agent had no access to the motives that he wasn't currently enacting. The resulting division in the agent's motives would be deeper than that in the motives of a sane and sober agent. But it would be a deeper version of the same fundamental division, between the motives that are being enacted and the motives that can at most modify that enactment.

As we have seen, this division in an agent's motives can lead to action that is irrational in relation to the totality of his desires and interests, as when it lets him get carried away in a debate, to his subsequent regret. But I think that the temporary irrationality of getting carried away can sometimes be exploited for more permanent gains in rationality. For an agent can get carried away with the better of his motives as well as the worse.

A colleague who studies rational choice tells me that he could never have quit smoking without indulging in some irrationality.[6] Although the long-run costs of smoking outweighed the long-run benefits, he says, the costs of smoking the next cigarette never outweighed the benefits of smoking that one cigarette, since he could always decide to quit after the next cigarette rather than before. In order to stop smoking in the long run, of course, he had to forego the next cigarette at some point, at an obvious sacrifice of utility. The only way for him to stop was thus to do something irrational. How did he manage to do it?

The answer, he tells me, was *not to think of himself as a smoker.* At the beginning, of course, not to think of himself as a smoker was incorrect, since he was

6 Thanks to Jim Joyce for this example.

still addicted to smoking, both physically and psychologically. I suggest, then, that he resorted to make-believe. He imagined that he was not addicted – that he didn't like the taste of cigarettes, wasn't in the habit of smoking them, had no craving for them – and he then enacted what he was imagining, pretending to be the non-smoker that he wanted to be. And I suggest that this make-believe succeeded because it excluded the smoker's tastes, habits, and cravings from the story that he was enacting. That story lacked the narrative background that would have made it intelligible for him to buy, light, or smoke the next cigarette.

I suggest, further, that my colleague got carried away with this make-believe, and that getting carried away was essential to his success at kicking the habit. His motives for smoking were relegated to externally constraining his enactment of a non-smoker's story. Those motives had proved irresistible when they were available at center-stage to motivate the next episode in the story; but when they were written out of the plot and left to operate, as it were, *ex machina*, they were unable to deflect the story from its natural conclusion.

I suggest, finally, that when my colleague got carried away with enacting an image of himself as a non-smoker, he was being motivated by an ideal. That's what an ideal is: the image of another person, or a currently untrue image of oneself, that one can get carried away with enacting.[7] To imagine oneself in that image, and to act accordingly, is to identify with and emulate the ideal.

An alternative to my conception of ideals would be to think of them as descriptions or images that motivate by way of one's desire to satisfy them and one's realistic beliefs about how to do so. According to this alternative conception, taking another person as one's ideal entails wanting to resemble him, which directly motivates behavior like his, conceived as a constitutive means to the desired resemblance. I doubt whether the motivational force of an ideal can flow directly from such a desire.[8]

Suppose that one idealizes a person for his generosity and wants to resemble him in this respect.[9] Insofar as this desire directly moves one to do generous things, those acts will not in fact be motivated by generosity, after all, and so one's attempted imitation of the ideal will be an obvious failure. Indeed, one would be unlikely to acquire or to learn generosity through acts motivated in this way. The desire to mold oneself in the image of a generous person will meet with better success if it moves one first to imagine being a generous person and then to enact this self-image, making believe that one is generous and using as props whatever motives one has that can be cast in the role of generosity. (Such props might be drawn from that fund of tenderness that Freud calls the libido.) Emulating generosity in this fashion, one comes closer to being and to feeling generous, and one has a better chance of becoming really generous, by gradually working one's way into the role. One can thus gradually adopt or assume the motive of generosity in

7 By "a false image of oneself", I mean a self-image that would not be true even if one enacted it. Of course, my colleague eventually became a non-smoker by pretending to be one, but at the outset his pretense was false.

8 For background to this section, see "On the Aim of Belief", in Velleman (2000a), pp. 256-72.

9 See Aristotle's discussion at *Nicomachean Ethics* 1105a ff.

a way that one never could by imitating it from the outside.

Although the desire to resemble an ideal can initiate this process only by motivating a deliberate turn toward make-believe, other attitudes can initiate it directly, because they already engage the imagination. In the former case, the desire to resemble an ideal depends for its motivational force on an assessment of how one falls short of the ideal and what one must do to close the gap. The desire may ultimately favor a process of conjuring up and enacting an idealized self-image, but only on the basis of a realistic calculation that the process will be conducive to a resemblance not yet attained. Now consider an attitude like respect or admiration for the ideal. Precisely because these attitudes are not goal-oriented motives, they tend to favor wishful thinking over purposeful activity. Admiring someone isn't a motive for bringing about anything in particular, and so it doesn't call for an instrumental calculation of the steps required to bring anything about. Wishfully picturing oneself in the image of an ideal is not a distraction from the business of admiring him: it *is* the business of admiring him. Emulation therefore flows directly out of admiration.

When a smoker draws on an ideal for motivation to quit, his behavior is in some respects irrational. He ignores various facts that would be relevant to fair-minded deliberation: the fact that he would enjoy the taste of a cigarette, that he is in habit of smoking, that he is even now craving a smoke, and so on. And he acts instead on various considerations that are figments of his imagination: that he feels fine without a cigarette, that he wouldn't enjoy one, that lighting up would be an uncharacteristic thing for him to do.

Yet his make-believe world is a world of make-believe reasons. His imaginative considerations guide him in the manner of reasons for acting, just as the facts would guide him if he acted on realistic grounds. These imaginative considerations serve as narrative premises in light of which only some actions make sense as the continuation of his story. And when an agent does what makes sense in light of a narrative premise, or *rationale*, he is acting for a reason, albeit one that isn't true.

What's more, this make-believe reasoning enables the agent to become more rational in the long run. For by pretending to be a non-smoker, he actually becomes a non-smoker, which is a more rational sort of person to be. As a smoker, he was deeply conflicted: his reasons for smoking were at odds with all of his other reasons for acting, although they were strong enough to prevail in a review of what he had reason to do next. He therefore chose to smoke, but always at the sacrifice of the many countervailing reasons that had been outweighed. In kicking the habit, he lost his reasons for smoking, leaving the field to his countervailing reasons, which can now guide his actions unopposed. Because his actual reasons have become less conflicted, he sacrifices less in doing what he actually has most reason to do.[10]

Indeed, the agent may have had sufficient reason to identify with a non-smoking ideal, even when he lacked sufficient reason to forego his next cigarette. Foregoing his next cigarette in his story as a smoker would have left the resulting dis-

10 I discuss these issues further in "Willing the Law".

comforts and inconveniences at center-stage, as salient repercussions to be faced. The second act of this story would have been "The Smoker Copes with Withdrawal" – an episode that's difficult to improvise without ending up in a third act entitled "The Smoker's Relapse." The difficulty of charting an intelligible course through the story of quitting as a smoker is what made for the rationality of continuing to smoke instead. The point of identifying with the ideal of a non-smoker was precisely to gain access to a different story, presenting a different set of reasons. That alternative story entailed not smoking the next cigarette, of course, but not smoking that cigarette was a different option for a non-smoker than it was for a smoker. For a smoker, not smoking that cigarette was a matter of changing course and facing the consequences; for a non-smoker, it was a matter of going on as usual. To be sure, the non-smoker in this case would be a merely make-believe non-smoker, who would experience twinges and shakes of what was in reality nicotine withdrawal. But those discomforts would not be expected repercussions to be faced and overcome; they would be inexplicable irritations to be ignored, if possible. And the smoker who wants to quit has good reason to prefer facing the consequent discomforts under the guise of irritations to be ignored rather than expected repercussions to be faced. Hence he had good reason for undertaking the pretense of being a non-smoker.

The smoker who wants to quit is like other agents who have reason to make themselves temporarily irrational – warriors who have reason to work themselves into a frenzy in order to frighten the enemy, or negotiators who have reason to become obstinate in order to win concessions. Unlike the warrior or the negotiator, however, the smoker does not have reason to arrange for something to interfere with his faculty for practical reasoning. On the contrary, the irrationality that the smoker has reason to cultivate requires the exercise of an intact deliberative faculty; it merely requires that faculty to operate on input from the agent's imagination rather than on his knowledge of the facts. When the agent's deliberative faculty operates in this way, he becomes insensitive to considerations that are genuine reasons for him to act, and so he becomes dispositionally irrational. And because he thereby neglects reasons against the action that he performs, he may end up performing an irrational action.

I have now argued, on the one hand, that it was rational for the smoker to undertake the activity of pretending to be a non-smoker, that this activity involved an exercise of an intact rational faculty, and that it resulted in the smoker's becoming a more rational agent. On the other hand, I have argued that the activity of pretending to be a non-smoker was irrational in the sense that it made the smoker insensitive to some of the reasons that actually applied to him, and consequently led him to do something that wasn't supported by the balance of actual reasons.

I think that such irrationality is often involved when an agent is motivated by a personal ideal – including the overarching ideals that embody Hume's general perspective or the Aristotelian virtues. Whether one is emulating an impartial observer or a virtuous human being, one may be engaged in make-believe and hence in an activity that's irrational in the respects described above.

Note, however, that I have not included Kant's Categorical Imperative in the list of moral ideals whose emulation tends to require make-believe.[11] The reason is that, in my view, Kantian moral theory manages to kick away that particular ladder.

The Categorical Imperative is an ideal image of the will, as acting on only those maxims which it can simultaneously will to be universal laws. But what moves this ideal will to act only on universalizable maxims? The answer is that it is restrained from acting on other maxims by respect for the law. And respect for the law is just respect for the Categorical Imperative, which is an ideal image of the will as acting only on universalizable maxims. To act out of respect for this ideal is therefore to emulate a will that acts out of respect for the very same ideal.

In the case of the Kantian ideal, then, emulation tends to rise to the level of attainment. What is ideal about the person we emulate is precisely that he is moved by an ideal, and indeed the same ideal by which we are moved. Hence to emulate him is already and really to resemble him, and so it is unlike emulating him with respect to a motive that doesn't rely on emulation. To do generous things by emulating a generous person is not yet to be generous, though it may be a means of learning generosity. But to do the moral thing by emulating a moral person really is to be moral, since enacting an moral image of oneself is what being a moral person consists in.

So we are not enacting a false conception of ourselves in emulating the Categorical Imperative, because we are making that conception true just by emulating it. Of course, we could get carried away with enacting that self-conception, by losing sight of our countervailing motives, so that they lapse into abeyance for want of reinforcement from our self-conception. Wouldn't we then be acting on a false self-conception and hence irrationally? Not necessarily. After all, the Categorical Imperative could be – come to think of it, I'm sure that it is –the image of a will that gets carried away with enacting that very self-image. The motivational division that underlies make-believe – the division between enacted motives and motives that externally modify such enactments – remains essential to our acting on the Categorical Imperative; but what gets enacted is not a false self-conception.

Insofar as we are Kantian moral agents, then, we are not just pretending. When we dream of our morally better selves, our dreams really can come true.

11 For my interpretation of the Categorical Imperative, see Velleman (1998, 1999).

CV

David Velleman is Professor of Philosophy at the University of Michigan, Ann Arbor. He is the author of *The Possibility of Practical Reason*, (Oxford 2000). Email: velleman@umich.edu

References

Anscombe, Elizabeth ([1957] 2000) *Intention*. Cambridge, MA: Harvard University Press.

Brakel, Linda (2002), "Phantasy and Wish: A Proper Function Account for Human A-Rational Primary Process Mediated Mentation," *Australasian Journal of Philosophy* 80: 1-16.

Harré, Rom (1979), *Social Being*. Oxford: Basil Blackwell.

Hollis, Martin (1977), *Models of Man*. Cambridge: Cambridge University Press.

Velleman, J. David (1989), *Practical Reflection*. Princeton: Princeton University Press.

Velleman, J. David (1998), "The Voice of Conscience", *Proceedings of the Aristotelian Society*, 57-76.

Velleman, J. David (1999), "A Rational Superego", *Philosophical Review* 108, 529 – 558.

Velleman, J. David (2000a), *The Possibility of Practical Reason*. Oxford: Oxford University Press

Velleman, J. David (2000b), "From Self-Psychology to Moral Philosophy", *Philosophical Perspectives* 14: 349-77

Velleman, J. David (forthcoming a), "The Self as Narrator", to appear in *Autonomy and the Challenges to Liberalism: New Essays*, ed. Joel Anderson and John Christman.

Velleman, J. David (forthcoming b), "Willing the law", in *Practical Conflicts* eds Monika Betzler and Peter Baumann, Cambridge University Press.

Velleman, J. David (ms), "A Sense of Self".

Walton, Kendall L. (1993), *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA: Harvard University Press.