

Self to Self

Selected Essays

J. DAVID VELLEMAN

New York University



CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521670241

© J. David Velleman 2006

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2006
Reprinted 2007

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data

Velleman, James David.

Self to Self : selected essays / J. David Velleman.

p. cm.

Includes bibliographical references and index.

ISBN 0-521-85429-6 (hardcover) – ISBN 0-521-67024-1 (pbk.)

1. Self. 2. Self (Philosophy) 3. Kant, Immanuel, 1724–1804 – Ethics. I. Title.
BF697.V45 2005
126–dc22 2005008114

ISBN 978-0-521-85429-0 Hardback

ISBN 978-0-521-67024-1 Paperback

Transferred to digital printing 2009

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party Internet websites referred to in
this publication, and does not guarantee that any content on such websites is,
or will remain, accurate or appropriate. Information regarding prices, travel
timetables and other factual information given in this work are correct at
the time of first printing but Cambridge University Press does not guarantee
the accuracy of such information thereafter.

Introduction

The title of this book comes from John Locke, who described a person's consciousness of his past as making him "self to himself" across spans of time. Implicit in this phrase is the view that the word 'self' does not denote any one entity but rather expresses a reflexive guise under which parts or aspects of a person are presented to his¹ own mind. This view stands in opposition to the view currently prevailing among philosophers – that the self is a proper part of a person's psychology, comprising those characteristics and attitudes without which the person would no longer be himself. I do not believe in the existence of the self so conceived.

To say that 'self' merely expresses a reflexive mode or modes of presentation is not to belittle it. The contexts in which parts or aspects of ourselves are presented in reflexive guise give rise to some of the most important problems in philosophy. They include the context of autobiographical memory and anticipation, in which we appear continuous with past and future selves; the context of autonomous action, in which we regard our behavior as self-governed; the context of moral reflection, in which we exercise self-criticism and self-restraint; and the context of the moral emotions, in which we blame ourselves, feel ashamed of ourselves, or want to be loved for ourselves. To understand what is presented to us under the guise of self in each of these contexts would be to gain some insight into personal identity, autonomy, the conscience, and the moral emotions – all important and complex phenomena.

¹ For an explanation of why I use 'he' to denote the arbitrary person, see my *Practical Reflection* (Princeton: Princeton University Press, 1989), p. 4, n. 1.

Many philosophers think that we can account for all of these phenomena at a stroke, by identifying a single thing that serves simultaneously as that which we have in common with past and future selves, that which governs our behavior when it is self-governed, that which we restrain when exercising self-restraint, and that which we blame, of which we feel ashamed, or for which we hope to be loved. I think that expecting a single entity to play the role of self in all of these contexts can only lead to confusion. Each context presents something in a reflexive guise, but not necessarily in the same guise, and certainly not the same thing.

That said, I still believe that there is much to be gained from a comparative study of selfhood in all of these contexts. Several of the essays in this volume undertake such a comparative study, while others confine themselves to selfhood in one context, with cross-references to essays about the others. The result is not a unified theory of the self, but it is, I hope, a coherent series of reflections on selfhood. In this Introduction, I will identify some of the subsidiary lines of argument uniting these reflections.

What Is a Reflexive Mode of Presentation?

Some activities and mental states have an intentional object: they are mentally directed at something. Of these, some can take their own subject as intentional object: they can be mentally directed at that which occupies the state or performs the activity. Of these, some can be mentally directed at their own subject conceived as such – conceived, that is, as occupying this very state or performing this very activity. A reflexive mode of presentation is a way of thinking that directs an activity or mental state at its own subject conceived as such.

The attitude of respect, for example, is directed at a particular person by some way of thinking about him. Sometimes it is directed at a person by the thought of him as the one holding this very attitude of respect. That way of thinking is a reflexive mode of presentation, and the resulting attitude is consequently called “self-respect.” In the simplest case, the reflexive mode of presentation is a first-person pronoun: the object of some respectful thought is picked out in that thought as “me,” and then the “self” in “self-respect” is just an indirect way of attributing an attitude that would be directly expressed with the first person. But there are also non-verbal modes of reflexive thought.

For example, a visual image represents things in spatial relation to an unseen point where its lines of sight converge. Insofar as vision implicitly

alludes to that point as the position of its own subject, its geometry constitutes a reflexive mode of presentation. Being visually aware of things involves being implicitly self-aware, because it involves this implicit way of thinking about the subject of vision as such. The reflexivity implicit in this awareness would naturally be expressed in the first person, with a statement beginning “I see. . . .” But what makes the awareness reflexive, to begin with, is not a use of the first-person pronoun. What makes visual awareness implicitly reflexive is the perspectival structure of the visual image, which secures the implicit reference to the subject of vision so conceived.

Whenever the self is spoken of, some reflexive activity or mental state is under discussion, with the word ‘self’ standing in for the mode of presentation by which the state or activity is directed at its subject as such. Strictly speaking, then, reference to the self *sans phrase*, in abstraction from any reflexive context, is incomplete. Talk of “The Self” is like talk of “The Subject” in that theory-laden sense which refers to a person in the abstract. Just as The Subject must be the subject of some activity or mental state, so The Self must be the self of some activity or mental state directed at its subject so conceived.

Talk of the self *sans phrase* can be harmless, of course, if the relevant state or activity is salient in the context. And some reflexive states and activities are of such importance to our nature that they can be made salient by little more than reference to the self. But our failure to specify a reflexive context when speaking of the self should not be taken to indicate that there is nothing to specify.

I distinguish among at least three reflexive guises under which a person tends to regard aspects of himself. These three reflexive guises correspond to at least three distinct selves.

First, there is the self-image by which a person represents which person and what kind of person he is – his name, address, and Social Security number, how he looks, what he believes in, what his personality is like, and so on. This self-image is not intrinsically reflexive, because it does not in itself represent the person as the subject of this very representation; in itself, it represents him merely as a person. It is made reflexive by some additional indication or association that marks it as representing its subject. It is like a photograph in the subject’s mental album, showing just another person but bearing on the reverse side “This is me.”²

² I discuss this issue further in “The Centered Self” (Chapter 11). See especially Appendix A.

A person's self-image cannot be intrinsically reflexive, in fact, if it is to embody his sense of who he is. Conceiving of who he is entails conceiving of himself as one of the potential referents for the pronoun 'who', which ranges over persons in general. From among these candidates neutrally conceived, it picks out the one he is, thus identifying him with one of the world's inhabitants. It therefore requires a conception of someone *as* one of the world's inhabitants, who can then be identified as "me."

Because a person's sense of who he is must contain a non-reflexive conception of himself as one of the world's inhabitants, it is the vehicle for those attitudes by which he compares himself to others or empathizes with their attitudes toward him. When he feels self-esteem, for example, he feels it about the sort of person he is, and hence toward himself as characterized by his self-image. When he indulges in self-hatred, he hates the object of his self-image, a person whom others might hate. As the repository of the characterizations grounding these self-evaluations, the self-image is sometimes referred to as the person's ego – not in the psychoanalytic sense but in the colloquial sense in which the ego is said to be inflated by praise or pricked by criticism. An inflated ego, in this colloquial sense, is an overly positive self-image.

Finally, a person's self-image is the criterion of his integrity, because it represents how his various characteristics cohere into a unified personality, with which he must be consistent in order to be self-consistent, or true to himself. Failures of integrity threaten to introduce incoherence into the person's conception of who he is; and in losing a coherent conception of who he is, the person may feel that he has lost his sense of self or sense of identity. This predicament is sometimes called an identity crisis.

When someone suffers an identity crisis, he may feel that he no longer knows who he is. The reason is not that he has forgotten his name or Social Security number; it's rather that the self-image in which he stores information about the person he is has begun to disintegrate under the strain of incoherence, either with itself or with his experience. Often such strain appears around features of his self-image that distinguish him from other persons and underwrite his self-esteem. The result is that his self-image seems to lose its power to set him apart from others in his eyes; and this result is what he is speaking of when he says that he no longer knows who he is.

Yet to say that a person has undergone an identity crisis, or no longer knows who he is, does not imply that there is any doubt, in our minds or in his, as to whether he is still the same person. His identity crisis is a crisis in his sense of identity, as embodied in his self-image; it is not a

crisis in his metaphysical identity – that is, in his being one person rather than another, or one and the same person through time. The qualities that are distinctive of the person, either descriptively or evaluatively, are crucial to his sense of who he is because that sense is embodied in a self-image representing him as one person among others, from whom he then needs to be distinguished by particular qualities. The fact that distinctive qualities are necessary to pick out the person who he is, and thus inform his sense of identity, does not indicate that those qualities play any role in determining his identity, metaphysically speaking.

Unfortunately, philosophers sometimes assume that the qualities essential to a person's sense of who he is are in fact constitutive of who he is and therefore essential to his remaining one and the same person, numerically identical with himself and numerically distinct from others. Here they conflate the self presented by a person's self-image with the self of personal identity, or self-sameness through time.

Self-sameness through time is the relation that connects a person to his past and future selves, as they are called. In my view, past and future selves are simply past and future persons in reflexive guise, or under a reflexive mode of presentation.³ The task of identifying a person's past and future selves is a matter of identifying which past and future persons are accessible to him in the relevant guise, or under the relevant mode of presentation – in short, which past and future persons are reflexively accessible to him. Past persons are reflexively accessible via experiential memory, which represents the past as seen through the eyes of someone who earlier stored this representation of it; and future persons are accessible via a mode of anticipation that represents the future as encountered by someone who will later retrieve this representation of it. These modes of thought portray past and future persons reflexively by implicitly pointing to them at the center, or origin, of an egocentric frame of reference, as the unseen viewer in a visual memory, for example, or the unrepresented agent in a plan of action. The unseen viewer in a visual memory is the self or "I" of the memory; the unrepresented agent in a plan of action is the self or "I" of the plan. Past and future selves are simply the past and future persons whom the subject can represent as the "I" of a memory or the "I" of a plan – persons of whom he can think reflexively, as "me."

These reflexive modes of thought are significantly different from the self-image that embodies a person's sense of self. To begin with, they

³ This claim is the thesis of "Self to Self" (Chapter 8).

are intrinsically reflexive, in the sense that their representational scheme is structured by a perspective whose point of origin is occupied by the past or future subject, whereas a self-image is the representation of a person considered non-first-personally but identified as the subject by some other, extrinsic means. Another difference lies in the extent to which these modes of thought actually constitute the self.

I have long defended the view that a person's self-image is self-fulfilling to some extent: thinking of himself as shy, or as interested in jazz, or as aspiring to cure cancer can be a part or a cause of his actually being shy, or being interested in jazz, or aspiring to cure cancer. Including these characteristics in his self-image can be partly constitutive of, or conducive to, possessing them in fact; and to this extent, the person can define himself by defining his self-image. I elaborate on this view of self-definition in several of the essays in this volume.⁴ As I point out, however, a person's powers of self-definition are limited. Although thinking that he has a characteristic can be one part or one cause of his actually having it, other parts and causes are invariably required. And although the self-image through which he defines himself can also be said to embody his sense of who he is, the fact of who he is lies strictly beyond his powers of self-definition. Thus, thinking that he is interested in jazz may or may not succeed in making him interested in jazz, while thinking that he is Napoleon will certainly fail to make him Napoleon.

By contrast, someone's first-personal memories and expectations determine which past and future persons are accessible to him in the guise of selves; and as Locke first pointed out, we have good reason to acknowledge connections of selfhood forged in this manner, whether or not they conform to the life history of a single human being. Such diachronic connections are the topic of the title essay in this volume (Chapter 8). There I argue, in support of Locke, that if a person could retrieve experiential memories that were stored by Napoleon at Austerlitz, then Napoleon at Austerlitz would genuinely be related to him as a past self; and when he reported one of those memories by saying "I commanded the forces at Austerlitz," he would be expressing a thought that helped to constitute its own truth, by giving him first-personal access to the relevant inhabitant of the past.

In sum, a person's identity is constituted by reflexive thought in two distinct instances. In the first instance, the person can to some extent

⁴ Empirical evidence for this view is summarized in "From Self-Psychology to Moral Philosophy" (Chapter 10). The view also figures in "The Self as Narrator" (Chapter 9), "The Centered Self" (Chapter 11), and "Motivation by Ideal" (Chapter 13).

fashion his own identity, because he can fashion his self-image and at the same time fashion himself in that image. In the second instance, the person's identity is given to him by the psychological connections that make past and future persons accessible to his reflexive thought.

The third reflexive guise under which a person is presented with a self is the guise of autonomous agency.⁵ Among the goings on in a person's body, some but not others are due to the person in the sense that they are his doing. When he distinguishes between those which are his doing and those which aren't, he appears to do so in terms of their causes, by regarding the former but not the latter as caused by himself. Yet even the latter goings-on emanate from within his own body and mind, and so when he disowns them, he ends up disowning parts of his own body and mind, as if the boundary between self and other lay somewhere inside the skin.

I think that in order to locate the self to whom autonomous actions are attributed, we have to ask which part or aspect of the person is presented to him in reflexive guise when he considers the causes of his behavior. Whatever is presented in reflexive guise to the agent's causal reasoning will be that to which such reasoning attributes his behavior when attributing it to the self. Clearly, what's presented in reflexive guise to causal reasoning is that which conducts such reasoning – that part or aspect of the person which seeks to understand events in terms of their causes. The self to which autonomous actions are attributed must therefore be the agent's faculty of causal understanding. Insofar as a person's behavior is due to his causal understanding, its causes will appear to that understanding in reflexive guise, and the behavior will properly appear as due to the self.

Most of my work prior to the essays in this volume was devoted to arguing that the actions traditionally classified as autonomous by philosophers of action are indeed due to the agent's causal understanding.⁶ Autonomous actions are actions performed for a reason, and reasons for performing an action, I argued, are considerations in light of which the action would be understandable in the causal terms of folk psychology. To act for a reason is to do what would make sense, where the consideration in light of which it would make sense is the reason for

⁵ The self of autonomy is the topic of "The Self as Narrator" (Chapter 9) and "Identification and Identity" (Chapter 14).

⁶ See my *Practical Reflection* (Princeton: Princeton University Press, 1989) and *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000).

acting. Thus, for example, one's being interested in jazz would explain why one might frequent nightclubs, and so one can frequent nightclubs not only out of an interest in jazz but also on the grounds of that interest, regarded as explanatory of one's behavior. When one's behavior is guided by such considerations, it is guided by one's capacity for making sense of behavior, which is one's causal understanding and is therefore presented in reflexive guise to that very understanding, as the self that causes one's behavior.

The essays in this volume elaborate on this theory of autonomy in a few, fairly modest respects. First, I explore what social psychologists have written about the self, pointing out that their research supports the aspect of my theory that seems most far-fetched to philosophers – namely, the claim that people are generally guided in their behavior by a cognitive motive toward self-understanding.⁷ Second, I point to this motive as effecting a crucial, hidden step in the process posited by Daniel Dennett to explain how a human being makes up or invents a self.⁸ I agree with Dennett in thinking that a human being makes up or invents a self in one sense; but I argue that in making up a self in that sense, a human being also manifests his possession of a self in another sense, by exercising genuine autonomy. The self that a human being makes up is the individuating self-conception that embodies his sense of who he is; the self that he thereby manifests is his capacity for understanding his behavior in light of that self-conception.

Dennett frames his notion of self-invention in terms of self-narration: the self-conception that a person develops is a sketch for the protagonist in his own autobiography. In these terms, the person's capacity for causal understanding gets redescribed as his capacity for coherent narration, which I call the self as narrator. In two further essays, I go on to explore implications for moral philosophy flowing from this narrative-based theory of autonomy.⁹

This completes my summary of the three reflexive guises under which we are presented with selves: the self-concept, the guise of past or future self,

⁷ "From Self Psychology to Moral Philosophy" (Chapter 10).

⁸ "The Self as Narrator" (Chapter 9).

⁹ "Willing the Law" (Chapter 12) and "Motivation by Ideal" (Chapter 13). In all of these essays, I assume that narrative is just a way of formulating our causal understanding of the narrated events. I have recently come to doubt this conception of narrative ("Narrative Explanation" *The Philosophical Review* 112 [2003]: 1–25). Although narrative conveys causal understanding of the narrated events, I have come to think that it also conveys a distinct mode of understanding as well. This conclusion complicates my view of practical reason in ways that remain to be explored.

and the guise of the self as cause of autonomous action. As I mentioned at the beginning, my strategy of identifying distinct selves, corresponding to these distinct reflexive guises, runs counter to the prevailing trend among philosophers, who prefer to theorize about a single, all-purpose self. I now turn to a summary of the arguments by which I attempt to resist this trend. I interpret the trend as a reaction against Kantian moral psychology, and so my arguments are largely interpretations and defenses of Kant.

In Kant's moral psychology, the governing *autos* of autonomy is rational nature, which a person shares with all persons. This rational nature includes none of the qualities that differentiate the person from others, none of the idiosyncratic attitudes and characteristics that inform his sense of individuality. It is therefore unfit to serve as the target of reflexivity in other contexts – as the target of self-esteem, for example – and so it strikes many philosophers as denuded, the mere skeleton of a self. These philosophers have consequently sought to flesh out a rival conception of the self that includes personal particularities, and they have then deployed this conception not only in contexts to which it is appropriate, in my view, but in others as well, including the contexts of personal identity and autonomy. I pursue three distinct strategies for resisting this trend, though I don't always distinguish among them.

First, I attempt to meet the trend head-on by arguing that it underrates the importance of bare personhood. I grant that each person has a detailed sense of his identity, representing those features of himself which he values as differentiating him from others. This individuating self-conception is that to which the person is true when he is true to himself, that which he betrays when he betrays himself, and that under which he esteems himself in feeling self-esteem. The distinctive features represented in this conception can even be said to define who the person is. Yet these features are not, for example, the object of the person's self-respect, since self-respect is an appreciation of his value merely as a person. Whereas self-esteem says "I am clever" or "I am strong" or "I am beautiful," self-respect says simply "I am somebody."

Of course, each person is not merely somebody but a concrete individual, and the qualities that flesh out his individuality are, as I have just granted, the focus of some reflexive attitudes, such as self-esteem. But the fact that some reflexive attitudes bear on the person's distinctive features does not entail that all such attitudes must do so as well, because there isn't a single thing on which all reflexive attitudes must bear. Assuming otherwise inevitably leads to underrating the importance of being somebody. Who I am, in particular, matters for many reflexive purposes;

but if all that mattered for reflexive purposes was who I am, then it would no longer matter that (as Dr. Seuss so wisely put it) I am a Who.

In two of these essays, I argue that the importance of being somebody is registered in human emotions that are often analyzed by philosophers as concerned with personal distinctiveness – namely, love and shame.¹⁰ The ordinary thought about love, reflected in most philosophical work on the subject, is that we love one another and want to be loved for who we are, in the sense of the phrase that I have just been using to invoke the qualities that differentiate us from others. Those same qualities are thought to be the basis for the negative emotion of shame.

I agree that personal distinctiveness is often in our sights when we feel shame, and always when we feel love, and I try to analyze precisely how it figures in these emotions. I argue, however, that its role is dependent on, and indeed unintelligible without, the role of bare personhood.

In my view, shame is anxiety that we feel about a threat to our socially recognized status as self-presenting creatures, a status that ultimately rests on the structure of a free will, in virtue of which we qualify as persons. This threat can arise from the exposure of particular discreditable qualities, of which we are then said to be ashamed, but it can also arise in the absence of any perceived demerit. We can therefore feel shame without there being anything about us of which we are ashamed. Such inchoate shame, I argue, is what we felt as children when pressed to perform for household guests, what we felt as adolescents when seen by our peers in the company of our parents, and what we feel as adults when subjected to various kinds of unwelcome attention ranging from racist epithets to excessive praise. These instances of shame are possible, I claim, because the object of anxiety in shame is not our distinctive personality but rather our social standing merely as self-presenting persons. Hence understanding shame requires acknowledging the importance of being somebody – in this case, the importance of being somebody to others.

Being somebody to others is also at the bottom of being loved, in my view. We often say that we want to be loved for who we are, again using that phrase which alludes to our particularities. Yet there is an ambiguity in the preposition that introduces this phrase – the ‘for’ in “for who we are.” Personal love is an essentially experiential emotion: it’s a response to someone with whom we are acquainted. We may admire or envy people of whom we have only heard or read, but we can love only the people

¹⁰ “Love as a Moral Emotion” (Chapter 4) and “The Genesis of Shame” (Chapter 3).

we know. So there is no question but that personal qualities experienced directly or indirectly – appearance, manner, words, actions, traits of character, and so on – are essential to eliciting love. The question remains, however, whether the love that's elicited by these qualities is an emotion felt toward or about those same qualities. Loving someone is a way of valuing him, but are we valuing him on the basis of those qualities that elicit our love? What is it to love someone *for* the way he walks and talks, the way he holds his knife and sips his tea, or (more loftily) for who he is?

I argue that to love someone for the way he walks or the way he talks is not to value him on the basis of his gait or his elocution; it's rather to value his personhood as perceived through them. The qualities that elicit our love are the ones that make someone real to us as a person – the qualities that speak to us of a mind and heart within – and the value that is registered in our love is therefore the value of personhood. Wanting to be loved is like wanting to be found beautiful: it's a desire that others be struck by our particularities, but in a way that awakens them to a value in us that is universal.

This account of love, like my account of shame, is an attempt to cope with paradoxes inherent in our ordinary understanding of the emotion. In the case of shame, the paradox is what I have called “inchoate shame,” in which we are shamed without there being anything that we are ashamed of. The paradox in the case of love is that, although it is a way of valuing people, it doesn't conform to any readily intelligible evaluations of them or value judgments about them.

Thus, I love my own wife and children as no others, and yet I know that other women and children are equally worthy of being loved by their own husbands and fathers. I do not honestly believe that mine are better or preferable; I don't even believe that they are better or preferable for me, as romantic soul-mates are supposed to be. Yet I treasure them above all. How can I value them especially without perceiving a special value in them? How can I believe that everyone, in deserving to be loved, deserves to be valued as special, if no one is especially deserving in this respect?

I am well aware that my view of love can be made to sound soft-headed and silly. Readers of my view sometimes think they can simply dismiss it with the remark that everyone knows love isn't like *that* – as if I did not already acknowledge the initial implausibility of the view. My response to these readers is that what “everybody knows” about love is deeply

problematic, as most children begin to suspect by the age of five or six, once they are told, for example, that everyone is special. If what we are taught to find plausible about love made any sense upon reflection, then philosophizing about love would be as pointless as philosophizing about humor or the weather. In fact, the truth about love had better be something fairly implausible to us, or the emotion itself will turn out to be absurd. Those who aren't troubled by the conventional wisdom will see no need for anything else; but then they should see no need for philosophy, either.

My second line of argument against the doctrine of a single, all-purpose self – which I interpret, in turn, as a reaction against Kantian moral psychology – is to humanize the latter theory. At the center of Kantian moral psychology is the attitude of respect for the law, which many readers and teachers of Kant interpret as deference to a purely formal rule of conduct, or the abstract concept of such a rule. This interpretation makes the moral agent appear to be fixated on a mere abstraction, as if lost in impersonal thought; and one natural reaction against this alienated conception of the moral agent is to insist that his attention be focused, not on abstract rules, but on particular people instead. I argue that Kant actually holds an intermediate view, which portrays the moral agent as attending neither to rules nor to particular people but to an ideal of the person.

In particular, I argue that respect for the law is respect for an ideal image of oneself: it's what Freud would describe as admiration for an ego-ideal.¹¹ The ego-ideal in Kantian ethics is that rational configuration of the will which is represented in the Categorical Imperative. The point is that admiring an ego-ideal is not a way of getting lost in thought; it's a way of finding oneself. The Kantian moral agent can therefore be seen as less of a space cadet and more of a well-centered person.

What's more, the Kantian moral motive – respect for the law – can be seen as a motive that would naturally develop out of our experience as particular people among others. According to Freud, admiration for an ego-ideal arises from love for the real people after whom the ideal was fashioned – parents or their surrogates, in most cases. My account of love enables me to explain how the love that we felt for our parents in

¹¹ I argue for this interpretation of Kant in "Love as a Moral Emotion" (Chapter 4) and "The Voice of Conscience" (Chapter 5).

childhood might give rise to respect for the rational will as represented in the Kantian ego-ideal. Love for our parents was our response to their loving care, in which they treated us as self-standing ends – a configuration of their wills that we then incorporated into an ego-ideal, for which we continue to feel the admiration that amounts, in my view, to Kantian respect for the law. Kantian respect for the law can thus be learned from the love between parent and child, which Freud was surely right to identify as the textbook for our moral education.¹²

In this second line of argument, I consider myself to be interpreting what Kant actually says. In the third line of argument, I propose a revision of Kant's moral theory, as I understand it, thus making a strategic concession to the current trend.¹³

Kant insists that immoral action is always contrary to practical reason, and this insistence seems insensitive to the many ways in which people's peculiar interests and commitments can give them reason to act immorally. If practical reason required the moral course of action on every occasion, it would sometimes require people to step outside the personal characteristics that define who they are. Although morality may demand such self-transcendence (or self-betrayal), practical reason does not, and so I propose to modify the Kantian view.

What practical reason requires, I argue, is that people develop interests and commitments that would not give them reason to act immorally; but if they develop their interests and commitments irrationally, then they may find themselves with reason to act immorally, after all. Self-transcendence is possible in such cases, with the help of ideals of the sort that are embodied in the moral law, according to my interpretation of Kant; but self-transcendence in these cases always involves some irrationality, contrary to orthodox Kantian doctrine.¹⁴

Note that in this third line of argument, I again grant that 'self' sometimes refers to a constellation of traits that, as I have put it, define who someone is. These traits constitute a person's identity in that understanding of the term in which a person's identity is his *sense* of identity, as embodied in his self-conception. In this context, I agree with the currently

¹² This is the ultimate conclusion of "A Rational Superego" (Chapter 6).

¹³ I argue for this revision in "Willing the Law" (Chapter 12) and "A Brief Introduction to Kantian Ethics" (Chapter 2).

¹⁴ See "Motivation by Ideal" (Chapter 13).

prevailing view that the self is rich in particularities, the qualities that differentiate one person from another. I merely deny that what serves as the self in this context is what serves as the self in all contexts.

My advocacy for Kantian moral psychology in some of these essays may seem to conflict with my advocacy in others for my own, more naturalistic theory of agency. Yet I believe that these two conceptions of ourselves, though different in spirit and vocabulary, are at bottom compatible and will eventually submit to unification. Let me conclude this Introduction by speculating as to how they might be unified.

To begin with, my theory of agency adopts the Kantian strategy of deriving normative conclusions in ethics from premises in the philosophy of action. I look for rational pressures toward morality in the nature of reasons for acting; and I explore the nature of reasons by considering what would make acting for reasons an exercise of self-governance, or autonomy.

As I mentioned earlier, I identify the self of self-governance with the faculty of causal reasoning, by which a person understands the determinants of his behavior. When the person's causal reasoning helps to determine his behavior, his understanding of its determinants becomes inescapably reflexive, so that his behavior turns out to be determined by something inescapably conceived as self.

The way in which a person's causal reasoning helps to determine his behavior, in my view, is by inclining him toward behavior of which he has an incipient causal understanding – behavior that he is already prepared to understand as motivated by his desires, expressive of his beliefs, guided by his intentions, and so on. That he has those desires, beliefs, and intentions is reason for him to do the things that he could understand as partly determined by them, because reasons for doing something are considerations in light of which doing it would make sense.

There is nothing remotely like this conception of reasons for acting in Kant's moral psychology. Yet the considerations that qualify as reasons, according to this conception, meet the Kantian requirement of being recognizable from a universally accessible perspective – namely, the perspective of causal understanding. What's more, they belong to a mode of reasoning that abhors exceptions, as does practical reason, according to Kant. In one of the following essays, I try to show how the causal self-understanding that guides practical reason, as I conceive it, militates against making an exception of oneself, by way of something

like a Kantian contradiction in conception.¹⁵ In another essay, I consider how the same mode of reasoning militates against something like a Kantian contradiction in the will.¹⁶

Naturalism in moral psychology has traditionally been associated with Hume. But we can be naturalists without settling for Hume's impoverished conception of human nature. I believe – though I don't pretend to have shown – that we can be naturalists while preserving the moral and psychological richness of Kant.

¹⁵ "The Centered Self" (Chapter 11).

¹⁶ "Willing the Law" (Chapter 12).