Introduction*

JON ELSTER

The idea that the individual person may be seen as – or actually is – a set of sub-individual, relatively autonomous 'selves' has a long history. The contributions to the present volume explore this idea in the light of recent developments in philosophy, psychology and economics. The conceptual strategies that have been used to make sense of this perplexing notion differ in many ways: with respect to how literally the notion of 'several selves' is taken, with respect to the principles of partition and with respect to the modes of interaction between the systems.

Some theories take the notion of a split self very literally, to the point of postulating different physical ('hardware') bases for the subsystems. Split-brain theories (Section VIII) are the most prominent example. Thomas Schelling (1983, pp. 95–6), in a speculative digression, suggests that 'the human being is not best modeled as a speculative individual but as several alternates according to the contemporary body chemistry. Tuning in and tuning out perceptual and cognitive and affective characteristics is like choosing which "individual" will occupy this body and nervous system.' This suggests a division of the self by different programmes ('softwares') using the same neurophysiological substrate. Somewhat further down on the scale of literalness is the Freudian theory of id, ego and superego. In some readings of Freud these are understood to be distinct and autonomous entities in a very strong sense, but in Ainslie's version they turn out to be little more than a manner of speaking (Section VII). The theories of self-deception and weakness of will offered by Davidson, Pears and Rorty in this volume retain some of the literal connotations of the

*I am grateful to Amos Tversky for helpful discussions.

divided self, but do not suggest that the relatively autonomous systems are durable, stable entities that have distinct functions in the life of the mind. The concept employed by Steedman and Krause in their contribution is even weaker. Their 'selves' are, as they say, more like aspects than like agents.

The two main strategies for concept formation in this field rely on, respectively, interpersonal and intertemporal phenomena to make sense of the notion of several selves. The obvious first idea is to ask whether subsystems within a person can relate to one another in ways analogous to the relations between different persons. This may amount to postulating a set of selves with different interests but similar status or force (Section III) or to a more asymmetrical notion of a hierarchical self (Section IV). Special cases are Freud's theory (Section VII) and the idea that *homo economicus* and *homo sociologicus* cohabit our minds (Section IX). A different strategy is to look at different 'time-slices' of the same person as so many selves (Section V). Again, this may or may not involve a hierarchy of agents. Finally one may toy with the notion of 'parallel selves' (Section VI), a notion explored in the contributions by Ainslie, Elster and Schelling.

Another way of differentiating between the approaches is by looking at the form of interaction postulated between the subsystems. I shall be arguing that *deception* and strategic *manipulation* are the central forms of interaction, the former having as its immediate goal to induce a belief and the latter to induce an action. Since one way of inducing an action is by inducing a belief, the two categories do not exclude one another, but manipulation can also take place by acting on the motives or directly on the opportunity set. Self-deception has to be an asymmetrical relation, in the sense that one could hardly have two subsystems mutually deceiving one another (Section IV). By contrast, the possibility of mutual strategic interaction ought at least to be considered, although I shall conclude that it is hardly plausible. Indeed, the unity of the multiple self may stem from such asymmetries.

I begin and end by considering some borderline cases at opposite ends of the spectrum. Some cases of split selves turn out to be little more than lack of integration or coordination (Section I). The left hand may not know what the right hand is doing, but this is not to say that the latter is fooling it (cp. also Section VIII). An extreme version of the multiple-self theory, on the other hand, is that of the infinitely

fragmented self (Section X). Hume had trouble finding more than a bundle of perceptions in his search for the self; a similar 'no self' theory is proposed in the Buddhist ideas discussed by Kolm in his contribution.

I. The loosely integrated self

Many apparent cases of a split or divided self turn out to be little more than failures of coordination and integration. Or what at any given occasion looks like a fissure in the unity of the self may only be a by-product of patterns that in the long run ensure the highest degree of unity. An analogy with the firm may be instructive.[1] Subunits within a firm may achieve considerable independence and autonomy. One subunit may proceed on the basis of information that another unit already knows to be outdated. In spite of the knowledge that this is liable to happen, the direction of the firm may still decide that the overall value of independent subunits outweighs the loss of efficiency. This may even hold if it anticipates that because of jealousy between subunits some of them may actively try to hide part of what they know in order to trap other units into making bad decisions. In the case of the self, one hypothetical analogue to the 'direction' is natural selection. That is, it might be the case that our cognitive and affective apparatus is an optimal package solution, given the constraints on what the nervous system could support and the goal of maximizing fitness.[2] This, however, is necessarily very speculative. Although one can tell a story of a sort to rationalize apparently suboptimal behaviour, both the existence of the alleged benefits and their explanatory power are often dubious.[3] Another possibility is that the direction could be a central planning agency within the person, delegating the less important tasks to habits and subroutines, knowing that this may occasionally lead one astray, but believing that on balance the outcome will be better than if the full power of the mind were brought to bear on every issue (since

[1] See Margolis (1982, pp. 42–3) for one use of this analogy. Classically, of course, the analogy was used the other way around. The firm was assumed to be a unitary actor with consistent goals, on the model of the rational individual agent.

[2] For some speculations along this line, see Nisbett and Ross (1980, pp. 191ff), citing Goldman (1978).

[3] For a spelling out of these doubts, see Elster (1983, pp. 157ff).

the full power would then at any given occasion be smaller).[4] This is close to what is argued by Amélie Rorty in her contribution below. Or, a third possibility, there may be no direction that bears responsibility for the lack of coordination. There might be less, or more, coordination than a rational direction would have chosen.

Let me survey some typical examples of coordination failures, beginning with beliefs. Contradictory beliefs may coexist peacefully for a long time, if they belong to different realms of life. As a child (and even a bit longer) I had two different beliefs concerning the origin of hot water in our house. On a practical level I believed, indeed knew, that the hot water came from a heater in the basement. There was not enough for everybody to have a bath in the morning, so we followed the operations of the heater with some attention. In addition I entertained the theoretical view that beneath the streets there ran two parallel sets of water pipes, one for hot and one for cold water. One day the two beliefs, hitherto separate, came into contact with each other, upon which one of them crumbled, never to be seen again.

Somewhat similar examples are provided by the child who believes in Santa Claus, yet asks the parents about the price of the Christmas gifts; by the Ethiopians who believe that leopards, being Christian animals, will never attack their domestic beasts on a day of fasting, yet do not fail to secure their enclosures on such days; by the Romans who believed in the divinity of their rulers, yet on important family occasions always turned to their traditional gods.[5] In such cases it is unclear whether we are dealing with different modalities of belief, or with separate beliefs that guide different spheres of life. On neither interpretation is there any need to postulate a split self. Nor do we need to make this strong assumption in cases where the formation of one belief, for which we have good evidence, is blocked by a strong a priori conviction incompatible with it. The television programme, *Candid Camera*, once recorded persons sitting on a bench in Central Park who suddenly saw a tree on the edge of their visual field walking towards them. Most reacted by shaking their head as if waking from a

[4] 'A system – any system, economic or other – that at *every* given point of time fully utilizes its possibilities to the best advantage may yet in the long run be inferior to a system that does so at *no* given point of time, because the latter's failure to do so may be a condition for the level or speed of long-run performance' (Schumpeter 1961, p. 83).

[5] For these and similar examples, see Veyne (1978, pp. 248, 561, 589, 669) and Veyne (1983, *passim*).

bad dream, and then went back to whatever they were doing. The thing couldn't happen, so it didn't happen. This is more like a sound piece of Bayesian reasoning than like self-deception. (But when the tree moved again, some left their bench to sit elsewhere, as if to escape from this persistent waking dream. This calls for a more complex analysis.)

Consider next some issues of motivation. An individual's preference can be inconsistent in various ways that do not imply any kind of split self. It is possible to make a person prefer A over B and C over D, even if A is essentially the same option as D, and B the same as C. For instance, 'Mr H. mows his own lawn. His neighbor's son would mow it for $8. He wouldn't mow his neighbor's same-sized lawn for $20' (Thaler 1980). The proposed explanation for this phenomenon is that people value out-of-pocket expenses differently from opportunity cost, thus creating a normatively unjustifiable presumption in favour of the status quo. Although this particular example may yield to another explanation (Section IX), many other cases certainly fit this distinction. Thus, credit card customers may be less deterred by a cash discount to non-users than by a surcharge to users, even if the two are substantively the same (Thaler 1980). If in such cases it is possible to induce preference reversal, it is not because two parts of the person have different preferences. Rather it is because *the* person reacts to the way in which the options are presented, and not simply to their substantive content.

Sometimes these phenomena occur as the result of mental compartmentalization, in the following sense. Whether as the outcome of deliberation or not, people often keep a mental account of their expenses and avoid spending too much in any given category. They may have, say, one attitude to money spent on going to the theatre and another to accidental losses of money. Thus if I go to the theatre to pick up a ticket costing $10 and on my way lose a ten-dollar bill, this may not stop me from going, but if I have bought the ticket and then lose it, I might not want to buy another one (Tversky and Kahneman 1981). Yet the two scenarios are substantively equivalent. Such 'framing' phenomena may seem irrational, yet some mental rules of thumb of this kind are often useful to facilitate decision-making. Compartmentalization allows for preference reversal, but this is not to say that the mind has separate compartments with different preferences.

II. Self-deception and weakness of will

In the philosophical literature these are the paradigmatic examples of a divided self (Davidson 1980, ch. 2; Pears 1984). In the present volume, four contributions deal with self-deception (Davidson, Elster, Pears, Quattrone–Tversky), one with weakness of will (Ainslie) and one with both (Rorty). Self-deception may also be involved in what, according to Ainslie, is the way in which people overcome their impulsiveness. This will be brought out by comparing his argument to the Quattrone–Tversky analysis of voting. Otherwise I shall not attempt to summarize the analyses, except for some classificatory remarks.

When philosophers refer to 'the problem or weakness of will' and 'the problem of self-deception', they usually have in mind the question how these phenomena are at all possible. Davidson and Pears have pioneered in offering non-mythical answers to that question. When non-philosophers refer to these problems, they are more likely to have in mind the question how weakness of will and self-deception can be overcome. Both questions turn upon the notion of the divided self. For these paradoxical phenomena to be possible, there must be some breakdown of internal communication in the mind. To restore communication, or to prevent the defective lines from doing serious damage, some further action is required. Whether this also needs a separate, further agent is more doubtful. While it might appear that a third party is needed to prevent the subversive action of one part of the self against another, it is more plausible to identify the referee with one of the parts – but operating at a different time. I return to this issue in several later sections.

Weakness of will, as traditionally conceived, is a problem of impulsive behaviour. It is clear, however, that impulsiveness is neither sufficient nor necessary for weakness of will. It is not sufficient, since the totally impulsive person, in whom there is no inner conflict, cannot be subject to weakness of will. That notion requires both that there is a conflict between two opposed wishes, and that the wish that the person himself judges to be the more decisive loses out. Nor does weakness of will always take the form of giving in to impulsive urges. As noted by Davidson (1980, p. 30), compulsive, rigid, rule-governed behaviour can also be a form of weakness of will, that is, acting against one's own

better judgment. This establishes a conceptual connection by subsuming the apparently opposite concepts of impulsive and compulsive behaviour under a common heading. The important insight offered by Ainslie is that there is a causal connection as well, in that compulsive behaviour may be seen as the overly successful attempt to control impulsiveness (Section VI).

Self-deception is one of a family of notions that also includes wishful thinking and other forms of improper influence of wishes on belief formation. Let me make two distinctions here. The first is between the mental operations that, however irrational, at least provide some fleeting satisfaction or pleasure, and those that do not even have that redeeming feature. The pleasure principle may be second best to the reality principle, but it does at least provide some pleasure, however precarious and short-lived. But what shall we say about the congenital pessimist, who constantly believes the world to be different from what he would wish it to be? Here the wirings of the pleasure machine have gone seriously wrong. (A similar distinction can be made with respect to preference formation, see Elster 1983, pp. 111–12.) The other distinction concerns two different operations of the pleasure principle. In addition to self-deception, which necessarily requires some duality in the mind, there can be wishful thinking that is a form of 'motivated irrationality', yet does not involve any kind of duality. A person might entertain a belief out of wishful thinking, and yet *that very same* belief might also be justified by the evidence available to him, had he only considered it. If the will to believe is strong, the process of evaluating the evidence may never start up at all. Bias may give rise to beliefs that by accident turn out to be not only true (which is irrelevant here), but unbiased. I confess to some uncertainty here. The cases which I have elsewhere described as demonstrating this possibility (Elster 1983, pp. 150–1), might turn out to be better characterized by saying that the person forms the belief by considering the evidence, but that he would have formed the same belief even had the evidence pointed in a different direction.[6]

To see the connection between impulse control and self-deception in

[6] Pears (1984, pp. 94ff) points to the difference between irrational intervention and failure to intervene rationally, and argues that irrational belief formation is to be explained in terms of the latter. I believe this captures many central cases, but I am still not sure it covers all important instances.

Ainslie's contribution, it is useful to begin by considering a closely analogous phenomenon in Quattrone and Tversky's chapter. One of their findings is that people sometimes fool themselves into thinking that voting can be instrumentally justified, even when the terms of the problem are such that this manifestly cannot be the case. What operates is a confusion of diagnostic and causal reasoning that magically magnifies the consequences of an individual act of voting so as to make it worth while.[7] 'If I vote, others like me are likely to vote too, so let me vote in order to bring it about that they vote as well.' Needless to say, the belief could never be consciously articulated in this way; some sort of self-deception is needed.

Ainslie argues that the same reasoning can help an individual overcome the problem of impulsiveness, which may be seen as an intra-personal collective action problem. By making present decisions diagnostic of later ones, the individual can bunch his choices in a way that allows for more self-control. In this case, however, it is not clear that the thinking is magical or irrational; at least it can be articulated consciously without losing its force. It corresponds to the following, well-known chain of reasoning. '(1) If I take a drink just this one time, I can abstain on the next occasions and no harm will be done. (2) But do I really have any reason to think that I shall behave differently on future occasions, which will be essentially similar to the present one? (3) On reflection, therefore, I had better abstain now since otherwise I shall almost certainly yield to temptation the next time.'

Is this irrational? Observe first that we are not here talking about a genuine causal impact of the present choice on later choices, as in cases of habituation or addiction. In this respect it corresponds to the voting problem, in which it is similarly assumed that voters do not exert any causal influence upon one another by setting an example etc. Yet the first choice will typically be known to the person at the moment the later choices are made, unless he engages in a piece of genuine or self-deceptive forgetting. When he is about to make the later choice, the situation will differ from the earlier one in that he now has information about an earlier choice (or about more earlier choices). This information constrains his *self-image* in a way that may ease or obstruct the

[7] Note that this does not turn upon altruist motivation. Altruism can also act as a multiplier on the benefits from voting, but there is of course nothing irrational about this.

prudential decision.[8] If this is self-deception it is of a benign kind, since it turns out to be self-fulfilling.[9]

III. Faustian selves

'Two souls, alas, do dwell within his breast.' It is a common fact that people are often torn between different desires. They want to do several things that as a matter of fact or a matter of logic are mutually exclusive. It would be absurd to elevate all such cases to the status of 'split selves', but some of them may exemplify that notion as well as any. If the opposed desires are not sorted out by the person to yield a consistent series of choices, but lead to behavioural inconsistencies of some kind, there is a prima-facie reason to suspect a deep-seated split.

Intransitive choices – choosing A over B, B over C and C over A – is one form of inconsistent behaviour. Steedman and Krause argue that such choices could be due to an Arrow-like problem of aggregation, an analogue a social-choice problem within the individual. As they point out, the plausibility of this move turns crucially upon the ordinality of the underlying 'aspect preferences'. If there are at least three aspects that the individual finds himself unable to compare in cardinal terms, and thus has to treat 'democratically', he could find himself in a Condorcet paradox. A somewhat trivial example would be an internalized social-choice problem, for example, if the ordinal (transitive!) preference of each member of my family form one aspect of my preference structure. More generally, one would have to look for cases in which the 'aspect preference' is based on ordinal information only, even if there is an (unknown) cardinal structure.

Choice reversal – first choosing A over B and then B over A – is an even more dramatic form of inconsistency. As mentioned in Section I, some instances of this phenomenon can be explained without any reference to a divided self. This also holds in cases where two corner solutions are deemed equally and optimally good. And, if the choice reversal is itself irreversible, because the preferences of the person simply have changed, there is no paradox at all. Hence, as a necessary but not sufficient condition, we shall have to look for cases in which A and B are alternately chosen.

[8] Cp. Føllesdal (1981) for an interpretation of Sartre along these lines.

[9] For the relation between self-deceptive and self-fulfilling beliefs see Elster (1984, pp. 48, 177) and Pears (1984, pp. 33ff).

In several important articles Thomas Schelling has discussed the notion that different selves might alternately win out in 'the intimate contest for self-command'. Thus: 'People behave sometimes as if they had two selves, one who wants clean lungs and a long life and another who adores tobacco, or one who wants a lean body and another who wants dessert, or one who wants to improve himself by reading Adam Smith's theory of self-command (in *The Theory of Moral Sentiments*) and another who would rather watch an old movie on television' (Schelling 1980, p. 58). Or again: 'To plead in the night for the termination of an unbearable existence and to express relief at midday that one's gloomy night broodings were not taken seriously, to explain away the nighttime self in hopes of discrediting it, and then to plead again the next night for termination creates an awesome dilemma' (Schelling 1983, pp. 107–8).

In his earlier work Schelling (1963) had pioneered in exploring the idea that in strategic interactions one may improve one's prospects by eliminating certain options from the feasible set, as when one bargainer gets his way by making certain concessions physically impossible or extremely costly. In the intra-personal case this can be understood in two ways. First, most obviously, I may try to protect myself against weakness of will by removing the source of temptation; that is, one self may try to ensure that the other self will not be exposed to temptation. In this case, the language of several selves does not seem to have much purchase. Next, more interestingly, I may try to make myself invulnerable against the strategies that I might later use to get my way. Here the first person singular seems inadequate. If two or more parts of a person are really engaging in mutual strategic manipulation, there would seem to be good grounds for referring to several selves. Schelling argues that there are such cases.

Is he right? Do we observe that the self that wants to stay sober hides the bottle from the self who wants to drink, while the latter hides the Antabuse pills from the former? Or, to use the bargaining analogy, do we observe that the self who wants to drink makes sure that if he is deprived of alcohol he will die – something that the other self is not likely to want to happen? My contention is that we do not. There may be some examples of mutual manipulation;[10] I certainly do not think

[10] Aanund Hylland has provided me with an example. Going to a party, he overheard a conversation between a woman who was trying to quit smoking and her companion. It

there is any logical impossibility involved in this notion. But as far as I know such cases are few and far between, and not very important. Observe that this carries no implication about which self is the more 'authentic' or ethically valuable one. The asymmetry with respect to the capacity for strategic behaviour helps us to identify *the* person (Frankfurt 1971), but it does not tell us what goals are the most valuable or autonomous (Elster 1983, pp. 21–2). The observer could well be on the side of the easy-going spontaneous self that is constantly being repressed by the excessively far-sighted self that is in charge. The observer might plead unsuccessfully, 'Give yourself a break', meaning that the other self should get a break. Also, being in charge does not mean getting one's way – the horse may throw the rider even if the latter is in command. The capacity to form second-order intentions, the capacity to form autonomous intentions and the ability to get one's way – these may, but need not, go together.

IV. Hierarchical selves

The asymmetry just pointed out is a hierarchical one. The self that constitutes *the* person is not the stronger or more decisive self, the self that gets its way: it is the self that entertains higher-order intentions about other selves. Horizontal divisions would yield selves neither of which (as in Steedman and Krause) or both of which (as in Schelling) have higher-order intentions about the other. A remotely similar asymmetry obtains in self-deception, as discussed by Davidson and Pears. The part of the person that wants another part to have a certain belief, not justified by the evidence, must itself have some beliefs about the other part, but not vice versa. The deceived ignores the existence of the deceiver, not only of the deception. At least this is true in the absence of learning about one's tendency to deceive oneself. If such learning occurs, the deceived self might try to change the ways of the deceiver, or at least to minimize the harm it can do, by seeking advice etc. It would not, however, try to influence the deceiver by deceiving it in turn; at least I cannot attach any sense to this idea.

transpired that she had asked her companion not to bring any cigarettes to the party, to prevent her from backsliding. When she got there, she was unable to stick to her decision, and left to buy a pack of cigarettes. As she left, she said to her companion: 'And if I ask you not to bring cigarettes the next time we go to a party, don't listen to me.' For another, hypothetical, example see Elster (1984, p. 41).

Consider the analogy with nations. Two opposed countries might engage in mutual strategic interaction. They might also engage in mutual deception by planting false intelligence, including intelligence that if believed would undermine the intelligence operations of the other. The intra-personal analogy to mutual strategic interaction is at least conceivable, but the analogy to mutual deception seems more than far-fetched. Hence if the deceived self decides to counter-attack, it must do so by other means. Perhaps it could persuade the subversive self that its 'altruistic' efforts (Pears 1984, p. 91) are in fact misguided; or, as I said, it could try to contain the damage. This would reestablish the supremacy of the deceived system. Being weak and knowing it is better than being weak unknowingly, although the best is to be without weakness. Again, the place in the hierarchy does not depend on getting one's way. The effort to get rid of self-deception may meet with small success. Yet this would be due to the weakness of the counter-attack, not to any measures taken *in order to* neutralize it. The last claim, of course, would be falsified if the deceiving system's awareness of the deceived system included awareness of the counter-subversive measures and if it remained unaffected by the knowledge that the deceived system did not really want to be deceived. Both of these assumptions are, however, highly speculative, bordering on un-intelligibility.

The distinction between horizontal and vertical divisions of the self has been formulated in the language of meta-preferences. A person may possess several first-order preferences, each of which evaluates the options from a certain point of view (e.g. morality, sympathy, self-interest). This gives rise to a horizontal division of the self. Depending on how the various preferences interact – by aggregation, bargaining or manipulation – choices will be produced that may or may not fall into a consistent pattern. Amartya Sen has suggested that we should also envisage a vertical division, with a ranking (that may only be a partial ordering) of the preferences themselves (Sen 1977). An individual might ask himself: 'Would I rather have (and act upon) preferences R than preferences R'' The outcome of many such pairwise choices will be a ranking of the preferences.

In one sense it is misleading to talk about meta-preferences. The basis for ranking the first-order preferences must itself be a first-order preference about what the person thinks he should do, all things

considered. We care about preferences because they produce things we care about – actions and outcomes. I can make no sense of the notion of a meta-ranking not thus anchored in first-order evaluations. Yet in another sense the concept may be useful. If in a given situation I know the choices I ought to make, all things considered, yet find myself unable to stick to my resolutions, I may undertake a process of planned character change. I first decide that I want to become the kind of person who just does the right thing, but then I decide that this is setting my sights too high. My ideal self simply is not to be found in the set of outcomes of feasible character changes. Hence I will have to make a decision about which of the selves in that set I most would like to become. This would involve comparing each possible self with the ideal self, in order to rank the options and make a decision about what I would like to become. Now the foundation for this ranking must be my conception of the ideal self. To ask me to compare the options with one another, with no reference to this bench-mark, would not make sense. Hence one of the preference orderings must be both referee and contestant. It goes without saying that it will always come out on top in the unrestricted set of selves, but if it is not itself in the set of feasible set of selves it can serve to guide the choice between those that are.

Hence the notion of a hierarchy of preferences depends, if I am right, on the notion of an asymmetric distribution of the capacity to have second-order intentions. If my day-time self and my night-time self were both able to behave strategically toward the other, we could not decide which of them to use as a bench-mark. There would be two sets of meta-rankings. However, once we have firmly identified *the* person with one of the selves, on the basis of that self's unique capacity to form higher-order intentions about the other, we can use the preferences of that self to construct *the* meta-ranking of the person.

V. Successive selves

The coherence and identity of a person centrally involves *time*, in at least three ways. First, and most obviously, the individual may become 'a different person' because he changes in some profound way. Derek Parfit (1973) has cited the example of a Russian nobleman who, in his idealist youth, views with horror the prospect of changing into a more cynical and – to him – altogether different self. The example does not

work well, however, since in such cases the older, cynical self does not usually disavow the earlier one. Rather he might see the youthful idealism as part of what made him the person he turned into; and he might cite the phrase to the effect that the person who is not radical in his youth is as much a subject for pity as the person who remains radical in his old age. For successive selves in a sharper sense we must look to cases of religious or political conversion, in which there is a relation of mutual disavowal between the pre-conversion and the post-conversion selves. Even in such cases, however, we might decide that the continuity runs deeper than the difference. The communist who turns into a militant anti-communist does remain in touch with his earlier self, in spite of the break; moreover, the fervour of the earlier self may continue to animate the later. My hunch is that in close examination of any actual case we would come to the conclusion that talk of several selves creates more confusion than illumination.

When a person changes, he may regret some of the choices he made before the change. Also, he may find that he does not want to stick to his earlier decisions when they had a scope extending beyond the change. These phenomena – regret and incontinence – may also occur in the absence of any character change. They can stem from inconsistencies within the person's (unchanging) attitude towards time. Consider first weakness of will, in the form of a high discounting of the future. If because of this myopic attitude a person is led to prefer a smaller immediate pleasure over a greater, delayed one, he may well experience regret and desolidarize himself from the choice. Consider next a more complex phenomenon, the person who has to allocate some scarce resource over more than two periods in the future. He may give a disproportionately high weight to the first period, less to the second, even less to third and so on. If the discounting has a non-exponential character, as set out in Ainslie's paper, he will not be able to stick to his first decision when the second period arrives. He will reconsider his choice, so as to give more weight to the second period (relatively to the third) than he originally planned to do (Elster 1984, Ch. II.5). Persons subject to either of these liabilities suffer from lack of integration, but there is no need to talk about distinct selves, except in the Davidson–Pears sense. If the person believes that he ought to do what is best, all things considered, and nevertheless fails to do so on a particular occasion, there must be some split in the mind that prevents

the first belief from having the influence it ought to have. But, although temporal inconsistency may involve a reference to a divided self, there is no need to talk about successive selves.

Steedman and Krause draw our attention to a third way in which time matters for identity: by the nested system of memories and anticipation. Pleasures have a life after death as well as a pre-natal life; some linger on in memory and provide continued satisfaction, while others yet unborn offer the pleasures of anticipation. Steedman and Krause offer a novel and valuable distinction between two ways in which future benefits matter in the present. First, the anticipation of future (first-order)[11] benefits may actually add to my pleasure in the present on a par with the memory of earlier experiences. Secondly, the pleasure that in the future I shall derive from those benefits matters to me now, because I view myself as extending over time with more than just momentary interests. The distinction may seem tenuous, but on reflection it is quite robust. I may be the sort of person who 'lives in the present', in the sense of not thinking much about specific past and future experiences. I concentrate all my attention on the matters at hand, enjoying them to the hilt. Yet this is quite compatible with having a prudential concern for my own future ability to enjoy similarly present-centred pleasures. The converse case may seem less plausible. Can we imagine someone who 'lives in the present' in the sense of not saving anything for the future, yet derives much satisfaction from anticipating whatever pleasures he expects to get later on? I believe we can, if we stipulate that the future benefits are non-convertible into present ones. Hence they are not subject to his weakness of will, and he may contemplate their arrival with pleasure untainted by inner struggle. If it were possible to have spring come in December, I might choose to do so, but since it is not, I have the daily satisfaction of seeing the days grow longer and the time for that sudden acceleration of nature approach.

In the absence of the one or the other form for concern about the future, can we plausibly speak of a divided self? First, note that myopia need not be a case of weakness of the will, as pointed out in Section II.

[11] If my anticipation also covers the future higher-order pleasures (from memory of earlier first-order pleasures, memory of anticipation etc.) a more complex construction is needed, similar to the one used by Becker (1977, pp. 270–1) in his discussion of interpersonal externalities in the utility functions.

Some people with short time horizons do not wish they had longer ones. If they do, we might want to say that (at any given point in time) they have a divided self; if not, that they suffer from lack of temporal integration. Such people could be so myopic that they do not even perceive that they are involved in an intra-personal, inter-temporal problem of collective action; and a fortiori do not have the motivation to arrive at the cooperative solution. (This involves cognitive myopia, rather than motivational. It is not unreasonable to think that the two often go together, although either may exist without the other.) Secondly, the absence of externalities in the utility functions in the successive periods is also, I believe, better seen as a lack of integration than as a succession of selves. The person who lives in the present, in the first of the two senses distinguished above, would be poorly linked up with his own past actions. He might remember them, but the memory would not be invested with much emotional significance. Yet there is no split in the sense of relatively autonomous entities each promoting its own interest, if need be at the expense of that of others.

The strategic element could be important in many of these cases. Consider first character changes over the life-cycle. The earlier self has two kinds of interests in the future. He cares about the accomplishment of his current plans, and about the kind of person he will turn into later on. In particular, he might – like Parfit's nobleman – be afraid that the later self might frustrate the plans laid by the earlier one. If he has no influence on what he will become, he must then entrust the safeguarding of those earlier plans to another person (his wife in Parfit's story) or to some institutional device for precommitment (as discussed at length by Schelling, 1983). If he can shape his future character, there could be a conflict between the two concerns. He might decide that the future self that would best ensure the realization of his current plans is one that he would not care to become. My current plan may be for my children to lead a happy life. To ensure this I may have to turn myself into a solid wage-earner, with both the income and the sense of responsibility (induced by work) that would be required, yet that person may be very different from the Bohemian character I should really have wanted to become.

Consider next weakness of will in the simple or the complex form (i.e. with exponential or non-exponential time-discounting). This has been the paradigm for writers on strategic manipulation of the self

(Thaler and Shefrin 1981; Winston 1980; Elster 1984, Ch. II). The goal of the manipulation could be to overcome regret-inducing behaviour, arising from time preferences *tout court*, or to prevent incontinence, that is, the inability to stick to past plans that arises out of non-exponential time preferences. Such manipulation has a paradoxical character: it is motivated by the future inability to relate to (what will then be) the future. Since by assumption the person will remain the same, he should be just as unable now to relate to the future as he will be later on. The answer, of course, is that the two situations must be different. The sacrifices of present benefits must be smaller at the time the manipulative scheme is set up than it will be at the time when it comes into operation. Essentially, it costs nothing to tell all my friends that I shall quit smoking on 1 January next year, or to throw away all bottles of whisky save one. Yet these moves may enable me to carry out my decision to quit by ensuring that certain options become unavailable at the time when I might want to choose them.

VI. Parallel selves

In addition to our immediate personal experience we often enjoy the vicarious experience provided by daydreaming, reading novels or writing them. In fanciful exaggeration we may say that the vicarious experience belongs to a parallel self, one that runs its course alongside the main self. In non-fanciful language, of course, the fictional self is embedded in the main self. When I am daydreaming, *I* am daydreaming. Yet the fanciful language can serve the function of pointing to the importance that satisfaction by proxy can take on. Sometimes the consumption or creation of possible worlds comes to dominate the life of the mind at the expense of one's engagement in the actual world. Instead of speaking of parallel selves, we might think of the person as communicating between parallel lives.

Consider first daydreaming. Daydreams come in many varieties, depending on whether their starting-point is in the past or in the future and whether they are constrained to branch off from one's real life. I may wish that I had been a general in Napoleon's army, or were born in the twenty-fifth century, and construct elaborate daydreams to flesh out such wishes. Most daydreams, however, are hooked up with my own life in some way. There are the might-have-beens of my past: if

only I had had the wit to reply in kind; if only I had had the courage to ask her to marry me. Although not all the might-have-beens turn into daydreams, some of them do and are frequently updated to keep abreast wth current developments in the real world. Sometimes the central counterfactual element in the daydream is not something I might have done, but what someone else might have done or some event that might have occurred: if only she had said yes; if only I had won the big prize in the lottery. And then there are the daydreams that might still come true, those which branch off from my life at some point in the future. These are especially prominent in youth, when the future is open and the borderline between plans and daydreams is easily blurred.

The basic flaw of daydreams as a source of satisfaction is well characterized by Ainslie: they suffer from a *shortage of scarcity*. With a few exceptions satisfaction and pleasure derive from the relief of tension. (Cp. also the contribution of Schelling below, and the 'opponent-process theory of motivation' to which he refers.) Tension is created by scarcity: of talent, time, knowledge and money. It is because we are not omniscient that the search for truth offers an occasion for pleasure; it is because we are not omnipotent that we find our deepest satisfaction in stretching our limited abilities to the limit. In daydreams there is no scarcity; or if there is, there is nothing we can do about it. We can do anything we want; we do not have to build airplanes since we can just as easily imagine that we are endowed with wings. True, we cannot know everything, but nor can we do anything to increase our knowledge. (We can of course imagine that we know the truth about Fermat's last theorem, but the pleasure derived from this remains shallow as long as we not know what the truth about it is.)

By writing a novel we can overcome these defects. Novels are subject to *constraints* that provide the scarcity lacking in daydreams. First, and most obviously, a novel is essentially finite and complete. When the artist's vision has been externalized and given to the public, he cannot go on adding details; nor can he answer critics by saying that they simply don't know the characters well enough. It is his task – and his constraint – to ensure that the readers know exactly what is needed to understand what is going on. Indeed, there is nothing more to be known about the characters than what appears in the novel itself. If the author thinks he has a private peephole into what his characters do off-

stage, he confuses the novel with an unconstrained daydream. Secondly, unlike daydreams, a novel must respect the laws of probability. In daydreams a mere possibility suffices to launch a train of events. One can call upon any coincidence to make things turn out the way one wants. A coincidence in a novel – like the one at the centre of *Middlemarch* – is rightly seen as a sign of authorial self-indulgence.

I argue in my contribution below that in Stendhal's case, the novel served largely as a means of vicarious satisfaction. In *Lucien Leuwen* he even experimented with multiple fictional selves – turning Lucien into the person he had wanted to be in his youth and Leuwen *père* into the character he wanted to become in his mature age. Observe, however, that this is not the same as having multiple plots in a novel. I tend to agree with Schelling when he writes that 'You cannot show two episodes and let each viewer choose.' Yet in *The French Lieutenant's Woman* John Fowles did exactly that, when he left the reader to choose between one ending in which the two main characters are united with one another and one in which they are not. The reason why I find this unsatisfactory is related to the role of constraints in fictional writing. In the beginning of a novel each action, choice or remark is heavily underdetermined by what we already know about the person. Their main purpose is to contribute to our knowledge of his character, that is, to narrow down the set of things that he can say or do at that point. A properly constructed novel ends at the point when the options open to the person are limited to the point of inevitability. Or more precisely: if in the penultimate period of the novel several options are open to a person, given what at that point the reader knows about his character, his final choice should improve our understanding of his character so as to make that choice the only possible one. *The French Lieutenant's Woman* ends before any such point is reached. Other novels sin in the opposite direction, by going on after that point has been reached. This is the case of all the novels whose authors could not resist the temptation to dwell on the bliss of the deserving and the misery of the undeserving. Our uneasy pleasure in reading about these inconsequential details is related to the pleasures of daydreaming; it is gorging oneself in a way that stills no hunger.

What, finally, about the vicarious experience provided by *reading* a novel? The operative constraint here is that of lack of knowledge: we want to see how it all turns out. Hence the twin dangers referred to in

the previous paragraph: that of frustration when our tension is not relieved and that of boredom when the narrative goes on after the relief with no new tensions being provided. The author writes under the constraint that he must provide the reader first with the knowledge constraint that sets up the tension and then with the knowledge that is necessary and sufficient to relieve it. If he succeeds, he has created a source of vicarious satisfaction both for himself and for the reader. The author and the reader – 'mon semblable, mon frère' – become truly parallel selves, since they live off the same daydream, the presence of each being a condition for the satisfaction of the other.[12]

VII. The Freudian legacy

Freud left us with a new language for talking about the divided self. On the one hand he introduced the division into conscious, preconscious and unconscious; in addition he proposed a distinction between id, ego and superego. The former is more like a distinction between territories, the latter approaches a distinction between agents. The exegetical and conceptual difficulties of understanding exactly what Freud intended by these notions are enormous. Fortunately, I can restrict myself to problems that impinge on the issues raised in the contributions to this volume. I shall consider two such issues: the relation between the conscious and the preconscious, and the idea that self-control can be a problem as well as a solution.

In our dictionary of mental categories there is room for something that is more than awareness and less than self-consciousness. Awareness is what animals and men enjoy in perception of external objects. When my dog watches me to see if I am going to slip a morsel of food to her, she is certainly aware of me. Self-consciousness is what men enjoy when they turn the mind inwards to watch its own operation. An instance is the effort to remember someone's name by bringing to mind all the circumstances in which one has met him. The intermediate category, which we may call consciousness, is what we possess when relating to external objects not immediately perceptible by the senses. Going back to Alaska from Florida we take a warm overcoat – an

[12] For other ways in which constraints are important for artistic creation, see Elster (1983, Ch. II. 7).

action not triggered by anything in the environment of choice. This ability to re-present what is physically absent is probably not an exclusive feature of men, but shared with some animals (Griffin 1984). It is, broadly speaking, what enables men to relate to the future and, if need be, to their future inability to relate to the future (Elster 1984, Ch. I). I am not saying that every form of consciousness is also awareness, and all self-consciousness also consciousness; nor am I denying these propositions. I am only arguing that there is a peculiar state of mind that makes mental representing possible, and that is closely related to what we usually refer to as consciousness.

There is another way of understanding the notion of consciousness. It may be taken to be a state peculiarly transparent to itself, so that a person, when asked if he remembers something of which he was conscious a few moments ago, could never truthfully say 'No'. This is Sartre's 'conscience (de) soi' – a knowledge of what one is doing that does not have the explicitly intentional structure of self-consciousness. This may appear mystical. Consider, however, an example from game theory. The 'common knowledge' condition among the players in a game can be set up by an umpire or experimenter telling all the players, in each other's presence, about the rules of the game. This does not mean that each player must have, with respect to each of the others, an explicit knowledge that 'I know that he knows that . . . I know the rules of the game', since this would involve a completed infinite regress. Yet, for any n, if one asked the player whether he had knowledge of degree n about the other, he would say that he had such knowledge.

Frequently, consciousness in the sense of re-presenting what is absent goes together with consciousness in the sense of pre-intentional transparency. Freud, or some Freudians, may be understood as saying that there can be consciousness in the first sense unaccompanied by consciousness in the second sense. This, if I understand him rightly, is how David Pears (1984, p. 79) interprets and defends Freud's doctrine of the preconscious. On this view, mental representations may exist and do their work, whatever that is, even when the person cannot tell whether he has them. (To assert that they exist but do not do any work when the person does not know that he has them, is not, I believe, to say anything.) The preconscious on this conception is not just a store-house from which mental entities or their precursors can be retrieved

when needed. It is itself the centre of a great many mental activities, such as representing, imagining, even choosing.

I feel deeply uncomfortable about this suggestion, but I do not feel my understanding of the problem goes far enough to buttress my uneasiness with arguments. Suffice it to say that the proposal does not appear to involve any logical contradiction (although I am not sure); it may be indispensable to account for certain mental phenomena that would otherwise be inexplicable (although in this domain such backward reasoning will never be conclusive); there could hardly be any sort of direct evidence for it (not even an analogy to the traces left by an electron in a cloud chamber); and it suffers from an almost total lack of structure (neither the *modus operandi* nor the scope of its operation is specified). Perhaps the best way of summarizing my uneasiness is that the preconscious, on this view, becomes detached both from the objective world (since the items in it are only *representations* of that world) and from the subject (since the person does not know whether he has them). Yet they are representations *of* the external world and belong in some sense *to* the subject. I seem to be able to handle either of these paradoxes by itself, but not both of them simultaneously.[13]

The achievement of George Ainslie's work, in this volume and elsewhere (notably Ainslie, 1982), is to make good analytical sense of what appeared to be the irreducibly metaphorical notions of id, ego and superego. The superego is a way of referring to an overly successful solution to the problem of self-control, that is, the problem of curbing the impulsiveness often referred to as the id. As briefly explained above, the central notion in his account of impulse control is that of *bunching of choices*. The other side of this coin, however, is that the self-control may turn into compulsive behaviour and rigid adherence to rules. The guiding principle 'Never suffer a single exception' is first applied to the specific kind of behaviour that one seeks to control, and then generalized and applied across the board. The superego on this view is not internalized parental authority, but an

[13] If animals have consciousness in the sense of having representations, they would appear to have it without possessing consciousness in the second sense. As I have defined the latter, it is necessarily accompanied by the potential for self-consciousness, in the sense of being able to relate to one's own (past) consciousness. If we deny self-consciousness to animals, how can we also ascribe to them consciousness in the first sense without getting into similar conceptual uneasiness? The short answer is that I do not know. More elaborate answers, in more speculative directions, might be suggested, but the ground is really too loose to give more than a momentary foothold.

endogenous by-product of strategies for self-control. The anxiety induced by the prospect of breaking a rule does not stem from infantile attitudes towards one's father, but from fear that the whole structure might unravel by a single violation of the rule, indeed of *any* rule. The person can, however, use a stratagem to give himself a break, without breaking down. He can orient himself by a system of *bright lines* or non-manipulable cues, which tell him when an exception is justified, given the circumstances, and when it is precisely the kind of temptation that motivated the rule in the first place. The autonomy of the person (or the strength of the ego) requires *loose bunching*. Bright lines are a form of mental book-keeping; hence the presently discussed loose bunching may not be unrelated to the reasons why the self tends to be only loosely integrated (cp. Section I above). An interaction between cognitive and motivational elements in creating behavioural slack seems plausible, even if the details elude us.

The two sets of issues I have been discussing have something in common. The ego (i.e. the person) is concerned with the future, while the id (the impulses) is guided by short-term pleasure. The id is climbing along a pleasure-gradient, which makes it as liable as other gradient-climbers to fall into the 'local-maximum trap' (Elster 1984, Ch. I; also Staddon 1983). Hence the id does not need any representation of the future. It scans the actual (as distinct from the potential, future, hypothetical, imagined) alternatives, and chooses the first one that will bring an increment of pleasure. To be sure, this is metaphorical language, treating the 'id' as a separate short-term maximizer competing with a long-term maximizer. Non-metaphorically, it amounts to saying that what is present and what is merely re-presented compete for our attention. I do not know whether those who advocate the possibility of preconscious representations believe that these can similarly compete with the actual, and not always lose out.

VIII. Split brain – split mind?

As a result of work on epileptic patients, it has been found that when the connections between the right and left brain hemispheres are severed, two semi-autonomous functional systems emerge.[14] The left

[14] The following is largely based on Springer and Deutsch (1981).

hemisphere controls speech, whereas the right is in charge of the visual and spatial processes. At least, this was the formulation that served as a vehicle for most of the earlier research. Later analysis suggests that other distinctions may be more appropriate. The left hemisphere is analytical, the right holistic; the left is based on sequential processing of information, the right on simultaneous processing.

Information comes to the right hemisphere from the left visual field, the left ear and the left hand; conversely information to the left hemisphere comes from the right side. In the normal brain the information is then pooled between the two hemispheres, but in split-brain patients this does not take place (or not in the same way). An answer to a question about a perceptual event will reflect only the information available to the left, speech-producing hemisphere. The patient will not be able to tell the nature of an object only presented to the left visual field. Yet the left hemisphere may express some awareness of that object, since it is aware of bodily reactions produced by the reception of information about the object in the right hemisphere. Thus a patient who was presented with the picture of a nude woman in the left field reacted emotionally to it, and then expressed a verbal explanation of what was going on. 'It is very common for the verbal left hemisphere to try to make sense of what has occurred in testing situations where information is presented to the right hemisphere. As a result, the left brain sometimes comes out with erroneous and often elaborate rationalizations based on partial cues' (Springer and Deutsch 1981, p. 33).

Split-brain is an extreme case of cognitive compartmentalization (cp. Section I), but I hesitate to talk about a divided self. The left and the right hemispheres do not seem to differ in goal or motivation. True, there have been conjectures about the 'egocentricity' of the left hemisphere (Springer and Deutsch, pp. 176–7), but not in a motivational sense. Rather, the idea seems to be that the left hemisphere is unwilling to admit that it may need cognitive assistance from the right. An experiment set up to find evidence for motivational conflict failed to do so (MacKay and MacKay 1982; cp. also Sergent 1983). Here one hemisphere, which could receive verbal stimuli and refer to them by non-verbal means, was exposed to a number. The other, verbal hemisphere would try to guess what the number was. The first hemisphere would then correct the guess by the non-verbal means at

its disposal. The experimenters failed, however, when they tried to turn the game into one of conflict rather than cooperation. The two hemispheres refuse to be drawn into conflict. The authors concluded that although the left and the right hand 'were substantially separate at the cognitive level, their priorities gave no evidence of being under the supervision of two independent normative systems.'

Split-brain experiments also may tell us something about the functioning of the normal brain. In particular, it has been suggested that 'Certain aspects of right hemisphere functioning are congruent with the mode of cognition psychoanalysts have termed primary process, the form of thought that Freud originally assigned to the system Ucs (unconscious)'.[15] It has also been suggested that even in the normal brain the verbal hemisphere engages in a good deal of constructive interpretation of behaviour initiated by the non-verbal one.[16] On this view, each person contains several mental systems – emotional, motivational and perceptual.

> Then, as maturation continues, the behaviours that these separate systems emit are monitored by the one system we come to use more and more, namely, the verbal, natural language system. Gradually, a concept of self-control develops so that the verbal self comes to know the impulses for action that arise from the other selves, and it either tries to inhibit these impulses or free them, as the case may be.[17]

Clearly, these conjectures are related to the discussion in Section VII. Conceivably, they might one day create physiological underpinnings for a suitably purified and de-mythologized Freudian theory. For the time being, however, they offer only a loose parallel between two highly speculative hypotheses.

IX. Homo economicus and homo sociologicus

Each of us seems to be split between a private and a public self. The 'economic man' within us strives for personal hedonic satisfaction. He

[15] Galin (1974), cited after Springer and Deutsch (1981).
[16] Cp. the partially related problem discussed in Pears (1984), Ch. IX.
[17] Gazzaniga and LeDoux (1978), cited after Springer and Deutsch (1981, p. 199).

regards other people as so many means to his own selfish ends – or as constraints and obstacles to his pursuit of happiness. The 'social man', by contrast, is governed by moral and social norms. He is kept on course by his concern for other people, and by their approval or disapproval of his behaviour. The problem is to understand the relation between these two homunculi that – like the short-term and the long-term interest – constantly vie for our attention.[18] A paradigm case is that of explaining voting behaviour, with regard to which social scientists have manoeuvred themselves into a situation of theoretical schizophrenia. To explain *that* people vote, an appeal to civic duty or similar normative concepts seems inevitable; to explain *how* they vote, the appeal to self-interest is usually deemed sufficient (Barry 1979). It is as if the voter, upon entering the voting booth, shed the social motivations that had carried him there. Surely, this cannot be the right conceptualization – but what is?

Howard Margolis (1982, esp. Ch. 4) has proposed a general theory of altruist behaviour to explain such phenomena. On his theory, an individual such as Smith behaves as if he were made up of two persons. S-Smith and G-Smith, who are concerned with selfish benefits and group benefits respectively. The rule that explains how Smith's income is allocated between selfish and public purposes has two parts. First, the larger the ratio of the marginal utility of public spending to that of private spending, the greater the tendency for Smith to allocate a marginal dollar to G-Smith. Next, the higher the proportion of Smith's income already spent on public purposes, the larger the tendency to allocate the next dollar to S-Smith. Margolis suggests (among other comparisons)[19] that this is related to the way in which the ego mediates between the superego (G-Smith) and the id (S-Smith). The second part of the rule, in particular, corresponds to the idea that a person must know when to give himself a break – when to temper the claims of duty. It also captures our intuitive notion of doing one's *fair share* of contributing to the common good. Suggestive as this proposal is, it does not turn upon any substantive notion of a divided self. Contrary

[18] For extensive discussions of the analogies between prudence (i.e. long-term selfishness) and altruism, see Nagel (1970) and Parfit (1984). Norm-guided behaviour is not, however, the same as altruistic behaviour. For a discussion, see Elster (1985).

[19] In a passage cited in note 1, he also compares the allocation of personal income between selfish and public spending to the allocation of profit between investment at home and overseas investment.

to what Margolis says, I think the most natural way of understanding the allocation rules is in terms of a unitary preference structure. The *person* seems to be firmly in charge.

Some cases, however, suggest a real conflict between the economic and the social self. The former seems to be able to exploit ambiguities in the norms to get its way. Consider again the lawn-mowing example from Section I. An alternative explanation (suggested to me by Amos Tversky) of that behaviour could be that mowing the neighbour's lawn would be incompatible with the man's self-image. He simply does not think of himself as the kind of person who mows other people's lawns for money. Yet one might easily imagine that he would mow the lawn in return for the neighbour's donating $20 to charity, and that this might make him feel justified in withholding a contribution of $20 that *he* would otherwise have made to charity. The situation is materially equivalent to the first, yet under the new description ('mowing for charity') the behaviour is more acceptable than under the first ('mowing for money'). I believe that this kind of normative reframing (Tversky and Kahneman 1981) is very frequent. It is 'as if' the economic man within us tried to manipulate the way in which choice situations are presented and described, so as to induce the social man to take the course of action that the economic man prefers. Analogously, in the conflict between the short-term and the long-term interest, it often looks 'as if' the former tries to persuade the latter that *this* occasion constitutes a genuine exception to the rule. After all, it would be ridiculous not to take a drink when a friend drops by unexpectedly, or when my candidate wins the presidential election, or when . . . (See also Elster 1985).

This is not a question of trade-offs. True, the question of how much it will take to bribe me into violating a norm is a meaningful one (North 1981, Ch. 5). Even if I wouldn't mow my neighbour's lawn for $20, I might do it for $100. If I do it, however, it may be at some cost to my self-respect. The impact of reframing is to make it possible to violate a norm without any cost to myself; in fact the behaviour that formerly appeared norm-violating may now become positively prescribed. Norms of cooperation, in particular, depend for their application on the specification of the reference group, and may yield different prescriptions if that group is redefined. What appears as cooperative behaviour with respect to my fellow union members may no longer do

so in the broader perspective of all workers (Olson 1982). The free rider may justify his behaviour to himself by placing it in a perspective that makes it appear as a form of society-wide or long-term cooperation. The pacifist who is asked, 'But who would fight the enemy if everyone acted like you?' may reply, 'If everyone acted like me, there would be no enemy to fight.'

I am not making the cynical point that a person may often be able to justify his behaviour to *others* by invoking norms on an *ad hoc* basis, exploiting the almost endless repertoire of norms to disguise the fact that he is moved by self-interest. My point is that a person must be able to live with his decisions – so he has to justify them to himself. There are constraints on the acceptable justifications. In particular, the need for consistency between the norms that are invoked in different situations may be as important as the consistency between the norm and the self-interest. Yet within these constraints a good deal of redefinition of norms is possible. My suggestion is that in addition to the head-on conflict between self-interest and social norms there is an insidious struggle that is more similar to self-deception and thus more closely related to the multiple self.

X. The 'no-self' theory

If the view that there can be a multiple self is carried to its extreme conclusion, it is more naturally labelled a 'no-self' theory. In the history of thought this view is associated especially with Hume; a more sophisticated and elaborate version is found in Buddhism.[20] Two Neo-Buddhist theories have recently been proposed by Serge-Christophe Kolm (1982) and by Derek Parfit (1984). I shall present a brief summary of Kolm's view, to provide a context for the chapter from his book, excerpted below.[21]

According to Buddhism, the human being at any given moment is made up of various elements (*dharma*), some of which constitute his body and others various mental states. Among the latter we find the belief in an enduring self, which is thought of both as the unchanging substance underlying the changing mental states and as the active centre of decision-making. Although the belief is an illusionary one, it

[20] For a good exposition, see Collins (1982).
[21] The following draws on Collins (1982) and on Kolm (1982).

is very difficult to shake it off, since it arises in a very natural, indeed compelling way. Hence Buddhism offers three doctrines with respect to the self. First, it contains a theoretical critique of the notion of an enduring self, together with a constructive analysis of the actual unity and continuity of the person. On Kolm's view, the unity of the person is merely a property of the causal chains that link together the successive mental states, so that an element can be ascribed to a person if it is sufficiently closely related to other elements that have already been imputed to him. Next, Buddhism offers an account of the emergence of the illusionary belief in the self. Among other things, the inherent logical difficulty of treating oneself as fully causally determined leads almost irresistibly to the invention of the notion of a free agent which is an active maker of decisions, not simply the aggregate of causally related mental states. Finally, Buddhism proposes a way of overcoming this spontaneously arising illusion, through study and meditation. The illusion is to be overcome not because it is a bad thing in itself to live under the sway of an illusion, but because this particular illusion generates so much unhappiness.

We should note the difference between a purely intellectual understanding of the non-self doctrine and the psychological or affective acceptance of the theory. The overt fetter of a theoretical belief in the self is of secondary importance compared to the 'selfishness inherent in the affective structure of experience' (Collins 1982, p. 101), indeed, excessive attachment to a theory of non-self is a sign that one has not liberated oneself from it affectively. 'Right view' in itself is simply one 'karmic agent' among others, a mental state causally producing other states of mind and not necessarily increasing peace of mind. For the latter, training and practice are required. There is a parallel here with psychoanalysis and its emphasis on the insufficiency (and sometimes the non-necessity) of *Bewusstwerden* as a condition of *Ichwerden*. There is also a vast difference, stated by Kolm in the final sentence of his book. Freud argued that 'where id was, Ego shall be'. Buddhism takes the further step of saying, 'where Ego was, consciousness shall be'.

Why should the belief in the existence of the self–ego lead to unhappiness? In the Zen version of Buddhism the answer is found in the corrosive effects of the habit of relating everything to self (cp. Elster 1983, Ch. II.2 and Smullyan 1980). If one is constantly thinking

about the impression one is making on other people, instead of just getting ahead with the task at hand, one will not make much of an impression on them. Similarly, in order to overcome such problems as stuttering, insomnia or impotence, one must above all avoid an inward-looking or self-conscious attitude. In such cases there is an interference between the goal one has set for oneself and the way in which one is trying to achieve it. The goal is within reach, if only one can forget about it. These are states that are essentially by-products (Elster 1983, Ch. II): they can come about as a result of action, but cannot be brought about deliberately by action. Much of the attractiveness of Buddhism derives from this simple moral point, that happiness tends to elude those who search for it and to fall into the lap of those who concentrate on achieving substantial goals. This view can be stated without reference to any epistemological or ontological doctrine concerning the self.

Kolm, in his contribution below, argues for a different view. The frustration of desires can be eliminated by exploiting the total plasticity of character. Since there is no permanent self opposing or constraining character changes, one may act strategically on the desire and preferences so as to achieve the optimum of happiness, or the minimum of suffering. It may be worth while pointing out that his argument presupposes that frustration is inherently bad, contrary to the view (set out in Section VI above) that some frustration is indispensable for happiness. Kolm quotes Benjamin Franklin to the effect that pleasure is the liberation from suffering, but neglects to draw the conclusion that sustained pleasure therefore requires a steadily renewed suffering or frustration. No doubt the 'egonomical' framework[22] could be extended to incorporate the need for an optimal amount of frustration.

XI. Summary

Barring pathological cases (which I have not discussed here) we ought not to take the notion of 'several selves' very literally. In general, we are dealing with exactly *one* person – neither more nor less. That person may have some cognitive coordination problems, and some motivational conflicts, but it is *his* job to sort them out. They do not

[22] The phrase 'egonomics' was coined in Schelling (1978).

sort themselves out in an inner arena where several homunculi struggle to get the upper hand.

Yet some of the motivational conflicts are so deep-seated and permanent that the language of a divided self almost irresistibly forces itself on us. Although only one *person* is in charge, he is challenged by semi-autonomous strivings that confront him as 'alien powers'.[23] To get his way, he may have to resort to ruse and manipulation. The relationship is essentially asymmetrical: it is that of an intentional person confronting causal forces within himself. There is one possible exception. One of the alien powers may in an almost literal sense try to deceive him, by preventing him from acquiring a belief that would interfere with the desire for short-term gratification. For this to be possible, the subsystem must itself have some minimal degree of rationality and intentionality. Yet the person may take cognizance of his tendency to deceive himself, and counteract it with measures to which the deceiving subsystem has no further reply.

All the cases I have come across are dichotomous or trichotomous (disregarding the general *n*-person case discussed by Steedman and Krause). Dual selves underlie the distinction between myopia and prudence; between economic man and social man; between the left and the right hemispheres; between the day-time person who wants to stay alive and the night-time person who wants to be relieved of suffering and anxiety. The tripartite self is at the core of Freud's anatomy of the mind. Actually, this conception amounts to saying that *the* person tries to mediate between the long-term and the short-term interests, or between the private and the public man. The autonomous individual does not want to identify fully with any of these extreme strivings. He wants to do what *he* thinks is best, all things considered, not to be the slave of his impulses or of the rules and norms he has set for himself.

[23] Marx (1845–6, p. 262). Marx saw a close relation – causally as well as conceptually – between such intra-personal 'reification' of mental powers and the inter-personal 'alienation' of the individual from society.

REFERENCES

Ainslie, G. (1982) 'A behavioral economic approach to the defence mechanisms: Freud's energy theory revisited', *Social Science Information* 21, 735–70.

Barry, B. (1979) *Economists, Sociologists and Democracy*, rev. edn, University of Chicago Press.

Becker, G. (1977) *The Economic Approach to Human Behavior*, Chicago: University of Chicago Press.

Collins, S. (1982) *Selfless Persons*, Cambridge: Cambridge University Press.

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Oxford University Press.

Elster, J. (1983) *Sour Grapes*, Cambridge: Cambridge University Press.

Elster, J. (1984) *Ulysses and the Sirens*, rev. edn, Cambridge: Cambridge University Press.

Elster, J. (1985) 'Weakness of will and the free-rider problem', *Economics and Philosophy* 1, 231–66.

Frankfurt, H. (1971) 'Freedom of will and the concept of a person', *Journal of Philosophy* 68, 5–20.

Føllesdal, D. (1981) 'Sartre on freedom', in P. A. Schilpp (ed.), *Sartre* (The Library of Living Philosophers), pp. 392–407. La Salle, Ill.: Open Court.

Galin, D. (1974) 'Implications for psychiatry of left and right cerebral specialization', *Archives of General Psychiatry* 1974, 572–83.

Gazzaniga, M. S. and LeDoux, J. E. (1978) *The Integrated Mind*, New York: Plenum Press.

Goldman, A. (1978) 'Epistemics', *Journal of Philosophy* 75, 509–24.

Griffin, D. (1984) *Animal Thinking*, Cambridge, Mass.: Harvard University Press.

Kolm, S.-C. (1982) *Le Bonheur-liberté*, Paris: Presses Universitaires de France.

MacKay, D. M. and MacKay, V. (1982) 'Explicit dialogue between left and right half-systems of split brains', *Nature* 295, 690–1.

Margolis, H. (1982) *Selfishness, Altruism and Rationality*, Cambridge: Cambridge University Press.

Marx, K. (1845–6) *The German Ideology*, in Marx and Engels, *Collected Works*, vol. 5, London: Lawrence and Wishart.

Nagel, T. (1970) *The Possibility of Altruism*, Oxford: Oxford University Press.

Nisbett, R. and Ross, L. (1980) *Human Inference*, Englewood Cliffs, N.J.: Prentice Hall.

North, D. (1981) *Structure and Change in Economic History*, New York: Norton.

Olsen, M. (1982) *The Rise and Decline of Nations*, New Haven: Yale University Press.

Parfit, D. (1973) 'Later selves and moral principles', in A. Montefiore (ed.), *Philosophy and Personal Relations*, pp. 137–69. London: Routledge and Kegan Paul.

Parfit, D. (1984) *Reasons and Persons*, Oxford: Oxford University Press.

Pears, D. (1984) *Motivated Irrationality*, Oxford: Oxford University Press.

Schelling, T. C. (1963) *The Strategy of Conflict*, Cambridge, Mass.: Harvard University Press.

Schelling, T. C. (1978) 'Egonomics, or the art of self-management', *American Economic Review: Papers and Proceedings* 68, 290–4.

Schelling, T. C. (1980) 'The intimate contest for self-command', *The Public Interest* 60, 94–118. Cited after the reprint in Schelling (1984).

Schelling, T. C. (1983) 'Ethics, law, and the exercise of self-command', in S. McMurrin (ed.), *The Tanner Lectures on Human Values IV*, Salt Lake City: University of Utah Press, 43–79. Cited after the reprint in Schelling (1984).

Schelling, T. C. (1984) *Choice and Consequence*, Cambridge, Mass.: Harvard University Press.

Schumpeter, J. (1961) *Capitalism, Socialism and Democracy*, London: Allen and Unwin.

Sen, A. (1977) 'Rational fools', *Philosophy and Public Affairs* 6, 317–44.

Sergent, J. (1983) 'Unified response to bilateral hemispheric stimulation by a split-brain patient', *Nature* 305, 800–2.

Smullyan, R. (1980) *This Book Needs No Title*, Englewood Cliffs, N.J.: Prentice Hall.

Springer, S. and Deutsch, G. (1981) *Left Brain, Right Brain*, San Francisco: Freeman.

Staddon, J. E. R. (1983) *Adaptive Behavior and Learning*, Cambridge: Cambridge University Press.

Thaler, R. (1980) 'Towards a positive theory of consumer behavior', *Journal of Economic Behavior and Organization* 1, 39–60.

Thaler, R. and Shefrin, M. (1981) 'An economic theory of self-control', *Journal of Political Economy* 89, 392–406.
Tversky, A. and Kahneman, D. (1981) 'The framing of decisions and the rationality of choice', *Science* 211, 543–58.
Veyne, P. (1978) *Le Pain et le Cirque,* Paris: Seuil.
Veyne, P. (1983) *Les Grecs ont-ils cru a leurs mythes?* Paris: Seuil.
Winston, G. (1980) 'Addiction and backsliding', *Journal of Economic Behavior and Organization* 1, 295–324.