Chapter I

Ulysses Revisited: How and Why People Bind Themselves

I.1. INTRODUCTION: CONSTRAINT THEORY

In this chapter I discuss why individuals may want to restrict their freedom of choice and how they achieve this end. Broadly speaking, they may want to protect themselves against passion, preference change, and (two varieties of) time-inconsistency. They do so by removing certain options from the feasible set, by making them more costly or available only with a delay, and by insulating themselves from knowledge about their existence.

In this section, I want to locate constraints that individuals impose on themselves within the broader field of what one might call "constraint theory." At a very general level, the present book illustrates the proposition that sometimes *less is more* or, more specifically, that sometimes there are benefits from having fewer opportunities rather than more. This idea must be seen on the background of the standard case, in which the exact opposite is true. Prima facie it would seem that nobody could have a motivation for discarding options, delaying rewards, or imposing cost on themselves. And in most of everyday life this intuition is obviously correct. Most people would rather have more money than less, more occupational options rather than fewer, rewards sooner rather than later, a larger range of potential marriage partners rather than a smaller one, and so on. Much progress in human history has in fact taken the form of removing material or legal restrictions on choice. Moreover, even when we don't benefit from having more opportunities, they usually do not harm us either, since we can always choose not to take them up (the "free disposal" axiom of general equilibrium theory). If I find some of the free meals offered by airlines unappetizing, I don't have to eat them.

In this book I discuss nonstandard cases in which the "more is better" assumption is invalid. It can be so for one of two reasons. On the one hand, the individual might benefit from having *specific* options unavailable, or available only with a delay, or at greater cost, and so on. (Although I always eat the meals the airlines offer me, I would pay them a bit extra for not serving me.) This is the topic of Chapters I and II. On the other hand, the individual might benefit just from having *fewer* options available, without the desire to exclude any specific choices. This is the main topic of Chapter III, where I argue both that artists need constraints and that the choice of constraints is largely arbitrary. True, the first reason for wanting to be constrained can apply here too, as when a film director decides to shoot in black and white so as not to be tempted by the facile charms of color photography. Yet the second reason for artistic precommitment is usually more important. The decision by a writer to use the format of the short story rather than the novel is not dictated by the desire to exclude any specific words or sentences, only by the desire to use fewer of them.

This second reason might also apply to social life more broadly. Erich Fromm argued that with the rise of the modern world and the progressive removal of restrictions on action, there has also emerged a "fear of freedom" – a fear of having too much choice, too many options, being subject to too little authority.[1] Along similar lines, Tocqueville said, "For my part, I doubt whether man can support complete religious independence and entire political liberty at the same time. I am led to think that if he has no faith he must obey, and if he is free he must believe."[2] The implication is not that people would *choose* to limit their options, but that they would benefit from having fewer rather than more. Many who like me grew up in the relatively austere 1940s and 1950s believe that children and teenagers in later decades would have benefited from having fewer opportunities and less money to spend. And there is a great deal of folklore to the effect that rich kids suffer irreversibly from having too many options, and that individuals who are very richly endowed in talents end up being jacks of all trades and masters of none. Although these beliefs may partly be sour grapes, casual observation suggests that they are not always only that.

At the same general level, the idea that less is more is susceptible of another interpretation, namely, that *ignorance is bliss*. Again, this

1. Fromm (1960).
2. Tocqueville (1969), p. 444. See also Elster (1999a), Ch. I.5.

idea must be considered on the background of a standard assumption to the contrary, namely, that *knowledge is power*.[3] In this case, too, historical progress has often taken the form of gaining new knowledge that enhances our mastery over nature, including, sometimes, human nature. Because this knowledge can also have destructive consequences, one might ask whether it might not sometimes be better to abstain from acquiring it. In *De Finibus* (V.xviii) Cicero has Antiochus interpret the Sirens episode of the *Odyssey* in this perspective:

> So great is our innate love of learning and of knowledge that no one can doubt that man's nature is strongly attracted to these things even without the lure of any profit. . . . For my part, I believe Homer had something of this sort in view in his imaginary account of the songs of the Sirens. Apparently it was not the sweetness of their voices or the novelty and diversity of their songs, but their professions of knowledge that used to attract the passing voyagers; it was the passion for learning that kept men rooted to the Sirens' rocky shores.

Cicero does not suggest, however, that Ulysses bound himself to the mast in order to remain ignorant, nor that the knowledge the Sirens offered would have been dangerous to him. Hence the analogy that is sometimes drawn between the Sirens episode in the *Odyssey* and the Fall in Genesis is somewhat halting.[4] The Serpent seduced Eve by offering her intrinsically corrupting knowledge, whereas the Sirens (in this reading) used the prospect of knowledge merely as a means of enticing their victims to the rocky shores.

In *Forbidden Knowledge*, Roger Shattuck pursues the theme of dangerous knowledge and blissful ignorance through a number of historical and fictional examples. From the history of science, he cites notably the moratorium on DNA recombinant research in the 1970s and objections to the Human Genome Project, arguing that there may be non-obscurantist reasons for blocking or halting the progress of knowledge.[5] Or consider another example: some years ago voices in the Norwegian government opposed exploratory oil drilling north of 62 degrees latitude. To those who argued that it could do no harm and might be useful to know whether there was oil in that region, these critics replied that if one found oil there would be an irresistible pressure on politicians to begin exploitation immediately. The critics

3. For a discussion of opposite proverbial sayings of this kind, see Elster (1999a), Ch. I.3.
4. For this analogy see Montaigne (1991), p. 543, and Shattuck (1996), p. 28.
5. Shattuck (1996), pp. 186–95, 210–17.

lost, and were proven right. I discuss half a dozen cases of this general kind at various places in this chapter. By and large, however, the emphasis in the book is on constraints that take the form of making known options less available rather than that of blocking knowledge about their existence.

The book as a whole is concerned with two types of beneficial constraints. First, there are constraints that benefit the agent who is constrained but that are not chosen by the agent for the sake of those benefits. This is the topic of I.5, and a central issue in Chapters II and III. The constraints may be chosen by the agent for some other reason, chosen by some other agent, or not be chosen by anyone at all but simply be a fact of life that the agent must respect. I shall refer to these as *incidental constraints*. Let me give two brief examples, both from the "fact of life" category. One concerns the need to shoot films in black and white before the invention of color photography. It has been argued, as we shall see in Chapter III, that this constraint made for greater artistic creativity. A similar argument has been applied to the social sciences. In a comment on James Coleman's work, Aage Sørensen claims that the invention of high-power computers came at the detriment of sociological theory, when and because "the data limitations and computational limitations that inspired Coleman to enormous creativity and imagination in developing and applying the models were removed."[6]

Second, there are constraints that an agent imposes on himself for the sake of some expected benefit to himself. This is the main topic of the present chapter, and an important topic of Chapters II and III as well. In *Ulysses and the Sirens* I referred to this phenomenon as "precommitment" or "self-binding." Others have used the terms "commitment" or "self-commitment." In the present volume, I often retain my earlier terminology. When the emphasis is on the constraints that are created rather than on the act of creating them, I refer to them as *essential constraints*.

Essential constraints are defined in terms of *expected* benefits, incidental constraints by the *actual* benefits they provide to the agent. (I ignore cases in which A constrains B with the intention of benefiting B but no benefits are in fact provided.) Whereas the establishment of essential constraints is always explained by the expectation of benefit, the actual benefits of incidental constraints may or may not enter

6. Sørensen (1998), p. 255.

into their explanation. In I.5 I discuss the view that emotional constraints on behavior emerge from natural selection by virtue of their beneficial impact on reproductive fitness. If that view is correct, the effects of the constraints have explanatory force. In Chapter II, I mention that consequences of constitutional arrangements that were not in the minds of the framers may come to be acknowledged at a later time, and then serve as reasons to maintain those arrangements if the grounds on which they were originally adopted no longer obtain. In that case, too, the effects of the incidental constraints would have explanatory force.

Because of the pervasive use of functional explanation in the social sciences, it is easy to commit one of two closely related fallacies: to confuse incidental and essential constraints, and to assume without argument that the benefits of incidental constraints always tend to explain them.[7] The human mind, it seems, is simply very reluctant to admit the idea of accidental or non-explanatory benefits.[8] In II.1 I mention some of my own past confusions, and I am not alone in this respect. Thus one task of the present book is to demarcate, as clearly as I can, intentional self-binding from other ways in which beneficial constraints can come about. Another task is to examine whether there can be constraints that are, as it were, essentially incidental. An agent might be *unable to make himself unable* to act in a certain way, and yet find himself constrained, to his benefit, by the force of circumstances or through an act of another agent.

As mentioned, the present chapter is mainly concerned with essential constraints, or self-binding in the standard intentional sense. More specifically, I shall discuss an agent's desire to create obstacles to his or her future choice of some specific option or options. In this perspective, precommitment embodies a certain form of *rationality over time*. At time 1 an individual wants to do A at time 2, but anticipates that when time 2 arrives he may or will do B unless prevented from doing so. In such cases, rational behavior at time 1 may involve precautionary measures to prevent the choice of B at time 2, or at least to make that choice less likely. The present chapter is a survey of the why and how of precommitment – of the *reasons* why people might want to precommit themselves and of the *devices* they have at their disposal.[9]

7. Elster (1983a), Ch. 2. 8. Elster (1983b), Ch. II.10.
9. In this book I assume a simple conflict between a short-term and a long-term interest. In the model presented by Ainslie (1992), the mind contains a whole population of

Reasons for precommitment

Devices for precommitment	Overcome passion	Overcome self-interest	Overcome hyperbolic discounting	Overcome strategic time-inconsistency	Neutralize or prevent preference change
Eliminating options	I.2 I.7		I.3 I.7	I.4	I.6
Imposing costs	I.2 I.7	I.5	I.3 I.7	I.4	
Setting up rewards	I.7		I.7		
Creating delays	I.2 I.7		I.3	I.4	
Changing preferences	I.2 I.7				
Investing in bargaining power				I.4	
Inducing ignorance	I.2 I.7		I.3	I.4	I.6
Inducing passion		I.5		I.4	

Table I.1

In I.2 I consider the traditional view that precommitment is an instrument to protect us against passion. Then I discuss the more recent argument that precommitment can help us overcome the problem

interests, with time horizons ranging from fractions of a second to a lifetime. In that case, more complex phenomena become possible, such as alliances between a short-term and a long-term interest against intermediate-range interests. Elsewhere I have used the following example to illustrate this idea: "I wish that I didn't wish that I didn't wish to eat cream cake. I wish to eat cream cake because I like it. I wish that I didn't like it, because, as a moderately vain person, I think it is more important to remain slim. But I wish I was less vain. (But do I think that only when I wish to eat cake?)" (Elster 1989a, p. 37).

of time-inconsistency, be it due to hyperbolic discounting (I.3) or to strategic interaction (I.4). In Section I.5 I consider the argument that passion can serve as such a device. Rather than being an obstacle to the rational pursuit of self-interest (I.2), passion can help us overcome our tendency to act according to immediate self-interest when doing so is against our long-term interest. In I.6, I consider some variations on the "Russian nobleman" case introduced by Derek Parfit, with main emphasis on why fundamentalists might want to insulate themselves from the modern world in order to prevent preference change. In I.7 I survey the numerous forms of self-binding strategies adopted by addicts. In I.8 I discuss some reasons why precommitment, when feasible, might not be desirable; and when desirable, might not be feasible. Here, I also discuss somealternatives to precommitment.

Not all devices for precommitment can serve all reasons for precommitment. Table I.1 indicates some of the main connections between reasons and devices for precommitment, and helps the reader to locate the sections where the various cases are discussed.[10]

I.2. PASSION AS A REASON FOR SELF-BINDING

When we act under the influence of passions, they may cause us to deviate from plans laid in a cooler moment. Knowledge of this tendency creates an incentive to precommit ourselves, to help us stick to our plans. Here, I use "passion" in an extended sense that covers not only the emotions proper such as anger, fear, love, shame, and the like, but also states such as drunkenness, sexual desire, cravings for addictive drugs, pain, and other "visceral" feelings.[11]

From Aristotle to some time in the twentieth century, the most frequent antonym of passion was *reason*, understood as any impartial – dispassionate or disinterested – motivation.[12] A person who wishes to behave justly toward others but fears that his anger

10. In this chapter as well as in the following, I ignore *randomization* as a form of individual or collective precommitment, to avoid repeating what I have written elsewhere on the topic (Elster 1989b, Ch. II). A brief survey of the issue is offered in the discussion of randomization in the arts (III.8).
11. For a discussion of the role of emotions in the explanation of behavior, see Elster (1999a), notably Appendix to Ch. IV. For the place of viscerality in the explanation of behavior see Loewenstein (1996, 1999).
12. For a fuller discussion see Elster (1999a), notably Ch. V.

might get the best of him is advised to precommit himself in one of the ways to be discussed shortly. Among modern economists, the most frequent antonym of passion is *rational self-interest*. A person who fears that anger might cause him to act in ways contrary to his self-interest would do well to avoid occasions on which this emotion might be triggered. A reasonable agent precommits himself against anger so as not to hurt others, whereas an agent moved by rational self-interest does so in order not to hurt himself. Later in this chapter, we find examples of precommitment motivated by either of these dispassionate attitudes. As we shall see, other cases also arise. An agent in the grip of passion may precommit himself against another passion, against the rational pursuit of self-interest, or against reason. A rational and self-interested agent may even precommit himself against his own rationality.

Effects of Passion

I shall distinguish among four ways in which passions may cause a discrepancy between plans – whether based on reason or on rational self-interest – and behavior. They may do so by distorting cognition (inducing false beliefs about consequences), by clouding cognition (blotting out awareness of consequences), by inducing weakness of will (options with worse perceived consequences are chosen over those with better consequences), or by inducing myopia (changing the decision weights attached to the consequences). Whereas the first two mechanisms involve cognitive irrationality, the last two need not. Whereas the third involves a motivational irrationality, the fourth need not. All but the second leave the agent with some capacity to respond to incentives.

(i) Passion may distort our thinking about the consequences of the behavior. This was in fact Aristotle's definition of emotion: "The emotions are those things through which, by undergoing change, people come to differ in their judgments, and which are accompanied by pain and pleasure, for example, anger, pity, fear, and other such things and their opposites" (*Rhetoric* 1378a 21–22). Although this does not provide a good definition of emotion – there are too many exceptions, some of them noted by Aristotle himself[13] – it accurately

13. Aristotle counts hatred as an emotion (*Rhetoric* 1382a 2–16), but also says that hatred can leave judgment unaffected (*Politics* 1312b 19–34). See also Elster (1999a), Ch. II.2.

captures many cases of emotionally induced wishful thinking and self-deception. Emotions can affect "probability and credibility estimates" concerning events outside one's control.[14]

This mechanism may also apply when the passion in question is a craving rather than an emotion. In an example from David Pears, a

> driver goes to a party and he judges it best to stop at two drinks in spite of the pleasure to be had from more, because there is nobody else to take the wheel on the way home. Nevertheless, when he is offered a third drink, which, we may suppose, is a double, he takes it. How can he? Easily, if the wish for a third drink biases his deliberation at the party before he takes it. For example, he might tell himself, against the weight of the evidence, that it is not dangerous to drive home after six measures of whiskey, or he might forget, under the influence of his wish, how many drinks he has already taken.[15]

(ii) The passion may be so strong as to crowd out all other considerations.[16] Before an unpleasant encounter, I may resolve to keep my cool. Yet when provoked to anger, I lash out without pausing to consider the consequences. It is not that I do not know the consequences or that I have false beliefs about them: I simply do not, when acting, keep them before my mind. This is Aristotle's conception of weakness of the will (or one of his conceptions), "admitting the possibility of having knowledge in a sense and yet not having it, as in the instance of a man asleep, mad, or drunk. But now this is just the condition of men under the influence of passions; for outbursts of anger and sexual appetites and some other such passions, it is evident, actually alter our bodily condition, and in some men even produce fits of madness. It is plain, then, that incontinent people must be said to be in a similar condition to these."[17]

(iii) I may know even at the time that I am acting against my better judgment. When offered the third drink at the party, the driver may accept it and yet think *as he does so* that he shouldn't. Although the

14. Frijda (1986), pp. 118–21.
15. Pears (1985), p. 12. Along similar lines Rabin (1995) argues that "we may over-eat not because we consciously sanction over-weighting current ... well-being over future well-being, but because we systematically deceive ourselves in ways that support immediate gratification."
16. The passions may also *preempt* all other considerations. As explained in LeDoux (1996) and summarized in Ch. IV.2 of Elster (1999a), there is a direct pathway from the sensory apparatus to the emotional apparatus in the brain that bypasses the thinking part of the brain entirely, so that when the sensory signal arrives to the latter some milliseconds later, the organism has already started to react.
17. *Nicomachean Ethics* 1147a.

reasons against drinking are stronger qua reasons than the reasons for drinking, the latter have a stronger causal efficacy qua sheer psychic turbulence. Something like this is the view of weakness of will that has been made prominent by Donald Davidson and that, in one way or another, is at the center of most recent philosophical discussions of the subject.[18]

A problem with this third view is the difficulty of finding reliable evidence that the agent really thought that, all things considered, he should not take the drink. It is easy enough to find independent evidence that the driver, *before* going to the party, did not want to have more than two drinks. He may have told his wife, for instance, "Stop me if I have more than two drinks." *After* the party, too, he may regret his behavior as contrary to his real interest and take steps to ensure that it doesn't happen again. But how can we know that this all-things-considered judgment exists at the very moment that he is accepting the third drink? By assumption, there is no observable behavior that can support this interpretation. How can we exclude, for instance, the possibility of a last-second preference reversal due to hyperbolic discounting (I.3)? The agent might retain an accurate appreciation of the consequences of his behavior yet weigh them differently from the way he did before. Because Davidson offers a transcendental argument – how is acting against one's better judgment at the time of action *überhaupt möglich*? – it is disturbing that the empirical premise is so hard to establish.[19]

(iv) A person in a state of passion may weigh the consequences of behavior differently from the way he does in a calmer mood. An addict, for instance, may have accurate beliefs about the disastrous effects of the drug on his or her body or purse, and yet ignore them because of an addiction-induced increase of the rate of time discounting.[20] The urgency and impatience often associated with emotion can have the same effect. If I have the choice between seeing

18. Davidson (1970).
19. Cp. Montaigne (1991), p. 1161: "I realize that if you ask people to account for 'facts,' they usually spend more time finding reasons for them than finding out whether they are true. . . . They skip over the facts but carefully deduce inferences. They normally begin thus: 'How does this come about?' But does it do so? That is what they ought to be asking." See also Merton (1987) for the need to "establish the phenomenon" before one sets out to explain it. For a fuller discussion, see Elster (1999d).
20. Becker (1996), p. 210; O'Donoghue and Rabin (1999a); Orphanides and Zervos (1998).

the person I love for ten minutes today and seeing him or her for one hour tomorrow, I might opt for the former option. The effect of passion in such cases is to induce myopia, rather than to distort and cloud cognition or to make us act against our better judgment. Note that the passions that induce myopia may themselves be either durable or transient. In the latter case, a *short-lived* passion causes the agent to have a *shortsighted* idea of his interests. Although the ideas of "momentary passion" and "immediate interest" are conceptually distinct, they are often causally linked (see also II.7).

In self-deception, weakness of will, and myopia – cases (i), (iii), and (iv) – the agent is *reward-sensitive*. This is not to say that he is rational, only that he is capable of exercising choice by weighing consequences against one another.[21] If the delayed negative effects of a certain behavior would be truly disastrous, the agent is less likely to fool himself into believing that they do not exist, less likely to accept them against his own better judgment, and less likely to let them be dominated by short-term reward. It is only in case (ii) that passion makes the agent entirely deaf to incentives beyond the desires of the moment.

Precommitment Against Passion

These differences have obvious implications for strategies of self-binding. When the agent is able to take account of incentives even in the heat of passion, precommitment can take the form of attaching a cost or a penalty to the choice one wants to avoid making. If you think you might get too drunk or too amorous at the office party, you can increase the costs of doing so by taking your spouse along. In *Lucien Leuwen*, Mme de Chasteller takes care to see Lucien only in the company of a chaperone, to make it prohibitively costly to give in to her love for him. By contrast, some passions are so strong that the only practical way of neutralizing them is to avoid occasions that trigger them.[22] In *La Princesse de Clèves*, the princess flees the court for the countryside to avoid the temptation of responding to the overtures of the Duc de Nemours; even later, when her husband is dead and she is free to remarry, she stays away. "Knowing how circumstances affect

21. For a fuller discussion, see Ch. 5.1 of Elster (1999b).
22. In Ch. 5 of Elster (1999b) I discuss whether there are cases in which this statement is true even when the word "practical" is omitted.

the wisest resolutions, she was unwilling to run the risk of seeing her own altered, or of returning to the place where lived the man she had loved."[23]

Anger is perhaps the most important of these blind-and-deaf passions. It may be unique among the emotions in its capacity to make us forget even our most vital interests. According to Seneca, anger is "eager for revenge even though it may drag down the avenger along with it."[24] "Who sees not," Hume asked, "that vengeance, from the force alone of passion, may be so eagerly pursued as to make us knowingly neglect every consideration of ease, interest, or safety?"[25] Clearly, if angry people are willing to disregard even a risk to their lives they will not be deterred by any additional disincentives. As we shall see in I.5, this disregard for consequences that characterizes the angry man may also serve his interest, that is, have good consequences. Here I shall focus on the need to contain anger, drawing heavily on various observations in Montaigne's *Essays*.[26]

It is a commonplace that other people can detect that one is angry or in love before one knows it oneself. When one is in love for the first time, as Madame de Rênal in Stendhal's *Le rouge et le noir*, one may live the emotion fully and innocently until one day the realization strikes: "I am in love." There is no self-deception involved, merely unawareness.[27] In anger, too, the emotion often has to reach a certain threshold before awareness occurs. At the same time, episodes of anger are often characterized by a "point of no return" beyond which self-control is of no avail.[28] The reason that anger is so hard to control, according to Montaigne, is that the second threshold occurs before the first. "The infancies of all things are feeble and weak. We must keep our eyes open at their beginnings; you cannot find the danger then because it is so small; once it has grown, you cannot find the cure."[29] In other words, the dynamics of anger (and of love) is subject to the dilemma illustrated in Figure I.1.

If this is right, and I think it often is, a self-control rule such as counting to ten is not likely to be a good remedy against anger. It is an *advice*, and not a very effective one; not a *device*. Although delay

23. Lafayette (1994), p. 108; see also Shattuck (1996), pp. 114–21.
24. *On Anger*, I.i.1. 25. Hume (1751), Appendix II.
26. See also Elster (1996) and Elster (1999a), Ch. II.3.
27. For more extensive discussions of unacknowledged emotions, see Elster (1999a), Chs. II.3, III.2, and IV.2.
28. Frijda (1986), pp. 43–45, 91, 241.
29. Montaigne (1991), p. 1154; see also Ekman (1992), p. 47.

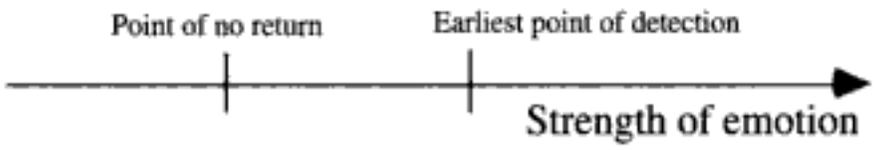

Fig. I.1

devices can be effective forms of precommitment (Chapter II), they have to be set up ahead of time rather than being left to the discretionary control of the agent at the moment he or she needs them. As Thomas Schelling writes, "If I am too enraged to mind my behavior, how can I make myself count to ten?"[30] In theory, delay devices might be used to counteract passion, in the wide sense that also includes cravings for addictive substances. If I want to limit my drinking to social occasions but do not trust myself to do so, I could keep my liquor in a safe with a delaying device so that I would have to set it six hours ahead of time to get access to it. In practice, I have not come across any examples of this strategy. Perhaps it is too expensive – a safe with a timer costs about $1,000 (see also I.8).

Legislation that requires a trial separation before a final divorce might seem to be an exception. The delay allows extramarital passion to calm down and reason to regain the upper hand. Yet with an exception that I discuss later, the delay is always imposed by the state rather than chosen by the spouses themselves at the time of marriage. The legal rights and duties of marriage come as a package. Even when the fact that two people marry shows that they prefer the "delay package" over mere cohabitation, they might have chosen an "instant package" with the possibility of divorce at will had that been available. Marriage would be an essential constraint rather than an incidental constraint only if the delay package was preferred both to cohabitation and to the instant package, because only in that case could restrictions on the freedom to divorce be the *motive* for marrying. Although I have in the past, along with others, used marriage as a standard case of precommitment, I now believe this to be a mistake, or at least misleading.[31]

30. Schelling (1999). Watson (1999) describes "the predicament of self-control" in similar terms: "Techniques of self-control often work by maintaining one's focus against . . . distractions. But employing those techniques already takes an amount of focus that tends to dissolve precisely where it is needed."

31. See also Montaigne (1991), p. 698: "We thought we were tying our marriage-knots more tightly by removing all means of undoing them; but the tighter we pulled the knot of constraint the looser and slacker became the knot of our will and

Gun-control legislation also works by imposing delays between the request for a gun and its delivery.[32] Although these laws serve to protect citizens against fits of murderous passion, it is less plausible to view them as instances of intentional precommitment. Legislators or voters in a referendum are probably much more concerned with protecting themselves and the population at large against others than with self-binding in a literal sense (see also Chapter II). Those who are bound may welcome the ties – or not.

Montaigne noted that when emotions emerge suddenly and strongly we cannot control them, whereas when they are weak enough to be kept under control we may not notice them. He did not think that the problem was necessarily insuperable: "If each man closely spied upon the effects and attributes of the passions which have rule over him as I do upon those which hold sway over me, he would see them coming and slow down a little the violence of their assault. They do not always make straight for our throat: there are warnings and degrees."[33] Yet almost all his practical advice takes a different form: we should avoid the occasions for strong emotions rather than try to stifle them when they arise. "For common souls like ours there is too much strain" in trying to resist or control the emotions.[34]

Emotions are triggered by external events, but only if they come to our knowledge. To prevent the emotions, therefore, we may either ensure that these events do not occur or, if they do, that we do not come to know about them. Montaigne adopted both strategies: "I shun all occasions for annoyance and keep myself from learning about things going wrong."[35] Concerning the first strategy, he refers to the example of King Cotys: "He paid handsomely when some beautiful and ornate tableware was offered to him, but since it was unusually fragile he immediately smashed the lot, ridding himself in time of an easy occasion for anger against his servants." For himself, he adds, "I have likewise deliberately avoided confusions of interests; I have not sought properties adjoining those of close relatives or belonging to

affection. In Rome, on the contrary, what made marriages honoured and secure for so long a period was freedom to break them at will. Men loved their wives more because they could lose them; and during a period when anyone was quite free to divorce, more than five hundred years went by before a single one did so." Montesquieu makes the same argument in *Lettres persanes* (Letter 116). See also Phillips (1988), Ch. 5.2.

32. I am grateful to David Laitin for bringing this issue to my attention.
33. Montaigne (1991), p. 1219. 34. Ibid., p. 1148.
35. Montaigne (1991), p. 1075.

folk to whom I should be linked by close affection; from thence arises estrangement and dissension."[36] In *On Anger*, Seneca offers similar advice:

> While we are sane, while we are ourselves, let us ask help against an evil that is powerful and oft indulged by us. Those who cannot carry their wine discreetly and fear that they will be rash and insolent in their cups, instruct their friends to remove them from the feast; those who have learned that they are unreasonable when they are sick, *give orders that in times of illness they are not to be obeyed*.[37]

The second strategy is to shield oneself from knowledge of events that might make one angry or otherwise emotionally disturbed.[38] Referring to an old man whose servants abuse him without his knowing it, Montaigne observes that it "would make a good scholastic debate: whether or not he is better off as he is."[39] I suspect Montaigne would say that he isn't, and that the key factor is that the old man did not *deliberately* blind himself to what his servants were doing. This is what Montaigne himself did: "I prefer people to hide my losses and my troubles from me. . . . I prefer not to know about my estate-accounts so as to feel my losses less exactly. Whenever those who live with me lack affection and its duties I beg them to deceive me, paying me by

36. Ibid., pp. 1147–48. 37. *On Anger*, III.xiii.5; my italics.

38. Tyler Cowen (personal communication) notes that some investment theories say that one should never look at one's portfolio. Whereas this advice could be based on the tendency to discount the future hyperbolically (see discussion at the end of I.3) or by a tendency to place excessive emphasis on recently acquired information (Bondt and Thaler 1985), it could also be justified by a tendency to react emotionally to good or bad news.

39. Montaigne (1991), p. 442. This observation prompts a couple of comparative remarks. First, note that the issue raised by Montaigne differs from the question whether it would be better to believe one's servants to be honest while in reality they are not or to believe them dishonest while in reality they are honest. Thus Gibbard (1986), p. 169, comments, "A jealous husband may . . . prefer a 'fool's hell' in which his suspicions rage but his wife is in fact faithful, to a 'fool's paradise' in which his suspicions are allayed but in fact he is unknowingly cuckolded." Second, we may compare Montaigne's dilemma with a similar conundrum raised by Tocqueville (1969, p. 317) in his discussion of American slavery. Tocqueville notes that "the Negro . . . admires his tyrants even more than he hates them and finds his joy and pride in a servile imitation of his oppressors" and asks whether he should "call it a blessing of God, or a last malediction of His anger, this disposition of the soul that makes men insensible to extreme misery and often even gives them a sort of depraved taste for the cause of their afflictions." Tocqueville and Montaigne both make the point that if well-being is bought at the cost of autonomy the price may be too high.

putting a good face on things."[40] Seneca provides another example: "The great Gaius Caesar . . . used his victory most mercifully; having apprehended some packets of letters written to Gnaeus Pompeius by those who were believed to belong either to the opposing side or to the neutral party, he burned them. Although he was in the habit, within bounds, of indulging in anger, yet he preferred being unable to do so."[41]

Another situation where ignorance may be bliss is in matters of marital faithfulness. Montaigne writes, "Curiosity is always a fault; here it is baleful. It is madness to want to find out about an ill for which there is no treatment except the one which makes it worse and exacerbates it."[42] In fact,

> We should use our ingenuity to avoid making such useless discoveries which torture us. It was the custom of the Romans when returning home from a journey to send a messenger ahead to announce their arrival to their womenfolk so as not to take them unawares. That is why there is a certain people where the priest welcomes the bride and opens the proceedings on the wedding-night to remove from the groom any doubts and worries about whether she came to him virgin or already blighted by an *affaire*.[43]

This is not like avoiding going on the scales to see if one has gained weight, or neglecting to make an appointment with a doctor to find out if one has some dread disease. Those self-deceptive practices are matters of individual information-avoidance, whereas Montaigne here is referring to custom and public policy.

The phenomenon of *regret-avoidance* can be an instance of either strategy. Consider an example suggested by Robert Sugden (personal communication). Suppose first that a driver who sees that traffic is very dense on the highway is deliberating whether to leave it for a smaller road. As the smaller road crosses the highway some miles further away, he knows that he will learn whether his decision was justified, and that if traffic on the highway is in fact going smoothly he

40. Montaigne (1991), pp. 731–32. There is another element at work too: "When I am on my travels, whoever has my purse has full charge of it without supervision. He could cheat me just as well if I kept accounts, and, unless he is a devil, by such reckless trust I oblige him to be honest" (pp. 1078–79). By "consciously [encouraging his] knowledge of his money to be somewhat vague and uncertain" (p. 1079), he worries less *and* ensures that there is less to be worried about. See also note 111.

41. *On Anger*, II.xxiii.4. 42. Montaigne (1991), p. 982.

43. Ibid., p. 983.

will feel regret. The regret is a possible cost associated with the choice of the small road that comes on top of the cost of being delayed. On some occasions, anticipation of regret may tip the balance in favor of staying on the highway.[44] Suppose next, however, that there is another small road, which is more circuitous than the first one but which does not cross the highway. The driver might then be tempted to take the slow road, anticipating that if he takes the fast road he might learn – to his regret – that traffic on the highway is in fact so smooth that he would have been better off staying there. Sometimes, people take action to avoid learning whether they have grounds for regret.

Let me mention some further strategies for containing the passions. One is to fight emotion with emotion. When the Argonauts sailed in the waters later visited by Ulysses, Orpheus sang so divinely that none of them listened to the Sirens. Or one might try to fight anger with some other emotion such as fear. In *On Anger*, Seneca raises a possible objection to his claim that it is impossible to retain one's anger: "Do not men sometimes even in the midst of anger allow those whom they hate to get off safe and sound and refrain from doing them injury?" and responds as follows: "They do; but when? When passion has beaten back passion, and either fear or greed has obtained its end. Then there is peace, not wrought through the good offices of reason, but through a treacherous and evil agreement between the passions."[45]

There is a related, but ambiguous, suggestion in Descartes' *Les Passions de l'âme*. He argues that some individuals are never in a position to use their willpower, "because they never let their will fight with its own weapons, using only those with which some passion provides it to resist other passions" (Article 48). Does he mean that these individuals can mobilize (occurrent) emotions at will to fight other emotions? Or that they can deliberately cultivate certain emotional dispositions that will trigger the occurrent emotions that are needed to fight others? The former idea is implausible: by and large, I side with the traditional view that occurrent emotions are involuntary rather than actively chosen.[46] The latter is also somewhat implausible. If emotional dispositions really are within the scope of

44. Loomes and Sugden (1987); Bell (1982).
45. *On Anger*, I.viii.7.
46. Elster (1999a), Ch. IV.3.

character planning, wouldn't it make more sense to get rid of those we find undesirable than to create new ones to counteract them?

Many writers within the Buddhist or Stoic traditions have indeed argued that we should fight the passions by getting rid of the tendency to experience them.[47] One of their concerns was to overcome the fear of death, which was also a central theme for Montaigne.[48] In an early essay, "To philosophize is to learn how to die," the emphasis is on how to master the fear of death by anticipating it. In another, "The taste of good and evil things depends on our opinion," he argues, "What we chiefly fear in death is what usually precedes it: pain." Montaigne first claims that this connection is self-deceptive: "It is our inability to suffer the thought of dying which makes us unable to suffer the pain of it . . . As reason condemns our cowardice in fearing something so momentary, so unavoidable, so incapable of being felt as death is, we seize upon a more pardonable pretext." This argument is soon given up, when Montaigne grants "that pain is the worst disaster that can befall our being." In the argument he then goes on to offer, however, the emphasis is more on ignoring or enduring pain when it occurs than on mastering the fear of pain before it occurs.[49]

In the later essays Montaigne came to view things quite differently. "To speak truly, we prepare ourselves against our preparations for death! Philosophy first commands us to have death ever before our eyes, to anticipate it and to consider it beforehand, and then she gives us rules and caveats in order to forestall our being hurt by our reflections and our foresight."[50] In the words of Nico Frijda, the "net effects of anticipation result from the opposing factors of stress produced by anticipatory fear and reduction of surprise with

47. To eliminate the passions, these writers recommended some form of character planning by purely psychic means, the mind acting on itself to change itself (Kolm 1986). Others have advocated or adopted the radical means of cutting off the passions at their physical root. See, for instance, a recent *New York Times* headline (April 5, 1996): "Texas Agrees to Surgery for a Molester. Soon to Leave Prison, Man Wants Castration to Curb His Sex Urge."

48. The main emphasis in these discussions is to eliminate fear as a source of suffering. Yet it is clear that fear was also seen as a source of undesirable behavior. In discussing why wills are rarely drawn up properly, for instance, Montaigne (1991), p. 93, says that it is "No wonder that [ordinary people] often get caught in a trap. You can frighten them simply by mentioning death; and since it is mentioned in wills, never expect them to draw one up before the doctor has pronounced the death-sentence. And then, in the midst of pain and terror, God only knows what shape their good judgement kneads it into!"

49. The quoted passages are in Montaigne (1991), pp. 58–59.

50. Ibid., p. 1190.

its possibilities of anticipatory coping."[51] One might say, therefore, that Montaigne, from viewing the net effect of anticipating death as positive, came to see it as negative. Yet his argument is actually more subtle: the main object of Philosophy is not to enable us to cope with an independently existing threat, but to reduce the fears she has herself created. Philosophers, like doctors or lawyers, pride themselves on their ability to put out the fires they have lit.[52]

Passionate Precommitment

Although I have been assuming that the earlier decision to precommit oneself against a later passion is itself a dispassionate one, this is not always the case. Sometimes, people precommit themselves in a moment of passion to prevent themselves from yielding to another passion at a later time. This case is illustrated by the 1997 Louisiana legislation on "covenant marriage," an optional form of marriage that is harder to enter and harder to leave than the regular marriage.[53] Under the traditional system, a couple is entitled to a no-fault divorce after six months of separation. Under the covenant marriage, two years are required. Commenting on the new option, Ellen Goodman cites Amy Wax, a law professor at the University of Virginia, as "worried that newlyweds would 'bind themselves by more stringent terms and live to regret it when Dr. Jekyll turns into Mr. Hyde.'" Goodman also writes, "The covenant marriage mandates premarital counseling. But even Barbara Whitehead, the author of 'The Divorce Culture,' acknowledges ruefully: 'It's impossible to get them to contemplate troubles, adversity, conflict, especially if it's their first marriage and they are fairly young. It's not a teachable moment.'"[54] In a state of infatuation, young people may overestimate the benefits and underestimate the costs of making themselves unable to yield to an extramarital passion later. The fact that by not choosing the covenant marriage when it is available one might send the wrong signal to one's partner might also contribute to excessive use of this option.

51. Frijda (1986), p. 293. 52. See also Montaigne (1991), pp. 1160, 1176.
53. In the debates over the British Divorce Reform Act of 1857, Gladstone unsuccessfully argued for a somewhat similar two-track system, according to which couples would have the choice between a Christian marriage without the possibility of divorce and civil marriage (which already existed) with divorce (Stone 1990, p. 379).
54. Goodman (1997).

Another example can be taken from Racine's *Andromaque*. When Hermione learns that Pyrrhus has definitely rejected her, she calls for Orestes and asks him to avenge her:

> While he still lives, fear lest I pardon him.
> Suspect my wavering anger till his death.
> Tomorrow I may love him if today
> He dies not. (1198–1200)[55]

Hermione claims that unless Orestes kills Pyrrhus immediately, she might come to love him again. Hence, the murder is a kind of precommitment: kill him, so that I cannot love him. After Orestes has carried out the assignment and tells Hermione what he has done, she responds as follows:

> *Hermione*
> Why did you murder him? What did he do?
> Who told you to?
> *Orestes*
> God! Did you not yourself
> Here, one short hour ago, ordain his death?
> *Hermione*
> Ah! how could you believe my frantic words?
> And should you not have read my inmost thoughts? (1542–46)[56]

First, Hermione tells Orestes to kill Pyrrhus as a precommitment against her possible backsliding. After the deed she tells him that he should have understood that she did not really want to be precommitted, and that her inauthentic self was the precommitting one, not the self against which the precommitment was directed. In her momentary passion for revenge, she precommits herself against her more enduring love.

55. Tant qu'il vivra, craignez que je ne lui pardonne.
Doutez jusqu'à sa mort d'un courroux incertain:
S'il ne meurt aujourd'hui, je puis l'aimer demain.

56. *Hermione*
Pourquoi l'assassiner? Qu'a-t-il fait? A quel titre?
Qui te l'a dit?
Oreste
O dieux! ne m'avez-vous pas
Vous-même, ici, tantôt, ordonné son trépas?
Hermione
Ah! fallait-il en croire une amante insensée?
Ne devais-tu pas lire au fond de ma pensée?

Second-Order Desires

The idea of precommitment is often linked to that of second-order desires.[57] Suppose a person wants to quit drinking, but finds himself torn between his desire to drink and his desire for all the things that drinking prevents him from doing. This conflict does not necessarily generate a second-order desire not to have the desire to drink. In general, when we desire two incompatible things we decide which desire is the more important and act on it. Thus so far I have been able to handle my desire for butter pecan ice cream without wishing I didn't have it. Suppose, however, that I find myself constantly acting against my better judgment, either in Davidson's sense or in Ainslie's sense (I.3). In that case – but in that case only – it would be rational to form a desire not to have a desire for butter pecan ice cream, and take steps to get rid of it.[58] Second-order desires, rather than being constitutive of what it means to be a person,[59] may simply reflect weakness of will.[60] Moreover, even a weak-willed person would not necessarily form and act on a second-order desire if he or she could deploy some other precommitment strategy. I could ask my wife to throw out butter pecan ice cream whenever she finds it in the fridge rather than try to get rid of my desire for it. What matters is what we do, not what we desire to do.

This statement goes against the view that in matters of personal morality, we should not even desire to do what, all things considered, we think we should not do. We should not wish other people to be less successful, or desire to get even when they insult us, or lust after their spouses. This attitude, which amounts to a self-imposed thought police, can do and has done great harm. I believe that what Kant says about envy is true of a great many other urges: "Movements of envy are ... present in human nature, and only when they break out do

57. The following owes much to discussions with Olav Gjelsvik.
58. This is a simplification. Even if I am successful in sticking to my better judgment, the effort to overcome temptation may be so strenuous that I am better off if I reduce it by investing in preference change. (I am grateful to Ole-Jørgen Skog for this point.)
59. This is the view advocated in Frankfurt (1971).
60. By contrast, the *potential* for having such desires may indeed be constitutive of personhood. (I am grateful to David Johnston for this point.) Also, there can be second-order desires without weakness of will if what I wish is to acquire a first-order desire I do not have, such as the desire for listening to classical music, rather than to rid myself of a desire that I do have. (I am grateful to Olav Gjelsvik for this point.)

they constitute the abominable vice of a sullen passion that tortures oneself and aims . . . at destroying others' good fortune."[61] Any guilt we might experience in having these feelings is pointless suffering.[62] The remedy is not to go the other extreme and advocate the acting out of all our desires, but rather to recognize the existence of conflicting desires, get our priorities right, and stick to them.

Although this analysis does not need the idea of second-order desires, it may require some kind of asymmetry between the conflicting urges. Suppose I have been a heavy drinker and am trying to quit, judging that my desire to drink is less weighty than my desire for all the things that drinking prevents me from doing. Suppose, moreover, that this is a judgment I make when sober, but that I make the very opposite judgment when under the influence of alcohol. Which self, if any, is right? To a first approximation, we may answer by using the capacity for self-binding as a criterion.[63] If we observe the sober self trying to bind the drunken self but never observe the drunken self trying to bind the sober self, we may reasonably identify the former self with the person's real interest. And there often seems in fact to be an asymmetry of this kind. By virtue of their strength the passions may induce a temporary neglect of the future that prevents the person who is in their grip from responding strategically to strategic moves he or she might later make to curb them. When the person tries to implement his long-term interest, he is aware of the obstacles created by his short-term interest, but not vice versa.[64] This asymmetry has nothing to do with second-order desires: the short-term interest is the object of a cognition, not of a desire.

61. Kant (1785), pp. 576–77. A vivid expression of this idea is given by a Tahitian pastor cited in Levy (1973), p. 332: "You think about sleeping with that *vahine*, about committing adultery with that *vahine*, but you do not commit adultery with her – there is no difficulty at all. Because the thought inside of you – all people have had that thought inside of them, there is not one man without that thing inside of him. What can be done? That thought exists within you, because it does not stop. It is the same as some machine that keeps running inside of you."
62. See also Elster (1999a), Ch. III.2.
63. Because of the issues mentioned in note 9, it is valid only to a first approximation.
64. See also Skog (1997), p. 268, for a statement to this effect. Cowen (1991), p. 362, argues, by contrast, that "the actions of an impulsive self are not limited to myopic forms of immediate gratification and may involve sophisticated strategic maneuvers." With the exception of the example from Merton cited later, Cowen's illustrations of this claim are hypothetical and, to my mind, unconvincing. Also, most of his discussion focuses on cases in which strategic action by the impulsive self would be *desirable*, rather than on cases in which it is likely to occur.

There are cases, nevertheless, in which cravings or emotions seem to be capable of taking strategic measures to ensure their own satisfaction. A case of this sort arose in the Belgian war trials after World War II. In most of the countries that had been occupied by the Germans, there was a tendency for similar crimes of collaboration to be treated more leniently in, say, 1947 than in 1944 or 1945.[65] Some collaborators were executed for crimes that a few years later would at most have gotten them twenty years in prison.[66] In Belgium, this pattern was anticipated and to some extent taken into account by the organizers of the trials. On the basis of the experience from WWI, "it was believed that after a while, the popular willingness to impose severe sentences on the collaborators would give place to indifference." Hence the Belgians wanted the trials to proceed as quickly as possible, before passion was replaced by a more dispassionate attitude.[67]

A similar case, also related to World War II, was noted by Robert Merton. In Tyler Cowen's summary,

> an example of impulsive precommitment is given in Robert Merton's study of social pressures. Merton notes the ephemerality of many persons' desires to contribute to the American war effort during the Second World War. The desire to contribute was strong only immediately after hearing radio appeals for funds. Merton's study of contributors revealed that 'in some instances, listeners telephoned at once precisely because they wished to commit themselves to a bond before inhibiting factors intervened'. After making such telephone calls, persons were required to fulfill commitments that had been undertaken by their impulsive selves.[68]

As these examples show, it is possible for an agent in the heat of passion to precommit himself against the predictable tendency for passion to abate after a while. Instances of such behavior may be rare, but they seem to exist. While they do not refute the overall difference in capacity for self-binding between passion on the one hand and

65. Tamm (1984), Ch. 7; Andenæs (1980), p. 229; Mason (1952), p. 187, note 36. The most thorough discussion is in Huyse and Dhondt (1993), p. 231, who consider and reject the hypothesis that the trend is an artifact of the most serious crimes having been tried first. See also Elster (1998).
66. Huyse and Dhondt (1993), p. 125.
67. Huyse and Dhondt (1993), p. 115. I assume that the more lenient attitude that emerged after a few years was based on a genuine desire for justice to be done, whereas what dominated in the initial phases was a desire for revenge disguised (to oneself or to others) as a desire for justice (Elster 1999a, Ch. V.2).
68. Cowen (1991), p. 363, citing Merton (1946), pp. 68–69.

reason or rationality on the other, they suggest that it is less stark than I, for one, have thought in the past.

These are cases of a passionate agent precommitting himself against dispassionateness. Earlier I discussed instances of how passion can induce agents to precommit themselves against passion. Both can be contrasted with the standard case of a dispassionate agent precommitting himself against passion. The case of a rational agent precommitting himself against rationality may also arise. A rational agent may take steps to make himself less rational on some future occasion when being irrational will enable him to get the upper hand in negotiating with other (rational) agents.[69] This idea must not be confused with the suggestion, considered in I.5, that an irrational *disposition* may be useful in dealing with others. Getting drunk on a given occasion if I believe it will help me get my way is not like acquiring a durable disposition toward irascibility or vindictiveness.

I.3. TIME-INCONSISTENCY AND DISCOUNTING

Time-inconsistency, or dynamic inconsistency, "occurs when the best policy currently planned for some future period is no longer the best when that period arrives."[70] To this definition, we may add that the preference reversal involved in time-inconsistency is not caused by exogenous and unforeseen changes in the environment, nor by a subjective change in the agent over and above the reversal itself. The reversal is caused by the mere passage of time. Once we learn that we are subject to this mechanism, we may take steps to deal with it, to prevent the reversal from occurring or from having adverse consequences for behavior.

This general phenomenon can be subdivided into *time-inconsistency caused by hyperbolic discounting*, discussed here, and *time-inconsistency caused by strategic interaction*, discussed in the next section. Apart from a certain formal similarity, the two have little in common. Hyperbolic discounting does not require interaction: it might apply to Robinson on his island before the arrival of Friday. Conversely, strategically induced inconsistency does not require discounting. As we shall see in I.4 and I.5, the two phenomena can interact, but either can exist without the other.

69. Schelling (1960). 70. Cukierman (1992), p. 15.

Varieties of Discounting

When individuals plan their behavior over time, they typically discount future welfare to a smaller present value. When faced with options whose welfare effects will be felt at various times in the future, they choose the one for which the sum of the present value of these effects is the largest. While there is widespread agreement on this general description,[71] there are two main views about the exact shape of the discounting function. Neoclassical economists usually assume that discounting is *exponential*, in the sense that the welfare t units of time into the future is discounted to present value by a factor of r^t, where $r(<1)$ is the one-period discount factor. Following the pioneering work of R. H. Strotz and George Ainslie, many psychologists and behavioral economists argue that discounting is *hyperbolic*, so that welfare t units into the future is discounted to present value by a factor of $1/(1+kt)$, with $k > 0$.[72] In the following, I shall assume the latter view, which seems to have a great deal of direct and indirect support.[73] Perhaps the central intuition behind this view is that individuals have a strong preference for the present compared to all future dates, but are much less concerned with the relative importance of future dates. If they receive a big sum of money today, for instance, they may decide to spend half of it immediately and allocate the rest evenly over their lifetime.

Time discounting may be undesirable on two very different grounds. First, an individual who discounts the future very heavily, with little ability to defer gratification, is unlikely to have a very good life. That is why we teach our children to be prudent and think about the future (see also I.5). These undesirable effects of discounting

71. Among other approaches to discounting, Loewenstein and Prelec (1992) assume that "intertemporal choice is defined with respect to *deviations* from an anticipated status quo (or 'reference') consumption plan." Skog (1997) assumes that discount rates fluctuate stochastically, and shows how this phenomenon, if anticipated, may give rise to precommitment behavior.
72. See notably Strotz (1955–56), Ainslie (1992), as well as the essays in Loewenstein and Elster, eds. (1992). For the present purposes I need not distinguish between hyperbolic discounting and the other non-exponential discounting functions discussed in Phelps and Pollak (1968), Akerlof (1991), Laibson (1994), and O'Donoghue and Rabin (1999a,b). These writers decompose overall discounting into a discounting of all future periods relative to the current period and a standard exponential discounting of all future periods relative to each other.
73. For direct support, see Ainslie (1992) and Laibson (1996a). The indirect support is that many of the precommitment strategies discussed here and in later chapters cannot be explained on the assumption of exponential discounting.

are independent of the exact shape of the discounting function. They may arise for non-exponential as well as for exponential discounting. Second, an individual who is subject to hyperbolic discounting is liable to time-inconsistency. The bulk of the present section is devoted to precommitment behavior motivated by the second problem.

Becker on Endogenous Time Preferences

First, however, let me discuss a recent argument by Gary Becker and Casey Mulligan to the effect that individuals may precommit themselves in response to the problem of high discounting per se. They offer

> a model of patience formation that combines the classical economists' insights with a particular view of what it means to be rational, a conception of rationality that is consistent with many kinds of human frailties, including defective recognition of future utilities. Rational persons may spend resources in the attempt to overcome their frailties. This simple idea provides the point of departure for our approach to endogenizing time preferences. Even rational people may 'excessively' discount future utilities, but we assume that they may partially or fully offset this by spending effort and goods to reduce the degree of overdiscounting.[74]

Along similar lines, Becker wrote in the introduction to his most recent book of essays that

> People train themselves to reduce and sometimes more than fully overcome any tendency towards undervaluation. The analysis in this book allows people to maximize the discounted value of present and future utilities partly by spending time and other resources to produce 'imagination capital' that helps them better appreciate future utilities . . .
>
> They may choose greater education in part because it tends to improve the appreciation of the future, and thereby reduces the discount of the future. Parents teach their children to be more aware of the future consequences of their choices. . . . Addictions to drugs and alcohol reduce utility partly through decreasing the capacity to anticipate future consequences. Religion often increases the weight attached to future utilities, especially when it promises an attractive afterlife.[75]

74. Becker and Mulligan (1997), p. 730. 75. Becker (1996), p. 11.

In their article, Becker and Mulligan also enumerate various investments that can shape "imagination capital." Besides the purely mental (but not costless) processes of image formation and scenario simulation, these include the purchase of newspapers and other goods that can distract one's attention away from current pleasures and toward future ones; spending time with one's aging parents in order to better appreciate the need for providing for one's own old age; the purchase of disciplinary devices such as a piggy bank or membership in a Christmas club, which help a person sacrifice current consumption; tearing up one's credit cards; investment in schooling, which focuses students' attention on the future; and spending parental resources on teaching one's children to better plan for the future.[76]

Some of these strategies (joining Christmas clubs and tearing up one's credit cards) are more plausibly seen as responses to the problem of time-inconsistency (see the next subsection). To assess the other strategies, let us distinguish among three causes of preference formation.

(1) The preferences of an agent A can be traced back to a rational choice by some other agent B for the purpose of shaping or changing A's preferences.
(2) The preferences of A can be traced back to a rational choice by A for some purpose other than preference-acquisition.
(3) The preferences of A can be traced back to a rational choice by A *for the very purpose of acquiring those preferences*.

Whereas Becker and Mulligan claim to be discussing case (3), most of their examples fall in (1) or (2). Among the examples that do illustrate case (3), the idea that people spend time with their aging parents in order to better appreciate the need for providing for their own old age is in my view nothing short of ludicrous. Nor, to my knowledge, is there evidence that people choose education or religion for the purpose of becoming more oriented to the future. If there is a connection, it is more likely the other way around: people who care more about the future are more likely to choose education or religion.[77]

76. Becker and Mulligan (1997), pp. 739–40.
77. Tocqueville (1969), p. 529. Although he also asserts the opposite causal chain (ibid., p. 547), he does not suggest that anyone would *choose* religion because of its spillover effects on discounting in secular matters. In fact, I believe that most religious doctrines would condemn the idea of choosing religion for this reason.

In fact, I believe the idea that people might engage in such behaviors for this purpose is conceptually incoherent. We cannot expect people to take steps to reduce their rate of time discounting, because *to want to be motivated by a long-term concern ipso facto is to be motivated by that long-term concern*, just as to expect that one will expect something to happen *is* to expect that it will happen or to want to become immoral *is* to be immoral.[78] If people do not have that motivation in the first place, they cannot be motivated to acquire it.

Let me expand on this argument, by comparing a cultivated taste for classical music with the putative cultivation of future-oriented time preferences. Considering the first case, I may believe today that I have the choice between two streams of experiences. If I abstain from cultivating a taste for classical music, the stream will be A, A, A .... If I undergo the initially aversive experience of listening to classical music, the stream will be B, C, C .... with $B < A < C$. A rational agent will make the investment if and only if the discounted value of the second stream is larger than that of the first, which may or may not be the case.

Considering the second case, suppose for vividness that I am offered a "discounting pill" at a price of $100, which will reduce my rate of time discounting. (The pill is a proxy for education, religion, psychotherapy, or any other activity that will affect my rate of

78. The first statement is obvious. For the latter, see Elster (1999a), Ch. IV.3. I believe the argument in Mulligan (1997), that parents can invest in becoming more altruistic toward their children, also presupposes the attitude that the investment is supposed to produce: to want to be altruistic *is* to be altruistic. It is instructive to consider two of Mulligan's verbal justifications for his model. First, he notes (p. 73), "Parents are aware of the effects of their actions on their 'preferences' and take those effects into account when determining what actions to take." But this statement, when true, implies the very opposite conclusion of what Mulligan wants it to show. If a selfish person predicts that certain activities will make him more altruistic, he will tend to stay away from them (just as an altruistic person will tend to stay away from activities that will predictably make him more selfish). Second, he argues (p. 77), "People may naturally tend to be selfish, but parents may also spend time and effort in self-reflection in order to overcome such a natural bias." This statement suggests that people may invest resources in overcoming (Davidsonian) weakness of will (I.2). Suppose that I value both my own consumption and that of my children. Faced with the choice between a consumption allocation that gives more to me and one that gives more to my children, I may decide that all things considered I ought to realize the latter, yet I always choose the former. To make myself do what I really want to do, I might then invest resources in strengthening the value I place on the consumption of my children, so that it is less easily undermined by my selfish desires. This is a coherent account, but it does not correspond to Mulligan's formal model. In fact, as far as I know no economic model of (Davidsonian) weakness of will has been proposed.

discounting.) If I take the pill I shall be motivated to save $50 out of my net weekly income of $500. My consumption stream will be 350, 450, 450 .... up to retirement and then continue 450, 450 ... because of the return on my savings. Let us call this stream **I**. If I do not take the pill, I shall spend all my current income on consumption goods. My consumption stream will be 500, 500, 500 .... up to retirement and then continue 100, 100 ... because I would now have to live on welfare thanks to my earlier profligacy. Let us call this stream **II**. Let us finally define stream **III** by a consumption of 450 in each period – a life of prudent saving and comfortable retirement without the initial expense of the pill. By assumption, I preferred **II** over **III**, as I was unwilling to defer gratification without the pill. To take the pill, I would have to prefer **I** over **II** and, by transitivity, to prefer **I** over **III**. As long as the rate of time discounting is positive, this cannot happen, since **I** and **III** differ only in that the first has a lower initial consumption.

Hyperbolic Discounting and Precommitment

I now turn to precommitment against inconsistency. Individuals who discount the future hyperbolically will tend to deviate from their plans unless they take precautionary measures. The person who decided to spread half of his allocation evenly over the rest of his lifetime may decide, at the beginning of the second year, to spend one-half of the half in that year and then allocate the remaining 25% over the rest of his lifetime. If he discounted the future exponentially, no such change of plans would occur.

The problem is set out in a more general form in Figure I.2. At time 1, the agent has a choice between a small reward that will be made available at time 2 and a larger reward that will be made available at time 3. The curves show how these future rewards are discounted (hyperbolically) to present value at earlier times. Before t^*, when the present value curve (II) of the larger reward is above that of the smaller reward (I), the agent intends to choose the larger reward. After t^*, however, the present value of the smaller reward dominates. At time 2, he therefore chooses the smaller reward. With exponential discounting, such preference reversal can never occur: if an option is preferred at one time it is preferred at all other times.

To exhibit pure cases of hyperbolic discounting, we must look for instances that do not involve passions or cravings. Ainslie argues

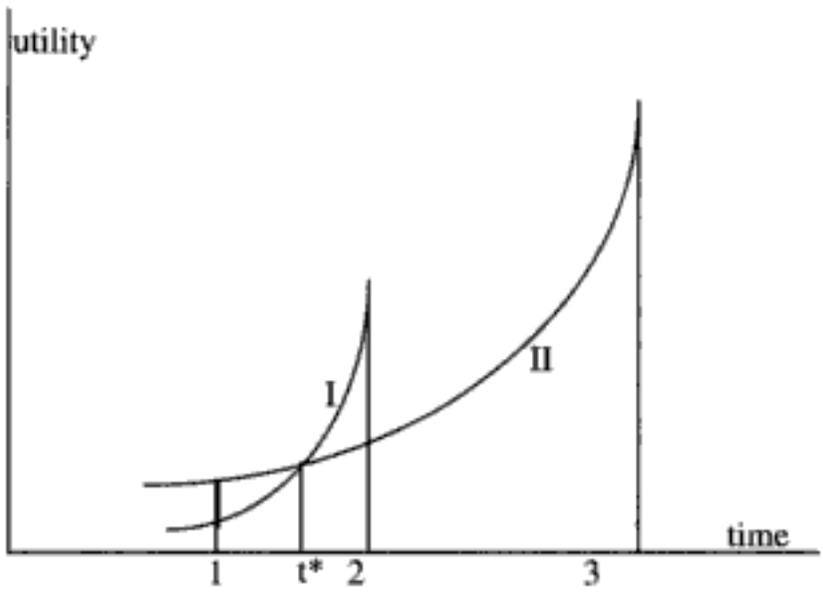

Fig. I.2

that hyperbolic discounting explains why people make resolutions to stop drinking, smoking, or gambling and then fail to carry them out, and in many cases he may well be right. Yet these behaviors are not decisive evidence for hyperbolic discounting, for the deviation from earlier plans could also be due to wishful thinking or to sudden craving triggered by cue-dependence (see I.2 and I.8). More clear-cut cases include procrastination, failure to save for Christmas or for one's old age, failure to go to bed early at night or to get up early in the morning, and failure to do physical exercise. In many of these situations, failure to keep one's resolution is plausibly due to the sheer passage of time. There may not be any passions, urges, or cravings of any kind involved (although in some cases guilt and guilt-induced denial could also contribute to procrastination). And by the argument sketched in I.2, we may reasonably assume that to a first approximation the pre-reversal preference embodies the "real" interest of the person.

The individual who finds that he constantly fails to carry out his plans, and suspects that something like hyperbolic discounting is the culprit, could respond in a number of ways. Here I shall consider only precommitment strategies. In I.8 I consider some of the strategies the person might adopt if self-binding devices are unavailable or have undesirable side effects.

Precommitment strategies include notably the first four devices of Table I.1.

(i) By making the choice of the early, smaller reward physically impossible, one obviously ensures that the larger, delayed reward is chosen.

(ii) By imposing a sufficiently large extra cost on the choice of the early reward, curve I is shifted downward so as to lie entirely below II. The larger reward is always preferred and will also be chosen at time 2.

(iii) By setting up a sufficiently large reward for the choice of the later reward, curve II is shifted upward so as to lie entirely above I. When feasible, this strategy may seem logically equivalent to the previous one, but psychologically the two may have different effects.[79] I postpone discussion of this strategy to I.7.

(iv) By the imposition of a mandatory and sufficiently long delay between the time of choice and the time at which the reward is made available to the agent, curve I is shifted to the right so that the present value of the smaller reward at time 2 is smaller than that of the larger reward. Again, this will ensure the choice of the larger reward.

To illustrate devices (i), (ii), and (iv), consider the problem of inadequate savings.[80] The formerly popular institution of Christmas clubs "offered the unusual combination of inconvenience (deposits were made in person every week), illiquidity (funds could not be withdrawn until late November), and low interest (in some cases, zero interest). Of course, illiquidity was the Christmas club's raison d'être since customers wanted to assure themselves of funds to pay for Christmas presents."[81] These institutions allowed members to pay a premium to protect themselves against their lack of willpower. They

79. There is one sense in which self-imposed rewards and self-imposed costs are conceptually different. One could ask a bank for a loan that is to be paid out only in case one chooses the later reward, and to be repaid (with interest) out of that reward. As Ole-Jørgen Skog has shown (personal communication), this strategy will, for certain parameter constellations, allow for a bootstrapping solution to the time-inconsistency problem. It is doubtful, though, that banks would give such loans without collateral. And if the person has collateral, he or she could also use it to fund self-imposed costs.

80. The second and third devices can also be illustrated by the mirror problem of excessive cash spending. A person might decide not to keep an ATM card and not have a PIN for his credit card, so that he can withdraw cash only during regular banking hours (a delay strategy). Alternatively, he could discard the ATM card but keep the PIN for his credit card, if the bank charges a fee for cash withdrawal on the card (the strategy of self-imposed costs). (I am grateful to John Alcorn for suggesting this example.)

81. Thaler (1992), p. 98.

illustrate the first device, that of making preference reversal physically impossible. Insurance companies may also embody that device, by refusing to reconvert annuities into cash or to have them used as collateral.

But the incontinent saver can also get the *bank* to pay the premium. Many banks allow for higher interest on accounts that can be drawn upon only once or twice a year, more frequent withdrawals being penalized. People who were afraid that they might not stick to their New Year's resolutions to save could use this device to protect themselves, and gain the higher interest as a bonus. This case illustrates the second precommitment device, as does the following proposal. Many people are afraid of going to the dentist. They make appointments, only to break them a day or two before they are due. To overcome this tendency, they might authorize their dentist to bill them thrice the normal fee for a canceled appointment. (See I.5 for the idea that an emotion of shame might act as a similar side bet.)

According to David Laibson,

> all illiquid assets provide a form of precommitment .... A pension or retirement plan is the clearest example of such an asset. Many of these plans benefit from favorable tax treatment, and most of them effectively bar consumers from using their savings before retirement. For IRA, Keogh plans, and 401 (K) plans, consumers can access their assets, but they must pay an early withdrawal penalty. Moreover, borrowing against some of these assets is legally treated as an early withdrawal, and hence also subject to penalty. A less transparent instrument for precommitment is an investment in an illiquid asset which generates a steady stream of benefits, but which is hard to sell due to substantial transaction costs, informational problems, and/or incomplete markets. Examples include purchasing a home, buying consumer durables, and building up equity in a personal business.[82]

In Laibson's formal model, illiquid saving instruments are defined by the fact that "a sale of this asset has to be initiated one period before the actual proceeds are received." This illustrates the fourth device. In a footnote, Laibson asserts that the results derived in that model also obtain if there is instead a premium on selling the asset. As explained earlier, precommitment can be achieved either by shifting the present-value curve downward or by shifting it to the right. This equivalence does not always obtain, however. We shall see in Chapter II that

82. Laibson (1997), pp. 444–45.

constitutions sometimes constrain behavior by requiring delays, but virtually never by imposing additional costs.

Laibson also notes that to be effective, these saving schemes require a combination of paternalism and self-paternalism. Even if individuals want to bind themselves, they may be unable to do so unless the government helps them out. "This market failure arises because the schemes are vulnerable to third party arbitrage: any consumer who is engaged in one of these schemes will have an incentive to use a third party to unwind the scheme or arbitrage against the scheme unless the government explicitly forbids such third party contracting."[83] This is not a case of straightforward paternalism, forcing or inducing the citizens to save for their retirement because the government believes that their time horizon is so short that they will not do so voluntarily. Rather, the government subsidizes the rate of interest, penalizes consumption, and forbids arbitrage in order to enable the citizens to do what they really want to do.

To conclude let me mention an important "ignorance-is-bliss" result by Juan Carillo and Thomas Mariotti. They show that in some cases problems created by hyperbolic discounting may be overcome by "strategic ignorance."[84] Specifically, they argue that "there is a tradeoff in the decision to acquire information. On the one hand, under full information, the agent can take the optimal action at the present date. On the other hand, due to perfect recall, this information is shared with all future incarnations." Because the agent can predict that the latter incarnations will be excessively present-oriented from his current point of view, he may not want them to be too well informed. Suppose, for instance, that the agent is afraid of HIV transmission through unprotected sex. Although he does not know how likely it is that the virus will be transmitted by a single act of intercourse, he has an initial subjective probability distribution over this outcome. Given those priors, his optimal action is to abstain from unprotected sex. Yet he also has the option (assumed to be costless) of finding out more, by asking a doctor or consulting statistical tables. He might then rationally decide to abstain from gathering that information, if it might have the effect of lowering the estimate of

83. Laibson (1996a).

84. Carillo and Mariotti (1997). Their model stipulates the two-factor model of discounting offered by Phelps and Pollak (1968) rather than hyperbolic discounting in the strict sense.

transmission and induce a future incarnation to engage in unprotected sex because that will be optimal *from the point of view of that incarnation*.

I.4. TIME-INCONSISTENCY AND STRATEGIC BEHAVIOR

The issue of strategically based time-inconsistency is closely related to the problem of making credible *threats* and *promises*. It is useful, I believe, to discuss these together with *warnings* and *encouragements*, the latter being an admittedly imperfect term for the phenomenon that is related to promises as warnings are to threats. All these phenomena involve a statement, a choice, an event, and an outcome, occurring at four successive moments in that order.

At time 1, the speaker A makes a statement to a listener B to the effect that if B makes a certain choice at time 2, a certain event will happen at time 3. In threats and promises, the event is under the control of A. In warnings and encouragements, it is the result of a causal chain outside A's control. Once B has made his choice and the event has or has not occurred, an outcome is produced at time 4. The two actors rank the possible outcomes according to their respective preference orderings. Although A's preferences could be based on altruistic or malicious concern for B's welfare rather than on A's self-interest, I limit myself to the latter case. Because the outcome occurs after utterance, choice, and event, the time discounting of the two actors can also be relevant. I first discuss strategic time-inconsistency without discounting, then with exponential discounting, and then with hyperbolic discounting. The focus throughout is on the precommitment devices available to A.

Cutting the Lines of Communication

First, however, I want to mention and illustrate another aspect of these interactions. In general, B, no less than A, must be supposed to be a strategic agent who is capable of precommitting himself.[85]

85. By contrast, I argued in I.2 that in intrapersonal ("multiple-self") precommitment cases, there is usually only one "self" that is capable of precommitting itself against actions that might be undertaken by the others.