Alchemies of the Mind
Rationality and the Emotions

JON ELSTER

CAMBRIDGE UNIVERSITY PRESS
1999
of the envious person is in fact to destroy the object of his envy or, if that is impossible, to destroy its owner. The action tendency of shame
is to hide or disappear; that of guilt, to make amends or to confess;
that of love, to approach and touch the other person. According to
Aristotle (II. 2), anger and hatred differ in that the action tendency
of the former is to make its object suffer, that of the latter to make it
suffer. Fear, as noted in I. 2, has two action tendencies: fight or
flight.

Although roughly accurate, this standard account requires some
modification. As noted in the discussion of envy (III. 3), we may dis-
tinguish between the action tendency and a wish or desire for a certain
state of affairs to obtain. In wrath, what matters is not simply that
the other suffer. To restore my self-esteem and sense of agency it is
necessary that I make him suffer. In jealousy, too, the humiliation of
the other has to come about by my agency. In hatred, in contrast,
what matters is that the hated person disappear from the face of the
earth, not that I make him disappear. In envy, similarly, what matters
is that the other not have the possession I envy him, not that he loses
it by my agency. Although a strong desire may seem almost indissoc-
iable from the tendency to realize it by action, the case of envy, in
which the most preferred state is that the other lose what he has but
not by my agency, shows that they are conceptually distinct.

Action tendencies usually go together with inhibitory tendencies.
Some inhibitions arise simultaneously with the action tendency they
inhibit (mechanisms of type B), whereas others are triggered by it
(type B'). An example of the former is when the sympathetic and
parasympathetic nervous systems are activated simultaneously, the
one producing the action tendency and the other a tendency to re-
strain or energy conservation. The latter category includes not
only self-control and control via social norms but also purely physi-
ological mechanisms.

In this connection, we may cite an important observation by Frijda:

It may . . . be that emotional impulse as evoked by relevant stimuli potentially
always is of maximal intensity, regardless of realistic or moralistic consider-
ations: this impulse is toned down by inhibition, as a permanent stabiliz-
counterforce governed by reality and morality. Sham rage, Mark and
Ervin's observations on unrestrained anger, enhancement of startle response
117. Ibid., pp. 155-61.

As an illustration, let us suppose that the desire for revenge is a
spontaneous impulse of this kind. In modern societies, this desire is
usually inhibited by social norms that become internalized at an early
age. In feuding societies, the impulse is magnified and focused by
strong social norms, as we have seen. I suggest, tentatively, that the
ancient Greeks displayed revenge behavior neither inhibited nor magni-
fied by social norms; that is, the presocial urge for revenge in something
like its pure form. To the extent that this conjecture is correct, at-
tempts to interpret ancient Greek society by relying on studies of
modern Mediterranean cultures may be somewhat off the mark.

Not all emotions have action tendencies. Hume asserts, plausibly,
that "pride and humility are pure emotions in the soul, unattended
with any desire, and not immediately exciting us to action." Also,
relief, regret, disappointment, sadness, grief, and most of the aesthetic
emotions do not seem to suggest any specific actions. Although small
children sometimes want to get on the stage to save the actor from an
impending danger, most works of art do not induce any action ten-
dency in readers, listeners, or viewers. In fact, it has been suggested
that the aesthetic emotions have an exceptional purity that is due to
their dissociation from action. 132

IV. RATIONALITY AND THE EMOTIONS

The relations between rationality and the emotions form an intricate
web. One may distinguish among three partly overlapping sets of
issues. First, we may try to determine the impact of emotions on the

118. Ibid., p. 408; references deleted.
119. This is not to deny that the Greeks disapproved of those who failed to avenge
themselves. It is clear from Demosthenes' speech against Meleager that he felt he
had to explain why he had not retaliated immediately when the latter slapped
him in public. Yet no one thinks we need to impute anything like a code of
honor to explain the disapproval of failures to take revenge. The disapproval
might be as spontaneous as the urge to revenge itself.
120. Cohen (1993) is an example.
121. Hume (1751), p. 357.
rationality of decision making and belief formation. Second, we may ask whether the emotions themselves can be assessed as more or less rational, independently of their impact on choice and belief formation. And third, we may ask whether emotions can be the object of rational choice, that is, whether people can and do engage in rational deliberation about which emotions to induce in themselves or in other people.

Concerning the first set of issues, the traditional view is that emotions interfere with rational choice. They are, as it were, sand in the machinery of action. More recently, several writers have argued for the revisionist view that emotions, far from interfering with rational decision making, may actually promote it. Thus one may argue that emotions help us make decisions by acting as tie-breakers in cases of indeterminacy and, more generally, improve the quality of decision making by enabling us to focus on salient features of the situation. Another revisionist idea is that rational belief formation is incompatible with emotional well-being (the "sadder but wiser" hypothesis). Others have taken positions that go against the traditional as well as the revisionist views, by claiming that emotions merely affect the parameters of choice without affecting the rationality of choice itself. On this view, emotions enter into decisions as costs and benefits associated with various options but not as psychic forces capable of shaping or distorting the mechanisms of choice.

Concerning the second set of issues, there are a number of arguments to the effect that emotions themselves can be assessed with a view to their rationality. One may argue that certain emotions are actions and hence can be assessed by the usual criteria of rational choice. Also, one may argue that some emotions are rational if they are appropriate or adequate to the cognitive appraisals that trigger them. Moreover, one may argue that emotions are rational if those appraisals themselves are rational. Finally, one may argue that rational emotional appraisals are those which make one happy. In considering these ideas I shall rely on a notion of rationality that is subjective through and through. I shall not discuss, in other words, whether emotions are objectively useful or adaptive, and, if they are, whether their usefulness explains why they exist.

Concerning the third set of issues, different approaches can be distinguished along two dimensions. First, we may ask whether the object of choice is an occurrence emotion or an emotional disposition: second, whether the choice is made by the person in whom the emotion occurs or by somebody else. The ensuing combinations are indicated by Table IV.1. Whereas examples in the lower left-hand cell are discussed in other chapters, the remaining cases are considered here.

### Table IV.1

<table>
<thead>
<tr>
<th>Induced by self</th>
<th>Schafir</th>
<th>Emotional dispositions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Induced by others</td>
<td>Oratory (V)</td>
<td>Character planning</td>
</tr>
<tr>
<td>Induced by others</td>
<td>&quot;Shaming&quot; (III.2)</td>
<td>Backer</td>
</tr>
</tbody>
</table>

Concerning the third set of issues, different approaches can be distinguished along two dimensions. First, we may ask whether the object of choice is an occurrence emotion or an emotional disposition: second, whether the choice is made by the person in whom the emotion occurs or by somebody else. The ensuing combinations are indicated by Table IV.1. Whereas examples in the lower left-hand cell are discussed in other chapters, the remaining cases are considered here.

### The Traditional View

Subjectively, rational choice involves three optimizing operations. The action that is chosen must be optimal, given the desires and beliefs of the agent. The beliefs must be optimal, given the information available to the agent. The amount of resources allocated to the acquisition of information must be optimal, given the desires of the agent and his beliefs about the expected costs and benefits of information. Given this definition, rationality can fail in two ways: by indeterminacy or by irrationality. Later, I discuss a possible positive role of the emotions in cases of indeterminacy. Here, I discuss how, on the traditional view, emotions might be a cause of irrationality in decision making. They can do so, obviously, by subverting the rationality of action, of belief formation, and of information acquisition.

In his discussion of emotionally induced irrationality, Frijda considers a number of phenomena that may be classified as follows: (i) Emotions affect "probability and credibility estimates" concerning

---

133. Elster (in press b), Ch. 3.3.
134. A good discussion of this issue is in Frijda (1986, pp. 475-9). Elsewhere (Elster in press a, section 1.5) I consider a view not discussed by Frijda, namely that emotions are useful self-binding devices in strategic interaction.

135. For fuller exposition see Elster (1989b, Chapter 1); Elster (in press b, Ch. 3.3).
events outside one's control. (iii) They 'cause some measure of belief in the efficacy of actions one would not believe in under other condi-
tions.' (iv) They induce fantasy behavior, as when a widow for many years after the death of her husband sets the table for two each day. (iv) They induce various forms of 'painful fantasies that, however painful, are yet sought by the subject himself.' Othello's jealousy is cited as an example. (v) They cause irrational behavior, such as 'anger at some deed that cannot be undone by the angry aggression, nor its recurrence prevented,' or 'desires and pursuits of someone whom one knows does not want to be pursued.' He notes that in trying to account for these phenomena, 'two kinds of explanation can be advanced that are advanced for unrealistic behavior generally, the one motivational, the other in terms of general strategic principles.' The
former explanation appeals to the tenacity of the mind to seek immediate gratification by illusions and magical actions, at the expense of long-term.
the latter rests on the premise that reactions that are subjectively irrational when considered as current emotions may be manifestations of a more general disposition that, on average and in the long term, is objectively adaptive. He concludes that 'both kinds of mechanism do apply.' The motivational explanation is the one that is relevant here.

Of Frieda's categories, the first three (in my classification) involve irrational belief formation, largely of the wishful-thinking variety. The importance of this phenomenon is obvious. It does not, however, exhaust the category of emotionally induced belief irrationality. Later in this section, and more extensively in V.2, I discuss how meta-emotions can induce belief changes that in turn generate new first-order emotions. The phenomena in category (iv) illustrate the mechanism of counterwishful thinking, briefly discussed in I.6 and II.13 and further discussed below. Those in category (v) illustrate irrationality at the level of action. Although Frida does not mention that emotions may subvert the process of information acquisition, the fact that passion can make us jump to conclusions is too obvious to need stressing. What should be stressed, however, is that this does not occur only in the face of threatening events, when the expected opportunity costs of gathering information are so high that delays would be irrational. Any strong emotion creates a tendency to act immediately, even if nothing would be lost and something might be gained by pausing to find out more about the situation. The urge of a guilty person to seek relief by making atonement may be so strong that he does not take the time to find out what form of atonement might be in the victim's best interest. A person in love may be too impatient to find out whether the love object might have some unknown and perhaps less-than-wonderful qualities.

In the philosophical literature the subversive influences of passion are usually described under the headings of self-deception or wishful thinking and weakness of the will. The latter phenomenon can take several forms: Acting without regard for the consequences of one's behavior, and acting against one's own better judgment. Frieda's irrational anger might fit the first case, the irrational pursuit the second. Macbeth's behavior is also a good example of the second — 'vaulting ambition' (I.7.27) making him do what he believes he should not. I believe the lack of regard for consequences — including the lack of concern for more emotion — is the most important mechanism by which emotion can subvert rationality. It can work in two ways, directly or indirectly. A person may simply be so caught up in visceral anger, fear, envy, shame, infatuation, or hybrid that he fails to think more than one step ahead. (I return to this idea below.) Alternatively, he may be subject to social norms according to which a visible concern for consequences is discreditable (III.5). In this case, fear of shame rather than no current shame itself — will motivate him to act as if he had no regard for the consequences.

GUT FEELINGS

Ronald de Sousa and Antonio Damasio argue that emotions, rather than being sand in the machinery of action, can actually promote rational behavior in situations of indeterminacy. Their argument is not only that a person without emotions would make irrational decisions. They also claim — and I believe this is their key idea — that in many situations a person would make no decision at all or delay it for very long, and that such an action or procrastination would be irrational. ('Deciding well also means deciding expeditiously, especially when time is of the essence.') As stated, the claim does not imply that emotions have a causal role in reaching a decision or in reaching it sooner rather than later. Emotional flatness and indecisiveness could simply be correlated as joint effects of a common cause that is necessary for the former and sufficient for

the latter or vice versa. It is clear, however, that both writers want to argue for the stronger thesis that emotions are causally involved in rational decision making.

Indeterminacy, in general, can take the form either of indifference or incommensurability. Writing R for weak preference ("a is at least as good as b"), indifference obtains when both a R b and b R a and incommensurability when neither a R b nor b R a. Thus indeterminacy of choice arises when two options are equally and maximally good, or when each of two options is strongly preferred to all other options but neither is weakly preferred to the other. De Sousa argues that emotions can serve as a tie-breaker in cases of indifference:

For suppose you are considering whether to take a fair bet. By definition, from the Bayesian point of view, a fair bet is equivalent to no bet at all. Its expected desirability is zero. Yet there is clearly a significant option between the choice to minimize the greatest possible losses (maximin) and the choice to maximize the greatest possible gains (maximax). And the choice between maximin and maximax strategies is obviously associated with such emotional and character traits as boldness or timidity.

Indifference is a knife-edge property, rarely observed in nontrivial real-life situations. Incommensurability is, I believe, a vastly more important source of indeterminacy. Consider for instance a customer who cannot make up his mind whether to buy car brand A or brand B. If he were indifferent, a one dollar discount on A should make him decisively prefer A. If he still cannot make up his mind, as will typically be the case, the two brands must be incommensurate. To choose, he may flip a coin − or consult his "gut feelings."

Before I proceed to discuss how these feelings enter into the decision-making process, let me use some passages from Boswell's Life of Dr. Johnson to distinguish between two cases:

Life is not too long, and too much of it must not be spent in idle deliberation how it shall be spent: deliberation, which those who begin it by prudence,

19. De Sousa (1987), p. 194. The example is somewhat defective, because in this case not making a decision is equivalent to making the decision not to gamble. It could be improved by stipulating that the agent is indifferent between two risky options, each of which has a higher minimum than the status quo. It is also defective for a different reason. The idea of a "fair bet" is usually defined in monetary terms rather than in utility terms. If the agent is indifferent between making the fair bet and not making it, the utility function must be linear in the relevant interval. Given the way cardinal utility functions are constructed, however, the presence of timidity or boldness would typically cause incommensurability.

190. We talked about the education of children; and I asked him what he thought was best to teach them first. JOHNSON: Sir, it is no matter what you teach them first, any more than what leg you shall put into your breeches first. Sir, you may stand disputing which is best to put in first, but in the mean time your breech is bare. Sir, while you are considering which of two things you should teach your child first, another boy has learnt them both.

191. He did not approve of late marriages, observing that more was lost in point of time, than compensated for by any possible advantages. Even ill-assorted marriages were preferable to cheerless celibacy.

192. The first passage illustrates the case (I) of pure incommensurability, in which one could not conceivably gather enough information to make an informed choice. In such cases, a person who could not make up his mind unless he had "just reasons" to go one way or another would simply be paralyzed. The other passages illustrate the case (ii) in which one might gather enough information to make an informed choice, although the opportunity costs of doing so would be prohibitive high. In a more important and realistic variety (iiib) of this case, the costs of gathering information might rather than would be too high. I might meet the perfect spouse by waiting a bit longer − but then again I might not. When I know that the expected costs of gathering more information exceed the benefits, the choice is easy. I just flip a coin or consult my gut feelings. When I do not know − and do not know whether it would be worthwhile to find out we are dealing with a more radical form of indeterminacy.

193. De Sousa writes that the "role of emotion is to supply the insufficiency of reason... For a variable but always limited time, an emotion limits the range of information that the organism will take into account..."
Alchemy of the Mind: Rationality and the Emotions

account, the inferences actually drawn from a potential infinity, and the set of live options from which it will choose. 144 It is not clear which of (i), (ia), or (ii) he has in mind. A reference to work by Daniel Dennett on information retrieval may suggest that he is mainly interested in (ii), but it is hard to tell. The standard view, in any case, is that the urgency of the emotions helps us make optimal decisions in case (iia). Because of their limited rationality, people always use rules of thumb to make short-cuts to a decision. "Emotion may well enhance utilization of these heuristics, in view of the desirability of rapid action or, more generally, the restriction of range of cue utilization." 145

What we may observe here, however, is not emotion doing what reason cannot do, but rather emotion doing what reason could do, only differently. De Sousa and others who argue along similar lines consistently present a strawman of rational-choice theory according to which a rational agent would always take account of all possible outcomes of all possible options. According to LeDoux, if you were a small animal faced with a bobcat and "had to make a deliberate decision about what to do, you would have to consider the likelihood of each possible outcome occurring or failing and could get so bogged down in decision making that you would be eaten before you made the choice." 146 What Philip Johnson-Laird and Keith Oatley call "impeccable rationality" enables the organism to decide which goals to pursue at any point in time, and to decide at each stage the best course of action in pursuit of those goals. No contingency is unanticipated, and performance is invariably optimal. 147 Below, I cite a passage from Damasio’s book. These authors all assume that rationality amounts to what I have called elsewhere an addiction to reason. 148 Some people do indeed have a craving to make all decisions on the basis of "just" or sufficient reasons. That, however, makes them irrational rather than rational. A rational person would know that under certain conditions it is better
to follow a simple mechanical decision rule than to use more elaborate procedures with higher opportunity costs. 149 If we disregard cases in which the twelve-millisecond delay of the cortex could be crucial, the organism might cope perfectly well by adopting and following mechanical decision rules, such as "when you hear a sound you cannot identify, stand still" or "when food tastes bitter, spit it out." In reality, of course, that’s not how we cope with novelty or bitter tasting food not because the program is unflexible but because natural selection has wired us differently. It is somewhat misleading, therefore, to assert that emotions are a "supplemental" principle that "fills the gap" between reflex-like behavior and fully rational action. 150

We can take this argument one step further. If we do not and cannot respond to emergencies by following a mechanical decision rule, it may be because our cognitive faculties are temporarily clouded by the emotional arousal caused by the emergency. The emotion serves as a functional equivalent for the rational faculties it suspends, by inducing the very behavior that isrationally required and that reason, if left undisturbed, could have come up with by itself. The emotions do solve problems, but problems that are to some extent of their own making. The capacity for the emotions to supplement and enhance rationality would not exist if they did not also undermine it.

In his (independently developed) argument for the rationality of the emotions, Damasio draws on findings from patients with specific brain lesions. Some patients who have suffered damage in their frontal lobes become emotionally flat and lose their ability to make decisions, while retaining other cognitive powers. Patients who have suffered damage to their somatosensory cortices display similar symptoms, although in their case there is also severe cognitive malfunctioning. From his analysis of these patients, and drawing on general neurophysiological data, Damasio concludes that their defective decision-making capacity is due to their lack of emotion. Not knowing much about the brain I cannot do full justice to his

---

144. De Sousa (1987), p. 195. Along similar lines, Johnson-Laird and Oatley (1992) argue that because the ideal of "impeccable rationality" assumes that "there are no surprises, no misunderstandings, no irreducible contacts," it cannot guide action in situations that are characterized by these features. Instead, "emotions enable social species to co-ordinate their behaviour, to respond to emergencies, to prioritise goals, to prepare for appropriate actions, and to make progress towards those goals... even though individuals have only limited abilities to cognize.

145. Frijda (1986), p. 121; see also p. 176.


149. Thaler (1981) argues that neglect of opportunity costs and excessive focus on out-of-pocket expenses is a frequent source of cognitive irrationality. The neglect of opportunity costs that are created by the fact that decision making takes time is also an important and pervasive source of irrationality.

150. For much the same, see De Sousa (1987), p. 194; Johnson-Laird and Oatley (1992), p. 206. The point I am making is well stated by LeDoux (1996, p. 175): "In responding first with its most likely-to-succeed behavior, the brain buys time. This is not to say that the brain responds automatically for the purpose of buying time. The automatic responses come first, in the evolutionary sense, and cannot count for the purpose of serving responses that came later" (LeDoux 1996, p. 175).
was discussing with the same patient when his next visit to the laboratory should take place. I suggested two alternative dates, both in the coming month and just a few days apart from each other. The patient pulled out his appointment book and began consulting the calendar... For the better part of a half-hour, the patient enumerated reasons for and against each of the two dates. Previous engagements, proximity to other engagements, possible meteorological conditions, virtually anything that one could reasonably think about concerning a simple date. Just as calmly as he had driven over the ice, and recounted that episode, he was now walking us through a firestorm of cost-benefit analysis, an endless outlining and flawless comparison of options and possible consequences. We finally did tell him, quietly, that he should come on the second of the alternative dates. His response was equally calm and prompt. He simply said: "That's fine." 155

In the gambling experiment, the brain-damaged patients consistently did worse than others. (This is the only case Damasio discusses in which the patient's failure was due to bad decisions rather than delaying them.) The game required subjects to draw cards from one of four decks. Each time the subjects drew a card from decks A and B they received a large sum of play money, and a smaller sum when they drew from decks C and D. When taking a card from A and B, they also sometimes had to pay back a very large amount of money. As they learn about the structure of the game, normal subjects mostly take cards from C and D and usually end up ahead. Brain-damaged subjects, by contrast, stick to decks A and B even though they regularly go bankrupt halfway through the game. Damasio's explanation is that these patients are highly anable to be motivated by mental representations of future states. Although they had normal skin conductance reactions to monetary loss, they differed from normal subjects in having no anticipatory responses in the period immediately preceding their selection of a card from a bad deck. 156

Brain-damaged patients, then, tend to be emotionally flat and to have defective decision-making capacities. It remains to characterize the relation between these two features. Damasio's strong causal claim is that "Reduction in emotion may constitute an... important..." 157

153. Damasio (1996), p. 31; see also pp. 55-66 for reports on similar behavior in another patient.
154. Davidson (1985) argues that "the perception of emotion and the experience of expression of emotion" are likely to be based on different neural control systems. Hence an alternative explanation of the behavior of the four-language patient might be that he was unable to detect the emotional reactions of others in his behavior.
155. This view presupposes that the emotions of shame and embarrassment would not exist in the absence of the norms. It is confusing, therefore, to see Damasio (1994, pp. 142-43) arguing that "the penitent, an attitude of social convention and moral codes, may be direct and immediate (physical or mental harm), or remote and indirect (future loss, embarrassment)," as if the emotions could exist independently of the norms.
Although the Mind: Rationality and the Emotions

The source of irrational behavior. Elsewhere, he makes the weaker claim that "The powers of reason and the experience of emotion decline together." This is to assert correlation, but not causation. To support the strong claim he first describes a decision-making problem that the owner of a business might confront, faced with the prospect of meeting or not with a possible client who can bring valuable business but also happens to be the archenemy of your best friend, and proceeding or not with a particular deal. The brain of a normal, intelligent, and educated adult reacts to the situation by rapidly creating scenarios of possible response options and related outcomes. He then argues that (what he takes to be) the rational-choice approach to this problem would involve impossibly complex calculations. The decision would take an "inordinately long time" or might never be made at all. Because we are, as a matter of fact, able to make such decisions quite rapidly and efficiently, something else must be going on:

Consider again the scenarios that I outlined. The key components unfold in our minds instantly, sketchily, and virtually simultaneously, too fast for the details to be clearly defined. But now imagine that before you apply any kind of cost/benefit analysis to the problem, something quite important happens: When the bad outcome connected with a given response option comes into mind, however fleetingly, you experience an unpleasant gut feeling. Because the feeling is about the body, I gave the phenomenon the technical term somatic state . . . and because it marks an image, I called it a marker.

That somatic marker . . . forces attention on the negative outcome to which a given action may lead, and functions as an automated alarm signal which says: Beware of danger ahead if you choose the option which leads to this outcome. The signal may lead you to reject immediately, the negative course of action and thus make your choice among other alternatives. The automated signal protects you against future losses, without further ado, and then allows you to choose from among fewer alternatives. There is


158. Ibid., p. 54. We may note, for future reference, that this claim about the covariation of emotion and reason differs from another, perhaps more dubious claim that he also makes. Referring to the contrast between sadness and happiness or, in extreme cases, between depression and mania, he writes that "because both the signal of the body state (positive or negative) and the style and efficiency of cognitive processes are triggered from the same system, they tend to be correlated" (ibid., p. 127). Depressive states go together with slow and inefficient mental functioning, and elated states with fast but not necessarily efficient functioning. The theory of "depressive realism" discussed below seems to go against this view.

159. Ibid., p. 170.

160. Once again, going by one's gut feelings is not the only way to cut through the maze of a complex decision problem. One can also, for instance, flip a coin. Damasio might counter that this procedure is inferior to going by gut feelings, which enable one not only to make swifter decisions but also better ones. But the coin-tossing heuristic is only the most simple of many rules of thumb that are used in complex decision-making problems. The best-known is perhaps Herbert Simon's idea of satisficing, embodied in such sayings as "never change a winning team" and "if it ain't broke, don't fix it." Also, medical diagnoses and prognoses can be very efficiently done by mechanical point systems that rely on a small number of variables. In fact, such methods almost invariably tend to perform better than intuition based on "gut feeling." In opposing gut feelings to hyperrational cost/benefit calculation Damasio is simply setting up a strawman.

This objection does not, however, affect Damasio's claim that in most complex decisions people do, as a matter of fact, consult their gut feelings. When confronted with a novel challenge for which no rule of thumb is available, some people procrastinate more or less indefinitely, while others, for better or for worse, make a snap decision based on some salient feature of the situation. Damasio claims (i) that more often than not this feature has great predictive value for making a good choice or at least avoiding a bad one, and (ii) that its salience is signaled by an occurring emotion. Concerning (i), it appeals to a regular reinforcement process, although he does not use that term. "Somatic markers are . . . emotions and feelings which have been connected, by learning, to predicted future outcomes of certain scenarios." On the basis of what I know of reinforcement theory it seems implausible that this mechanism could guide the decision whether to deal with a businessman who is the enemy of one's best friend. For reinforcement to establish behavior, it should ideally
Although paying bills on time is certainly important in consequential terms, I do not need my emotions to tell me that it should take priority.

Rather than telling more of my own counteranecdotes, let me try to address Damasio's argument and evidence more directly. He argues plausibly that (i) lack of emotionality causes (ii) defective social behavior, and that (iii) lack of ability to be motivated by the representation of absent events causes (iv) defective decision making. He makes a plausible argument, moreover, that (i) and (iv) tend to go together. 165 It remains to be shown that (i) is the cause, or a cause, of (iv). His argument seems to be that (i) is the cause, or a cause, of (iii). Future events gain motivational significance through the somatic, emotional markers attached to their representation. Although nothing in the data he presents excludes this counterintuitive hypothesis, I cannot see that they exclude the alternative view that the same brain damage that induces (i) also causes (iii) and thereby (iv). There is correlation, but for the time being we cannot tell whether there is also causation. Moreover, the causal mechanism — reinforcement — that is supposed to link emotions to decisions seems too coarse for the range of cases it is supposed to explain.

Damasio's theory may be contrasted with the view put forward by Robert Frank in Passions within Reason. 166 Both assert that criminal behavior can be explained by some emotional deficiency. Moreover, both hypothesize a link between emotions and the ability to be motivated by the future. Yet their arguments are very different. Damasio claims that the behavior of at least some "developmental sociopaths or psychopaths" fits the general pattern he is describing. While being "the very picture of the cool head we were told to keep in order to do the right thing," they also act "to everybody's disadvantage, including their own... They are, in fact, yet another example of a pathological state in which a decline in rationality is accompanied by diminution or absence of feeling." 167 They lack, presumably, the emotional markers that enable them to be motivated by the future. (Here I go beyond what Damasio says, but the imputation seems plausible.)

Frank, too, finds the key to criminal behavior in the inability to be motivated by future consequences of present behavior. His
mechanism, however, is entirely different. In most of us emotions such as guilt weigh in as current representatives of the future payoffs and thus impart greater motivational force to the latter. "If the psychological reward mechanism is constrained to emphasize rewards in the present moment, the simplest counter to a [short term] reward from cheating is to have a current feeling that tugs in the opposite direction." Damasio would assert to Frank’s proposition that “the widespread and chronic impulsiveness of criminal offenders may ... be interpreted as support for the claim that emotional competencies underlie moral behavior.” But for entirely different reasons. When the data are in Damasio may turn out to be right, but for the time being I find Frank’s view more plausible. Yet the discussion would be incomplete if we ignored that many emotions have the opposite effect — they tend to reduce our ability to take account of the future consequences of present action, not to enhance it. As Bacon said, “affection beholds principally the good which is present; reason looks beyond and beholds likewise the future and sum of all.”

**Sadder but Wiser**

Mania and depression, in the clinical sense, are extreme versions of the more ordinary emotions of exaltation and sadness. These, in turn, are stronger versions of sentiment and low spirits. As we move inward in this way, away from the extremes, we presumably reach a ground state with a neutral emotional tonality. A natural working assumption is that this would also be the cognitive ground state, the state that is, in which belief formation is unaffected by motivational bias. This assumption would fit well with two other assumptions that we routinely make, that depressed people tend to believe that things are worse than they in fact are and that those in more exuberant moods tend to believe they are better.

---

166 Frank (1988), p. 82.
170 Ibid., p. 162.
171 Brothers (1995, p. 1112) objects to Damasio that he “considered the somatic portion of a social experience simply to signal reward or punishment,” whereas “evidence from amygdala stimulation is in favor of specific, and variegated emotional states (e.g., guilt, feeling outraged) rather than simply reward or punishment.” I have no competence to judge who is right; my objections to Damasio’s theory are of a more expository nature.


---

**Rationality and the Emotion**

The second of these routine assumptions is, I believe, valid. Let me illustrate it by a personal example. For about fifteen years, from 1968 to 1984, I worked on a book on Karl Marx, and eventually felt I got to know him quite well. He was, very obviously, a very emotional person. Moreover, his emotions equally obviously distorted his thinking, both in what he wrote about the communist society and about the process of getting there. His mind seems to have been shaped by two implicit assumptions: Whatever is desirable is feasible, and whatever is desirable and feasible is inevitable. The first shows up in his refusal to consider trade-offs between values, and in his beliefs that all good things go together. The second underlies his unswerving belief in the imminent and inimitably necessary communist revolution.

These observations, although true, provide only part of the truth. The other side of the coin is that the emotions of rage, indignation, and hope provided the indispensable motivation for Marx’s political and theoretical work. They kept him going through years of exile and misery in London and sustained his enormous scholarly labors no less than his tireless organizational work. It would seem absurd and unrealistic to wish that he had the same level of motivation without the correlative cognitive distortions. The motivation for achievement may interfere, as it did in his case, with the efficacy of achievement. The emotions that provided him with a meaning and a sense of direction to life also prevented him from going steadily in that direction. Again, not all good things go together. To achieve much, one has to believe one can achieve more than one can.

The first of the routine assumptions, by contrast, is probably wrong. One of the more striking psychological findings over the past fifteen years is that the emotional ground state may not be the cognitive ground state. Rather, the only persons who are capable of taking an unbiased view of the world are the depressed. They are “sadder but wiser.”

The results are far from final. The effects may not be as strong in real-life settings. In some contexts, the polarizations described in the
his allegiance to life itself. Life has to have substance if anything is to have sense, including adherence to the impartial system, but if it has substance, then it cannot grant supreme importance to the impartial system, and that system’s hold on it will be, at the limit, insincere. 178

Similarly, we may have to accept that cognitive rationality can only be achieved at the cost of lacking anything we want to be rational about.

EMOTIONS AS PSYCHIC COSTS AND BENEFITS

To the extent that economists consider the role of emotions in behavior, they tend to view them simply as a source of preferences. The most typical approach is to assume that a given action may have emotional costs and benefits as well as material costs and benefits. 180 In choosing among various options, the agent maximizes an inclusive utility function in which all costs and benefits are considered together. On some occasions, the material benefits associated with one option may be large enough to offset any negative emotions that it might induce; on other occasions the emotional costs may be decisive; in the general case some compromise between material and emotional satisfaction will be sought. When emotions are viewed in this way, the standard apparatus of indifference curves and trade-offs automatically applies. Well-known analyses of emotions by Gary Becker, Robert Frank, and Jack Hirshleifer all rely on this approach. 81

There are two ways in which the pleasure and pain associated with the emotions may enter into the utility function. To illustrate them, let me refer to Gary Becker’s analysis of guilt (see below for a fuller discussion). First, the guilt itself is a cost. Even if I do not have any money with me, I may cross the street to avoid coming face-to-face with a beggar whose visible misery would induce the unpleasant feeling of guilt. Second, the guilt may induce behavior that is costly

178. See Fiske (1980) for a review.
179. For reasons of space, I limit myself to the relation between emotion and material self-interest. Similar issues arise for the relation between emotions and impartial motivations. For instance, a wealthy liberal might on impartial grounds prefer to send his children to a public school, but his emotional attachment to the children might induce a preference for a better-quality private education.
180. For Becker, see subsequent text. For discussions of Frank (1988) and Hirshleifer (1989), see Fiske (in press a), section 1.
in the material sense. If I do have money with me, I know that I come face-to-face with the beggar: I would give him something to alleviate my guilt. More accurately, I would give up to the point at which the marginal utility of money in alleviating my guilt equals its marginal utility for other purposes. If crossing the street is costly, this would also have to be taken into account.

In many other analyses, the encounters that trigger the emotion are taken for granted and not subject to choice. The question of choice arises only because the agent has to weigh emotional satisfaction against other satisfactions, as in the choice of how much to give to the beggar. In modeling envy, we may assume that the agent is willing to invest resources in making his rival worse off up to the point at which he derives more utility from making himself better off. In modeling altruism, we can make a similar assumption. Economic analyses of regret also assume that agents weigh satisfaction from actual outcomes and emotions generated by counterfactual beliefs. Strictly speaking, none of these analyses need to rely on valence, in the sense of subjective feelings of pleasure and pain. All that is needed is that we can draw indifference curves that reflect the trade-offs involved. We may think of emotional valence as the underlying mechanism behind these trade-offs, but it need not be directly reflected in the formal analysis. In a modeling perspective, "emotional altruism" is indistinguishable from "reason-based altruism."

To assess the validity of the cost-benefit model of the emotion, we may consider the case of guilt. Assume that a person is tempted to steal a book from the library. If he feels guilty about doing it, he may abstain. If he steals the book and then feels guilty, he may return the book to the library. On the assumption that guilt is to be modeled at a cost, both the abstention from stealing and the return of the book would be explained by a simple cost-benefit analysis.

This approach has the great advantage that it allows us to account for the undeniable existence of a trade-off between moral emotions and self-interest. The world is not made up of two exclusive and exhaustive categories, those who would always steal a book whenever there was no risk of detection and those who would never do so. Many people would go ahead and steal the book if but only if its value to them was sufficiently high and/or its value to others sufficiently small. To model such behavior, we can talk "as if" guilt and interest add up to an inclusive utility, with the marginal disutility from guilt being an increasing function of (say) the number of people on the waiting list for the book and the marginal utility from interest.

Rationality and the Emotions...

a decreasing function of (say) the time the agent expects to use the book.

Independently of its predictive adequacy, I submit that this model is conceptually flawed. If guilt were nothing but an anticipated or experienced cost, an agent whose guilt deterred him from stealing or retaining the book should be willing to buy a guilt-erasing pill if it was sufficiently cheap. I submit that no person who is capable of being deterred by guilt would buy the pill. In fact, he would feel guilty about buying it. For him, taking the pill in order to escape guilt and be able to steal the book would be as morally bad as just stealing it. He would not see any morally relevant difference between stealing the book in a two-step operation (taking the pill to steal the book and stealing it in a one-step operation). There is a strict analogy between this argument and a point that I have made elsewhere, namely that a person who discounts the future very highly would not be motivated to buy a pill that would reduce his rate of time discounting. To want to be motivated by remote consequences of present behavior is to be motivated by remote consequences of present behavior. Similarly, to want to be immoral is to be immoral. A person willing to take the guilt-erasing pill would not need it.

We need, therefore, a model that can account for the trade-off between guilt and interest and yet does not imply that a reluctant agent would buy the guilt-erasing pill. I conjecture that the model would involve some kind of nonintentional psychic causality rather than deliberate choice. Suppose that the agent is initially unwilling to steal the book, but that as its value to him increases he finally decides to do so. Suppose conversely that the agent has stolen the book, but that as its value to others increases, he finally returns it to the library. In the first case, suppose that its value to others is 10 and that he decides to steal it just when its value to him reaches 15. In the second case, suppose that its initial value to him is 15 and its initial value to others is 5. On the cost-benefit model, he would return it when its value to others reaches 10. On the catastrophe model, he might not do so.

19. Technically, the dependent variable has to be continuous rather than the dichotomous choice between stealing the book and not stealing or returning it. We can use an example in which the choice variable is continuous, such as the amount to give to the beggar in the street.
Alchemies of the Mind. Rationality and the Emotions

until its value to others reached 15. The reason for this asymmetry is found in the mechanism of dissonance reduction. An individual who is subject to several motivations that point in different directions will feel an unpleasant feeling of tension. When on balance he favors one action, he will try to reduce the tension by looking for cognitions that support it; when he favors another, he will look for cognitions that stack the balance of arguments in favor of that action. Thus the timing of the switch in behavior will be path-dependent.

Dissonance theory is more realistic than the cost-benefit model in that it views individuals as making hard choices on the basis of reasons rather than on the basis of introspections about how they feel. Although the person who has stolen the book but feels guilty about it may try to alleviate his guilt, he would do so by coming up with additional reasons that justify his behavior rather than by accepting a guilt-emitting pill. It is a fundamental feature of human beings that they have an image of themselves as acting for a reason (see also V.2). Guilt: in this perspective, acts not as a cost but is a psychic force that induces the individual to rationalize his behavior. Beyond a certain point, when the arguments on the other side become too strong and the rationalization breaks down, a switch in behavior occurs. Although we may well say that the switch occurs when the guilt becomes unbearable, we should add that the point at which it becomes unbearable is itself influenced and in fact delayed by the guilt.

What this example suggests is that emotions can have a dual role in behavior, by affecting the rewards that are traded off against each other as well as the shape of the trade-off itself. In Chapter III I made somewhat similar arguments about the role of shame and envy in behavior. When one is suffering from intense shame, it is hard to imagine that the state will last forever. The overwhelming desire is for immediate release. Shame has a causal effect on the evaluation and perception of other rewards ever and above its own role as a negative reward. In the case of envy, the urgent wish for the destruction of the envied object or its possessor may induce destructive behavior that leaves the agent worse off rather than better off. These cases—guilt, shame, and envy—are not all similar. There is no particular reason to expect all emotions to interact with interest in the same way. Yet the cases all support the claim that the simple cost-benefit model is too simple.

Some of the remarks I have made about shame and guilt suggest that emotions could be modeled as temporary preferences. The person whose boss is a beggar in the street can feel an urge to give him money, or

the person who is in the grip of shame and feels an urge to kill himself may be viewed as undergoing a short-term change of preferences. It is in fact an important feature of many occurrences, that they have a relatively short duration. Anger, for instance, tends to "spend itself" quickly.116 Aristotle comments that "men become calm when they have spent their anger on someone else. This happened in the case of Ergopluheus. Though the people were more irritated against him than against Callisthenes, they acquiesced him because they had condemned Callisthenes to death the day before" (Rhetor. 1390b12—13). In trials of collaborators in German-occupied countries after World War II, those who were tried later generally received milder sentences even when the crimes were similar.117

Yet some emotions have additional horizontal character. In Becker's analysis of love (see below), the reason why a prudent man would take care to avoid low-income women is presumably that he might contract a lifelong disposition to share his income. One might question, perhaps, whether the relationship between spouses typically involves emotions in the full sense of the term. Marital love may involve concern for the welfare of one's spouse, but not the strong arousal and action tendencies that we associate with "impenetrable" (III.4). In her study of this emotion, Dorothy Tennov found that the typical duration of an episode was from eighteen months to three years, with some episodes lasting only a few weeks and others a whole lifetime. Revenge behaves another counterexample to the idea that emotions can be modeled as momentary preferences. In societies where blood feuds are common, revenge can be a lifetime obsession (III.4). The "prideful emotions," contempt and hatred, can also be very durable and frequently all-consuming. These "standing" emotions shape preferences in a durable manner. When they are all-consuming, we can best model them by a lexicographic preference ordering. For the person who is in a state of linearity or in pursuit of revenge, there is no trade-off between satisfaction of the emotion and material interest. Ordinary economic activities are pursued only to the extent that they promote—or do not come at the expense of—the emotional goal of the agent. Unlike the person who is in the grip of an acute emotion of shame or

117. Tannah (1964), Chapter 7, Anderes (1967), p. 249; Mowen (1964), p. 177, n. 36. The most thorough discussion is in Huys and Chomdt (1953), 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 Ettin (1968).
anger, people who are subject to an all-consuming durable passion are perfectly capable of acting in an instrumentally rational fashion. As Aristotle noted, the angry man is irrational whereas the man animated by hatred is not. The emotion may be grounded in an irrational belief, but that is another matter.

We may conclude, therefore, that the interaction between emotion and interest cannot be modeled in terms of competing ends and benefits. Concerning the short-lived emotions, the model correctly predicts that these will be a trade-off between emotional rewards and other rewards, but it fails to incorporate the fact that the trade-off itself may be shaped by emotion. Concerning the durable emotions, the model ignores that the pursuit of emotional satisfaction may be so fundamental to a person’s life that all other considerations become secondary. In brief summary, the short-lived passions undermine the theory of the rational actor, whereas the durable ones undermine the theory of homo economicus.

ACTIONS OR PASSIONS?
The traditional view rests on the premise that emotions are involuntary, suffered in a passive mode (whence “passion”) rather than chosen in an active mode. From Sartre onwards, a number of writers have denied this assumption. Here I discuss the two authors who have produced the most sustained arguments for the view that emotions are chosen rather than undergone, Roy Schafer and Robert C. Solomon. Although both claim that emotions are invariably rational, it follows from their views that emotions, being actions, are at least capable of being assessed as rational or irrational. Schafer does not acknowledge this implication, but Solomon does.

Schafer set out to create a “new language for psychoanalysis,” in which we would refer to emotions by verbs and adverbs – “action language” – rather than by nouns. As part of this project, he wanted to show that emotions are “enacted” rather than “undergone.” Because of the psychoanalytic framework, much of what he has to say

Rationality and the Emotions

is opaque, at least to me. I shall try, however, to extricate some arguments from his analysis that can be discussed independently of that framework.

To refute the idea that emotions occur involuntarily, Schafer begins by stating a paradox that we encountered in IV.2: “Even when we hold ourselves accountable for our emotions (e.g., ‘I hate myself for feeling so envious’), we usually go on believing that we are in fact finding or encountering our own emotions, thereby taking the paradoxical position of being responsible for what we cannot help feeling.” Now, there is a long tradition for arguing that one can be responsible for what one cannot help. Aristotle argued, for instance, that “to the unjust and to the self-indulgent man it was open at the beginning not to become men of this kind, and so they are such voluntarily, but now that they have become so it is not possible for them not to be so.” The same argument applies to the emotions. Yet it seems implausible to assume that when people feel guilty about their emotions it is because they believe there is something they could have done at some earlier time to develop different emotional dispositions. I may feel guilty about my inappropriate mirth at a funeral or my lack of happiness on my wedding day without knowing what I could have done to develop a disposition to react differently, or even without believing that there was something I could have done to develop it.

The fact that we do blame ourselves for our current emotions (or lack of them) is a prima facie argument in favor of Schafer’s thesis. Yet in blaming ourselves we might merely be misguided, just as Schafer argues that we are misguided in thinking that emotions occur involuntarily. The meta-emotion of guilt triggered by our first-order emotions or lack of them simply be irrational or inappropriate, in a sense to be discussed shortly. The idea, admittedly, is speculative and does not provide a compelling refutation of his argument. To supplement it, I shall try to refute two of his other arguments and later argue directly against the thesis itself.

Schafer is concerned with refuting the idea of a lawlike connection between triggering conditions and emotions. “Should this apparently

176. See Deleuze (1957) for discussions of this idea in early Greek thinking. Gordon (1987) is an excellent modern defense of the traditional view.
178. Schafer does not even say explicitly that emotions are chosen (this favorite verb is “enacted”) but the idea seems to follow from the general tenor of his discussion (notably Schafer 1970, pp. 101, 156).
182. Spivak and Silver (1987) comment that they “do believe that some people at some moment have been responsible for their emotions – although perhaps this responsibility is irrational, limited to those who have studied Aristotle’s theory.”
lawlike behavior stand up to examination, we should have to accept a passive conception of emotional experience: for then we could attribute activity legitimately only to a person’s getting into, staying in, or putting an end to those situations that must engender one emotion or another. Although I agree that the proper causal framework is that of mechanisms rather than lawlike regularities, I do not think this issue is related to that of activity versus passivity of the emotions. Schacter’s argument implies the absurd conclusion that the rational consumer is entirely passive, as we can predict his behavior with lawlike regularity: When prices go up, he buys less. A proponent of the view that emotions are actions might argue that agents predictably choose to have the emotion that will best promote their ends.

Schacter does not claim that the nonlawlike character of the emotions implies that emotions are chosen, only that their lawlike character would exclude that conclusion. To show that emotions are indeed chosen he relies on the fact that an individual can choose how to perceive the situation that triggers the emotion. Discussing the variety of responses to the extreme circumstances that the Nazis created for their victims, he writes that “in circumstances that seem identical to us as independent observers, some people conduct themselves more reasonably, resourcefully, patiently, boldly, or decently than others,” because “at least in some important respects the ‘stronger’ ones have defined their circumstances differently.”

Although the statement is not entirely clear, it can be read as asserting that people choose their emotions by deciding how to perceive the situation. I would use the belief-emotion connection to argue for the very opposite conclusion: It is precisely because people cannot “decide to believe” that they cannot decide which emotions to have.

Over the past twenty years Robert C. Solomon has developed the view that emotions are actions. The Myth of the Passions has so thoroughly indoctrinates us with its notion of passivity that we are no longer capable of seeing what we ourselves are doing. Once the Myth is exploded, however, it is obvious that we make ourselves

angry, make ourselves depressed, make ourselves fall in love.”

He claims, moreover, that emotions are always intelligible and often rational. “Every emotion is a subjective strategy for the maximization of personal dignity and self-esteem.” Like Schacter, he acknowledges that emotions are experienced as if they happen to us, but argues that this way of perceiving them is itself a matter of choice — “a vehicle of irresponsibility, a way of absolving oneself from those fits of sensitivity and foolishness that constitute the most important moments of our lives.”

The influence of Sartre, here as elsewhere, is evident. Another strong influence, as will be evident from the next paragraph, is Nietzsche.

Solomon’s claim about the rationality of emotions is probably best conveyed by some examples. Anger promotes self-esteem because it is always tinged with self-righteousness, except when it is directed against oneself, as in guilt. The latter emotion contributes to self-esteem because “the ability to admit and atone for our mistakes is . . . essential to wisdom and personal dignity.” Anxiety promotes self-esteem by virtue of being a “perverted form of self-aggrandizement.” Contempt does so by making oneself appear superior. Depression does so (at least in intention) by “a Cartesian method: Doubt everything until you find at least one value or task that remains unchallengable.” Despair contributes to self-esteem by encouraging self-indulgence and self-pity. Dread makes its contribution by virtue of being “an apt excuse for paralyzing curiosity.” Duty and morality “can be used as weapons against apparently superior forces.” Embarrassment allows one to be the center of attention, “even at the expense of dignity (but to the benefit of self-esteem).” Envy, in symbiotic alliance with resentment, “seeks to rob the superior man of his virtues and possessions” in an ideologically justified way. Fear may involve “protection of a self-image.” Gratitude “places oneself in the dignified position of passively receiving.” Hate, which presupposes self-esteem, seeks to expand it by confrontation with powerful and evil opponents.

194. Ibid., p. 386.
196. He first developed his theory of emotion as action in Solomon (1973). In Solomon (1978), which I cite in the revised edition of Solomon (1993), he modified some of the more outrageous claims from the article while still upholding the central argument. An explicit enumeration of the themes from the article that are abandoned in the book appears in the appendix in Solomon (1980), a reprint of Solomon (1973).

198. Solomon (1993), p. 160; Solomon (1994) distinguishes between two senses of rationality: The first corresponds roughly to the idea of intuitional or, as we shall, although left undefined, seems to correspond more or less to the idea of rationality used here.
201. From Solomon (1993), Chapter 8.
202. Solomon defines as guilt what I have defined as shame and vice versa. The reference in the text uses my terminology.
Alchemies of the Mind: Rationality and the Emotions

My résumé of the catalogue stops here, about halfway through Solomon’s alphabetical list of some 111 try-odd emotions. It might be clear enough from this condensed sample that his approach is both insightful and arbitrary. Many of his Nietzschean statements about how emotional states serve some further psychic purpose illustrate the “alchemies of the mind” further discussed in V.2. The vignette of “a woman who continues to patronize a shop which she knows has cheated her” because her small losses are made up by the self-righteous satisfaction of her continuing indignation203 rings true enough. Rather than supporting his theoretical argument, however, the story directly undermines it. It suggests that the woman chooses to get into a certain situation because it is, predictably—indeed, independently of her will—generates the gratifying emotion of indignation. Alternatively, the emotional gratification might act as a reinforcer on her behavior. On the basis of this and similar examples, one could interpret Solomon as arguing that people choose situations because of their emotional gratification they provide, which is of course different from his official view that they choose the emotions themselves.

Even on that more plausible interpretation, however, Solomon’s claim that all emotions always exist to promote self-esteem and personal dignity, and that they even maximize these values, cannot be taken seriously. The statement: that all emotional reactions can be explained by their contribution to self-esteem is a reckless generalization from a few selected cases. Most depressions are not vehicles for self-esteem; most cases of grief and sadness are not occasions for self-pity, and so on. The idea that an agent, faced with a given external situation, screens all possible emotional reactions and then selects the one that maximizes self-esteem and dignity is nothing short of absurd.

To be sure, someone steeped in the hermeneutics of suspicion can always come up with a story in which any emotional reaction can be seen as an ultrable self-serving strategy. “Most emotions involve . . . strategies for the maximization of self-esteem that would shame a professional confidence man and a prurient awareness of psychological intrigue that would impress even Dr. Freud.”204 Yet in my opinion, the ingenuity is all Solomon’s. He offers no evidence whatsoever for the mind’s capacity to engage in these strategic calculations. In fact, it is hard to square his views about the strategic nature of the emotions with two other views that he asserts, namely that the emotions are “undeliberated, unarticulated, and unreflective,”205 and that they are myopic because the purposes they serve tend to be short-sighted.206 Nor does Solomon offer any evidence for the supremacy of self-esteem over all other values or, more generally, for the idea that emotions exist to serve some psychic purpose.

Now, one cannot refute a conclusion by showing that arguments that have been made for it do not work. Turning now to a more direct approach, I can offer the following arguments in favor of the view that emotions are by and large, passively undergone rather than actively chosen or enacted. (i) Those who see emotions as actions either have to argue that this view applies to nonhuman organisms and small children as well as to adult humans or to argue for a radical discontinuity. In my opinion, both horns of the dilemma are highly unattractive. (ii) Even in adult humans, there is very strong evidence that “[b]ecause emotions can occur with anspic onset, through automatic appraisal, with little awareness, and with involuntary response changes in expression and physiology, we often experience emotions as happening to us, not chosen by us. One can not simply elect when to have which emotion.”207 (iii) If we can choose our emotions without any costs and constraints, why do we not choose always to be happy? (iv) If there are costs and constraints on which emotion can be chosen, what are they and where do they come from? Also, would not these constraints at costs precisely reflect the involuntary nature of the emotions? (v) Most emotions are triggered by beliefs, which cannot be chosen. (vi) Emotions that are not triggered by beliefs, such as panic and phobias, are even less plausibly viewed as chosen for the purpose of maximizing self-esteem.

In my opinion, therefore, the views of Schafer and Solomon are not only counterintuitive: They are false. There is, nevertheless, an element of truth in the view that the emotions are under our voluntary control. When an emotion arises from some external stimulus, we can “let it happen,” amplify it by giving full rein to its expression, or try to limit it, for example by directing our attention elsewhere.208 We can also create emotions without any external stimulus. Within limits

204. Ibid., p. 151.
205. Ibid., p. 151.
206. R. C. Solomon (1980), p. 265. This claim is not among those that are repudiated in the appendix.
Rationality and the Emotions

Izec. definitions of the emotions (IV.2), moralized assessments are not vulnerable to the same objections.

Let me recall the main examples cited here and in previous chapters of putatively irrational emotions:

1. Anger at those whom we have harmed
2. Anger at those who have helped us
3. Love of those whom we have helped
4. Love of those who have harmed us
5. Anger at those whom we bore
6. Hating those who fail to requite our love
7. Anger at a person who has proved us to be wrong
8. Anger at those who frustrate our goals even if they did not and could not have been expected to foresee the consequences of their actions
9. Guilt for emotions, thoughts, or events outside our control
10. Shame for emotions, thoughts, events, or character traits outside our control
11. Pride for events outside our control
12. Contempt for features of a target individual that are outside his control
13. Admiration for features of a target individual that are outside his control
14. Agapic love, that is, love which lacks cognitive support
15. Panics or phobias that lack cognitive support
16. Jealousy, which is always irrational, because it presumes cognitive antecedents that undermine the love which is a condition for the jealousy itself

With the exception of (9), (12), and (16), these seem all to be reasonably well-established phenomena. In cases (8), (12), (9), (6), and (7), damage to our self-esteem induces a hostile emotion towards another person. Without denying its reality, I find case (9) more opaque. In cases (8) through (13) an emotion is triggered by a truncated cognition, which focuses on a favorable or unfavorable outcome without any concern for its causal history. Cases (9) and (10) are especially puzzling. These emotions are not only inappropriate but painful. We see irrational anger is induced by the need to repair damage to our self-esteem and irrational pride actually enhances it, irrational
guilt and shame seem to be sources of countless suffering. One person feels ashamed when he goes bald or starts losing his sense of hearing and tries to hide it from others. Another feels guilty when he learns that a friend has died in a car accident and realizes that he could have prevented it by calling him up and delaying his departure by a few minutes. We may well wonder, What's in it for them? Case (13) is somewhat different: the others, in that people subject to these fears usually know they are unfounded.

In many of these cases, the irrational emotion goes together with irrational belief formation. Because our self-esteem may not allow us to harbor emotions that we cannot defend to ourselves and to others, we invent some kind of belief to justify even the most irrational reactions. Cases (1) and (4) are a unique example of this mechanism. Arguing from first principles, the process must have four stages. First, the emotion occurs. Second, there is an unconscious recognition that it is unjustified. Third, the recognition causes some kind of mental discomfort or distress. Finally, there is the invention of a justification to reduce the discomfort. In practice, we only observe the first and the last stage, which may occur almost simultaneously. I return to the interaction between emotion and cognition below and then again in V.2.

It remains to determine the criteria for an emotion being "appropri- ate" or "rational." To some extent, this may be a cultural matter. For the Greeks, cases (9), (12), and (13) were not instances of irrationality. In our own society, however, many may not view case (13) as irrational at all. I doubt, however, that culture can provide the whole explanation. I conjecture that there are no societies in which cases (1) and (3) are not seen as inappropriate by the agent and those around him. Case (1), in particular, seems relatively to a purely cultural explanation. All societies have norms for what constitutes justified harm. The norms may vary across societies, but there are no societies that do not have some norms of this kind. It follows that in all societies there will be cases in which people hurt others without being justified in doing so. If they get angry with a person when and because they have hurt him without justification, they are irrational. Although case (1) is instantiated in all societies, it will have different instantiations in different societies. By contrast, there may be some societies in which, say, case (3) is never instantiated, not because individuals do not feel contempt towards the ugly or disabused, but because they experience no discomfort in doing so. These are difficult issues and I am far from confident that I have got the right end of the stick.

Rationality and the Emotions

Irrational Emotions and Belief-Irrationality

One might also argue that emotions are irrational if their cognitive antecedents embody some form of belief irrationality. This proposal can be spelled out in a number of ways, which capture different varieties of irrational emotions. The most important case, of which I shall distinguish three subcases, is that of irrational emotions caused by emotionally induced belief irrationality.

Consider first, however, cases of belief formation by cold rationality, due to defective cognitive processing of various kinds.22 During the Second World War, Londoners were persuaded that the Germans systematically concentrated their bombing in certain parts of the city, because the bombs fell in clusters. This invalid inference, which reflected a lack of understanding of the statistical principle that random processes tend to generate clustering,23 probably made the Londoners living in those areas more fearful than they would otherwise have been. One might say, therefore, that their fear was irrational or irrationality strong. Conversely, the idea that lightning never strikes twice in the same place may induce an irrational lack of fear.

In the case just discussed the process that results in this irrational emotion does not itself originate in an emotion. Even if it does, emotionally caused beliefs need not be instances of motivated irrationality. Emotions may affect belief formation qua sheer arousal, regardless of their content and direction. It is difficult to make correct inferences when one is in the throes of a strong passion. Assume, for instance, that a salesman is unable to add up sums correctly because he is constantly distracted by thoughts about a woman he just met. Sometimes the errors in the sum might benefit the customer, sometimes himself. There is no reason to expect a pattern either way.24 If his trembling arithmetical hand leads him to believe that the customer owes him a great deal of money, any elation that might ensue would be irrational because grounded in an irrational belief.

A third variety arises when an emotion affects belief-formation and thereby generates another emotion—by virtue of its content rather than by virtue of its arousal properties. Consider the following generic tragic scenario. Emma loves Henry, who spurns her. The calculus

221. For a survey see Nisbett and Ross (1980).
224. We may note in passing that this case shows the need to distinguish between "hot" and "motivated" irrationality—two categories that have been treated as synonyms by several writers, including myself (see Rosen 1996, p. 48, n. 25).
attachment to Iago that makes him want to believe him (unlike the case of Paul discussed above), and believing Iago will make him feel bed. Although his perceptions of a racial prejudice may enter into his willingness to believe the worst, they cannot fully account for it—
and in any case there are many other instances of irrational jealousy in which there is no such prejudice at work. It is enough to cite the extensive analyses of jealousy in A la recherche du temps perdu.212

A more mundane illustration of the same mechanism is the following. Walking through a wood at night, we suddenly hear a faint sound and become worried that it might represent a danger. Our immediate fear leads us to interpret all sorts of other innocent sounds as ominous signs, until we start running in full-fledged panic. "In such an emotional inferno one cannot be reassured; that is not what one seeks. One acts to intensify whatever emotion is felt, turning fear into terror, anger into fury, disgust into revulsion, distress into anguish."

211 This is the phenomenon of counterbalanced thinking, which I have discussed in various sections (I.6, II.3, III.3). The mechanism seems mysterious, and yet it is hard to deny that it exists. As in irrational anger, an initial emotion induces a belief that justifies and even strengthens it, generating an "emotion attack." The difference with anger, to repeat, is that it's hard to see what's in it for the organism. In that respect, irrational jealousy and fear may be classified with irrational guilt and shame.

RATIONAL EMOTIONS AND THE GOOD LIFE

I have discussed whether and in which sense occurrent emotions can be rational. A different issue is whether emotional dispositions can be rational. With respect to this question one might argue that rational emotions are those that make one's life go as well as it could, given external circumstances. This idea can be spelled out in two ways. On the one hand, we can ask which set of emotional dispositions would be optimal to have. On the other hand, we can ask which dispositions it would be optimal to develop. Although one cannot acquire an emotional disposition in the way in which one can raise one's arm, by just doing it, one can engage in various actions that will, more or less predictably, produce it.

215 These cases of substitution do not represent displacement of affect in the Freudian sense. Displacement of anger from one object to another is one of the defense mechanisms (V.2, and supposedly motivated by the need to ward off unacceptable impulses. Substitution, by contrast, is caused by the importance of the original object and the need for a new one. Thus substitution also differs from what happens when "motional experience motivates search for causes explaining that experience... Why this angry mood? or what, precisely, in the other's action made one white with anger?" (Frijda 1986, pp. 235-37). The goal of substitution, by contrast, is to create an object for the emotion, not to discover it.

Rationality and the Emotions

Encounters (and their outcomes) and dispositions are not independent of each other. Encounters with the world are the most effective way of acquiring a disposition or changing it. Meditation and inner exhortation, by themselves, have limited force. Stendhal observed, "Everything can be acquired in solitude, except character"; also, "To have a strong character one must have experienced the effect produced by others upon oneself; therefore we need the others" (il faut les autres). Aristotle, too, argued that dispositions or habits are acquired by action: "[By] doing the acts that we do in our transactions with other men we become just or unjust, and by doing the acts that we do in the presence of danger, and being habituated to feel fear or confidence, we become brave or cowardly. The same is true for appetites and feelings of anger; some men become temperate and good-tempered, others self-indulgent or irascible, by behaving in one way or the other in the appropriate circumstances. Thus, in one word, states arise out of like activities." 210

Conversely my dispositions may affect the number of my encounters with others and shape the outcomes of such encounters in a direction that is more or less favorable to me. If others know that I am irascible, they will not deal with me if they can avoid it. (A benevolent disposition, on the other hand, may cause others to seek me out.) If they cannot avoid it, their fear of my temper will induce them to make concessions they would not otherwise have made. (A benevolent disposition, on the other hand, may induce me to make concessions I would not otherwise have made.) The net result of these two opposing effects is, in general, indeterminate.211 Irascible (and benevolent) people may or may not benefit materially from their disposition.

Given these preliminaries, the problem at hand can be defined as follows: In a parametric environment, which actions will bring about the encounters, outcomes, and dispositions that maximize net expected welfare over time? Although it is intuitively obvious that this question has no answer, it may be inductive to understand exactly why life could never be the solution to a maximization problem. Two main obstacles are, I submit, those of uncertainty and of by-products. Moreover, the emotions themselves can get in the way.

210. Stendhal (1828), Fragments 1, 48.
211. Nichomachean Ethics 1105b 10–13. The word translated as “state” is “hexis.” Sometimes rendered as “disposition.”
212. In “Is it in (proc. 3, section 12), I utilize Robert Frank and Adolf Ziegler for their exclusive focus on the benefits of a disposition to be angry.