JON ELSTER

RATIONALITY, EMOTIONS, AND SOCIAL NORMS

The set of human motivations can be classified in an indefinite number of ways. Many of the seventeenth- and eighteenth-century moralists drew a main distinction between *interests* and *passions*.[1] La Bruyère also included reason, as a sort of weak, younger brother: "Rien ne coûte moins à la passion que de se mettre au-dessus de la raison: son grand triomphe est de l'emporter sur l'intérêt".[2] In *Democracy in America*, Tocqueville suggested a different tripartite division of the motivations, into interests, passions, and social norms.[3] A similar distinction occurs in Pascal, for instance, in his discussion of the different motivations people may have for dueling.[4] In this article, I shall offer yet another trichotomy, by distinguishing among rationality, emotions, and social norms as the mainsprings of action. As several of the distinctions I have mentioned, the typology is neither exhaustive nor exclusive, nor does it proceed from first principles. It is strictly inductive, and, whatever value it might have, it derives from its ability to help us understand otherwise puzzling aspects of human behavior.

Covering a vast terrain and painting with a wide brush, the present article is largely programmatic. For the same reason, the exposition is somewhat schematic. I begin with brief statements of each the three approaches to behavior. I then proceed to the three pairwise comparisons that can be made among them, to point out conceptual differences as well as causal connections. The final section, on the relationship between rationality and the emotions, is by far the lengthiest and, in a way, the main *raison d'être* of the article. In this section, I try to go beyond the programmatic and schematic, to offer a more fine-grained analysis of the sense in which emotional life itself can be rational or irrational.

1. RATIONALITY[5]

The theory of rational choice is first and foremost a normative or prescriptive theory. It tells people how to choose and to act in order to achieve their aims as well as possible.[6] It offers also, but only

Synthese **98**: 21–49, 1994.

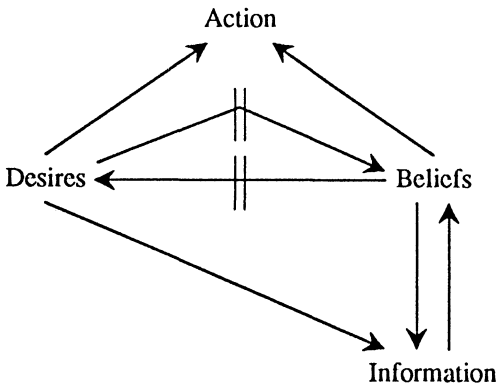

Fig. 1.

secondarily, an explanatory account of human behavior. In this perspective, the hypothesis is that one can explain how people act by assuming that they follow the prescriptions of the normative theory.

The basic structure of rational-choice explanation of behavior is set out in Figure 1. It involves three distinct conditions. First, for an action to be rational, it has to be the best means of satisfying the desires of the agent, given his beliefs. In itself, this is a very weak requirement. If I want to kill a person and I believe that the best way of doing so is to make a doll representing him and stick a pin through it, then according to this weak definition I act rationally if I make the doll and pierce it with a pin. We would hardly be satisfied with this conclusion, however, not because the homicidal desire is irrational (it may be immoral, but that is another matter), but because the beliefs are transparently ill-founded.

Second, therefore, we need to stipulate that the beliefs themselves are rational, in the sense of being grounded in the information that is available to the agent. These may be beliefs about factual matters or about general lawlike connections. In particular, they will include beliefs about the *opportunities* available to the agent. In fact, rational-choice theory is often stated in terms of desires and opportunities rather than desires and beliefs. In that version, the theory says that a rational agent chooses the most preferred element in his opportunity set. In some simple choice situations, this formulation is adequate enough. In general, however, we need to take account of the fact that the full set

of objective opportunities available to the agent may not be known to him. An automobilist arriving in an unknown city without a map will not know the full set of paths that will take him through it.

In such cases, the agent must use whatever information he has, to form some belief or subjective estimate of the alternatives. The fact that it is subjective does not in itself detract from its rationality. On the contrary, *the concept of rationality is subjective through and through*. To be rational does not mean that one is invariably successful in realizing one's aims: it only means that one has no reason, after the fact, to think that one should have acted differently. Nor does a rational belief have to be true: it must only be well grounded in the available information. Beliefs are rational if they are formed by procedures that in the long run tend to produce more true beliefs than any alternative procedure, but on any particular occasion the belief thus formed may not correspond to the facts. This being said, belief formation *is* vulnerable to distorting influences of various kinds. Some of these are more in the nature of mistakes, as when we get sums wrong in arithmetic. Others, however, belong to the category of *motivated irrationality*, as when the adding-up errors made by a salesman systematically (although non-intentionally) work out to his favor.[7]

However, a belief is not made rational simply by being well grounded in the available information. If the automobilist is in a hurry, he should perhaps buy a map to acquire more information about the feasible paths. The third condition for rational behavior, therefore, is that the agent should acquire an optimal amount of information, or, rather, invest an optimal amount of time, energy, and money in gathering such information. Clearly, it will often be irrational not to invest any time in collecting information. If one is buying a house or a car, one should compare several options and investigate each of them in some depth. Equally clearly, there are occasions when there is a danger of gathering too much information. If the doctor makes too many tests before deciding on treatment, the patient may die while in his care. Between these extremes, there exists an optimal level of search, a 'golden mean'. (Whether this optimum can be known is another matter.)

As depicted in Figure 1, there are several factors that determine the amount of information that a rational agent will gather. The agent's beliefs about the expected costs and expected value of gathering the information will obviously matter. His desires – i.e., how important the decision is to him – will also enter into the calculus. Indirectly,

therefore, the desires of the agents will enter into the process of belief formation. However, the blocked arrow from desires to beliefs is intended to indicate that a direct influence, as in wishful thinking, is inadmissible. The blocked arrow in the opposite direction represents a more controversial claim, viz., that the desires of a rational agent should not be shaped by his beliefs. On this criterion, the fox in the fable of the sour grapes was the victim of irrational preference formation.[8]

2. SOCIAL NORMS[9]

A defining feature of social norms, in the conception offered here, is that they are *not outcome-oriented*. The simplest social norms are unconditional: Do X, or: Don't do X. More complex norms are conditional on the past behavior of the agent or of other people: If you have done Y, then do X, or: If others have done X, then do X. Still more complex norms might say: Do X if it would be good if everyone did X. Social norms are either unconditional or, if conditional, not future-oriented. Rationality, by contrast, is clearly forward-oriented. 'Let bygones be bygones' is a key prescription of the theory. For norms to be *social*, they must be shared by other people and partly sustained by their approval and disapproval. They are also sustained by the feelings of embarrassment, anxiety, guilt, and shame that a person suffers at the prospect of violating them. Section 5 elaborates this point.

Social norms can be distinguished from a number of other, related phenomena. First, they differ from moral norms, as some moral norms, as those derived from utilitarian ethics, are squarely consequentialist. Second, social norms differ from legal norms. Legal norms are enforced by specialists who do so out of self-interest: they will lose their job if they do not. By contrast, social norms are enforced by members of the general community, and not always out of self-interest. Third, social norms are not convention equilibria, since people need no motive beyond self-interest to adhere to conventions. Fourth, social norms differ from private norms, the self-imposed rules that people construct to overcome weakness of will.[10] Private norms, like social norms, are non-outcome-oriented and are sustained by feelings of anxiety and guilt. They are not, however, sustained by the approval and disapproval of others, since they are not, or not necessarily, shared with others. Finally, norm-guided behavior must be distinguished from habits and

compulsive neuroses. Unlike social norms, habits are private. Unlike private norms, their violation does not generate self-blame or guilt. Unlike neuroses and private norms, habits are not compulsive. Unlike social norms, compulsive neuroses are highly idiosyncratic. Yet what in one culture looks like a compulsive neurosis may, in another society, be an established social norm.

To fix our ideas, I shall give some examples of social norms. *Consumption norms* regulate manners of dress, manners of table, and the like. As shown by Proust's masterful account of life in the Guermantes circle, conformity with such norms can be vitally important to people, even though nothing of substance is at stake. *Norms against behavior* '*contrary to nature*' include rules against incest, cannibalism, homosexuality, and sodomy. *Norms regulating the use of money* often become legal rules, such as laws against buying and selling votes. Often, however, they remain informal, as the norm against buying a place in a bus queue. *Norms of reciprocity* enjoin us to return favors done for us by others. Gift-giving is often regulated by these norms. *Norms of retribution* enjoin us to return harms done to us by others. Rules regulating revenge are often highly elaborate.[11] *Work norms* regulate, among other things, the effort put in by the workers: neither a chiseler nor a ratebuster be. *Norms of cooperation* include what one may call 'everyday Kantianism': cooperate if and only if it would be better for all if all cooperated than if nobody did. Another such norm is a 'norm of fairness': cooperate if and only if most other people do. Finally, *norms of distribution* regulate what is seen as a fair allocation of income or other goods. In democratic societies, the norm of equality is especially strong.

3. THE EMOTIONS[12]

The most striking feature of the emotions is their component of arousal. The level of arousal may be low, but it cannot be zero. The idea that art is 'emotion recollected in tranquillity', for instance, implies that the emotions are not experienced in tranquillity. When we say that a person is 'emotional', we certainly intend the opposite of a calm and composed attitude. For some purposes, the arousal in itself suffices to explain behavior. If a rifleman's hand shakes because he is excited, he is likely to miss his goal, whether the excitement stems from joy, anger, fear,

or hope. However, most implications of emotions for behavior depend on their specific character and intentional object.

Emotional experiences have three further properties. First, they have different qualitative, phenomenological properties. Second, they are inherently positive or negative, in a sense to be discussed. Third, they usually have an intentional object – they are *about* something.[13] Because of the presence of arousal, qualitative feel, and inherent pleasantness or unpleasantness, the state of 'emoting that p' differs from other intentional states, such as the state of 'desiring that p' or 'believing that p'.

I do not know if other people see colors as I do, nor if their emotional states are the same as mine. When they experience shame, do they feel what I feel when I am ashamed? The question cannot be answered, and may not even be a meaningful one. However, I can tell from introspection that my various emotional states differ in their phenomenological qualities, and I have no reason to doubt the reports of others that theirs do, too. Moreover, there seems to be rough intersubjective agreement as to how many different states there are, and which perceived situations tend to elicit them. (Again the comparison with colors is instructive.) Joy and grief, love and hate, hope and fear, anger and shame, envy and *Schadenfreude*, disappointment and regret, disgust and contempt, for instance, seem to be internally homogeneous and externally well differentiated emotional states. There may well be others, but the limits of introspection and the fuzziness of language may prevent their identification.

Emotions can be negative or positive. In saying this, I have in mind that an emotion can be good or bad for the person who has it at the moment he is having it, not that it can have good or bad adaptational consequences or arise from goal-satisfying or goal-frustrating conditions. We could try to place the immediate subjective experiences on a continuum from pleasure to pain, so that specific cases of shame and grief, for instance, might involve the same amount of pain, and differ only in their phenomenological properties and intentional objects. From introspection, this does not seem right. The attractive or repulsive properties of an emotion are part of its qualitative feel, not something added to it. It might at least seem that we can compare the attractiveness of different emotional states by asking ourselves which state we would rather be in. But the answers to such thought experiments, even if they were reliable, would probably not be valid. We cannot simply

ask people to say whether they would rather be ashamed than grieving, because each of these states could occur with different degrees of intensity. But we cannot specify the intensity either, because there is no intersubjectively meaningful measure. We would have to specify the situations that are supposed to give rise to the emotions. But then people's preferences over the emotional states would be inextricably intertwined with their attitudes to the eliciting situations.

Finally, emotions are 'about' something, they have an intentional object or target. They differ in this respect from mere feelings such as nausea or vertigo. They also differ from what are usually called moods, relatively undifferentiated and untargeted states of contentment and discontentment. Earlier, I listed fourteen instances, ranging from joy to contempt, of internally homogeneous and externally differentiated emotions. Of these, only a few – joy, grief (or sadness), anger – also exist in the form of moods. However, on reflection we may find that what we think of as a mood really is a heightened disposition to have occurrent emotions. 'He is in a happy mood' may refer to an uninterrupted state of well-being, or to an enhanced tendency to experience pangs of well-being. Reasoning from first principles suggests that the latter is the more plausible idea. The life of the mind is a succession of events, not an enduring state. On this understanding, the reason why moods appear to lack an object or a target is simply that the object is constantly changing. In any case, moods – be they states or dispositions – should be distinguished from the durable dispositions that we attach to a person's character rather than to transient circumstances in which he finds himself. You do not have to be an irascible person to be in an angry mood, and neither is a necessary condition for anger.

Following Hume, one commonly distinguishes between the object and the cause of an emotion.[14] Sometimes, the two coincide, as when a person slaps my face and I get angry at him or when I feel envy at a colleague's winning the prize I coveted. Often, however, the coincidence does not obtain, or obtains only in a somewhat rough fashion. A crucial fact about the emotions is that they have the capacity to alter and distort the cognitive appraisal that triggered them in the first place. *The object of an emotion is the emotionally distorted picture of its cause.* This feedback from emotions to their cognitive origins is a key to the dynamics of the emotions, and explains how they can escalate and get out of hand.

4. RATIONALITY AND SOCIAL NORMS[15]

The sharp contrast I have drawn between rationality and social norms suggests that people who are motivated by norms are irrational. Those who resist this conclusion typically try to reduce social norms to some indirect form of outcome-oriented, rational behavior. I shall discuss three varieties of reductionism. The first, (A), says that norms are nothing but *ex post* rationalizations of self-interest; the second, (B), that people follow norms because of a(n outcome-oriented) fear of being punished if they do not; and the third, (C), that non-consequentialist behavior can be explained by the fact that it often has good consequences. Although each argument can account for some norms, none explains all of them; nor do they jointly account for all norms.

(A) Is it true, as argued by early generations of anthropologists and sociologists, that norms are in the saddle and people merely their supports? Or is it true, as argued by more recent generations, that rules and norms are just the raw material for strategic manipulation or, perhaps, for unconscious rationalization?

Social psychologists have studied norms of distribution to see whether there is any correlation between who subscribes to a norm and who benefits from it. Some findings point to the existence of a 'norm of modesty': high achievers prefer the norm of absolute equality of rewards, whereas low achievers prefer the norm of equity, i.e., reward proportional to achievement. More robust, however, are the findings that suggest people prefer the distributive norms that favor them. This corresponds to a pattern frequently observed in wage discussions. Low-income groups invoke a norm of equality, whereas high-income groups advocate pay according to productivity.

Conditional norms lend themselves easily to manipulation. There is, for instance, a general norm that whoever first proposes that something be done has a special responsibility for making sure that it is carried out. This can prevent the proposal from ever being made, even if all would benefit from it. A couple may share the desire to have a child and yet neither may want to be the first to propose the idea, fearing that he or she would then get special child-caring responsibility. The member of a seminar who suggests a possible topic for discussion is often saddled with the task of introducing it. The person in a courtship who first proposes a date is at a disadvantage. The fine art of inducing others to make the first move, and of resisting such inducements,

provides instances of instrumentally rational exploitation of a social norm.

Some have said that this is all there is to norms: they are tools of manipulation, used to dress up self-interest in more acceptable garbs. But this cannot be true. Some norms, like the norm of vengeance, obviously override self-interest. In fact, the cynical view of norms is self-defeating. "Unless rules were considered important and were taken seriously and followed, it would make no sense to manipulate them for personal benefit. If many people did not believe that rules were legitimate and compelling, how could anyone use these rules for personal advantage?".[16] Or again, "if the justice arguments are such transparent frauds, why are they advanced in the first place and why are the given serious attention?".[17] If some people successfully exploit norms for self-interest purposes, it can only be because others are willing to let norms take precedence over self-interest.

(B) When people obey norms, they often have a particular outcome in mind: they want to avoid the disapproval – ranging from raised eyebrows to social ostracism and physical violence – of other people. Suppose I face the choice between taking revenge for an offence to my sister and not doing anything. The cost of revenge is that I might in turn be the target of counter-vengeance. This outcome is not certain, for – depending on the specific rules in force – the avenger might also obtain satisfaction by killing my brother or my cousin, but it remains a distinct possibility. And even if I am not killed in the present round, I might be the target of later acts of retaliation. The cost of not taking revenge is, at worst, that my family and friends desert me, leaving me out on my own, defenselessly exposed to predators. At best, I will lose their esteem and my ability to act as an autonomous agent among them. A cost-benefit analysis is likely to tell me that revenge (or exile) is the rational choice. More generally, norm-guided behavior is supported by the threat of social sanctions that make it rational to obey the norms. Akerlof argues, along these lines, that in India it is rational to adhere to the caste system, even assuming that 'tastes' are neutral.[18]

We may counter the claim that people obey norms because of the sanctions attached to violations of norms by asking why people would sanction others for violating norms. What's in it for them? One reply could be that if they do not express their disapproval of the violation, they will themselves be the target of disapproval by third parties. When there is a norm to do X, there is usually a 'meta-norm' to sanction

people who fail to do X, perhaps even a norm to sanction people who fail to sanction people who fail to do X. As long as the cost of expressing disapproval is less than the cost of receiving disapproval for not expressing it, it is in one's rational self-interest to express it. Now, expressing disapproval is always costly, whatever the target behavior. At the very least, it requires energy and attention that might have been used for other purposes. One may alienate or provoke the target individual, at some cost or risk to oneself. Opportunities for mutually beneficial transactions are lost when one is forbidden to deal with an ostracized person. By contrast, when one moves upward in the chain of actions, beginning with the original violation, the cost of receiving disapproval falls rapidly to zero. People do not usually frown on others when they fail to sanction people who fail to sanction people who fail to sanction people who fail to sanction a norm violation. Consequently, some sanctions must be performed for motives other than the fear of being sanctioned. In a system of norms, there must be an unmoved mover.

(C) Some social norms can be individually useful, such as the norm against drinking or overeating. Moreover, people who have imposed private norms on their own behavior may join each other for mutual sanctioning, each in effect asking the others to punish him if he deviates, while being prepared to punish them if they do not punish him. Alcoholics Anonymous provides the best-known example. "Each recovering alcoholic member of Alcoholics Anonymous is kept constantly aware, at every meeting, that he has *both* something to give *and* something to receive from his fellow alcoholics".[19] Most norms, however, are not social contracts of this kind.

Another argument for the view that it is individually rational to follow norms is that they lend credibility to threats that otherwise would not be believable. They help, as it were, to solve the problem of time inconsistency. Vendettas are not guided by the prospect of future gain but triggered by an earlier offence. Although the propensity to take revenge is not guided by consequences, it can have good consequences. If other people believe that I invariably take revenge for an offence, even at great risk to myself, they will take care not to offend me. If they believe that I will react to offence only when it is in my interest to react, they need not be as careful. From the rational point of view, a threat is not credible unless it will be in the interest of the threatener to carry it out when the time comes. The threat to kill oneself, for instance, is not rationally credible. Threats backed by a code of honor

are very effective, since they will be executed even if it is in the interest of the threatener not to do so.

This observation, while true, does not amount to an explanation of the norm of vengeance. When a person guided by a code of honor has a quarrel with one who is exclusively motivated by rational considerations, the first will often have his way. But in a quarrel between two persons guided by the code, both may do worse than if they had agreed to let the legal system resolve their conflict. (Mafiosi seem to do better for themselves in the U.S. than in Sicily.) Because we are talking about codes of honor that are shared social norms, the latter case is the typical one. The rationality of following the code then reduces to the desire to avoid sanctions, discussed above. In any case, one cannot rationally decide to behave irrationally, even when one knows it would be in one's interest to do so. To paraphrase Weber, a social norm is not like a taxi from which one can disembark at will. Followers of a social norm abide by it even when it is not in their interest to do so. In a given situation, following the norm may be useful, but that is not to say that it is always useful to follow it. Moreover, there is no presumption that its occasional usefulness can explain why it exists.

5. EMOTIONS AND SOCIAL NORMS

Emotions and social norms are both poorly understood. These two failures of the social sciences are not independent of each other; rather, they are as closely related as are the phenomena themselves. Emotions and social norms are, as it were, sister stepchildren. Emotions are involved in all social norms, as their external or internal enforcers. Social norms regulate the expression of emotions and sometimes the emotions themselves. Emotional reactions to emotional states are often, therefore, mediated by social norms.

When Lazarus asserts that it is "surprising to find Averill (1983) also maintaining that rather than being destructive, anger, too [i.e., in addition to shame and guilt] upholds social norms",[20] he fails to see that anger and shame are two sides of the same coin. The expression of anger and indignation tends to call forth the intensely unpleasant emotion of shame in the person at whom they are directed. The anticipation of this emotion is the ultimate enforcer of social norms. I need to say a few words here about the relationship between shame and guilt. With some reservations stated below, I tend to accept the view

that both shame and guilt are forms of self-blame, but that the former is global or characterological (I am a bad person) and the latter specific or behavioral (I committed a bad action).[21] We might expect, perhaps, that violation of a behavioral norm would induce guilt rather than shame. The fact that this is not the case enables us to infer that adherence to a social norm is seen as a particularly important and intimate part of the person. Violations of the norm tends to elicit emotions in others that induce shame rather than guilt in the violator. "The disgust/contempt of another also forces us into a global attribution. It is very difficult to envision someone making a specific attribution in the face of another's disgusted/contemptuous look because the look says, 'You disgust me'".[22]

I now consider the other side of the relation: the regulating influence of norms on the emotions. The issue has two aspects. On the one hand, there are normative expectations about which emotions one should *feel* under specific circumstances. On the other hand, there are norms regulating what emotions one is expected to *express*. I begin with the latter issue, which is more accessible and has been more widely discussed.

Ekman has observed that emotions are subject to what he calls cultural "display rules", which specify "who can show what emotion to whom, when". He goes on to offer the following examples:

> The prohibition against showing anger, or the rule to substitute sadness for anger, is learned so well by some middle-class American girls, that later, if liberated, it requires some struggle to "get their anger out". Other display rules are learned more by example, by observing what others do or following implicit instructions of those who manage events when emotion is made the occasion for public ceremony. The performance of such display rules may not be as good, but errors are usually overlooked. An example of this type of display rule is that at beauty contests a winner may cry but not the losers. At funerals, one can note almost a "pecking order" of grief expressions based on the rights to mourn. A man's secretary cannot look sadder than his wife unless she intends to state something quite different about the true nature of their relationship.[23]

Hypocrisy is organized around such display rules. Rather than being the homage that vice pays to virtue, most hypocritical behavior is dictated by social norms. In some societies, such as the former USSR or China under Mao, hypocrisy has had an extreme or pathological character: everybody knew that everybody's enthusiasm for fulfilling the plan or anger at the class enemy was entirely faked, and yet one would lose one's job or be expelled from the party if one failed to

conform. But under normal conditions, hypocrisy is part of what makes the world go round, not so much by forcing us to express emotions we do not have as by keeping us from expressing those we do. There are even norms regulating the emotions that are appropriate to express as sanctions for norms violation. Although society would be a horrible place if norms of politeness and minimal helpfulness were not respected and enforced, it would not be much better if norm-violators were consistently terrorized. These meta-norms are sustained by meta-emotions, that is, by expressing disapproval of moralizers who express their disapproval too strongly.

Norms directed at the emotions themselves, not simply their expression, constitute a deeper issue. Emotions, like thoughts, are not under one's immediate control. Hence, injunctions about what to feel or not to feel, as well as rules about what to think and not to think, might seem to be pointless. Yet Christianity enjoins us to refrain from even thinking about our neighbor's wife, and middle-class mothers tell their daughters that good girls do not even think about sex. If taken seriously, such injunctions can produce a state of hopeless confusion and guilt in the recipient.[24] The same is true about the social norms that regulate the emotions. Failure to conform to the norm that one ought to grieve when a close relative dies, or to be happy on the day of one's wedding, tends to induce guilt. This presents a puzzle. As I said, I tend to accept the view that guilt is attached to actions rather than to character. Moreover, I also agree with the accepted view that guilt tends to motivate one to atone for the bad action, whereas the action tendency of shame is to hide or disappear. Failure to feel grief or happiness, however, would seem to reflect on one's character rather than being attributable to a specific action. The puzzle, then, is why we feel guilt rather than shame when we violate a norm about how to feel, especially as there is no way in which one can atone. Part of the solution may lie in the old-fashioned idea that shame attaches to publicly observable behavior, whereas guilt is a more private affair.[25] But as long as this idea is not integrated with what I called 'the accepted view', the puzzle remains.

Let me compound the puzzle. In one of the rare discussions of the norms that regulate emotions, Hochschild cites the following example. "Each of two mothers may feel guilty about leaving her small child at day care while working all day. One mother, a feminist, may feel that she should not feel as guilty as she does. The second, a traditionalist,

may feel that she should feel more guilty than, in fact, she does feel".[26] It follows from what I have said that the first mother would feel guilty about her guilt. But does that make sense? Isn't it much more plausible that a feminist would rather be *ashamed* at feeling guilt in this situation?[27] Although I am reasonably confident in my intuitions about this case as well as the grief-happiness cases of the previous paragraph, I am unable to locate the relevant difference in the eliciting situations.

It is not quite true, of course, that our emotions are outside our control. I shall have more to say about this in the next section, when I discuss whether emotions may be seen as rational actions. Here I shall just point to the relevance of the distinction between occurrent emotions and emotional dispositions. The former are to a very small extent subject to our will. The latter can, to a larger degree, be shaped by conscious character planning, such as Montaigne tells us he practiced to overcome his fear of death.[28] I find it impossible to imagine, however, that we might train ourselves to experience grief or happiness on the appropriate occasions. A more likely response to the norm-violation is to make pseudo-atonement by producing a feeling of pseudo-grief or pseudo-happiness, that is, to will the appropriate feelings by the kind of intimate hypocrisy in which we all indulge from time to time.

6. RATIONALITY AND THE EMOTIONS

Rationality and the emotions are often thought to be opposites. 'Rational' and 'emotional' are opposed to each other as character descriptions. The emotions are supposed to interfere with our ability to form rational beliefs and to make rational choices. They are like sand in the machinery of action. It does not take much reflection to see that this picture is inadequate. The emotions may themselves be subject to rationality criteria. They may facilitate cognition rather than obstruct it. And, finally, the emotions have an indispensable role to play in providing a sense of meaning and direction in life. Without the emotions there would be little reason to act at all.

Figure 2 suggests a way in which to insert emotions into the standard model and, at the same time, state the question of the rationality of desires.

A gloss on the arrows going to and from the emotions follows. Emotions depend on beliefs because of the need for a cognitive appraisal of the eliciting situation before any emotional reactions can be

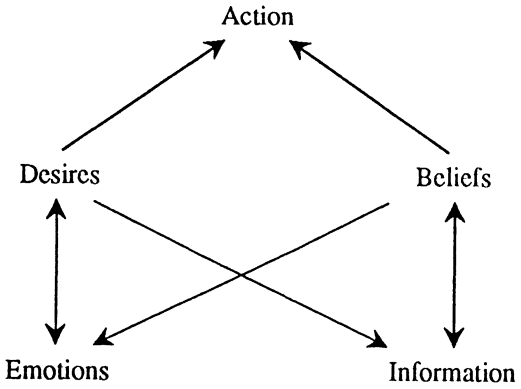

Fig. 2.

triggered. I will not get angry or ashamed at my Albanian host who passes me a cup of tea under his left arm if I do not know that among Albanians this is a way of expressing contempt. I will not get upset with your lying, unless I believe that you are in fact knowingly uttering a falsehood with the intent to deceive me. Causal attributions also matter for the emotion that occurs when something goes wrong or well in our lives.

That emotions can depend on desires is no less obvious. Many emotions arise as a direct result of the satisfaction or frustration of desires. Lazarus, for instance, distinguishes between goal-incongruent emotions (anger, fright-anxiety, guilt-shame, sadness, envy-jealousy, disgust) and goal-congruent ones (happiness, pride, love/affection, relief).[29] Conversely, an emotion can trigger a desire to change or maintain the situation that has triggered it. In many cases, we observe both mechanisms in succession: the frustration or satisfaction of a desire causes an emotion, which in turn causes a new desire to appear. Suppose I would like to marry Patricia, who seems to prefer Henry. The frustration of my desire may cause jealousy, which in turn may induce a desire to hurt Henry or to make him appear inferior in Patricia's eyes.

However, the arrows added in Figure 2 do not necessarily have the same interpretation as those in Figure 1. Clearly, the arrows going to and from the emotions have a causal interpretation, but can they also be interpreted as optimality conditions? The question can be decomposed in two parts. First, what meaning if any can we attach to the

idea that an emotion is rational, given certain beliefs and desires? (Actually, I shall be more concerned with irrational emotions than with rational ones.) Second, what meaning if any can we attach to the idea that a desire is rational, given certain emotions?[30]

I shall approach the first issue by asking a somewhat more general question: What meaning if any can we attach to the idea of rational emotions? I shall consider six possibilities. One, emotions are actions; rational emotions, therefore, are simply a variety of rational action. Two, emotions are rational if they are appropriate in kind to the occasion. Three, emotions are rational if they are appropriate in degree to the occasion. Four, emotions are rational if they are based on rational beliefs. Five, emotions are rational if they are based on rational (i.e., autonomous) desires. And six, emotions are rational if they make us happy.

One, consider Figure 1 again. We observe that both 'action' and 'information' are assessed as rational in light of the desires and beliefs of the agent. Here, 'information' is really shorthand for an *action*, viz., the act or process of gathering information. We are asking whether the agent invested an optimal amount of energy, time, and money into gathering information, and whether he looked in the right places. We may express this fact by saying that every action is accompanied by a prior "shadow action" of information-gathering.[31] It is natural to ask, then, whether emotions, if they also are to be assessed as rational in the light of beliefs and desires, might be conceptualized as actions. Solomon claims, for instance, "that emotions are rational and purposive rather than irrational and disruptive, are very much like actions, and that we choose an emotion much as we choose a course of action".[32] He offers the example of a man whose wife wants to go to a party while he does not. While she resigns herself to staying home, sighing occasionally, he goes into a rage about some shirts she has failed to pick up from the dry cleaners. "The anger can be explained, not in terms of what it is 'about' or what causes it, but in terms of its *purpose*. The husband, in this case, has *used* his anger to manipulate his wife. He has become angry 'about' the shirts *in order to* get his wife's mind off the party and in order to stop her irritating reminders".[33]

To me, the example (and the theory it is supposed to illustrate) is unconvincing. In the first place, there must be some *awareness* on the husband's part that his anger is likely to distract his wife. The link between anger and its effect cannot be due to reinforcement, which is

a non-intentional causal mechanism. Moreover, this awareness must be unconscious. The conscious knowledge about the effect of anger might induce him to fake it, and the faking might in turn bring about the real thing, but it could not bring about anger directly. Solomon recognizes as much. Moreover, the unconscious must be able to act strategically – to take one step backward (by accepting the unpleasant emotion of anger) in order to take two steps forward (redirecting his wife's attention). This assumption is implausible in itself, and contradicts what Solomon says elsewhere. It is implausible in itself, because it makes the unconscious too much like the conscious mind. I know of no evidence that the unconscious is capable of strategic behavior. On the contrary, it does not seem to be able to go beyond gradient-climbing: seeking the path of least resistance or of greatest pleasure. Also, the idea of a strategic unconscious is inconsistent with what Solomon says about the essentially myopic nature of emotional reactions. "Emotions serve purposes and are rational; but because the purposes emotions serve are often short-sighted, they appear to be nonpurposive and irrational on a larger view. For the sake of a passion, we destroy careers, marriages, lives".[34]

Two, it follows that any approach to the rationality of emotions must take a different route. Perhaps emotions are more like beliefs, which are said to be rational if they stand in a certain relationship to the available information, even though they cannot be chosen at will. One cannot decide to adopt a certain belief for the reason that having it will have useful consequences (e.g., that my marriage will be better if I choose not to believe that my wife has a lover), nor to have a certain emotion for similar instrumental reasons. We are then led to consider the idea that an emotion might be more or less appropriate or adequate to the occasion that triggers it. Now, this cannot simply mean that it is the kind of emotion that is expected and approved by others on that kind of occasion: that is a question of social norms, not of rationality. Nor can it mean that the emotion has some kind of necessary or conceptual connection with its eliciting conditions, because then emotions could never be irrational or inappropriate. We are looking for a notion of intrinsic but non-conceptual appropriateness and (especially) inappropriateness.

Consider the case of 'displaced' emotion, such as the man who is angry with his family because he has received a dressing down at his job.[35] His reaction is inappropriate because the causal chain goes in

the wrong direction: the emotion arises independently and then latches on to an object that has done nothing to provoke it. Or consider the man whose anger is transformed into righteous indignation when he redefines the goal-obstructing behavior of another as *illegitimate* goal-obstruction. Another example would be the transmutation of envy into indignation: if somebody else has what I covet, he must either have done something immoral to get it or have actively tried to prevent me from getting it. In the latter two examples, the emotion acts on the *perception* of the object that generated it, differing in this respect from the first example in which the emotion attaches itself to an object that did not under *any* description have a causal role in generating it. The underlying mechanisms are also quite different. The first example owes nothing to the need to justify one's emotions, as do the later examples, and everything to the need to express them. Yet we can use the feature they have in common to define one approach to the problem of the rationality of the emotions: an emotion is irrational or inappropriate if there is an emotionally caused divergence of the intentional object of the emotion from its cause.

Here is a different kind of example. When I have friends visiting from out of town, I am usually apprehensive about the weather. If it rains throughout their stay, I feel guilty. That would seem to be an inappropriate emotion: there's nothing I could have done about the weather. Or suppose that my child suffers an accident walking to school, and I feel a pang of guilt because I had in fact been thinking about driving her to school on that day but decided not to.[36] Although I could have done something to prevent the accident, I had no reason to think there would be one. In both cases, I blame myself, although nobody else would. If they did, it would be an irrational reaction, similar to the man who blamed his wife for being away the day of the burglary. It seems to make sense to say that if I think it would be irrational for A to blame B for X, then it is irrational for me to blame myself for X if I am in B's situation.[37]

Broad offered the following account of two ways in which emotions can fail to be appropriate:

> Some kinds of emotional quality are *fitting* and others are *unfitting* to a given kind of epistemological object. It is appropriate to cognize what one takes to be a *threatening* object with some degree of *fear*. It is inappropriate to cognize what one takes to be a fellow man *in undeserved pain or distress* with *satisfaction* or with *amusement*. Then again,

an emotion which is fitting *in kind* to its epistemological object, may be unfitting in *degree*, i.e. inordinate.[38]

We may easily agree that it is morally wrong to feel amusement at someone else's undeserved pain, but is it irrational? My inclination is to say that it such reactions are pathological rather than irrational. The psychopath who is angry at anyone who frustrates his desires, however inadvertently and innocently, is sick, not irrational. The sign of his being sick in fact is that he does not display the irrational tendency of seeking to justify his reaction by finding fault in the victim.

Three, let us now consider Broad's second suggestion: emotions are inappropriate or irrational if they are disproportionate in strength or degree to their intentional object. Let us first exclude phobias. If I am very afraid of, say, spiders, while knowing perfectly well that they are harmless, the only appropriate degree of fear is zero. Such emotions belong to pathology rather than irrationality. What Broad has in mind is the case in which some non-zero degree of the emotion is appropriate. Let us next exclude the case (discussed above) in which an emotion is disproportionate to its cause, although adequate to its object. Moreover, we should obviously exclude the case in which the inappropriateness is defined in terms of the real features of the object (spiders are not really harmful, although I believe they are). Finally, we should disregard the role of social norms in defining what is the appropriate intensity of an emotion.

What is left are cases such as the following. I know that there is some risk associated with crossing a street, even on a green light. It is possible, although unlikely that a car might come at great speed and hit me before I have time to retreat. This belief, together with the desire not to get hit, might cause me to feel some modicum of fear, not only in the aseptic sense of 'fearing that', but also in the sense of 'fear and trembling'. In itself, there need not be anything inappropriate about this feeling, especially if the street is in São Paulo rather than in Oslo. However, it ceases to be appropriate if the fear is so great as to paralyze me, leaving me helpless on the street corner for hours. Now, I can see two grounds on which such fears might be said to be irrational. First, one might appeal to consistency criteria for emotions. If my fear is greater on this occasion than on others which I believe to carry the same risk, it might be said to be irrationally strong.[39] Some people's fear of nuclear power plants might be irrational in this sense.[40] Also,

if I come to believe that the risk is smaller than I thought, it would be irrational if my fear grew stronger. Second, one might appeal to the debilitating effects of such fears. I will examine that suggestion in a moment. Let me first note, however, that I can see no other grounds for claiming that emotions are disproportionate to their object. There is no inherently 'right' degree (or range) of anger, joy, envy, or grief that a given occasion could elicit.

Four, according to the second proposal, emotions are irrational if their intentional object is itself shaped or distorted by emotion. This is a special case of a more general idea, viz., that emotions are irrational if they are based on irrational beliefs about the situation that elicits them. These beliefs need not (as was assumed in that proposal) owe their irrationality to the motivating force of an emotion. There may be no emotion at all; or the force of the emotion may be independent of the motivational content; or the motivation may not have any emotional force.

In the first place, not all irrational beliefs have emotional causes. There is also 'cold' irrationality, due to defective cognitive processing of various kinds.[41] During the Second World War, Londoners were persuaded that the Germans systematically concentrated their bombing in certain parts of their 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,[42] probably made the Londoners living in those areas more fearful than they would otherwise have been.

In the second place, not all emotionally caused beliefs are 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 sums correctly because he is constantly distracted by thoughts about a woman he just met. Sometimes the addition errors might benefit the customer, sometimes himself: there is no reason to expect a pattern either way. By contrast, we can imagine a motivated irrationality that led him to obtain sums that systematically erred to his benefit.

In the third place, motivated irrationality can stem from desires that have no occurrent emotional backing. Suppose I have a calm, reflective desire (with zero arousal level) to go to a concert tomorrow, and that it causes me to believe, through wishful thinking, that I may get a

ticket. (In fact, evidence available to me indicates that tickets are invariably sold out long in advance.) The belief and the desire will then jointly induce the emotion of hope. And once there is hope, it may act to reinforce the belief and, in turn, itself. But the process which results in this unfounded emotion need not itself originate in an emotion.

Five, on the preceding proposal, an emotion can be tainted by the irrationality of the beliefs on which it is based. However, emotions are also based on desires. We may ask, therefore, whether emotions can be similarly tainted by irrational desires. Now, there is no agreement on what one might mean by an irrational desire. Many would deny that the idea makes any sense at all. Let us nevertheless consider the following two proposals.[43] On the one hand, a desire might be said to be irrational if it is heteronomous – if it is shaped, that is, by some non-autonomous psychic mechanism such as conformism, anti-conformism, adaptation, or counter-adaptation.[44] On the other hand, irrational desires might be those that lead to needless frustration or misery, a suggestion that is further discussed below. Sometimes, these suggestions lead in the same direction, as when a person constantly adjusts his desires so to want what he does not have. In other cases, they yield opposite conclusions. Unconscious adaptation of desires to what is possible is a non-autonomous mechanism making for greater happiness.

To illustrate the first proposal, consider an adolescent who always wants to do the opposite of what his parents want him to do. Having understood this mechanism, the parents tell him to do the opposite of what they want him to do. Later, on learning that he was led to do what they wanted him to do, he becomes angry because his desire was frustrated. That anger, we might say, is irrational, because it derives from the frustration of an irrational desire. Conversely, emotions that derive from the satisfaction of a non-autonomous desire might also be viewed as irrational. If my strongest desire is to be noticed by my favorite movie star, the joy I feel when my antics catch her attention in a crowd would, on this account, be an irrational one.

Six, a final suggestion is that rational emotions are emotions having which makes one happy, with the concomitant idea that irrational emotions are those that induce needless suffering. Regret and envy, for instance, might seem to be irrational in this sense. We teach our children not to cry over spilt milk or to envy the success of others, but instead to get on with their lives with no distractions by backward or sideward glances. We also teach them to be neither excessively risk-

taking and fearful nor excessively risk-loving and thrill-seeking: risk-averse people enjoy life less and risk-takers may not have much life to enjoy. Some parents try to inculcate a Polyannish attitude of seeing the good side of all things so never to suffer pangs of disappointment. Grief, repentance, remorse, shame, and guilt might also seem to be pointless sources of misery, although here the argument clearly becomes more strained.

This proposal differs from the preceding ones in that it applies to emotional dispositions rather than to occurrent emotions. In this respect, it also fits with the idea that rationality is a guide to action and not only an evaluative criterion. Compared to our ability to control our occurrent emotions, we can do more to change our emotional dispositions. That we *can* change them does not imply, however, that it is rational to do so. In the first place, it might be more efficient to avoid the occasions on which these emotions tend to occur than to eliminate their tendency to occur on these occasions. In the second place (reinforcing the first point but also valid independently), there are a number of costs associated with character planning that could make it more rational to leave the emotions as they are. There may be some monetary costs: paying a psychiatrist, buying books of self-help. Sometimes, the time and energy invested in changing oneself may be more usefully spent enjoying oneself. It makes no sense to spend life preparing oneself for death. Crying over spilt milk may, after all, be no worse than crying over one's tendency to cry over spilt milk.

Other issues are more complicated. One arises from the fact that there may exist couplings between the emotions that make it impossible to change a misery-inducing tendency without simultaneously getting rid of another, life-enhancing one. Is it really possible to enjoy hope without suffering disappointment if the hoped for event fails to materialize? To love another person without grieving if he or she dies? To feel pride at one's achievements without feeling guilt at one's failures? If there are such couplings (and I believe there are), the rationality of changing one's emotional dispositions would depend on whether one's stream of experiences is by and large positive or by and large negative. In a society with much poverty and illness, one might have more to gain by getting rid of the negative emotions than by losing the positive ones. It may not be an accident that Stoicism was associated with turbulent times and Buddhism with one of the poorest societies in the world. However, the matter is complicated by the fact that it is often

possible to shape *both* one's emotional propensities and the stream of experiences to which one is exposed. Instead of taking the first as a parameter and the second as a control variable (as in standard decision theory) or the second as parameter and the first as a control variable (as in theories of character planning[45]), the two ought to be chosen simultaneously, in one overall maximizing operation. But this very soon gets too complicated: we might as well take life as it comes.

The elimination of guilt, repentance, and remorse raises a different set of issues. Let us assume that such elimination is in fact possible, and that it would reduce the net suffering after all costs have been taken into account. One might still wonder whether it would not remove all substance and character from the person. In a recent article, Bittner has staged a useful confrontation between Spinoza and Williams on whether it is "reasonable to regret things one did".[46] Whereas Spinoza believed that it is possible and desirable to confront one's past wrongdoings without suffering nor loss of character, Williams believes that this ideal would incorporate "an insane concept of rationality".[47] I do not feel equipped to resolve this question, but I would like to connect it with the issue of rational desires that I raised earlier. Consider a dissident under a stable totalitarian regime, who suffers because his desire for freedom is frustrated. There are two ways in which he might get rid of his suffering: by ceasing to desire freedom or by no longer suffering even though his desire is frustrated. If we were to say that rationality dictates the first strategy, we would in fact embrace the totalitarian practice of branding dissidents as crazy. Neither Spinoza nor Williams would advocate this first strategy. They might disagree, however, with regard to the feasibility of the second.

The elimination of shame, as distinct from guilt, seems a more plausible desideratum. On the account suggested above of the difference between these two emotions, guilt is the more constructive. The guilty person wants to make amends, the shameful one to hide and disappear. The feeling of shame is so intensely painful that a person may go to great lengths to avoid it, including the attribution of blame to others for the bad outcomes of his actions. The more bearable feeling of guilt is less likely to induce such irrational substitutions. But before we conclude in favor of the elimination of shame and its replacement by guilt, we might ask whether it would be consistent with the ideal of personal substance and character.[48] Often, guilt and atonement are adequate responses. But some wrongdoings go to the core of a person,

and then perhaps only deep shame can lead to a transformation. The elimination of shame might yield a more shallow concept of personal responsibility.

Indirectly, shame is involved in a great deal of unnecessary suffering because of its importance in sustaining social norms. I believe that the relentless pressures for conformity and the cruel sanctions meted out to non-conformists are prominent sources of needless misery. When children come home from school crying because classmates have detected some minor deviation of dress or speech, or Chinese women were tortured by foot-binding, the misery is in one sense wholly avoidable. If the relevant norms had not existed, sustained by feelings of shame, nobody would have been worse off and some would have been very much better off. Yet this fact does not allow us to infer that any given individual would be better off if he managed not to feel ashamed of not participating in some absurd practice. He would still be the target of the sanctions by others not similarly liberated.

Let me try to pull these ideas together. An idea of rational emotions may be constructed in several ways. We may say: emotions are a variety of X, and we know what a rational X is. Or: emotions are based on X and Y. We know what a rational X and a rational Y are. Therefore, emotions are rational if they have a rational basis. These are derived notions. Non-derived notions are more interesting. I have canvassed three such ideas: rationality as happiness-inducing, rationality as consistency, and rationality as appropriateness. Of these, the first, although valuable, is incomplete: happiness is not what life is all about. The second captures a part, but not a very large one, of our intuitions. The third seems to me to be the most interesting. It is based on the idea that *emotions can interfere with themselves* and, in doing so, become incongruent with the occasion that gave rise to them. To be angry need not be irrational, but anger transmuted by its own momentum into righteous indignation is.

I now turn, much more briefly, to the second question I raised above: What meaning if any can we attach to the idea that a desire is rational, given certain emotions? For instance, if I am jealous of Henry, is it rational for me to want to hurt him? Is it rational for me to want him look inferior in Patricia's eyes?

I begin with the second question. If an emotion is caused by the frustration of a desire, I think it fits well with intuition to say that it is rational to desire the removal of the obstacle that is in my way. If I

have reason to believe that Patricia will turn to me once Henry has been made to look a fool, it is rational for me to want him to look a fool. However, that conclusion also follows on what I have called the aseptic or cognitive notion of emotion. Suppose I fear that my house will be burgled some time over the next few years, unless I do something about it. That fear – the belief that there may be a burglary and my desire for there not to be a burglary – makes it rational for me to desire to do something about it, e.g., by getting a burglar alarm.[49] To explain this connection, there is no need to bring in actual feelings, that is, a state of fear as arousal. Similarly, even if there is arousal (as there may be in the jealousy example), there is no need to invoke it to rationalize the desire. What we are dealing with in such cases are really desires that are connected to each other as means to an end rather than desires originating in emotions.

In fact, arousal can induce behavior that is instrumentally irrational with regard to the desire that is being frustrated. If I want to hurt Henry and in fact do hurt him, that might easily turn Patricia against me. Miss Bingley's behavior toward the end of *Pride and Prejudice* fits this pattern. Being jealous of Elizabeth Bennet, she talks unfavorably of her to Darcy: "Persuaded as Miss Bingley was that Darcy admired Elizabeth, this was not the best method of recommending herself; but angry people are not always wise". More generally, the action tendency that is associated with emotion in the full-blooded (aroused) sense need not be instrumentally irrational in promoting the desire whose frustration or satisfaction was at the origin of the feeling. When I am hurt, there is a natural tendency to lash out at the offending object. But what is natural need not be rational. If I stumble over a stone, kicking back does me no good. If I am ashamed, there may be a natural desire to hide myself from the view of others, an action that may lead them to think even less of me. Fear can freeze me to the spot when I ought to run or vice versa.

A natural response is that if these tendencies are 'natural', they must serve some purpose by virtue of which they have been selected by natural evolution. I am very skeptical of functionalist analyses of the emotions, whether based on their socially useful consequences or their fitness-enhancing effects, but it would take me too far to go into this question here. I shall only warn against confusing the objective notion of fitness with the subjective notion of rationality. Even if, say, Miss Bingley's behavior could be explained by an analysis of the circum-

stances in which early hominids frequently found themselves, it would not make it an iota more rational. Given her desires and her beliefs, her badmouthing of Elizabeth Bennet is irrational; she just cannot help herself.

If we go back to Figure 2, therefore, we can now interpret the two-way arrow between desires and emotions as follows. On the one hand, emotions may be rationalized by desires in the sense of being appropriate reactions to the frustration or the satisfaction of a desire. On the other hand, a desire can be rationalized by an emotion if it is directed toward maintaining conditions that satisfy the desires that elicited the emotion or removing conditions that frustrate them. However, as we have seen, to the extent that an emotion (qua arousal) is causally involved in the production of desires, there is in general no reason to expect them to be rational in the latter sense.

NOTES

[1] Hirschman (1977).
[2] La Bruyère, *Les Caractères*, Chap. IV.77.
[3] Elster (1993, Chap. 3).
[4] Billacois (1986, p. 239).
[5] For fuller discussions, see Elster (1983, Chap. I; 1985; 1986; 1989a, Chap. I).
[6] These aims include the material interests of the agent, but are not restricted to them. Although La Bruyère does not spell out what the category of *reason* is intended to include, we may conjecture that it covers any kind of impartial concern, whether consequentialist or non-consequentialist. In that case, 'rationality' as understood here and 'reason' as understood by him (as understood by me) will overlap, but neither category includes the other.
[7] Pears (1985).
[8] Elster (1983, Chap. III).
[9] For fuller expositions, see Elster (1989b, Chap. 3; 1989c).
[10] Ainslie (1992).
[11] Elster (1990).
[12] Useful surveys of the emotions are Frijda (1986) and Lazarus (1991). My own work in this area is limited to a study of envy (Elster, 1991).
[13] Most writers also cite the fact that emotions have "action tendencies" – an impulse to run, fight, hide, confess, or destroy (the reader can supply the emotions!). With some hesitation I exclude this feature from the list of features invariably or usually associated with the emotions, partly because I think there are too many emotions that have no action tendencies associated with them, such as sadness or grief, and partly because I think that the tendencies, where they exist, are more properly seen as part of the phenomenological feel of the emotion (Frijda, 1986, pp. 237–39).
[14] Hume (1960, p. 277).

[15] For fuller expositions, I again refer to Elster (1989b, Chap. 3; 1989c).
[16] Edgerton (1985, p. 3).
[17] Zajac (1985, p. 120).
[18] Akerlof (1976); see also Abreu (1988).
[19] Kurtz (1979, p. 215).
[20] Lazarus (1991, p. 24).
[21] Tangney (1990); Lewis (1992).
[22] Lewis (1992, p. 112).
[23] Ekman (1980, p. 87).
[24] Levy (1973).
[25] According to Tangney (1990, p. 102), this view "has been soundly criticized from many quarters" and "has been largely abandoned in favor of a more complex conceptualization".
[26] Hochschild (1979, p. 567).
[27] By contrast, I believe the second mother *would* feel guilty about her lack of guilt.
[28] *Essais*, Bk. I.20.
[29] Lazarus (1991, Chaps. 6, 7).
[30] For a similar distinction, see de Sousa (1987, p. 6): "[O]ne role of the emotions is to ground assessments of rationality regardless of whether the emotions are themselves rational".
[31] See also Elster (1985).
[32] Solomon (1980, pp. 251–52: this is a reprint with a new appendix of an essay first published in 1973). Other advocates of this view include Sartre (1936) and Schafer (1976).
[33] Solomon (1980, p. 263: italics in original).
[34] Ibid., p. 265. There is some waffling here. Although the passage says that emotions are "often" short-sighted, he offers no evidence that they may not always be. Some evidence of that kind would be needed, however, to support his claim that "[e]motions are not irrational; people are irrational" (ibid.).
[35] The psychoanalytical idea that apparently inappropriate emotional reactions can be traced back to childhood experiences that render them more intelligible (Fenichel, 1945, p. 20; Lazarus, 1991, pp. 165ff.) is a more general version of this observation.
[36] See Kahneman and Tversky (1982) for a seminal study of the conditions under which counterfactual scenarios contribute to our emotional reaction to actual events.
[37] But see Williams (1981, Chap. 1) for a different view.
[38] Broad (1971, p. 293).
[39] Again (see note 35 above), psychoanalytic theory would explain such idiosyncratic emotions by invoking the force of early experiences in shaping current reactions.
[40] Since consistency cuts both ways, however, one might also conclude that their fear of other, comparable risks is irrationally weak.
[41] For a survey, see Nisbett and Ross (1980).
[42] Feller (1968, p. 161).
[43] Elster (1983, Chap. I).
[44] Elster (1983, Chap. III) has a survey of these mechanisms.
[45] See notably Kolm (1982).
[46] Bittner (1992). His notion of regret includes repentance, guilt, and remorse, as well as non-moral phenomena such as regretting that one did not buy IBM shares when they

were at their all-time low. However, contrary to what he asserts, his argument does not really apply to the non-moral emotions.

[47] Williams (1981, p. 29).

[48] I owe this point to an incisive question by Audun Øfsti.

[49] Gordon (1987, pp. 73–79).

REFERENCES

Abreu, D.: 1988, 'On the Theory of Infinitely Repeated Games with Discounting', *Econometrica* **56**, 383–96.

Ainslie, G.: 1992, *Picoeconomics*, Cambridge University Press, Cambridge.

Akerlof, G.: 1976, 'A Theory of Social Custom, of which Unemployment may be one of the Consequences', *Quarterly Journal of Economics* **90**, 599–617.

Billacois, F.: 1986, *Le duel dans la société française des XVIe-XVIIe siècles*, Éditions de l'École des Hautes Études en Sciences Sociales, Paris.

Bittner, R.: 1992, 'Is it Reasonable to Regret Things One Did?' *Journal of Philosophy* **LXXXIX**, 262–73.

Broad, C. D.: 1971, 'Emotion and Sentiment', in D. C. Cheney (ed.), *Broad's Critical Essays in Moral Theory*, Allen and Unwin, London, pp. 283–301.

Edgerton, R.: 1985, *Rules, Exception, and the Social Order*, University of California Press, Berkeley.

Ekman, P.: 1980, 'Biological and Cultural Contributions to Body and Facial Movement in the Expression of the Emotions', in A. Rorty (ed.), *Explaining the Emotions*, University of California Press, Berkeley, pp. 73–102.

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

Elster, J.: 1985, 'The Nature and Scope of Rational-Choice Explanation', in E. LePore and B. McLaughlin (eds.), *Action and Events: Perspectives on the Philosophy of Donald Davidson*, Blackwell, Oxford, pp. 60–72.

Elster, J.: 1986, 'Introduction', in J. Elster (ed.), *Rational Choice*, Blackwell, Oxford, pp. 1–33.

Elster, J.: 1989a, *Solomonic Judgements*, Cambridge University Press, Cambridge.

Elster, J.: 1989b, *The Cement of Society*, Cambridge University Press, Cambridge.

Elster, J.: 1989c, 'Social Norms and Economic Theory', *Journal of Economic Perspectives* **3**, 99–117.

Elster, J.: 1990, 'Norms of Revenge', *Ethics* **100**, 862–85.

Elster, J.: 1991, 'Envy in Social Life', in R. Zeckhauser (ed.), *Strategy and Choice*, MIT Press, Cambridge, Massachusetts, pp. 49–82.

Elster, J.: 1993, *Political Psychology*, Cambridge University Press, Cambridge.

Feller, W.: 1968, *An Introduction to Probability Theory and its Applications*, 3rd ed., Vol. 1, John Wiley, New York.

Fenichel, O.: 1945, *The Psychoanalytical Theory of Neurosis*, Norton, New York.

Frijda, N.: 1986, *The Emotions*, Cambridge University Press, Cambridge.

Gordon, R. M.: 1987, *The Structure of Emotions*, Cambridge University Press, Cambridge.

Hirschman, A.: 1977, *The Passions and the Interests*, Princeton University Press, Princeton.

Hochschild, A.: 1979, 'Emotion Work, Feeling Rules, and Social Structure', *American Journal of Sociology* **85**, 551–75.
Hume, D.: 1960, *A Treatise on Human Nature*, L. A. Selby-Bigge, ed., Oxford University Press, Oxford.
Kahneman, D. and A. Tversky: 1982, 'The Simulation Heuristics', in D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment Under Uncertainty*, Cambridge University Press, Cambridge, pp. 201–09.
Kolm, S.-C.: 1982, *Le bonheur-liberté*, Presses Universitaires de France, Paris.
Kurtz, E.: 1979, *Not-God: A History of Alcoholics Anonymous*, Hazelden Educational Services, Center City, Minnesota.
Lazarus, R. S.: 1991, *Emotion and Adaptation*, Oxford University Press, Oxford.
Levy, R.: 1973, *The Tahitians*, University of Chicago Press, Chicago.
Lewis, M.: 1992, *Shame*, Free Press, New York.
Nisbett, R. and L. Ross: 1980, *Human Inference*, Prentice Hall, Englewood Cliffs.
Pears, D.: 1985, *Motivated Irrationality*, Oxford University Press, Oxford.
Sartre, J.-P.: 1936, *Esquisse d'une théorie des émotions*, Hermann, Paris.
Schafer, R.: 1976, *A New Language for Psychoanalysis*, Yale University Press, New Haven.
Solomon, R.: 1980, 'Emotions and Choice', in A. Rorty (ed.), *Explaining the Emotions*, University of California Press, Berkeley, pp. 251–82.
Sousa, R. de: 1987, *The Rationality of Emotion*, MIT Press, Cambridge, Massachusetts.
Tangney, J. P.: 1990, 'Assessing Individual Differences in Proneness to Shame and Guilt: Development of the Self-Conscious Affect and Attribution Inventory', *Journal of Personality and Social Psychology* **59**, 102–11.
Williams, B.: 1981, *Moral Luck*, Cambridge University Press, Cambridge.
Zajac, E. P.: 1985, 'Perceived Economic Justice', in H. P. Young (ed.), *Cost-Allocation*, North-Holland, Amsterdam, pp. 119–53.

Department of Political Science
University of Chicago
Chicago, IL 60637
U.S.A.