BOOK REVIEWS

Alchemies of the Mind: Rationality and the Emotions. JON ELSTER. New York: Cambridge University Press, 1999. xi + 450 p. Cloth $75.00, paper $19.95.

Alchemies of the Mind is devoted to uncovering and exploring the dynamic structures of emotional life. On a familiar conception, an emotion is an attitude towards an object or state of affairs, laden perhaps with affect. Jon Elster argues that such a view is atomic, static, and simplistic. Emotional episodes usually have structures that are not atomic but complex in their contents, and are not static but dynamic and evolving. He proposes that "emotional life may be a succession of episodes, each of which has an internal structure, rather than a simple succession of experiences" (59). This idea permeates the book. The result is a rich and nuanced work that reveals many complexities of the human mind.

Elster investigates such intriguing psychological states as shame, envy, malice, indignation, pity, schadenfreude, contempt, pridefulness, and romantic love. There are a few reasons why Elster concentrates on these complex emotions. Such emotions have the "greatest explanatory relevance" for social and political analysis (61). And, though science can help us to understand simpler emotions (anger, fear, disgust, parental love), it has not shed much light on the more deeply human emotions (48). This is why "prescientific insights into the emotions are not simply superseded by modern psychology in the way natural philosophy has been superseded by physics" (50). Interestingly, Elster finds La Rochefoucauld's views about unconscious thought and motivation "probably more valuable than anything found in twentieth-century psychology" (84). Any successful analysis will require a great deal of subtlety—perhaps more than can be controlled for or created in a scientific setting—as well as intuitive insight into the human heart and mind. So Elster goes where the insights seem greatest, and where they intersect with his own interests, which happily are broad ranging and deep. He finds his sources of insight into the emotions in literature, proverbs, and historical texts. His case studies range over humiliation in ancient Greece, the exquisite subtleties of seventeenth-century French society, the envy-driven practices of nineteenth-century America, and vengeance killings in medieval Iceland and nineteenth-century Corsica. He discusses Aristotle,

0022-362X/04/0109/484–91

Shakespeare, Jane Austen, Stendhal, and Tocqueville. He draws on the French moralists—Montaigne, Pascal, La Rochefoucauld, and La Bruyère—for insight into how emotions cause other emotions, affect judgment, and induce what he calls "transmutation of motivations" (76). Most of the book is concerned with such internal changes, the "alchemies" of the mind.

Once the phenomena have been identified and illustrated, Elster investigates the *mechanisms* behind them. These are "frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences" (1, italics omitted). An explanation by mechanism takes the form "A sometimes causes B." Elster points out that if this were all that could be said for them, mechanisms would be "near-useless" (10). But mechanisms identify particular causal patterns that make them explanatorily useful. They offer more explanatory power than do appeals to statistical regularities, which can mask distinct causal mechanisms (44ff.). An example of a mechanism is wishful thinking: sometimes wanting x causes believing that x is true. Though this is hardly a lawlike phenomenon—more often than not, wanting x does not result in believing x is so—still, it is a phenomenon we understand. The same tension between the belief that not-x and the desire that x can *also* produce the different and incompatible effect known as adaptive preference formation, whereby a desire is changed to match a belief. We cannot predict which of the two mechanisms, if either, will be triggered. But sometimes certain factors might nudge us toward a prediction. Consider the absent lover. Will "absence make the heart grow fonder," or will it be "out of sight, out of mind"? Elster approves of La Rochefoucauld's resolution: "Absence lessens moderate passions and intensifies great ones, as the wind blows out a candle but fans up a fire" (40). In addition to the difficulty of predicting which of two opposing mechanisms will become active, there are other ways that the net effect of the mechanisms is indeterminate. A single cause might produce two determinate but opposite effects. Or a cause might trigger an effect that then itself triggers an opposite effect.

To some it will seem intuitively true that such mechanisms describe standard causal patterns in human psychology. They seem good inputs for further generalizing, for deeper and broader explanations of how the mind works. But others will be dismayed. They worry that "mechanisms" are no better than made-up stories. Elster is sensitive to this worry: "Stories are ad hoc and arbitrary; mechanisms are not" (15). He does not explicitly say why not. His book as a whole, with a great many examples analyzed and interpreted, constitutes the answer. The internal alchemies he identifies are not haphazard. There is for the

most part an engine that drives them all, namely the human need to act and feel for good reason, and to see oneself as acting and feeling for good reason. But one wants more justification of the methodology. Is it justification enough that we cannot deny that these are patterns in human psychology? Are they the sort of patterns that psychologists accept as phenomena and then try to explain?

Elster does not think that appeal to mechanisms is ideal. It is essential to understanding psychology only because there is often no explanation by laws to be had; it would be better if such laws could be found. It is not completely clear whether he thinks the need for mechanisms is a feature of our ways of knowing or of the world. At least some appeals to mechanisms seem to be the result of not knowing enough about the world (cf. 36). But he also seems to leave open the possibility that mechanisms may turn out to be the best description of our basic psychology. The indeterminacy built into them may be just a fact, rather than a product of our ignorance. In any case, we crave understanding of the emotions and their effects. Our best method of learning as much as we can, says Elster, is to follow the path of mechanisms.

His discussion of the nature of the emotions opens with the following caveat: "I prefer to leave the concept [of emotion] open-ended and ambiguous, in the hope that at some future time we may come to understand it better." Because there may not be a common causal mechanism, it might "turn out that the unruly category of 'the emotions' encompasses several, internally homogeneous classes of phenomena" (241). In the meantime, Elster adopts an ecumenical view. He prefers to admit many states into the category of the emotions: admiration, liking, aesthetic awe, aesthetic surprise, being "needy of honor" (which produces "an anguish of incompleteness" (228)), hybris (by which he means the pleasurable humiliation of others (214)), and in general any state that seems to be "emotive," as well as some that are not. He provides a valuable discussion of the deficiencies inherent in a cost/benefit modeling of the emotions (155ff., 301ff.). Emotions are portrayed as tightly bound vehicles of rationalizations: examples abound in which an "emotion works backward, to invent its own justification" (130) by causing the very beliefs that provide justification for the emotion. In general, Elster relies on a phenomenological, or at least a pretheoretical, intuitive understanding of emotions (243). He does not provide a general theory of the emotions, nor is doing so his aim. He avoids constructing fully general theories because of the tremendous diversity of his subject matter. He is more interested in the consequences of the emotions, in their interrelation with reason and rationality, and in the ways the emotions generate

other mental states, especially other emotions. The lack of a tightly constructed theory leaves Elster with much room to indulge us with a great many distinct observations, speculations, and arguments. Sometimes we are presented with the interesting just because it is so. Yet he is probably right that what he calls the "hodge-podge" nature of his taxonomizing is truer to the complexity of the facts than "a more parsimonious and coherent account" would be (329). All this leaves much to chew on, perhaps disagree with, in the nitty-gritty.

Consider his discussion of hatred in a section on Aristotle. "In hatred, my hostility is directed toward another person or a category of individuals who are seen as intrinsically and irremediably bad. For the world to be made whole, they have to disappear" (65). Emphasizing that he is talking about an occurrent emotion and not a disposition, Elster argues, against Aristotle, that hatred is painful. But he agrees with Aristotle's claim that hatred is not as great an impediment to instrumental rationality as anger is. No argument is given for this view, though of course it is a truism that hatred can be cold and even "unemotional." It seems counterintuitive, however, to say that hatred essentially permits rational behavior in a way that anger does not. Consider gay bashing in the military. This is not Elster's example, but the motivation behind that behavior seems to satisfy his definition of hatred. Surely such hatred can "cloud the mind as anger and fear do"? Perhaps he—and Aristotle—would respond that this case involves an extreme kind of anger, and not hatred proper. But then one wants to know more about what is at stake in this disagreement, and how we might adjudicate it.

In Elster's view, Aristotle reveals many truths about the emotions. And foremost among these is that emotions are shaped by social interaction. Elster provides an extended discussion of social emotional phenomena, especially shame, envy, and the "cluster of emotions related to the pursuit and defense of honor" (139). This supports the general thrust of his work: that we are essentially social creatures, caring more about our relative standing in our social relations than we do about our absolute well-being. Moreover, Elster wants to explain the mechanisms that sustain and shape social norms. He argues provocatively that "the emotion of shame is not only a support of social norms, but *the* support" (145). He tempers this with the observation that norm-guided behavior might be upheld by a passion for the norm that is independent of fearing shame. Wrath, for example, may sustain practices of vengeance (155).

Elster's view is that social norms do not operate merely as a system of material sanctions. Sanctions work not because of the costs imposed on the norm violator, but because of the costs incurred by the sanc-

tioner. "The more it costs me to refuse to deal with you, the stronger you will feel the contempt behind my refusal and the more acute will be your shame.... [T]he costs to the sanctioner are what makes the sanction really painful to the target. It tells him that others see him as so bad that they are willing to forego valuable opportunities rather than to have to deal with him" (146). So sanctions work because of contempt and of shame. This is compelling and surely illuminates not a few cases of norm enforcement. But I wonder how central it is: it seems possible to sanction without shaming, just as one can shame without incurring costs. Investigating the extent of this is an interesting empirical question. Still one cannot deny that "the anticipation of shame acts as a powerful regulator of behavior" (154). It is, as Elster stresses, extremely difficult to endure the contempt of others.

One reason Elster thinks shame is so effective at norm-guidance is that avoidance of shame, unlike avoidance of guilt, "cannot take the easy option of self-deception, it has to use the hard option of behavior modification" (154). But taken at face value, it cannot be true in general that one cannot avoid shame by being self-deceived. Self-deception about whether properties one finds shameful apply to one seems possible. Take, for example, being a gossip, a bore, having bad posture, or being cheap and stingy. "It is hard to deny behavior," Elster says (154). But this misses a whole stock of characters who think they are charming but are boring, and so on. In addition, it seems possible for such self-deception to affect one's ability to see how others perceive one. It is true that once one experiences shame—once shaming is successful—one may not be able to hide what one takes to be the shameful property from oneself. Shaming is in this sense transparent, and such cases are paradigms of social control. Elster's account of the shame-based control of social behavior requires that the shame "is triggered by the contemptuous or disgusted disapproval by others of something one has done" (149). But Elster also allows that shame can be "induced by something that is done to one or by entirely unrelated events" (150), and that one can feel shame when there is no disapproval present or perceived (though he calls this shame "irrational"). It is not clear how these forms of shame play a role in maintaining social norms.

The most interesting mechanisms, the centerpiece of the book, are those whereby a motivation is transformed, either through (a) the unconscious process of transmutation, highly motivated by the need for self-esteem, or through (b) the conscious misrepresentation of that motivation, motivated by the need for the esteem of others. There is much grey area between the two (336, 374), and self-deception hovers uncomfortably between, and can sometimes morph into, the

two (335). Knowledge of one's motivations can be so undesirable that there are internal and external pressures to misrepresent them, to oneself and to others. Similarly, transmutations are driven by the need to avoid painful meta-emotions (feelings of shame, guilt, or anger whose objects are other emotions) and to experience something better in its place. So the pure envy of another's good fortune can turn to righteous indignation that it was got by immoral means (98, 316). The envy then remains unconscious, but one also gains a virtuous feeling. In another frequently discussed example, a shameful love can turn into hate.

Elster identifies *amour-propre*, the desire for esteem and self-esteem, as the central motivating force behind the formation of many emotions, both pleasant and unpleasant. He writes of its "corrosive and transmuting effects" (77)—corrosive because it can pull us away from doing or seeking what is in our best interest (cf. 90ff.). Central to *amour-propre* is the belief that we act for reasons. "It is an important part of our self-image that we believe ourselves and want ourselves to be swayed by reason rather than by passion or interest" (91). Indeed, it is because the heart must pretend to act by reason (that is, impartially) that "the head is always fooled by the heart" (91). Elster aims to show that many irrational emotional states are in fact motivated in this way (286). Take the following: hatred of those whom we have unjustly injured; anger at those who have helped us; anger at those whom we bore, or who have proved us to be wrong; anger at those who do not return our love. In each of these cases, Elster argues that the hostile emotion is caused by a mechanism involving damage to one's self-esteem. Regarding those who irrationally hate those whom they have harmed, he says, "admitting that they behaved badly is intolerable to their self-esteem. Instead, they engage in fault finding, so that they can say about the other, to themselves and to third parties, 'He only got what he deserved'" (106).

While one might quibble with some of his examples and detailed taxonomy of types of emotional transmutations, it is hard to deny the general phenomenon. And it is hard to deny the importance of explaining transmutations if one hopes to understand the mind. Yet, as Elster argues, we do not have such an explanation; transmutations are far from understood.

> *Exactly how* does it happen that people fool themselves into thinking that they do what they do for other motivations than those which really animate them? [...] What is needed is a theory or a mechanism-generating framework that can explain the role of the various motivations as inputs, engines, and outputs of transmutation. *But there is no such theory or*

> *framework*. Neither of the two dominant theories of motivated attitudinal change, Festinger's theory of cognitive dissonance and Freud's theory of defense mechanisms, addresses the issues I have identified, although each of them touches on closely related questions (362–63).

Elster's main complaint against cognitive dissonance theory is that it does not adequately account for the process whereby dissonance is reduced. In particular, it does not recognize that emotions such as shame can provide the motive for such reduction. He urges those who would develop cognitive-dissonance theories to recognize that emotions themselves can induce dissonance reduction (365–66). But he does not say just what the status of the emotions would be in such a theory. For example, it is not clear whether the emotions in question might motivate in virtue of their negative affect, or in virtue of encapsulating a set of beliefs that are themselves involved in the cognitive dissonance.

Problems with the Freudian theory of defense mechanisms are not as easily remedied, according to Elster. The theory does not address emotional transmutations, and it is not clear how it could do so. Another charge is that of "taxonomic chaos": there is little agreement on the types of defense mechanisms. He adds to this the charge that defense mechanisms cannot be directly observed, while emotions can be directly experienced from a first person perspective. Thus we have much stronger reasons to believe in the existence of the emotions than we do to believe in defense mechanisms. "We may be unsure about how to classify them and how to distinguish them from one another, but the reality of, say, anger and fear could not possibly be denied" (368). This is of course right. But even if we grant that emotions are evident in a way that defense mechanisms are not, it seems to me that we must be struck by the fact that emotional transmutations are, in this respect, in no better a position than defense mechanisms. For while one may observe that one emotion is followed by another, transmutations are not themselves directly observable, even from the first person perspective. Even so, there are crucial differences between Elster's unconscious mechanisms and the Freudian unconscious. The notion of transmutation is potentially more open to interesting psychological theory that could someday be tested. It is also not as deeply theoretical; it draws upon our common-sense psychology. The question is which of these theories are more likely to be fruitful, to lead to an empirical research program. That transmutation will do so seems plausible.

A point that perhaps gets to the heart of Elster's criticism of Freud,

and to a central feature of his own methodological approach, is this. Freud's defense mechanisms, he says, have "no intuitive plausibility" (368). It seems implausible that defense mechanisms, or the Freudian notion of projection, for that matter, are brute mechanisms. Elster wants a rationalistic explanation. He wants it to *make sense* how and why a mechanism such as projection provides relief for the subject. Explanations via mechanisms must make their subject matter intelligible. I think this is part of the reason Elster urges that we look to plays and novels "as the closest thing to a controlled experiment involving high-stakes human emotions" (108). He suggests that we refer to fictional characters as if they were real, "as if, that is, they could provide illustrations or counterexamples for this or that abstract proposition" (125). This suggests that the standards by which we are to judge the contributions of literary and historical sources are the standards of intelligibility. As he puts it, we are to see if the experiments explored through literary characters *ring true* (109).

Elster challenges much of the methodology of current work on the emotions (the hasty inferences sometimes drawn from neurological studies, for example). He also challenges much work that is not overtly about the emotions but should be, such as rational choice theory and economics. He lays a large array of cases before us. He is unafraid of the difficult cases, the complex cases, and the highly puzzling ones; we have a feast of phenomena to consider. Elster is attracted by complexity, and compelled by its being subtle and elaborate. Two or more mechanisms often yield more than the sum of understanding provided by either one alone. Add to this complexity the essential indeterminacy at the heart of all mechanisms, and one faces a daunting theoretical challenge. My sense is that Elster would not have it any other way. Indeed, part of the value of his work is that he continually—and rather enthusiastically—reminds us how little we really know about psychology and about our emotional lives in particular. The facts are both fascinating and genuinely mysterious. *Alchemies of the Mind* provides a map of fertile areas for further inquiry, replete with novel and deep insights. Not only do we learn that the emotions are far more complex than we might have thought, but we also learn how diverse are their "large-scale, systematic effects" (407) on virtually all aspects of human life.

STEPHANIE BEARDMAN

Barnard College