Emotional Alchemy


Jon Elster needs no introduction to the readers of this journal. He is an outstanding social scientist who has had a great impact on the work of many in the field. He has done pioneering work in rational choice theory and social and political theory. Moreover, Elster is one of the most prolific writers in the intersection of philosophy and the social sciences. In the previous decade alone he has written five monographs, edited eight volumes and published over 70 articles on various topics. Elster’s numerous contributions often are original and inspiring. However, given the sheer volume of his publications repetitions are inevitable.

*Alchemies of the Mind*, Elster’s contribution to the theory of emotions falls in this category. Those already familiar with Elster’s work will find that some of the themes and observations in this book are more or less the same as those that he treated in the seventies and eighties in such landmark monographs as *Explaining Technological Change*, *Ulysses and the Sirens*, and *Sour Grapes*. This does not mean that there is nothing new or worthwhile in this book. On the contrary, there are many interesting points and passages. I will discuss some of them below.

The book consists of five chapters and a coda in which Elster sums up the lessons that can be learned from *Alchemies of the Mind*. Chapter one, entitled “A Plea for Mechanisms”, restates the position that Elster advocated in *Nuts and Bolts for the Social Sciences* (with some slight modifications) and elsewhere. His claim is that explanation in the social sciences should not be conducted by quoting law-like statements that subsume the explanandum under a general principle of the form “if A than always B”. There are two related reasons for this. First, there are no principles of this form with a sufficient degree of generality that qualify as a real explanation (p. 2). Secondly, many of the proposed “laws” in the social sciences are statistical correlations. These correlations cannot predict or explain in individual cases. Consider the following example that Elster borrows from Carl Hempel (p. 44):

(1a) If the barometer is falling, it almost certainly will rain.
(2a) The barometer is falling.
(3a) It almost certainly will rain
(1b) If the sky is red at night, it almost certainly will not rain.
(2b) The sky is red tonight.
(3b) It almost certainly will not rain

All four premises may be true in a particular case, but the conclusions cannot both be true.
Because of these reasons, Elster is sceptical that there could ever be acceptable law-like explanations in the social sciences. Instead, he opts for explanations using mechanisms. Mechanisms are low-level “...causal patterns that are triggered under generally unknown conditions or with indeterminate consequences.” (p. 1). In individual cases, one of several, mutually exclusive, mechanisms may be triggered by the initial conditions. For example, a loosing streak in a game of roulette can cause some people to increase their bets because they believe that the chance of winning the next game goes up (the so-called gambler’s fallacy). However, one can also observe the reverse effect. A loosing streak can cause people to reduce their bets (p. 7). These two incompatible tendencies are triggered by the same initial conditions, thus making the result unpredictable. Alternatively, the initial conditions might trigger several non-exclusive mechanisms, which may have opposite effects. An example is the following. A high marginal tax rate lowers the opportunity cost of leisure. Therefore, it creates an incentive to consume more leisure and work less. On the other hand, a high tax rate lowers one’s disposable income, thus creating an incentive to work more to generate more income. Which of these two incentives will be the greater depends on the strength of both mechanisms (p. 7). Elster believes that many, if not all, social and emotional phenomena are determined by clusters of such mechanisms. Hence, it is useless to look for broad, general explanations. Instead, we should aim at identifying the mechanisms that determine the explanandum. This sets a modest but feasible research agenda for the social sciences that merits further thought if not emulation. Note that one implication of this quest for mechanisms rather than laws is that one gives up on the idea that social science can generate predictions. All it can do is explain after the fact.

The scepticism about prediction sets the stage for the remainder of the book. According to Elster, emotions are both the result of mechanisms and can trigger mechanisms. Emotions are discernible, often occurring causal patterns that are indeterminate in their triggering conditions or their consequences. Given this indetermination, it is at best extremely difficult to establish the existence of such emotional mechanisms in laboratory experiments let alone in real life. Therefore, Elster suggests, we should look not just at the work of psychologists but also (perhaps more so) at other sources of knowledge of human nature. In particular, we should take to heart the observations of philosophers and literary authors.

Chapters two and three are entitled “Emotions Before Psychology” and, “Social Emotions in Historic Context” respectively. These chapters form the bulk of the book and they are an application of the mechanism approach. Elster discusses in detail the opinions and observations of many authors, among whom are Aristotle, Montaigne, Pascal, La Rochefoucauld and La Bruyère, as well as Shakespeare, Racine, Mme. De Lafayette, Jane Austen, Stendhal and George Elliott. In these chapters, Elster illustrates the many subtle connections between the emotions, actions and beliefs.
Chapter three deals with the social emotions of shame, envy and honor. Elster pays a lot of attention to the historical context in which these authors made their observations in order to infer the relative historical importance of these emotions.

The example of envy is particularly interesting in this respect. It demonstrates both the many subtle and sometimes irrational effects of the emotions as well as the effects of a particular social and cultural environment on the emotions. For example, suppose you have something that I desire. This might cause me to envy you – a belief together with a relevant desire triggers an emotion. However, this emotion itself might trigger another emotion. In contemporary society, there are strong social norms against such envy. Consequently, I might be ashamed about my envy. This second-order emotion of shame can be very painful. Therefore, it might cause some changes in my beliefs. I may come to believe that you do not really deserve that enviable object. An emotion affects a belief. This then might lead me to be righteously angry with you, which is not shameful at all. My beliefs have changed my original emotion into one that attracts less shame.

However, under different social and cultural circumstances very different effects might be observed. According to Elster, the ancient Greeks were in general not ashamed of envious emotions. Aristotle, for example, only discusses the effect of envy on action. He seems unaware that envy might trigger other emotions. According to Aristotle, the envious person will try to destroy the object that he envies. Elster adds that such a person might also revise her beliefs (e.g., coming to believe that the desired object really is an inferior product), thus reducing the pain of envy. The emotional dialectic of envy and shame is typical for our modern society where there are strong norms against envy.

Chapter four, “Rationality and the Emotions”, deals with the relation between rationality and the emotions. It starts out with two sections on the nature of the emotions, which draws from the observations made in chapters two and three. Elster argues, quite convincingly as far as I can judge, that there is no plausible theory that covers every emotion. The phenomena we label as ‘emotions’ share resemblances of analogy, rather than resemblances of kind (p. 239ff). Elster attempts to list some general characteristics of occurrent emotions (as opposed to dispositional emotions, such as irascibility) but acknowledges in the same pages that these characteristics do not hold for all emotions, especially aesthetic emotions. In doing so, he debunks several prominent theories of the emotions. For example, the appraisal theory of emotions claims that all emotions are triggered by a cognitive appraisal of the situation (p. 245, 269). However, Elster observes that this theory is at a loss to explain the joy one experiences when listening to a joyful piece of music. The emotional response to such music is instantaneous and unmediated by a belief that this is a joyous piece of music. In fact, if there is such a belief it usually follows from the emotion, not the other way around.
At one point in this chapter, Elster considers the claim that emotions are rational in the sense that they can contribute to making a rational choice in cases of indeterminacy. Ronald de Sousa in his *Rationality and the Emotions* has advocated this view. The idea is that emotions are “gut-feelings” that make us go for one of two or more indeterminate options. Emotions assist rationality in situations where rationality alone would result in indeterminacy.

Elster is sceptical of this claim. He argues that indeterminacy of rational choice is the result of either indifference or incommensurability of the options. This seems incomplete to me, as I will explain below. Elster endorses a test that Joseph Raz and others have proposed for establishing whether two options are indifferent or incommensurable (p. 288). Suppose an agent cannot make up her mind between A and B. She is indifferent between A and B, according to this test, if she would prefer A+, a more attractive but otherwise similar version of A to B. (Think of A+ as A plus a sum of money). If she still cannot make up her mind, A and B must be incommensurable. I believe this test is inadequate. By proving that A+ is preferred to A and A+ preferred to B you still do not have conclusive evidence about the nature of the relation between A and B. (Thanks to Jurriaan de Haan of University of Amsterdam for pointing this out).

Consider the following example. Suppose you have two equally smart and deserving children, A and B, who both have been admitted to a top-notch university. Unfortunately, you only have money for one of them. Suppose you cannot make up your mind – which seems only natural because this situation seems very much like a tragic choice given the overall importance of good university education for success in later life. Suppose that you now are presented with opportunity A+ where A goes to a top ten university and you have enough money left to send B the next year as well. Clearly A+ is preferable to both A and B. However, this does not imply that you were indifferent between A and B in the first place. It is clear, however, that this test assumes that indifference and incommensurability are aspects of the relation between the values of rival options.

Elster subsequently argues that true indifference is rare at best in real life. Incommensurability is far more prominent. His examples of (types of) incommensurability come from Boswell’s *Life of Doctor Johnson*. In one example Dr. Johnson explains that our limited cognitive capabilities never enable us to make the ideally rational choice for a certain kind of life because we can never assemble sufficient information. In another example Dr. Johnson disapproves of marrying late in life because the opportunity costs of collecting sufficient information, i.e., waiting for the right person who might or might not come along, are too high. Both examples are off the mark given the test. The test for incommensurability after all assumes that indifference and incommensurability hold between the values of the options of choice. Elster’s examples on the other hand deal with knowledge of the options. Therefore,
indeterminacy is not exhausted by indifference and incommensurability. One should distinguish a third category of epistemic indeterminacy (with all its subclasses).

This has consequences for the view that emotions assist rational choice in settling problems of indeterminacy. Elster finds this claim implausible, arguing that it makes a straw man of rational choice. I agree with Elster that this is the case where it is assumed that a rational agent could not deal with indeterminacies of value. However, I am not so sure the same is true of indeterminacies of knowledge. It is true that there is an opportunity cost to assembling more information for most difficult decisions in life. The magnitude of that cost is in most cases unknown to agent. Therefore, it seems rational to find out how much effort it would cost to get more information. However, there are opportunity costs related to such an inquiry as well. In short, there is a regress here. The rational agent could never be in a position in which she knows that further inquiry is too costly. I do not really see how a plausible account of rational choice, even a satisficing one, could avoid it. If there is radical indeterminacy of knowledge as exemplified in the examples of Dr. Johnson, rationality is at a loss. The only recommendation it could give is to find out what the opportunity costs are, but at the same time, it is recognized that doing so may be too costly. It is, therefore, plausible to assume that the emotions play a key-role in minimizing the costs of such otherwise rational inquiries by simply prompting the agent to choose.

The title chapter, chapter five, is Elster at his best. Here he applies some of the observations made in earlier chapters. It deals with mechanisms of transmutation and misrepresentation of the real motivations of agents. The basic idea is as simple as it is pervasive. Just as it is sometimes undesirable or shameful to represent one’s real motives to others, it is often similarly undesirable to be honest to oneself about one’s motivations. Nobody likes to think of himself as an egoist. Blatant self-interest might be misrepresented to others as concern for the general good. Similarly, it can be transmuted for oneself to such a concern. Another example is the dialectic of envy discussed above. Since envy is shameful, it is attractive to misrepresent one’s envy as righteous anger to others. Moreover, if the feeling of shame results in the formation of the belief that the other did not deserve his good fortune, the envy might transmute into genuine righteous anger.

In this context, Elster discusses the role of reason, understood as an impartial constraint on the sort of arguments and claims one can put forward. For example, an argument to the effect that this policy should be accepted because it serves my interests is not an argument at all. One should at least represent one’s arguments in a more impartial way, arguing for example that this policy will be good for everybody is far more convincing, even if other parties realize that you are motivated by your own interests. Elster gives beautiful examples of forms of misrepresentation and transmutation in the political arena, ranging from speeches by Greek envoys, to recent union negotiations in
Norway. The phenomena of misrepresentation and transmutation, although quite well-known, have never attracted the attention of theorists in a systematic way. There are no plausible theories that explain exactly how misrepresentation and transmutation work. Moreover, many of the underlying mechanisms are poorly understood at best. Given their pervasive importance in politics there is significant work to be done. For example, misrepresentation and transmutation put stress on the feasibility of ideals of deliberative democracy.

The final part of the book, entitled “Coda”, tries to draw some general conclusions from the entire enterprise. It is clearly written as an afterthought. Here Elster repeats the themes that kept coming up in the previous chapters. He argues for what he calls “historical psychology”. Such an approach to the emotions will study the social-historical influences on the emotions as well as the effects of emotions a broader framework than could be done in either the laboratory or the social sciences alone. It will include neurophysiology, psychological experiment, but also social science, history, literature and philosophy as its sources of information. *Alchemies of the Mind* gives us a glimpse of the possible results of such a study.

While reading this book, I could not shake off the impression that it is written too hastily. The chapters are more or less independent in spite of many cross-references and are rather mixed in quality. Chapters two, three and four contain long well-organized lists of phenomena illustrated by beautiful anecdotes and then move on to the next topic without showing his reader the greater picture. Only chapter five contains a clear focus. Part of this is due to the subject matter: the human emotions are too complex and too diverse to be curtailed into a neat and tidy simple theoretical framework. Another part of it is due to Elster’s scientific approach. The stress on mechanisms has many implications for the form of this work. Elster sees it as his main task to point out the indeterminate emotional mechanisms. The resulting overall picture is inevitably as complex as the clusters of mechanisms one could distinguish.

This will not come as a surprise to those who are familiar with Elster’s other writings. Elster often gives you the impression that there is a beautiful necklace of important ideas in his books. He provides you with the beads: invariably interesting conceptions and well chosen examples. However, when you start looking for the thread you find there is not one long enough or strong enough to make the necklace. This might be disappointing to some, but, ah, those pretty beads...!