CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

Chapter I

A Plea for Mechanisms

I.1. INTRODUCTION

Are there lawlike generalizations in the social sciences? If not, are we thrown back on mere description and narrative? In my opinion, the answer to both questions is no. The main task of this chapter is to explain and illustrate the idea of a *mechanism* as intermediate between laws and descriptions. Roughly speaking, mechanisms are *frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences*. They allow us to explain, but not to predict. An example from George Vaillant gives a flavor of the idea: "Perhaps for every child who becomes alcoholic in response to an alcoholic environment, another eschews alcohol in response to the same environment."[1] Both reactions embody mechanisms: doing what your parents do and doing the opposite of what they do. We cannot tell ahead of time what will become of the child of an alcoholic, but if he or she turns out either a teetotaler or an alcoholic we may suspect we know why.

Over the years, I have increasingly come to view the ideal of lawlike explanation ("covering-law explanation") in history and the social sciences as implausible and fragile. Early on, I was struck by Paul Veyne's discussion of the idea of providing a nomological explanation of Louis XIV's unpopularity. Suppose we start from the generalization that "any king imposing excessive taxes becomes unpopular." To take care of counterexamples, we would soon have to modify the statement by introducing exceptions and qualifications, the cumulative effect of which would be to "reconstitute a chapter of the history of the reign of Louis XIV with the amusing feature of being written in the

1. Vaillant (1995), p. 65.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

present and the plural" rather than in the past tense and the singular.[2] Later Raymond Boudon offered forceful arguments in the same vein.[3]

I was even more struck by the total lack of consensus among the best practitioners in the social sciences and by the numerous failures of prediction – both the failures to predict and the predictions that failed. My own studies of collective bargaining[4] and of the allocation of scarce goods[5] entrenched this generally skeptical attitude, bordering on explanatory nihilism. The downfall of Communism in Eastern Europe and its subsequent reemergence provide two stunning examples of massive social changes that were virtually unanticipated by the scientific community. The virulent civil war in the former Yugoslavia offers another. What pulled me back from the nihilist conclusion was the recognition that the idea of a mechanism could provide a measure of explanatory power.

As I reached this conclusion I discovered that it had been anticipated by Nancy Cartwright's claim that "the laws of physics lie." Hence the resort to explanation by mechanism in the social sciences may not be due to their less developed state or to the complexity of their subject matter, but to more general facts about human understanding or about the world. The following passage will convey some of the flavor of Cartwright's argument:

> Last year I planted camellias in my garden. I know that camellias like rich soil, so I planted them in composted manure. On the other hand, the manure was still warm, and I also know camellia roots cannot take high temperatures. So I did not know what to expect. But when many of my camellias died, despite otherwise perfect care, I knew what went wrong. The camellias died because they were planted in hot soil. . . .
>
> So we have an explanation for the death of my camellias. But it is not an explanation from any true covering law. There is no law that says that camellias just like mine, planted in soil which is both hot and rich, die. To the contrary, they do not all die. Some thrive; and probably those that do, do so *because* of the richness of the soil they were planted in. We may insist that there must be some differentiating factor which brings the case under a covering law: in soil which is rich and hot, camellias of one kind die; those of another thrive. I will not deny that there may be such a covering law. I merely repeat that our ability to give this humdrum explanation precedes our knowledge of that

2. Veyne (1971), p. 198.
3. Boudon (1984).
4. Elster (1989a).
5. Elster (1992).

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

law. On the Day of Judgment, when all laws are known, these may suffice to explain all phenomena. But in the meantime we do give explanations; and it is the job of science to tell us what kinds of explanations are admissible.[6]

Cartwright's example relies on what I call *type B mechanisms*. Briefly defined (see I.2 for a fuller discussion), they arise when we can predict the triggering of two causal chains that affect an independent variable in opposite directions, leaving the net effect indeterminate. I contrast them with *type A mechanisms*, which arise when the indeterminacy concerns which (if any) of several causal chains will be triggered. An example from the natural sciences of type A mechanisms can be taken from fear-elicited behavior in animals.[7] Environmental stimuli can trigger one of three mutually incompatible fear reactions: fight, flight, or freeze. We know something about the conditions that trigger these reactions. Thus, "in response to a painful shock, animals will typically show increased activity, run, jump, scream, hiss or attack a suitable target (e.g., another animal) in their vicinity; but, in response to a stimulus associated with shock, the animal will most likely freeze and remain silent. [The] brain mechanisms that mediate these two kinds of reactions are quite distinct."[8] But although we can identify the conditions that trigger freeze versus either fight or flight, we do not know which will trigger fight versus flight. "Rather than thinking in terms of two systems for reaction to different classes of punishment, it makes better sense to imagine a single fight/flight mechanism which receives information about all punishments and then issues commands *either* for fight *or* for flight depending on the total stimulus context in which punishment is received."[9] But to say that the independent variable is "the total stimulus context" is equivalent to saying that the two responses are triggered under "generally unknown conditions." Cartwright's example and the flight–fight example provide robust instances of mechanisms in the natural sciences.[10]

In developing the idea of a mechanism I proceed as follows. In I.2 I provide a somewhat more precise definition of the notion of a mechanism. In I.3, I discuss proverbs as a source of insight into

6. Cartwright (1983), pp. 51–2. The substance of Cartwright's book being concerned with quantum mechanics, this homely example obviously cannot convey more than "some of the flavor" of her argument.
7. I am indebted to Nils Roll-Hansen for suggesting this example.
8. Gray (1991), p. 244.
9. *Ibid.*, p. 255.
10. In IV.2 I give examples of type A and B mechanisms that are observed in the physiological expression of the emotions.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

mechanisms. In I.4 and I.5 I discuss the privileged place of mechanism reasoning in Montaigne and Tocqueville. Sections I.3 through I.5 overlap to some extent with the more systematic sections, and some readers may find them confusing or redundant. I believe, however, that it is worthwhile showing that the idea of a mechanism, far from being a novel or radical innovation, has deep roots in commonsense psychology as well as in the writings of some of the greatest social thinkers.[11] In particular, because mechanisms are so central to Tocqueville's thinking and Tocqueville so central for the study of mechanisms, I draw on his work throughout this chapter. In I.6 I discuss some pairs of psychological mechanisms in more detail. In I.7 I indicate how these elementary mechanisms may form building blocks in constructing more complex explanations. In I.8 I discuss some conditions under which it may be possible to move beyond the ex post identification of mechanisms to predictive statements ex ante. Section I.9 offers a few conclusions.

I.2. EXPLAINING BY MECHANISMS

Let me begin by clearing up a terminological ambiguity. In *Explaining Technical Change*, I used the term "mechanism" in a sense that differs from the one adopted here.[12] In that work I advocated the *search for mechanisms* as more or less synonymous with the reductionist strategy in science. The explanation of cell biology in terms of chemistry and of chemistry in terms of physics are strikingly successful instances of the general strategy of explaining complex phenomena in terms of their individual components. In the social sciences this search for mechanisms (or for "microfoundations") is closely connected with the program of *methodological individualism* – the idea that all social phenomena can be explained in terms of individuals and their behavior.

In that earlier analysis, the antonym of a mechanism is a *black box*.[13] To invent an example at random, suppose somebody asserted that

11. My experience when finding the idea of a mechanism in these writers can itself be analyzed in terms of mechanisms. On the one hand, the discovery produced a "recognition effect" that made me feel good. As these great writers had the idea, there must be something to it. On the other hand, the discovery produced a "humiliation effect" that made me feel bad. As they thought of it first, I have less of which to be proud. I am not sure about the net effect, but I think it is positive.
12. Elster (1983a).
13. But see Suppes (1970), p. 91, for the point that "one man's mechanism is another man's black box"; along similar lines, see also King et al. (1994), p. 86.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

unemployment causes wars of aggression and adduced evidence for a strong correlation between the two phenomena. We would hardly accept this as a lawlike generalization that could be used in explaining specific wars, unless we were provided with a glimpse inside the black box and told *how* unemployment causes wars. Is it because unemployment induces political leaders to seek new markets through wars? Or because they believe that unemployment creates social unrest that must be directed towards an external enemy, to prevent revolutionary movements at home? Or because they believe that the armament industry can absorb unemployment? Although many such stories are conceivable, some kind of story must be told for the explanation to be convincing, where by "story" I mean "lawlike generalization at a lower level of aggregation."

In the present analysis, the antonym of a mechanism is a scientific *law*. A law asserts that given certain initial conditions, an event of a given type (the cause) will *always* produce an event of some other type (the effect). An example: If we keep consumer incomes constant, an increase in the price of a good will cause less of it to be sold ("the law of demand"). Again, we may ask for a story to support the law. One story could be that consumers maximize utility. Gary Becker showed, however, that the law of demand could also be supported by other stories, such as that consumers follow tradition or even that they behave randomly.[14]

In more abstract terms, a law has the form "If conditions $C_1, C_2, \ldots C_n$ obtain, then always E." A covering-law explanation amounts to explaining an instance of E by demonstrating the presence of $C_1, C_2, \ldots C_n$. At the same abstract level, a statement about mechanisms might be "If $C_1, C_2, \ldots C_n$ obtain, then sometimes E." For explanatory purposes, this may not seem very promising. It is true, for instance, that when there is an eclipse of the moon, it sometimes rains the next day, yet we would not adduce the former fact to explain the latter. But consider the idea that when people would like a certain proposition to be true, they sometimes end up believing it to be true. In this case, we often do cite the former fact to explain the latter, relying on the familiar mechanism of wishful thinking.

This is not a lawlike phenomenon. Most people hold some beliefs that they would like to be false. Ex ante, we cannot predict when they will engage in wishful thinking – but when they do, we can recognize it after the fact. Of course, the mere fact that a person adopts a

14. Becker (1962).

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

belief that he would like to be true does not show that he has fallen victim to wishful thinking. Even if the belief is false or (more relevantly) inconsistent with information available to him, we cannot infer that this mechanism is at work. He might, after all, just be making a mistake in reasoning that happens to lead to a conclusion that he would like to be true. To conclude that we are indeed dealing with wishful thinking, more analysis is needed. Is this a regular pattern in his behavior? Does he often stick to his beliefs even when evidence to the contrary becomes overwhelmingly strong? Does he seem to be strongly emotionally attached to his belief? Can other hypotheses be discarded? By standard procedures of this kind we can conclude, at least provisionally, that wishful thinking was indeed at work on this particular occasion. In doing so, we have offered an explanation of why the person came to hold the belief in question. The mechanism provides an explanation because it is *more general* than the phenomenon that it subsumes.

In my earlier terminology, going from a black-box regularity to a mechanism is to go from "if A, then always B" to "if A, then always C, D, and B." In this perspective, mechanisms are good because their finer grain enables us to provide better explanations. Understanding the details of the causal story reduces the risk of spurious explanations, that is, of mistaking correlation for causation. Also, knowing the fine grain is intrinsically more satisfactory for the mind.[15] On the view set out in the present chapter, the move from theory to mechanism is from "if A, then always B" to "if A, then sometimes B." In this perspective, mechanisms are good only because they enable us to explain when generalizations break down. They are not desirable in themselves, only *faute de mieux*. Because fine grain *is* desirable in itself, I also urge the further move to "if A, then sometimes C, D, and B."

Mechanisms often come in pairs. For instance, when people would like the world to be different from what it is, wishful thinking is not the only mechanism of adjustment. Sometimes, as in the story of the fox and the sour grapes, people adjust by changing their desires rather than their beliefs.[16] But we cannot make a lawlike statement to the effect that, "Whenever people are in a situation in which rational principles of belief formation would induce a belief that they would like to be false, they fall victim either to wishful thinking or to adaptive preference formation." To repeat, most people hold some beliefs that they

15. On both points, see Elster (1983a), Chapter I.
16. Elster (1983b).

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

would like to be false. Or take another pair of mechanisms: Adaptive preferences versus counteradaptive preferences (sour grapes versus forbidden fruit). Both phenomena are well known and easily recognizable: Some people prefer what they can have; others tend to want what they do not or cannot have. Yet it would be absurd to assert that all people fall in one of these two categories. Similarly, some people are conformists, some are anticonformists (they do the opposite of what others do), and some are neither.

When the paired mechanisms are mutually exclusive, they are what I call type A mechanisms. An explicit recognition of this phenomenon is found in a discussion of the gambler's fallacy and its nameless converse:

> When in a game there is a 50% chance of winning, people expect that a small number of rounds will also reflect this even chance. This is only possible when runs of gains and losses are short: a run of six losses would upset the local representativeness. This mechanism may explain the well-known gamblers fallacy: the expectation that the probability of winning increases with the length of an ongoing run of losses. The representativeness heuristic predicts that players will increase their bet after a run of losses, and decrease it after a run of gains. This is indeed what about half the players at blackjack tables do. . . . But the other half show the reverse behaviour: they increase their bets after winning, and decrease them after losing, which is predicted by the availability heuristic. After a run of losses, losing becomes the better available outcome, which may cause an overestimation of the probability of losing. [The] repertoire of heuristics predicts both an increase and decrease of bet size after losing, and *without further indications about conditions that determine preferences for heuristics, the whole theoretical context will be destined to provide explanations on the basis of hindsight only.*[17]

Yet paired mechanisms can also operate together, with opposite effects on the dependent variable. Even when the triggering of these mechanisms is predictable, their net effect may not be. These are what I call type B mechanisms. For an example, consider the impact of taxes on the supply of labor:

> A high marginal tax rate lowers the opportunity cost or "price" of leisure, and, as with any commodity whose price is reduced, thereby encourages people to consume more of it (and thus do less work). But, on the other hand, it also lowers people's incomes, and thereby may induce them to work harder so

17. Wagenaar (1988), p. 13; italics added.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

as to maintain their standard of living. These two effects – the substitution and income effects, in economists' parlance – operate in opposite directions, and *their net effect is impossible to predict* from theory alone.[18]

As in Cartwright's camellia example, the separate effects are robust propensities, but the net effect is more contingent. The *indeterminacy* associated with mechanisms can, therefore, take two forms. With type A mechanisms we may not be able to predict which of two opposing mechanisms will be triggered. With type B mechanisms we may not be able to assess the net effect of two opposing mechanisms when both are triggered.[19] This is a special case of Max Weber's observation that "the actors in any given situation are often subject to opposed and conflicting impulses, *all of which we are able to understand*. In a large number of cases we know from experience it is not possible to arrive at even an approximate strength of conflicting motives. . . . *Only the actual outcome of the conflict gives a solid basis of judgment*."[20]

A further distinction may be made between cases in which the two opposing mechanisms are triggered simultaneously by the same cause and cases in which one is triggered by the other.[21] I refer to these as mechanisms of type B_1 and B_2 respectively. The camellia example and the income-substitution example are B_1 mechanisms. For an example of a B_2 mechanism, consider the behavior of a person who faces some barrier or impediment to his goal. According to Jack Brehm, this threat to his freedom of action will induce *reactance* – a motivation to recover or reestablish the freedom:

18. Le Grand (1982), p. 148; italics added.
19. From an abstract point of view, type A mechanisms may be subsumed under type B mechanisms. We may define a type B mechanism as the propensity for a cause C to produce two oppositely signed effects E_1 and E_2 of generally unknown magnitudes m_1 and m_2, with an indeterminate net effect $m_1 \cdot E_1 + m_2 \cdot E_2$. Type A mechanisms arise when one of m_1 or m_2 is constrained to be 0. More generally, it may be impossible to distinguish observationally between $m_1 \cdot E_1 + m_2 \cdot E_2$ and $m^* \cdot E_1$, where $m^* < m_1$. What looks like one of the two outcomes in a type A mechanism might be the net effect of a type B mechanism. Thus, "We could assume . . . a general tendency to derogate whatever one might have had but for some reason now cannot have – i.e., a 'sour grapes' effect. This implies that the 'sour grapes' effect is likely to occur whenever there is a behavioral elimination . . . and that it will tend to *subtract from reactance effects*" (Brehm 1966, p. 36; my italics). Similarly, what looks like a weakly conformist behavior might be the net effect of a strong conformist tendency and a weak anti-conformist one: A child might simultaneously want to imitate the parents and to differentiate himself or herself from them. Subsequently, I ignore this complication.
20. Weber (1968), p. 10; my italics.
21. An application of this distinction to Marx's theory of the falling rate of profit is in Elster (1985), pp. 123–24.

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

Since one way of re-establishing one's freedom would be to obtain the goal object despite the barrier, we might expect, considering reactance alone, that the tendency to take the goal object would be a direct function of the magnitude of the barrier. But this would ignore the effort necessary to overcome the barrier, which also increases with the magnitude of the barrier. To the extent this effort is noxious, it will reduce one's tendency to try to overcome the barrier. Thus *there is a predictive impasse* in that the effects of the barrier and the consequent reactance oppose each other and it is not possible to say which, if either, will be the stronger.[22]

Another paradigm case of a B_2 mechanism is the "opponent-process system" stipulated by Richard L. Solomon. According to his theory, the onset or termination of an initial experience of pleasure or pain generates an oppositely signed experience of pain or pleasure.[23] Euphoria and withdrawal in drug addiction illustrate the pleasure–pain sequence. The pain–pleasure sequence is illustrated by the relief a woman experiences upon learning that her fear of cancer was ungrounded. (See Fig. I.2 in section I.8 for illustrations.)

I have asserted that we cannot tell, in general, under what conditions a given mechanism will be triggered or, in the case of several mechanisms that operate simultaneously or successively, what their net effect will be. In doing so, I may appear to dismiss a large psychological literature demonstrating the operation of these mechanisms under specific conditions. Consider, for instance, the availability and representativeness heuristics cited in the gambling example.[24] For each of these mechanisms it is possible to specify conditions under which it will predictably come into play. This is in fact the general

22. Brehm (1966), p. 76; my italics.
23. Solomon and Corbit (1974); Solomon (1980a,b). In most of these writings, Solomon asserts that the opponent process is triggered by the *onset* of the main or primary process, although its effects are masked as long as the stimulus responsible for the primary process is maintained. In some of his examples, however, it makes more sense to assume that the opponent process is triggered by the *termination* of the primary process. It seems more plausible to assume that the relief of the woman who learns that she does not have cancer is triggered by that piece of news, than to assume that the relief was somehow present from the moment she learned that she might have cancer. To take two of Solomon's other examples, it does not seem intuitively reasonable that the sadness induced by the absence of a lover or the irritation felt when sexual intercourse is interrupted by a telephone call would somehow have been present from the very beginning, although masked by the stronger primary process. Recent work on addiction indicates, however, that aversive effects from drugs may be present from the beginning (Gardner 1997), as already suggested by Solomon and Corbit (1974), p. 137.
24. Tversky and Kahneman (1974).

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-64279-8 - Alchemies of the Mind: Rationality and the Emotions
Jon Elster
Excerpt
More information

strategy of experimental work: Manipulate conditions so that unambiguous results can be produced. Yet a set of sufficient conditions that can be realized in experimental situations may not often appear in real-life cases.[25] Knowing that C_1, C_2 ... C_4 are sufficient for X to occur and D_1, D_2 ... D_5 are sufficient for Y to occur does not help us to predict what will happen in the presence of C_1, C_3, D_2, D_4. If we know that "if C_1, then sometimes X" and "if D_4, then sometimes Y," we should be ready for either effect.

To repeat, I am not advancing explanation by mechanisms as an ideal or a norm. Explanation by laws is better – but also more difficult, often too difficult. (See also I.8.) Moreover, as should be clear by now, I am not suggesting that mechanisms can be identified by formal conditions analogous to those that enter into the formulation of laws. "If p, then sometimes q" is a near-useless insight. Explanation by mechanisms works when and because we can identify a particular causal pattern that we can recognize across situations and that provides an intelligible answer to the question, "Why did he do *that*?"

I.3. PROVERBIAL MECHANISMS

The study of proverbs is a good introduction to the study of mechanisms, for several reasons. Proverbs that become entrenched in a culture can be expected to have specific mnemonic features. They will be simple and robust rather than complicated and hedged with qualifications. Moreover, they will not survive unless they illuminate behavior that is widely observed.[26] Finally, it is proverbially true that for any proverb one can find one that asserts the opposite.[27] To explore this source I consider some of the 1,500 proverbs collected in the *Random House Dictionary of Proverbs and Popular Sayings*.[28] Occasionally, I also mention proverbs from other sources, notably

25. Parducci (1995), p. 37, n. 2; p. 45.
26. Mieder (1993), pp. 20–23, cites fifty-five different definitions of proverbs that he elicited from various individuals. The one that best fits my purposes here is that "a proverb has been passed down through many generations. It sums up, in one short phrase, a general principle, or common situation, and when you say it, everyone knows exactly what you mean." See also Elster (in press e)
27. This statement is a slight exaggeration. Although there are many proverbs about proverbs (Mieder 1993, pp. 5, 19) and the existence of contradictory proverbs is a well-recognized fact (*ibid.*, pp. 25–28), it has not itself become embodied in a proverb.
28. Titelman (1996).

© in this web service Cambridge University Press www.cambridge.org