Self-poisoning of the mind

Jon Elster*

Chaire de Racionalité et Sciences Sociales, Collège de France

Rational-choice theory tries to explain behavior on the assumption that individuals optimize. Some forms of irrational behavior can be explained by assuming that the individual is subject to hedonic, pleasure-seeking mechanisms, such as wishful thinking or adaptive preference formation. In this paper, I draw attention to psychic mechanisms, originating in the individual, which make her worse off. I first consider the ideas of counterwishful thinking and of counteradaptive preference formation and then, drawing heavily on Proust, the self-poisoning of the mind that occurs through the operation of amour-propre.

Keywords: amour-propre; wishful thinking; adaptive preferences; cognitive dissonance; reactance; Proust

1. INTRODUCTION

In this paper, I shall discuss what Scheler (1972), inspired by Nietzsche, called ‘self-poisoning of the mind’. I shall not limit myself to the ‘ressentiment’ that was Scheler’s main example, but also consider other counterhedonic mental mechanisms—belief changes or preference changes that originate in the agent and make her worse off. I shall draw heavily on the writings of Marcel Proust, which provide many insights into these matters.1

As background for the discussion, consider the standard model of rational choice (figure 1). The theory of rational choice, or rational action, is primarily normative. It tells people what to do to achieve their aims as well as possible. It can also be used for explanatory purposes, by assuming that people follow the prescriptions of the theory and then determining whether the observed behavior conforms to the prescribed behavior. In a fuller exposition, the preceding statements would have to be qualified in various ways, but for my ends here these nuances do not matter.

Figure 1 represents the explanatory or causal version of the theory. The heavily drawn lines represent both causal relations and optimality relations. The action, for instance, is optimal in the light of the desires and beliefs that cause it. The lightly drawn lines represent causal relations that are not also optimality relations. Thus, the line from desires to beliefs represents wishful thinking, self-deception and other forms of motivated belief formation. The line from beliefs to desires or preferences represents mechanisms such as the reduction of cognitive dissonance or adaptive preference formation. Dissonance reduction in fact applies more widely—it can have the effect of aligning beliefs on desires and even on emotions, as we shall see.

Although aligning one’s beliefs on one’s desires is intrinsically irrational, it does embody a form of short-term optimizing. Believing that the world is as you would like it to be provides some kind of immediate satisfaction, or at least removes the discontent that is produced when beliefs and desires diverge. In the long run, of course, one might be very badly off acting on beliefs adopted on hedonic grounds. As a Norwegian proverb has it: pissing in one’s pants gives brief warmth.

Aligning desires on beliefs is neither rational nor irrational. There are no rationality criteria for processes of preference formation or for the outcome of such processes, except for the requirement that preferences be logically consistent. At the same time, the outcome of adaptive preference formation or of dissonance reduction can be seen as a form of optimizing, as they make the agent better off.

The alignments of desires and preferences on one another tend to improve the welfare of the agent, at least in the short run. Some processes of belief adjustment or preference adjustment seem, however, to make the agent worse off, not better. As Amos Tversky once remarked (in conversation), they embody dissonance production rather than reduction. The puzzle is to explain how they can arise. Tversky conjectured that they might be the outcome of a ‘crossing of the wires in the pleasure machine’, a metaphor suggesting that the production of dissonance is a biochemical phenomenon on a par with some forms of mental illness. Without denying that this may indeed be so, I shall offer some alternative suggestions. More specifically, I shall draw on the French moralists (Elster 1999, ch. II.3) to argue for the importance of amour-propre in the self-poisoning of the mind. While amour-propre is not itself an emotion, anything that threatens it can provoke strong emotional reactions. Anything that can bolster it may also induce strong emotions.

In §2, I consider the mechanisms of counteradaptive preference formation and counterwishful thinking. The latter has received some attention from philosophers (Pears 1984, pp. 42–44; Mele 2001, ch. 5), but not, to my knowledge, from social scientists. The former has received brief attention from economists (von Weizsäcker 1972), but not, to my knowledge, from other quarters. Rather than reviewing the literature, I shall suggest some approaches that draw on the French moralists and Proust.
individuals, the idea that they might be the plaything propre could itself be due to amour-propre. For some suggested that the fear of being entrapped by amour-propre tends to make us overvalue whatever things are foreign to me, lacking in me or are not mine’.

Unlike Pascal, Montaigne did not explicitly explain these counterhedonic tendencies as the result of leaning over backwards to deflect self-suspicion. For a very explicit description of that mechanism, we may turn to Proust.

I had long since been prepared, by the strong impression made on my imagination and my faculty for emotion by the example of Swann, to believe in the truth of what I feared rather than of what I should have wished. And so the comfort brought me by Alber-tine’s affirmations came near to being jeopardized for a moment, because I was reminded of the story of Odette. But I told myself that, if it was only right to allow for the worst, not only when, in order to understand Swann’s sufferings, I had tried to put myself in his place, but now, when I myself was concerned, in seeking the truth as though it referred to some one else, still I must not, out of cruelty to myself, a soldier who chooses the post not where he can be of most use but where he is most exposed, end in the mistake of regarding one supposition as more true than the rest, simply because it was more painful.


The text is remarkable and perhaps unique in that it invokes not only the suspicion of wishful thinking in oneself, but also suspicion of that suspicion itself. These vertiginous doubts and doubts about doubts are indeed characteristics of Proustian jealousy (Grimaldi 1993; Landy 2004). For my purposes here, however, I want only to stress the basic counterhedonic mechanism. Suspicion of one’s tendency to believe the best may cause one to believe the worst.

Another possible mechanism for generating counterhedonic phenomena is, surprisingly, the reduction of cognitive dissonance. In A Theory of Cognitive Dissonance, Festinger (1957, pp. vi–vii) explains the origin of the theory as follows:

The fact [...] which puzzled us was that following the [1934 Indian] earthquake, the vast majority of the rumors that were widely circulated predicted even worse disasters to come in the very near future. Certainly the belief that horrible disasters were about to occur is not a very pleasant belief, and we may ask ourselves why rumors that were ‘anxiety provoking’ arose and were so widely accepted. Finally a possible answer occurred to us—an answer that held promise of having rather general application: perhaps these rumors predicting even worse disasters to come were not ‘anxiety provoking’ at all but rather ‘anxiety justifying’. That is, as a result of the earth-quake these people were already frightened, and the rumors served the function of giving them something to be frightened about (my italics).

It is worthwhile mentioning that many of the post-earthquake rumours involved natural calamities.
other than earthquakes, such as cyclones and floods (Festinger 1957, p. 238). If the rumours had simply predicted more earthquakes, they could have had a rational basis in the risk of aftershocks. By contrast, a belief in the risk of a cyclone following an earthquake has no rational foundation.

Festinger’s case is not an isolated one because rumours tend in fact to be on the pessimistic rather than on the optimistic side (Ploux 2003, p. 63). They mostly express counterwishful thinking rather than wishful thinking. Festinger’s explanation of this tendency seems to presuppose that the belief in the rumour causes a net decrease in psychological discomfort. The decrease in discomfort caused by the consonance between emotion and belief must more than offset the increased discomfort caused by holding a ‘not very pleasant belief’. In other words, the state of being afraid for a reason must be less painful than the state of being afraid for no reason. It is not obvious to me that this is true, nor how one would go about determining whether it is true.

The belief in a just world (Lerner 1980), an offshoot of the cognitive dissonance theory, also has apparently counterhedonic effects. To the extent that people assume that the world is fundamentally just, they are motivated to ‘blame the victim’, even palpably innocent victims such as young men who had drawn an unlucky number in the draft lottery. In fact, even those who drew the unlucky number tended to blame themselves (Rubin & Peplau 1973). The self-blame of rape victims illustrates the same phenomenon. Once again, however, it is somewhat counterintuitive that the comfort these individuals draw from believing that their fate is just should dominate the discomfort produced by their belief that they are to blame.

Finally, we might try to explain counteradaptive preference formation as the result of reactance (Brehm 1966). Imagine a child who prefers toys A, B and C in that order. If a parent suggests that she might want to choose A, the suggestion may cause her to choose B. On one interpretation, autonomy of choice trumps welfare. (One should not say that autonomy is preferred to welfare, as that preference itself would simply be one aspect of welfare.) It has been argued, for instance, that non-compliance of patients with regard to medical instructions might be due to reactance (Fogart 1997).

The ultimate explanation of reactance is probably to be found in the amour-propre of the agent. La Rochefoucauld provided an early analysis of this phenomenon:

The scorn for riches displayed by the philosophers was a secret desire to recompense their own merit for the injustice of Fortune by scorning those very benefits she had denied them; it was a private way of remaining unsullied by poverty; a devious path towards the high respect they could not command by wealth

(Maxim 54).

Before I pursue the further development by Nietzsche of this idea, let me note that La Rochefoucauld had been anticipated, and in a sense disproved, by Thales. According to Aristotle (Politics 1259 b),

[Thales] was reproached for his poverty, which was supposed to show that philosophy was of no use. According to the story, he knew by his skill in the stars while it was yet winter that there would be a great harvest of olives in the coming year; so, having a little money, he gave deposits for the use of all the olive-presses in Chios and Miletus, which he hired at a low price because no one bid against him. When the harvest-time came, and many were wanted all at once and of a sudden, he let them out at any rate which he pleased, and made a quantity of money. Thus he showed the world that philosophers can easily be rich if they like, but that their ambition is of another sort.

In his retelling of the story, de Montaigne (1991, p. 153) explicitly asserts that when he condemned money-making, Thales ‘was accused of sour grapes like the fox’. Although Thales wanted to ‘show the world’ that the accusation was unfounded, one could also imagine that he had made a fortune in order to demonstrate to himself that his philosophy was not the product of sour grapes. Not content with thinking that he could have acquired riches had he wanted to, he might have decided to actually acquire them to deflect self-suspicion. I return to this point.

Nietzsche was heavily influenced by La Rochefoucauld (Donnellan 1979). His description (Nietzsche 1967, I.14) of the ‘workshop’ in which the transmutation of values takes place may well have been inspired by the Maxims:

It is a careful, crafty, light rumor-mongering and whispering from every nook and cranny. It seems to me that people are lying; a sugary mildness clings to every sound. Weakness is going to be falsified into something of merit. [...] And powerlessness which does not retaliate is being falsified into ‘goodness,’ anxious baseness into ‘humility,’ submission before those one hates to ‘obedience’ (of course, obedience to the one who, they say, commands this submission—they call him God). The inoffensiveness of the weak man—cowardice itself, in which he is rich, his standing at the door, his inevitable need to wait around—here acquires a good name, like ‘patience’, and is called virtue itself. That incapacity for revenge is called the lack of desire for revenge, perhaps even forgiveness.

There are two ideas at work in the sentence I have italicized. One is the transmutation of ‘I cannot do x’ or obtain x’ into ‘I do not want to do x or to obtain

3. TRANSMUTATIONS
By ‘transmutation’ I shall mean any change in beliefs or preferences that is caused by a threat to the
x’. The other is the transmutation of the latter into ‘I want not to do x or not to obtain x’, that is, the transmutation of passive negation into active negation (Elster 1993, ch. 2). The first mechanism is illustrated by the incapacity for revenge turning into the lack of desire for revenge, the second by the lack of desire for revenge turning into forgiveness, that is, the desire to abstain from revenge. Later, I shall suggest a further transmutation.

It is perhaps not clear why the outcome of these two transmutations should be referred to as self-poisoning of the mind. But consider the following example. Peter is attracted to Anne, but she does not require his love. As a consequence, he ceases to desire her and persuades himself that she is in fact positively undesirable. It is this downgrading of what he cannot have (‘sour grapes’) that constitutes the self-poisoning. If Peter had simply contented himself with redirecting his desire to another woman more inclined to requite it, no self-poisoning would be involved.

We can see that this downgrading is unattractive and might induce all kinds of ugly behaviours, but is it counterhedononic? Does it necessarily have a negative impact on the welfare of the agent? As we shall see shortly, Proust offers one example in which the downgrading enhances welfare and one in which it detracts from it. We can nevertheless make, I think, a general argument for the claim that downgrading tends to have counterhedononic effects. Whenever Anne’s name comes up in conversation, Peter is likely to react with derogatory remarks that have no basis in facts, only in her rejection of him. Others may notice his attitude and suspect its basis and, as a result, come to dislike and avoid him. Indirectly, therefore, Peter’s reaction induces a loss of welfare. We shall shortly see an example of this mechanism in Proust.

Proust offers several examples of the transmutation of ‘I cannot have it’ into ‘I do not want it’. The first and least consequential is also the most amusing. It occurs in the context of an exchange between Mme de Gallardon and Oriane, Princesse des Laumes (the future Duchesse de Guermantes):

‘Oriane, don’t be angry with me’, resumed Mme de Gallardon, who could never restrain herself from sacrificing her highest social ambitions, and the hope that she might one day emerge into a light that would dazzle the world, to the immediate and secret satisfaction of saying something disagreeable, ‘people do say about your M. Swann that he’s the sort of man one can’t have in the house; is that true?’

‘Why, you, of all people, ought to know that it’s true’, replied the Princesse des Laumes, ‘for you must have asked him a hundred times, and he’s never been to your house once’. (I, p. 330.)

The next episode occurs at Balbec, where the Narrator observes the behaviour of two bourgeois wives towards an old and noble lady:

Whenever the wives of the notary and the magistrate saw her in the dining-room at meal-times they put up their glasses and gave her an insolent scrutiny, as minute and distrustful as if she had been some dish with a pretentious name but a suspicious appearance which, after the negative result of a systematic study, must be sent away with a lofty wave of the hand and a grimace of disgust.

No doubt by this behaviour they meant only to show that, if there were things in the world which they themselves lacked—in this instance, certain prerogatives which the old lady enjoyed, and the privilege of her acquaintance—it was not because they could not, but because they did not want to acquire them. But they had succeeded in convincing themselves that this really was what they felt; and it was the suppression of all desire for, of all curiosity as to forms of life which were unfamiliar, of all hope of pleasing new people (for which, in the women, had been substituted a feigned contempt, an artificial brightness) that had the awkward result of obliging them to label their discontent satisfaction, and lie eternally to themselves, two conditions for their being unhappy. But everyone else in the hotel was no doubt behaving in a similar fashion, though their behaviour might take a different form, and sacrificing, if not to self-importance, at any rate to certain inculcated principles and mental habits the thrilling delight of mixing in a strange kind of life. Of course, the atmosphere of the microcosm in which the old lady isolated herself was not poisoned with virulent bitterness, as was that of the group in which the wives of the notary and magistrate sat chatting with impotent rage (II, p. 38; my italics).

In this text, the transmutation of ‘I cannot’ into ‘I do not want to’ is explicitly cited as a cause of poisoning, bitterness and unhappiness. Although, as noted initially, the alignment of desires on beliefs is usually thought to induce dissonance reduction and greater contentment with one’s fate, here the very opposite effect occurs. The element of self-deception, of lying to oneself, may be responsible. Whereas wishful thinking may be free of self-doubts, self-deception rarely is.

The contrast could not be greater with the following example, in which the same mechanism is said to be conducive to happiness rather than to unhappiness. The episode involves the absurdly self-contented father of the Narrator’s friend Bloch.

M. Bloch senior [...] lived in the world of half-truths where people salute the empty air and arrive at wrong judgments. Inexactitude, incompetence do not modify their assurance; quite the contrary. It is the propitious miracle of amour-propre that, since few of us are in a position to enjoy the society of distinguished people, or to form intellectual friendships, those to whom they are denied still believe themselves to be the best endowed of men, because the optics of our social perspective make every grade of society seem the best to him who occupies it, and beholds as less favored than himself, less fortunate and therefore to be pitied, the greater men whom he names and calumniates without knowing, judges and—despises without understanding them. Even in cases where the multiplication of his modest personal advantages by his amour-propre would not suffice to assure a man the dose of happiness, superior to that accorded to others, which is essential to him, envy is always there to make up the balance. It is true that if envy finds expression in scornful phrases, we must translate ‘I have no wish to know him’ by ‘I have no means of
knowing him’). That is the intellectual sense. But the emotional sense is indeed, ‘I have no wish to know him’. The speaker knows that it is not true, but he does not, all the same, say it simply to deceive; he says it because it is what he feels, and that is sufficient to bridge the gulf between them, that is to say to make him happy. (Il, pp. 129–130; my italics).

The last few sentences are somewhat impenetrable. I find it hard to make sense of them. Yet, the overall idea of the passage seems clear: the upgrading of one’s own small advantages may, if necessary, be supplemented by the downgrading of the greater advantages of others, to produce happiness. As suggested by the comments on the bourgeois wives of Balbec, the latter mechanism may not by itself yield the same result.

Proust’s reference to the downgrading mechanism as envy is, however, distinctly idiosyncratic. Envy presupposes the recognition of the value of the envied object, not the denial of its value. The action tendency of envy is to destroy what you cannot get, not to denigrate it. The following passage shows that Proust was perfectly aware of this standard understanding of envy, and of its effect on the envious. It occurs in a comment on the reactions induced by the worldly successes of the Narrator’s friend Albertine:

Albertine’s ‘successes’ in society excited the envy of certain spiteful mothers, furious at seeing her received like one of the family by the banker’s wife, even by Andrée’s mother, neither of whom they themselves really knew. They therefore went about telling common friends of those ladies and their own that both ladies would be very angry if they knew the facts, which were that Albertine repeated to each of them everything that the intimacy to which she was rashly admitted enabled her to spy out in the household of the other, a thousand little secrets which it must be infinitely unpleasant to the interested party to have made public. These envious women said this so that it might be repeated and might get Albertine into trouble with her patrons. But, as often happens, their machinations met with no success. The spite that prompted them was too apparent, and their only result was to make the women who had planned them appear rather more contemptible than before (Il, p. 289).

Envy, notoriously, is a self-poisoning mechanism, which exacerbates rather than alleviates the pang caused by the perception of another’s greater success or fortune. The cause, in this case, is the perception by others that one is envious. In other cases, the agent’s own perception that she is harbouring this stigmatized emotion may be sufficient to make her unhappy.

I conclude this Proustian catalogue by some comments on the complex case of Legrandin, a character whose outwardly anti-snob attitude hides deep inward snobbery. The Narrator cites his grandmother’s surprise at ‘the furious invective which [Legrandin] was always launching at the aristocracy, at fashionable life, and “snobishness”—“undoubtedly”, he would say, “the sin of which Saint Paul is thinking when he speaks of the sin for which there is no forgiveness”’ (I, p. 67). From the context, it seems that the grandmother may have thought that Legrandin ‘doth protest too much’. If so, this impression is confirmed later, when the Narrator innocently asks Legrandin whether he knows the Guermantes family. The acuity of the Narrator’s analysis of Legrandin’s response justifies, I hope, a lengthy quotation:

[At] the sound of the word Guermantes, I saw in the middle of each of our friend’s blue eyes a little brown dimple appear, as though they had been stabbed by some invisible pin-point, while the rest of his pupils, reacting from the shock, received and secreted the azure overflow. His fringed eyelids darkened, and drooped. His mouth, which had been stiffened and seared with bitter lines, was the first to recover, and smiled, while his eyes still seemed full of pain, like the eyes of a good-looking martyr whose body bristles with arrows.

‘No, I do not know them’, he said, but instead of uttering so simple a piece of information, a reply in which there was so little that could astonish me, in the natural and conversational tone which would have befitted it, he recited it with a separate stress upon each word, leaning forward, bowing his head, with at once the vehemence which a man gives, so as to be believed, to a highly improbable statement (as though the fact that he did not know the Guermantes could be due only to some strange accident of fortune) and with the emphasis of a man who, finding himself unable to keep silence about what is to him a painful situation, chooses to proclaim it aloud, so as to convince his hearers that the confession he is making is one that causes him no embarrassment, but is easy, agreeable, spontaneous, that the situation in question, in this case the absence of relations with the Guermantes family, might very well have been not forced upon, but actually designed by Legrandin himself, might arise from some family tradition, some moral principle or mystical vow which expressly forbade his seeking their society.

‘No’, he resumed, explaining by his words the tone in which they were uttered. ‘No, I do not know them; I have never wished to know them; I have always made a point of preserving complete independence; at heart, as you know, I am a bit of a Radical. People are always coming to me about it, telling me I am mistaken in not going to Guermantes, that I make myself seem ill-bred, uncivilized, an old bear. But that’s not the sort of reputation that can frighten me; it’s too true! […]’

If I asked him, ‘Do you know the Guermantes family?’ Legrandin the talker would reply, ‘No, I have never cared to know them’. But unfortunately the talker was now subordinated to another Legrandin, whom he kept carefully hidden in his breast, whom he would never consciously exhibit, because this other could tell stories about our own Legrandin and about his snobishness which would have ruined his reputation for ever; and this other Legrandin had replied to me already in that wounded look, that stiffened smile, the undue gravity of his tone in uttering those few words, in the thousand arrows by which our own Legrandin had instantaneously been stabbed and sickened, like a Saint Sebastian of snobbery: ‘Oh, how you hurt me! No, I do not know the
Guermantes family. Do not remind me of the great sorrow of my life. And since this other, this irrepressible, dominant, despotic Legrandin, if he lacked our Legrandin’s charming vocabulary, shewed an infinitely greater promptness in expressing himself, by means of what are called ‘reflexes’, it followed that, when Legrandin the talker attempted to silence him, he would already have spoken, and it would be useless for our friend to deplore the bad impression which the revelations of his alter ego must have caused, since he could do no more now than to mitigate them. (I, pp. 126–127; my italics.)

Legrandin’s suggestion that his absence of relations with the Guermantes was a matter of choice rather than of necessity illustrates the simple transmutation of ‘I cannot’ into ‘I do not want to’. His further claim that people told him that it was a mistake not to visit the Guermantes—as if he could easily have done so—suggests that he is subject to a third transmutation, which to the statement ‘I do not want to know them’ adds ‘But I could have known them had I wanted to’. This is a self-deceptive factual statement, not a matter of preference. It is as if Peter, having first been rejected by Anne and then downgraded her, had added that ‘Of course I could easily have married her had I wanted to’. I suspect that this joint transmutation of preferences and beliefs is quite common. Hence, even though a sour grape reaction is not in itself irrational, it may go together with irrational belief formation. The fox probably could not have persuaded himself that he could have reached the grapes by jumping high enough, but in interpersonal relations constraints are rarely so hard that they cannot be undone by a self-deceptive rewriting of the script.

All these transmutations serve to bolster the amourexplosive of the agent. (i) The transmutation of ‘I cannot have it’ into ‘I do not want to have it’ restores a sense of agency. (ii) The further transmutation into ‘It is not worthwhile having anyway’ cements the wisdom of the rejection. (iii) The final transmutation into ‘I could have had it if I had wanted it’ weakens any suspicion of sour grapes. Yet, the bolstering remains fragile. For Legrandin to truly persuade himself and others that he could have frequented the Guermantes had he wanted to, he would have had to follow the example of Thales: obtain an invitation, visit them once and then never again. For Legrandin, such a course of behaviour would be unthinkable. Had he been invited by the Duchesse de Guermantes, he would have gone there again and again, persuading himself that ‘he was yielding to the attractions of her mind, and her other virtues, which the vile race of snobs could never understand’. (I, p. 127)

The self-poisoning effect of transmutations is due largely to the fact they are so obvious to others. Legrandin provides a further illustration of this mechanism. Both the mother and the father of the narrator see through him, the former being ‘greatly delighted whenever she caught him red-handed in the sin, which he continued to call the unpardonable sin, of snobbery’ (I, p. 128) and the latter deliberately ‘torturing him’ (I, p. 131) with requests for information that his snobbery would not allow him to give out.

The general idea that helps us understand these phenomena is that the unconscious is never wise. In some cases, as in wishful thinking, unconscious reactions may cause temporary alleviation or gratification. In other cases, illustrated by Legrandin’s reflexes that give him away before he can catch himself, they may cause instant and irreversible damage. On these lines, there is a story (which I have been unable to track down) told about Sigmund Freud, who was invited to meet a person, Dr X, who was prominent in the international Jewish movement. During their conversation, Dr X asked him, ‘Tell me Dr Freud, who is in your opinion the most important Jewish personality in the world today?’ Freud answered politely, ‘Why, I think that must be yourself, Dr X’.

When Dr X replied, ‘No, No’, Freud asked, ‘Wouldn’t ‘No’ had been enough?’ Double negation can be equivalent to affirmation.

ENDNOTE


REFERENCES


