You have been presented with an exciting opportunity. Before you on the table are two boxes, B1 and B2. B1 is transparent ; you can see that it contains $1,000. B2, which is opaque, contains either $1,000,000 or nothing.

- B1 : $ 1,000
- B2 : $ 1,000,000 or nothing

You have a choice between two actions : taking what is in both boxes or taking only what is in the second box. Before you make your choice, the following background information is carefully explained. The content of the second box is determined by a superlative predictor who has successfully predicted the choice of all (almost all) those who were previously placed in this situation. His prediction is based on an in-depth psychological study of the individual ; you have already been examined by him, and a detailed profile of your basic personality and character traits has been constructed. After making his prediction, the predictor acts as follows :

1. If he predicts that you will take what is in both boxes, he puts nothing in the second box.
2. If he predicts that you will take just the second box, he puts $1,000,000 in the second box.

Now it is your turn. You have five minutes to reflect. What should you do ?

One line of reasoning that seems utterly compelling goes as follows. The predictor has already consulted your psychological profile, made his prediction, and either placed $1,000,000 in B2 or left it empty. Nothing you do now can affect his prior decision. The content of B2 is fixed and determined. Consider the two possibilities : either there is $1,000,000 in B2 or there is $0 in B2. In the first case, you will get $1,001,000 if you take both boxes, whereas you will get (only) $1,000,000 if you take just B2. In the second case, you will get $1,000 if you take both boxes, but nothing if you take just B2. In either case, you are $1,000 richer if you take both boxes. So clearly taking both boxes is the rational thing to do.

But another argument seems equally forceful. Given the predictor's astonishing past record of predictive success, it is virtually certain that he will correctly predict your choice. Thus, if you take both boxes, almost certainly he will have predicted this, will have left B2 empty, and you will get only $1,000. Similarly, if you take just B2, almost certainly he will have predicted this, will have placed $1,000,000 in B2, and you will get $1,000,000. The choice is between $1,000 and $1,000,000. Clearly, taking just B2 is the rational thing to do.

Doris Olin, *Paradox*, Acumen, 2003, p. 105-106.

You and I have been arrested for drug running and placed in separate cells. Each of us learns, through his own attorney, that the district attorney has resolved as follows (and we have every reason to trust this information) :

(1) If we both remain silent, the district attorney will have to drop the drug-running charge for lack of evidence, and will instead charge us with the much more minor offense of possessing dangerous weapons. We would then each get a year in jail.

(2) If we both confess, we shall both get five years in jail.

(3) If one remains silent and the other confesses, the one who confesses will get off scot-free (for turning state's evidence), and the other will go to jail for ten years.

(4) The other prisoner is also being told all of (1)-(4).

How is it rational to act ? We build into the story the following further features :

(5) Each prisoner is concerned only with getting the smallest sentence for himself.

(6) Neither has any information about the likely behavior of the other, except that (5) holds of him and that he is a rational agent.

There is an obvious line of reasoning in favor of confessing. It is simply that whatever you do, I shall do better to confess. For if you remain silent and I confess, I shall get what I most want, no sentence at all ; whereas if you confess, then I shall do much better by confessing too (five years) than by remaining silent (ten years). We can represent the situation by table 4.5, and the reasoning in favor of confessing is the familiar dominance principle (DP).

	you confess	you don't confess
I confess	<5,5>	<0,10>
I don't confess	<10,0>	<1,1>

Table 4.5 : The Prisoner's Dilemma.

In table 4.5 <0, 10> represents the fact that on this option I go to prison for zero years, and you go for ten years ; and so on. The smaller the number on my side of the pair (the left side), the better I am pleased. It is easy to see that confessing dominates silence : confessing, as compared to silence, saves me five needless years if you confess, and one if you do not.

Since you and I are in relevantly similar positions, and, by (6), we are both rational, presumably we shall reason in the same way, and thus perform the same action. So if it is rational for me to confess, it is rational for you to do likewise ; but then we shall each go to prison for five years. If we both remain silent, we would go to prison for only one year each. By acting supposedly rationally, we shall, it seems, secure for ourselves an outcome that is worse for both of us than what we could achieve.

On this view, rational action in some circumstances leads to worse outcomes than other courses of action.

Even if this is depressing, it is not as it stands paradoxical : we all know that irrational gambles can succeed. What is arguably paradoxical is that the case is one in which the failure of rationality to produce the best results is not a matter of some chance intervention, but is a predictable and inevitable consequence of so-called rational reasoning. How, in that case, can it be rational to be "rational" ?

R. Sainsbury, *Paradoxes*, 3rd ed., Cambridge University Press, p. 89-91.