1 The Unexpected Examination

The teacher tells the class that sometime during the next week she will give an examination. She will not say on which day, for, she says, it is to be a surprise. On the face of it, there is no reason why the teacher, despite having made this announcement, should not be able to do exactly what she has announced: give the class an unexpected examination. It will not be totally unexpected, since the class will know, or at least have good reason to believe, that it will occur sometime during the next week. However, surely it could be a surprise, or unexpected, in this sense: that on the morning of the day on which it is given, the class will have no good reason to believe that it will occur on *that* day, even though they know, or have good reason to believe, the teacher's announcement. Cannot the teacher achieve this aim by, say, giving the examination on Wednesday?

The class reasons as follows. Let us suppose that the teacher will carry out her threat, in both its parts: that is, she will give an examination, and it will be unexpected. Then the teacher cannot give the examination on Friday (assuming this to be the last possible day of the week); for, by the time Friday morning arrives, and we know that all the previous days have been examination-free, we would have every reason to expect the examination to occur on Friday. So leaving the examination until Friday is inconsistent with giving an *unexpected* examination. For similar reasons, the examination cannot be held on Thursday. Given our previous conclusion that it cannot be delayed until Friday, we would know, when Thursday morning came, and the previous days had been examination-free, that it would have to be held on Thursday. So if it were held on Thursday, it would not be unexpected. Thus it cannot be held on Thursday. Similar reasoning supposedly shows that there is no day of the week on which it can be held, and so supposedly shows that the supposition that the teacher can carry out her threat must be rejected. This is paradoxical, for it seems plain that the teacher can carry out her threat.

Something must be wrong with the way in which the class reasoned; but what?

R. M. Sainsbury, *Paradoxes*, third edition, Cambridge University Press, 2009.

2 Newcomb's Paradox

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.