Newcomb's Problem as a Prisoner's Dilemma

An interesting observation, due to philosopher David Lewis, is that the Newcomb Problem can be thought of as a special case of the Prisoner's Dilemma in which you and Jones are similar enough that you're highly likely to end up making the same decisions, even though you haven't communicated with one another.

Suppose, for example, that Jones is your clone. The two of you are molecule for molecule identical, and have grown up in parallel environments. As a result, the probability that you'll make the same decision in this case is 99%. Now imagine yourself facing 10,000 days of prison. You have two boxes in front of you. The small box contains a voucher, which reduces your prison time by 1,000 days; the large box contains a voucher which will either be counted as reducing your prison time by 9,000 days, or declared void. You are offered two choices: take both boxes or take just the large box.

How is the value of the voucher in the large box determined? Jones is in a situation analogous to yours. If he decides to two-box, the voucher will be declared void; and if he decides to one-box, it will be counted as giving you a 9,000 day reduction. (Similarly, if you decide to two-box, the voucher in Jones's large box will declared void, and if you decide to one-box, it will be counted as giving him a 9,000 day reduction.) Since there is no communication between you, your choices are causally independent. But since you are clones, the choices are not probabilistically independent: it is 99% likely that you will both end up making the same choice.

We have constructed a version of the Prisoners' Dilemma in which defecting is, in effect, two-boxing and keeping quiet is, in effect, one-boxing. In this sense, the Newcomb Problem is a special case of the Prisoners' Dilemma.

Numerical Input

2/2 points (ungraded)

On the assumption that the probability that Jones and you will make the same decision is 99%, what is the expected value of defecting? (Note: in an outcome where you spend n days in prison, your payoff is -n.)

-8910 ✔ **Answer:** -8910

-8910

On that assumption, what is the expected value of keeping quiet?

-1090 ✔ **Answer:** -1090

-1090

Explanation

Let D_Y be the proposition that you defect, D_J the proposition that Jones defects, Q_Y the proposition that you keep quiet and Q_J the proposition that Jones keeps quiet.

1. Expected value of defecting:

$$EV(D_Y) = v(D_Y D_J) \cdot p(D_J|D_Y) + v(D_Y Q_J) \cdot p(Q_J|D_Y) = -9,000 \cdot 99 + 0 \cdot 0.01 = -8910$$

2. Expected value of keeping quiet:

$$EV(Q_Y) = v(Q_Y D_J) \cdot p(D_J|Q_Y) + v(Q_Y Q_J) \cdot p(Q_J|Q_Y) = -10000 \cdot 0.01 + -1000 \cdot 0.99 = -1090$$

Submit

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Week 5 / Newcomb's Problem as a Prisoner's Dilemma

Add a Post

Show all posts by recent activity

There are no posts in this topic yet.