

ON NEWCOMB'S PROBLEM

Zvonimir Šikic, Zagreb

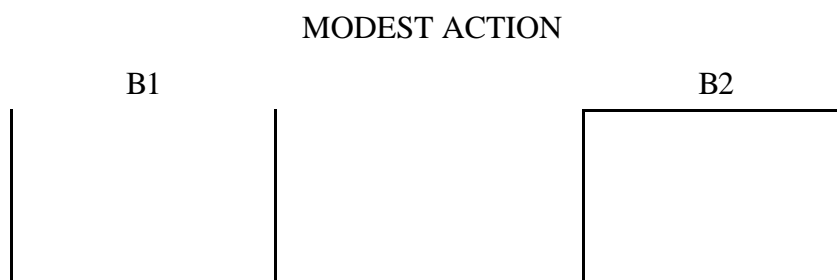
1. Nozick's presentation of the problem

Robert Nozick in his [N] presented the following problem which he attributed to W. Newcomb.

“Suppose a being in whose power to predict your choices you have enormous confidence. (One might tell a science-fiction story about a being from another planet, with an advanced technology and science, whom you know to be friendly, etc.) You know that this being has often correctly predicted your choices in the past (and has never, so far as you know, made an incorrect prediction about your choices), and furthermore you know that this being has often correctly predicted the choices of other people, many of whom are similar to you, in the particular situation to be described below. One might tell a longer story, but all this leads you to believe that almost certainly this being's prediction about your choice in the situation to be discussed will be correct.”

There are two boxes, B1 and B2. B1 contains either $D = \$1,000,000$ or nothing. B2 contains $d = \$1,000$. What the content of B1 depends upon will be described in a moment. You have a choice between two actions.

(M) You may open only the first box and take what is in it. I call this action modest.



(\bar{M}) You may open both boxes and take what is in them. I call this action non-modest.

NON-MODEST ACTION



The being has a choice between two actions:

- (G) If the being predicts you will open only the first box, he does put the \$1,000,000 in the first box. I call this action generous.

GENEROUS ACTION



- (\bar{G}) If the being predicts you will open both boxes, he does not put the \$1,000,000 in the first box. Is call this action non-generous.

NON-GENEROUS ACTION



The situation is as follows. First the being makes its prediction. Then it puts the \$1,000,000 in the second box, or does not, depending upon what it has predicted. Then you make your choice. What do you do? Nozick offered the following arguments.

“There are two plausible-looking and highly intuitive arguments which require different decisions. The problem is to explain why one of them is not legitimately applied to this choice situation. You might reason as follows.

Two Boxes Argument: *The being has already made his prediction, and has already either put the \$1,000,000 in the first box, or has not. The \$1,000,000 is either already sitting in the first box, or it is not, and which situation obtains is already fixed and determined. If the being has already put the \$1,000,000 in the first box, and I take what is in both boxes I get \$1,000,000 + \$1,000, whereas if I take only what is in the first box, I get only \$1,000,000. If the being has not put the \$1,000,000 in the first box, and I take what is in both boxes I get \$1,000, whereas if I take only what is in the first box, I get no money. Therefore, whether the money is there or not, and which it is is already fixed and determined, I get \$1,000 more by taking what is in both boxes rather than taking only what is in the first box. So I should take what is in both boxes.*

One Box Argument: *If I take what is in both boxes, the being, almost certainly, will have predicted this and will not have put the \$1,000,000 in the first box, and so I will, almost certainly, get only \$1,000. If I take only what is in the first box, the being, almost certainly, will have predicted this and will have put the \$1,000,000 in the first box, and so I will, almost certainly, get \$1,000,000. Thus, if I take what is in both boxes, I, almost certainly, will get \$1,000. If I take only what is in the first box, I, almost certainly, will get \$1,000,000. Therefore I should take only what is in the first box.”*

To emphasize the pull of each of these arguments Nozick added the following remarks.

“Two Boxes Argument: *The being has already made his prediction, placed the \$1,000,000 in the first box or not, and then left. This happened one week ago; this happened one year ago. B2 is transparent. You can see the \$1,000 sitting there. The \$1,000,000 is already either in B1 or not (though you cannot see which). Are you going to take only what is in B1? To emphasize further, from your side, you cannot see through B1, but from the other side it is transparent. I have been sitting on the other side of B1, looking in and seeing what is there. Either I have already been looking at the \$1,000,000 for a week or I have already been looking at an empty box for a week. If the money is already there, it will stay there whatever you choose. It is not going to disappear. If it is not already there, if I am looking at an empty box, it is not going to suddenly appear if you choose only what is in the first box. Are you going to take only what is in the first box, passing up the additional \$1,000 which you can plainly see? Furthermore, I have been sitting there looking at the boxes, hoping that you will perform a particular action. Internally, I am giving you advice. And, of course, you already know which advice I am silently giving to you. In either case (whether or not I see the*

\$1,000,000 in the first box) I am hoping that you will take what is in both boxes. You know that the person sitting and watching it all hopes that you will take the contents of both boxes. Are you going to take only what is in the first box, passing up the additional \$1,000 which you can plainly see, and ignoring my internally given hope that you take both? Of course, my presence makes no difference. You are sitting there alone, but you know that if some friend having your interests at heart were observing from the other side, looking into both boxes, he would be hoping that you would take both. So will you take only what is in the first box, passing up the additional \$1,000 which you can plainly see?

One Box Argument: *You know that many persons like yourself, teachers and students, etc., have gone through this experiment. All those who took only what was in the first box, including those who knew of the two boxes argument but did not follow it, ended up with \$1,000,000. And you know that all the shrew dies, all those who followed the two boxes argument and took what was in both boxes, ended up with only \$1,000. You have no reason to believe that you are any different, vis-à-vis predictability, than they are. Furthermore, since you know that I have all of the preceding information, you know that I would bet, giving high odds, and be rational in doing so, that if you were to take both boxes you would get only \$1,000. And if you were to irrevocably take both boxes, and there were some delay in the results being announced, would not it be rational for you to then bet with some third party, giving high odds, that you will get only \$1,000 from the previous transaction? Whereas if you were to take only what is in the first box, would not it be rational for you to make a side bet with some third party that you will get \$1,000,000 from the previous transaction? Knowing all this (though no one is actually available to bet with), do you really want to take what is in both boxes, acting against what you would rationally want to bet on?"*

Nozick added that he has put this problem to a large number of people, both friends and students in class.

"To almost everyone it is perfectly clear and obvious what should be done. The difficulty is that these people seem to divide almost evenly on the problem, with large numbers thinking that the opposing half is just being silly. Given two such compelling opposing arguments, it will not do to rest content with one's belief that one knows what to do. Nor will it do to just repeat one of the arguments LOUDLY and s l o w l y. One must also disarm the opposing argument; explain away its force while showing it due respect."

My aim in this article is to do just that.

2. The resolution of the problem

Two boxes argument and one box argument can both be put in game theoretic terms. You think of Newcomb's problem as a game between you and the being. You have two strategies M and \bar{M} . The being has two strategies G and \bar{G} . Your payoffs are the following:

		the being		
		G	\bar{G}	
you	M	D	0	$d = \$ 1,000$
	\bar{M}	$D + d$	d	$D = \$ 1,000,000$

The being has made its move although you don't not know what it is. It is your turn now. Which strategy do you choose?

Two Boxes Argument is an argument from the *Dominance Principle* (DP). Strategy \bar{M} (two-boxer's strategy) dominates strategy M (one-boxer's strategy), and you should play dominant strategy \bar{M} .

One Box Argument is an argument from the *Expected Utility Principle* (EUP). If you believe that the being almost certainly will have predicted your strategy, e.g. with probability 0.99, then the expected utility depends on the strategy you choose. If you choose M (one-boxer's strategy):

$$EU(M) = 0.99 \times \$1,000,000 + 0.01 \times \$0 = \$ 990,000.$$

If you choose \bar{M} (two-boxer's strategy):

$$EU(\bar{M}) = 0.01 \times \$1,001,000 + 0.99 \times \$1,000 = \$11,000.$$

You should be modest and play M , because $EU(M) > EU(\bar{M})$.

Is it possible that two well respected principles of choice, DP and EUP, are in conflict? I would not say so. I think that EUP is incorrectly applied in *One Box Argument* above.

Let me give you another example. Suppose you are playing a game with four possible outcomes ○, ●, □, ■, with probabilities 0.4, 0.3, 0.2, 0.1 and with utilities \$1, \$2, \$3, \$4:

X	○	●	□	■
$P(X)$	0.4	0.3	0.2	0.1
$U(X)$	\$1	\$2	\$3	\$4

Your expected utility in this game (which is the reasonable payment to play the game) is:

$$EU = 0.4 \times \$1 + 0.3 \times \$2 + 0.2 \times \$3 + 0.1 \times \$4 = \$2.$$

You may ask the following question. Is it more profitable to bet on circles or on squares? I suppose you see it is the same. If you bet on circles you are engaged in the following game:

CIRCLE GAME

X	○	●	□	■
$P(X)$	0.4	0.3	0.2	0.1
$U(X)$	\$1	\$2	\$0	\$0

$$EU(\text{CIRCLE}) = 0.4 \times \$1 + 0.3 \times \$2 = \$1$$

If you bet on squares you are engaged in the following game:

SQUARE GAME

X	○	●	□	■
$P(X)$	0.4	0.3	0.2	0.1
$U(X)$	\$0	\$0	\$3	\$4

$$EU(\text{SQUARE}) = 0.2 \times \$3 + 0.1 \times \$4 = \$1$$

Expected utilities of both games are the same, hence neither of them is more profitable. This is the right answer.

Suppose that someone proposes you the following answer. You should calculate the conditional probabilities of circles and their expected utilities:

QUASI-CIRCLE GAME

X	○	●	□	■
$P(X \text{circle})$	4/7	3/7	0	0
$U(X)$	\$1	\$2	\$3	\$4

$$EU(\text{QUASI-CIRCLE}) = 4/7 \times \$1 + 3/7 \times \$2 = \$10/7$$

Then, you should calculate the conditional probabilities of squares and their expected utilities:

QUASI-SQUARE GAME

X	○	●	□	■
$P(X \text{square})$	0	0	2/3	1/3
$U(X)$	\$1	\$2	\$3	\$4

$$EU(\text{QUASI-SQUARE}) = 2/3 \times \$3 + 1/3 \times \$4 = \$10/3$$

The expected utility of quasi-square game is greater, hence it is more profitable to bet on squares. I hope you agree that this argument does not make sense.

If you are not convinced, think about a simpler game:

X	○	●
$P(X)$	1/3	2/3
$U(X)$	\$6	\$3

$$EU = 1/3 \times \$6 + 2/3 \times \$3 = \$4$$

Is it more profitable to bet on white or on black? I suppose that now you definitely see it is the same:

WHITE GAME

X	○	●
$P(X)$	1/3	2/3
$U(X)$	\$6	0

$$EU(\text{WHITE}) = 1/3 \times \$6 = \$2$$

BLACK GAME

X	○	●
$P(X)$	1/3	2/3
$U(X)$	0	\$3

$$EU(\text{BLACK}) = 2/3 \times \$3 = \$2$$

But if you calculate the conditional probabilities of white and black, and expected utilities corresponding to them, you conclude that it is more profitable to bet on white:

QUASI-WHITE GAME

X	○	●
$P(X \text{white})$	1	0
$U(X)$	\$6	\$3

$$EU(\text{QUASI-WITHE}) = 1 \times \$6 = \$6$$

QUASI-BLACK GAME

X	○	●
$P(X \text{black})$	0	1
$U(X)$	\$6	\$3

$$EU(\text{QUASI-BLACK}) = 1 \times \$3 = \$3$$

This argument does not make sense.

But this same nonsense is exactly what is supposed to be *One Box Argument* for Newcomb's problem. You have four possible outcomes MG , $M\bar{G}$, $\bar{M}G$ and $\bar{M}\bar{G}$, with corresponding utilities D , 0 , $D + d$ and d , where $D = \$1,000,000$ and $d = \$1,000$. If you

choose strategy M you have conditional probabilities $P(MG | M) = P(G | M) = p = 0.99$ and $P(M\bar{G} | M) = P(\bar{G} | M) = \bar{p} = 0.01$, with corresponding expected utility:

QUASI- M GAME

X	MG	$M\bar{G}$	$\bar{M}G$	$\bar{M}\bar{G}$
$P(X M)$	0.99	0.01	0	0
$U(X)$	D	0	$D + d$	d

$$EU(\text{QUASI-}M) = pD = 0.99 \times \$1,000,000 = \$990,000$$

If you choose strategy \bar{M} , you have conditional probabilities $P(\bar{M}G | \bar{M}) = P(G | \bar{M}) = \bar{p} = 0.01$ and $P(\bar{M}\bar{G} | \bar{M}) = P(\bar{G} | \bar{M}) = p = 0.99$, with corresponding expected utility:

QUASI- \bar{M} GAME

X	MG	$M\bar{G}$	$\bar{M}G$	$\bar{M}\bar{G}$
$P(X \bar{M})$	0	0	0.01	0.99
$U(X)$	D	0	$D + d$	d

$$EU(\text{QUASI-}\bar{M}) = \bar{p}(D + d) + pd = 0.01 \times \$1,001,000 + 0.99 \times \$1,000 = \$11,000$$

You should be modest and play M , because $EU(\text{QUASI-}M) > EU(\text{QUASI-}\bar{M})$.

We have seen that this argument does not make sense. You should not choose between QUASI- M GAME and QUASI- \bar{M} GAME but between M -GAME and \bar{M} -GAME:

M -GAME

X	MG	$M\bar{G}$	$\bar{M}G$	$\bar{M}\bar{G}$
$P(X)$?	?	?	?
$U(X)$	D	0	0	0

$$EU(M) = P(MG) \times D$$

\bar{M} - GAME

X	MG	$M\bar{G}$	$\bar{M}G$	$\bar{M}\bar{G}$
$P(X)$?	?	?	?
$U(X)$	0	0	$D + d$	d

$$EU(\bar{M}) = P(\bar{M}G) \times (D + d) + P(\bar{M}\bar{G}) \times d$$

The main problem is to calculate $P(MG)$, $P(M\bar{G})$, $P(\bar{M}G)$ and $P(\bar{M}\bar{G})$. In the formulation of Newcomb's problem it is presupposed that $P(G|M) = P(\bar{G}|\bar{M}) = p = 0.99$ and $P(\bar{G}|M) = P(G|\bar{M}) = \bar{p} = 0.01$. What is missing in the formulation of Newcomb's problem is $P(M)$, the probability that you are modest one-boxer.¹⁾ If $P(M) = m$ were given then:

$$\begin{aligned}
 P(MG) &= P(M) \times P(G|M) = m p & P(M\bar{G}) &= P(M) \times P(\bar{G}|M) = m \bar{p} \\
 P(\bar{M}G) &= P(\bar{M}) \times P(G|\bar{M}) = \bar{m} \bar{p} & P(\bar{M}\bar{G}) &= P(\bar{M}) \times P(\bar{G}|\bar{M}) = \bar{m} p
 \end{aligned}$$

$$EU(M) = P(MG) \times D = m p D$$

$$EU(\bar{M}) = P(\bar{M}G) \times (D + d) + P(\bar{M}\bar{G}) \times d = \bar{m} \bar{p} (D + d) + \bar{m} p d = \bar{m} (\bar{p} D + d)$$

Hence, if you bet on M you are engaged in the following game.

M - GAME

X	MG	$M\bar{G}$	$\bar{M}G$	$\bar{M}\bar{G}$
$P(X)$	$m p$	$m \bar{p}$	$\bar{m} \bar{p}$	$\bar{m} p$
$U(X)$	D	0	0	0

$$EU(M) = m p D$$

If you bet on \bar{M} you are engaged in the following game.

\bar{M} - GAME

X	MG	$M\bar{G}$	$\bar{M}G$	$\bar{M}\bar{G}$
$P(X)$	$m p$	$m \bar{p}$	$\bar{m} \bar{p}$	$\bar{m} p$
$U(X)$	0	0	$D + d$	d

$$EU(\bar{M}) = \bar{m} (\bar{p} D + d)$$

Which one is more profitable for you? It depends on m , p , D and d . The last three parameters are constants prescribed by the formulation of Newcomb's problem. Hence, which game is more profitable for you depends on the value of your m .

It is easy to calculate that M is more profitable for you if and only if your m is higher than $(\bar{p}D + d)/(D + d)$, while \bar{M} is more profitable for you if and only if your m is lower than $(\bar{p}D + d)/(D + d)$. The main point is that *Expected Utility Principle* allows \bar{M} to be more profitable, for suitable m , and that there is no conflict between DP and EUP.

The really important question is what is your m . If you are a rational being who understands DP your m should be 0, which means that you should be a two-boxer. In that case $m < (\bar{p}D + d)/(D + d)$ and EUP also recommends that you should be a two-boxer. DP and EUP do not quarrel about that.

If you are an irrational being who does not understand DP your m could be high, even 1, which means that you could be a one-boxer. In that case $m > (\bar{p}D + d)/(D + d)$ and EUP also recommends that you should be a one-boxer, in spite of DP. But there is no contradiction here, because you are an irrational being not constrained to behave rationally.²⁾

You might be perplexed by the fact that to be irrational is more profitable than to be rational. (Namely, $EU(M)$ for high m is greater than $EU(\bar{M})$ for low m .) But there are a lot of games in which it is more profitable to be irrational. Think about the following offer:

If you do not know that $7 + 5 = 12$ I give you \$1,000,000 but if you do know that $7 + 5 = 12$ I give you nothing.

Or think about the following variant of Newcomb's game:

You are allowed to be a one-boxer only if you do not know that $7 + 5 = 12$ and you are allowed to be a two-boxer only if you do know that $7 + 5 = 12$.³⁾

3. Newcomb's problem and Prisoner's dilemma

I will discuss two more questions. First one is rather technical and addresses the connection between Newcomb's problem and Prisoner's dilemma. Is it really true that Newcomb's problem is simply a version of Prisoner's dilemma, as sometimes asserted (cf. [S] p.70). It is definitely not true and I'm going to prove that.

The general form of Prisoner's dilemma is shown in the following matrix:

		me		
		<i>C</i>	<i>D</i>	
you	<i>C</i>	$(R \ R)$	$(p \ R^+)$	$p < p^+ < R < R^+$
	<i>D</i>	$(R^+ \ p)$	$(p^+ \ p^+)$	

We may cooperate *C* or defect *D*. If we both cooperate we both get the cooperation reward *R*. If you cooperate and I defect you get the cooperation punishment *p* and I get even bigger reward *R*⁺. If I cooperate and you defect it is the other way around. If we both defect we both get the defection punishment *p*⁺ (slightly better than *p*).

The movement diagram below reveals *DD* as the only equilibrium outcome of the game. It means that *D* is the dominant strategy for both of us.

		me			
		<i>C</i>	<i>D</i>		
you	<i>C</i>	$(R \ R)$	\rightarrow	$(p \ R^+)$	$p < p^+ < R < R^+$
	<i>D</i>	\downarrow		\downarrow	
		$(R^+ \ p)$	\rightarrow	$(p^+ \ p^+)$	

The main point of the game is that equilibrium *DD*, which is dominant for both of us, is not Pareto optimal since both of us would be better at *CC*. (Individual rationality which leads to *DD* is contradicted by group rationality which leads to *CC*.)

Let us look at Newcomb's problem. Its general form is shown in the following matrix:

		the being		
		G	\bar{G}	
you	M	D	0	$0 < 0^+ < D < D^+$
	\bar{M}	D^+	0^+	

Your payoffs are given. What about the being's payoffs? They are not given in the formulation of Newcomb's problem. But we do know that the being prefers MG to $M\bar{G}$ and $\bar{M}\bar{G}$ to $\bar{M}G$, i.e. his payoffs should be given by the matrix of the following form:

		the being		
		G	\bar{G}	
you	M	x	x^-	$x^- < x \quad y^- < y$
	\bar{M}	y^-	y	

What we do not know is whether $x < y$, $y < x$ or $x = y$. But even without that knowledge we know that the general form of Newcomb's game is given by the following matrix:

		the being		
		G	\bar{G}	
you	M	$(Dx) \leftarrow$	$(0x^-)$	$0 < 0^+ < D < D^+$
	\bar{M}	$(D^+y^-) \rightarrow$	(0^+y)	
		\downarrow	\downarrow	$x^- < x \quad y^- < y$

Its movement diagram is different from the movement diagram of Prisoner's dilemma. Hence, these are quite different games. Notice that you have the dominant strategy \bar{M} , while the being has no such strategy. In Prisoner's dilemma we both had one. The equilibrium outcome $\bar{M} \bar{G}$ could even be Pareto optimal if $x < y$, i.e. if the being is less willing to reward modesty of one-boxers than to punish non-modesty of two-boxers. It is quite clear that interpreting Newcomb's problem as a kind of Prisoner's dilemma makes no sense at all.

4. How comes that there are one-boxers?

The last question I should answer is how does it happen that quite a number of rational beings who do understand DP (e.g. Newcomb himself, Bar-Hillel, Margalit [BM], Brams [B] etc.) are one-boxers? My answer is that they endorse a wrong interpretation of Newcomb's problem, as will be explained in a moment.

In all interpretations of Newcomb's problem your payoffs are given by the following matrix:

		the being		
		G	\bar{G}	
you	M	D	0	$0 < 0^+ < D < D^+$
	\bar{M}	D^+	0^+	

You do not know which strategy the being has chosen and the problem is which strategy, M or \bar{M} , is more profitable for you.

Let me start with two simple minded and patently wrong interpretations.

(1) The being does not see into the future and does not know which strategy you will choose.

In terms of game theory, you and the being choose your strategies simultaneously, without communication beforehand.

If you apply DP you realize that \bar{M} is more profitable for you, because $D < D+d$ and $0 < d$. If you apply EUP you realize it must be \bar{M} again, because $gD + \bar{g} 0 < g(D+d) + \bar{g} D$

for any $g = P(G)$. It must be \bar{M} any way and there is nothing controversial about that. Of course, Newcomb's problem is very controversial so this interpretation is not acceptable.

(2) The being sees into the future and does know which strategy you will choose. In terms of game theory, you and the being do not choose your strategies simultaneously. You move first, or equivalently, you tell your choice to the being beforehand.

If you choose M the being chooses G and you get D . If you choose \bar{M} the being chooses \bar{G} and you get d . (There is no point to apply DP in this situation, because this principle is designed for simultaneous moves.) Hence, it is more profitable for you to choose M and there is nothing controversial about that. As already said, Newcomb's problem is very controversial so this interpretation is also not acceptable.

The patently wrong interpretations should be refined. Before that, I would like to propose an explanation of the fact that *"to almost everyone it is perfectly clear and obvious what should be done. The difficulty is that these people seem to divide almost evenly on the problem, with large numbers thinking that the opposing half is just being silly."* My explanation is that people switch to one of the interpretations, (1) or (2), and stay close to them even when they refine them realizing that they are not acceptable. I discuss the typical refinement of (2) and then an atypical refinement of (1), arguing that the last one is acceptable, whereas the first one is not. Let me start with the refinement of (2).

(2') The being sees into the future through the fog and knows only with some probability, e.g. $p = 0.99$, which strategy you will choose.

If you choose M the being chooses G with probability $p = 0.99$, \bar{G} with probability $\bar{p} = 0.01$, and you expect to get $pD = \$990,000$. If you choose \bar{M} the being chooses \bar{G} with probability $p = 0.99$, G with probability $\bar{p} = 0.01$, and you expect to get $\bar{p}D + d = \$11,000$. Hence, it is more profitable for you to choose M .

This is the *one box argument*. There is nothing controversial about it, if you accept this interpretation of Newcomb's problem. I suppose this is the refined interpretation that one-boxers like Newcomb, Brams, Bar-Hillel, Margalit etc. have in their minds. But one of the main points of Newcomb's problem is that it is still an interesting problem even if the being is

an absolute predictor, i.e. even if $p = 1$. But then our interpretation (2') reduces to (2) and we have already rejected that interpretation. Hence (2') is not acceptable.

My final interpretation is a refinement of (1):

(1') The being does not see into the future and does not know which strategy you will choose, but the being knows you extremely well (your past behavior, your character, your psychological profile etc.), hence it is extremely good in predicting your choice. In terms of my previous discussion the being is extremely good in predicting your m .

Which strategy is more profitable for you, M or \bar{M} , depends on the value of your m . If you are a rational being who understands DP your m should be 0.⁴⁾ In this interpretation Newcomb's problem is still an interesting problem even if the being is an absolute predictor, i.e. even if $p = 1$. In this extreme case (1') does not reduce to (1). The expected utilities of M and \bar{M} are mD and $\bar{m}d$, in this extreme case, whereas there is no sensitivity to m in case (1). Hence, contrary to (2), (1) has a really interesting refinement (1') and this is the reason why (1') is acceptable and why we can say that it is the right interpretation of Newcomb's problem.

- 1) I think the incorrect application of EUP in *One Box Argument* could originate from this omission. Namely, the problem could be resolved only if $P(M)$ is given. Because it is not given people switch to quasi-resolution which does not need $P(M)$. (But cf. also chapter 4. below.)
- 2) There is a simpler *meta-argument* which proves that \bar{M} is the only rational strategy. Namely, the only rational assumption is that *DP-recommended strategy* = *EUP-recommended strategy* (rather than *DP-r.s.* \neq *EUP-r.s.*) and we have seen that this is true if and only if $m < (\bar{p}D + d)/(D + d) \approx 0$, i.e. if and only if $\bar{m} \approx 1$.
- 3) Note that substituting "DP is a valid principle" for " $7 + 5 = 12$ " restores the original Newcomb's problem.
- 4) And you should bear your envy of the irrational millionaires whose m is 1.

References:

- [B] Brams, S., *Newcomb's problem and prisoner's dilemma* (596-619), *Journal of Conflict Resolution* 19, 1975.
- [BM] Bar-Hillel, M., Margalit, A., *Newcomb's paradox revisited* (295-304), *British Journal of the Philosophy of Science* 23, 1972.
- [N] Nozick, R., *Newcomb's Problem and Two Principles of Choice*, (114-146), in *Essays in Honor of Carl G. Hempel*, ed. N. Rescher et al. D. Reidel, 1969. Reprinted in Robert Nozick, *Socratic Puzzles*, Harvard University Press, 1997.
- [S] Sainsbury, R. M., *Paradoxes*, Cambridge University Press, 1995.