A

James Tartaglia

🏛 Keele University (United Kingdom)

🆔 0000-0002-5251-1908

✉ j.tartaglia@keele.ac.uk

Free will and believing in determinism[1]

📰 **Abstract:**

The article addresses the issue of free will and determinism through a discussion of Newcomb's paradox, presented as a dialogue between the spirits of Lady Luck and Fate. I argue that commitment to determinism, which is suggested by materialist metaphysics, is in contradiction with our experience of freedom of choice. Newcomb's paradox describes the dilemma of choosing between either one or two boxes in order to maximise the quantity of money these boxes contain, which has been determined by the machine predicting what your decision will be. The character of Heather, faced with this dilemma, symbolizes humanity grappling with the issue of free will in the face of determinism. I claim that our experience of freedom prevents us from believing that determinism is true when we properly reflect on the issue, just as Heather could not believe in determinism when she had to make the choice in the situation described in the paradox.

---

1 Adapted from Chapter 3, Part 4 of James Tartaglia's *Inner Space Philosophy: Why the next stage of human development should be philosophical, explained radically (suitable for wolves).* Winchester, UK: Iff Books (publication date: 28 June 2024).

Lady Luck and Fate are spirits having a conversation about free will and determinism. They are trying not to attract the attention of the Philosophy spirit since they find her both frightening and boring. Newcomb's paradox was first brought to the attention of philosophers in a 1969 article by Robert Nozick (Nozick, 1969). The paradox is usually discussed in the context of decision theory, but one of my teachers, J.J. Valberg, used it to draw conclusions about free will and determinism in his unpublished book entitled *Will* (2011)—unlike Valberg, who accepts determinism and is a compatibilist, the conclusion I draw is that determinism must be false.

LADY LUCK: If you promise to keep your voice down, I'll tell you how I know determinism isn't true, why it just can't be. It came as quite a shock to me when I found out because ever since the scientific revolution of the 17th century, I've been encouraged to think of myself as backed up by the deterministic laws of physics, as pretty much the same thing as them, really. It was a new image for me at the time, thoroughly up to date—not everyone's got the wherewithal to remake themselves like that; it's the reason for my longevity. Still, it took some getting used to. One minute, you think you're a mystical lady-spirit playing with human lives as if they were chessmen; the next, you're the inevitable temporal unfolding of natural forces. Well, I never really gave up on the first self-image; it's so much more *me*, but I did throw myself into the 'unfolding of nature' thing. Why not? If you can't beat 'em, join 'em! But it doesn't work; I found that out from a physicist called William Newcomb … a physicist, eh, you'd have thought it'd be a … nope, mustn't say the word.

Newcomb imagined me having perfect knowledge of the physical universe and using it to predict a person's decision. This person – let's call her Heather – is given two choices, and I get to predict which choice she's going to make – which is easy enough when you know everything about the physically determined universe, right? According to my new image, all that vast knowledge was only supposed to be my

own self-awareness anyway … know thyself and all that, no biggie. So, knowing how the bits and pieces of Heather's brain change as she's making her choice, I can work out the effect this'll have on how she'll move her mouth and tongue (i.e., what she'll say when she announces her choice), so I can predict her decision every time, easy-peasy.

Heather's choice is about whether she wants the money in two boxes (B1 and B2) or just one (B2). B1 always contains $1000, but what B2 contains depends on my prediction. If I predict she's going to choose to have box B2 only, then I'll reward her by putting $1,000,000 in it. But if I predict she's going to be greedy by choosing both boxes, then I won't put any money in box B2; she'll just get the $1000 from box B1.

The fun starts once I've made my prediction, and so either put the million in B2 or not, depending on the prediction. Once that's done, Heather has 5 minutes to choose whether she wants both boxes or only one. Well, you can imagine the humming and harring that one causes! Her first reaction is always going to be that she should choose box B2 only; that's just common sense. She knows everything I've told you, the whole shebang, so she figures that she'd better choose B2 only to get the million. She thinks that if she chooses both boxes, then I'll have predicted she would, so she won't get the million.

But then, if she's got an ounce of sense, the next thing that's going to occur to her is:

HEATHER (as remembered by Lady Luck): Hang on a cotton-picking minute; Lady Luck's *already* made the prediction. The prediction is a *past event*! Lady Luck either put the million in the box or she didn't; that's already happened, and nothing can change it now. So I might as well choose both boxes because if the million's already in B2, it'll be there whatever I choose now; I know the extra thousand from B1 isn't much, but there's no risk involved, so I might as well have it if I can. My decision can't change what she's already done, so … I'll have both boxes, please—final answer.

LADY LUCK: Now, you might think this is a kind of paradox—that's how Newcomb saw it—since her motive for choosing only the one box seems just as strong as her motive for choosing both. What I've come to realise, however, is that if Heather is even contemplating the choice of both boxes, then she can't really believe my predictions are always correct. Just thinking about choosing both boxes, once she fully understands the situation, can only mean that she doesn't trust me. And it makes perfect sense for her to choose both if she doesn't trust me, of course, since if she goes for only box B2, then she might end up with nothing—in the scenario where I falsely predicted that she'd choose both boxes, and so didn't put any money in B2.

So, choosing both boxes only makes sense if you don't trust me, or rather, if you don't think determinism is true—that amounts to the same thing for Heather, as for most of my people these days, because they think I'm the inevitable flow of physical reality, or backed up by it, or whatever. But if Heather does trust me, then she'll choose only one box since she'll expect me to have known she'd make that choice, and so she'd also expect me to have already put the million in the box.

To sum up, then, if Heather trusts me, she should choose one box, but if she doesn't, she should choose both. Since she's supposed to trust me, she ought to choose one box.

Well, I'd been thinking this over for some time when I had the bright idea of trying it out experimentally, just to be sure. There was a businesswoman called Heather, whom I was working with at the time—she made billions with an internet company, a nasty piece of work. She told me she was trying to change the future of friendship. Well, I assured her it was her fate to succeed, that the suffering she caused would be outweighed in the long run, that she should ignore all risks … you know, the usual stuff I whisper in their inner spaces. Anyway, one night, I took her to a dream world to try my experiment. When we got to the bit where she had 5 minutes to decide, what I discovered astonished me. It turned out she wasn't even capable of trusting me!

FATE: That's what I was telling you earlier.

LADY LUCK: Shut up a minute, will you? We're just getting to the good bit. You see, Heather was a very clever woman, so she understood the scenario well enough to understand that, at the end of the day, trusting me just doesn't make sense. This is how she explained it—she said it so nicely; I'll repeat her exact words.

> HEATHER (as remembered by Lady Luck): If I'm to trust that you always get the predictions right, then I have a choice between B2 only to get the million dollars, or B1 and B2 to get only a thousand dollars. The choice I make right now will decide how much money I get because whatever that choice is, it'll be the choice you predicted. But that's impossible; I can't make a choice about something which happened in the past. So, I'm afraid I simply can't trust you. It's paradoxical to experience freedom while believing in determinism, and since I can't renounce the experience of freedom—I can't deny that I currently have a choice to make; as a plain matter of fact, I'm here, and I do—so you leave me with no alternative except to renounce determinism.

LADY LUCK: She explained that well, don't you think, Fate? I saw straight away that she was right, so I had to admit to myself that I wasn't really an all-knowing, deterministically unfolding natural world—or, as humans like to think of it, a deterministic physical reality which allows *them* (!) to *in principle* (!) predict what's coming next. It was a bit deflating, I'll admit, but it made a lot of sense since I certainly didn't always get things right when I was a mysterious lady-spirit. Sometimes, I'd tell a military leader he was heading for glory, that his army was invincible, and then … oops!

FATE: Are we talking about the same Heather you handed over to me, the Aussie internet tycoon?

Lady Luck: The very same; now you know why I dropped her.

Fate: She didn't do very well with me either. She kept walking under ladders, didn't salute magpies, didn't have lucky underpants, so hardly a day passed when I wasn't irritated by her. Before long, I snapped and got her to sell up her shares in the company at the worst time possible. Well, even that didn't change her behaviour. She didn't start talking to me more than before, she wasn't begging me for better luck, so I'm afraid I got a little vindictive—I persuaded her to apply for a job as a philosophy lecturer, after that her downfall was assured.

Philosophy: Aha, I thought I heard somebody mention my name … I've been invoked !!!

Lady Luck (groaning): Oh no, not you …

Commentary

I am arguing that we cannot consistently believe that determinism is true once the practical implications of living in a deterministic world are considered. We see Heather's discovery of this in her final speech. If she has complete faith in the prediction machine (Lady Luck / determinism), then she will be fully confident that the machine has accurately predicted whether she will choose to have both boxes or only one. When she makes her decision, then, she will believe that if she chooses one box only, she will subsequently open the box to find $1,000,000 inside, whereas if she chooses two boxes, she will subsequently open the box to find nothing inside. To believe this, however, is to believe that she currently has a choice over whether the money is in the box. From her current epistemic situation, it is as if there were two buttons she must choose between pressing, one which gives her $1,000,000 and one which does nothing. This is something she cannot truly believe, however, because she knows the question of whether the

money is in the box or not has already been settled. She knows that at the moment she makes her choice, the money will either already be in the box or the box will already be empty and that her choice cannot affect this—and yet, if she fully trusts the predictor, it cannot help seeming to her that her choice will indeed affect this, in the sense that it is currently up to her to choose between two different outcomes. If she chooses only one box, after all, it will be because she wanted the money and felt that she needed to do something to get it, i.e. make the right choice. The conclusion of the argument is that while experiencing free will, as we all do, we cannot fully trust the prediction machine and hence cannot consistently believe in determinism. The portrayal of determinism as "Lady Luck" reflects my view that the thesis of determinism, which is typically portrayed as a rather scientific thesis, is actually rooted in ancient superstitions such as astrology (see Tartaglia, 2020, chapter 6, section 4).

Bibliography

Nozick, R. (1969). Newcomb's Problem and Two Principles of Choice. In N. Rescher (Ed.), *Essays in Honor of Carl G. Hempel* (pp. 114–146). Dordrecht: Springer.

Tartaglia, J. (2020). *Philosophy in a Technological World: Gods and Titans.* London: Bloomsbury.