Topic 1: Time Travel and Free Will

Contents

---

1. Time Travel

VIDEO: Welcome to Week 1!

See here for Tina Fey's take on the Time Traveler's Convention.

Consistent and Inconsistent Time Travel Stories

Back to the Future is an inconsistent time travel story.

At the beginning of the film we are told that, on October 25, 1985, George McFly is respected by none, and leads a miserable life. Later in the film we are told that George's son, Marty, travels back in time to 1955 and meets his young father. As a result of their meeting, George conquers his fears, and blossoms into a bold and courageous man. Towards the end of the film, we are told that Marty returns to October 25, 1985, and is faced with a very different situation from the one that was described at the beginning of the film. George is respected by all, and leads a happy life.

It is not hard to see why this story is inconsistent. At the beginning of the film we are told that George has a miserable life at time t. At the end of the film we are told that George has a happy (and therefore non-miserable) life at time t. In other words: what we are told at the beginning of the film conflicts with what we are told at the end of the film.

For a time-travel story to be consistent, it must never give us conflicting descriptions of a single point in the narrative's timeline. If at some point in the story we are told that, on October 25, 1985, the world is thus and so, then this is a fact that must be respected throughout the narrative, regardless of what happens to the characters in the story as they travel through time.

QUESTION: Couldn't the Timeline Change?

A World-Travel Interpretation?

We could try to make the story consistent by thinking of it as a story about travelling between worlds, rather than a story of time travel within a single world. The idea is that, when Marty reaches 88 miles per hour in the DeLorean, he travels to the 1955 of a different universe, and then comes back to the 1985 of that other universe. So George is miserable in 1985 in the original universe, and his alternative-universe counterpart is happy in his own, other-worldly 1985. Then there would then be no inconsistency.

But if that is really what happened, then it seems like we should feel very differently about events in the film than the film itself encourages us to feel. For instance, it seems like the film is supposed to have a happy ending. But on the alternate-universe interpretation, the ending is tragic. At the end, Marty goes to the 1985 of the alternate universe. He seems to think he's been reunited with his real family, and is happy to be home. But he hasn't; he's meeting these alternate-universe copies of his real family for the first time. His real family is still miserable back in his home universe. They are probably even more miserable now that Marty has disappeared. It seems that Marty will live out his days as an impostor in this new universe, his old universe abandoned. This makes *Back to the Future* a total downer!

The Moral

The one-universe interpretation of *Back to the Future* is inconsistent. You might be tempted to think that *all* time travel stories set in a single universe are inconsistent. Not so! But it takes some ingenuity to make a time travel story completely consistent. That's why consistent time travel stories can be fun to watch. Good examples include Twelve Monkeys and The Time Traveler's Wife. (And maybe Primer, though it is so complicated it is hard to tell whether it is consistent or not. Check it out and try to decide!)

Even if not all time-travel stories are inconsistent, the fact that *some* of them are leads some people to think that time-travel is logically impossible. Read on...

VIDEO REVIEW: Consistent and Inconsistent Time Travel Stories

Exercises

1. Determine whether the following narrative can be interpreted as a consistent time travel story (barring travel between worlds): You are an avid gambler. One day at the races you decide to take a risk, and bet all your savings on Easy Goer. Tragically, you lose; Master Charlie wins the race, by a nose. Because of your recklessness, you end up leading a life of hardship. Towards the end of your life, however, you catch a lucky break: an old aunt bequeathes you a time-machine. So you travel back to the time in which you had placed the disastrous bet. You head straight for the tracks, and see your younger self standing in line, anxiously awaiting the chance to place a bet on Easy Goer. You stand next to him and whisper into his ear "I've got some inside info, buddy. Master Charlie's got the goods!" Your younger self acquiesces, and bets his money on Master Charlie. He wins big, and spends the rest of his life in luxury and style.

- Consistent
- Inconsistent
  - This narrative cannot be interpreted as a consistent time-travel story because it makes contradictory claims about your life after the race. Towards the beginning of the narrative we are told that you live in poverty after the race. Towards the end of the narrative we are told that you live in luxury after the race.

2. Determine whether this story can be interpreted as a consistent time travel story (barring travel between worlds):

- Consistent
- Inconsistent
  - Correct Answer

3. Determine whether this story can be interpreted as a consistent time travel story (barring travel between worlds):

- Consistent
- Inconsistent
  - Correct Answer

Optional Activities

- Watch *Twelve Monkeys*, and think about how to interpret it as a consistent time travel story.
- **Advanced Option**: Watch *Primer* and determine whether it can be interpreted as a consistent time travel story or not.

Self Causation

Exercise

Determine whether the following narrative can be interpreted as a consistent time travel story (barring travel between worlds): On the eve of your 21st birthday, an elderly stranger shows up at your doorstep and tells you how to build a time-machine. After several years of work, you succeed in constructing a working version of the machine, and lead a fascinating life of time-travel. Towards the end of your life you use your time-machine to travel to the eve of your 21st birthday and tell your younger self how to build a time-machine.

- Consistent
  - According to this narrative, something is a cause of itself. (Specifically: the encounter of your older and younger selves causes your younger self to build a time machine and, later and life, travel back in time to encounter his younger self.) That means that the scenario depicted by the story is weird, and possibly even incompatible with the laws of physics. But there is nothing outright inconsistent about it. There is no time t and proposition p such that we are told that p is the case at t at one point in the narrative, and told that p is not the case at t at a different point in the narrative. So the narrative can be interpreted as a consistent time-travel story.
- Inconsistent

QUESTION: Isn't That 'Consistent' Case Paradoxical?

The Grandfather Paradox

Consider Bruno. Bruno hates his grandfather, and not without reason. Grandfather was a terrible man: he was a gambler and a drunk. Not just that. Grandfather orchestrated a large-scale fraud, which left hundreds of people in poverty. Nothing would please Bruno more than killing Grandfather. But there is a problem: Grandfather died many years ago, long before Bruno was born. Undeterred, Bruno builds a time machine and travels to September 13, 1937, a time before Grandfather had children, and before he orchestrated the fraud. Nothing will stop him from killing Grandfather now!

After weeks of careful planning, the moment of truth has arrived. Bruno has climbed up a church steeple, and positioned himself at the belfry with a sniper's rifle. Grandfather is on his morning walk. Bruno spots him, and aims his rifle with the precision of an expert gunman. Grandfather stops to tie his shoelaces. The church bells start ringing. Noon has arrived. Bruno caresses the trigger. Grandfather stands still for a moment, as a breeze ruffles the trees nearby. Bruno prepares to shoot. . . What happens next?

Suppose Bruno kills Grandfather. Then Grandfather will never have children. So Bruno's mother will never be born. So she will never have Bruno. So Bruno won't go back in time. So he won't kill Granfather. So if he kills Grandfather, he won't kill Grandfather. Contradiction! Paradox!

So it can't be that Bruno kills Granfather. But what's to stop him?

Here is a tempting thought: it must be that Bruno could never have been in a position to kill Grandfather. He must not have been able to go back in time, contrary to our initial assumption. According to this thought, our story is a *reductio* of the idea of time-travel — that is, it shows that the assumption that time-travel is possible leads to contradiction, and thus shows that time-travel is an inherently contradictory idea, like a square circle.

I don't think that's the right way of thinking about the situation. This is what I think we should say instead. When we hear the story, we know that somehow Grandfather will survive. We know this because it follows from the earlier part of my narrative that Grandfather will live long enough to have children, and the only way for this to happen is for Bruno's 1937 assassination attempt to fail. But why will Bruno fail? It will probably be for some perfectly mundane reason. Perhaps he will lose his nerve at the last minute; perhaps he will get distracted by the barking of a nearby dog; perhaps he will slip on a banana peel; perhaps his trigger will get stuck.

Of course, the issue is not settled by the story so far. Our story offers us no particular clue as to what stops Bruno; in fact, the way the story is told makes any particular reason seem very arbitrary. But arbitrary is not the same as inconsistent. What we know is that, as long as the story is consistent, the assassination attempt will somehow fail. It will fail, because it already has failed.

VIDEO REVIEW: The Grandfather Paradox

Could an all-powerful God make Bruno kill Granfather?

Not even an all-powerful God could make a world in which Bruno kills his grandfather before his grandfather has children.

How do we know this? Because not even an all-powerful God could bring about an inconsistent state of affairs, and a world in which a grandfather gets killed before having children would be inconsistent.

Here is a picture that I sometimes find helpful. Imagine that the way God makes the world is by starting with a grid of space-time points, and then deciding what physical properites to fill each space-time point with (in some God places positive charge; in some She places mass). Since we are assuming that our God is all-powerful, She can fill each space-time point however She likes (or leave it blank). On some ways of filling out space-time points, the world will be such that someone kills his grandfather. On other ways of filling out space-time points, the world will be such that someone travels back in time and kills someone before that person has children. But there is no way of filling out space-time points whereby the world is both such that someone kills his grandfather and such that he does so by travelling back in time to kill him before his grandfather has children.

(*A comparison*: Suppose you are allowed to paint each of the points on a sheet of paper any color you like. Could you draw a figure that is both a circle and a sqaure? You cannot: there is no way of filling out points on a page that would yield such a figure.)

What about Contemporary Physics?

I hope to have convinced you that a world in which Bruno kills his grandfather before his grandfather has children would be inconsistent.

Since the laws of physics are consistent, this means that any scenario consistent with the laws of physics is a scenario in which it is not the case that Bruno kills his grandfather before his grandfather has children. But how does this work? How, exactly, does it come about that paradoxical scenarios get ruled out by the laws of physics? (Why is any scenario in which Bruno travels back in time to kill Grandfather also a scenario in which he either looses his nerve, or slips on a bannana peel, or *something*?)

The first thing to note is that the particular way in which a system of physical laws rules out paradoxical scenarios need not be especially interesting. (To see this, consider the question of how the laws of physics rule out a scenario in which I draw a figure that is both a circle and a square.)

But is there anything more substantial to be said?

The simplest interesting way for a system of physical laws to rule out a scenario in which someone kills his own grandfather before his grandfather has children is for the laws to exclude time-travel altogether. As it turns out, however, it is not clear that this is true of the actual laws of physics, since there are solutions to Einstein's equations which allow for time-travel. (See here for more information.)

Some systems of laws that allow for time travel turn out to exclude paradoxical scenarios in interesting ways. If you'd like to know more, have a look at these toy examples. Even better: have a look at the Bonus Section below, in which MIT physicist Alan Guth tackles the question of whether it is physically possible to build a time machine.

VIDEO: WiPhi Summary

If you'd like to see a fun summary of the material so far, check out the video below. It was animated by Damien Rochford as part of Wi-Phi, a project which works with the Khan Academy to make philosophy more accessible to non-philosophers.

---

2. Free Will

What is it to act freely?

A natural answer is that one acts freely if and only if one could have done otherwise.

This answer consists of two parts: a sufficient condition and a necessary condition:

- **Sufficient Condition for Acting Freely (SUF)**:
  - If a person could have done something other than X, then she acted freely in doing X.
- **Necessary Condition for Acting Freely (NEC)**:
  - If a person acted freely in doing X, then she could have done something other than X.

(SUF) says that having been able to do something other than X is *sufficient* for having acted freely. It entails, for example, that if I could have stayed home rather than attending the piano recital, then I definitely acted freely in attending the recital.

(NEC) says that having been able to do something other than X is *necessary* for having acted freely. It entails, for example, that if someone forced me to attend a piano recital by chaining me to one of the chairs in the concert hall, then I definitely didn't act freely in attending the recital.

Generally speaking, a **sufficient condition** for Y is a condition whose satisfaction is enough to guarantee that Y will occur, and a **necessary condition** for Y is a condition that must be satisfied in order for Y to occur. For instance, a sufficient (but not necessary) condition for being a mammal is being a walrus, and a necessary (but not sufficient) condition for being a mammal is being an animal.

(If you'd like to know more about sufficient and necessary conditions, you can check out this video. If you'd like to do exercises beyond the ones below, you can find some here.)

Exercises

1. Being divisible by 2 is _____ for being divisible by 4.

- NECESSARY
  - A number cannot be divisible by 4 without being divisible by 2. So, it is a necessary condition of a number's being divisible by 4 that it be divisible by 2. There are some numbers, like 6, that are divisible by 2 but not divisible by 4. So being divisible by 2 isn't sufficient for being divisible by 4.
- SUFFICIENT
- NEITHER NECESSARY NOR SUFFICIENT
- BOTH NECESSARY AND SUFFICIENT

2. Being divisible by 10 is _____ for being divisible by 5.

- NECESSARY
- SUFFICIENT
  - If a number is divisible by 10, it has to be divisible by 5. So being divisible by 10 is sufficient for being divisible by 5. Some numbers (e.g., 25) are divisible by 5 but not divisible by 10, so being divisible by 10 isn't necessary for being divisible by 5.
- NEITHER NECESSARY NOR SUFFICIENT
- BOTH NECESSARY AND SUFFICIENT

3. Growing up on the same street is _____ for being best friends.

- NECESSARY
- SUFFICIENT
- NEITHER NECESSARY NOR SUFFICIENT
  - Growing up on the same street is neither necessary nor sufficient for being best friends, since some people who are best friends didn't grow up on the same street, and some people who grew up on the same street aren't best friends.
- BOTH NECESSARY AND SUFFICIENT

4. A bouncer forces you to leave the club by pushing you out the door. Under such circumstances, it is natural to think that you did not act freely in leaving the club. Which of the following principles could be used to deliver this result?

- (SUF)
- (NEC)
  - When X is leaving the club, (NEC) tells us that the only way of acting freely in leaving the club is by being such that you could have done otherwise. But in this case you couldn't have done otherwise because you were being pushed out the door by the bouncer. So (NEC) can be used to get the result to that you didn't act freely in leaving the club.
- Both
- Neither

5. After several hours of good fun, you decide to leave the party. Unbeknownst to you, your hosts were finding you extremely obnoxious, and were about to kick you out just as you decided to leave. So, had you instead decided to stay, you wouldn't have been able to: you would have been forced to leave anyway. Does (NEC) entail that you didn't act freely in leaving the party?

- Yes
  - Yes. (NEC) entails that you didn't act freely in leaving the party, since you couldn't have done otherwise. Is this the right result? Some people think it is not. For further discussion, see the optional section 1.4 on Frankfurt cases.
- No
- Question is ambiguous

6. Ulysses ordered his men to bind him to the mast of his ship. When the Sirens arrive, he desperately tries to follow them into the sea, but is unable to do so because he is tied to the mast. So he stays on board. Does (SUF) entail that Ulysses acted freely in remaining bound to the mast?

- Yes
- No
  - (SUF) does not entail that Ulysses acted freely (since once he was bound to the mast he couldn't have done anything other than remain bound to the mast). Notice, however, that Ulysses may well have acted freely in asking his men to tie him to the mast. If so, there is a certain sense in which he acted freely in staying aboard his ship.
- Question is ambiguous

Does Bruno Act Freely?

Here, again, are (SUF) and (NEC):

- **Sufficient Condition for Acting Freely (SUF)**:
  - If a person could have done something other than X, then she acted freely in doing X.
- **Necessary Condition for Acting Freely (NEC)**:
  - If a person acted freely in doing X, then she could have done something other than X.

(SUF) and (NEC) together entail:

`Bruno acted freely when he shot and missed if and only if he could have done otherwise.`

So: could Bruno have done otherwise?

An argument

Let us add some details to our story. Let us suppose that the assassination attempt was well-documented in the local press. It is known that the would-be assassin fired a single shot. The bullet missed Grandfather — but only barely: it burnt a scar on his left cheek.

Bruno has seen that scar in photographs on countless occasions. He can remember it clearly. "A hideous scar on a hideous face," he thinks, as he takes his position at the belfry, and loads a single bullet into his rifle. He prepares to take aim, and remembers one of the old newspaper articles about the shooting. He has read it so many times that he can recite the basic facts from memory: "The assassination attempt occurred just after noon, on September 13, 1937. The church bells were ringing. A single shot was fired. The bullet missed, but only barely, burning a scar on Grandfather's left cheek." Bruno's train of thought is interrupted when the church bells start ringing. Noon has arrived. Bruno caresses the trigger. Grandfather stands still for a moment, as a breeze ruffles the trees nearby. Bruno prepares to shoot...

On this version of the story, we know exactly how Bruno will act: he will take a single shot and miss—but only barely. This means that any possibility consistent with *what we know* about the rest of the story is a possibility in which he does not kill Grandfather. So it is tempting to think that Bruno couldn't have done otherwise. And, of course, someone who accepts (NEC) would claim that Bruno only acts freely in shooting and missing if he could have done otherwise. So someone who accepts (NEC) might be tempted to think that Bruno failed to act freely.

Does the argument succeed?

As it turns out, there are two big problems with the argument above.

The **first problem** is that it is not clear that (NEC) is correct. (You can check out the optional discussion on Frankfurt cases if you'd like to know more).

The **second problem** is that, even if one assumes that (NEC) is correct, it is not clear that the sense of 'possibility' in which it is right to say that Bruno could have done otherwise is the sense of 'possibility' that is important for free will, and that is relevant to (NEC).

To see this, it is useful to consider an analogy. We meet a friend for breakfast in New York and she tells us that, fed up with the noise and the rising housing prices, she attempted to leave on a train-trip to Alaska last night. We, her listeners, know that she did not succeed in making the trip to Alaska; here she is in New York, after all, telling us her story. So any possibility consistent with what we know about the actual situation is a

possibility in which she did not succeed in making the trip to Alaska. But none of this entails that our friend failed to act freely, or that her trip had to fail.

As our friend tells us the story, we the listeners have information about the story's future: we know that our friend will fail to make the trip to Alaska, despite her attempt. There is therefore a sense of 'possible' in which we—the listeners—could say that it is not 'possibility' for the trip to succeed. But this is not the sense of 'possible' that is relevant to free will. What matters for free will is whether our friend was in a position to act otherwise at the time, regardless whether or not we — the listeners — happen to know that things won't actually turn out that way.

The lesson of our example is that a scenario can count as a 'possibility' for the purposes of free will, even if we are able to exclude it on the basis of information we have about the story's future.

Now return to the case of Bruno. We — the readers — have information about the story's future: we know that Grandfather will live to have offspring. So, again, there is a sense of 'possible' in which we — the readers — could say that it is not 'possible' for the assassination attempt to succeed. But, as before, this is not the sense of 'possible' that is relevant to free will. As in the case of the train-catching friend, what matters is whether Bruno was in a position to act otherwise at the time (for instance, whether he was in a position to shoot slightly to the right and hit Grandfather between the eyes), regardless of whether we — the readers — are able to exclude such possibilities on the basis of information we have about the story's future.

We know enough about the story to know that Bruno's assassination attempt will, in fact, fail. But why should we accept the further claim that there isn't an alternative possibility (in the sense of 'possibility' that matters for free will) in which Bruno decides to to aim slightly to the right and the assassination attempt succceeds? If there is such an alternative possibility, then (SUF) entails that Bruno acted freely when he failed to kill Grandfather.

Exercises

1. Suppose we learn that your train-chasing friend bought a ticket, but we have no idea whether she got past security. Should one expect the following to be true? If she got past security, she managed to make the trip to Alaska.

- Yes
- No
  - No. One should expect the statement to be false. We know that she is in New York the next morning. So if she did get past security, there must have been a different reason why she failed to make the trip to Alaska.
- Maybe

2. We later learn that she didn't get past security. Should one expect the following to be true? Had she gotten past security, she would have made the trip to Alaska.

- Yes
- No
- Maybe
  - We know that she didn't get past security. On the basis of the information we have, it's hard to know what would have happened had she made it through. Maybe she would have made the trip, maybe she wouldn't have. The story just doesn't give us enough information.

3. Here is the story of Death in Damascus One day a travelling merchant met Death in the souks of Damascus. Death appeared surprised, but she quickly recovered her characteristic cool and intoned with cadaverous solemnity, "Prepare yourself; I am coming for you tomorrow."

The merchant was terrified, and fled that very night to Aleppo.

The next day, the merchant woke up and — horror of horrors! — found Death at her bedside. Her voice quaking, she managed to squeak, "I thought you were looking for me in Damascus!"

"No, I was merely shopping in Damascus," said Death. "That's why I was surprised to see you: it is written that our final meeting is in Aleppo."

Could the merchant have stayed in Damascus, and avoided her date with Death? Think about this for a moment before reading on...

Whether the merchant could have avoided her date with Death is not fully specified by the story so far, but here's a way of interpreting the story so that, yes, the merchant could have avoided her date with Death; it's just that she wasn't going to, and Death knew this.

The idea is to think of the merchant as analogous to your friend who tries to go to Alaska, and as Death as analogous to you. Death knows that she will meet the merchant on the appointed date the way that you know that your friend will fail to make the trip to Alaska. But it doesn't follow that the merchant had to meet Death in Aleppo; just as it doesn't follow that your friend had to stay in New York. If the merchant had stayed in Damascus, she might have avoided Death; similarly: had your friend gotten on the train she might have succeeded in getting to Alaska.

Suppose we interpret the story in such a way that the merchant could have stayed in Damascus. Does (SUF) entail that she acted freely in travelling to Aleppo?

- Yes
  - (SUF) is the claim that if you could have done otherwise in doing F, then you acted freely in doing F. So, when we interpret the story in such a way that the merchant could have stayed in Damascus, (SUF) entails that she acted freely in travelling to Aleppo.
- No
- Maybe

Determinism and Free Will

For the laws of nature to be **determinisitic** is for them to allow us to deduce from a full description of the state of the universe at any given time, a full description of the universe at any future time.

On one standard interpretation of quantum mechanics, the laws of quantum mechanics are fail to be deterministic. So there is some reason for thinking that we actually live in a non-determinsitic world. (The issue is not at all settled, however, because there are also deterministic interpretations of quantum mechanics.)

Regardless of whether the actual laws of nature turn out to be deterministic, we can consider the question of whether determinism is compatible with free will. More specifically, we can consider the following question:

`Assuming (NEC) is true, could somone act freely in a world with deterministic laws?`

The answer, as we shall see, depends partly on what one takes a law to be.

Humeanism

On one way of thinking about about laws, a law of nature give us information about how the world will, in fact, develop. But it does not tell us that the world *must* develop that way. More specifically. it does not tell us that alternate futures are not *possible*, in the sense of 'possibility' that is relevant to free will; it just tells us that alternate futures will not, in fact, come about.

This is sometimes called a **Humean** conception of the laws, after philosopher David Hume. On a Humean conception of the laws of nature, the answer to our question may well be 'yes': determinism might well be compatible with free will, even if (NEC) is true.

One way to see this is to consider the following case, which is analogous to the case of Bruno and the case of your Alaska-bound friend.

Suppose that we know the initial conditions of the universe, and suppose that the laws of nature enable us to deduce that Susan will raise her right arm on October 7, 2019, at 11:41:23 am (EST). This gives us information about the future. There is, therefore, a sense of 'possibility' in which we — we, who know the laws and the initial conditions — could say that it is not 'possible' for Susan to fail to lift her arm. But, as in the case of Bruno and in the case of your Alaska-bound friend, this is not the sense of 'possible' that is relevant to free will. What matters for free will is that Susan be in a position to act otherwise, whether or not we are able to exclude this possibility on the basis of what the laws has taught us about the future.

On a Humean conception of the laws, knowing the laws of nature is like befriending a time-traveler. The friend travels to the future, and returns to give us information about what the future is like. The laws of nature differ from the time-travelling friend in that they give us information about the future by using information about the past and present. But in both cases the information is information about what the future will actually bring, and not necessarily information about what it is possible and not possible for the future to bring, in the sense of 'possibility' that is relevant for free will.

Moral: if Humeanism is right, then determinism need not be incompatible with free will, even on the assumption that (NEC) is true.

A Shameless Plug

David Lewis has famously argued for a Humean conception of the laws of nature. (See his 'Humean Supervenience Debugged'.)

My own view is that Lewis is *almost* right, but not quite. If you'd like to learn more, you can check out my book. (It's a book written with professional philosophers in mind, so the material is more challenging than anything we've discussed here. . .)

---

3. More on Free Will

What Alternative Possibilities are Relevant?

As I noted earlier, there is some reason for doubting whether (NEC) is correct. (If you'd like to know more, see the optional discussion of Frankfurt in section 1.4.)

In contrast, I see no obvious reason for doubting (SUF):

- If a person could have done something other than X, then she acted freely in doing X.

As it stands, however, (SUF) is slightly under-determined. For (SUF) tells us that if there is a *possibility* in which someone does something other than X, then she is acts freely in doing X. But we are not told what sorts of scenarios are to be counted as possibilities.

We made a little progress on this issue earlier, when we noted that something can count as a possibility for the purposes of (SUF) even if we — we, the theorists — are in a position to exclude it on the basis of information we have about the actual future of the relevant situation. But that still leaves matters quite open.

Suppose, for example, that I am chained to my chair. Am I free to walk away? Intuitively, not. But there need be no *logical contradiction* in a scenario in which a miracle occurs, my chains disappear, and I stand up and walk away. So there is a *logical possibility* in which I do walk away. Since we don't want such a possibility to count as relevant for the purposes of free will, we want to make sure it is somehow excluded when it comes to applying (SUF).

How might one revise (SUF) so as to make clear which possibilities are relevant?

I don't think there's an easy answer to this question, but my guess is that (SUF) should be revised roughly as follows:

- **Revised Sufficient Condition for Free Action (RSUF)**:
  Suppose there is a possibility satisfying the conditions below in which a person does something other than X. Then the person acted freely in doing X.
  - The alternate possibility is just like the actual world at all times prior to the relevant action.
  - At the time of the relevant action, it's okay for the alternate possibility to include a 'miracle': a violation of the actual physical laws. After the 'miracle' has taken place, however, events must continue to take place in accordance with the laws of physics.
  - The 'miracle' must occur entirely within the subject's brain, and it must be such as to not change her psychological profile. (It must, for example, leave her beliefs and desires facts intact.)
  - The 'miracle' must cause the subject to make a decision different from the one she actually made, and it must be the case that, as a result of this decision, she does something different from what she actually did.

VIDEO REVIEW: The Sense of 'Could' Relevant to Free Will

Exercises

1. A scenario is *logically possible* if it can be described without contradiction. A scenario is *physically possible* if it involves no violation of a physical law. Describe a scenario which is logically possible but not physically possible.

- Suppose it is a law of physics that nothing can travel faster than the speed of light. Then a scenario in which I travel from Earth to Alpha Centauri in two minutes is incompatible with the laws of physics, and therefore physically impossible. But it is not logically impossible because my travelling to Alpha Centauri in two minutes does not entail a logical contradiction.

2. Describe a scenario that is possible in the sense relevant to action, according to (RSUF), but not physically possible.

- The answer to this question depends a bit on what exactly the physical laws are like. But suppose the laws are deterministic. Then, if I in fact wore a pink shirt today, a scenario in which the initial conditions of the universe are the same, but I wear a blue shirt today, is inconsistent with physical laws. But such a scenario is still possible in the sense relevant to action, according to (RSUF), because

that scenario is just like the actual scenario, except for a small miracle that occurs in my brain when deciding what shirt to wear today. Nothing else has to change because I have a blue shirt in my closet, right next to the pink one, you see.

Disappointed?

A common reaction to this account of the possibilities relevant to free will: something seems disappointing about it. Maybe the videos below can help explain why.

VIDEO: This Is Not Physics

VIDEO: Maybe We're Changing the Subject

---

4. Frankfurt Cases

Advanced Topic [Optional]

In our discussion of free will we considered a necessary condition on free action:

- **Necessary Condition for Acting Freely (NEC):**
  If a person acted freely in doing X, then she could have done something other than X.

A philosopher named Harry Frankfurt proposed a famous counterexample to (NEC). (These days, counterexamples of this kind are called 'Frankfurt Cases'.)

Consider Ilda, who is about to commit a terribly evil act: she is going to kick a puppy. A dachshund puppy. She imagines the upcoming kick with glee; she rubs her hands in anticipation. "It's going to be so great!" she says to herself.

Unbeknownst to Ilda, she is being monitored by a genius neuroscientist: Dr. Evil. Using advanced technology, Dr. Evil is constantly scanning Ilda's brain to see if she is going to go through with her intention to kick the puppy. If, at any time, Ilda shows signs of wavering, Dr. Evil is ready to intervene: she will use her powerful mind-control rays to make Ilda kick the puppy after all.

But, as it happens, Dr. Evil never needs to intervene, as Ilda never waivers. From beginning to end, Ilda is set on kicking the puppy, and she does.

Now, Ilda could not have done other than kick the puppy; had she tried, Dr. Evil would have immediately intervened and made her kick the puppy anyway. But it is natural to think that Ilda nonetheless acted freely when she kicked the puppy. After all, Ilda is obviously responsible for kicking the puppy — we all think Ilda can be held accountable for her terrible act. But if she's responsible, she had to act freely.

What this shows, says Frankfurt, is that it is possible to not be able to act otherwise, but still act freely. That is exactly what happened in Ilda's case.

Exercise

Construct another Frankfurt case. This time, no mind control. [Hint: when we first introduced (NEC) in section 2, we considered an exercise involving a party. . .]

- You decide to leave the party, and you do. You had been annoying your hosts, however. So had you instead decided to stay, they would have forced you to leave anyway.

Frankfurt's Theory of Action

If Frankfurt is right, and (NEC) is false, when do you act freely? Frankfurt himself went on to develop a theory. His theory is that whether you act freely is not about what possibilities are open to you; instead, it's all about having the right kind of mental state when you act.

What kind of mental state is that?

Here's the idea. A person has many desires: a desire to go get ice cream, a desire to be good at football, a desire to finish her homework, and so on. Some desires are about desires: you might desire, for instance, to desire to read great works of literature. Maybe you don't, in fact, desire to read great works of literature; maybe you find them super boring. But maybe you wish you did desire to read great works of literature. You wish you were the kind of person who enjoyed that sort of thing.

Frankfurt calls desires like that, which are themselves about desires, **second-order desires**. Frankfurt's theory (not exactly, but roughly speaking) is that you act freely when you act in accordance with your second-order desires. When you act that way, according to Frankfurt, you are acting in a way that you, yourself, endorse. You are not being just pushed around by random desires; you are instead the author of your own actions, and hence you are acting freely.

This theory seems to get the right result in many cases. Consider Maria, who is addicted to heroin. Maria desires heroin a lot: that's part of her addiction. But she doesn't desire to desire heroin. She hates that she desires heroin. She wishes for nothing more than to kick the habit. She fights her addiction with all her might; but sometimes the addiction wins out, and she finds herself shooting up nevertheless. Maria, it seems, does not act freely when she shoots up. And that is exactly what Frankfurt's theory says. So, point Frankfurt.

But there are some cases which don't fit quite so well with the Frankfurt theory. Consider the case of Huck Finn. In Mark Twain's story, Huck is travelling with an escaped slave named Jim. Huck and Jim are good friends, and help each other out. But Huck feels bad about helping out Jim. He was raised in Mississippi, before the civil war, in a slaveholding society. Everyone around him has taught him that it is right and proper to own slaves, and that helping an escaped slave is like theft: you are running away with someone's property. Huck believes it is unethical to help Jim out. So he desires to not desire to help Jim. Nevertheless, Huck's compassion for Jim wins out, and he does help Jim, contrary to his second order desire.

It seems like Huck is morally praiseworthy for helping Jim. When he helps Jim, it seems like it is his real, compassionate self shining through the evil, nineteenth century, slaveholding society indoctrination. So it seems like Huck is responsible for what he does. So, arguably, Huck is acting freely when he helps Jim. But that's not what Frankfurt's theory says.

A Question to Ponder

Is the Huck Finn case a decisive counterexample to Frankfurt's theory?

In thinking about this question, you might consider watching this video, by Yale philosopher Joshua Knobe.

---

5. Further Resources

- An early version of these lecture notes was published in Spanish as Rayo, A. 'Viajes a través del tiempo: ¿Qué nos enseña la ciencia ficción acerca del determinismo y el libre albedrío?', Investigación y Ciencia, October 2009.
- Many of the ideas discussed above were originally set forth by David Lewis, who was a professor of philosophy at Princeton University until his untimely death in 2001. If you'd like to read more about time travel, I recommend his 'The Paradoxes of Time Travel'. If you'd like to learn more about determinism and free will, I recommend his 'Are We Free to Break the Laws?'.

- For an excellent online resource on time-travel, check out John Carroll's A Time Travel Website.
- For a fun survey of a lot of the philosophy literature on time travel, see Frank Artzenius's 'Time Travel: Double Your Fun'.
- I suggested above that determinism may not be incompatible with free will. Not all philosophers agree with this view, however. See for instance, Peter van Inwagen's 'The Incompatibility of Free Will and Determinism'.
- The Death in Damascus story, in one version or another, has existed for a long time, but it was introduced to the philosophical literature in a paper on decision theory by Allan Gibbard and William Harper called 'Counterfactuals and Two Kinds of Expected Utility'.
- For a blog devoted to free will, check out gfp.typepad.com.
- The problem of finding a sense of 'possible' appropriate for action is closely related to the problem of finding a sense of 'possible' that gets our judgments about what would would have happened, had things gone differently, right. Our old friend David Lewis tackles this problem in 'Counterfactual Dependence and Time's Arrow'.
- The classic paper in which Frankfurt Cases were introduced is 'Alternative Possibilities and Moral Responsibility'.
- Frankfurt's first presentation of his theory of free will is 'Freedom of the Will and the Concept of a Person'.
- Nomy Arpaly discusses Huck Finn to make points closely related to the one above in her seminal book *Unprincipled Virtue*).
- For full versions of this week's lectures, see the next units in this subsection.

VIDEO: Lecture 2 in Full

VIDEO: Lecture 3 in Full

---

6. Bonus Section: Meet the Expert

Bonus: How to Build a Time-Machine

Is it physically possible to build a time-machine? Prof. Alan Guth tackles that question in the video below.

VIDEO: Alan Guth on Time-Travel