Image: Mireille Clavien

Free Will and Control

Are we puppets?

Causation ≠ Determination

- Causation (or "efficient" causation)
  - "***C*** caused ***E***" means ***C*** brought ***E*** about, or made it happen. A cause, we might say, is a *source* of the effect.

- Determination (or "physical" determination)
  - If ***C*** determines ***E***, then ***E*** *must* occur, given that ***C*** occurs.
  - In every possible world where ***C*** occurs, ***E*** also occurs.

Don’t mix them up, or you get this mess …

> ...the well-known dilemma of determinism. One horn of this dilemma is the argument that if an action was **caused or necessitated**, then it could not have been done freely, and hence the agent is not responsible for it. The other horn is the argument that **if the action was not caused, then it is inexplicable and random**, and thus it cannot be attributed to the agent, and hence, again, the agent cannot be responsible for it. In other words, if our actions are caused, then we cannot be responsible for them; if they are not caused, we cannot be responsible for them. Whether we affirm or deny necessity and determinism, it is impossible to make any coherent sense of moral freedom and responsibility.

- Paul Russell, *Freedom and Moral Sentiment*, 1995, p.14

Features of causation

1. Causes and effects must actually occur, i.e. have “real existence”. Something isn’t caused unless it occurs. And merely possible events, that don’t actually occur, never cause anything.
2. A cause must be prior to (or perhaps simultaneous with) its effect. Future events cannot cause past ones.
3. Causation is mysterious.

Contrasting features of determination

1. One possible event may determine another, even if neither event actually occurs.

   E.g. my dropping this Ming vase 6 feet above a concrete floor determines the breaking of the vase. (I did not drop the vase.)

2. Future events can determine past events.

   E.g. the past states of the solar system can be calculated from the present state, using Newton's laws.

3. Determination is well understood.

Determination

- Event A determines event B if and only if B is a logical consequence of A, together with the laws of physics.

  "A determines B" iff (Laws & A) $\Rightarrow$ B

  So determination is logical necessitation, given the laws of physics.

Determin*ism*

- Philosophers generally assume that every event in space-time has a cause.

- The doctrine of determinism goes further than this, and says that every event is *determined* by its prior causes.
- In other words, given the past and the laws of physics, only one future is logically possible.

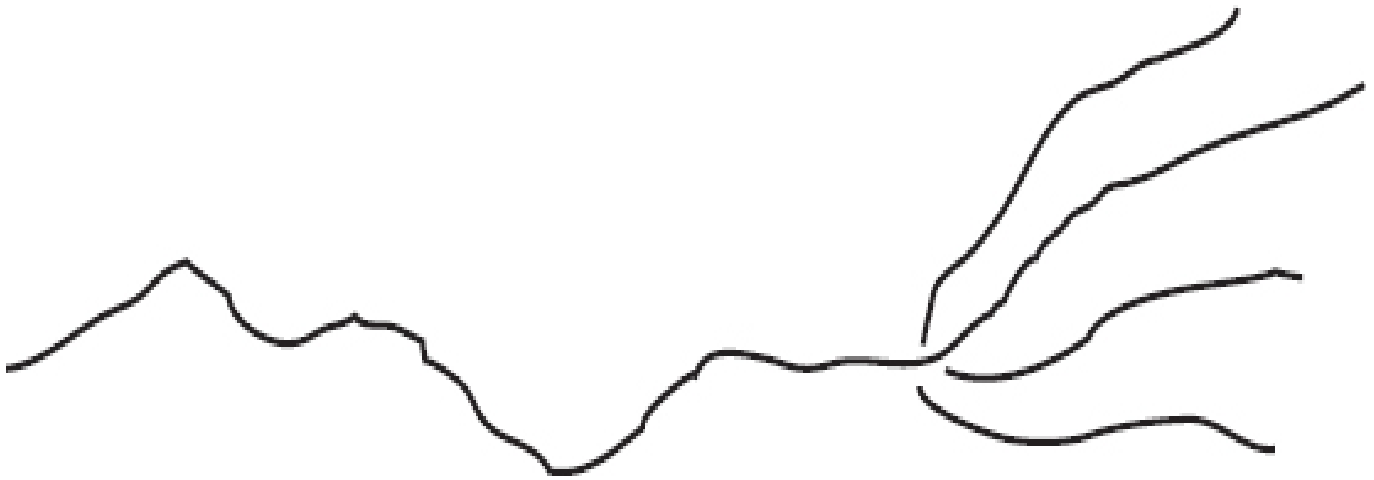
Laplace's Demon

"An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes."

(Pierre-Simon Laplace, *A Philosophical Essay on Probabilities,* 1812)

Forking paths

- “Thus, according to determinism, although it may often seem to us that we confront a sheaf of possible futures, what we really confront is something like this:”



Simple argument for incompatibilism

1. Free will requires that, when a person is deciding what to do, there are at least two possible choices, and hence two possible futures that result from those choices.
2. According to determinism, there is only one possible future, that's compatible with the actual past (up to that moment) and the laws of physics.

If determinism is true, then we have no free will

Assumptions

- This argument assumes that the past, for a given time *t*, is 'fixed' in some sense. And so are the laws of physics.
- This 'fixedness' doesn't mean that the past or the laws are *necessary*. Maybe the way things turned out last year were (at the time) partly a matter of chance?
- Rather, the idea is that present actions *have no effect on* the past, since causation flows from the past to the future.
  - (Rather like a chemical plant discharging into a river-- it has no effect on the *upstream* water quality.)
- Also we cannot change the laws.

Compatibilism

- *Compatibilism* is the view that free will is compatible with determinism.
  - I.e. compatibilism says that even (and perhaps only) a perfectly deterministic system can have free will.

- The most basic disagreement among philosophers, concerning free will, is the one between compatibilism and incompatibilism.

Incompatibilism

- *Incompatibilism* says that free will is incompatible with determinism.
  - A deterministic system cannot have free will
  - I.e. a system with free will cannot be deterministic

- The most important argument for incompatibilism is the ‘consequence argument’ of van Inwagen.

Compatibilism is counter-intuitive

"This is a wretched subterfuge with which some persons still let themselves be put off, and so think they have solved, with a petty word-jugglery, that difficult problem, at the solution of which centuries have laboured in vain, and which can therefore scarcely be found so completely on the surface."

- Kant, talking about compatibilism, *Critique of Practical Reason*, 1788.

William James on compatibilism:

"quagmire of evasion"

"eulogistic terminology,"

"mere word-grabbing game played by the soft determinists."

"they make a pretense of restoring the caged bird to liberty with one hand, while with the other we anxiously tie a string to its leg to make sure it does not get beyond our sight."

- William James, "The Dilemma of Determinism", 1884

Simple version of compatibilism

a future is "open" to an agent if, given that the agent *chose* that future, it would come to pass.

- the reason we are interested in "open" futures is that we are interested in *modifying the way people behave*.
- One important way in which we modify behavior is by punishing behavior we dislike. (Rewards too.)
- Punishing people is effective (and only effective) at causing people to make different choices in the future.
- Free action (could have done otherwise) = *punishable* action

The notion of *control*

- van Inwagen's consequence argument focuses on the notion of control.
- The argument assumes that even a compatibilist has to distinguish the states of affairs that are under a person's control from those that are outside one's control.
- van Inwagen defines an *untouchable* state of affairs as one that **not possibly within my causal influence**.
  - "*x* is and always was unable to do anything about *y*, and *x* would never have been able to do anything about *y*, no matter what knowledge *x* might have had and no matter how lucky *x* might have been"

The ‘Principle’

- If Up and U($p \rightarrow q$), then Uq.

- N.B. one consequence of the Principle is that:

(Up & Uq) $\Rightarrow$ U(p & q) (*Agglomeration*.)

(The proof of this begins by noting that the following sentence is logically necessary and hence untouchable: ($p \rightarrow (q \rightarrow (\mathrm{p} \wedge \mathrm{q}))$)

Proof of Agglomeration

Premises: If U(p) and U($p \rightarrow q$), then U(q).

U(($p \rightarrow (q \rightarrow (\mathrm{p} \wedge \mathrm{q})$))

Conclusion: If (Up & Uq) then U(p & q)

Original version (1975)

- In the case of an event *p* that actually occurred, "I had control over *p*" was originally understood by van Inwagen as "I had the power to render *p* false".
  - (And "I can render p false" was defined to mean there is some action I can perform, the performance of which is *nomically sufficient* for ¬p.)
  - "I cannot render p false" was often written N(p), and 'N' was called the 'power necessity' operator.
  - All this seems to mix up causation and determination!

Problem for Agglomeration

- Suppose an indeterministic coin was *not* tossed, and hence *did not land heads*.
- Let proposition ***p*** **= 'the coin landed heads'** (so $\neg p$ is true as things stand)

- At one time I could have tossed the coin, let's say. Did I have the power, at that time, to render $\neg p$ false?

- If I had tossed the coin, then it *might* have landed heads, so that my tossing it *might* have led to $\neg p$ being false, but this action *isn't nomically sufficient* for $\neg p$ being false.

Agglomeration fails

- So, according to van Inwagen's definitions, at no time did I have the power to render $\neg p$ false.
- Hence, N$\neg p$ holds.

- Also, if we let ***q*** **= '<u>the coin landed tails</u>'**, then similar reasoning shows that N$\neg q$ holds.
- But $(\neg p \ \& \ \neg q)$ is equivalent to $\neg(p \vee q)$.
- And if I had tossed the coin, then $(p \vee q)$ would have been true, so $\neg(p \vee q)$ would have been false.
- Hence N$(\neg p \ \& \ \neg q)$ is false.

Problem fixed

- New version: y is an “untouchable” fact for x

- “*x* is and always was unable to do anything about *y*, and *x* would never have been able to do anything about *y*, no matter what knowledge *x* might have had and no matter how lucky *x* might have been”

- In this sense, the coin’s not landing heads is *not* an untouchable fact, and the problem is avoided.

The consequence argument

1. The laws of physics (L), and ancient history (H) are both untouchable.
2. (L $\wedge$ H) logical entails a description of any voluntary action P. (If determinism is true.)
3. If *x* is logically necessary, then *x* is untouchable.
4. The Principle: If *x* and *x* $\rightarrow$ *y* are both untouchable, then *y* is untouchable.

$\therefore$ 5. (L $\wedge$ H) is untouchable. (from 1, and Agglomeration)

$\therefore$ 6. (L $\wedge$ H) $\rightarrow$ P is untouchable (from 2 and 3)

$\therefore$ 7. P is untouchable. (from 4, 5 and 6)

The control argument

- What does it mean for one system to *fully control* another? How should this be defined?

- X fully controls Y iff the movements of X both *cause* and *determine* those of Y?

- In that case, determinism entails that every person is fully controlled by the past.

Designer and Tool (Producer and Victim)

"Designer designs Tool (in some of the stories, in the way the maker of a robot designs his robot or a god creates a human being; in other stories, by employing techniques of behavioral engineering)."

(SEP: "Arguments for Incompatibilism")

- If Designer both causes and determines what Tool does, then Designer is fully in control of Tool, and Tool is not morally responsible for anything he does.
- If determinism is true, then every human is just like Tool, in all relevant respects.

Part 2

Indeterministic systems can’t be free either (?)

“incompatibilism also hides a mystery”

- “… the indeterminism that seems to be required by free will seems also to destroy free will.”

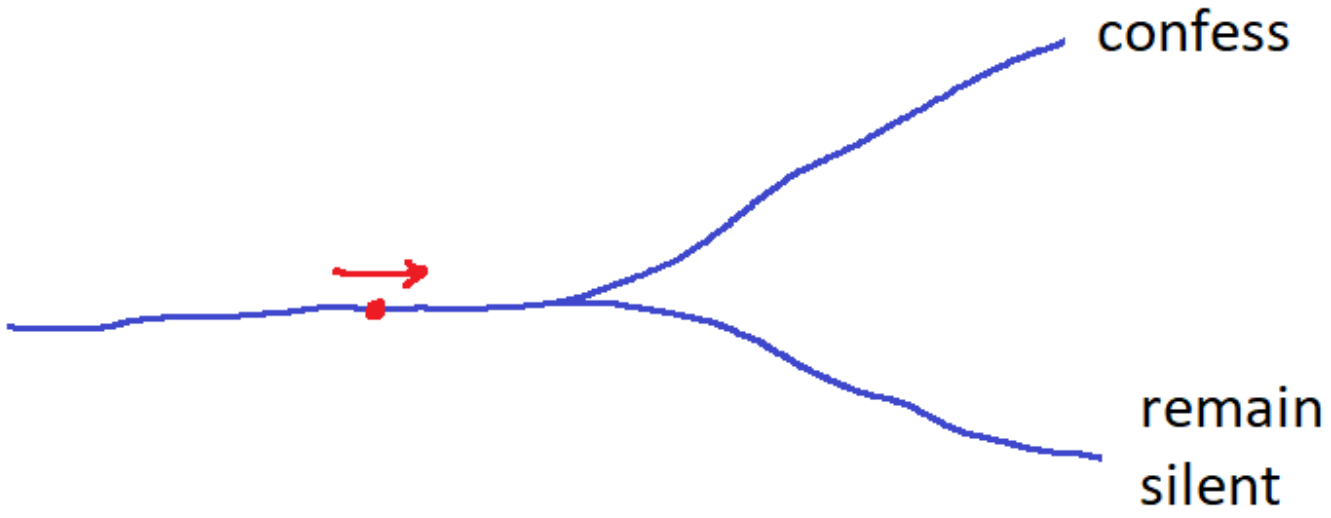
  (van Inwagen reading , p. 15)

- “… there would seem to be no possibility of its being up to Jane (or to anyone else) what the outcome of an indeterministic process would be.

  (p. 16)

What’s the argument for this?

“Let us suppose that a certain current-pulse is proceeding along one of the neural pathways in Jane’s brain and that it is about to come to a fork. And let us suppose that if it goes to the left, she will make her confession, and that if it goes to the right, she will remain silent.”



(This is Laplace,
not his demon)

“And let us suppose that it is undetermined which way the pulse will go when it comes to the fork: [even Laplace’s demon would not know].”

"Now let us ask: Is it *up to Jane* whether the pulse goes to the left or to the right? If we think about this question for a moment, we shall see that **it is very hard to see how this could be up to her**."

- N.B. Beware of what looks like a kind of "knowledge argument here"

We cannot see how this outcome was up to Jane

∴ The outcome was not up to Jane

Causation or determination?

"Nothing in the way things are at the instant before the pulse makes its "decision" to go one way or the other **makes it happen** that the pulse goes one way or goes the other. If it goes to the left, *that just happens*. If it goes to the right, *that* just happens. There is no way for Jane to *influence* the pulse. There is no way for her to **make it go** one way rather than the other."

- Notice how, in this argument, van Inwagen denies that Jane could be *causing* the pulse to go one way or the other. But why?

- Remember this is supposed to be an *indeterministic* system: one where the early states don't *determine* the later ones. (The early states still *cause* the later ones.) Why does van Inwagen deny that Jane causes the pulse to go one way?
- I think it's because he's picturing a kind of physical description of the system, and *within such descriptions* causation is absent.
  - Remember how the arrow of time, real existence and causation are all absent from physics?

Indeterministic causation

- If a system is indeterministic, then the laws and initial state do not determine its actual history (by definition).
  - (There are many possible histories consistent with the laws and initial state.)
- Yet a single real history comes to exist. We might ask: "How does that get determined?"
  - It doesn't get *determined*! (Remember that determination is a logical consequence relation between states of affairs.)
  - It does get "selected", or something, by the causal process itself.
  - N.B. "determined" is ambiguous between "selected" and "logically entailed".

If Jane did something to *make* the pulse go to the left, then obviously its going to the left would *not* be an undetermined event.

It is a plausible idea that it is up to an agent what the outcome of a process will be only if the agent is able to arrange things in a way that would make the occurrence of *this* outcome **inevitable** *and* able to arrange things in a way that would make the occurrence of *that* outcome **inevitable**.

Nozick on the intelligibility of FW

"… we want to know how [free will] works.

According to the view currently fashionable, we adequately understand a psychological process only if we can simulate that process on a digital computer. … Any process of choosing an action that could be understood in this sense would appear not to be a process of free choice. …

Robert Nozick, *Philosophical Explanations*, 302-3

Nozick on the intelligibility of FW

- … Suppose that this is so. Does the fact that we cannot, in this sense, understand what a free choice is, indicate some defect in the notion of a free choice or rather is the defect in the view that this mode of understanding is the sole mode? Is the result, that we cannot understand what a free choice is, an *artifact* of this method of understanding?" (p. 303)

- *Yes*, the latter. Such an abstract representation of a physical process leaves out causation, which selects a history non-deterministically.

Do indeterministic events "just happen"?

"Nothing in the way things are at the instant before the pulse makes its "decision" to go one way or the other **makes it happen** that the pulse goes one way or goes the other. **If it goes to the left, *that just happens*. If it goes to the right, *that* just happens**."

- Indeterministic events are often described as being "arbitrary," "capricious," "random," "irrational," and "uncontrolled."
  - I think this is a result of confusing the abstract model with reality.

The fallacy of misplaced concreteness

- This is the "error of mistaking the abstract for the concrete"

"The enormous success of the scientific abstractions has foisted onto philosophy the task of accepting them as the most concrete rendering of fact … Thereby, modern philosophy has been ruined."

Alfred North Whitehead, *Science and the Modern World*, 1926.

Some more Nozick

“Making some choices feels like this. There are various reasons for and against doing each of the alternative actions or courses of action one is considering, and it seems and feels as if one could do any one of them. In considering the reasons, mulling them over, one arrives at a view of which reasons are more important, which ones have more weight. One decides which reasons to act on …

The reasons do not come with previously given precisely specified weights; the decision process is not one of discovering such precise weights but of assigning them. The process not only weighs reasons, it (also) weights them.

At least, so it sometimes feels. This process of weighting may focus narrowly, or involve considering or deciding what sort of person one wishes to be, what sort of life one wishes to lead.”