

Free will &
determinism,
part II

Recall from last time that the consequence argument involved four principles:

The no choice principle: If I have no choice about p, and no choice about whether if p, then q, I have no choice about q.

We have **no choice about events which happened in the distant past.**

We have **no choice about what the laws of nature imply.**

Determinism: the laws of nature + the state of the universe at a time determine a unique future.

The argument was an attempt to show that it follows from these four principles that we have no choice about any of our future actions.

But of course van Inwagen is **not** trying to show that there's no such thing as free will; so why does he give an argument which has this as its conclusion?

The idea is that **if the first three principles are true** -- the No Choice Principle + the claims that we have no choice about the laws of nature or the past -- then, what the argument shows is that **if** determinism is true, **then** we have no free will.

So to demonstrate the incompatibility of free will and determinism, we assume the truth of determinism, and argue from there to the absence of free will. This is a style of argument called **conditional proof**. To prove the truth of a statement

if p, then q

we assume p as a premise, and argue from this premise, using only other true premises, to q as our conclusion. If we can construct a valid argument with p + some true statements as premises for q, it follows that the conditional statement

if p, then q

must be true. Here p = the truth of determinism, and q = the denial of the existence of free will.

The idea is that **if the first three principles are true** -- the No Choice Principle + the claims that we have no choice about the laws of nature or the past -- then, what the argument shows is that **if** determinism is true, **then** we have no free will.

So to demonstrate the incompatibility of free will and determinism, we assume the truth of determinism, and argue from there to the absence of free will. This is a style of argument called **conditional proof**. To prove the truth of a statement

if p, then q

we assume p as a premise, and argue from this premise, using only other true premises, to q as our conclusion. If we can construct a valid argument with p + some true statements as premises for q, it follows that the conditional statement

if p, then q

must be true. Here p = the truth of determinism, and q = the denial of the existence of free will.

We can lay out the consequence argument in premises as follows:

The consequence argument		
1	The laws of nature + the state of the universe at a time imply a unique future. In particular, (let's suppose) the laws imply that if DINOSAUR is the case, then so is DECISION.	Determinism (premise assumed for conditional proof)
2	If the laws of nature imply that X is the case, then we have no choice about X.	No choice about the laws of nature
3	We have no choice about the fact that if DINOSAUR, then DECISION.	1, 2
4	We have no choice about events which happened in the distant past.	No choice about the past
5	We have no choice about DINOSAUR.	4
6	If I have no choice about p, and no choice about whether if p, then q, I have no choice about q.	No choice principle
C	We have no choice about DECISION.	3, 5, 6

If premises 2, 4, and 6 are true, then it follows that if Determinism is true, there is no free will. Hence, if these premises are true, then Incompatibilism is true.

The consequence argument

1	The laws of nature + the state of the universe at a time imply a unique future. In particular, (let's suppose) the laws imply that if DINOSAUR is the case, then so is DECISION.	Determinism (premise assumed for conditional proof)
2	If the laws of nature imply that X is the case, then we have no choice about X.	No choice about the laws of nature
3	We have no choice about the fact that if DINOSAUR, then DECISION.	1, 2
4	We have no choice about events which happened in the distant past.	No choice about the past
5	We have no choice about DINOSAUR.	4
6	If I have no choice about p, and no choice about whether if p, then q, I have no choice about q.	No choice principle
C	We have no choice about DECISION.	3, 5, 6

If the consequence argument succeeds, it establishes only the conditional claim that if determinism is true, then we have no free will (and so also that if we have free will, the world must not be deterministic).

One way to think about this is in terms of the following chart illustrating four possible combinations of views about free will and determinism:

free will + determinism	no free will + determinism
free will + no determinism	no free will + no determinism

In terms of this chart, what the consequence argument establishes, if successful, is that the upper left box does not describe a way the world could be.

It does not, by itself, establish that we do or do not have free will; and it does not by itself establish that determinism is false, or that it is true.

free will + determinism	no free will + determinism
free will + no determinism	no free will + no determinism

In terms of this chart, what the consequence argument establishes, if successful, is that the upper left box does not describe a way the world could be.

It does not, by itself, establish that we do or do not have free will; and it does not by itself establish that determinism is false, or that it is true.

However, it does limit our choices. One option is that we **deny the existence of free will**; a second option is that we **accept the existence of free will, but say that this is only possible in an indeterministic world**.

The first of these is an extremely difficult option to accept, for at least two reasons.

First, it is simply extremely hard to believe that I have no choice about what I will be doing 5 seconds from now. And second, it seems that if there is no free will there is no moral responsibility; and it certainly seems that people are sometimes morally responsible for their actions.

So at this point the view that free will exists, but requires the falsity of determinism, looks like a pretty attractive option.

So at this point the view that free will exists, but requires the falsity of determinism, looks like a pretty attractive option.

However, upon reflection this view too can seem difficult to accept, because it can seem difficult to see how the falsity of determinism could make room for free will.

A good way to introduce the first sort of argument is by way of van Inwagen's example of Jane's decision. We are imagining that Jane is deciding whether to say something, and we are granting for now the incompatibilist's idea that if this decision is free, it must not be determined. Then we can imagine the case going like this:



Make a Mental Picture of Yourself

Each contemplated action will, of course, have antecedents in Jane's cerebral cortex, for it is in that part of Jane (or of her body) that control over her vocal apparatus resides. Let us make a fanciful assumption about these antecedents, since it will make no real difference to our argument what they are. (It will help us to focus our thoughts if we have some sort of mental picture of what goes on inside Jane at the moment of decision.) Let us suppose that there is a certain current-pulse that 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. And let us suppose that it is undetermined which way the pulse will go when it comes to the fork

We now imagine the current pulse traveling through Jane's brain.

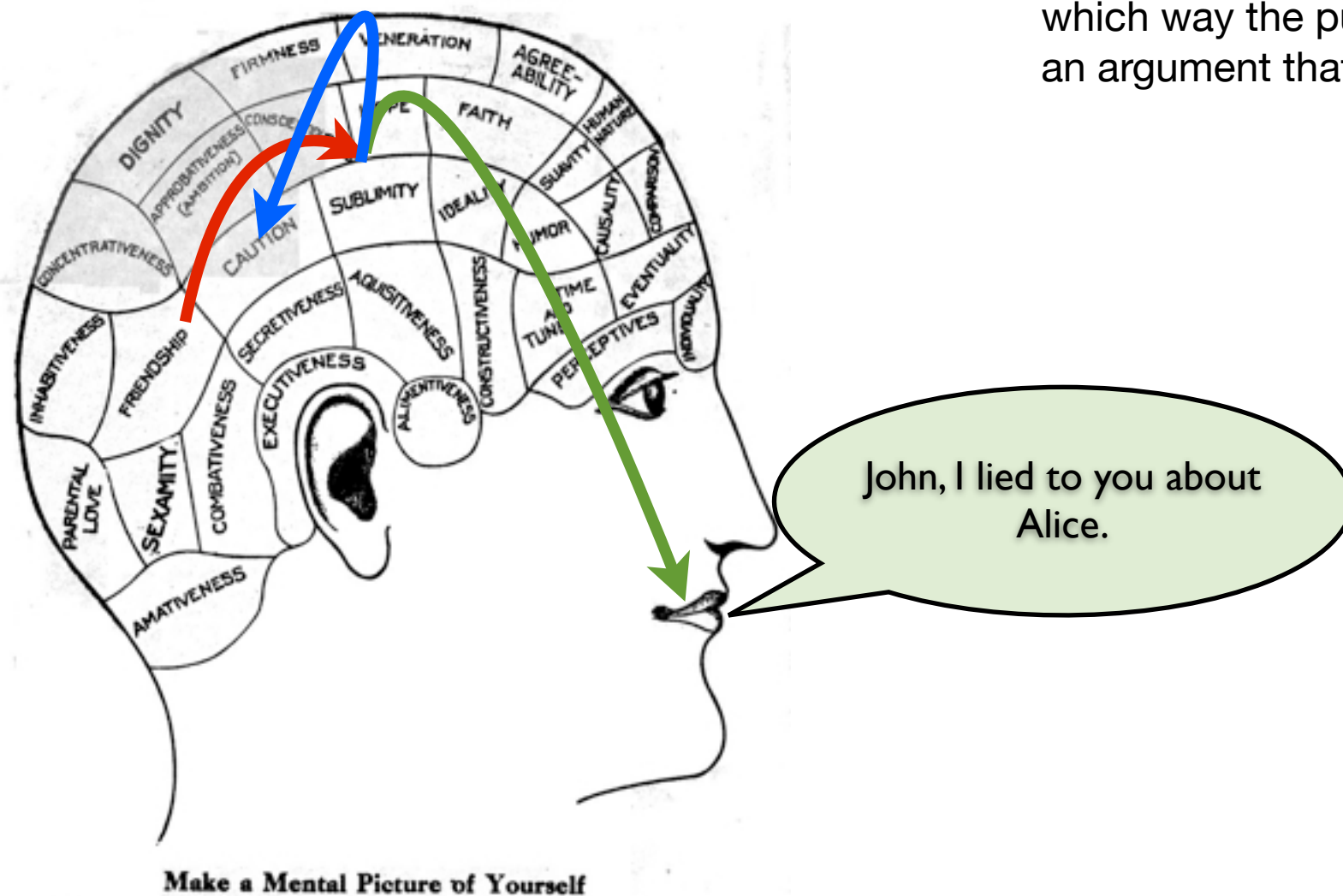
The pulse could go one of two ways. Which way it goes will determine whether or not Jane speaks; and which way will go is not determined by the the laws of nature + the state of Jane's brain (or the state of anything else).

We now imagine the current pulse traveling through Jane's brain.

The pulse could go one of two ways. Which way it goes will determine whether or not Jane speaks; and which way will go is not determined by the laws of nature + the state of Jane's brain (or the state of anything else).

Each contemplated action will, of course, have antecedents in Jane's cerebral cortex, for it is in that part of Jane (or of her body) that control over her vocal apparatus resides. Let us make a fanciful assumption about these antecedents, since it will make no real difference to our argument what they are. (It will help us to focus our thoughts if we have some sort of mental picture of what goes on inside Jane at the moment of decision.) Let us suppose that there is a certain current-pulse that 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. And let us suppose that it is undetermined which way the pulse will go when it comes to the fork

The key question, now is: is Jane free to decide which way the pulse will go? van Inwagen gives an argument that she is not:

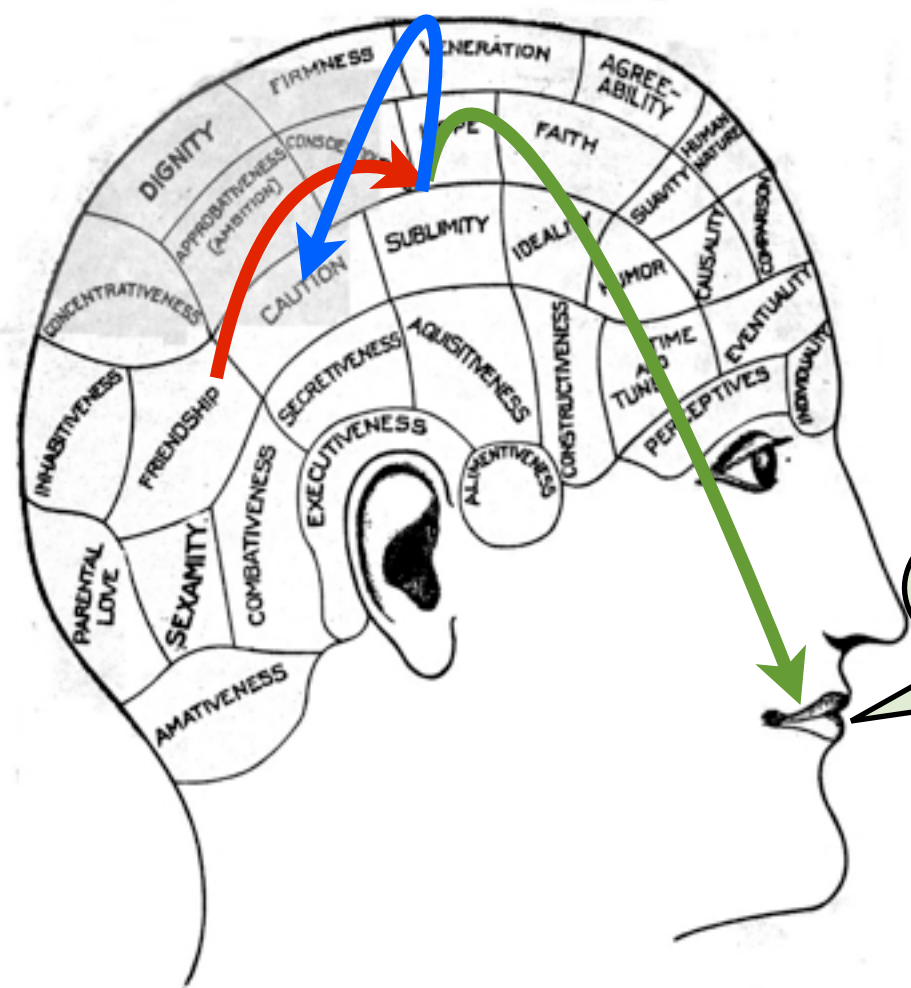


We now imagine the current pulse traveling through Jane's brain.

The pulse could go one of two ways. Which way it goes will determine whether or not Jane speaks; and which way will go is not determined by the laws of nature + the state of Jane's brain (or the state of anything else).

The key question, now is: is Jane free to decide which way the pulse will go? van Inwagen gives an argument that she is not:

Now let us ask: Does Jane have any choice about 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 she could have any choice about that. 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. Or, at least, there is no way for her to make it go one way rather than the other and leave the "choice" it makes an undetermined event. 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 the only way to have a choice about the outcome of a process is to be able to arrange things in ways that will make it inevitable that this or that outcome occur. If this plausible idea is right, then it would seem that there is no way in which anyone could have any choice about the outcome of an indeterministic process. And it seems to follow that if, when one is trying to decide what to do, it is truly undetermined what the outcome of one's deliberations will be, then one could have no choice about that outcome.



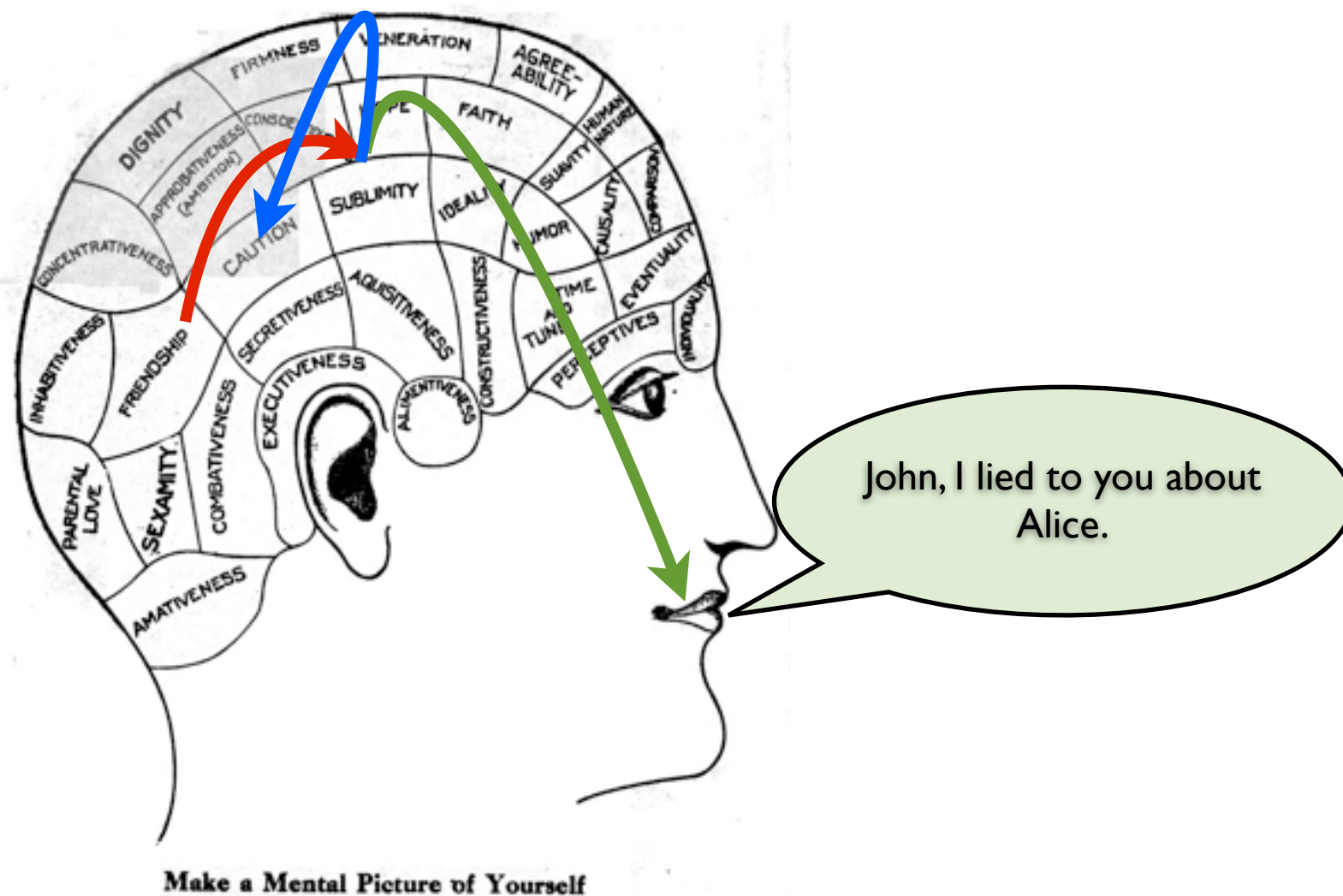
John, I lied to you about Alice.

Make a Mental Picture of Yourself

We now imagine the current pulse traveling through Jane's brain.

The pulse could go one of two ways. Which way it goes will determine whether or not Jane speaks; and which way will go is not determined by the laws of nature + the state of Jane's brain (or the state of anything else).

The basic idea seems to be this: if we think about the causal chain leading up to some putatively free action A of Jane's, then, if A is really free and incompatibilism is true, there must be some event, E, in this causal chain which is not determined by prior events plus the laws of nature. Further, it seems that for A to be free, Jane must have had a choice about whether E happened. But it is hard to see how Jane could have had a choice about whether E happened, since the entire state of the universe prior to E, including everything Jane does and thinks, is consistent both with E happening and with E not happening. But then it was **not up to Jane** whether E happened at all.

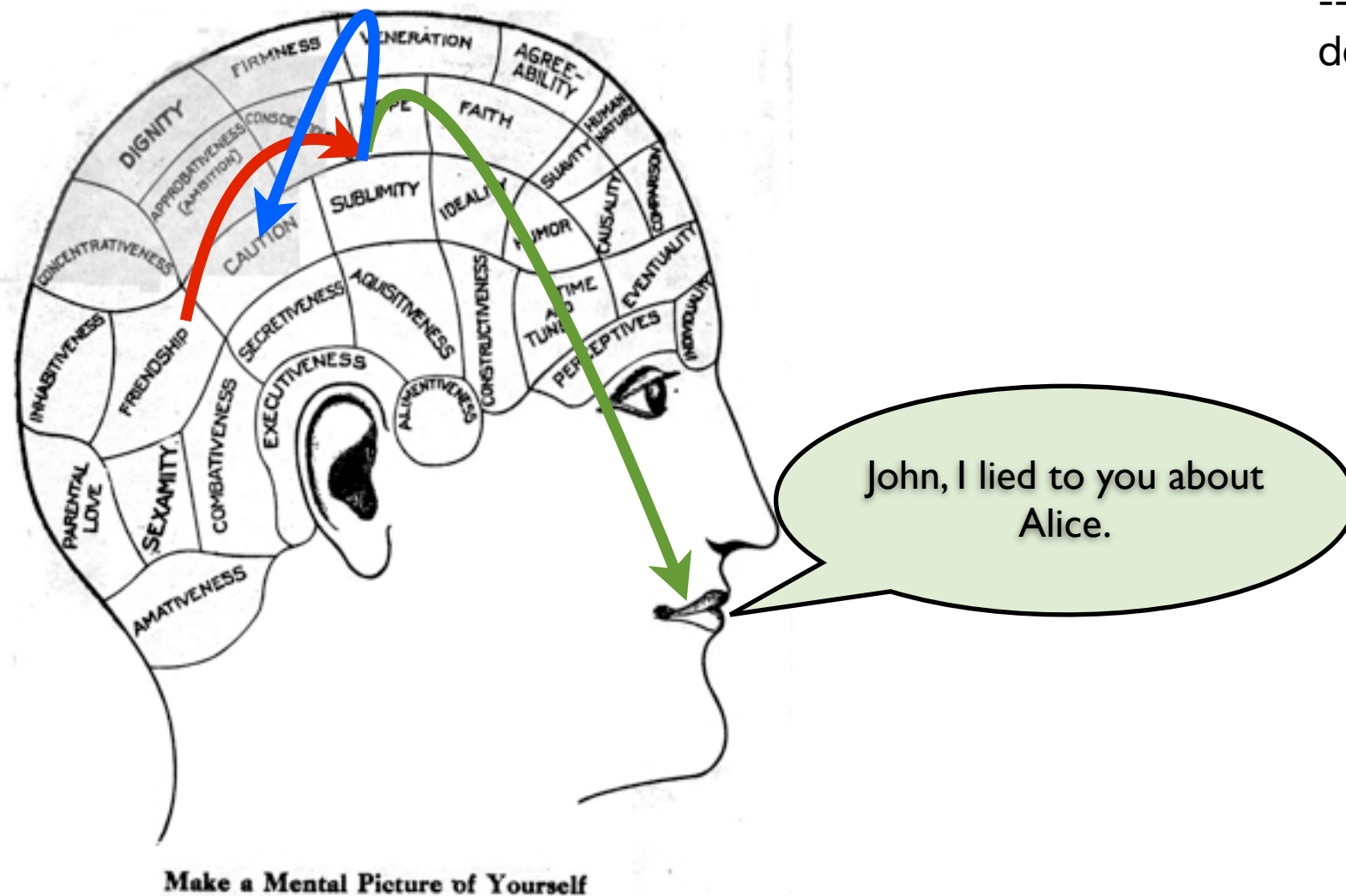


The basic idea seems to be this: if we think about the causal chain leading up to some putatively free action A of Jane's, then, if A is really free and incompatibilism is true, there must be some event, E, in this causal chain which is not determined by prior events plus the laws of nature. Further, it seems that for A to be free, Jane must have had a choice about whether E happened. But it is hard to see how Jane could have had a choice about whether E happened, since the entire state of the universe prior to E, including everything Jane does and thinks, is consistent both with E happening and with E not happening. But then it was **not up to Jane** whether E happened at all.

This argument seems to rely on some principle like the following:

If nothing determines whether someone chooses A or B, the choice of A or B is random, and hence not a free choice.

If any principle of this sort is true, this is serious trouble for the incompatibilist who wants to believe in free will. After all, this sort of principle seems to show that free will **requires** determinism -- or at least requires that human actions be determined.



This argument seems to rely on some principle like the following:

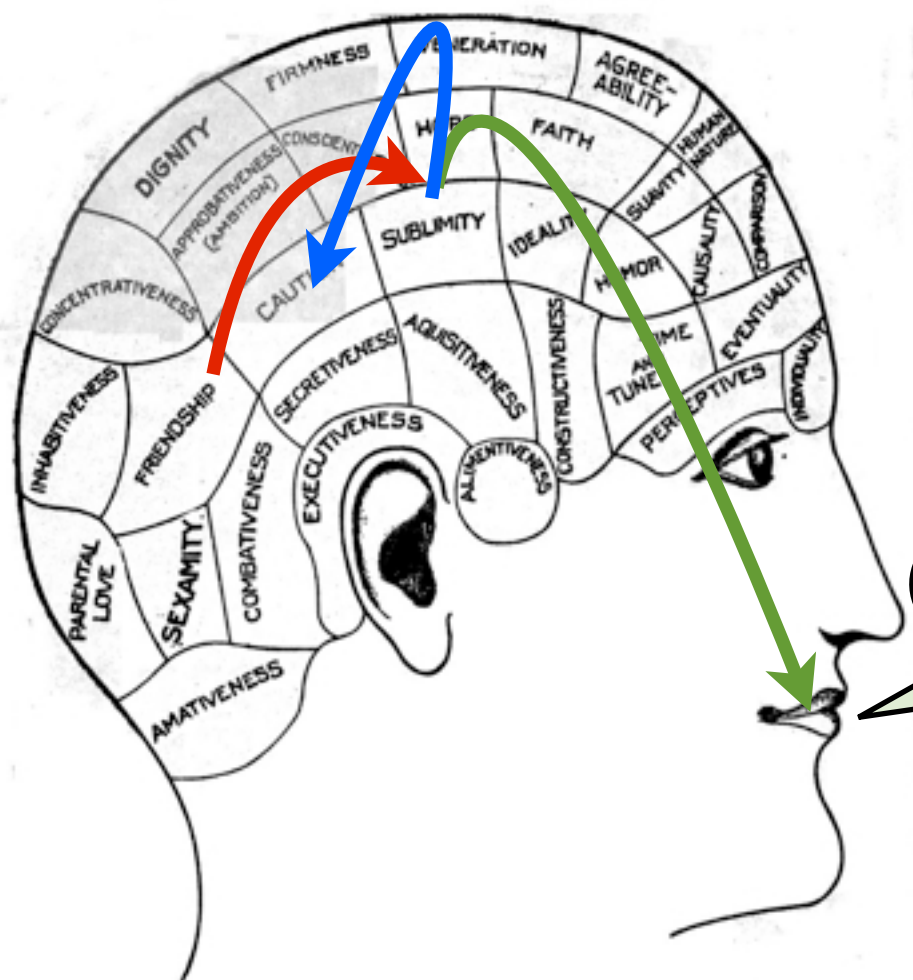
If nothing determines whether someone chooses A or B, the choice of A or B is random, and hence not a free choice.

If any principle of this sort is true, this is serious trouble for the incompatibilist who wants to believe in free will. After all, this sort of principle seems to show that free will **requires** determinism -- or at least requires that human actions be determined.

Something like this was the view of David Hume. He thought that genuine free will requires that one's actions be determined by one's prior motives, and that the idea that there is a conflict between determinism and free will just rests on a confusion:



'Tis commonly allow'd that mad-men have no liberty. But were we to judge by their actions, these have less regularity and constancy than the actions of wise-men, and consequently are farther remov'd from necessity. Our way of thinking in this particular is, therefore, absolutely inconsistent; but is a natural consequence of these confus'd ideas and undefin'd terms, which we so commonly make use of in our reasonings, especially on the present subject.



John, I lied to you about Alice.

Make a Mental Picture of Yourself

This was part of Hume's argument for **compatibilism**: the view that determinism is consistent with the existence of free will. But, as van Inwagen says, one might take the example of Jane and her brain to show something quite different.

This was part of Hume's argument for **compatibilism**: the view that determinism is consistent with the existence of free will. But, as van Inwagen says, one might take the example of Jane and her brain to show something quite different.

But now a disquieting possibility suggests itself. Perhaps the explanation of the fact that both compatibilism and incompatibilism seem to lead to mysteries is simply that the concept of free will is self-contradictory. Perhaps free will is, as the incompatibilists say, incompatible with determinism. But perhaps it is also incompatible with *indeterminism*, owing to the impossibility of anyone's having a choice about the outcome of an indeterministic process. If free will is incompatible with both determinism and indeterminism, then, since either determinism or indeterminism has to be true, free will is impossible.

Recall our chart of possible views about free will above:

free will + determinism	no free will + determinism
free will + no determinism	no free will + no determinism

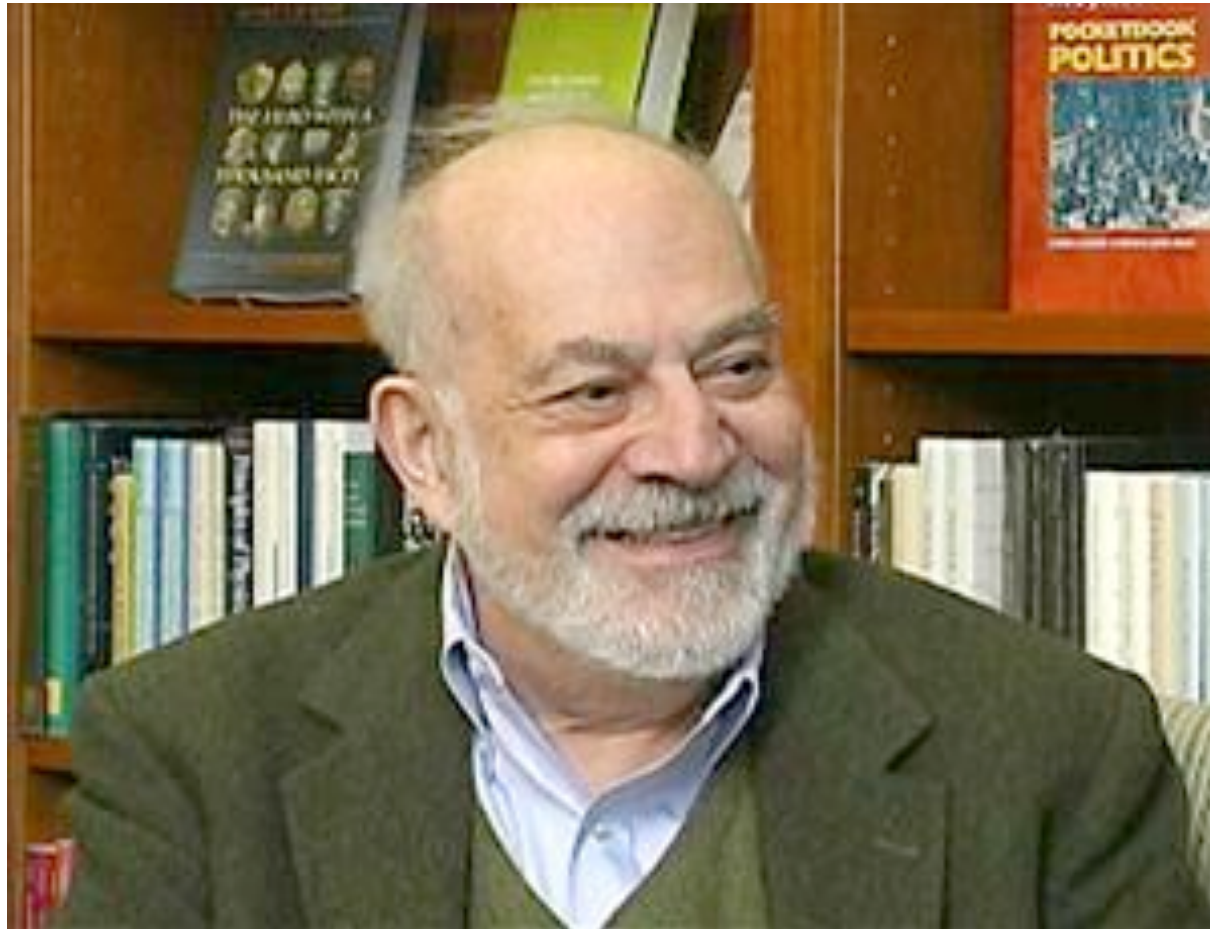
The consequence argument seems to rule out the top left square. But now the example of Jane makes it hard to see how indeterminism could make room for free will, which seems to rule out the bottom left square.

This leaves us only with the right column of the chart - both boxes of which involve denying the existence of freedom of the will.

But before we accept this conclusion, perhaps we should re-examine the compatibilist view that was the target of the consequence argument.

A powerful reason for re-examining compatibilism is brought out by an example due to Harry Frankfurt.

A powerful reason for re-examining compatibilism is brought out by an example due to Harry Frankfurt.



“Suppose someone - Black, let us say - wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something *other* than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do....

Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. It would be quite unreasonable to excuse Jones for his action ... on the basis of the fact that he could not have done otherwise. This fact played no role at all in leading him to act as he did.... Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it.”

“Suppose someone - Black, let us say - wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something *other* than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do....

Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. It would be quite unreasonable to excuse Jones for his action ... on the basis of the fact that he could not have done otherwise. This fact played no role at all in leading him to act as he did.... Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it.”

Suppose, at time T, that Black decides that he wants Jones on a bus out of South Bend by some later time, T+3. He hopes that Jones will get on the bus of his own accord, but, if he doesn't, plans to force him onto the bus.

Now suppose that, at time T+1, Jones is deliberating about whether or not to take a bus out of South Bend. He goes back and forth, but eventually decides to board the bus.

At time T+2, Jones boards a bus leaving South Bend. This seems (certainly, at least, to Jones himself) to be a free action. Since it never got to time T+3, Black never had to execute his nefarious plan.

But now think about Jones' decision making at time T+1. At that time, it was already determined that Jones would be boarding a bus out of South Bend. After all, Jones must either decide to board the bus, or not. In the former case he gets on the bus, and in the latter case, Black forces him on the bus, so again he gets on the bus. So facts fixed prior to his decision making process determined that he would get on the bus. Nonetheless, his decision to get on the bus seems to be a free action.

This seems to be some sort of evidence that free actions can coexist with determinism, and therefore seems to be an argument **for compatibilism**.

“Suppose someone - Black, let us say - wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something *other* than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do....

Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. It would be quite unreasonable to excuse Jones for his action ... on the basis of the fact that he could not have done otherwise. This fact played no role at all in leading him to act as he did.... Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it.”

But now think about Jones' decision making at time $T+1$. At that time, it was already determined that Jones would be boarding a bus out of South Bend. After all, Jones must either decide to board the bus, or not. In the former case he gets on the bus, and in the latter case, Black forces him on the bus, so again he gets on the bus. So facts fixed prior to his decision making process determined that he would get on the bus. Nonetheless, his decision to get on the bus seems to be a free action.

This seems to be some sort of evidence that free actions can coexist with determinism, and therefore seems to be an argument **for compatibilism**.

Can an incompatibilist say that Jones was free despite the fact that, prior to his decision, the state of the universe determined what he would do?

Should the incompatibilist instead say that Jones was not free? Do they then have to deny Frankfurt's plausible-seeming claim that Jones is morally responsible for his action?

This is why the problem of free will seems so hard: no matter what view one takes of the relationship between free will and determinism, one faces some difficult challenges.

Basically, we have three options; and each of these options has its costs.

Deny that there is any such thing as free will.

Make sense of the fact that it seems so clear that what we will do in the near future is something that we have a choice about.

Either deny that anyone is ever morally responsible for anything, or explain how moral responsibility is possible without free will

Affirm the existence of free will, and say that this is compatible with determinism being true.

Explain how it can be that I have no choice about p, and no choice about the fact that if p, then q, and yet have a choice about q.

Affirm the existence of free will, and say that this is incompatible with determinism being true.

Explain how an action can be undetermined without being random, and hence not free.

Make sense of Frankfurt's example; either explain why Jones is really not free, or why Jones's being free is not a genuine example of a free action determined by factors outside of the agent's control.

This is why free will seems to be such a great mystery. It is clear that one of the views listed at left must be true, but it is hard to see how any of them could be.