

Recall:

The Consequence Argument

- (1) What happens next is a necessary consequence of the past and the laws of nature.
- (2) I can't change the past.
- (3) I can't change the laws of nature.
- (4) I can't change the fact that what happens next is a necessary consequence of the past and the laws of nature.
- (5) I can't change what happens next.

And to make the argument *valid* we needed:

- (β) If we can't change X or that Y necessarily follows from X, then we can't change Y.

The classical compatibilist blocks the argument by claiming.

I could do X (freely) = If I had wanted/chosen to do X, I would have.

This seems to make the hidden premise (β) false. But is this *analysis* a good one?

Counterexample 1: Compulsions

Danielle's Dog Hang-up. Danielle had a traumatic experience with a blond-haired dog which makes her *incapable of wanting* to touch a blond-haired dog. For her birthday Danielle is given the option picking either a blond or a black-haired Lab by touching it. She (happily) touches the black-haired Lab.

Claims:

- If, somehow, she had *wanted* to touch the blond-haired dog, she would have.
 - Nonetheless, she *couldn't* have touched that dog (because she couldn't have *wanted* to).
- So, it is claimed, the analysis is wrong.

Counterexample 2: Manipulation

Mad Neurosurgeon. An evil neurosurgeon can manipulate only Mark's *desires* via electronic manipulation of his brain states. The neurosurgeon uses Mark like a puppet in this way to fire a gun at the neurosurgeon's arch-nemesis.

Claims:

- If Mark had *wanted* not to shoot, he wouldn't have.
 - Nonetheless Mark *couldn't* have *not* fired the gun (of his own free will).
- So, it is claimed, the analysis is wrong.

"New" Compatibilism

In response to these problems, some compatibilists "refine" their analysis. A general strategy: keep the *conditional* form of the analysis, but change what free action, or power to choose is conditional on. So in general...

I could have done X =
If, I would have done X.

The only problem is: what goes in for the blank? A very interesting recent suggestion involves appealing to what we "think we have reason to do". That is,

Reason-Responsive Compatibilism

I could do X (freely) = If I *thought* I had most *reason* to do X, then I would do X.

Does this do a good job with Danielle? Does it deal with the Neurosurgeon? Are there other problems for it?

This is just one example of a new form of compatibilism, and there are many, many more. You can probably come up with views like this on your own.

The important point is that whatever analysis we end up with had better be *extremely* persuasive. Otherwise, principle beta might be true, and incompatibilism would follow. What should we believe then?