

# Frankfurt's compatibilist theory of free will

Jeff Speaks

March 19, 2009

1	Free will and second-order volitions . . . . .	1
2	Frankfurt on the principle of alternate possibilities . . . . .	3

Last time we discussed van Inwagen's consequence argument, which is the most important argument for incompatibilism about free will and determinism.

Harry Frankfurt is a prominent defender of a compatibilist view of free will. We will be talking about one piece which can be read as a counterexample to an apparent incompatibilist principle ('Alternate Possibilities and Moral Responsibility') and one which explains a view of free actions which has the consequence that free will is compatible with determinism ('Freedom of the Will and the Concept of a Person').

## 1 Free will and second-order volitions

In "Freedom of the Will and the Concept of a Person", Frankfurt presents a theory of free action. Note that this is something which van Inwagen did not try to do; van Inwagen argued for the claim that free will is inconsistent with determinism, but never tried to say what about certain actions makes them free (other than the necessary condition that they cannot be determined by facts about which the subject has no choice).

Frankfurt's argumentative strategy can be thought of like this: he will defend a theory of the form

$$A's \phi \text{ing is free iff } A's \phi \text{ing is } F$$

for some property of actions  $F$ . Let's suppose we are convinced that his theory, or one like it, is correct. Then we can raise the question of compatibilism in a new form: if the world were deterministic, could any actions have property  $F$ ? If the answer is 'Yes,' as he thinks, then Frankfurt's theory of free will will be an indirect argument for compatibilism.

To understand Frankfurt's view, we'll have to get clear on a few of the terms that he uses, beginning in §I of the article:

- *first-order desire*: a desire to perform some action. A desire to eat a bag of potato chips is a first-order desire; a desire for world peace is not.
- *will*: a first-order desire which is effective, i.e. that causes one to do what one desires to do. A desire to eat a bag of potato chips is one's will, in Frankfurt's sense, if that desire brings one to actually eat the bag of potato chips.
- *second-order desire*: a desire to have a certain desire. A desire that I should desire celery rather than potato chips is an example of a second-order desire.
- *second-order volition*: a desire that a certain desire be one's will, i.e., a desire that a certain desire bring one to action. The above example can be turned into an example of a second-order volition if I desire, not just to have the desire for celery, but that the desire for celery rather than potato chips be effective in bringing me to eat celery rather than potato chips.

In §II, Frankfurt uses these notions to give an analysis of the concept of a person. We will be interested in his use of these notions to explain the nature of free action. He attempts this in §§III-IV.

Here is what Frankfurt says about free will:

“It seems to me both natural and useful to construe the question of whether a person's will is free in close analogy to the question of whether an agent enjoys freedom of action. Now freedom of action is . . . freedom to do what one wants to do. Analogously, then, the statement that a person enjoys freedom of the will means . . . that he is free to want what he wants to want. More, precisely, it means that he is free to will what he wants to will, or to have the will that he wants. Just as the question about the freedom of an agent's action has to do with whether it is the action he wants to perform, so the question about the freedom of the will has to do with whether it is the will that he wants to have.

It is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will.”

How this illuminates the distinction between the willing and the unwilling addict.

Some strengths of Frankfurt's analysis of freedom of the will:

1. It explains freedom of the will in terms of a (relatively) un-mysterious concept, desire.
2. It offers a plausible account of the difference between the willing and unwilling addict.
3. It explains our intuition that human beings, but not lower animals, have free will. Lower animals lack free will because they lack the second-order volitions which are constitutive of free will.

4. It explains why freedom of the will is worth wanting, in terms of desire satisfaction.

One problem for the analysis: third- (and higher-) order desires. A way around this problem.

Another interesting fact about this analysis of free will, given the kinds of views we have discussed so far, is that on this view free will appears to be perfectly compatible with determinism. As Frankfurt puts it:

“My conception of the freedom of the will appears to be neutral with regard to the problem of determinism. It seems conceivable that it should be causally determined that a person is free to want what he wants to want. If this is conceivable, then it might be causally determined that a person enjoys a free will. There is no more than an innocuous appearance of paradox in the proposition that it is determined, ineluctably and by forces beyond their control, that certain people have free wills and that others do not.”

Is Frankfurt right that in a deterministic world one can have desires about one's own desires?

## 2 Frankfurt on the principle of alternate possibilities

We now turn to Frankfurt's more direct objection to incompatibilist views of free will.

One reason why you might find the consequence argument for incompatibilism convincing is that you might find something like the following principle plausible:

*The principle of alternate possibilities (free will version)*

A person's act is free if and only if that person could have done otherwise.

The natural thought is that if determinism is true, that means that no one ever could have done otherwise, since the laws of nature determine a unique future; and therefore that, given the principle of alternate possibilities, no one ever acts freely.

Frankfurt argues that this principle is false. If he is right, this does undercut one sort of incompatibilist intuition about free will. We will return to the question of whether and how an argument for the falsity of this principle is a problem for incompatibilism per se.

Frankfurt's argument against the above principle proceeds via an argument against the following closely related principle:

*The principle of alternate possibilities (moral responsibility version)*

A person is morally responsible for their action if and only if that person could have done otherwise.

Is it true that if the PAP's moral responsibility version is false, the free will version is as well? Why one might have some doubts about this, based on cases of moral responsibility without free will.

However, these sorts of worries are not ultimately of great importance, since the case that Frankfurt presents as an example of moral responsibility without alternate possibilities might rather be thought of directly as an example of free will.

Beginning on p. 835, Frankfurt develops the example of Jones<sub>4</sub>, which is the most important of his attempted counterexamples to the principle of alternate possibilities:

“Suppose someone — Black, let us say — wants Jones<sub>4</sub> 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<sub>4</sub> 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<sub>4</sub> is going to decide to do something *other* than what he wants him to do. If it does become clear that Jones<sub>4</sub> is going to decide to do something else, Black takes effective steps to ensure that Jones<sub>4</sub> 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<sub>4</sub>, 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<sub>4</sub> 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<sub>4</sub> 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.”

Does this principle refute the principle of alternate possibilities (free will version)?

Let's suppose for the sake of argument that it does. Why should the incompatibilist care? We can consider three different possible answers to this question.

1. Perhaps we construct some argument from the falsity of the principle of alternate possibilities to the falsity of incompatibilism. But it is not easy to see how such an argument might go.
2. Perhaps, in the case of Black and Jones, we can simply stipulate that determinism is true. Why, one might ask, could this matter if Jones can't do otherwise anyway? But the incompatibilist should clearly not be convinced by this.
3. Perhaps this sort of argument casts some doubt on the motivation for being an incompatibilist; for example, perhaps it shows that there is some problem with the consequence argument for incompatibilism.

To see why this might be so, consider the following new instance of the consequence argument, in which ‘BEGINNING’ is a name for the state of the universe, including the dispositions of Jones and Black, at  $t$ , before Jones begins to deliberate. Now consider some later time  $t^*$ , during which Jones is deliberating about whether to perform the action in question. It seems that we can argue as follows:

1. Jones at  $t^*$  is not free with respect to BEGINNING.
  2. Jones at  $t^*$  is not free with respect to L.
  3. Jones at  $t^*$  is not free with respect to (BEGINNING & L). (1,2)
  4. Necessarily, if (BEGINNING & L) then ACTION. (set up of the case)
  5. Jones at  $t^*$  is not free with respect to the fact that if (BEGINNING & L) then ACTION. (4)
  6. If we are not free with respect to (BEGINNING & L), and we are not free with respect to the fact that if (BEGINNING & L) then ACTION, we are not free with respect to ACTION. (instance of the No Choice Principle)
- 
- C. Jones at  $t^*$  is not free with respect to ACTION. (3,5,6)

If the incompatibilist grants that Jones’ action is free in Frankfurt’s example, he must say that the conclusion is false. Since the argument is valid, he must reject a premise. He cannot reject premise (6), since this is an instance of the No Choice Principle, which is crucial to the main argument for incompatibilism. So he must reject one of (1), (2), and (4). But which one?