

**On the Equivocation of Moral Responsibility: A Summary and Response to  
Frankfurt's Critique of the Principle of Alternate Possibilities**

DAVID RAY DIDIODATO

March 20th, 2010

**Introduction**

The field of metaphysics is plagued with problems – for which many the resolution appears, if not impossible, then at least exceedingly difficult to achieve. This is perhaps nowhere more explicit than in the problem surrounding the supposed ‘free-will’; and, in particular, in reconciling it with the seemingly deterministic world within which it is thought to exist. The imperative for the resolution of this particular metaphysical problem is grounded largely on the social desire to allocate moral responsibility: for if one *could not have done otherwise*, as determinism seeks to imply, ‘moral responsibility for ones actions’ becomes a difficult concept to sell.

At the heart of this problem lies a principle which seeks to demonstrate just this point: This principle, known as the Principle of Alternate Possibilities, suggests that “a person is morally responsible for what he has done only if he could have done otherwise”.<sup>1</sup> Attempting to demonstrate the inadequacy of this principle, Frankfurt provides a counter-example wherein an individual is suggested to be both ‘unable to do otherwise’ and yet still remain ‘morally responsible’ for their actions. This paper shall examine both the content and validity of Frankfurt’s argument.

To carry out this task of exposition and criticism I have divided this paper into two sections. The *first*, ‘*Frankfurt’s Critique and Revision of PAP*’, will be largely expositional and will outline the basic structure and logic of Frankfurt’s criticism, and

subsequent revision, of the Principle of Alternate Possibilities. This is of methodological necessity insofar as it will lay the background foundation upon which my critique shall be grounded. *Part II* of this paper, then, will provide this critique. I will argue, first, that Frankfurt's counter-example to PAP has an *implicit presupposition of determinism* whereby an individual's actions can be *entirely determined* by reference to their biological drives, psychological states, and physiological nature. With this premise in mind I will then argue that Frankfurt's counter-example to PAP fails insofar as it *equivocates on the criteria for allocating 'moral responsibility'*. Let us begin, then, by turning first to the logic and structure of his argument itself:

### **Part I: Frankfurt's Critique and Revision of PAP**

As mentioned above, the Principle of Alternate Possibilities is an attempt to qualify the allocation of moral responsibility to actions such that one is responsible only if they *could have done otherwise*.<sup>2</sup> The content of the principle hinges, therefore, on the manner by which one chooses to interpret 'could have done otherwise'; and, more particularly, on the specific examples in which 'could have done otherwise' seem to be manifest. Frankfurt, although arguing against the *unconditional* truth of the principle, is nonetheless sympathetic to many of its applications, particularly when 'could *not* have done otherwise' would involve matters of coercion.<sup>3</sup> However, if PAP is to be accepted as a metaphysical and moral truth it must be capable of withstanding *all* possible examples. It is on this point where Frankfurt believes the principle to be inadequate.

Frankfurt's method of critique is to provide a counter-example. This example, if it is to succeed, will be required to provide a situation wherein an individual '*could not have done otherwise*'; and yet whom, at the same time, our intuitions would hold *morally*

*responsible for their act.* Such an example would demonstrate that PAP is necessarily false.

His counter-example<sup>4</sup> is as follows: Black, a master of manipulation, coercion,<sup>5</sup> and the ability to judge the future actions of others, desires that Jones perform a particular action *X*. Also being lazy, however, Black does not want to go to the trouble of manipulating Jones so as to perform *X* unless absolutely necessary. His plan, therefore, is to wait until the moment of Jones' decision: if Jones decides to perform *X* on his own accord, all the better for lazy Black; if Jones does not, however, then Black (drawing upon his ability to judge future actions) will immediately manipulate Jones such that he performs *X* anyways. Thus far, the essential idea Frankfurt is attempting to defend is that Jones *could not have done otherwise*: his performing of *X* will occur whether he personally decides to do *X* or not.

From here there is only one additional step required for Frankfurt to complete his critique: namely, to find a way of holding Jones morally responsible for performing *X* within the context of his (unfortunate) situation. This Frankfurt believes can be done if Jones decides to do *X* on his own: in this case Black will not interfere, and thus there will be no tampering or coercion. Morally speaking, it seems Jones will be just as responsible for performing his action as he would have been if Black did not exist at all.<sup>6</sup>

Frankfurt's critique is therefore complete: for insofar as Jones was both 'unable to do otherwise' and yet, at the same time, morally responsible for his action, it follows that PAP must be false. Frankfurt attempts to revise the principle, however, so as to take into account the consequence of his critique. He suggests PAP should be altered such that an individual "is not morally responsible for what he has done if he did it only because he could not have done otherwise":<sup>7</sup> in other words Jones, if he decides to perform *X* on his

own, is morally responsible for his action because he performed it, not *only* because he could not have done otherwise, but *also* because he *wanted* to perform it himself. If Jones had decided to *not* perform X, however, and Black were to manipulate him such that he did, we can say that Jones would *not* be morally responsible insofar as he performed X *only* because he could not do otherwise.

But is such a revision necessary? I will now argue that it is not insofar as Frankfurt's critique fails to demonstrate the falsity of PAP to begin with.

## **Part II: In Response to Frankfurt**

The basic structure of my argument is twofold: I will show, first, that Frankfurt's counter-example has an underlying presupposition of determinism; and second, by drawing upon that insight, I will demonstrate that Frankfurt equivocates his criteria for allocating 'moral responsibility' regarding Jones' decision to perform or not perform X.

### *a) That Determinism is Implicit in Frankfurt's proposed Counter-Example*

First, then, I will argue to Frankfurt's presupposition of Determinism: The challenge of anyone wishing to proceed philosophically by employing examples to defend or critique a position is to structure the example such that the premises one wishes to define follow with *absolutely certainty*: for if the circumstances of the situation can be shown to have a margin of error such that the premises are not believable, it follows that the conclusion drawn therefrom may be questioned.

In Frankfurt's counter-example he is concerned principally with ensuring the *absolute certainty* of two premises: the *first*, is that Jones *could not have done otherwise than X*; and the *second*, is that *if Jones performed X on his own accord, he would be*

*morally responsible*. The second of these premises does not concern us at present, as Frankfurt does not provide a ‘logical’ argument in defense of it. He relies, rather, on the assertion that ‘intuition’<sup>8</sup> demonstrates Jones’ responsibility. (Although this premise does not concern us here that is not to say it should necessarily be accepted: it could well be argued that intuition is an unreliable arbiter concerning the allocation of truth or moral responsibility: think Nietzsche’s suggestion that “convictions are more dangerous enemies of truth than lies”).<sup>9</sup>

Frankfurt’s first premise, however, is more problematic to establish with certainty, and he seems to recognize this fact himself as he provides a thorough argument in its defense. The problem to be overcome is *doubt* – namely, doubt as to whether Jones *could have done otherwise* – and Frankfurt, so as to achieve this task, proceeds (sensibly in my view) by paralleling the method employed by the champion of overcoming doubt: Descartes.<sup>10</sup>

Descartes, to find certainty, employed example after example of progressively insurmountable obstacles to his possessing certain knowledge: for if the last and greatest of these could be overcome, certainty of knowledge would be implicitly possible thereby. Frankfurt is not seeking to defend ‘certain knowledge’ in the epistemological sense; but, rather, in the assertion that *Jones could not have done otherwise*. He therefore provides a series of progressively insurmountable coercive techniques theoretically employed by Black to ensure Jones’ compliance.

First, a ‘terrible threat’ is suggested.<sup>11</sup> This seems immediately problematic, however, as a threat (supposing it is sufficiently terrible) only provides a *reason* for Jones to do X; but Jones may not be reasonable, and thus not perform X on account of it. Next it is suggested that Black could drug or hypnotize Jones in such a manner as to make him

conform to his wish.<sup>12</sup> But it could still be argued that even while under the influence of a certain drug or hypnosis Jones would still possess enough autonomy to decide to act in a manner intuiting moral responsibility. To solidify his case, therefore, Frankfurt makes his final regress and asserts definitively that Black could “manipulate the minute processes of Jones’s brain... [such that] that causal forces... determine that he chooses to act and that he does act in the one way and not in any other”.<sup>13</sup> Even Descartes, I think, is convinced: Jones *really* could not have done otherwise.

My intent, however, is not to dispute the validity of Frankfurt’s conclusion; but, rather, to reflect on a presupposition which is implicit in its method: namely, *determinism*. This presupposition underlies *all* of Black’s proposed coercion techniques: In the first case, of employing a ‘threat’, it is implied that Jones’ decision (if it is rational, as the example implies)<sup>14</sup> will be altered as a result of the new information he has of the potential consequences of the threat: thus, he is not ‘free’ to do what he wishes; but, rather, is determined by his emotional/rational/physiological drives and values. If he were not, then the institution of a threat would not provide any reason to believe he would act differently.

The remaining three examples are much easier to appropriate the deterministic presupposition: the use of drugs to manipulate Jones’ actions necessarily presupposes that psychochemical relations determine brain functioning; and, naturally, suggesting that brain processes can be themselves manipulated causally leaves no doubt that Jones’ brain states (and, therefore, actions) are purely a result of causal relations: thus Jones is at all times determined by physical laws, and never ‘free’ in the ‘metaphysical’ sense employed by van Inwagen.<sup>15</sup>

Indeed, it seems that instituting determinism is of methodological necessity for Frankfurt’s argument: for insofar as it is the *certainty* of Jones’ action that is to be

defended, it seems intuitive that employing an argument of causal-necessity is the most effective means by which to proceed. And, this is no small matter: for it is the very believability of Frankfurt's central premise which is at stake – and indeed, if determinism is to be rejected, based on the arguments Frankfurt has provided in his example,<sup>16</sup> we need not believe his assertion that *Jones could not have done otherwise*. Frankfurt's argument *requires* determinism.

*b) On the Equivocation of 'Moral Responsibility' in Frankfurt's Counter-Example*

Having satisfied ourselves with the first of Frankfurt's premises (that Jones could not have done otherwise) it now remains to reflect on the second: namely, that Jones is morally responsible if he chooses to perform action *X* by himself, without being manipulated by Black; and that, conversely, he is *not* morally responsible if he chooses to not do *X*, but is subsequently made to do so by Black.

When Frankfurt appealed to his intuition to propose this premise it seemed non-problematic as it was supposed Jones was 'free' when choosing to either perform or not perform *X*. However, as the above reflection has made explicit, Frankfurt's argument relies upon the truth of determinism and the rejection, therefore, of the idea that Jones was ever free: indeed, following from the above, it must be inferred that Jones' brain, thoughts, and actions are entirely determined by his biological drives, psychological states, and physiological nature: if they are not, Jones cannot be said with certainty to have been unable to do otherwise.

This realization seems problematic, however, for Frankfurt's conclusion: for it was on account of Jones' being 'manipulated' and 'coerced' by Black that he was thought to be *not* morally responsible. If, however, Jones was equally manipulated and coerced by his

physiological constitution, and the physical laws which determines it, even when choosing to perform *X*, it seems Frankfurt is equivocating on the criteria by which he allocates moral responsibility: indeed, it seems the criteria is altogether lacking, as Jones can be regarded as equally manipulated and coerced when choosing to perform *X* by himself, or when being made to do so by Black.

It follows, therefore, that Frankfurt's counter-example to PAP must be rejected. This is not to provide a positive argument in favor of PAP (although, in all fairness, I can at present see any reason for why it should not be accepted), but only, rather, to show that Frankfurt's proposed critique is inadequate.

## **Conclusion**

In conclusion, I have outlined an argument given by Frankfurt which proposed a critique of the Principle of Alternate Possibilities by providing a counter-example whereby an individual was thought to be *unable to do otherwise* and yet, at the same time, be held *morally responsible for their actions*. To validate his premises, however, Frankfurt was required to demonstrate that an individual could be manipulated to perform an action with an absolute level of certainty. This was achieved, but only by implying the individual acts in accordance with deterministic laws.

I have argued that insofar as this is the case Frankfurt equivocates his criteria for allocating moral responsibility in his example: for insofar as it was on account of coercion that the Jones was thought to be *not* morally responsible, it seems irreconcilable to hold him responsible at all, when his actions are entirely determined by his physiological constitution. Frankfurt's argument against PAP is, therefore, inadequate.

## Bibliography

Descartes, Rene. Meditations on First Philosophy: With Selections from the Objections and Replies. Translated and edited by John Cottingham. Cambridge: Cambridge University Press, 1996.

Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." In Peter van Inwagen and Dean W. Zimmerman edit Metaphysics: The Big Questions. Singapore: Blackwell Publishing, 2008.

Nietzsche, Friedrich. Human, All Too Human. Translated by Marion Faber and Stephen Lehmann. New York: The University of Nebraska Press, 1984.

Van Inwagen, Peter. "The Mystery of Metaphysical Freedom." In Peter van Inwagen and Dean W. Zimmerman edit, Metaphysics: The Big Questions. Singapore: Blackwell Publishing, 2008.

---

## Endnotes

<sup>1</sup> Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility" in Peter Van Inwagen and Dean W. Zimmerman ed., Metaphysics: The Big Questions (Singapore: Blackwell Publishing, 2008), 471.

<sup>2</sup> Ibid.

<sup>3</sup> Ibid., 473.

<sup>4</sup> The totality of his argument is found on Frankfurt, 476.

<sup>5</sup> Ibid., 476. I believe these terms to be of particular significance as they are both necessary for the cogency of Frankfurt's argument; yet, at the same time, devastating to it as they imply a determinism, as will be elaborated in *Part II*.

<sup>6</sup> Ibid., 476-7.

<sup>7</sup> Ibid., 478-9.

<sup>8</sup> Ibid., 472.

<sup>9</sup> Friedrich Nietzsche, Human, All Too Human, trans. by Marion Faber and Stephen Lehmann (New York: The University of Nebraska Press, 1984), 234.

<sup>10</sup> Rene Descartes, Meditations on First Philosophy: With Selections from the Objections and Replies, trans. and ed. by John Cottingham (Cambridge: Cambridge University Press, 1996), 12-15.

<sup>11</sup> Frankfurt, "Alternate Possibilities and Moral Responsibility", 476.

<sup>12</sup> Ibid.

<sup>13</sup> Ibid.

<sup>14</sup> Frankfurt, "Alternate Possibilities and Moral Responsibility", 476.

<sup>15</sup> Peter van Inwagen, "The Mystery of Metaphysical Freedom," in Peter Van Inwagen and Dean W. Zimmerman ed., Metaphysics: The Big Questions (Singapore: Blackwell Publishing, 2008), 460.

<sup>16</sup> Frankfurt, "Alternate Possibilities and Moral Responsibility", 476.