

# ALTERNATIVE POSSIBILITIES, SELF-DETERMINATION AND RESPONSIBILITY

TIHAMÉR MARGITAY  
ZSOLT ZIEGLER

Department of Philosophy and History of Science  
Budapest Univ. of Technology and Economic Sciences  
margitay@filozofia.bme.hu  
zsolt.ziegler@filozofia.bme.hu

## ABSTRACT

We shall argue that flickers are indispensable in Frankfurt-type examples for metaphysical reasons. Frankfurt's argument against the principle of alternate possibilities (PAP) rests on a distinction between self-determined and coerced behavior of the agent. However, such a distinction requires at least a flicker-type difference in the two sequences of events realizing the two types of behavior in the counterexamples used in the argument. Firstly, we analyze the original Frankfurt argument briefly and point out its preconditions. After that, pursuing the debate over the role of flickers we reconsider Hunt's flicker-free example and show that it cannot meet the preconditions of Frankfurt's argument because of Leibniz's Law, that is, because Hunt's case makes a distinction between self-determined and coerced action without there being any difference between the events realizing the two. Finally, we shall conclude that the Frankfurt-type argument against PAP is ineffective for logical and metaphysical reasons since it cannot work without relying on examples involving flickers, i.e., alternate possibilities.

**Keywords:** responsibility, action, alternate possibilities, Frankfurt-type arguments, Frankfurt-type examples, Leibniz Law, Frankfurt, Hunt

## INTRODUCTION

Harry Frankfurt's 1969 paper "Alternate Possibilities and Moral Responsibility" purports to refute the principle of alternate possibilities (PAP) that "*a person is morally responsible for what he has done only if he could have done otherwise*" (829). Frankfurt offers a counterexample in which the agent had no alternatives (could not have done otherwise), yet he is clearly responsible because he acted on his own. In this paper, we analyze the sequences of events the agent's action comprises of in the

Frankfurt-type examples, and we shall conclude that Frankfurt's project is deemed to fail for metaphysical reasons. The Frankfurt-type counterexamples are purported to satisfy two conditions that are inconsistent. They should be examples of clear-cut responsibility based on self-determination *and* they should contain no alternatives. But, in fact, either they are not clear-cut cases of responsible self-determination or a certain kind of alternate possibilities is present in them. This inconsistency of the Frankfurt-type examples derives from the metaphysical structure of the sequence of events leading to the agent's action. The Frankfurtian argument requires two different chains of events: one realizing a self-determined action and another realizing a forced one. But if there is no alternative then the two collapses into one. If the two are different then they are different because of germs of alternatives.

We will also argue that the debate on intuitions concerning the robustness of flickers misses the point because it is the *metaphysical function* of flickers that matters in the Frankfurt-type examples. Finally, some lessons shall be drawn from the failure of Frankfurt's project for responsible agency.

#### FRANKFURT'S ARGUMENT AND HIS COUNTEREXAMPLE

The principle of alternate possibilities is a central thesis of the libertarian account of moral responsibility entailing the incompatibility of determinism and responsibility. If determinism is true then agents have no alternatives (no one could have done otherwise) and, according to the principle, no one could be morally responsible.

Frankfurt-type counterexamples are to show that the agent's self-determination, by itself, is enough to confer responsibility on him for his action irrespective of whether he had alternative possibilities to act. If he acts exclusively by virtue of his self-determination then he is responsible even if he could not have done otherwise. This is a compatibilist undertaking to prove that a lack of alternatives in a deterministic world is compatible indeed with responsibility.

In the well-known original example of Frankfurt (1969), Black wants Jones to kill a person, White. Let us suppose that Jones actually does on his own what Black wants him to do without any interference from Black. Libertarians would say that Jones is responsible for killing White only if he has an option not to do so. Nevertheless, if Jones were about to act differently then Black would know it and would get Jones to kill White by the help of his fancy neurological device (often called neuroscope). In this counterfactual scenario in which Black intervenes, Jones also kills White, just like in the actual scenario. Consequently, Jones does not have alternatives. He does the same in both cases—either on his own or by coercion. Nevertheless, Frankfurt claims, in the actual scenario Jones is morally blameworthy for killing White in spite of the fact that he could not have done otherwise. Therefore, alternate possibilities are not required for moral responsibility.

Frankfurt (1969:836) argues for Jones' responsibility in the following way:

...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, or to withhold the praise to which it would normally entitle him, 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. He would have acted the same even if it had not been a fact. Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it.

This argument rests on that Jones acts on his own *independently* of Black's intrigue. Jones' self-determination (that is, his own reasons and deliberation leading to his decision and action) alone brings about his action *without* any coercion.<sup>1</sup> Coercion is only triggered by some change in Jones' state, by some deviation of the expected course of events that would lead him to some alternative action. The role of (counterfactual) coercion is to exclude Jones' alternative action by forcing the sequence of events back to track leading to the murder. Frankfurt (1969:830) puts the general argument lying behind his counterexample in the following way:

[t]here may be circumstances that constitute sufficient conditions for a certain action to be performed by someone and that therefore make it impossible for the person to do otherwise, but that do not actually impel the person to act or in any way produce his action. A person may do something in circumstances that leave him no alternative to doing it, without these circumstances actually moving him or leading him to do it—without them playing any role, indeed, in bringing it about that he does what he does.

It is clear from this and the previous passages that the argument presupposes a distinction between a self-determined action and a forced one.

The obvious responsibility of Jones and so the cogency of the argument based on the counterexample springs from that Jones performed a purely self-determined action. His action was brought about solely by his deliberation and decision, and coercion had no role in it. Counterfactual coercion had the only role to exclude alternate possibilities. For this reason the example assigns the role of coercion excluding alternate possibilities to a *counterfactual* intervener who has absolutely

<sup>1</sup> Hunt suggests (1997:203) the following formal explication of the relationship of the two causal factors: „[L]et  $C_{\text{g}}$  comprise all the conditions which actually contributed toward S's Aing at t. Suppose further that there is a set of conditions  $C_{\text{N}}$ , *disjoint* from  $C_{\text{g}}$  given which S cannot do otherwise than A at t. The foregoing specifications, including most prominently the presence of a  $C_{\text{N}}$  which is *disjoint* from  $C_{\text{g}}$  define what I shall call a 'Frankfurt scenario.'” (our italics)

no role whatsoever in the actual scenario in which Jones on his own brings about the action.

#### THE FLICKER OF FREEDOM STRATEGY

One possible way—and indeed a prevalent one in the literature—to challenge Frankfurt’s case is to argue that Frankfurt fails to show the irrelevance of the alternatives since a prior sign of an alternative decision of the agent (called flicker) must always be present in Frankfurt-type examples to call for the intervention of the evil manipulator. Black follows Jones’ thoughts by his neuroscope in order to see whether or not Jones is willing to kill White. The neuroscope has to detect a sign of Jones’ intention not to kill White. This suggests that Jones was at least able to become inclined not to kill White before Black could intervene. Thus Jones does have a choice of some sort even if Black is ready to intervene if necessary. Flickers represent alternatives for the agent. To preclude Jones’ alternative *actions*, a counterfactual intervener has to rely on a trigger, thus it is not possible for him to preclude all the alternatives (i.e. the trigger itself) of Jones. So Frankfurt-type examples do include alternatives.

According to libertarians, Jones does have alternative possibilities (he could have been, at least, inclined not to kill White), and this is precisely what explains why we hold him morally responsible. On this interpretation of the Frankfurt examples, we hold Jones responsible in the actual scenario not because he acted on his own but because he has not *tried* to do something else even if he could not have done otherwise eventually.

For Frankfurtians this rejoinder rests on *irrelevant* and *contingent* components of the examples. They tried to play down the significance of the flickers, and, eventually, devised examples without flickers. Fischer (1995: 134-47) contended that flickers are not sufficiently robust to ground responsibility. His point is that the kind of alternatives found in the Frankfurt-type examples are too “flimsy and exiguous” to play a part in grounding moral responsibility. For example, when Jones’ thought of not killing White begins to form at the unconscious level and when it is realized by a certain neural pattern then the neuroscope can intervene and turn Jones’ neural firings back to the path leading to his killing White. This unconscious mental episode could be, say, a bad feeling about the murder. If we take this interpretation of the flicker phenomenon then Fischer’s point seems to challenge the libertarian rejoinder. It would be counterintuitive to say that the possibility of some unconscious mental episode can ground our moral assessments (see also Fischer 2008).

In the last several years, it was part of the flicker debates to argue about the sufficiency of the “size” of the flicker phenomenon.<sup>2</sup> We, nevertheless, do not want take

<sup>2</sup> About the dialectical role of the flickers in this debate, see Timpe 2006.

part in this debate. We think that flickers have a different function in the examples and their robustness does not matter. We will argue that flickers play a metaphysical role in making coercion independent self-determination possible.

As a result of the robustness debate, philosophers devised examples with more and more insignificant flickers and finally they came up with examples involving no flicker at all. Preventing the flicker criticism, they wanted to eliminate even the flicker-type alternate possibilities, e.g., by a blockage (Hunt 1997 and 2006). Hunt's blockage-example attempts to retain all the important characteristics of Frankfurt's original example but avoid the occurrence of flickers. In a blockage case, all alternative sequences of events other than the neural firings and muscle contractions of the causal chain that brings about the required decision and action are blocked. The agent is not able to do or try to do anything other than what he has decided for and what he actually does. The blockage is designed to leave the agent's self-determined decision and action intact and to preserve the agent's moral responsibility.

Nevertheless, the blockage-example failed to convince the critiques since it seems to them that the agent's particular decision and action is forced by the blockage. The blockade does not only determine that the agent has no alternatives but it can also determine his decision and action. While compatibilists and libertarians could not agree on the reason *why* the agent is held responsible in the original Frankfurt-examples, now libertarians cannot see any responsibility in the examples without flickers at all. As if agents would lose responsibility as flickers are diminishing. Eventually the debate seemed to get stuck in the contrary intuitions.

#### AN ANALYSIS OF THE METAPHYSICAL FUNCTION OF FLICKERS

The significance of flickers might be proved without any reference to intuitions. We think that flicker-type alternatives have a metaphysical function in distinguishing the self-determined action from the compelled one in the Frankfurt examples. In order to demonstrate the metaphysical role of flickers, let us examine the structure of the blockage-example.

First of all an important tacit assumption should be made explicit: it is presupposed in the blockage example that the actual neural pathway *is* activated because otherwise Jones could also remain inactive and he would not kill White despite of the blockade. So *ex hypothesi* Jones' process of self-determined action can involve either the stimulus running through the pathway leading to murder or its turning away from this path at a neural junction when Jones changes his mind *but it cannot* involve the extinction of the stimulus.

Of course, the stimulus cannot turn away because of the blockade. Jones' state of mind and brain at the time of the installation of the blockage activates exactly the same neurons and muscles of Jones as the blockade. That is self-determined and forced action can consist of exactly the same sequence of events. After the elimina-

tion of the flickers, the stimulus bumping into the blockage is exactly the same as its running directly toward the action.

Note that the blockade by itself cannot bring about Jones' decision and action. But once the blockade is in place *and* some stimulus from the early phase of the sequence of events realizing his decision and action is present in Jones' brain then the blockade and the initial stimulus bring about exactly the same decision and action as Jones' self-determination does. To maintain the analogy with the original example, let us assume that the blockade is mounted when, say, Jones realizes first time at  $t_1$  that he dislikes White.<sup>3</sup> From this point two sets of factors can bring about the same deliberation, decision and action of Jones: (1) the stimulus present at  $t_1$  and the blockade, (2) Jones' own reasons and deliberation leading to his self-determined decision resulting in his self-determined action.

It should be clear that the problem is not that they *together* bring about the same effect but that they can bring about the realization of his action *separately*. His self-determination on the one hand, and the blockade on the other (plus other necessary factors) constitute two sets of circumstances each of which are sufficient to bring about his action. His action is overdetermined by these two sets of factors.

Overdetermination has the consequence that the division of labor between the two sets of causes required by Frankfurt's argument does not hold anymore. Notably, the blockade by itself can make Jones realize his action *and* it does not only prevent him from doing something else. Similarly, his self-determination by itself can make him realize the same action *and* not something else. So when Jones kills White in the Hunt scenario, the two sets of conditions (why he did what he did and why he could not do otherwise) cannot be separated. It is impossible to attribute Jones' decision and action to his self-determination and his not doing something else to the blockade.

In Frankfurt's argument Jones clear-cut responsibility rested on that his action was purely self-determined, independent of any coercion. His action being self-determined and independent of coercion presupposes a distinction between self-determined and coercive action.

Well, how would it be possible to distinguish a sequence of events realizing Jones' self-determined action from a sequence of events realizing Jones' action that he was forced to perform? They can be distinguished either by the causal factors that bring about them or by the events they consist of. In Frankfurt's example the two types of sequence of events were different regarding both aspects. However in Hunt's example the sequences of events cannot be different from either point of view.

As to the distinction by the causes, in Frankfurt's example the coercive force is triggered only by the flicker. Without a flicker, Jones' decision and action are brought about by his self-determination, if a flicker occurs then Jones is forced by Black to

<sup>3</sup> Only that much is clear from the original Frankfurt example that Black begins his project before Jones' decision, and he is only watching in the background while Jones' deliberating and deciding for the murder.

kill. Jones' self-determined action is clearly different from his coercive action because the two have clearly different causal antecedents. This is not so in Hunt's example. The two sets of causal factors—Jones' self-determination and the blockade—are both present and causally effective in bringing about Jones' decision and action. His action cannot be traced back to either of these sets of causes. Therefore the causal antecedent of his action cannot make a difference between self-determined and coercive action. The concurrent presence of two sets of causal factors derives from the structure of Hunt's flicker-free counterexample. Both sets of factors should be there and they both should be causally effective if one of them is to produce the action and the other is to exclude even flicker type alternatives *at the same time*.

If both sets of causes must be there then a self-determined action can be distinguished from a forced one only if the sequence of events realizing the first is different from the sequence of events realizing the second. However, in the Hunt example, there *can* be no difference between the two sequences of events because the blockade lets only one sequence of events happen that is exactly the same that is determined by Jones' self-determination. And where there can be no difference, there should be no distinction according to Leibniz' Law. (Leibniz 1969: 308) Then it is merely a terminological issue, a rhetoric device (called framing) to call one and the same sequence of events as self-determined or forced action. Of course the entities, Jones and the blockage can be distinguished but once the blockage is in place, the Jones-blockage complex can produce only one relevant sequence of events, as opposed to the Jones-Black complex in the original Frankfurt example that can produce two different sequences of events and the flicker is the difference between them. This is why it is impossible to distinguish the agent's self-determined action from his forced action in the blockage case while it is possible in Frankfurt's example including flickers. The flicker itself is the difference on which the distinction between the sequence of events realizing his self-determined and forced action can be based.

It is a crucial condition of the application of Frankfurt's argument to the blockade example that the fact that the blockade is there should not play any role in Jones' action. However, the blockade example does not meet this condition without the distinction between the self-determined and forced action that was made possible by flickers.

If such a distinction cannot be made then it is mere rhetoric to say that in the blockade example Jones' action is a self-determined one. It makes even less sense to claim that the agent acted exclusively on his own *independently of coercion*. This knocks down Frankfurt's argument for Jones' responsibility because, as we have seen above, it rested on this premise. In sum, the blockade example is not an example of self-determination and thus no clear-cut example of responsibility, consequently no counterexample to the principle of alternate possibility.

Our analysis can illuminate the intuition stalemate in the compatibilist-libertarian debate. In the blockade example, we have only one possible sequence of events and

that, when viewed as a self-determined action, will establish the responsibility of the agent while the same sequence, when viewed as a forced action—because of the blockage—will exempt him from any responsibility. This is the reason for the stuck in contrary intuitions. It is due to a metaphysical mistake: parties in the debate want to make a distinction where there is no difference. Even though flickers as action germs may be insufficient to ground moral responsibility, as Fischer contends, they play a significant role in distinguishing self-determined action from forced one. Thus the significance of flickers is not the matter of intuition—as many philosophers seem to think.

So Frankfurt's argument does not work in the blockage example. If both self-determination and coercion are present and they bring about exactly one and the same sequence of events, then this sequence of events cannot realize a purely self-determined and coercion independent action. For the action to be brought about independently of coercion as it is required by Frankfurt's argument, coercion should be removed from the scenario when Jones acts voluntarily. If coercion is removed then it needs some trigger in order to prevent Jones' alternative action. Coercion must be triggered by some flicker, that is, by some deviation from Jones' original, self-determined chain of events leading to his killing White.

This implies that *the possibility of flickers must be built in an example in which the agent can act on his own independently of any coercion and he is deprived of alternatives*. The possibility of flickers (that is the agent can be inclined to do otherwise) is necessary for coercion independent self-determination without alternatives. If the agent had no alternatives and he did what he did as a self-determined action independently of coercion, then, at least, he could have been inclined to do otherwise, even if he could not have done otherwise eventually because of some coercion.

## AN OBJECTION

So far we argued that the blockade and Jones' self-determination produce one and the same sequence of events. Therefore, on the basis of the actual sequence of events, it is impossible to distinguish between events realizing self-determined action and events realizing inevitable (i.e., forced) action. Now we should turn to an objection to our conclusion.

Namely, perhaps there is another reason independent of the features of the actual sequence of events why we should take the only actual sequence of events as a self-determined action instead of a forced one. At this point the blockage-builder may come up with a different argument about how to use his case as a counterexample to PAP. He may contend that his example is a case of self-determined action despite of the presence of the blockage. The blockade, so the argument goes, is causally irrelevant because the sequence of events would be the same if the blockade were

removed. (Hunt 1998:218) Jones' considerations and neural processes leading to his decision and his action would be exactly the same without the blockage. Therefore Jones' decision and his action is the result of his self-determination and rather than the blockade. The latter rules out only the alternatives but does not contribute causally to the action.

First we should note that this rejoinder refers to a counterfactual situation. It is true that Jones' decision and action are independent of the blockade in the counterfactual situation in which the blockade is removed but not in the actual situation in which it is in place.

It is easy to show the weakness of the blockade-builder's contention. As we have showed earlier *ceteris paribus* the blockade (and the stimulus at  $t_1$ ) brings about exactly the same sequence of events realizing Jones' decision and action that Jones' self-determination does. Thus we can come up with a counterfactual argument like the blockade-builder: When Jones realizes that he dislikes White at  $t_1$ , the blockade would produce exactly the same sequence of event (i.e., Jones killing White) even if Jones had no reasons and did not want to kill White at all. Therefore Jones' self-determination is causally irrelevant. It is clear that solely on the basis of the *counterfactual situation* it is impossible to argue the causal irrelevance either of the blockade or of Jones' self-determination in the *actual situation*. In the blockade case both are causally relevant and Jones' action is overdetermined by them.

The conditions eliminating alternatives in the original Frankfurt example are counterfactual while the similar conditions in the blockade example are actual. The other side of this is that the reason why we should take Jones' action self-determined refers only to actual circumstances in the original Frankfurt example but involves a counterfactual premise in the blockade example. The first can properly establish his self-determination and responsibility in the actual situation while the second cannot.

## CONCLUSIONS

Frankfurt and the Frankfurtian project trade on the separation of what leads to an action from what blocks alternative actions. In order to produce a scenario in which the agent can act on his own independently of coercion, coercion should be effective only in blocking alternatives.<sup>4</sup> If there is nothing to be blocked, that is, if there is no flicker that would—without coercion—lead to alternative action, then no coercive factor should be present.

It follows from our argument from Leibniz' Law that flickers are not contingent

<sup>4</sup> In the compatibilist version of PAP, Frankfurt suggests (838) the opposite: „a person is not morally responsible for what he has done if he did it only because he could not have done otherwise”. But of course we discuss and defend the original PAP, the PAP of the Libertarians attacked by Frankfurt.

but necessary components of Frankfurt-type examples. Because if there can be no flicker when alternatives excluded then there can be no difference (like in the blockage-example) grounding the distinction between the sequence of events constituting a self-determined action and the sequence of events constituting a forced action. If there can be no difference between self-determination and coercion then there cannot be self-determination that is independent of coercion. Therefore if there can be no flicker when alternatives excluded then there can be no coercion-independent self-determined action in which responsibility can be grounded. Consequently, there can be no flicker-free counterexample to PAP in which the agent is responsible because he acted on his own even though he had no alternatives. Flickers are necessary in any counterexample that can work in a Frankfurt-type argument against PAP.

The Frankfurtian argument failed to refute PAP by a counterexample. We agree with Frankfurt that a coercion-free self-determination indeed grounds responsibility. In Frankfurt's original example, Jones is responsible because he did what he did on his own, free from any coercion. However, as the above argumentation showed, the agent can perform a self-determined action that is independent of any coercion *only if* he is able to become inclined to do otherwise. If there is responsibility because of coercion-free self-determination then there must always be alternate possibilities as well. In principle Frankfurtian compatibilism cannot achieve its goal and it is a dead end.

The failure of refutation of PAP, by itself, does not provide further support for it. However, the reason *why* the Frankfurtian argument failed can reinforce belief in PAP. The project failed because it turned out that, on the one hand, if it is a self-determined and coercion-free action entailing responsibility then it should involve alternatives at some point. On the other hand, however, examples in which alternatives are eliminated are not cases of self-determined and coercion-free action and, thus, no clear-cut examples of responsibility anymore. This suggests that responsibility, in fact, depends on alternatives.

We can go one step further and conclude from the failure of the blockade-type examples that the agent bears responsibility by virtue of his self-determination *only if* he has also alternatives. Our first premise is the uncontroversial connection between self-determination and responsibility that is the common ground of the debate between compatibilists and incompatibilists. (1) Self-determination can establish responsibility *only if* it is coercion-free.<sup>5</sup> The second premise is a version

<sup>5</sup> Frankfurt (1969) himself admits after having analyzed the Jones<sub>1</sub> – Jones<sub>3</sub> cases that coercion-independent self-determination is the only kind of self-determination that can establish responsibility without provoking objections. He concedes this for the sake of the argument. (Though he suggests a modification to PAP at the end of his paper (838) that runs contrary to his earlier concession).

of our conclusion above: (2) Self-determination can be coercion-free only with (at least, flicker-type) alternate possibilities. Then it follows from these two that *a responsible self-determinate action must involve alternate possibilities*. An agent is responsible for performing a self-determined action only if he could have been inclined to do otherwise.

Eventually the criticism of Frankfurt's argument seems to supply us—surprisingly enough—an argument bearing out what he wanted to refute, namely that alternate possibilities are indeed necessary for responsibility.

#### ACKNOWLEDGEMENTS

We are grateful to István Danka for his painstaking and helpful comments on earlier versions of his paper.

#### REFERENCES

- Fischer, John M. 1995. *The Metaphysics of Free Will*. London: Blackwell.
- Fischer, John M. 2008. Responsibility and the Kinds of Freedom. *The Journal of Ethics* 12: 203-228.
- Frankfurt, Harry G. 1969. Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy* 66: 829-839.
- Hunt, David P. 1997. Moral Responsibility and Unavoidable Action. *Philosophical Studies* 97: 195-227.
- Hunt, David P. 2006. Freedom, Foreknowledge and Frankfurt. In: Widerker and McKenna 2006:159-184.
- Leibniz, Wilhelm Gottfried. 1969. Discourse on Metaphysics, Section 9 in.: Loemker 1969.
- Loemker, L. 1969 (ed. and trans.) G. W. Leibniz: Philosophical Papers and Letters. 2nd ed. Dordrecht: Reidel.
- Timpe, Kevin 2006. The Dialectic Role of the Flickers of Freedom. *Philosophical Studies* 131:337-368.
- Widerker, D. and McKenna, M. (eds.) 2006. *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Aldershot: Ashgate.