Original Paper UDC 159.947.2: 159.947.23: 159.947.24
Received September 20th, 2007

Michael Pauen
Humboldt-Universität zu Berlin, Institut für Philosophie, Berlin School of Mind and Brain,
Unter den Linden 6, D-10099 Berlin
pauen@mail.com

Self-Determination

Free Will, Responsibility, and Determinism

Abstract

An analysis of our commonsense concept of freedom yields two "minimal criteria": (1) Autonomy distinguishes freedom from compulsion; (2) Authorship distinguishes freedom from chance. Translating freedom into "self-determination" can account for both criteria. Self-determination is understood as determination by "personal-preferences" which are constitutive for a person. Freedom and determinism are therefore compatible; the crucial question is not whether an action is determined at all but, rather, whether it is determined by personal preferences. This account can do justice to the most important intuitions concerning freedom, including the ability to do otherwise. Waiving determination, by contrast, would violate the minimal criteria rather than providing "more" freedom. It is concluded that self-determination provides everything that we can ask for if we ask for freedom.

Keywords

free will, self-determination, responsibility, determinism, alternative possibilities

The ability to act freely and to take responsibility for one's own decisions is certainly among the most distinctive features of human beings. But since actions occur in the physical realm that is governed by natural laws, the question comes up whether free will and determinism are compatible. The view that they are not is supported by strong and suggestive arguments. Paradoxically, however, the view that *in*determinism and freedom might be incompatible has recently received increasing attention, too. Taken in isolation, the latter position might comfort compatibilists, but if combined with traditional incompatibilism, it leads to a serious problem: According to this position that has been elaborated by Galen Strawson, Thomas Nagel, and, recently to some extent by Peter van Inwagen, the basic criteria of freedom itself are mutually incompatible. Thus it is impossible, even in theory, that free actions exist, no matter what science will tell us about the goings on in the natural world.

This position sounds disturbing. It would lead us to concede that one of our basic commonsense concepts that plays a constitutive role in our view of humans as responsible agents is incoherent and therefore useless. I think, however, that we should *not* give up the search for a coherent analysis. Ideally, such an analysis meets two requirements: First, it is strong enough to do justice to our commonsense intuitions concerning freedom. Second, it can be defended against the demand for stronger requirements.

In the following paper, I will demonstrate that a proposal that comes pretty close to such a "strong analysis", as I will call it, is in fact available.

Two widely undisputed "minimal criteria" for freedom will provide the basis for my proposal. According to these requirements, that I will outline in the first section, freedom has to be distinguished from compulsion and from chance likewise. It will turn out that this distinction can be made in a very straightforward way if we translate freedom into "self-determination". If this is so, then, of course, we should say something about the "self". I will do this in the second part. The third section is dedicated to the first requirement of a "strong analysis": I will demonstrate that the present proposal can account for one of the most important intuitions concerning freedom, namely the so-called "Principle of Alternative Possibilities", and that it provides a basis to reject the "Consequence-Argument". As far as the demand for stronger requirements is concerned, a definite answer is obviously impossible, although it should be possible to defend the proposal at hand against actual demands for stronger account. I think that the most obvious demands concern an elimination of determinism. That's why I will show in the fourth and final part that introducing indetermination does not give us a stronger analysis but leads to a conflict with the minimal criteria. I conclude that the self-determination account of freedom comes quite close to a "strong analysis".

I. The Concept of Freedom

Initial Considerations

In what follows, I will treat freedom basically as a property of actions. Actions are, of course, performed by persons, but the freedom of a person seems to result from the freedom of the actions she performs – not the other way around. I will not go into the intricacies of action theory, rather, I will assume that actions are mental or bodily doings that have some psychological explanation referring to the person whose doings they are. Actions may be based on explicit decisions or acts of will. If so, a decision that leads to a free action may also be said to be free, and vice versa. In any case, I will make an important distinction between wishes and decisions: The latter unlike the former require that the state of affairs, they refer to, is made real. You may have the *wish* to stop smoking and still continue to smoke, but you cannot be said to have *decided* to stop smoking unless you really do so.

Still, the suggested conceptual framework does not imply any commitment to a special account of freedom. It does not imply, in particular, a commitment to the traditional "freedom of action" account. If you think that freedom is a property of actions, you may still think that an action is free if and only if it was agent-caused or if the underlying act of will was undetermined.

It seems undisputed that free actions need not be performed in the complete absence of external constraints. Thus, if we say that some person *p* was free to do *x* under conditions *c*, it may be true that she was *not* able to do *w* or *v* due to certain external circumstances, although she could have done *y* instead. This difference might become important if *x* and *y* are very narrowly related while *w* and *v* provide "real" alternatives. The defendant might have had the alternative to stab rather than shoot his victim but he had no choice to refrain from killing altogether. It would therefore be misleading or at least unclear to say that he was free when he committed the crime – even if he was able to do otherwise in some sense. We can account for this consideration if we mention

the alternatives explicitly. We might say that *p* was free when he did *x* rather than *y* under conditions *c*. Thus, an action whose freedom is discussed should be determined explicitly in relation to the available alternatives. Another advantage of this way of formulating the "freedom statement" is that it makes it unnecessary to scrutinize the whole etiology of *x*. What we need, is only an answer to the question why it was *x* and not *y* that was performed. There may be a long story to be told if you ask why the camel's back broke; still it may have finally been up to a single straw to break the camel's back rather than sparing it under conditions *c*.

One of the subjects of my paper is the relationship of freedom and determinism. In discussing this point I will make no assumptions whether or not determinism is a positive fact in our world. I use determinism as a hypothesis that allows us to explore the relationship between free actions and facts about our world. I will not discuss possible impacts of a physical realization of mental properties, because I think that interest in this question if based on the assumption that physical realization leads to determinism and determinism is the real issue. Besides that, I think that even a dualistic world might be deterministic.

Finally, I would like to say a word about the difference between compulsion and determination. Compulsion implies some kind of determination, that is, it implies that a particular kind of doing is brought about or prevented by factors external to the agent whose doing it is. Still, there are many cases of determination that are no cases of compulsion. Although the performance of a computer might be determined by its program, we would not say that the program *forces* the computer to do what it does. The important difference seems to be that compulsion unlike determination requires that an opposite will was overcome. Parents can force a child to do *x* if the child prefers *y*, but we would not say that the child was forced if it had the wish to do *x* anyway.

Authorship and Autonomy

But let's get back to the concept of freedom. What are the "minimal criteria" that have to be met by any action in order to count as free? I have already said that I will begin with two basic features. It may be true that a "strong analysis" eventually leads to more demanding requirements but I will start with those features that seem to be almost universally accepted as necessary conditions.

First, a negative criterion: Freedom implies the absence of compulsion. If we say that the defendant was free to do *x* rather than *y* under conditions *c*, this implies that he was not forced to do *x* rather than *y*. Thus we can say that free actions must comply with the *principle of autonomy.*

The second, positive feature is of equal importance. Free actions have to be distinguished from random events. We would not say that the defendant was free to do *x* rather than *y* in conditions *c* if it was only a random neural activity that brought about *x* rather than *y*. The obvious way to make this distinction is to say that free actions can be ascribed to an author. Thus, free actions must comply with the *principle of authorship.* I will give a more detailed account of the criteria of authorship below. Presently, I would like to stress two points only: First, it seems that only intentional beings, that is, persons with certain desires, beliefs, and dispositions particularly concerning the objectives and the consequences of what they are doing, can count as authors in general. Second, it appears that a person counts as the author of a particular event only

if she acted out of her own desires, beliefs, and dispositions, rather than just accidentally causing the event. Another way to phrase this constraint is to say that the author's desires, beliefs, and dispositions should contribute to an explanation of why she did *x* rather than *y* under conditions *c*.

We can summarize both criteria in a single requirement if we say that actions that are free in the minimal sense above are *self-determined*. In fact, nothing counts as self-determined unless the two criteria above are met. First, self-determination requires the autonomy. We would not say that an action is self-determined if we know that it was brought about by force. Second, self-determination requires authorship. That's what we mean if we say that someone determines herself. Trivially, an activity that is brought about unintentionally or just by chance does not count as self-determined just because the determination by the self is missing. Taken together, these necessary conditions are also sufficient: If an action meets the minimal criteria of autonomy and of authorship, it will count as self-determined.

Note well that self-determination, although it rules out certain kinds of determination, namely external determination, seems to imply other forms of determination, namely determination by the self whose action it is. Thus, whatever the features may be that are constitutive for a particular person *p*: If it is due to these features, say certain beliefs, desires, and dispositions, that *p* did *x* rather than *y* under conditions *c*, then this action would count as self-determined, just for conceptual reasons. I will argue that this remains true even if the individual features, together with external factors, determine what *p* does.

II. Personal Capabilities and Personal Preferences

It is completely unclear what it means to act in a self-determined way, as long as it remains to be spelled out what the "self" is. Note that the "self" that is at issue here is not a particular, let alone a non-physical entity, like the Cartesian or Ecclesian mind. "Self" is just an umbrella term for those features and capabilities that are constitutive for an individual agent. Even if you disagree with the account of these features that I will give below and even if you doubt that self-determination really captures genuine freedom, you have to say something about what it takes to be an agent and what the features are that constitute a particular individual agent in contrast to other individual agents. Without a concept of an agent you would not be able to say whose action is not free and who is the subject of determination or compulsion. A second important point is that the criteria below don't commit you to any *empirical* claim concerning the existence of selves, much less of self-determined actions. Even if you accept the above analysis as well as the features below, this would give you only a *standard* for empirical investigations whether or not selves and self-determination exist.

I think that, by and large, the relevant features fall into two categories. First, there are those more general abilities that *every* conscious being has to have in order to count as a self that is able to determine her own actions. In what follows, I will call them "personal capabilities". Second, you need "personal preferences" that distinguish one particular self from other selves. These features are needed in order to determine whether or not a person is the author of a particular action.

1. Personal Capabilities

Let me start with a specification of those "personal capabilities" that everyone needs in order to count as a self in the sense that is required here. Since self-determination implies authorship, a robust and intelligible connection between the action and the beliefs, desires, and dispositions that are constitutive for an agent is required. It follows, first, that the agent must be rational in a weak sense, such that an intelligible connection can be established between the action and his particular beliefs, desires, dispositions – no matter what these beliefs, desires, and dispositions may be. Without such a connection, it could not be made intelligible that she did *x* rather than *y* in situation *c* and the action would fail to meet the requirement of authorship. Random "decisions" cannot be reduced in an intelligible manner to underlying preferences since, trivially, any random decision is compatible with any given set of preferences in any situation. Rational decisions, in turn, may involve a kind of rational calculus that balances competing preferences, e.g. antagonistic long term and short term desires, thus enabling an agent to determine which of the available options fits best to her overall set of preferences. This would include, second, the ability to assess the relevant consequences of each of these options. Imagine that *p* gave a bottle of whisky to *q* who died shortly after drinking the whole bottle at once. Whether or not *p* counts as the author of the killing of *q* depends upon whether *p* knew the consequences of what she did. If *p* was aware of the fact that *q* was an alcoholic, that he was seriously ill, that he would drink the whole bottle at once which, in turn, would kill him, then we might say that it was *p*'s action to kill him.

I will call an agent who meets these requirements a "rational agent". Consequently, if it turns out that *p* is *not* a rational agent she will *not* be able to act in a self-determined manner in general.

2. Personal Preferences

Rational Principles

Needless to say that not every action of a rational agent will be self-determined. But how could we spell out those criteria that might help us to decide whether or not a *particular* action of a *particular* agent is self-determined? In order to do so, we need preferences that are constitutive for an individual person. In what follows, I will call these preferences "personal preferences". Personal preferences are those beliefs, desires, and dispositions that are distinctive for a particular agent. They help us to make a connection between an individual self and a particular action. I have already said that not all the features, say the desires, beliefs, and dispositions that an agent actually has, count as personal preferences in the sense that is required here. But how do we distinguish those features that are part of the self from those that are not? How do we distinguish between personal and non-personal preferences?

The answer to this question is of central importance because it determines whether or not the present account goes beyond weak concepts like the traditional freedom of action account. Intuitively, strong, stable, and well grounded rational beliefs may be typical examples for personal preferences while transient dispositions or psychological addictions may be typical examples for non-personal preferences.

First, in order for something to qualify as a personal preference rather than a transient motive, it should have a certain temporal stability that is, it should

last for some time. It is difficult to specify the relevant sense of temporal stability, but let's say that a minimal criterion is that the feature in question persists for more than one day and determines more than one action. You will not count as a passionate lover of Italian Operas when you are listening to an opera for the first time in your life.

Taken by itself, this criterion is clearly insufficient: Compulsions are often very stable although they do not count as personal preferences. Thus, we need a second criterion. One might feel tempted to demand rationality in a stronger sense, such that not only the choice between the existing preference but these preferences themselves have to comply with rational principles. Rational agents, after all, should accept rational principles. Therefore, they can be said to act in a self-determined manner if they follow those principles. And if you think that moral principles can be rationally justified, then your actions should count as self-determined if they meet moral principles. The problem with this view is that it would lead us to conclude that one never acts in a self-determined manner if one acts irrationally. Again, given that moral principles can be rationally justified, we had to conclude that no one acts in a self-determined manner if he violates moral principles and nobody would be responsible for immoral acts.

I agree that rationality, together with the required temporal stability, is a sufficient criterion for a rational agents' personal preferences, but I don't think that it is necessary – not only because it seems counterintuitive to say that you are responsible only for your moral actions, but also because I don't think that humans are rational beings only. It seems evident to me that irrational temptations may be part of one's self.

A possible way out of this situation has been suggested by John Martin Fischer (1994, 164–168). Fischer proposes to replace the above criterion of rationality by a criterion that he calls "weak reasons-responsiveness". According to Fischer, you are weakly reasons-responsive with respect to a certain action, if *you are able, in principle,* to respond to reasons as far as this action is concerned, *even if you do not in your present execution of this action.* Even if you follow an irrational temptation in your present action, you may still be weakly reasons-responsive, provided that it is possible for you to respond to much weightier reasons in an otherwise similar situation.

Unfortunately, this requirement is too weak because it can be met by actions that are clearly neither free nor self-determined. There may be situations in which even an addict would respond to reasons, otherwise only a few addicts would have ever decided to undergo a withdrawal treatment. Still we would not say that addicts who would make such a rational decision under appropriate conditions act freely even in those cases when they succumb to their addiction.

Possible Subject to Self-Determination

It would seem, then, that we need an alternative. I think that we have two possibilities: Let's call the first one "identificationist" and the second one "liberal". Both are compatible with the present account, but I have a clear preference for the second one. According to the first one, something qualifies as a personal preference if and only if it is a possible subject to approval by the person whose preference it is. It is not required that a personal preference has been actually approved; the idea is only that, should the person start to reflect upon the feature in question, then she would accept it, maybe even

"wholeheartedly". The rationale behind this idea is that there is a core of preferences that constitute an agent's personality. Thus, for an action to count as free, it must depend upon these very core features. Note that this view does not require that you can decide against a core feature, because such a decision would amount to a decision against an aspect of the core of your self. As far as I can see, there are two ways to find out, whether or not a certain feature qualifies as a personal feature according to the identificationist standard. Either you have to ask the person explicitly whether she accepts this feature, or you have to imagine whether or not the person would approve a certain feature in a hypothetical scenario. Your answer would be based on a combination of general psychological knowledge and particular insights concerning the person in question. One of the problems with this option is that it may lead to implausible results: What about a person who identifies with her addiction? The only way to treat this problem that I can see is to stipulate that addictions etc. do not qualify as personal features in general, but this is clearly an *ad hoc* solution.

This is one of the reasons why I prefer the other option. In this case, we do not ask for approval, rather, the focus is on the possibility of a self-determined decision. For something to qualify as a personal preference, it should be a possible subject to self-determination, too. The idea is that it would be unintelligible to treat *p*'s doing *x* rather than *y* as self-determined, while insisting, at the same time, that his doing was determined by factors that are beyond *p*'s control.

But how could we find out whether this criterion is met without ending up in a vicious circle: In order to determine whether something qualifies as a self-determined decision we have to appeal to personal preferences, while the identification of personal preferences, in turn, seems to require knowledge about self-determined decisions. Note, however, that it is *not* required that each candidate for a personal preference is or can be *approved*. The requirement is only that a personal preference *can* be subjected to such a decision and that means that the result of a process of decision-making can be implemented even if the person opts against the preference in question. And this criterion can be verified without reference to actual decisions.

Theoretical considerations might be sufficient in certain paradigm cases. Rational beliefs are paradigm examples. I take it that my belief that *x* is *F* qualifies as a rational belief if I would reject it in the light of convincing evidence that *x* is not *F*. Provided that I'm a rational agent, the rejection of the belief would count as a self-determined decision and the belief would qualify as a personal preference. Likewise, my rational belief that stealing is reprehensible should be a possible subject to a self-determined decision. Conversely, physical or psychological addictions are paradigm cases for features that are *not* subject to self-determined decisions. It is a defining feature of an addiction that it will persist even if I wish to get rid of it. Since all these assessments can be made without reference to actual or hypothetical self-determined decisions of the person in question, there is no circle at least in the paradigm cases.

But what about non-paradigm cases? I assume that psychological or neuroscientific investigation concerning human decision processes and the underlying neural mechanisms can help us in these cases. As a result of such investigation, it may turn out that certain dispositions are *not* amenable to self-determined decisions in general while others are. Extended knowledge about the neurobiology of addiction might be particularly helpful. As a result, it might be established that if a certain behavior or the perception of certain objects

correlates with activity in neural area *a* or with a behavioral pattern *b,* this could indicate that the person in question is not able to make self-determined decisions on behalf of those desires or dispositions. If certain areas in my brain light up in an fMRI scanner when I hear a Verdi-aria, this might indicate that I cannot make a self-determined decision concerning my love for this kind of music. It is obvious that such assessments are fallible and that there will remain a considerable number of doubtful cases; but that is what you have to expect in free will questions anyway. Still, the examples show that assessments concerning personal preferences can be made independently from any actual or hypothetical decisions of the author, even in the non-paradigm cases. Thus there is no vicious circle in these cases either.

The main difference between the liberal and the identificationist option is that the latter, unlike the former, accepts features that are no possible subjects to change, in principle. If I would wholeheartedly approve my love for Italian operas, then this feature would count as a personal preference, according to an identificationist position. The liberal position, by contrast, would deny this, given that I would not be able to decide *against* this feature. It will turn out that this liberal requirement is particularly helpful for a defense of the *Principle of Alternative Possibilities*. If my decision to do *x* rather than *y* in situation *c* is based on personal preferences that I cannot change, then it seems difficult to deny that I could not have acted otherwise. Although I'm not sure that the *Principle of Alternative Possibilities* is really incompatible with the identificationist position, the principle seems at least difficult to defend on such an account. Because I think that the *Principle of Alternative Possibilities* captures an important intuition concerning freedom, I think that the liberal position is preferable.

All this does not mean that personal preferences are subject to random changes. Self-determined decisions depend upon one's personal preferences, even if these decisions, in turn, concern personal preferences. If I change my former belief that abortion is acceptable, then this will be a self-determined decision only if I have other beliefs and dispositions that make it reasonable to make this decision, say because I started to reflect upon assumptions that seemed to justify my former belief or because I have acquired new information about the cognitive capabilities of embryos, etc. This qualification is important because it shows that beliefs or desires that I never dreamt of changing may be personal preferences, even in the liberal sense. The criterion is that these features *would* change, *should* I wish to do so; still it would be perfectly unreasonable to make such a decision, given the whole system of my other beliefs and desires. Thus, my self-determined decision will be to keep this belief.

In addition, I would like to stress the differences between the present proposal and Frankfurt's theory of second-order volitions. One difference is that the present proposal does not imply different orders of decisions. Freedom does not depend upon certain relationships within a hierarchy. The problems with such a hierarchy are notorious: It is difficult to see how a second-order decision should guarantee the freedom of a first-order decision just because the former has a certain position within the hierarchy of decisions. Either, the hierarchical relation is decisive, but then the second-order decision will need a third-order decision in order to qualify as free and so forth, thus we would end up in an infinite regress. Or the second-order decision meets a certain criterion, but then the position within the hierarchy cannot be decisive and we might ask why the criterion does not work for first-order decisions as well.

Thus, we would, and I think we *should*, get rid of the whole hierarchy, at least as a requirement of free decisions.

This is one of the reasons why I think that the absence of a hierarchy it is an advantage of the present proposal. The absence of any hierarchy follows also from the symmetry between personal preferences and decisions concerning these features: Every personal preference can be part of a decision on actions or other personal preferences. Another difference is that the present self-determination account does not require that personal preferences or other elements or results of the process of decision making are approved. The only requirement is that the result of a self-determined decision process can be implemented, no matter what the result will be.

Finally, let me stress why the present proposal goes beyond the traditional "freedom of action" account. It does so because it provides criteria that allow us to identify actions that are not self-determined although they conform with an act of will. According to the present proposal, such an action is not free if the act of will in question is determined by non-personal preferences, that is, by features that are no possible subjects of self-determined decisions. Consequently, this account can be defended easily against the typical objections that might be brought forward against freedom of action accounts: Psychological or physiological addictions are non-personal preferences since they are not subject to self-determined decisions. Actions that are determined by such features do not count as self-determined, according to the present proposal, although they may conform with an underlying act of will and thus would count as free according to the standard freedom of action account.

Another Nefarious Neurosurgeon

But what would happen, if one of the notorious nefarious neurosurgeons would change *p*'s personal preferences tomorrow night, implanting her a completely new system of beliefs, desires, and dispositions? While the belief that stealing is unacceptable is constitutive for her present self, let the neurosurgeon implant her the belief that stealing is acceptable, given the unjust distribution of property in the present society. Assume also that, although *p* is able to make a self-determined decision concerning this belief, she will keep it, given her new system of beliefs, desires, and dispositions. It would be intuitively unplausible to say that the actions she performs afterwards are *p*'s free actions. But the present account seems committed to this very position: If what results from the neurosurgeon's intervention is a personal preference, then the ensuing actions have to be counted as self-determined.

On reflection, however, things turn out to be a bit more complicated. Note, first, that it was required that personal preferences have some kind of temporal stability. This might solve the problem immediately after the nefarious neurosurgeon's intervention. Thus, the present account would not be committed to say that an action that is performed immediately after the intervention is free, if it is determined by one of the manipulated features. But what about "her" actions two months after the manipulation? No matter what the criteria of temporal stability are, the present account is in fact committed to the view that any action that has been determined by personal preferences is a self-determined action, no matter how these features came into being. Note that the emergence of preferences prior to *p*'s becoming a rational agent is not, and for that matter *cannot* be, a subject to self determined decisions, either.

It follows that the action in question has to be treated as self-determined. Another question, however, remains open: Whose action is it? If we say that *p*'s

personal preferences make up her self, then a fundamental change in her personal preferences amounts to a fundamental change of her self. Consequently, we cannot attribute the actions that depend upon the manipulated features to her previous self. In other words, we cannot say anymore that these actions are actions of her previous self that despised stealing, although they might count as the free actions of the person that resulted from the neurosurgeon's intervention. I think that this does justice to our intuitions. It would seem, then, that the present account can deal with this thought experiment in a satisfactory way.

III. Intuitions – The Principle of Alternative Possibilities and the Consequence-Argument

Everything that has been said so far is based on the minimal criteria that we started with. I think that these minimal criteria are almost universally accepted as *necessary* conditions, but there are many philosophers who think that they are not *sufficient*. Genuine freedom, so a libertarian might argue, requires more than the ability to act in a self-determined manner.

In the following section, I will scrutinize the demand that stronger criteria are necessary in order to capture what we really mean if we talk about freedom in a strict sense, rather than mere self-determination as it was characterized above. The main question will be whether the present account can do justice to the most common intuitions concerning freedom, particularly to those intuitions that seem to support the demand for stronger criteria. But even if stronger criteria are not *necessary,* we might want to know whether they are *possible*, that is, whether there are stronger and maybe more convincing accounts available that comply with the minimal criteria above. I will discuss this point in section IV.

One of the most widely shared intuitions concerning freedom is the so called *"Principle of Alternative Possibilities."* In addition to autonomy and authorship, freedom seems to require that *p*, even if she did *x* rather than *y* under conditions *c*, could have done *y* rather than *x*. The underlying intuition is quite strong: If *p* was *not* able to do *y* rather than *x*, how can we say that she was free when she actually did *x*?

It seems clear ever since G.E. Moore (1912) that the crucial question is what we mean if we say that someone "can" do otherwise in a given situation. So what would be an adequate interpretation of "*p* could have done *y* rather than *x* in situation *c*, although she did *x* rather than *y*"?

According to the most widely accepted interpretation, "being able to do *y* rather than *x* in situation *c*" requires that *y* could happen rather than *x* under identical conditions. The idea is that any situation in which it is *determined* that *p refrains* from doing *y* is a situation in which *p* is *unable* to do *y*. Consequently, the statement "*p* could do *y* in situation *c*" would be true in a deterministic world only if *p* actually did *y* in situation *c* and false if she didn't. Saying "*p* could have done otherwise than she actually did in situation *c*" might be true in a nondeterministic world only. It is obvious that, on this reading, the principle is incompatible with the above self-determination account.

Probably the most straightforward theory of freedom that has been developed along these lines is Chisholm's "agent-causation" account. According to this view, saying that *p* could have done otherwise when she did *x* rather than *y*

under conditions *c* implies that *p* could have done *y* rather than *x* under exactly the same conditions, that is, *no matter what her dispositions, beliefs, and desires may be.* Even if *p*'s desires, beliefs, and dispositions are such that *x* is the only rational alternative for her in conditions *c*, she should be able to do *y* instead – otherwise her action is not free. Thus, self-determination in the sense that was outlined above would be insufficient because it implies a dependency between the agent's preferences and her actions. Likewise, the *Principle of Alternative Possibilities* as it is understood by the proponents of agent-causation seems to rule out determinism: If it is determined that *p* does *x* rather than *y* in situation *c* then she seems unable to do otherwise under these very conditions.

According to the agent causation account it is an empirical question whether or not free actions exist. If there are cases of undetermined agent causation then there are cases of free actions. However, several philosophers have argued that freedom is impossible, no matter what might be true about our world, because the relevant criteria are incompatible. The crucial point is that the *Principle of Alternative Possibilities*, as understood above, seems to rule out authorship, and vice versa. If the former requires that *p* could have done *x* rather than *y,* or *y* rather than *x* likewise without a change in her personal preferences, then it is difficult to see how the principle of authorship can be met. Remember that, according to the principle, *p* can count as the author of action *x* rather than *y* only if *p*'s preferences give us an explanation of why she did so. But if we can give a true explanation of why it was *p* who did *x* rather than *y* in conditions *c*, then we cannot give another true explanation of why it was *p* who did *y* rather than *x* under exactly the same conditions. Only a bad explanans is compatible with two contradictory explananda. It follows that if doing *x* rather than *y* under conditions *c* counts as *p*'s action, then doing *y* rather than *x* cannot count as *p*'s action, too, under these very conditions. Note that this is not to deny that *y* may happen under these conditions. Of course, *p* may be able to behave differently under conditions *c* if she lives in an indeterministic world, but then at least one of the alternatives will not count as *her* action. If it was *p*'s *action* to do *x* rather than *y*, then doing *y* rather than *x* under these very conditions cannot count as her action – not for empirical but for conceptual reasons.

It would follow that, whatever *p* may do, it will not qualify as her free action. If her action depends upon her personal preferences then the action complies with the principle of authorship but violates the *Principle of Alternative Possibilities*. If it does not depend upon her beliefs, desires, and dispositions, then it might comply with the *Principle of Alternative Possibilities* but the principle of authorship would be violated.

It seems that the agent-causation account and the underlying interpretation of the *Principle of Alternative Possibilities* lead us into a severe dilemma. The dilemma is situated not on the empirical but rather on the conceptual level. Free actions seem impossible in principle, not because a certain requirement, say the absence of determinism, is not met in our world, but because two basic criteria for freedom, the *Principle of Alternative Possibilities* and the principle of authorship, are incompatible. It would appear, then, that nothing can meet both criteria at the same time, even in theory, and no single action has ever been and will ever be free. What is more, we would have to admit that our commonsense concept of freedom is inconsistent. Freedom turns out to be a benevolent illusion of human beings who, in reality, just are the marionettes of a relentless fate – or so it seems.

This position was originally developed by Thomas Nagel (1986) and Galen Strawson (1998, 1989). In a recent paper, even one of the proponents of incompatibilism, Peter Van Inwagen, tends to subscribe to this view:

»Free will seems to be incompatible with both determinism and indeterminism. Free will seems, therefore, to be impossible. But free will also seems to exist. The impossible therefore seems to exist. A solution to the problem of free will would be a way to resolve this apparent contradiction« (2002, 169).

Frankfurt's Objection

It may certainly be true that one of our commonsense concepts is incoherent, even if it plays such a crucial role as the concept of freedom does. Still, we should prefer analyses of such concepts that do not lead us into an incoherence – provided that such analyses are available. We should do so in particular because an incoherent concept of freedom would be completely useless. In this case, we could make no sensible statements whatsoever concerning the existence of free actions, since it would be completely unclear what we are looking for, respectively, what it is that does not exist in our view. Consequently, Van Inwagen's claim that "free will seems … to be impossible" would be as misleading as the above statement that "no single action has ever been and will ever be free." If the concept of free will is incoherent, then such statements are vacuous because nobody would be able to say what he is looking for if he looks for freedom. Conversely, we can say that freedom of will does not exist in our world only if we have a coherent idea of what "freedom of will" means, that is, if we can give a sketch of what would qualify as a free action.

Seeing all this, one might feel a temptation to give up the *Principle of Alternative Possibilities* altogether in order to save the concept of freedom. This move might appear even more attractive because would remove one of the most serious obstacles to a reconciliation of freedom and determinism. If freedom does not require the ability to do otherwise, then even a determined action might be free. In fact, Harry Frankfurt has tried to demonstrate that *p* can act freely in doing *x* rather than *y* in conditions *c* even if he could *not* have done *y.*

Imagine that, unbeknownst to *p*, a so called "counterfactual intervener" has implemented a mechanism in *p*'s brain which would prevent *p* from doing *y,* given the faintest hint that he might choose to do so. Still, as long as *p* actually does *x*, the mechanism remains completely passive. Now, assume that *p*'s doing *x* rather than *y* in conditions *c* would qualify as a free action according to your favourite account of freedom, as long as the mechanism is not able to interfere. Merely adding the mechanism's *ability* to interfere, should *p* consider to do otherwise, doesn't seem to change anything as long as there is no *actual* interference. However, even under these conditions *p* can't do otherwise because the mechanism *would* interfere before *p could* decide to do so.

It would appear, then, that *p* acts freely if he does *x* rather than *y* in conditions *c*, although he is not able to do *y* rather than *x* under these conditions. The consequences should be clear: If the *Principle of Alternative Possibilities* is not a necessary requirement for freedom, then we could not only evade the above dilemma but could also refute one of the most serious objections against a compatibilist account of free will.

Frankfurt's examples are very suggestive, but is it really true that he has presented a convincing objection against the *Principle of Alternative Possibili-*

ties? I do not think so, although I doubt that the standard objection against Frankfurt, the "flicker of freedom" strategy, goes through. I will not discuss this objection here because I think that we can reply to Frankfurt in a much more straightforward way. The *Principle of Alternative Possibilities* requires that, if *p* did *x* rather than *y* under conditions *c*, he could have done *y* rather than *x under these conditions*. It is not really obvious how we have to interpret the identity requirement concerning the background conditions *c*. But no matter how this requirement is understood, it should be beyond dispute that the counterfactual intervener that might prevent *p* from doing *y* is part of the background conditions. Thus, if the intervener becomes active, then the background conditions change from *c* to, say, *c'*. So, we have two different sets of background conditions: Conditions *c* if the mechanism remains passive and conditions *c'* if the mechanism intervenes. It seems clear that *p* cannot do *y* rather than *x* under conditions *c*', but since he is forced by the mechanism, we would not say that his action is free. We have no freedom and the *Principle of Alternative Possibilities* is violated. But what about conditions *c*? If the mechanism remains passive, then *p* is free *and* able to do otherwise because nothing will prevent him from doing so unless the background conditions change from *c* to *c'*. It follows that Frankfurt's objection can be dismissed because the alleged inability to do otherwise requires a change in the background conditions and thus ignores one of the most important requirements of the *Principle of Alternative Possibilities*.

But why is Frankfurt's objection so suggestive? I think the reason is that the change in the background conditions is concealed. Since the conditions for the mechanism's interference are determined before *p* will start with his process of decision-making, it may seem that nothing really changes, no matter whether or not the mechanism interferes. But even a determined change is a change, and a comparison makes it obvious that *c* differs from *c*'. What remains constant is the *rule* that governs the mechanism's activity in either situation, but this does not affect the difference between *c* and *c'*. Only if you ignore this difference, you may be led to believe that *p* can be free even if she can't do otherwise. But if you recognize this difference, you have to reject Frankfurt's example and the *Principle of Alternative Possibilities* remains in force.

That seems to be quite bad news, though. If the principle remains valid, then the incoherence that was noted above persists, too. But even if we could evade this conceptual problem, we would be left with the incompatibility of freedom and determinism and its empirical consequences.

Another Look at the Principle of Alternative Possibilities

On reflection however, doubts arise whether the above reading of the *Principle of Alternative Possibilities* is adequate. I take it that any acceptable interpretation has to treat the principle in such a way that it remains relevant for the question whether or not an action is free. Consequently, the interpretation should make sure that it cannot be said that *y* could happen rather than *x* in conditions *c* although it was not up to *p* to do *y*. Given that the actual outcome was *x* rather than *y* in conditions *c*, there are at least three interpretations of the *Principle of Alternative Possibilities*. According to these interpretations, saying that *p* is able to do *y* rather than *x* in conditions *c* could mean

(a) that the outcome *could* have been *y* under otherwise unchanged conditions;
(b) that the outcome *could* have been *y because p*'s preferences could have changed in such a way that they could explain why *p* did *y* rather than *x*;
(c) that the outcome *could* have been *y if p*'s preferences were such that they could explain why *p* did *y*.

It should be obvious that option (a) underlies the positions we have discussed so far in this section. The problem with this option is that if it was *p*'s action to do *x* rather than *y*, then, even if *y* could have happened under these very conditions, this would have occurred completely independent from *p*'s personal preferences. If you consider that these preferences constitute *p* then you cannot say anymore that *p* could have been the author of the fact that *y* rather than *x* happened in conditions *c*. Thus, even if *y* could have happened, it would not have been up to *p* to do so. The only thing we could say in this case is that something else might have happened instead. But this "something", whatever it might have been, was not an action that can be ascribed to *p*. And that is why it has no relevance whatsoever for your assessment of the original action.

Since interpretation (a) treats the *Principle of Alternative Possibilities* in such a way that it looses its relevance for the question whether or not an action is free, it does not comply with the criterion mentioned above, thus it seems justified to dismiss it. So what about interpretation (b)? It seems that, on this interpretation, *p* would have the ability to do *y* rather than *x* in situation *c*. At the same time, the interpretation leaves room for the required connection between the event and the agent's preferences, thus what happens would count as *p*'s action. All this is possible because of the previous change in *p*'s preferences, say in some situation *c*' at a certain time before *c*. But if what happens in situation *c* depends on a previous change in *p*'s preferences then it becomes questionable whether this interpretation has any advantage. Even if you agree that the ability to do otherwise may be contingent upon a previous decision to do so in situation *c'*, it seems clear that this only moves the problem to our assessment of situation *c*': The decisive question would be whether or not *p* was able to do otherwise in situation *c*'.

I conclude that interpretation (b) is of no help, either. So we are left with option (c). The rationale behind this interpretation is that "being able to do otherwise" cannot mean "anything else may just happen under identical conditions". What we need is *p*'s ability to perform another *action* than she actually performed, otherwise what happens in the counterfactual situation cannot be ascribed to *p* and would have no relevance for our assessment of the factual situation. If doing *x* rather than *y* is *p*'s action in situation *c*, then an occurrence of *y* rather than *x* in situation *c* will count as *p*'s action only if *p*'s preferences have changed – otherwise we would not be able to explain this event with reference to *p*. Consequently, we may not only permit a change in *p*'s personal preferences, rather, such a change is required in order to make sure that what happens in the counterfactual situation counts as *p*'s action. We would then have to interpret the demand for the ability to do otherwise as follows: Given that the author's preferences had been different, would she be able perform a different action?

Interpreting the *Principle of Alternative Possibilities* in this way implies a shift of focus from the outcome of the situation to the process of decision-making. What is at issue, then, is the relationship between the agent and the outcome. Asking whether different preferences could lead to different outcomes is ask-

ing whether the outcome depends upon the preferences rather than upon the external conditions. And if you consider that the agent is constituted by his personal preferences, then it turns out that the question is whether the outcome depends on the agent. Since this amounts to saying that the outcome was up to the agent, it would appear that the above criterion is met.

But does this interpretation really capture what we mean if we ask whether someone could have done otherwise? It clearly does. Saying that it was up to the agent whether *x* or *y* would happen is saying that the agent *could* do either *x* or *y*. Consequently, it would still be true to say that he *could have done y* even after he did *x*. It follows that the present interpretation is not a compromise that we have made in order to save the above theory of freedom. Rather, it does capture the entire meaning of the principle. In addition, it has been demonstrated that the seemingly stronger alternatives (a) and (b) have to be rejected because they don't provide an adequate interpretation within the context of the free will debate.

It should be noted that, superficial similarities notwithstanding, the present suggestion is not affected by the standard objections against Moore's conditional analysis. According to Moore (1912) saying that someone could have done otherwise is just saying that he would have done otherwise had he chosen to do so. The obvious reply is that it is unclear whether *p* had the ability to make the requested choice, say in cases of psychological addictions etc. Moore himself anticipated this reply and introduced a second order choice. It is, however, easy to see that this strategy is threatened with the same regress problem that puts Frankfurt's account into question. The present suggestion is not affected by this problem because personal preferences provide a very different way of dealing with those objections. This criterion blocks features like psychological addictions that may motivate the objection that the agent was not able to make a different choice: Remember that, in order to qualify as a personal preference, a feature must be a possible subject to a self-determined decision against it. This is, by the way, the reason why I think that the "liberal" option to determine personal preferences is better than its "identificationist" rival. In addition, it has been demonstrated on a more theoretical level that a rational choice that can be ascribed to an author requires certain abilities and preferences. It follows that these preferences, although they may determine what the agent's choice is, cannot be said to disable the agent from making a choice. Quite the contrary, they give her the ability to do so. I concede that there may be many cases where it is dubious whether a certain feature qualifies as a personal preference. Maybe you can doubt in *every single* case that a candidate feature really qualifies as a personal preference. However, you cannot coherently believe that agents have *no* features in general that qualify as personal preferences in the sense above, because otherwise you would loose the justification to talk about rational agents and their actions.

It would seem, then, that the present account can do justice to the *Principle of Alternative Possibilities*, that it can be defended against the standard objections against the conditional analysis and, finally, that it can block further demands, primarily because such demands, due to their incompatibility with the requirement of authorship, can be met by random events only.

The Consequence-Argument

In addition, the above line of reasoning can provide a basis for an answer to another challenge of compatibilism, namely the so-called "*Consequence-Ar-*

gument" that has been brought forward by Ginet, Wiggins, van Inwagen, and Lamb. In van Inwagen's words, the argument goes as follows:

»If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us." (Van Inwagen 1983, 16)

This argument, too, seems to show that determinism rules out freedom. The central premise of the argument is that necessity rules out choice. Van Inwagen takes this to be obviously true. If it were, then it should be obviously wrong or even self-contradictory to say "I know what his choice will be", at least if this claim is understood in a strict sense. However, I don't think that it is, and I have already said why: A rational agent's choice is not a random event. Choice must be connected to the agent who makes the choice, otherwise there is no reason to say that it was *his* choice. So if you know the situation and if you know the agent's preferences then you should know what his choice will be. But that does not mean, that no choice was made. This would be the case only if what happened did not depend upon the agent's preferences. He will have no choice if *external* factors determine what will happen, or in the case of random events. However, he does have a choice if what happens depends upon his personal preferences, that is, upon his whishes, desires, and beliefs.

I would like to stress that nothing will change if it turns out that the outcome of a certain situation is determined by events that happened prior the agent's birth. This is, after all, what you have to expect in a deterministic world. So demonstrating that freedom and determinism are compatible implies demonstrating that an action may be free even if it is determined by events that happened prior to the agent's birth.

But if events that happened prior to the agent's birth determine what he does, how can we respond to the objection that the agent did have no choice and, therefore, did not act freely? Provided that the action meets the minimal criteria above, the response is not difficult to find: The objection omits that, during the process described, a rational agent with personal preferences came into being, and that it were the agent's personal preferences that determined what happened. Since determination by the agent's preferences is everything we need for a self-determined choice, it follows that, contrary to what Van Inwagen's premise assumes, the conceptual criteria for a self-determined choice are met. In addition, extending the demand for self-determination to the time before the agent's becoming a rational agent would lead us immediately into a conceptual incoherence because self-determination prior to the self's coming into existence is conceptually impossible. The consequences have been described above: The concept of freedom would become useless and we could not make any judgment concerning the freedom of any action.

Still, you might feel an intuitive resistance against the idea that the determination of personal preferences by events that happened prior to the agent's birth does not interfere with her freedom. Note, first, that unlike events that happen *after* the person in question became a rational agent, events that happen *before* this time cannot be said to exert force upon the agent or to interfere directly with the agent's personal preferences. This is true for the simple reason that the agent has yet to come into existence. That having been said, I agree that many events prior to an agent's birth have an impact on the personal features that emerge later and thus may prevent her from acting freely indirectly. It may be that, due to some misfortune in her early childhood, *p* did not develop

the "personal capabilities" that are required for a rational agent. Consequently, she cannot act in a self-determined manner and cannot be held responsible for what she is doing including, of course, the fact that she is no rational agent. However, the opposite may also be true: Events that happened during her development may have *enabled* her to act in a self-determined way. This would be the case if these events brought all those personal capabilities and preferences into existence that are mandatory for a self-determined action. Even if this whole process is determined, it is still true that it brought about all that is required for a self-determined action. It may well have been determined that *p* would develop a love for Italian operas. But *if* this love is a personal preference, then her decision to go to the opera rather than watching a movie may be self-determined. Note, in addition, that on the "liberal" account of personal preferences that I have endorsed, those preferences are possible subjects to self-determined decisions. Consequently, *p* would be able do make a decision in favor of or against her love for Italian operas. If you still think that determination at this point interferes with the freedom of ensuing actions, waiving determination should enhance freedom. It will turn out below in section IV that this is not the case.

Finally, let me stress that, the compatibility with determinism notwithstanding, the present account leaves considerable room for actions that *appear* to occur at random from a folk psychological point of view. It may indeed be an important part of our commonsense intuitions concerning freedom that we can decide to make a dramatic change in our life. The present proposal can account for this intuition. First, my decision to refrain from such a dramatic change will count as free only if I could do otherwise in the sense that was explained above, that is, if it is up to me to make the opposite decision. Second, although the present account denies that random decisions can count as free, it accepts decisions that look like a random decision from a commonsense point of view. My decision to turn my life upside down may qualify as self-determined even if it would appear completely unpredictable even to my closest friends. Still, the decision might be explained with reference to my personal preferences, say, because there was a hidden dynamics within these preferences that led to this decision.

IV. Freedom and Determinism

It would seem, then, that there is considerable evidence that the present account is strong enough to do justice to some of the most widely shared intuitions concerning freedom, namely the *Principle of Alternative Possibilities* and the *Consequence-Argument.* Still, you might suspect that the present self-determination account is too weak, because it is compatible with determination. Genuine freedom, so you might think, is incompatible with determination. I have already tried to show that some of the most important arguments that are brought forward in favor of the alleged incompatibility of freedom and determinism can be rejected. Nevertheless I think that it is useful do demonstrate in a more systematic fashion that getting rid of determination doesn't help: Eliminate determination wherever you want – you won't get "more" freedom.

In order to show this, let's assume a deterministic world with a chain of events beginning at some time *t'* before *p*'s birth that ultimately leads to a self-determined decision in the sense described above at time *t*. If you think that such

a decision or the related action isn't free because it is determined, then there should be at least *one* link in the chain whose interruption gives you freedom. In what follows, I would like to demonstrate that this is not the case.

First, eliminate determination at some time *t* immediately before *p* acquires those properties that become personal preferences and determine her decision to do *x* rather than *y* in situation *c,* and assume that *c* happens a considerable time after *t*' at *t*. It would follow that *p*'s decision is not determined anymore by events that happened before her birth. Since *p* has acquired the properties in question some time before she makes the decision, they meet both the stability requirement and the requirement that *p* has to be able to make a self determined decision concerning these properties. Consequently, the properties qualify as a personal preferences and the action counts as free, as far as the present account is concerned. Still, it is somewhat unclear to whom the action can be ascribed, particularly if there was a fundamental change in *p*'s preferences. I have discussed a similar problem above, regarding the nefarious neurosurgeon. Things look different, however, if you take an incompatibilist point of view. Since it is possible, in principle, to predict p's eventual action already at some time *t'* after *t'* but before *t*, that is, before *p* herself knows what she will do, the decision at time *t* remains determined. It is still the "consequence of the laws of nature and events in the past", the difference is just that the past is not so remote and the causal chain is a bit shorter. But if you think that the decision at time *t* wasn't free before the causal chain was interrupted then you have no reason to assume that it is free afterwards, since it is still determined. Of course, *p* could decide to get rid of the feature in question in the time between *t'* and *t*, but she could do so anyway, according to the present account: Remember that it is a requirement of the liberal account of personal preferences that they are subject to self-determined decisions. In any case, interrupting determination at this point doesn't give us "more" freedom.

It might seem that the problem is the long interval between *t'* and *t*. So, let's move a bit forward in time and assume that the random change in *p*'s preferences takes place *immediately before the decision.* Again, the decision is not determined by events that happened before *p*'s birth; in addition we can reject the first premise of the *Consequence-Argument* because the relevant action can not be regarded anymore as the "consequence of the laws of nature and events in the remote past." However, since there is no interval between the random change at *t* and the decision at *t'*, *p* would be left without any control over the preference in question, which, by the way, would also fail to meet the above stability requirement. Thus, the principle of authorship would be violated: While *p* was deeply convinced so far that steeling is reprehensible, this conviction may have vanished instantly when she saw the cashier. And because all this happened immediately before she decided to leave the shop without paying for the goods in her basket, *p* had no chance to make another self-determined decision concerning the changed preference. As far as *p* is concerned, all this does not differ from an external event like a manipulation by an nefarious neurosurgeon. It is therefore difficult to see why eliminating determination at this point should give *p* "more" freedom: Why should the chance that one of my central convictions might vanish at random enhance the freedom in those situations where I act according to this conviction?

But maybe we still got the wrong point in time. Since I have already said why eliminating determination of the decision by personal preferences does not help either, my third suggestion is an interruption *during* the process of

decision-making. Consequently, there should be at least one situation *during* this process where it is really open what will happen. One part of the process would be detached from the rest. This means that any result that might have been achieved during the first part of the process would loose its effect on the second part and the ensuing decision. Assume that, during the first part of the process, you have found good and almost decisive reasons to do *x* rather than *y*. Interrupting the process afterwards would make these considerations void as far as the outcome of the process is concerned. It seems clear that such an interruption would lead to a destruction of the whole process of decision making rather than giving us freedom. Of course, disrupting the process might waive the effects of force or compulsion, but force or compulsion are incompatible with freedom anyway.

Fourth, you might eliminate determination *after* the process of decision-making, but it should be obvious that this would be of no help either, since it would detach *p*'s doing *x* from her previous decision. So even if *p* has finally decided to do *y*, it might happen that *x* comes about. I assume that this isn't either what you expect if you ask for freedom.

Kane's Approach

But maybe this line of reasoning is all too simple. Robert Kane has made a very elaborate suggestion that seems to resolve the conflict between authorship and indeterminism. Kane accepts that freedom requires an intelligible connection between the action and the agent; still he thinks that indeterminism is mandatory for free actions in order to provide "ultimate responsibility". In his view, conflicts between moral reasons and selfish motives are paradigm cases of free will. The decision might then depend upon whether the agent believes that the moral reasons or the selfish motives are more relevant, and it is this act of belief that is not determined and, consequently, cannot be predicted. But since the eventual decision can tell us whether the moral or the selfish reasons prevailed, we can then explain why the person acted in the way she did. Thus we will be able to provide an intelligible connection and the requirement of authorship is met. Kane concedes that there is some circularity in his account since the explanandum (*p*'s doing *x* rather than *y*) is part of the explanans, because only the eventual action allows us to determine which kind of reasons prevailed. However, Kane thinks that the explanation is still informative.

On reflection however, several objections come to mind. It is at least unclear whether conflicts can count as paradigm cases for free decisions. It is difficult to accept that we don't act freely if we feel clear moral obligations. This would mean that someone who feels a temptation to kill someone but finally refrains from doing so is free while someone who feels only the moral obligation without being tempted to act against it, is not.

The important point, however, is that Kane has only *transferred* the problem to another level rather than solving it. Note that our desire to explain *p*'s doing *x* rather than *y* results from the assumption that such an explanation is required in order to decide whether or not *p* was the author of this action. This does not require the complete causal history of *p*'s action. What is necessary, however, in order to satisfy the requirement of authorship is an identification of those preferences that make it intelligible that *x* is done rather than *y* in situation *c*. If these preferences can be ascribed to *p* then it was *p*'s action, otherwise it wasn't. Now, Kane is very clear about this point. According to

him, *p*'s doing *x* rather than *y* depends upon an act of belief "that a rather than b" concerning the conflicting preferences, and this belief, in turn, can *not* be explained. We can, of course, conclude the content of this belief from the eventual decision, but this doesn't tell us *why p* believed "that a rather than b". And since *p*'s action depends upon this belief, we are still left without an explanation why *p* did *x* rather than *y*. Kane, after all, insists that the belief and, therefore, the decision is not determined. But for the very same reason, we will not be able to give the relevant explanation and, consequently, we will not be able to ascribe it to *p* that *x* rather than *y* was done. And this means that the requirement of authorship is *not* met. Note that this is not saying that *p* has nothing to do with the outcome of this situation. Of course he has – *x* as well as *y*, after all, are attractive options from his point of view. That's the reason why the conflict emerged. It may also be true that it was *p*'s decision to do either *x* or *y* rather than *w*. The question, however, was whether it was up to him to decide between *x* and *y*. And this is not the case because this depended upon a random event, namely the belief "that a rather than b". It would seem then, that Kane is not able to show that freedom requires indeterminism. Inserting indetermination doesn't give us a stronger account of freedom – it just destructs the indispensable connection between the agent and his action.

There are certainly a great many ways to enhance the present account, but to add indeterminism is not one of them. It would seem then, that the minimal self-determination account meets the two more demanding criteria for a "strong analysis" that I've introduced at the beginning: First, it does justice to our most important intuitions concerning freedom. And second, it can be defended against the demand for stronger requirements. The most popular allegedly stronger strategies, particularly the introduction of indeterminism and the concept of agent-causation, fail because they don't even meet the minimal criteria.

References

Chisholm, R. M. (1982) "Human Freedom and the Self". In: G. Watson (ed.), *Free Will.* Oxford.

Dennett, D. C. (1984) *Elbow Room. The Varietes of Free Will Worth Wanting,* Cambridge MA.

—— (2002) "I Could Not Have Done Otherwise – So What?" In: R. Kane (ed.), *Free Will,* Oxford.

Fischer, J. M. (1994) *The Metaphysics of Free Will,* Oxford: Blackwell.

Frankfurt, H. G. (1969) "Alternate Possibilities and Moral Responsibility", *Journal of Philosophy,* 66: 828–39.

—— (1971) "Freedom of the Will and the Concept of a Person", *Journal of Philosophy* 68: 81–96.

Haji, I. (2002) "Compatibilist Views of Freedom and Responsibility". In: R. Kane (ed.), *The Oxford Handbook of Free Will,* Oxford: Oxford UP.

Kane, R. (1989) "Two Kinds of Incompatibilism", *Philosophy and Phenomenological Research,* 50: 219–254.

Moore, G. E. (1912) *Ethics,* Oxford.

Nagel, Th. (1986) *The View from Nowhere,* New York: Oxford UP.

Pauen, M. (2000) "Painless Pain. Property-Dualism and the Causal Role of Phenomenal Consciousness", *American Philosophical Quarterly,* 37: 51–64.

Strawson, G. (1998) "Free Will", *Routledge Encyclopedia of Philosophy.*

—— (1989) "Consciousness, Free Will, and the Unimportance of Determinism", *Inquiry,* 32: 3–27.

Van Inwagen, P. (1983) *An Essay on Free Will,* Oxford: Oxford UP.

—— (2002) "Free Will Remains a Mystery". In: R. Kane (ed.), *The Oxford Handbook of Free Will,* Oxford: Oxford UP.

Wolf, S. (1980) "Asymmetrical Freedom", *Journal of Philosophy,* 77: 151–166.

Michael Pauen

Selbstbestimmung

Freier Wille, Verantwortung und Determinismus

Zusammenfassung

Eine Analyse unseres auf dem gesunden Menschenverstand beruhenden Freiheitskonzeptes ergibt zwei „minimale Kriterien": 1) Autonomie bedeutet einen Unterschied zwischen Freiheit und Zwang; 2) Urheberschaft bedeutet einen Unterschied zwischen Freiheit und Zufall. Die Auslegung von Freiheit als „Selbstbestimmung" kann für beide Kriterien in Anspruch genommen werden. „Selbstbestimmung" wird verstanden als Bestimmung anhand „persönlicher Vorlieben", die für die betreffende Person konstituierend sind. Freiheit und Determinismus sind also kompatibel. Die Schlüsselfrage ist nicht, ob unser Handeln überhaupt determiniert ist, sondern eher, ob dies durch persönliche Vorlieben geschieht. Diese Erklärung kann den meisten freiheitsbezogenen Intuitionen gerecht werden, einschließlich der Fähigkeit, anders [als gewohnt] zu handeln. Im Gegensatz dazu würde der Verzicht auf eine Determinierung eher das genannte Minimalkriterium verletzen, als „mehr" Freiheit zu ermöglichen. Der Verfasser kommt zum Schluss, dass Selbstbestimmung die Verwirklichung aller unserer Ansprüche ermöglicht, wenn wir Freiheit fordern.

Schlüsselbegriffe

Freier Wille, Selbstbestimmung, Verantwortlichkeit, Determinismus, alternative Möglichkeiten

Michael Pauen

L'autodétermination

Libre arbitre, Responsabilité et Déterminisme

Résumé

L'analyse de la conception commune de la liberté produit deux « critères minimaux » : 1) L'autonomie distingue la liberté de la contrainte ; 2) La responsabilité distingue la liberté du hasard. Interpréter la liberté comme « autodétermination » correspond aux deux critères. L'autodétermination se comprend comme une détermination par les « préférences personnelles », constitutives de la personne. La liberté et le déterminisme sont ainsi compatibles. La question essentielle n'est pas de savoir si une action est déterminée ou pas, mais plutôt de savoir si elle est déterminée par les préférences personnelles. Cette explication est juste à l'égard des intuitions les plus importantes concernant la liberté, y compris le pouvoir d'agir autrement. Abandonner la détermination, par contraste, violerait les critères minimaux au lieu de procurer « davantage » de liberté. Dans la conclusion, il est indiqué que l'autodétermination procure tout ce qu'on peut demander si on demande la liberté.

Mots-clés

libre arbitre, autodétermination, responsabilité, déterminisme, possibilités alternatives