This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Southwest Philosophy Review*, January 2003, 221-29.

FRANKFURT STYLE EXAMPLES

JAMES CAIN

OKLAHOMA STATE UNIVERSITY

Frankfurt style examples (FSEs) were originally formulated by Harry Frankfurt[1] as a thought experiment to be used in refuting the principle of alternate possibilities (PAP) which states that *a person is responsible for what she has done only if she could have done otherwise*. More recently FSEs have become a cornerstone of the case built by Fischer and Ravizza for the doctrine of semicompatibilism (the view that moral responsibility is compatible with determinism) and for an elaborate theory of moral responsibility.[2] Fischer and Ravizza present FSEs as a powerful tool that can be used to investigate the metaphysical nature of free will and moral responsibility, and can be used to justify a theory of moral responsibility in terms of an 'actual sequence' account of 'guidance control' as opposed to an 'alternative sequence' account of 'regulative control'. In the writings of Eleonore Stump, a libertarian, FSEs are also presented as a powerful and legitimate tool for investigating the roles played by the will, the intellect and power to act otherwise in accounts of moral responsibility.[3] If these writers are correct, FSEs can be used to gain insight into deep metaphysical issues tied to our nature as beings susceptible to moral assessment.

FSEs have been criticized by defenders of PAP and by those who reject semicompatibilism.[4] My efforts here will add to the criticisms of FSEs, though in a different way than has been customary. I will make no appeal to the *freedom* of the will and no appeal to claims concerning conditions for moral responsibility, nor will I attempt a defense of PAP. If

one is to establish by means of a FSE that people can exercise their will without there being an available alternative sequence in which they do not so exercise their will then that FSE must express a metaphysical possibility.[5] I argue that before we can accept the metaphysical possibility of FSEs we need a fuller knowledge of the nature of the will (that is, a fuller knowledge of the nature of deliberation, choice, decision-making, voluntary action, etc.).

Typically FSEs involve two people, an agent and a controller. The agent has an opportunity to perform a morally significant action. The controller wants the agent to perform that action and, to insure that the agent does so, puts into place a special mechanism, a 'Frankfurt device'. Should there develop a sufficiently strong tendency on the agent's part not to perform the action (i.e., what takes place in the agent deviates from acceptable parameters), the mechanism will bring it about that the agent performs the action after all. As the example develops however the agent never exhibits such a tendency and, without any influence from the mechanism, does the action in a fashion that has all the usual hallmarks of being voluntary. Because the mechanism is never utilized it is argued that the agent is morally responsible for the action. On the other hand, the presence of the mechanism guaranteed that the agent could not have done otherwise. Thus, the argument concludes, PAP is false.[6]

FSEs assume that for any suitable possible history H that extends the past beyond the point at which the Frankfurt device has been readied for action[7], if H first deviates from the acceptable parameters at time t, then (1) there are possible histories which up to time t are identical to H but then continue in such a way that the agent does the action, and (2) the Frankfurt device is capable of actualizing such a history. What the device must do may depend on the point at which the deviation takes place. It might have to cause a decision to be made, or it might have to bring it about that the agent adheres to a previous decision, or it might bring

about the action in some other fashion. To simplify discussion I will only be concerned with the special case in which the deviation occurs before a decision is made to do the action and the device must cause such a decision to be made.

Typically in FSEs the device operates (if it is ever triggered) by manipulating the micro-level processes taking place within the brain, e.g., it controls the patterns of neural firings. Initially I will focus on FSEs that operate on the brain by controlling the movements of atoms within the brain. Let us begin then with FSEs that meet the following specifications. There is a device in place that is to insure that the agent will decide to do X. If the history of the agent's brain falls within certain parameters the device is not triggered. If at a given time the agent deviates from the parameters, then the device becomes triggered and brings it about that the atoms in the agent's brain move exactly as they move in some history in which the agent decides to do X. Then (to fill in this version of the FSE) since the atoms, under the influence of the mechanism, move exactly as they do in a possible history in which the agent decides to do X, *it follows that* in this case too the agent decides to do X. Thus the device guarantees that the agent could not have done otherwise than to decide to do X.

This kind of FSE assumes that, under appropriate circumstances, if the atoms in the agent's brain could have moved a certain way in the course of the agent's deciding to do X, then if the atoms in the agent's brain do move that way then a decision has been made to do X. This may seem to be a plausible assumption, but it is problematic. An analogy may help to show why. Suppose a magnet M is about to be dropped into a jar of iron filings, however there is a possibility that M will lose its magnetic power before it gets into the jar. Imagine that there is a 'controller' who wants to insure that M will act as a magnet. He constructs a device so that should M begin to lose its magnetic power the device is triggered; once triggered the device

causes the movement of each iron filing and of M itself to be exactly like it would be under a scenario in which M had magnetic power. Should we conclude that if the device kicks in then M exercises magnetic power? The answer is that the example does not give us enough information. It all depends on how the controller's device works. If it works by *magnetizing* M then the answer is in the affirmative. If the device moves M and the iron filing by some other means than causing M to be magnetic the answer is negative.

Now reconsider the last FSE. We sometimes refer to the act of deciding as an exercise of the will. Some even speak of a faculty of the will and of will power. Suppose for the moment that this is not just a fashion of speech and that there really is a kind of power of the will and that coming to a decision is an exercise of this power. Then before we can accept the FSE under consideration we must confront a problem analogous to that raised in the example with the magnet. If we are to get an example in which the agent can be caused by a device to decide to perform X then it must be the case, not merely that the agent's atoms move as they might move if he decided to perform X, but his activity must arise (in part at least) through the power of the will (just as the movement of the iron filings must be due in part to the magnetic attraction of M if M is to magnetically move the filings). More generally, *regardless of whether one accepts the existence of a faculty of the will*, FSEs of the type we are investigating depend on the assumption that under appropriate circumstances a duplication of the atomic movements that occur during the process of deciding will automatically yield a process of decision-making. Since the process of deciding will presumably involve causal features, we must ask whether a duplication of the movement of atoms that takes place in deciding will automatically also duplicate enough of those causal features so that what results is an act of deciding.

I propose then to look at several possible accounts of decision-making with this problem in mind. We might categorize these accounts as physicalistic and dualistic. Within physicalistic accounts we may distinguish accounts that are 'bottom-up' from those that are not. By a 'bottom-up' account of the world I have in mind accounts that hold that micro-level events and processes (of the sort physics would strive to study) are in some fundamental sense responsible for everything that happens, and that all causation ultimately arises at the micro-level.

Bottom-up physicalistic accounts seem to be the most promising for successfully constructing FSEs: if everything that happens ultimately happens because of micro-level processes, and the device employed in the FSE manipulates these micro-level happenings then it might seem there would be no limit to the things these Frankfurt devices could bring about. But even here there may be difficulties for FSEs depending on the nature of decision-making. An illustration of the difficulty may be had by looking at two types of events. It is not part of what constitutes something as an event of the first type that it involves causal processes of a specific sort. Examples of events of this type include movement, change of shape, and change of size. Consider orbiting: a space ship could orbit the earth by moving in a path around the earth as a result of the earth's gravitational pull, or it could orbit an asteroid by using rocket engines to propel itself through a path around the asteroid. The causal mechanism by which the orbiting takes place does not in part constitute what it is to orbit. If the decision-making process is an event of this sort then FSEs seem in theory quite plausible.

The second type of event is one whose causal features are integral to its very nature. If I rub a diamond across the surface of a piece of glass the diamond will scratch the glass. At the micro-level the atoms near the surface of the glass will be rearranged in a certain configuration. Suppose that instead I rub my fingernail across the glass and the atoms near the surface become

rearranged in exactly the same configuration as they might have had had the diamond scratched it, should we say that scratching of the glass took place? The answer is that it depends on what caused the atoms to reconfigure. If it was caused by the pressure and hardness of my fingernail then perhaps the answer is in the affirmative. If however the reconfiguration of atoms at the surface had another cause that did not involve something rubbing against it, then scratching of the glass did not take place.

Plausibly the scratching of glass is the sort of event whose causation arises fundamentally at the micro-level. Imagine an FSE-like example here: I want to see if my fingernail can scratch a certain surface. A controller builds a device so that if I start to move my fingernail across the surface and there is an indication that my fingernail may not scratch it then the device will cause the atoms on the surface of the glass to move just as they might if a process of scratching was reconfiguring them. If the way the device works is by directly manipulating the atoms on the surface and it does not in part work via the usual mechanisms involved in scratching (e.g., the device does not harden my fingernail), then the device is not one that will bring it about that my fingernail scratches the surface.

If the process of deciding is like *orbiting* or *changing shape* and a bottom-up account of decision-making is correct then FSEs are in theory plausible. If on the other hand decision-making is more like *scratching* or *cutting* then the plausibility of FSEs depends on its being made plausible that the device can bring about the same exercise of causal powers that (partially) constitute decision-making and can do so in such a way as to guarantee that the decision in question is made. I do not think that those who fashion FSEs have done much to make this plausible. Take an example. Suppose that the causal processes that partially constitute decision-making in humans include processes whose causal outcomes are indeterminate (perhaps these

take place as quantum-level processes[8]). Then it is hard to see ahead of time that Frankfurt devices could be fashioned so that they control the movements of elementary particles in narrowly prescribed ways *and* they do it via these indeterminate causal processes.

It is important to stress that nothing in my argument depends on the nature of the device. The problem would remain even if the role of the device were played by God or a demon. To see this point, return to the example of the magnet. Suppose that as magnet M is dropped into the jar it loses its magnetic power. Imagine an evil deceiver decides not to re-magnetize M but instead directly moves all the iron filings in the jar so that they move exactly as they would have had M been re-magnetized. This would not be a case in which M magnetically attracts the filings. So too, if decision-making essentially involves certain causal processes, then even if an evil demon moves the atoms in the brain as they might move if a given decision were made, no decision will be made if the essential causal processes are absent. And if decision-making essentially involves the exercise of causal powers whose outcomes are by nature indeterminate, then if the demon tries to bring about the decision through having those powers activated the resulting process will be indeterminate.

Next consider accounts of the physical world that are not totally 'bottom up' but include 'top-down' features. Perhaps a good example is an account of physical objects as composites of matter and form in the tradition of Aristotle. An object might be said to have certain powers in virtue of having a substantial form, where these powers of the object fundamentally arise at the macro-level. If such an account turned out to hold for the will, there seems to be little reason to think in advance that FSEs can be successfully constructed. If the power exercised in deciding is a power that fundamentally resides in a formal macro-level feature of a person, then merely to be told that the device brings about the same sequence of atomic movements that are brought about

by the exercise of this power does not give us any reason to think that the same power is at work. And if decision-making turns out to essentially involve the exercise of macro-level causal powers whose outcomes are by nature indeterminate, then if the device is to bring about the decision through having those powers activated the resulting process will have indeterminate features. It is not clear in advance that such processes can be controlled to yield predetermined decisions.[9]

Finally there are dualistic accounts of decision-making. I will simply say a little about accounts in the tradition of Cartesian dualism. Typically in such accounts it is held that decision-making involves the exercise of a mental power (the will) by the mind, a nonphysical substance. It might be thought that variant FSEs can be constructed to handle dualism: instead of there being a device controlling the happenings in the brain, the controlling device could be designed to operate on the happenings in the mind.[10] This suggestion may seem to pick up plausibility from a way some people tend to look at the mind, namely as a mere receptacle for qualia, the mental life consisting of a stream of qualia. Obviously mechanisms can sometimes cause us to have qualia of certain sorts, or even streams of qualia. The television is such a mechanism. So one might think that all we need is a mechanism to produce the right sort of qualia in the mind and we will have produced an exercise of the will. But this picture of things is not very plausible after all. For an FSE to work it will have to duplicate the causal features of the mind that come into play in decision-making, and it is not at all clear that these can be duplicated merely by replicating qualia. And it is not clear without argument that whatever the manipulation device did it could be said to cause the occurrence of a *decision* in the mind of the agent.[11] To return to an earlier point, if decision-making essentially involves causal processes then the device would have to bring these causal processes about. And if the causal processes were intrinsically

indeterministic then the device would have to bring about indetrministic processes. We have little reason to think *a priori* that if decision-making involved indeterministic processes then these processes could be controlled by a Frankfurt device in such a way as to bring about a foreordained decision.

By way of summary and conclusion, we have noted that there are several accounts that might be given of the nature of decision-making: bottom-up physicalistic accounts, physicalistic accounts that involve top-down features, and dualistic accounts in which an agent exercises a (nonphysical) mental power in deciding. In all three cases the same difficulty arises: if decision-making is essentially a causal process then in order for a Frankfurt device to bring about such a process it must do so via the causal powers that are part of decision-making. But there is seemingly no guarantee, given our current knowledge of the nature of decision-making, that such processes could even in theory be duplicated by Frankfurt devices that are capable of predetermining the outcome of the processes within narrow bounds.[12]

NOTES

[1] Frankfurt (1969).

[2] Fischer (1994) and Fischer and Ravizza (1998).

[3] Stump (1990, 1996 and 1999). Stump attacks PAP on several grounds, some of which do not appeal to FSEs (and are, I believe, weighty considerations); see, e.g., Stump (1990), pp. 257 and 262 ff. However, my concern will not be with the correctness of PAP, but with the legitimacy of FSEs as a tool for exploring the metaphysics of free will and moral responsibility.

[4] Objections have been raised concerning various features of FSEs. One set of objections involves the timing of Frankfurt devices; see for example Widerker (1995a and 1995b) and Kane

(1998), esp. 142-42, which are anticipated in Blumenfeld (1971), 340. Roughly put, these objections hold that if the device allows the agent enough time to come to the decision desired by the controller then either the agent will have time to make a contrary decision or act in some other way that is morally significant, and thus the FSE will not be a counterexample to PAP. For responses to these kinds of objections see, for example, Fisher (1994) 134-47, and Stump (1999). Another approach questions the philosophical significance of FSEs. For example, van Inwagen (1983, Chapter 5, and 1999) allows that FSEs might be counterexamples to PAP but they are not counterexamples to some important modified versions of PAP; also see Blumenfeld (1971). My approach raises the question whether we should ever call whatever it is that the device brings about a *decision*.

[5] Fischer and Ravizza (1998) and Stump (1999) hold that FSEs can be used to establish that people can exercise their will without there being an available alternative sequence in which the will is not so exercised. Another interesting case in which FSEs must be taken to express a metaphysical possibility is found in Eshleman (1997) which presents an argument that presupposes that God has the power to bring about FSEs. Since God's power cannot extend beyond what is metaphysically possible, his argument depends upon the metaphysical possibility of FSEs. There *may* be cases in which the mere *conceptual* possibility of FSEs is of philosophical importance: if one accepts that Frankfurt's (1969) original examples are conceptually possible then one will not be able to accept PAP as an *analytic* truth (though perhaps one could accept PAP as true on grounds that do not merely rely on a conceptual understanding of PAP). The extent to which FSEs are conceptually possible will be left an open question. I thank Alastair Norcross for a helpful criticism of my original formulation of the point under discussion in the text.

[6] I will follow the literature in speaking of a 'mechanism' or a 'device' when developing FSEs. But, as is generally recognized, the role of the 'device' in FSEs may be played by the presence of "any condition under which it will be maintained that [the agent] cannot do otherwise" (Frankfurt [1969], p 836). Frankfurt (1969, pp. 835-36) mentions hypnosis, threats, potions, brain implants, and "natural forces involving no will or design at all" as candidates that could play this role (depending on which account of *inability to do otherwise* one wishes to adopt). Stump (1999, p. 306) and Eshleman (1997) allow for the possibility of God playing the role of the Frankfurt device. Nothing I say will depend on there being a mechanical device as opposed to one of these other sorts of possibilities.

[7] I add the qualification 'suitable' to 'possible history' because one who devises FSEs may not want to consider all possible histories. For example, often FSEs involve a mad scientist who implants a Frankfurt device. One possible history that occurs after the device has been implanted might involve the scientist changing her mind and deactivating the device. This would not count as a 'suitable possible history' for the purposes of the FSE. I need not spell out what counts as suitable since my objection will apply however the notion is spelled out.

[8] For a highly speculative example of such an account see Robert Kane (1998).

[9] Another, more recent, top-down account of agency has been propounded by Timothy O'Connor (2000). O'Connor suggests that our way of conceiving of agency involves the positing of an active power exercised in agent causation; this active power is a macro-level emergent property. According to O'Connor, no event can determine how an agent is going to exercise this power (if it exists). O'Connor does not try to prove the existence of such an emergent property (he describes himself as doing descriptive metaphysics). Nonetheless his work raises an important issue. Unless those who propose FSEs can rule out the hypothesis that

there is such an emergent macro-level property of persons as having the power to exercise agent causation, then they will have to argue that the way this causal power is exercised by the agent must be susceptible to manipulation by a Frankfurt device. Though I cannot go into O'Connor's views here, I merely note that this would be a difficult task.

[10] Eleonore Stump (1999, p. 306, fn. 19), though not herself a Cartesian dualist, suggests that even Cartesian dualists are not immune to FSEs: God could play the role of the controller and directly manipulate what goes on in the mind of the agent.

[11] Again, this holds even if we replace the device by a demon or God.

[12] I would like to thank Alastair Norcross for his thoughtful comments on the paper.

BIBLIOGRAPHY

Blumenfeld, David. 1971. The Principle of Alternate Possibilities. *Journal of Philosophy* 68: 339-45.

Eshleman, Andrew. 1997. Alternate possibilities and the free will defence. *Religious Studies* 33: 267-86.

Frankfurt, Harry G. 1969. Alternate possibilities and moral responsibility. *Journal of Philosophy* 66: 829-39.

Fisher, John Martin. 1994. *The Metaphysics of Free Will: An Essay on Control*. Oxford: Blackwell Publishers.

Fischer, John Martin and Ravizza, S.J., Mark. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Kane, Robert. 1998. *The Significance of Free Will.* Oxford: Oxford University Press.

O'Connor, Timothy. 2000. *Persons and Causes: The Metaphysics of Free Will*. Oxford: Oxford University Press.

Stump, Eleonore. 1990. Intellect, Will, and the Principle of Alternate Possibilities. In Michael D. Beaty (ed.), *Christian Theism and the Problems of Philosophy*. Notre Dame: University of Notre Dame Press. 254-285.

_____. 1996. Libertarian Freedom and the Principle of Alternative Possibilities. In Jeff Jordan and Daniel Howard-Snyder (eds.), *Faith, Freedom, and Rationality*. Lanham, Maryland: Rowman and Littlefield Publishers, Inc. 73-88.

_____. 1999. Alternative Possibilities and Moral Responsibility: The Flicker of Freedom. *The Journal of Ethics* 3: 299-324.

van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

_____. 1999. Moral Responsibility, Determinism, and the Ability to Do Otherwise. *The Journal of Ethics* 3: 341-350.

Widerker, David. 1995a. Libertarian Freedom and the Avoidability of Decisions. *Faith and Philosophy* 12. 113-18.

_____. 1995b. Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities. *The Philosophical Review*, 104. 247-261.