

An Agent-Causal Theory of Free Will and Moral Responsibility: Tim O'Connor

Kenton Machina

[This is from a draft textbook I'm writing, and is not in final form. It should not be cited publicly as a representation of my own views. Copyright reserved.]

O'Connor's explanation of how we exercise free will

One of the best accounts of agent-causal libertarian free will that I am aware of has been developed by Timothy O'Connor, currently a faculty member in Philosophy at Indiana University.¹ I will summarize O'Connor's theory to promote thinking about what an agent-causal version of free will might be like.²

According to O'Connor, as a person begins the process of authoring a decision, he first "represent[s] possible courses of action to himself" while being aware perhaps of some "desires and beliefs concerning those alternatives." For example, you might think about going to class or skipping class in order to study for an exam. You have the desire to go to class in order to find out more about the current topic of discussion, and the desire to do well in that class. But you also have the desire to do well on your exam, and you have the belief that you will not do well unless you study for it during the next hour. Additionally, you have other beliefs about the value of attending class, the relative importance of the two courses in your education, and so on. All this gives you reasons for going to class and reasons for skipping in order to study. So, you "represent to yourself" these alternative courses of action: going to class versus skipping class and studying, and you are perhaps aware of at least some of your relevant desires and beliefs about these alternatives.

Next, according to the theory, the person "brings about" an "executive state of intention to act in various ways." An "executive state of intention to act" is an intention that will bring about the intended action so long as it is not interfered with or abandoned. The act that the person then performs "flows from" that intention, which appears to me to mean that the act is caused by the existence of the intention.³ Applying this to the case of your dilemma over whether to go to class or study, apparently what happens, according to O'Connor, is that after "representing to yourself" the two alternative courses of action, going to class and studying, and taking into account whatever relevant beliefs and desires come to mind, you form an effective intention to do one of the two things, and then your behavior results from that intention's having been formed.

Where exactly does free will enter into this story? Your free will gets exercised when you form the effective "executive" intention. The formation of that intention is *caused*, but it is caused by *you*, and not by anything else—not by your desires, not by your personal background, not by your

¹His book, *Persons and Causes*, Oxford U. Press, 2000, is an excellent resource. It is the source for my description of his views.

²For a different theory of contra-causal free will see the work of Robert Kane, found in numerous places.

³All O'Connor quotations thus far from page 72.

hormones, not even by your own reasoning. ***Nothing causes you to form that particular executive intention at that particular time.*** You just form it, by exercising your decision-making powers. So, it is truly a “contra-causal” type of freedom that O’Connor is describing.

How are you able to form this intention? You have the “power” of free will! That “power” is what generates the intention. Once the intention has been generated, the rest is regular causation—the formation of the intention causes you to behave. If you form the effective, “executive” intention to go to class, then that is what you will attempt to do so long as you do not change your mind and form a new intention, or abandon the old intention. So, once the “executive intention” gets formed, free will has done its job, and the rest of the process that makes your muscles move, and gets you to class does not involve any additional free will activity (unless you change your mind). Although O’Connor does not quite say this, I assume he believes that the “executive intention” is represented in the human brain as some sort of brain state, some sort of pattern of neuron firings or something like that. The results that follow, to get you to class—the muscle expansion and contractions, the focusing of the eyes, and so on, would be caused by your brain without any additional input from free will (unless you changed your mind). (If the formation of the executive intention doesn’t produce any physical changes in the brain, then I don’t see how it can bring about any behavior.)

O’Connor agrees that his theory of free will does not explain why a particular executive intention is formed by a person at a particular time under particular circumstances. There simply *is no* such explanation for a free choice. In fact, it is exactly that *lack* of explanation that makes appeal to free will different from normal kinds of explanation. It’s what makes free will contra-causally free. If we could explain in the usual way how various factors caused you to chose to go to class rather than studying, your choice would not count as an exercise of your (contra-causal) free will. Instead, there would have been some normal causes that brought about your choice to go to class. It then would have been those causes that brought about your decision, rather than *you yourself* bringing about the decision. To ask for an explanation of why you chose to go to class, and why that particular choice was made at the time it was made is to ask for an explanation that does not attribute the choice to your exercise of your own power of free will.

The critic of O’Connor’s account (as O’Connor himself brings out) finds this *totally* unsatisfactory. The critic says that explanation by appeal to free will doesn’t really explain anything—it leaves everything mysterious. O’Connor is aware of this objection and replies to it. He replies that the critic is correct in claiming that appeal to free will does not explain in detail why a particular intention gets formed, rather than some other intention. However, appeal to free will can say this much: the particular intention gets formed at that time *because the person formed it at that time*. The critic responds by pointing out that this doesn’t tell us *why* the person formed it at that time. O’Connor agrees but does not see that lack of explanation as a problem. On the contrary, O’Connor believes this lack of causal explanation is exactly what one should hope for if one wants to have the power to author one’s own decisions.

But what about the role of influences on our decision-making process? Everyone is always being influenced in many ways, and that this creates questions about what it means to be in control of your own choice-making. It would be a serious mistake on O’Connor’s part to create a theory of free choice-making that did not provide any explanation of how influences work. Actually, O’Connor does not talk very much about influences, but what he does say is very direct: an influence on you affects the probability that you will choose one way rather than another. That is, it does not cause you to choose one way rather than another, but it does make some choices more likely than they

otherwise would have been. I assume influences never can make a particular choice 100% certain, since then the probability would be so strong that there was no longer any possibility that you could form any intention other than the one you do in fact form. In that case, the influence would seem to have caused the choice, and so the choice would not count as being made freely for O'Connor.

Similarly, on O'Connor's view, you do have *reasons* to make whatever free choices you make, but having those reasons does not *cause* you to make the choices. If you were *caused* to make a choice by your having reasons, then on O'Connor's view it would not be *you* that formed the executive intention. Instead, it would be your having the reasons you had that would explain why you made the choice you made. That would mean the next time you had those reasons in a similar situation, you would choose exactly the same way again. But on O'Connor's view, such regularity of choice-making, based on the causal power of reasons, is not free will—it is opposed to free will (because the exercise of free will for O'Connor has to be uncaused). So, according to O'Connor, **one cannot fully explain why a particular free choice is made by pointing to the reasons the person had for making that choice.** Having those reasons is *not* enough to bring about the formation of that specific intention at a particular time in the person's life.

However, the reasons that a person (consciously) has for doing something *influence* the person by making it *more likely* they will form an intention to do the thing they have the most reasons to do. As far as I can tell, O'Connor does not have more to say about exactly how this works. It sounds as though this change in the likelihood of a particular choice is an *effect* on the person, *caused* by the person's becoming aware of having reasons to act in a certain way. If so, then the power to make a free choice in a person is causally affected by the person's previous thinking. This sounds like a common sense view of how things work, but O'Connor does not provide any detailed explanation of exactly what is going on. How reasons (or influences) can cause a change in likelihood of a free choice without causing the person to make that choice remains mysterious to me.

It might help to make O'Connor's position clearer if we consider an example. Think about this one for a moment:

Suppose something (perhaps a conversation, perhaps something you see on TV, perhaps an experience you have) *causes* you to recognize that you have a good reason to sign up for a new cell phone plan. That thought in turn then causes you to form the executive intention to find a new plan. Then, acting from that intention, you actually do go looking for a new plan.

Here, I have given a simple, common sense cause and effect explanation for why you go out looking for a new plan. One thing leads to another in my story, in a perfectly natural way. But this is exactly the sort of explanation O'Connor wants to *rule out* in cases where we act freely, because this explanation traces the causes of your behavior back to something other than *you*. If it was a conversation you had that caused you to see the reason you had, then your going out to find a new plan was caused by a process that was set in motion by that conversation. To O'Connor, that makes it clear that your going out looking was *not* the result of your free choice, but rather the result of having had the conversation. If the cause of your going out looking was the conversation, the cause was not *you*. According to O'Connor, in that case you did not choose freely, because your choice was caused by something other than *you*.

So, it turns out to be extremely important to O'Connor's account of free choice that when you recognize that you have a reason to do X, that does *not* cause you to do X. Instead, it only *influences* you to do X by making it *more likely* that you will choose to do X.

This all means that the very same person under exactly the same circumstances with exactly the same reasons and influences might choose differently the next time the situation arises. It would be up to them to decide on that next occasion whether to make the same choice again, or a different one. There simply is *no* explanation for why they chose they way they did at the time they did, other than the explanation that they exercised their power of free will and that's how it turned out.

O'Connor recognizes how frustrating it is to the skeptic that he keeps on saying there is no further explanation available for why people choose the way they do, when their choices are free. But he points out that in the branch of contemporary physics known as quantum mechanics there is a similar lack of explanation for why particular subatomic events occur at the specific times that they occur.⁴ From this fact he takes comfort, since it suggests that his account of free will is no more mysterious than that. If the physicists can live with the idea that no explanations exist for why certain tiny events happen, and that is considered acceptable science, why can't we live with the idea that no explanations exist for why people freely choose one way rather than another?

So, for O'Connor, the key to our freedom of choice, the key to our free will, and thus the key to our moral responsibility, is that *we* are sometimes not caused by anything to form the intentions we in fact form. We author our intentions, when we freely form them. We are in control of forming them. Nothing is causing us to form them. We might be influenced to form them, but we are not caused to form them. This gives us the right kind of personal control to make us morally responsible. Our moral responsibility rests on the our powers of contra-causal free will.

Nevertheless, it is a mistake to think that O'Connor claims that free actions are uncaused. As we have seen, they are brought about by the formation of executive intentions, and that, too, is caused, even when it is the result of the exercise of free will. The intention comes into being because it is caused to come into being. The cause is the person, the agent. But nothing causes the agent to form the intention. That is where the agent's freedom appears; that is where contra-causal free will plays its role.

Objections

O'Connor's story about free will fits well with one commonly held view about how people make choices of their own free will. Choices (i.e., executive intentions) are simply brought about by the person, without any further explanation for how the person manages to do this. If you have a desire for more explanation than that, you'd appreciate knowing more about exactly how this works, but the whole point of the appeal to agent-causal free will is that at some crucial point in the choice-

⁴Page 76. More about quantum mechanics later.

making process, there really isn't any more explanation to be had.⁵ There simply are no additional causal mechanisms at work, hiding in the background, making things happen. End of story.

If you appreciate a good scientific explanation for an event, you will notice that a contra-causal account does not provide anything like what would today be recognized as a good scientific explanation for the crucial choice-making events that ground our freedom and responsibility. There is an important reason why that is so: in constructing good scientific explanations for events, we expect to provide detailed descriptions of cause and effect patterns that the event fits into. That is, we expect to be able to say how the event fits with the detailed, well-established patterns of cause and effect according to which things always work. So, in the case of choice-making, we would expect to talk about how people who are brought up in certain environments, and who have certain brain structures, and who have had certain experiences, thoughts, and the like, would make certain kinds of choices. We would see ourselves as products of our past experiences and thinking, interacting with the basic biological equipment we have been developing since birth. We would look for cause and effect patterns that hold for all similar people. We would expect that if we could somehow get a complete list of all the influences on a person, along with a good theory of how environmental influences interact with genetic makeup, to produce specific forms of brain development, we would be on the track to explaining why that person would choose a certain way in a certain specific circumstance.

Of course, we do not expect to be able to actually pull all this off in a real life case because we cannot get all the relevant information, but if we are committed to full causal explanations of human choice-making, this is the sort of thing that we would expect to aim for. And all of this cause and effect explaining, all of this looking for reliable patterns in nature, all of this is *directly contrary* to contra-causal free will as an explanation for choice-making. That is why contra-causal free will is called "contra-causal"! If contra-causal free will is real, then the search for a cause-and-effect explanation of choice-making is bound to fail at some point, for there simply is no such explanation for some crucial step in the free choice making process. To be sure, O'Connor does allow for influences to play some role in choice-making—influences affect the probabilities regarding how a person will choose. But clearly the point of contra-causal free will is that free choice is *not* the product of such influences.⁶

How can we decide what to think about this? If the fields of psychology, sociology, or neurophysiology were so far advanced right now that they could actually provide us with detailed well-confirmed theories that would explain much of our choice-making in terms of our past states,

⁵In the case of Robert Kane's particular theory about free will, the crucial uncaused events often happen earlier than in O'Connor's theory. They happen when a person makes a character-forming free choice. Kane's approach has the advantage that at the time of making most everyday choices, everything is caused in the normal way by the person's thoughts and desires or brain processes, or whatever psychologists would say causes us to do what we do – there are no executive intentions being caused "by the person." Still, if you want full causal explanations for why things happen, you won't be satisfied by Kane's approach either, since he does require that the person be able to make the crucial character-forming choices without being caused by anything to make one decision rather than another.

⁶Analogous remarks apply to Kane's theory. When a person freely decides what kinds of values to adopt, or what sort of person to be, the crucial thing that makes the decision "free" is that the person making it is not caused by anything to decide to be a certain way. Kane refers to these crucial decisions as "self-forming acts."

brain events, and the like, the question would have a ready answer. The answer would be, of course, that no such thing as contra-causal free will exists in people. We would have causal explanations of enough choice-making to be strongly suggestive of the conclusion that all choice-making is caused, and so there would be no good evidence in favor of the existence of contra-causal free will.

However, as you presumably already know, we are not in fact now in possession of such scientific theories. At present we have *partial* theories about some things that are related to the question, but we are nowhere near being able to explain in detail why people make the choices they do. To many scientists working in these fields, it seems as though progress is being made toward the goal of explaining human thinking, including human choice-making, but we are clearly still a long way from the goal. The fact that we are not near the goal makes it impossible to rule out contra-causal free will on the basis of developed scientific theories.⁷

Nevertheless, O'Connor's idea that human beings have a power of free will that enables them to stand outside the reach of the laws of nature when they form executive intentions does raise some serious questions about how free choices can be sufficiently explained. These questions about explaining free choices arise because of the way that O'Connor's theory handles (or, perhaps, fails to properly handle) influences. These questions serve as a basis for an interesting objection to O'Connor's view.

In order to set the stage for these questions, first consider how cause-and-effect explanation works in everyday life when free choice is not involved. How would we react if someone claimed that an ordinary natural event had no cause?

Louis Pojman gives us a nice illustration:⁸

Imagine how you would feel if, on visiting your dentist for relief of a toothache, she were to conclude her oral examination with the remark, "I certainly can see that you are in great pain because of your toothache, but I'm afraid that I can't help you, for there is no cause of this toothache." Perhaps, she calls her partner over to confirm her judgment. "Sure enough," he says, "this is one of those interesting noncausal cases. Sorry, there's nothing we can do for you. Even medicine and pain relievers won't help these noncausal types."

Your natural reaction to this case is to suppose that the dentists are incompetent. Maybe they do not know what the cause of the toothache is, but surely there is a cause. There just has to be a cause. Toothaches, you suppose, do not just happen for no reason at all. This reaction shows that you assume causal determinism is true for toothaches. Let's pursue this just a bit more with another example from Pojman:

⁷These remarks are in no way intended as a criticism of these fields of scientific study. Some scientific questions are hard to answer. Maybe the questions about why we choose what we choose are in fact answerable scientifically, and we just do not yet know the answers. Or, maybe there are causes for all our choices, but we will never be able to discover what they are because there is no way to ethically conduct the relevant experiments.

⁸ In his text *Philosophy: the Pursuit of Wisdom*. Belmont, CA: Wadsworth Publishing Co.(1994).

In Melbourne, Australia, weather forecasts for a twenty-four-hour period are exceedingly reliable. The predictions based on the available atmospheric data and the known meteorological laws are almost always correct. On the other hand, on Star Island off the New Hampshire coast, the official forecasts for a twenty-four-hour period are more often wrong than right. Suppose someone came along and said, "There is an easy explanation for the success of the Australian forecasts and the lack of success of the Star Island forecasts. In Melbourne the weather is caused by preceding conditions, but on Star Island, more often than not, the weather has no cause. It's cut off from what happened before.

As Pojman points out, we would all reject the proposed explanation for the inaccuracy of the Star Island forecasts. Instead of supposing there is no cause for the weather on that island, we would instead say something consistent with causal determinism for weather—something like

The forecasts are less reliable on Star Island because of the greater complexities of the factors that have to be taken into account and the greater difficulty of observing them, but not because sufficient causal factors do not exist.

These examples illustrate that, for a very wide range of natural events, we all assume that normal cause and effect patterns hold, and that these relations explain why things happen. When we are unable to actually provide an adequate cause and effect explanation of such events in terms of natural patterns that events can be expected to follow, we blame our ignorance—we do *not* say there simply was no cause. In our everyday lives, we generally seem to assume something like complete causal determinism, except perhaps when it comes to accounting for choices made by persons.

But what about human choice-making? Is that causally determined as well, in the same ways that other events are caused? Not if people have the power of contra-causal free will.

To understand what is at stake here, it is helpful to construct an example similar to those about the toothache and the weather. It might go something like this:

"Hey, I just heard that Jessica broke up with her boyfriend."

"Yeah, it happened over the weekend."

"I was really surprised. I thought they were getting along great. What happened?"

"Oh, Jessica just decided to do it. There really wasn't anything that led up to it."

"Wow! But surely she must have had some reason. She seemed so happy with that guy."

"Oh, sure, there were some things she thought about for why they should stay together, and why she might be better off if she broke it off. She knew it would really hurt him, and she knew she would be lonely for a while. But she decided she would break up with him."

"You mean nothing bad had happened to make her do that?"

"That's right. Jessica just woke up one morning having firmly decided that she would be break up with him. Nothing caused her to decide that. It just happened."

"But surely there must have been something going on. Maybe something she wasn't aware of, something hidden. Something had to make her decide that she would be better off without him."

“Nope. Nothing at all made her decide that. It just happened for no reason at all. There wasn’t anything subconscious, or anything like that.”

“But something must have made her decide to break up. It’s just crazy to think she simply decided without something, some reasoning, some event, some desire making up her mind.”

“No it isn’t crazy. It happens all the time. Choices just get made by us when we decide to make them. Nothing causes us to make our choices. That’s just the way we are.”

In this story, the one who reports about Jessica’s mental state describes what it would be like to make a decision that might well seem to fit O’Connor’s description of a free choice.

Giving an “explanation” O’Connor’s way, in the most precise form possible, we might improve a bit on what was said in the story, and add “Given her beliefs about the relationship and all of her desires, there was about a 90% chance that she would stay in the relationship with him.” If someone says, “What ultimately then made her choose to break up?” the only answer seems to be that *nothing* made her choose that way—she simply exercised her power to choose. If Jessica had been put in exactly the same situation again, she might have decided to stay with the guy. Nothing at all made her go against the 90% probability, rather than with it.

In fact, it’s difficult to see how there can be any probabilities at all, if ultimately it’s up to Jessica to decide, without being caused to decide one way rather than another by her own prior thinking or reasons. How do these probabilities work? I assume they are represented somehow in changes in Jessica’s brain structure, caused by prior events and states. But once they are in place, how do they influence her choice without causing it?

These questions suggest that there is a problem with O’Connor’s theory, since they appear to have no good answers. But, this line of inquiry leads to a way of developing this objection that makes it more telling: it’s not just that O’Connor’s view has difficulty explaining how influences work – it’s more than that; the view has difficulty explaining how the person’s desires, beliefs, reasoning, and the like could even be relevant to the choice that the agent makes. That is, it’s not just that we have trouble understanding the *process* by which the influence of these things could affect the probability of a given choice, but we have trouble understanding how it is even possible that these things *play any role at all* in free choice-making on O’Connor’s theory.

Look at the story about the dumping of the boyfriend again. In the story the reporter claims that the dumper did the dumping “for no reason”. That illustrates the objection to O’Connor’s theory. Does his theory imply that the agent causes executive intentions to be formed for no reason at all? If so, then what happened to the supposedly important role of the agent’s beliefs, desires, motives, and so on?

O’Connor’s response

One of the strengths of O’Connor’s account of contra-causal free will is that he is keenly aware of this problem and attempts to solve it. We have already seen part of that proposed solution: according to O’Connor, influences make it more likely that certain executive intentions will be formed, without causing any particular executive intention to be formed. I have already complained that this was not a terribly helpful explanation. But O’Connor has more to say about all this. He hopes to provide a substantial role for reasons that move person to make their choices, while still denying that reasons cause the formation of the freely formed executive intentions. Reasons are

among the most important influences on choice-making, and surely we need an account of how these particular “influences” are relevant to free choice-making. Surely, the reasons have to play *some* important role. Otherwise, it seems a person could constantly issue freely formed executive intentions that are totally disconnected from the person’s thinking, totally unrelated to the reasons the person is aware of for doing one thing rather than another. That sort of free will doesn’t seem very attractive as a basis for moral responsibility.

Here is O’Connor’s proposed solution, as I understand it: the *reasons* a person has for choosing a specific action are to be thought of in terms of the *desires* that the person has prior to forming the “executive intention” to act. For example, when you had a reason to skip class and study for your exam, that reason existed because you had various desires—desires like the desire to do well on your exam, the desire to have enough time to study for it, and so on. (Perhaps O’Connor is thinking that each desire generates a reason.) Probably one important way that influences work on us is to create or strengthen certain desires. (Advertising would often seem to work that way.) Perhaps this is part of what is going on when influences make a particular intention more likely to be formed—by creating a desire, or strengthening a desire, the influence does its work. O’Connor doesn’t say that, but I think it might fit well with the things he does say, and would do something to answer the first objection that complains about lack of explanation of the process by which influences work. So, in trying to deal with the second objection, we may have stumbled on an answer to the first objection. But, let’s go on to see more about O’Connor’s way of handling the second objection.

When the person freely creates an “executive intention” to do some particular thing, as an exercise of the person’s free will, according to O’Connor that intention *includes* reference to the desires that are thereby going to be satisfied. So, for example, when you formed the executive intention to go to class rather than studying for the exam, your intention was *not* merely the intention to go to class. Instead, it was the intention to go-to-class-in-order-to-satisfy-the-desire-to-do-well-in-that-course. **That is, the executive intentions formed by the exercise of free will *must* include a description of the desires that are going to be satisfied by the carrying out of the intention.** In this way, the intentions are *automatically* tied to the desires the person has, and if you had not been influenced to have a particular desire, then you would not have formed the intention to do something in order to satisfy it. This shows why your reasons are relevant to your free choices! The free choices are explicitly choices to satisfy certain specific desires that the agent has and that motivate the choice.

This means when you exercise free will, according to O’Connor’s view, you have to be consciously aware of what desires you have that are going to be satisfied by carrying out your intention. Otherwise, you would be unable to form the complex intention O’Connor’s account of free choice requires. Forming this complex intention is something O’Connor takes very literally—the person who forms the intention does so quite consciously, and could tell someone what their intentions are. “I’m going to go to class so that I can do well in that course because it’s important to me.” Or, “I’m going to skip class and study for my exam because I think it is more important that I do well on the exam and I need the time to study in order to do well.” Merely forming the intention to go to class, without building in the motive, is not enough for a free choice.

My problem with O’Connor’s replies

I assume that there are other influences on a person’s free choice making, besides desires. For example, the energy level a particular person may feel on a particular day will presumably influence

what things they decide to do that day, and it is not clear to me that this influence necessarily proceeds only by creating, strengthening, or weakening desires the person has. This gap in the theory troubles me. However, I leave that to you to think about. It might be the basis for a continuation of the first objection. I will turn instead to objections to O'Connor's idea that free choice must involve the formation of executive intentions that include the desires that are intended to be satisfied.

This question will suggest that there is trouble for O'Connor's approach here: Do not we often form intentions to do certain things without really knowing why? Can we not engage in self-deception? And aren't we sometimes moved to do things by unconscious forces of which we are unaware? People sometimes buy a new car, claiming to themselves that they are doing so because they "need" a new car, when in fact they do not "need" a new car. They really are buying the car because they are getting a psychological high out of it, but they will not admit that is the real reason, not even to themselves. Or, a person avoids doing things with the individual who lives down the hall, saying to themselves that the other person is stuck up, when in truth the person is just jealous of that individual. So, how can we form the kinds of intentions O'Connor thinks we should form in order to be acting freely? We often don't know the real influences, the real reasons, that move us.

O'Connor responds to this problem in an interesting way. He agrees that people sometimes do not know their own real motives for acting, and so would not be able to form an executive intention that properly includes those desires that are being satisfied by carrying out the intention.⁹ But he claims that this is not a problem for his theory, because in such cases, when the person does not know their own real motives, **they are in fact not choosing freely.**

This move on O'Connor's part is interesting, but it raises more questions. It means that unless I know what desires I intend to satisfy by my choice, my choice is not coming from an exercise of my free will! Apparently, free will is the power to bring about executive intentions that include an *accurate* description of the desires that are going to be satisfied by carrying out the intention. He says that he is "unable to conceive an agent's directly controlling his own activity without any awareness of what is motivating him."

Applying this new information to your decision to go to class, we get a result that might surprise you. If you say to yourself "I'll go to class so that I'll do well in the course," but your real motive for going to class rather than studying for your exam is different, your decision to go to class was not free after all. *That is, you are not the true origin of your decision when you do not know your true motives.* For example, if your real motive for going to class is that you hate studying for the other course, and so you are actually just finding an excuse to avoid studying, but you will not admit this fact to yourself, then on O'Connor's account, your choice to go to class could not count as a freely made choice. It was not a choice *you* made. Instead, it was made for you by the unconscious forces acting inside of you. The only executive intentions that count as products of the power of free will are those that accurately represent the desires that will be fulfilled by carrying out the intention.

I think this move on O'Connor's part raises a number of important difficulties. I say that partly because I wonder how frequently people do not understand their real motives. If people hardly ever really understand their true motives, then O'Connor would be committed to saying that people hardly ever make free choices. But more importantly I am troubled because I wonder how often

⁹ Pages 87-88.

any of us will be able to tell whether *someone else* has properly understood their own true motives.¹⁰ It seems to me there will be many cases in which we will not know whether the other person's intentions were freely formed. If I cannot tell whether other people are being honest with themselves about their real motives, then I cannot tell whether other people are acting from freely formed intentions. So, we may have here a theory of moral responsibility that cannot be applied in a practical way.

Moreover, in many cases of morally significant action, where questions of moral responsibility loom large, I would argue that there simply is no conscious formation of an executive intention to do such and such in order to satisfy such and such desire. The driver who kills a pedestrian through simple inattention (because he allowed his mind to wander) does not form an executive intention to be inattentive in order to satisfy some desire. So, if moral responsibility requires freely formed conscious intentions to do something in order to satisfy some desire, as O'Connor requires, I think that many cases of moral responsibility will be lost.

It seems to me this may be a significant problem. Do I really have to delve into the possibility that someone else is not being honest with themselves before I can tell whether they are morally responsible? Do I have to know how much of what they did resulted from conscious deliberation or conscious choosing on the basis of their real reasons? Do they even have to have such complicated intentions in order to be morally responsible? For example, do I have to know whether the firefighters who climbed the World Trade Center towers in the futile rescue effort on September 11, 2001, were being honest with themselves about their real reasons for doing so before I can tell whether they are praiseworthy? Do I have to know whether they even thought through why they were doing their job before I can tell that they count as deserving of praise? I think not.

I am thus not convinced that O'Connor has a good explanation for how influences that give us reasons are able to influence free choices without causing them, and how our reasons are even relevant to our free choices. This is not just O'Connor's problem, since there is a general problem for the contra-causal free choice view in trying to explain influences that somehow influence without causing. This has been a long-standing thorn in the side for theories of contra-causal free choice.

¹⁰There is scientific evidence that people can be deeply mistaken about the reasons for their behaviors, and that much of our choosing one course of action over another has little or nothing to do with what we consciously think about. (Bargh and Chartrand 1999, 54:462-479) Also (Wegner and Wheatley 1999, 54:480-492). O'Connor's picture of how we exercise free will may therefore be inapplicable in many cases to real people.