MUST KNOWERS BE AGENTS?

Linda Zagzebski

I. Act and Agency

Acts are performed only by certain kinds of beings—agents (from the Latin agent: that which is acting), although that tells us neither what an act is nor what an agent is, only that “act” and “agent” are defined corelatively. Ethics is a field primarily concerned with the evaluation of human agents, their acts, and the consequences of their acts, but again, this does not tell us what ethics does, only that certain things it does are connected to certain other things it does. In particular, it does not tell us what the scope of ethics is, which might turn out to be greater or less than we think.

I assume that most human beings are agents at a fairly early age, and it is possible that some other animals are agents also, although I will not discuss animal agency here. An agent is the kind of being that acts. To act is to exert power and, at least typically, to bring about a certain kind of effect through the exercise of that power. I say “at least typically” because successfully bringing about an effect of a certain kind might be treated as constitutive of agency—presumably, a power is not a power unless it is reliably effective. But we are somewhat forgiving about the degree of efficacy required of human agents since human power is obviously not infallibly efficacious, yet no one takes that fact to be incompatible with the existence of human agents. An agent is still an agent if he occasionally is ineffective in bringing about the effect associated with her power, and she is exercising her agency even on such occasions. The extent to which she can be ineffective and remain an agent will be one of the issues addressed in this essay.

An agent, then, is a being that brings about certain kinds of effects through the exercise of a power. But to do so is not sufficient for acting since many artifacts and inanimate substances bring about similar effects, but they do not act. We may even say they have “powers,” although it is likely that a chemical agent (note the term) has a power only in an extended sense. In the strict sense only conscious beings have powers. It is interesting to consider why that should be the case. What difference does it make if a being with the capacity to produce effects is a conscious agent?
tion presupposes an understanding of power. Our notion of active power is more conceptually basic than that of causation, Reid claimed; it is presupposed by our knowledge of ourselves as rational and moral agents. The idea of causation is derived from the idea of agency and responsibility. Presumably, we would not have acquired the idea of causation if we were merely witnesses to nature. Reid maintained that agency even appears in the operation of our intellectual powers, including those operative during perception. The faculty of perception is an original power of the mind. The true cause of perception is the agent exercising this power to produce an effect. Sensations and impressions are not the real causes, much less the objects perceived.

I mention Reid’s strong view on the place of agency in perception not to endorse it, but to highlight a question that I think deserves more attention in epistemology: What is the place of acts and agency in the acquisition of epistemic states, particularly those that are evaluatively positive, such as justified belief, responsible belief, and knowledge? In this essay I am particularly interested in the relevance of agency to the acquisition of knowledge. It is uncontroversial that processes and events of some kind lead up to a human being’s coming to know something. Coming to know is an event, and the issue of what causes that event is important, not only because some form of the causal theory of knowledge might be true, but because understanding the cause of something almost always helps us to understand it better. Is knowledge best understood in the model of event causation or on the model of agent causation? That is another question I want to begin investigating in this essay.

While knowledge and justification are often connected with causation in the contemporary literature, agent causation is rarely mentioned. Although Aristotle and Aquinas referred to the “act of knowing,” nowadays knowing is more commonly construed as a state rather than as an act. This may be due in part to the fact that perceptual knowledge is commonly taken to be the paradigm and perception is usually understood as a relatively passive state, or at any rate, as a state one acquires prior to the exercise of one’s agency. It may also be partly due to the fact that the range of acts and the corresponding range of agency has narrowed significantly in modern philosophy for reasons that derive from ethics rather than the philosophy of perception. The importance of the act in Kant was gained at the price of narrowing its scope to a single tightly circumscribed deliverance of will. The mind itself is now often viewed as a passive information processor rather than as an active agent. This view has led to a shift in the prototype of the act in modern philosophy. We no doubt find it curious that in Aquinas the act par excellence is a mental act since in contemporary discussions the prime example typically given of a basic act is the raising of one’s arm. So these days when we think of an act we usually think either of an act of will or of a willed bodily movement. Cognitive and perceptual acts only make sense on this view if preceded by acts of will. The broader Aristotlean category of the voluntary and the even broader category of acts both voluntary and non-voluntary have generally disappeared from discussion. Granted, the fully intentional act is in many ways the most interesting kind of act, but that does not mean that there are no interesting differences between acts and non-acts. If there are such differences, the epistemic arena is one place in which we would expect them to appear.

So far I have identified three sets of questions about agency that I want to begin investigating:

1. What are the conditions for being an effective agent? What determines that an agent is effectively exercising her agency on a particular occasion? Must she be reliable? Is her efficacy determined by what she is able to do in counterfactual circumstances?
2. Is there any important difference between an effect arising from the act of an agent, whether voluntary or non-voluntary, and events brought about by a non-agent? In particular, does it make any significant difference to epistemology?
3. Is knowledge best understood on the model of event causation or on the model of agent causation?

II. Agency and Counterfactual Conditions

What are the conditions for being an effective agent? If I am right in my conjecture that effectiveness is part of the concept of an agent, then the conditions for being an effective agent are the conditions for being an agent. An effective agent is one whose acts are successful in reaching their ends. Presumably effectiveness comes in degrees, and so my degree of effectiveness as an agent is partly a matter of the proportion of my successes to failures in achieving the ends of my acts. But my effectiveness as an agent is also a matter of the extent to which my successes can be credited to me rather than to something else. An effective agent is one who reaches her end because of her act, the exercise of her power. This rule out both accidental success and success that is non-accidental but due to something other than the agent. An effective agent produces a high proportion of good dishes over mediocre ones and does so because of what she does in exercising her cooking ability rather than by chance or because someone else is guiding her every step of the way. An effective agent gets the credit for her culinary successes. An effective teacher produces a high proportion of students knowledgeable or skillful in the subject of his course, and he does so because of what he does in exercising his teaching ability rather than by chance, or because his students are simply bright enough to learn on their own or for some other reason. An effective teacher gets the credit for his students’ success in learning.

The same point applies to our effectiveness as moral agents. An effectively compassionate agent is one who produces a high proportion of successes at alleviating suffering and who does so because of the exercise of his own power in reducing suffering. He gets the credit for the alleviation of suffering that follows from his efforts. In fact, any end an agent has is something he can be effective or ineffective at bringing about. His effectiveness is both a matter of reliably producing the intended effect and of doing so because of the exercise of his own power rather than because of the many other conditions that are also operative. An effective agent gets the credit for the effect.

One of the ends agents have is to get to the truth. Getting truth is probably the primary epistemic end of agents, or at least a very important end. It is also arguably
a "natural" end of belief formation. Epistemically effective agents therefore have a high proportion of successes in reaching truth and avoiding falsehood. And their successes must be credited to themselves rather than to something else.9

Must an agent satisfy counterfactual conditions for effective agency? If so, must she satisfy them on each occasion in which she is exercising agency? Reid maintained that in the case of the active powers (as opposed to the intellective powers), an agent does not have the power to do something unless he has the power not to do it. This is strikingly similar to the Principle of Alternate Possibilities (PAP) often proposed by non-determinists as a condition for moral responsibility: A person is not responsible for her act unless she could have done otherwise. According to PAP a counterfactual condition is necessary for the attribution of moral responsibility to an agent's act. Reid's principle is stronger since he proposes that alternate possibilities are necessary for the power an agent must have to act as an agent, not just for the subset of acts for which he is responsible. Therefore, if PAP fails, Reid's principle fails also.

Thirty years ago Harry Frankfurt presented a famous thought experiment that arguably leads to the conclusion that PAP is false. In this section I want to look at Frankfurt-style cases, but not for the usual purpose: moral responsibility is not the focus of this essay. But the moral of Frankfurt cases, I believe, is important because it can be generalized to apply to many principles that offer counterfactual conditions for the application of some property. Knowledge is often defined in a way that includes counterfactual conditions. Many philosophers who propose such accounts intend these conditions to specify what it takes for the knower to get credit for her belief. Since the idea of credit is similar to that of responsibility, we might expect criteria for responsibility to have an analogue in criteria for epistemic credit, and we might also expect that any problems in the former may be reflected in problems in the latter. Other proponents of counterfactual accounts of knowledge separate epistemic credit from knowledge. But I think that even these accounts can be illuminated by examining epistemic parallels to Frankfurt cases. That is because I believe that Frankfurt has identified a very general problem in counterfactual conditions for any property, whether or not it has anything to do with responsibility. My intention, then, is to see how epistemic Frankfurt-style cases can illuminate both the connection between counterfactual conditions and epistemic credit, and the broader issue of the extent to which counterfactual conditions are necessary for knowledge. My hope is that these thought experiments will make it easier to answer the questions posed at the end of section I.

Case 1: Standard Frankfurt case

Black, an evil neurosurgeon, wishes to see White dead but is unwilling to do the deed himself. Knowing that Mary Jones also despises White and will have a single good opportunity to kill him, Black inserts a mechanism into Jones's brain that enables Black to monitor and to control Jones's neurological activity. If the activity in Jones's brain indicates that she is on the verge of deciding not to kill White when the opportunity arises, Black's mechanism will intervene and cause Jones to decide to commit the murder. On the other hand, if Jones decides to murder White on her own, the mechanism will not intervene. It will merely monitor but will not affect her neurological function. Now suppose that when the occasion arises, Jones decides to kill White without any "help" from Black's mechanism. In the judgment of Frankfurt and most others, Jones is morally responsible for her act. Nonetheless, she seems to be unable to do otherwise since if she had attempted to do so, she would have been thwarted by Black's device.10

Discussion of cases like this has generated a large literature. Non-determinist defenders of PAP have argued that Frankfurt's thought experiment fails to demonstrate the falsehood of PAP, while many determinists have argued that these cases successfully falsify PAP.11 I have argued that Frankfurt cases are successful in demonstrating the falsehood of PAP, but they are unsuccessful in supporting determinism.12 Frankfurt cases succeed in showing that principles that offer counterfactual conditions for the application of some property can fail because of the possibility of a counterfactual manipulator, and so the counterfactual condition is not strictly necessary for the application of the property in question. But it would be too hasty to conclude, as Frankfurt does, that the counterfactual condition can fail systematically, much less that it is irrelevant. The reason for this might be that counterfactual conditions are usually proposed not because actual conditions are literally inadequate, but because it aids our understanding of conceptually abstract properties such as responsibility, power, causality, and knowledge to think of them in terms of what happens in non-actual circumstances. If I am right about this, counterfactual conditions can fail even when they are relevant and perhaps even when their failure must be selective.

To make the point, let me give a harder Frankfurt case (F case) that changes nothing in the standard case except background conditions.

Case 2: Altered Frankfurt case

In the standard F cases the device is set to go into operation a maximum of once, but it is not needed in the case in question because Jones makes what Black considers the "right" choice on her own. But suppose that Black has been systematically manipulating Jones's choices all along. Every time Jones is about to make a choice, if it is the one Black wants, the device does nothing, but if it is not the one Black wants, the device makes Jones choose the way Black wants her to. And let us suppose that Jones has been living with this device for many years. A multitude of her choices have been manipulated and changed by Black, unknown to Jones. And suppose this is one of those times that Jones makes the choice Black wants and so the device does not go into operation.

Is Jones responsible in this case? Perhaps she is; I am not going to argue that she is not. My point is that the case is harder and it can be made harder still. This could be the only time in her life that Jones has made a choice on her own without the intervention of the machine. If so, we might worry that the counterfactual manipulator is not only manipulating the circumstances, he may be manipulating the person. If we hesitate in ascribing responsibility, I suspect that that is because we think that Jones's lack of responsibility for all her other choices can infect her responsibility for
the choice the one time the device is not needed. A person who never satisfies PAP might be a different kind of being than one who usually does. She is arguably not an agent, the kind of being who are morally responsible. Perhaps moral agency, like virtue, is the sort of thing that takes practice.

But whether or not we ultimately decide that Jones is responsible in case 2, the fact that this case is harder suggests that the counterfactual condition can still be a good sign of the presence of the target property even if it is not necessary that it ever be satisfied. If so, what is really essential to responsibility needs to be extracted from a story that explains why we sometimes think the property obtains even when the counterfactual condition is not satisfied. This is important because the wrong moral to draw from these cases is that the counterfactual condition is irrelevant. Frankfurt was right that he had described a case in which alternate possibilities are not necessary for responsibility, but then he leaped to the conclusion that therefore there is nothing blocking the acceptance of determinism. In my view, what Frankfurt cases show is that whereas alternate possibilities are not strictly necessary, they are usually associated with responsibility because they are a sign of something that really is necessary—the presence of agency. Agency is a necessary condition for responsibility. If it is lacking, so is responsibility.14

If I am right about this, we can apply the same point to properties other than moral responsibility for which there are allegedly counterfactual conditions. Epistemic credit is a property that is closely allied to responsibility and in addition to being interesting in its own right, it may be an ingredient in knowledge. Can we use epistemic Frankfurt-style cases to test the need for something like PAP in cases of epistemic credit and/or knowledge?

Case 3: Epistemic Frankfurt case

Suppose that Jones is very good at identifying vintages of Bordeaux. In particular, she has no trouble distinguishing a '94 Château Margaux from very similar wines. Black knows that Jones is going to be tasting different vintages of Margaux without knowing in advance the year of the vintage she is tasting. He has installed a device in her head that can make Jones believe that the next wine she tastes is a '94 Margaux whether it is or not. (Never mind why Black would want to do such a thing.) When Jones tastes the next wine, if she appears about to judge that it is a '94 Margaux, the device does nothing. But if Jones is about to judge that it is anything else, the device will interfere with her tasting sensations and will lead her to think it is a '94 Margaux. Now suppose that she tastes a '94 anyway and believes it is a '94, and Black's device does nothing but monitor what is going on in Jones's nervous system. Jones's tasting faculties and taste memory are working fine and she comes to have a true belief in the normal way.

My intuition in this case is that Jones gets epistemic credit, and for the same reason that she is morally responsible in the standard Frankfurt case. Furthermore, I am willing to say that she has knowledge. She knows she is tasting a '94 Margaux. The device does not operate and its very existence is an accidental feature of Jones's epistemic situation. As in the standard F case, the counterfactual manipulator has no effect on our inclination to judge that Jones has the property we are inspecting. But notice that she not only fails the test of alternate possibilities, she also fails some well-known counterfactual conditions for knowledge. For example, she may fail the Nozick conditions since she would have had the same belief even if it had been false.15

We can make up a harder epistemic F case, parallel to the harder regular F case as well.

Case 4: Altered epistemic Frankfurt case

Suppose that it will serve Black's sinister purposes if Jones forms a specific set of beliefs. Some of these beliefs are true, but many of them are false; their truth or falsehood is irrelevant to Black's purposes. Jones's beliefs are systematically manipulated by the device Black has installed in her head. Many times the device has forced Jones to form a false belief that she would not have formed on her own. Other times it forces her to form a true belief that she would not have formed on her own. Still other times it permits her to form a true belief on her own in such a way that in the absence of the device we would not hesitate to say that Jones is credited with getting the truth and knows the proposition in question. Now suppose Jones forms the true belief that she is tasting a '94 Margaux in this way.

My reaction to the altered epistemic F case is the same as to the altered F case. While I do not propose that Jones does not know the identity of the wine she is sipping, I find it a more difficult case. Perhaps believing on her own, like acting on her own, is something that at least in many cases requires a background of practice in believing/acting on her own. Perhaps very simple sorts of perceptual knowledge do not require such a background, and I will have more to say about the simplest perceptual knowledge in section III, but believing out of an acquired power of taste discrimination does seem to be the sort of perceptual knowledge that may require such a background, or at least the possibility that it is required is enough to make some of us wonder that Jones does not know in case 4.

Epistemic Frankfurt cases have something in common with the standard evil demon scenarios, but there is an important difference. The F cases involve manipulation of agency in a way that does not appear in the skeptical scenarios. The evil demon gives the agent misleading sensory inputs that inevitably result in false beliefs, but the agent's control over her reasoning process is not altered. This is like deception, whereas the F cases are cases of coercion. Extended manipulation of the reasoning process itself undermines our ability to initiate our own cognitive projects. The counterfactual manipulator in epistemic F cases therefore attacks the agent's epistemic responsibility, whereas the evil demon does not.16

I conclude that the moral of the epistemic F cases is the same as the F cases. Cases 1 and 3 show that the counterfactual condition is not strictly necessary for the target property, but cases 2 and 4 show that it is not irrelevant. It is a sign of something deeper: The agent gets credit for reaching the end. The agent must be an agent, and the fact that she gets a true belief must be due to her. Whether or not her belief is voluntary, her agency is central to acquisition of the belief. Causal processes that bypass her agency take away her epistemic credit, and they also take away her knowledge.
Notice that the application of epistemic F cases to knowledge need not go through the concept of epistemic credit. Some accounts of knowledge analyze it in terms of counterfactual conditions, and some of those conditions fail in the epistemic F cases whether or not the reason is that the agent lacks epistemic credit. I have already remarked that case 3 might fail Nozick's conditions even though it is intuitively a case of knowledge, and it no doubt fails the conditions of other theories as well. The conditions for relativism are a special and interesting case because they are tied to effectiveness, but not to the "agent gets the credit" aspect of effectiveness. Instead, they are tied to the first condition for effective agency mentioned above—that the agent must have a high proportion of successes over failures. However, the counterfactual manipulator can arguably make the agent unreliable when we think she has knowledge and reliable when we think she does not have knowledge. In case 3 where the device is set to operate a maximum of once, Jones is arguably still reliable and she does have knowledge, as the theory predicts. In case 4 she is unreliable and, as the theory predicts, we hesitate to say she has knowledge because of the worry that the machine has tampered with her agency. But my experience with proposing this case to others leads me to think that the intuitive judgment here is unclear. If Jones is able to form true beliefs in the normal way through the exercise of her own power when the machine is not operating, perhaps it is reasonable to say that she has knowledge even if the machine has made her unreliable.

Furthermore, consider the case of the benign manipulator (Case 5), who makes Jones believe only truths. In such a case Jones is both highly reliable and satisfies some counterfactual conditions for knowledge, but it is doubtful that she has knowledge when she acquires a machine-produced true belief. An agent relativist who shares this intuition might say that the problem here is that Jones is not reliable. What is reliable is the machine operating in her. If so, the problem I am raising is not a problem for agent relativism itself, but for the view that reliability entails counterfactual conditions that are subject to Frankfurter-style manipulation. A careful agent relativist could therefore accommodate the intuition I have about case 5 and perhaps case 4 as well. In the latter case the agent relativist could say that the agent herself is not unreliable. What is unreliable is the complex of agent plus machine.

Similarly, the process relativist could say that what is unreliable in case 4 is the process used by the complex of agent plus machine. In case 5, if we assume that the benign manipulator does not generate the belief directly, but makes the process the agent is using or her faculty reliable, there is no problem in concluding that she has knowledge. After all, even ordinary knowledge often makes use of aids to our faculties: eyeglasses, hearing aids, and one day, perhaps, computer chips installed in our brains to aid our memory. These devices make it easier for us to get knowledge; they surely do not take it away. So if the benign manipulator in case 5 installs a device that is comparable to these aids, only better, there is no difficulty. But there are ways the benign manipulator could operate that would threaten Jones's epistemic agency or even eliminate it. The device could bypass Jones's perceptual and cognitive functions entirely. Suppose it implants true beliefs in her head overnight, counteracting any false beliefs she unhappily acquired during the day and adding many other true beliefs besides. The process is as reliable as a human-generated process can be. Does Jones have knowledge?

Here I think our agreement on the target property we are trying to analyze breaks down. Some philosophers will probably be willing to say that Jones has knowledge in this instantiation of case 5. But if so, they are not treating knowledge as something we earn or even something we contribute to through our own powers; it is something we are blessed with. I recognize that it is not obvious that knowledge is a unitary concept, and it may be a flexible enough concept to include instances of knowledge as a gift rather than as something we merit. But I think we should try to see how far we can get with a unitary concept, and if we are going to aim for a unitary concept, it should be one of merit, not blessedness. If so, Jones does not have knowledge when her true beliefs are produced wholly through the action of a benign manipulator.

I conclude that we should be wary of making manipulable counterfactual conditions necessary for either epistemic credit or knowledge (case 3), nor are they sufficient for knowledge (case 5). But neither should we ignore the importance of the close association that ordinarily obtains between those conditions and knowledge.

III. Knowledge, Agency, and Virtue

Let us now go back to giving a partial answer to the three sets of questions posed at the end of section I. I have already answered the general question of what it takes to be an effective agent. An effective agent is reliably successful in reaching her ends and she does so through the exercise of her own power. Frankfurter-style cases show that efficacy is associated with the satisfaction of counterfactuals, but cases 1 and 3 show that their satisfaction is not necessary on every occasion in which agency is operative, and case 5 shows that it is not sufficient. The satisfaction of appropriate counterfactuals is not constitutive of agency, but is a sign of it.

For the same reason, it is not strictly necessary that I be reliably effective in order to effectively exercise agency on a particular occasion. In cases 3 and 4 the agent is unreliable but she is arguably successful in exercising her agency, although some of us hesitate. Our hesitation, however, does not stem from her lack of reliability per se, but from the worry that the counterfactual manipulator has interfered with her power to be an agent. And case 5 shows that reliability is not sufficient. Again, I think that the fact that an act of belief is that of a reliable agent is a sign of what we are looking for—that the act or belief really belongs to the agent; that she gets credit for it.

What difference does it make to epistemology if a causal process is brought about by the exercise of agency? Since getting the truth is one of my ends, I am an effective epistemic agent to the extent that I am reliably successful in reaching the truth and do so because of the exercise of my epistemic powers. My epistemic success is due to me. In case 5 the benign manipulator makes Jones believe only truths, and we considered a causal process that bypasses Jones's perceptual and cognitive faculties completely. My intuition is that Jones neither gets epistemic credit nor has knowledge.
in this case. So it matters epistemically that her agency is not operative. It also matters epistemically that her agency is operative in cases 3 and 4. In case 4 the intuition that she knows and gets epistemic credit is weaker than in case 3, but that is because of doubts about her agency. If she really is acting as an agent, the intuition that she knows is fairly strong even though it is clear that she is unreliable. Agency seems to be enough to make up for the lack of reliability, whereas even the presence of reliability cannot make up for the lack of agency.

In stressing the importance of agency in getting epistemic credit and knowledge, it must be admitted that human agents are not pure agents and it is unreasonable to expect otherwise. Our beliefs, like our desires, often come unexpectedly as the result of causal processes that are largely external to ourselves, and there is nothing abnormal about that. Many philosophers are willing to say that some of these beliefs constitute knowledge. In particular, some perceptual knowledge may be in this category. The examples of acquired wine discrimination show that not all perceptual beliefs can be in this category, and probably most perceptual beliefs utilize some degree of learned discrimination. But perhaps the simplest cases of perception do not require the operation of agency. What should we say about these cases?

I have argued elsewhere that agency operates counterfactually even in simple perceptual cases: If an intellectually virtuous agent had indications that her perceptual ability or her perceptual situation was in some way deviant, she would withhold or withdraw perceptual judgment until she could investigate. If this is right, agency operates even in the simple cases of automatic perceptual belief formation at the second-order level, the level of reflectiveness. But let us look once again at the Frankfurt cases since I think we can use them to illuminate the place of agency in evaluating simple perceptual beliefs.

So far I have said nothing about what happens from the agent's viewpoint when the Frankfurt device operates. Although we can only guess at the phenomenology of the device during operation, I imagine that the victim has the experience of impulsively deciding/believing something unexpected. She is about to decide not to kill White when she suddenly decides to kill him after all. Or she is about to believe that the wine she is sipping is a '95 when she abruptly decides it's a '94 instead. Since all of us are subject to changes of mind and sudden impulses, this will not necessarily seem peculiar unless it happens very often or if the decision/judgment seems to the agent to be out of character, something she can hardly imagine herself doing/believing. But when an agent suddenly acquires a belief or suddenly makes a decision, she should subsequently reflect about her own belief/judgment. I think that agency requires this. One of the central features of agency is self-reflectiveness, and since one of the aspects of self-reflectiveness is the second-order desire for self-integration, agents need to tell themselves some story about the unexpected act or belief. I am not suggesting that this is something we should do constantly, and certainly not obsessively. But at some point we should assess our sudden beliefs and decisions, at least those that have any important consequences or implications for our view of ourselves. A sudden belief that comes out of nowhere, like a sudden urge, ought to be either endorsed or repudiated. Agency does not require that we do one rather than the other, but it does require that we do one or the other, probably not for every sudden belief and impetuous act, but certainly for some of them. I suggest that when the machine operates in cases 2 and 4, a test of whether the machine has eliminated Jones's agency is whether she reflectively endorses or repudiates her machine-produced beliefs/acts after the fact.

But, you will ask, cannot the device also operate on the second level, the level of endorsement? Yes, of course it can, and if it does, and if there is no higher level of reflectiveness at which the machine does not operate, then it is likely that Jones has indeed lost her agency.

In some respects our simplest perceptual beliefs are like the beliefs produced by the counterfactual manipulator. They come upon us without warning and without any effort on our part. In normal situations there is nothing suspicious about them and we have learned to expect to have perceptual impressions almost all the time, so the beliefs formed from these impressions are not like suddenly acquiring the urge to kill. If I am normal, these beliefs are easy to integrate into my view of myself and my environment, unlike the urge to kill or the sudden belief that airplanes are following me. Perceptual beliefs are typically unimportant, and there is no great need for reflective endorsement in many cases. But when the consequences of believing them are serious, reflective endorsement is called for. If Jones is a professional wine taster she should reflect about the grounds of her belief when tasting wine, at least when something of importance hinges on her judgment. If some of her beliefs are unknowingly machine-produced, she should reflect about them if she has reason to suspect that there is something out of the ordinary in the way she got them. Since reflectiveness preserves her agency, she can get epistemic credit even when the belief is machine-produced. It is even possible that she gets credit for her beliefs in case 5 if she later endorses them when the machine is not operating. For the same reason, she can get epistemic credit for non-voluntary perceptual beliefs. An agent reflects about her beliefs from time to time, particularly when they are either suspicious in their origin or of special importance. And this includes perceptual beliefs. True perceptual beliefs earn the believer epistemic credit when the agent exercises her agency over them at the level of reflective endorsement.

The view I am proposing on the place of agency in belief is similar to Christine Korsgaard's interpretation of autonomy in desire. Korsgaard argues that according to Kant, autonomy is compatible with acting out of desire as long as the reflective mind endorses the bidding of desire. In this way we are self-determining even when we act instinctively. Similarly, I am suggesting that we can be autonomous agents even in the simplest perceptual knowledge by endorsing the bidding of our pre-reflective minds. The connection between the reflective endorsement of belief and the second-order endorsement of desire has been explored in some detail by Keith Lehrer. Lehrer calls the positive evaluation or endorsement of desire "preference" and the positive evaluation or endorsement of belief "acceptance." I am not suggesting that the place of agency in knowledge and responsible belief is limited to such second-order endorsements, but it is a way in which agency can extend even to those parts of the self that are initially acquired non-voluntarily. Since it is likely that some perceptual beliefs—those requiring the most meager conceptual resources—are in this category, it means that agency can extend even to such perceptual beliefs.
I suspect, then, that self-determination, autonomy, and agency operate on a much wider scale than is included in intentional action or even the broader class of voluntary action. If I am right about this, not only is it a mistake to focus on the voluntary/non-voluntary distinction in analyzing epistemic responsibility, it is a mistake in the analysis of moral responsibility as well.

This position obviously needs considerable refinement and a number of objections need to be answered. The scope of the self is a difficult matter, and the Kantian view of the self is notoriously narrow. Why think that our first-order desires and beliefs are any less a part of our selves than our will or second-order endorsements? Why does the latter have authority over the former? It is far too facile to identify the self only with those desires/beliefs with which we identify at the second level. After all, a person who has numerous first-order desires or beliefs with which he does not identify has a different self than he would have had if he had not had them. He has, we would say, a fragmented self. And there may even be a sense in which he is responsible for the desires/beliefs with which he does not identify or even explicitly repudiates. That might explain why even such desires/beliefs are in some sense his own. Some epistemologists wish to extend the scope of knowledge to include many such beliefs. I've already said that knowledge may not be a unitary concept and there may not be any way to resolve some of the disputes about the application of "knowledge" to cases in which agency clearly does not apply, but I am suggesting that we need not assume that agency does not apply when an epistemic state is initially acquired non-voluntarily. Furthermore, I suggest that part of the reason it is hard to decide whether knowledge or epistemic responsibility applies to beliefs about which agency does not apply is the vagueness of the boundaries of the self.

We can now answer the second question posed at the end of Part I: Is there any important difference between an effect arising from the act of an agent, whether voluntary or non-voluntary, and events brought about by a non-agent? In particular, does it make any significant difference to epistemology? We have seen that the non-voluntary acts and beliefs of agents can differ in important ways from events that are produced by non-agents. If it is the act or belief of an agent, the agent's subsequent reflective endorsement makes it voluntary on the second level. The agent either does or does not make the belief her own. Even non-voluntary acts/beliefs can therefore earn the agent credit (or blame), and in the case of beliefs, they may constitute knowledge.

This brings us to question (6): Is knowledge best understood on the model of event causation or on the model of agent causation? I have already suggested that epistemic credit is earned by an agent only when her agency is operative, either in the initial acquisition of the belief, or in her later reflective endorsement of the belief or beliefs like it. Since on my view epistemic credit is a component of knowledge, I am also willing to say that knowledge requires the operation of agency. But throughout this essay I have not rested my case that the agent does or does not have knowledge in the various Frankfurt cases on the fact that she does or does not have epistemic credit. In case 3, for example, it seems to me that the agent has knowledge. It also seems to me that she gets epistemic credit. But I am not suggesting that she has knowledge because she gets credit. Those epistemologists who separate epistemic credit, either in the form of justifiedness or responsibility, from knowledge will no doubt have more complicated responses to these cases. They might say, for instance, that agent causation need not be operative in knowledge, but it does need to be operative in generating justified or responsible belief. But I will not try to sort out here the various possibilities that emerge when various forms of epistemic credit are separated from knowledge. My position is that the fact that an agent has knowledge is "up to her," to use Reid's words. She need not be responsible in the sense that requires alternate possibilities or even voluntariness, but she needs to be exercising her agency either at the first- or second-order level.

Aquinas defines virtue as the perfection of a power. Within the context of this essay that would mean that virtue is a property that makes agents effective. In a broad sense of virtue there can be physical virtues, culinary virtues, veering virtues, and so on. Epistemic virtues make us effective epistemic agents. An effective epistemic agent is one who reliably reaches her epistemic end and who reaches her epistemic end because of her, not by chance or because of something outside of her. Need she be exercising a virtue in getting knowledge? I have said elsewhere that that is too strong a requirement. She does need to be exercising a power and she needs to get to her end because of her power; she must be exercising her agency. She need not intentionally aim at her end, however. She need not even be acting voluntarily. In fact, it might turn out that she need not even be generally reliable in reaching her end, assuming that in case 4 we ultimately judge that she knows. But she needs to be an agent. Her agency is critical in explaining how it is that she ends up with her true belief. A causal sequence that leaves out her agency is not good enough.

At the beginning of this essay I remarked that ethics is concerned with the evaluation of agents and their acts, but that does not tell us the full range of ethics since both agency and act could be more or less extensive than we think. I have argued that agency is operative in getting epistemic credit and knowledge. The scope of agency includes those evaluative aspects of belief investigated by epistemology. In other work I have argued that it is artificial to separate epistemology from ethics. The role of agency in beliefs as well as in acts further supports this position.

Notes

1. To complicate matters further, it is likely that not everything an agent "does" is an act, although it is difficult to draw a systemic distinction between a non-voluntary act and a non-act done by an agent. I will not pursue this distinction here. See Jonathan Bennett, The Act Itself (Oxford: Clarendon Press, 1995), beginning of chapter 2.

2. Agent causation is sometimes even thought to be necessary to explain event causation since the latter generates a regress ending in a non-event: an agent.

3. Susan Sa Té Meyer denies that Aristotle contrasted agent causation with event causation since on Aristotle's view, every efficient cause is a substance, whether or not it is an agent and, in fact, the effect is a substance also. See "Self Movement and External Causation," in Self-Motion, Mary Louise Gill and James Lennox, eds. (Princeton, NJ: Princeton University Press, 1994). But for the purposes of this essay it is worth pointing out that whether or not Aristotle had a notion of event causation, his notion of efficient causation...
was more like what we mean by agent causation than event causation. The causation due to human agents is a subclass of the substance causation found in nature. A broader definition of agent causation is given by William Rowe, who defines "agent causation" as "the idea that the primary cause of an event is a substance." Rowe claims that Reid used "agent causation" in a narrower sense (The Cambridge Dictionary of Philosophy, Robert Audi, ed. 1995, p. 13). On Rowe’s definition, Aristotle’s notion of causation would qualify as agent causation.


5. Not only did Aquinas place a great deal of importance on the cognitive act, but Eleonore Stump argues that he believed most acts of intellect are not causally determined. See “Aquinas’s Account of the Mechanisms of Intelectual Cognition,” Revue Internationale de Philosophie 21 (1967), 387-397.

6. There is more than one sense of ends used in the history of ethics and sometimes they are not clearly distinguished. For the purposes of this essay, ends are those at which the agent consciously aims, or they can be natural ends. Either interpretation is permissible.

7. These two senses of chance are distinguished by Wayne Riggs in “What are the ‘Chances’ of Being Justified?” The Monist 81, 3 (July 1998): 452-72.

8. Of course, no effect is brought about by a single cause, whether the cause is an event or an agent exercising a power. But some causes are much more salient than others, and that leads us in many of the most interesting situations to designate one or a small number of causes as “the” cause. As far as I can see, the vagueness of this usage does not affect the argument of this essay.

9. Wayne Riggs addresses the issue of the need for the knower to get get credit for her belief in “Reliability and the Value of Knowledge,” forthcoming, Philosophy and Phenomenological Research.

10. This adaptation of Frankfurt’s example using a neurological device is similar to some of the cases described by John Martin Fischer. An early use of this example appears in “Responsibility and Control,” Journal of Philosophy 85 (January 1988): 24-40.


12. The most extensive deterministic defense of the success of Frankfurt’s counterexamples to PAP has been given in a number of places by John Martin Fischer. See Metaphysics of Free Will, (Cambridge, Mass.: Blackwell, 1994), chapter 7.


14. It would be fair to point out that in drawing this conclusion I have moved backward in the order of explanation. Agency is a vague concept, just as vague as responsibility. It does not help us understand what responsibility is to be told that it requires agency. In contrast, PAP at least has the virtue of being clear and reasonably precise. I agree with this point, but deny that it is an objection. I am not offering any part of an account of responsibility here. My point is that we should not worry excessively about the failure of PAP. Its failure permits us to look more deeply at the property PAP was aiming to elucidate.

15. Perhaps she does not fail the Necessity conditions since he requires that the method of belief formation must be kept constant. She might, therefore, satisfy the following condition: If the belief p has been true and she has used the same way of arriving at whether p, she would not have believed that p. Arguably, in the epistemic F case, if p has been true she would not have used the same way of arriving at whether p.

16. I thank Abrol Fairweather for this point.

17. In fact, she may not be reliable on some ways of constraining reliability, in which case our basic epistemic F case is a counterexample to such theories.

18. Of course, unless the benign manipulator is omniscient, he will not be able to fully accomplish this, but we need only assume that the benign manipulator has far greater knowledge than Jones.

19. Compare what we would say about the parallel moral case. If the benevolent manipulator makes Jones do only right acts, a right machine-produced act does not earn Jones any moral credit.

20. Linda Zagzebski, Virtues of the Mind, (Cambridge: Cambridge University Press, 1996), 286, and in the “Reply to Alston” in the Symposium on Virtues of the Mind, Philosophy and Phenomenological Research. Careful readers will no doubt notice that this is a counterfactual condition and perhaps it also can fail due to the action of a counterfactual manipulator.


23. This naturally raises the question of whether alternate possibilities are necessary at the second level. I will leave that question aside for this essay.
