

Seeing What You're Doing

John Gibbons

Do we have privileged access to what we're intentionally doing? Well, that might depend on how you think about privileged access. One way to think about privileged access is to try to identify a true formal principle. One thing you'll need to do when identifying the formal principle is to specify the relevant range of propositions to which you have privileged access. These ranges are usually specified by subject matter: propositions about your own current, conscious propositional attitudes, propositions about your own sensations, or perhaps, propositions about what you're currently, intentionally doing. In addition to specifying a range, you need to decide which way the arrow goes. Many formal principles are modeled on one of the following:

(BF) Nec. $(p \in R) (Bp \rightarrow p)$

(FB) Nec. $(p \in R) (p \rightarrow Bp)$

(BF), also known as infallibility, says that for every proposition in the range, belief is sufficient for truth, while (FB) reverses direction and says that the facts are sufficient for belief. There are many ways of modifying these two basic principles. In either of the principles, you can replace the notion of belief with being justified in believing or with being in a position to know. In (FB), you can add to the antecedent that the agent is attentive and has the relevant concepts, and so on for a wide variety of modifications.

If you do succeed in discovering a true formal principle of this sort, your work is far from done. Suppose it were true that you couldn't be wrong about any proposition about your own mind. We would need an explanation of this astonishing fact. The astonishing fact by itself couldn't explain anything. And when we go looking for an

explanation of the astonishing fact, the natural place to look is at the idea that the facts in virtue of which it's true that p are not independent of the facts in virtue of which you believe that p (or are justified in believing or a position to know or whatever). Giving an account of the way in which these sets of facts are not independent is the real work of giving an account of privileged access.

If one of these formal principles does turn out to be true, we have every right to be surprised. At least *almost* every such principle that anyone's ever come up with has turned out to be false. Those less timid than I am might be tempted to say that *every* such principle has turned out to be false. Clearly, the failure of the principles does not provide grounds for doubting that we have privileged access to our thoughts, actions, or what have you. At most it provides grounds for thinking that we have not yet specified what that privilege consists in. But it might not even provide grounds for that. If the real work is done by the explanation, why do we need the principles in the first place?

Here I think things get a little tricky. Of course, you can't have an explanation of a principle without a principle. But the question is whether you can have an explanation of privileged access without a principle, or at least, something like one. Maybe we should try to explain something like this. We have a way of knowing about these things whether we use that way of knowing on a particular occasion or not. And our way of knowing about these things is different from everyone else's way of knowing about them because our way of knowing about them depends on the fact that these things are ours. And the fact that these things are ours means that the facts are not independent, in the relevant way, of our take on the facts.

No explanation of this conception of privileged access could ever get off the ground without some specification of what *these things* are supposed to be. But giving some idea of what the relevant things are is not fundamentally different from specifying the relevant range of propositions in a formal principle. If the things are our intentional actions, then restrict the range of the variable “p” to propositions of the form, “I am currently A-ing.” And you can’t explain privileged access without some idea of the way in which the facts are not independent of our take on the facts. And figuring that out will involve, among other things, figuring out whether that connection is more like (BF) or more like (FB).

So you do need something like a formal principle. But it’s hard enough to give an account of what’s going on in those cases where things go smoothly without having to worry about every little thing that can go wrong. So it’s sometimes tempting to trade your formal principles in for some more informal cousins. Perhaps something like this.

(UBF) Usually, $(p \in R) (Bp \rightarrow p)$

(UFB) Usually, $(p \in R) (p \rightarrow Bp)$

Every modification of the formal principles will yield a corresponding informal principle. Unfortunately, the precise interpretation of these informal principles is not absolutely clear. But the basic motivation is relatively straightforward. We don’t want to worry about every possible counterexample to the corresponding formal principle. We just want to think about the normal case, or the usual case, or the cases where things go smoothly. If we could give a non-trivial specification of the normal case, we’d build that into the restriction on the quantifier and we’d be back to a formal principle. Relying on the informal principles or a shared understanding of the normal case does change the

rules of the game: not every counterexample counts. But the basic questions remain the same: what do we have privileged access to and what connection between the facts and our take on the facts explains it? But if this is the motivation, it's not clear that retreat to the informal principles is necessary.

This paper is structured around a formal principle about our knowledge of our own intentional actions, a principle I know to be false. Here it is.

(AK) If you're intentionally A-ing, you know that you're A-ing.

While I think that this is false, I think that it's onto something. I think that it's close to the truth and that the truth it's close to is important. Basically, I think that we have a way of knowing about our own intentional actions whether we use that way of knowing on a particular occasion or not. And our way of knowing about what we're doing is different from everyone else's way of knowing about what we're doing because our way of knowing depends on the fact that these actions are our own. And the fact that these actions are ours means that the acts are not independent of our take on the acts. This is what we need to explain in giving an account of our knowledge of our own actions.

If the point of the formal principle is to tell you what needs explaining, it looks like (AK) can do that even if it is false. We want our account of our knowledge of our own intentional actions to apply to everyday cases of making dinner, going to the bank, and writing philosophy papers. And it looks like the connection between our actions and our take on our actions is much more like (FB) than it is like (BF). Cases where you falsely believe that you're intentionally A-ing are relatively easy to come by. You think you're phoning home but you press seven where you should have pressed eight. But cases where you're intentionally A-ing without knowing it are more difficult to find.

The reason I'm not tempted by an informal cousin of (AK) is that the counterexamples to it that interest me (no doubt there are others) do not involve any sort of repression, subconscious desires, inattentiveness, conceptual confusion or anything like that. It's not clear that they involve any irrationality of any sort on the part of the agent. So I'm not sure that even our shared understanding of the normal case rules them out. So I'll stick with (AK). In the first part of the paper, after presenting some counterexamples to (AK), I'll consider and reject some suggestions about how we might restrict the range of (AK) to bring it even closer to the truth. This might look like a fight about which false formal principle we ought to like better. But the real issue here is about the range of our privileged access, and there are trade-offs to make. Usually the way it goes with these things is that the broader the range, the more you have access to, but that access is less secure.

In the second part, I'll try to sketch an outline of an account of how we know what we're doing. This will be a story about the particular way in which the facts are not independent of our take on the facts. In the third part of the paper, we return to the counterexamples to (AK), because I take them as evidence for my theory. The principle fails exactly where you would expect it to fail if the account of the second section is correct. The goal of all this is not to defend the use of formal principles in accounting for privileged access or to defend (AK) in particular. You do have to start somewhere, and perhaps formal principles are just as good as informal ones for this purpose. But the crucial question is not about the principles. It's about how things go when things go well.

Part I

1.1 Cataloguing the Counterexamples

Let's start with some counterexamples to (AK). This one comes from Davidson.

[I]n writing heavily on this page I may be intending to produce ten legible carbon copies. I do not know, or believe with any confidence, that I am succeeding. But if I am producing ten legible carbon copies, I am certainly doing it intentionally.¹

Davidson withholds judgment on the proposition that he is making ten copies, and he's perfectly justified in doing so. If this is a counterexample to (AK), it's an epistemically blameless counterexample. How does the example work? Well, the chances of success have to be sufficiently in doubt that withholding judgment is a reasonable thing to do. But if the chances of failure are sufficiently high, we'll think it's unreasonable for you to intend, and at least in many cases, even if you succeed, we won't think that you did so intentionally.

Suppose you buy a lottery ticket, and despite what any sane person would reasonably expect, you actually win. You tried to win the lottery. You did everything you could to win the lottery. And you won the lottery. But you didn't intentionally win the lottery. Winning the lottery wasn't up to you; it wasn't under your control; and it was just an accident that your attempt was successful. To the extent that we think that there's an intimate connection between intentional action and voluntary control, there will be a temptation to say that when you succeed against all odds, the success was just lucky, and so, unintentional.

But there's a great deal of room between being justified in believing you'll succeed and being justified in believing that you'll fail. This is the room we occupy when we reasonably withhold judgment, and this is the room Davidson's example exploits. Once you see how the example works, it's relatively easy to think of cases

¹ DD, "Intending" p. 92

where you're intentionally getting in touch with an old friend, intentionally buying something online, or, if your computer is anything like my computer, intentionally turning on your computer without knowing that you're doing any of these things. Let's call cases like this "carbon-copy cases."

A particularly clean version of a carbon-copy case would contain the following three features. First, there's enough reason to suspect an intervention that belief in success is not justified. Second, no intervention occurs so success is under your control and you do what you do intentionally. And third, success is not immediately apparent. It's hidden under nine other carbon copies or at the other end of the telephone or what have you.

Grounds for doubting success in carbon-copy cases are based on grounds for doubting that your attempts will be successful. They're not in any way based on doubts about what you're trying or intending to do. When Davidson presented the original carbon-copy case, he was primarily interested in arguing against the Strong Belief Requirement.

(SBR) Intending to A entails believing that you will A.

If you like (SBR), you might object to Davidson's description of the story. Maybe he's wrong about what he intended. Maybe he really only intended to try to make ten carbon copies. This way of saving (SBR) involves rejecting what Michael Bratman calls The Simple View:

(SV) Intentionally A-ing entails intending to A.

If Davidson didn't really intend to make ten carbon copies but he did intentionally make ten carbon copies, then (SV) is false. The literature on all this is fairly extensive and I'm

not going to try to tell you what to think. Here's what matters for our purposes. The conjunction of (SBR) and (SV) seems to entail the following.

(AB) Intentionally A-ing entails believing that you will A.

This connection between action and belief is necessary (though of course not sufficient) for (AK). If you reject either of the conjuncts in the face of carbon-copy cases, you should expect epistemically blameless counterexamples to (AK). Which conjunct you reject is entirely your business.

There's another kind of example, also directed originally against the strong belief requirement, that causes trouble for (AK). Here's Michael Bratman on his bicycle.

I might intend now to stop at the bookstore on the way home while knowing of my tendency toward absentmindedness – especially once I get on my bike and go into “automatic pilot.” If I were to reflect on the matter I would be agnostic about my stopping there, for I know I may well forget. It is not that I believe I will not stop; I just do not believe I will.²

The worry here is not that he might try to stop at the bookstore and fail. This is not a carbon-copy case. In these cases, which we might as well call “auto-pilot cases,” the worry is that he might not try to do what he knows he has most reason to do. And worrying about this is perfectly compatible with being quite sure that if he tries he will succeed.

There are many ways of losing control, and going into automatic pilot is simply the most benign. If you reasonably take seriously the possibility that you will be drunk, or depressed, or outraged, or simply inattentive or forgetful, you can sensibly withhold judgment on the proposition that you'll succeed, even while you firmly intend to let reason reign. Again, sufficient evidence that you'll fail can make the intention

² Bratman, *Intention, Plans, and Practical Reason* p. 37

unreasonable and keep you from acting intentionally. But at least sometimes grounds for doubt can spoil belief before they spoil intention.

Auto-pilot cases are most clearly at home when you have a present intention for future action. Using these to cause trouble for (AK), which concerns present belief about present action, is only a little bit tricky. Suppose Bratman does manage to stop at the bookstore on the way home. When he first starts off on his bicycle, worrying about going into auto pilot, he is intentionally going to the bookstore. Does he know that he is? Well, maybe he does. It can be true that you're going to the store at a certain time, even if you never make it to the store.³ But here's something else that he's doing intentionally: he's taking the first steps in a plan that will in fact eventually lead him to the bookstore. Being Michael Bratman, he knows all about plans, so this may well be a perfectly good description of what he's trying to do. Given that his forgetfulness does not intervene, his execution of his plan to get to the bookstore is under his control, and so, intentional. But given his genuine grounds for doubt, he's not justified in believing that this is what he's doing.

1.2 Avoiding The Overreaction

I've catalogued some of the counterexamples to (AK) because once you see how they work, you see how relatively easy it is to generate further counterexamples of the same sort. Given sufficiently elaborate mechanisms for lighting cigarettes, mechanisms that obstruct success from view, we can easily tell stories in which people are intentionally lighting cigarettes without knowing it. So there's no easy and obvious way of restricting the "A" in (AK). We can't say, for example, that for all actions A except

³ Anscombe, p. 39

those involving carbon paper or bicycles, if you're intentionally A-ing then you know that you are. The examples don't essentially depend on these sorts of things.

But there is a response to the counterexamples that might actually tempt someone. The Overreaction, as I call it, is to take the examples as showing that our own intentional actions do not form an epistemic natural kind. This is the rejection of the idea that we have privileged access to what we're doing, at least if that includes going to the store, making dinner, or lighting a cigarette. This is why I think it's important to think about the role formal principles play in our thinking about privileged access. If the only thing privileged access could be is something a couple of qualifications short of infallibility, then of course we don't have that. But even if we thought that some principle in the neighborhood of infallibility were true of some domain, we wouldn't think that that's what privileged access consists in. The privilege consists in the kind of non-independence between the facts and our take on the facts that explains the truth of the principle.

So the question is not whether ignorance or error is possible. The permanent possibility of ignorance or error about just about anything is an essential feature of the human condition. The question is whether there's a way of knowing about what you're doing that's available to you as agent, unavailable to anyone else, and which depends on the fact that the actions are yours. If that's the question, you should be moved by the following sort of exchange.

“What are you doing?”

“Making dinner.”

“How do you know? You’re cutting a couple of carrots. You could be doing anything. Maybe you’re checking to see how sharp your knife is. Or maybe you’re making lunch for tomorrow.”

“Oh my god, you’re right! Let me check to see how I sliced the carrots. Maybe that way I can figure out what I’m doing with them.”

Formal principles aside, it really is tempting to say that if you didn’t know what the carrots were for, you weren’t intentionally making dinner in the first place. We have a pretty good idea of how to account for most people’s knowledge of their own address. This is the kind of thing that people tend to be interested in, and it’s the kind of fact that tends to be readily available. An account of our knowledge of our own actions that’s modeled on an account of our knowledge of our own address is guaranteed to leave out something important about the relation between knowledge and action. If you find someone who really doesn’t know their own address, you’ve found someone who needs to go acquire further empirical information. If you find someone who really doesn’t know what they’re doing, you’ve probably found someone who isn’t really doing anything.

So we should be suspicious of the overreaction. When we look at the details, grounds for suspicion increase. One way to restrict the range of (AK) is to exclude not only actions that involve bicycles and carbon paper, but all actions that involve anything outside the body. On this view, we should restrict our attention to our privileged access to our bodily actions like raising your arm, wiggling your ears, and moving your legs in a walking like motion. Since walking requires having your feet on the ground, we could probably tell a carbon-copy style story about that. Simply block the ground from view,

block your tactile access to the ground with some anesthesia, and raise the grounds for doubt of success to the appropriate level.

But if this is why we're worried about walking, we should be equally worried about raising our arms, wiggling our ears, and moving our legs in a walking like motion. In the hastily sketched thought experiment that called walking into question, the grounds for doubt were grounds for doubting that your feet were on the ground. But if you can't feel your legs, we can just as easily call into question whether your legs are moving at all.

I readily admit that these sorts of cases with their laboratory settings, anesthesia, and so on are not what we had in mind by our shared understanding of the normal case. But if this is how you eliminate the counterexamples to moving your legs, it seems that you have thereby eliminated the counterexamples to walking. I'm not calling into question the idea that we do have privileged access to our bodily actions. I'm only calling into question the epistemic significance of the line around the body.

If you want an example that doesn't involve so much equipment, think about wiggling your ears. Some people can't wiggle their ears. Some people, I suppose, wiggle their ears on a regular basis. I only very rarely wiggle my ears. So when I do so, after not having done it in a while, I feel the need to go check in a mirror to see if that's what I'm really doing. In fact, I've just come back from looking in the mirror, and not surprisingly, that's exactly what I was doing despite the fact that I didn't know that I was.

So, of course, restricting attention to bodily actions is safer, and the number of counterexamples to your principle does diminish dramatically. But compared to the distinction between intentional and unintentional actions, the distinction between bodily actions and the rest is relatively unimportant from the point of view of privileged access.

When things go well, our access to the fact that we're walking does not seem significantly different from our access to the fact that we're moving our legs. And when things do not go well, there may be fewer counterexamples to (AK) restricted to bodily actions, but the examples are of the same kind as the ones that plagued (AK) in the original.

So much for retreating to bodily action. The more common retreat from the world leaves the body behind as well. While it seems that from the point of view of the person, parts of your body are parts of you, it can also seem that from the point of view of the mind, the body is just the closest part of the external world. If you think that privileged access is restricted to the mental, then on one very familiar conception of the mental, we should restrict privileged access to non-factive mental states and events, for our purposes to intentions and attempts.

Difficulties for this view will come from two directions: Does it explain enough? And does it explain too much? I think we should all agree that there is something like the appearance of privileged access to our ordinary intentional actions under their ordinary intentional description. If someone asks you how you know you're making dinner, you don't go looking at the carrots to find out what they're for. You go looking for an account of privileged access that accounts for the case. If we think that real privileged access is restricted to intentions or attempts, we need to explain at least the appearance of privileged access to our actions. Maybe we have very good access to our intentions and attempts and only pretty good access to our actions. I'll argue that the most natural explanation of this along these lines either doesn't explain enough or ends up explaining too much.

So according to what I'll call the two-factor theory, we have some kind of privileged access to our intentions and attempts. There's no need to make exaggerated claims about infallibility or indubitability here. Whatever access we have to our beliefs and desires, we have to our intentions and attempts. And we'll just let the explanation of this access wait for another day. There are all sorts of interesting and important differences between intentions and attempts, but I'm going to ignore them all and just focus on attempts. Feel free to rewrite as you read. So, according to the developing picture, while you're acting, you have very good privileged access to a proposition of the following form.

(T) I am currently trying to A.

You replace "A" with a description of the action under which it will be intentional if everything goes according to plan.

In addition to having very good privileged access to things like (T), we do quite often have ordinary empirical knowledge of the corresponding try conditionals.

(TC) If I'm trying to A, then I'm A-ing.

Believing (TC) is not a matter of believing that you can do anything: you replace "A" with a particular description of action. And it's not a matter of believing, e.g., that you can always make dinner. It's more like thinking that in the particular circumstances in which you find yourself, whether or not you make dinner is up to you. If you try here and now, you will succeed.

Now the most straightforward version of the two-factor theory would have it that you infer that you're A-ing from your belief in (T) plus your belief in (TC). But this is not the only option. The idea, as I understand it, is something like this. Whatever goes

on in the heat of the moment, we epistemologists in the calm of our studies can take the facts in virtue of which you know what you're doing and sort them into two categories. First, there's the familiar privileged access to your non-factive mental states that puts you in a position to know something like (T). And second, there's familiar empirical knowledge of the external world and your place in it that puts you in a position to know something like (TC). Maybe you don't even have to believe (TC). Maybe circumstances in which you're entitled to ignore the possibility that (TC) is false are circumstances in which you can move directly from something like (T) to the claim that you're A-ing.⁴

Though we'll have to hire someone to work out the details, the basic picture does seem to have it all. The epistemic building blocks do not involve any kind of unfamiliar privileged access to the external world. Nevertheless, the resulting knowledge does deliver a first person/third person asymmetry. Since your knowledge of what you're doing is based on, or anyway somehow connected to, your knowledge of what you're trying or intending to do, your privileged access to these latter facts will explain the asymmetry and give us at least pretty good access to our actions. But since our knowledge of what we're doing is also based on or connected to empirical knowledge of our external circumstances, all the familiar possibilities of ignorance and error are in play. According to the picture, some of our privileged access to our minds transfers to our knowledge of our actions, but not too much.

I know that we have some kind of privileged access to our intentions and attempts. But I don't think that the two-factor picture can be the whole story about our knowledge of our own actions. Briefly put, the worry is this. If some, but not too much, of our privileged access to our intentions and attempts transfers to our actions when we

⁴ This seems to be Falvey's position.

know the relevant try conditional, why doesn't an equal amount of our privileged access to our beliefs transfer to the world when we know the relevant belief conditional?

Somewhat less briefly, here it is again. It's part of the two-factor picture that whatever access we have to our intentions and attempts, we have to our beliefs and desires. So in an ordinary, everyday case in which you believe that *p*, you have very good privileged access to a proposition of the following form.

(B) I believe that *p*.

We can replace "p" with an ordinary proposition about the external world. Maybe it's the proposition you express when you tell someone your address. In addition to our very good access to things like (B), it seems that we can have ordinary empirical knowledge of the corresponding belief conditionals.

(BC) If I believe that *p*, then *p*.

We can certainly have ordinary empirical knowledge of belief conditionals applied to others. I don't know what her address is, but I'm sure she does. If she thinks it's *X*, then that's what it is. But if I can know this about her, what would keep me from knowing it about myself?

According to the two-factor theory, very good access to (T) plus ordinary access to (TC) adds up to pretty good access to what you're doing. By parity of reasoning, very good access to (B) plus ordinary access to (BC) should add up to pretty good access to your address. I have no problem with the idea that people typically know their own address; they know where the closest grocery store is; and they know the color of their car. But I do have a problem with modeling our knowledge of our own actions on our knowledge of our own address. If this is all there is to pretty good access to our own

actions, the two-factor theory doesn't explain enough. If pretty good access to our actions amounts to anything like what we think it amounts to, then it looks like the two-factor theory is saddled with this kind of access to the world, and it ends up explaining too much.

In addition to beliefs and belief conditionals, there are also appearances and seem conditionals. In ordinary circumstances, you have very good access to facts like this.

(S) It seems to me that p.

And we can have empirical knowledge of seem conditionals applied to others, so why not ourselves?

(SC) If it seems to me that p, then p.

Replace "p" with an empirically difficult proposition, and the resulting seem conditional feels risky. Replace it with an empirically obvious proposition, and it seems safe. In any case, special knowledge of (S) along with ordinary knowledge of (SC) does not add up to somewhat special knowledge of the visible world.

Perhaps the two-factor theorist can explain why some, but not too much, privileged access transfers in one of these cases but not the other two. I'm not trying to prove that no two-factor theory could ever work. After all, almost any theory according to which knowledge of action somehow involves knowledge of intention can be factored into two parts. There's the part where you know what you're trying to do, and then there's everything else. But I am suggesting that despite appearances, our knowledge of (T) and (TC) does not explain the difference between knowledge of action and knowledge of address. The real work remains to be done.

2.1 The Basic Idea

Explaining our privileged access to some domain is a matter of giving an account of the source of our knowledge, an account that shows how the facts are not independent of our take on the facts. There are many kinds of non-independence: logical or conceptual, normative, causal, constitutive, and no doubt others. Once we know what kind of non-independence is involved, we'll want to know whether the knowledge depends on the facts, the facts depend on the knowledge, or both depend on some third thing. And what we want to know is how things go when things go smoothly. So for this part of the paper, we forget about the carbon-copy and auto-pilot cases, and just focus on the easy cases. We assume that success is readily apparent in the way it quite often is. We assume both that you're intentionally A-ing and that you believe that you're A-ing. And we ask why that belief constitutes knowledge.

One way to think about the difference between knowledge of action and knowledge of address is to start by thinking about the patterns of reasoning that could plausibly lead to the relevant conclusions. Here you start out thinking about the difference between (T) and (TC) on the one hand and (B) and (BC) on the other. The problem with this is that you seem to commit yourself too early to a picture of the source of our knowledge of action.

Another way to think about the difference between knowledge of action and knowledge of address is to start by thinking about the difference between the facts known. You can live on Elm Street without knowing you live on Elm Street. But then, you can make ten carbon copies without knowing that you're doing so. But still, whether you live on Elm Street doesn't depend on anything you know. My son's address was the

same as mine as soon as he was born. He may well have been a genius at that time, but his address did not depend on it.

So forget for a moment about knowledge of action, and think about how much you have to know on any particular occasion in order to make dinner. When chopping, ordinary empirical knowledge of the carrots plays an indispensable role in enabling you to intentionally chop carrots in the first place. If you couldn't see them or feel them, you might still end up chopping your carrots by waving your knife around randomly. But you wouldn't chop them intentionally.

I'm not now asking whether any particular item of knowledge is necessary in every case of intentionally making dinner. Since you can have dinner without carrots, knowledge of carrots is not, generally speaking, a necessary condition for knowing that you're making dinner. So maybe you know different things on different nights that you make dinner. But every night that you make dinner, you know a great deal about what's going on in your kitchen. This is ordinary empirical knowledge about carrots, knives, their relations to each other, and their relations to you. We might even go so far as to call it observational knowledge. And this is the knowledge you use, on this particular occasion, to make dinner.

But how do we know that this ordinary empirical knowledge doesn't also let you know what you're doing? I don't think we should worry that if our knowledge of our own action has its source in perception then it won't explain privilege. If the knowledge in virtue of which you are able to intentionally A on this occasion is sufficient, or anyway close to sufficient, for you to know that you're A-ing, then your take on the facts will not be independent of the facts. They'll both depend on the same thing.

Since it's your knowledge of the carrots that plays the dual role of letting you chop them and letting you know that you're chopping them, there's no reason to think that this knowledge will be anywhere close to sufficient for me know what you're doing. So we have a first person-third person asymmetry that's independent of the asymmetry based on your knowledge of your own intentions.

This is the story I'd like to try out. Like the two-factor theory, my basic epistemic building blocks are relatively unproblematic. Unlike the two-factor theory, I don't think our privileged access to our actions is transferred from our privileged access to our intentions. I think the ordinary empirical knowledge plays an essential role in the explanation of the privilege. And unlike the two-factor theory, I think there's a fairly obvious explanation of the scope and limits of the privilege. The fact that you're intentionally making dinner is a knowledge-dependent fact. If you didn't know enough, the fact would not obtain. But the fact of your address is not knowledge-dependent. So there is no knowledge in virtue of which it obtains already there to let you know that it obtains.

So on my picture, ordinary empirical knowledge plays an essential role in the explanation of our privileged access. Doesn't that mean that we have ordinary empirical knowledge of what we're doing? And if ordinary empirical knowledge is observational knowledge, doesn't that mean that Anscombe was wrong when she said that we have non-observational knowledge of our own intentional actions? And if the picture is incompatible with Anscombe, isn't that reason enough to be suspicious of the picture?

I am generally tempted by the idea that being incompatible with Anscombe is a lot like being incompatible with the facts. So I'm worried about this even if you're not.

But I do think that our knowledge of what we're doing is empirical knowledge. I think there's a perfectly good sense in which we can usually see what we're doing. Seeing what you're doing in this sense doesn't contrast with doing it. It contrasts with doing it by feel because something's in your way. And I think that the knowledge acquired when you can see what you're doing is exactly the kind of knowledge that plays the dual role of making intentional action both possible and known. But I'm not sure that Anscombe was wrong about anything because I'm not at all sure that her notion of observational knowledge is the same as our notion of empirical knowledge.

Let's start with a case where Anscombe would say that we do know what we're doing on the basis of observation. You walk into a room, and the lights go on, but there's no one else in the room. You step out of the room, and the lights go off. In fact, you're turning the lights on and off. There's a sensor built into the door. But you didn't know you were doing this. And you weren't doing it intentionally. And these two facts are not completely unrelated. When you move into and out of the room, you do this intentionally. And, of course, you know that you're doing it. What is the crucial feature of this knowledge that distinguishes it from the knowledge you eventually acquire that you were turning on the lights?

Is it the fact that the knowledge is somehow non-empirical or somehow not derived from experience? Suppose that after you've entered the room you need to do some a priori theorizing from intuitively obvious first principles in order to figure out what you've done. This would not capture what's special about our knowledge of what we're doing. Suppose that as you're entering the room, you need to gaze into your mind, catalogue your various beliefs, desires, and intentions, and use your a priori knowledge of

the principles of practical reason to formulate a likely hypothesis about what you're up to. This non-empirical knowledge would not capture what's special about ordinary knowledge of what we're doing either.

When you're walking into and out of the room, you already know what you're doing. You don't need to gaze outward at the world in order to find out what you're doing, and you don't need to gaze inward at the mind or upward or wherever you gaze when doing your a priori theorizing. The crucial fact about knowledge of action in the easy cases is not its source. It's the immediacy. And by that I don't just mean that you can find out very quickly. I mean you already know. If this is at least part of what Anscombe has in mind in talking about non-observational knowledge of our own actions, I think my picture can capture this part.

Another thing Anscombe thinks you know without observation is the cause of certain involuntary bodily movements. Here's Anscombe's example. A crocodile leaps up and barks at you. This startles you causing you to involuntarily jump back and you automatically know why. Of course Anscombe thinks that you observe the crocodile and the crocodile's barking. But you don't observe the barking making you jump.⁵ I'm pretty sure this is not the Humean idea that you can never see anything cause anything. It's supposed to be restricted to a certain kind of first-person case. Maybe the idea is that you can perceive the cause, the crocodile bark, but you can't observe the effect.

I'm not at all sure that what's crucial for Anscombe in this section is something about the nature of observation.⁶ But if it is, we should be very leery about identifying her notion of non-observational knowledge with our notion of non-empirical knowledge.

⁵ P. 15

⁶ but things are complicated...

Maybe we have non-empirical knowledge of the particular causes of particular effects. But you should be careful about attributing that view to others.

Anscombe exegesis aside, here's one thing that I am tempted to say about the crocodile case. As you're jumping back, you already know what startled you. So there's a perfectly good sense in which you don't find out what startled you by observing your own reaction and then searching the world for a likely cause. This is something you might do if it were someone else's reaction. Though there is a kind of first person-third person asymmetry, I don't think our knowledge of what startles us is in any serious sense non-empirical. If you hadn't seen or heard or somehow observed the crocodile, you wouldn't have been startled in the first place. And if the empirical knowledge necessary on this occasion for being startled is sufficient on this occasion for knowing what startled you, then your knowledge of what startled you is immediate, and so in one sense non-observational, but nevertheless empirical.

I think that our knowledge of what we're currently perceiving provides the best analogy for our knowledge of our own intentional actions. How do you know that you see that p? Not on the basis of an inward glance. What you're looking at is out there in the world. And not on the basis of a second outward glance either. Generally speaking, once you see that p, no further glances, calculations, or inferences are necessary. The conditions in virtue of which you see that p in the first place are usually sufficient for you to see that you see.

The applicability of the account to the case of perception shows that the picture does not essentially rely on the idea that we are the authors of our own actions or that our actions in some important sense come from us. Whatever the important sense is in which

our actions come from us, it must be consistent with the fact that what we do is not completely independent of the layout of the world. If there is an important sense in which our actions come from us, we should say that in that sense our perceptions do not come from us. They come from the world, though, of course, this will have to be consistent with the fact that what we see is not completely independent of where we point our eyes and what we pay attention to. Any account of our knowledge of action that did essentially depend on the idea that our actions come from us would not extend to our knowledge of perception. This is precisely a respect in which action and perception differ.

Since I'm going to try to blur the line between action as active and perception as passive, I think that the extension to the case of perception is a benefit. At this level of abstraction, or metaphor, my picture is based on the following idea. Your intentional actions and your awareness of those actions are elements of the same perspective, or as I'd rather put it, they both happen in the same mind. This is a respect in which action resembles perception. In whatever sense intentional action is a mental event, perception is as well. Like many other cases of privileged access, it's the fact that the facts and the take on the facts happen in the same mind that's ultimately responsible for the relevant non-independence.

2.2 Seeing What You're Doing

So much for the very rough sketch of the basic idea. Now for the only slightly less rough sketch of how it's supposed to work. So the next time you're making dinner, I want you to watch what you're doing. You're chopping the carrots and looking at the carrots. Keep your eyes open; stop looking at the carrots; and start looking at your action

instead. If you're like me, nothing happens. Either you could see what you were doing all along and you can't see the chopping without seeing the carrots, or you can't see yourself chopping the carrots at all. No amount of refocusing or shifting attention will bring the action into view if it wasn't in view already.

According to Anscombe, observing your own actions is not necessary for knowing what you're doing. According to Brian O'Shaughnessy,⁷ observing your own intentional action while performing it is impossible. Whatever they have in mind in talking about observing your own action, it can't be this perfectly normal thing we call seeing what you're doing. The perfectly normal thing is what O'Shaughnessy is talking about when he says that perception is the handmaid to action. We put perception "to the services of action as an accomplice in actively helping to lay before the action its object and its field of action."⁸ You see the carrots and that helps you chop them.

The questions are whether this use of perception amounts to seeing what you're doing in a sense that involves perception (if not observation) of your own action, and whether this perception of your own action, if there is such a thing, can play a role in our ordinary knowledge of what we're doing. I don't know how to answer the first question in the negative. There's something very peculiar about the idea that our own actions are invisible to us. Other people can see you cutting the carrots. Why can't you? Is there something blocking your view, perhaps, the carrots? Other people can see both the carrots and you chopping them. Why do the carrots block your view and not theirs? At least in the easy cases, in order to know what you're doing, you don't need any more

⁷ O'Shaughnessy, "Observation and the Will"

⁸ p. 374

observations than you need to do it in the first place. Maybe this is because nothing else would count as observing yourself doing it.

It's unlikely but possible that the idea that you can't see yourself acting comes from a peculiar view about what you can see. Can you see that there's a judge on the bench, or that there's been a car accident, or that you're out of carrots? Someone might say no: in the strict and philosophical sense, all you really see are sense data or surfaces or whatever. But this robs the claim that we don't know what we're doing on the basis of observation of its philosophical significance. You can't see your own actions for the same reason that you can't see your own carrots: neither are sense data or surfaces or whatever. But we should put our own intentional actions on the first person side of some interesting first person-third person asymmetry. This puts them on the same side as the carrots.

Perhaps what motivates the idea that you can't observe yourself acting is not some restriction on the objects of perception, but the idea that observation is essentially passive. Sometimes it seems that this is one of the things that moves O'Shaughnessy. When you try to observe the action, it becomes just another happening in the external world and it loses its meaning, and so, ceases to be an intentional action.⁹ Unfortunately, O'Shaughnessy also likens trying to observe your own action to trying to do two completely independent things at once. If this is the difficulty, then it looks as though trying to observe your own action is one of the things that you're doing, and so, it begins to look active.

If the problem of observing your own action is like the problem of doing two things at once, it clearly doesn't apply to the ordinary thing I'm calling seeing what

⁹ p. 378

you're doing. There's no problem doing two things at once if the two things are holding a knife and chopping the carrots. These are not two completely independent things. But seeing what you're doing in the ordinary sense is no more independent of and no less necessary for chopping the carrots than is holding the knife.

So if there is a problem with observing our own actions, it must have something to do with the passivity of observation. "We could say that observation departs from the generally and typically true maxim: To understand the world it is necessary to change it. (I have in mind bird watching.)"¹⁰ Perhaps bird watching is an activity, so you're active with respect to the observing. But you're not active with respect to the birds. If this is a general feature of observation, maybe we can explain why we can't observe our own actions. Maybe, if you observe x then you are passive with respect to x. If this were true, it would be easy to see why x can't be one of your own intentional actions.

Unfortunately, this simply cannot be it. The job of perception in her role as handmaid to action is to lay the carrots before you, even as you chop them. Here you're observing the carrots even though you're active with respect to them. If you can both observe and be active with respect to the carrots, why can't you both observe and be active with respect to the chopping of them? If we build passivity into the meaning of "observe," then why can't you both perceive and be active with respect to the chopping? Maybe the basic idea here really is just that in order to know what you're doing, you don't need any more observations than you need to do it in the first place. If that's the idea, then I agree.

While I am willing to use O'Shaughnessy's expression "handmaid to action" in order to pick out the kind of perceptual process I have in mind, I don't really like the

¹⁰ p. 376

description because I don't think it accurately portrays the relation between perception and action in the usual case. I think that relation is much more like the relation between holding the knife and chopping the carrots than the description suggests. This is fairly obvious in at least some cases. Looking for something, e.g., rooting through the secondary silverware drawer for a pickle fork, involves both perception and action in a way that it just wouldn't be looking for a pickle fork if you didn't have them both. Here's it's not too disturbing to think of the perceptions (perceivings? the looking??) as part of the action itself along with the movements you make.

I think this is the general case and not just something that holds when we use words like "looking" in the description of the action. What's the difference between finding your way to the bank and going to the bank in those ordinary situations in which you pick up and use ordinary information about your location relative to the bank in the course of going there? "I found my way to the bank" certainly suggests or implies that there was some difficulty. But "I went to the bank but didn't find my way there" suggests or implies that someone else took you.

If "finding your way" and "going to the bank" are two descriptions of the same action, we should give up the idea of perception as handmaid to action. Call it "finding your way" and the action itself involves picking up information about the world on the basis of experience. But calling the same thing "going to the bank" cannot demote the process of acquiring information to a mere passivity distinct from and subservient to the real action.

When you're chopping carrots, you're looking over here and paying attention to that. This active interest in what's going on around you plays an important role in getting

your carrots and not your fingers chopped. I say we should think of all this apparently mental activity as part of the process of chopping the carrots. And I also say that your perception of the carrots, the knife, and your fingers amounts to or results in your seeing that you're cutting carrots in exactly the same way that your seeing a judge on the bench, in ordinary circumstances where everything goes smoothly, amounts to or results in your seeing that the judge is on the bench.

So you can see that you're chopping the carrots by seeing the carrots. But can you see that you're making dinner if dinner isn't ready yet? I don't see why not. You can see that there's been a car accident even if you don't see the accident occur. And you can see that a rock is about to fall on someone's head even if you immediately close your eyes because you can't bear to watch. These cases don't seem fundamentally different from the case in which you see that the judge is on the bench without being able to see the judge's credentials.

I'm not going to rely on any particular story about how seeing that works. No doubt, your background knowledge of what judges look like plays an important role in your seeing that the judge is on the bench. And no doubt, much more background knowledge than just this is involved on any particular occasion. But I don't think we should think of the role of all that background knowledge as premises in an argument whose conclusion is that the judge is on the bench. It can seem to you that there's a judge on the bench even if you know it's just an actor. Your background knowledge is playing an important role here as well. If you didn't know what judges look like, it wouldn't even seem that way to you.

Your background knowledge of what judges look like seems to be playing the same role in the case where you know it's an actor and the case where you know it's a judge. In both cases, by the time she gets to you, she already looks like a judge. But if the background knowledge is playing the same role, and if, in the case of the actor, there is no inference to the conclusion that it's a judge, then even in the case where you see that it is a judge, the background knowledge does not play the role of premise in an argument.

I don't know that much about the role of background knowledge in coming to see that p. I know what it's not, and I think I know it when I see it. And I think that things other than knowledge can play that role. Our expectations can influence how things seem, even if those expectations are unfounded. And we all know that what we want can influence how things look, even if this isn't always a good thing from the purely epistemic point of view.

But if what we want, and not just our knowledge of what we want, can influence how things look, then presumably, our intentions, and not just our knowledge of our own intentions, can influence how things look as well. And at least in the case of our own actions, this influence may well be a very good thing from the purely epistemic point of view. So the plan is in motion. You'll finish chopping the carrots, move on to the rest of the vegetables, and then crush the garlic while you let the onions sweat. You see the flurry of activity against the background of your knowledge of what carrots and onions look like. And of course you see it against the background of your knowledge of what happens to fingers when they meet up with sharp knives. But you also see it against the background of the very plan that's in motion. So it looks to you as much like making dinner as it does like chopping carrots. So seeing that that's what you're doing is

certainly no more difficult than seeing that there's a judge on the bench. And if it's your plan and your action, then seeing that that's what you're doing is hardly any more difficult than doing it in the first place.

Part III

3.1 What Do You Have To Know?

In the easy cases, the story goes like this. In order to make dinner in the first place, you need to be able to see the carrots and onions. Once you've got this, plus what you've already got, namely, ordinary empirical knowledge plus a plan in motion, you can see not only that you're chopping the carrots but that you're making dinner. Since the knowledge in virtue of which you're making dinner is enough to let you know that you're doing it, no further observations are necessary. And once you're cooking, you already know that you are. Since the facts and the take on the facts are both made possible by the same underlying knowledge, we have a kind of non-independence that explains privilege.

Of course, you can't always see what you're doing. If you're tightening a bolt underneath a pipe, the pipe might get in the way so that you have to do it by feel. But the question about the relation between observational knowledge and knowledge of your own actions is not in any way restricted to the sense of sight. Any information you pick up from any of the senses, including your awareness of your own body through proprioception and kinesthesia, can play the dual role of making action both possible and known.

So far, our picture is this. On Monday, you know what you're doing on the basis of your knowledge of carrots and onions. On Tuesday, it's your knowledge of pasta and pesto. If, generally speaking, the idea is that the knowledge necessary for you to

intentionally A in the first place is sufficient, or anyway, close to sufficient, for you to know what you're doing, then our next question must be this. What, generally speaking, do you have to know in order to intentionally A?

I've argued elsewhere that the means-ends beliefs in use on a particular occasion must constitute knowledge in order for your achievement of that end to be intentional. Suppose you think that doing M will get you E. If that belief is false, then you won't get E, at least, not by M-ing. And if it's just an accident that your belief is true, and you act on that belief, then it will just be an accident that you get what you want, and your getting it will not be intentional under that description.

Some of the evidence for this comes from the following three lottery stories. The first we've already seen. You buy a ticket in a fair lottery and you win. Here the winning is too lucky, accidental, or out of your control to count as intentional. In the second case, you're justified in believing that the lottery is rigged in your favor, but you're mistaken about this. You buy a ticket, and just by chance, you happen to win. Here you have a justified, true belief that if you buy a ticket you will win, but this belief does not constitute knowledge. And of course, your winning here is just as lucky, accidental, out of your control, and unintentional as it was in the first case. In the third case, you know the lottery is rigged. You know that if you buy a ticket you will win. And acting on this knowledge, you buy a ticket and intentionally win the lottery.

So I think the question of what you have to know in order to intentionally A is pretty much the same as the question of what you have to believe in order to intentionally A. And here, I'm not sure there is much of an easy answer. If intentionally A-ing required believing that you will A, then that's what you'd have to know. Unfortunately,

as the carbon-copy and auto-pilot cases show, nothing this tidy seems to be true. In the carbon-copy case, you do believe that you have a pretty good chance of making ten carbon copies by pressing very hard. But if that's what you have to believe, then that's what you have to know.

Suppose that your grounds for thinking that you have a pretty good chance of making ten copies, namely, your past experience with carbon paper, are sufficiently independent of the facts in virtue of which it's true that you have a pretty good chance of making ten copies. Maybe on this occasion, you have a pretty good chance not because the carbon paper is functioning normally, but because the Martians intervene or what have you. If the facts that ground the belief are sufficiently independent of the facts that make the belief true, then we have a Gettier example, and your justified true belief is not knowledge. But if what makes it true that you have a pretty good chance of making ten copies is not normal functioning but Martian intervention, then we also have a deviant causal chain case. This is a case where you try to A; this causes you to A; but you don't intentionally A because of a non-standard connection between attempt and success. Needless to say, an exact specification of the distinction between deviant and non-deviant causal chains is very difficult to come by. But it's safe to assume that Martian intervention will fall squarely on the deviant side of the divide.

We might give up on the search for the one thing you always have to believe whenever you act intentionally and look instead for what you can't believe. Maybe, if you intentionally A, you can't believe that you won't A. While this, or perhaps a normative version of this (you shouldn't believe that you won't) is probably a good restriction on intending to A, it doesn't seem to work for intentionally A-ing. Here's

another example from the battle over the simple view.¹¹ Suppose I build a fence and it's important to us that it not be easily knocked over. Quite confident in my fence-building skills, I think that the best way to convince you that it can't be pushed over is to try my best to do so. I don't try to push it over thinking that I will. I try to push it over thinking that I won't. But when I do push it over, and no Martians intervene, I do so intentionally.

This looks like a case where I intentionally A without intending to A, so it's supposed to cause trouble for the simple view. For our purposes, what matters is that it drives a pretty dramatic wedge between intentional action and belief in success. Here we seem to have intentional action with belief, maybe even justified belief, in failure. If we're really desperate, we could always try this. If you intentionally A, you might believe that you'll fail, but you can't know that you'll fail. Since this follows from the factivity of knowledge, we ought to believe it. But we'd have to be pretty desperate to come out and say it.

So I wish there were something tidier I could tell you about what you have to know in order to intentionally A. But it's not as though there's some tidy truth about belief and things only get messy when we start talking about knowledge. All the stories that make the mess in the first place were originally about belief. Perhaps the tidiest, most general thing we can say is something like this. Whatever means-ends beliefs are in use on any particular occasion must constitute knowledge in order to intentionally achieve the end.

I take it that the point of acquiring all that ordinary empirical information about carrots, knives, their relation to each other, and their relation to us is to update our means-ends beliefs about what we have to do next. Perhaps for basic actions, whichever those

¹¹ Mele?

turn out to be, beliefs about how to perform them are not required. In this case, knowing how is sufficient. But everywhere else, knowing how consists in part of knowing that, the knowledge of the means-ends beliefs, plus of course, knowing how to take the means. So the ordinary empirical knowledge that both lets you know what you're doing and makes you able to do it in the first place, makes you able to do it by letting you know how to do it.

There's one further thing about the messiness of the relation between intentional action and belief about success. Any account of privileged access in general or of our knowledge of our own actions in particular, must not only give an account of how things go when things go smoothly. It should also have a story, or at least leave room for a story, about the kinds of things that can go wrong. So suppose for a moment that my story is true. What makes you know how to do it lets you know that you're doing it. When would we expect (AK) to fail? When would we expect to find people intentionally A-ing without knowing that they were doing so? We would expect to find it in exactly those cases where the knowledge in virtue of which they're able to A is not sufficient for them to know that they're A-ing. And this is exactly what we find in carbon-copy and auto-pilot cases.