

Knowledge by intention? On the possibility of agent's knowledge

Anne Newstead

University of New South Wales

1. Introduction

Action theorists are typically interested in those actions done with, or in some cases preceded by, a certain kind of knowledge. Let's call such knowledge "agent's knowledge" to signify that it is an agent's knowledge of her own action.¹ Famously, in *Intention* Anscombe claimed that

(AK) Agents have non-observational knowledge of their own intentional actions.²

How is (AK) to be understood? Consider a simple example of a toddler drawing a figure on paper. The toddler may answer the question "What are you doing?" by explaining, "I am drawing a house." Presumably, the toddler knows what she is doing not by looking at the figure on the paper, which may indeed barely resemble a house. Rather, she knows what she is doing in virtue of knowing her intention and trusting that it is being implemented. Though there may be other possible descriptions of her action, she counts as knowing what she is doing so long as she is aware of her action under some description that fits it.

Is (AK) true? How far can we go in defending (AK)? Many theorists have thought that (AK) is obviously false or else in need of some important restrictions. Searle (1983: 90)

agrees that "at any given point in a man's conscious life he knows without observation the answer to the question 'what are you now doing?'" However, he allows that knowing the answer to the question "what are you now doing?" may consist simply in knowing without observation what one is *trying* to do. To see what one is *actually* doing, however, requires observation in Searle's view. Unless we identify tryings with actions, this falls short of the thesis that agents know their intentional actions non-observationally.

More radical opposition to (AK) stems from opposition to the idea of privileged self-knowledge generally. Gopnik (1993) follows Ryle in arguing that the idea that subjects have privileged knowledge of their intentional states is an illusion. Theorists influenced by this line of thought will claim that agents can only know about their intentional actions in the same way as everyone else: by observation. In order to account for the qualitative difference between an agent's knowledge of her action and observer's, we can say that the agent has more data (as Gopnik does) or perhaps different data -- such as sensory feedback from one's own body (as Pickard 2004 does) -- than observers do.

Despite the objections, there is a stubborn intuition that counts in favour of (AK). The idea is that so long as one is acting intentionally, one will (potentially) have access to one's intention and interpret one's action in the light of that intention. To view an action as a *mere* observer would be to view it blankly, without awareness of the intention it fulfils. It would then become questionable to what extent the movement constituted one's own intentional action at all. Movements that one undertakes but cannot interpret as fulfilling some intention of one's own are not one's intentional actions, but mere *happenings*. So it does seem that

agents must be able to become aware of their own intentional actions in a way that differs from being *mere* observers.³ Obviously, a defence of (AK) requires more than an appeal to intuition.

Imagine someone who accepts that agents apparently have some non-observational awareness of their intentional actions, but disputes that this awareness amounts to knowledge. This is the kind of sceptic whose position I address in this essay. Whereas the debate in action theory has by and large focused on the issue of whether an agent's knowledge of her action is based on *observation* or not, I will focus on the issue of whether we should agents have *knowledge* of their actions at all in a non-observational way. To see just how far the defence of (AK) can go, I will construct two models of agent's knowledge: a causal-reliabilist model (presented in §2) and a teleological, reasons-based model (presented in §3). Both are undeniably cases where an agent is apparently aware of her action without observation. My contention will be that these cases are also -- in the right circumstances -- cases of knowledge. It turns out that progress in epistemology -- in the form of an appreciation of the fallibilist response to scepticism -- allows us to defend (AK) from objections.⁴

2.1. A causal approach

Can the causal theory of knowledge be extended to knowledge of action? In the case of action, the object of knowledge is that one is ϕ -ing (in the present case) or that one is going to ϕ (in the future case), where " ϕ " represents some suitably basic description of one's action.⁵ In the present case, the causal approach yields:

(CTAK-present) The causal theory of agent's knowledge for the present case. S knows that she is ϕ -ing just in case: i) S is ϕ -ing; ii) S believes she is ϕ -ing; iii) S's belief that she is ϕ -ing is caused in the right way by the fact that she is ϕ -ing.

What counts as the right way for S to acquire her belief that she is ϕ -ing? S's belief that she is ϕ -ing must be "directly" (non-deviantly) caused by her action of ϕ -ing. Presumably, if as on (CTAK-present), the action of ϕ -ing is the direct cause of the subject's belief that she is ϕ -ing, then the agent's belief and knowledge comes *after* her action.⁶ But the kinds of actions we are interested in are ones in which such knowledge comes *before* completion of the action. In Anscombe's favourite example, someone knows what she is doing (e.g. writing a certain word) *before* she has completed the action. Furthermore, the delay between performance of an action and the receipt of sensory feedback from the action of some 50-250 milliseconds makes it unlikely that such knowledge of the action is derived from feedback from the actual performance of the action.⁷ This suggests we amend our theory as follows:

(CTAK-present)*. S knows that she is ϕ -ing just in case: (i) S is ϕ -ing; (ii) S believes she is ϕ -ing; (iii*) S's belief that she is ϕ -ing is *caused by her intention* of ϕ -ing.

This approach will restore the anticipatory character of agent's knowledge provided we can assume that the agent's action is caused by a prior intention to perform that action. However, in some cases, the intention of ϕ -ing can merely be embodied in the action of ϕ -ing without being preceded by a conscious prior intention to ϕ .⁸ Nonetheless, we do think that an agent knows what she is doing in such a case. Where there is apparently no (conscious) prior intention to ϕ , the agent's knowledge that she is ϕ -ing will be roughly contemporaneous with her performance of the action.

Our formulation of (CTAK*-present) should leave it open whether the agent's knowledge reflects just an intention-in-action or also a prior intention. In essence, the picture (CTAK*-present) presents is:

(Figure 1) *The causal theory of agent's knowledge, present case.* (The arrows indicate direct causal links.)

The optional part of the diagram is the agent's prior intention to ϕ and the agent's anticipatory knowledge. In those cases where the prior intention is present, the agent will also have some anticipatory warning that she is going to ϕ before she actually does. Nonetheless, one might say her knowledge that she *is* ϕ -ing is still derived from her intention-in-action, because only once the intention is embodied in action can she know she is performing the intentional action.⁹

In the future case, matters are different. Knowing what one is going to do in the future requires having a prior intention to ϕ before one actually does ϕ , where this prior intention is an efficient cause of the action of ϕ -ing. So in the future case:

(CTAK-future)*. S knows (at time t) that she is going to ϕ just in case: (i) S is in fact going to ϕ at some later time t^* (later than t); (ii) S believes (at t) she is going to ϕ ; (iii*) S's belief that she is going to ϕ is caused by her intention to ϕ at some point in the future.

The future case raises deep issues concerning the philosophy of time, specifically whether there can be a *fact* of the matter as to what someone is going to do if they have not yet done it. Proponents of an "open future" view will reject the assumption. But this leads to the counterintuitive consequence that one can never know what one is going to do until one actually starts doing it. So suppose there can be facts about what one is going to do. Even so

S's knowledge of what she is going to do looks *incurrable*, since if she does not end up ϕ -ing in the future, she can say that was nevertheless *going* to ϕ until she changed her mind, or something intervened. This contrasts with knowledge of what one is doing, which can be undermined by pointing out that, although the agent believes she is ϕ -ing and intends to be ϕ -ing, she is not actually ϕ -ing.

In the future case, the intention to act precedes both the knowledge and the action it describes, leading to the following picture:

(Figure 2) *The causal theory of agent's knowledge (future case)*

The crucial element common to both (CTAK*-present) and (CTAK*-future) is that they place an intention as the common cause of both the agent's knowledge and her action. In the present case, the intention-in-action causes both the knowledge and the action. In the future case, the prior intention causes both the foreknowledge and the action. This way of viewing matters gives a new sense to Anscombe's claim (2000: 87) that an agent's knowledge of what she is

doing, now understood as knowledge of intention, is "the cause of what it understands". Does the amendment also restore the anticipatory character claimed for agent's knowledge? As we saw, it preserves the sense in which agent's knowledge is anticipatory only in those cases where the intentional action is preceded by a conscious prior intention.

2.2. Objections to the causal approach

The causal theory makes several important assumptions that will strike some epistemologists as objectionable. First, the causal theory as stated is unabashedly externalist. Second, the theory requires a belief in the reality of mental causation. Third, the causal theory is in no way sceptic-proof. Let me consider these objections and reply to them in turn.

The causal theory is externalist. It allows that an agent can have a true belief that she is ϕ -ing, but it says nothing about the agent's *reasons* for this belief. Instead the causal theory concentrates on the aetiology of the belief: it is somehow caused by the agent's having the intention to ϕ .¹⁰ But it does not require that the agent is conscious of an intention to ϕ as a reason for believing that she is ϕ -ing (or is going to ϕ shortly). Therefore the causal theory allows that the agent may lack an understanding of why her belief is true. For epistemologists of an internalist bent, this lack of understanding on the subject's part may be enough to disqualify the subject from having knowledge.

Though the causal view presented here is externalist, there is no reason why an internalist cannot add requirements to the causal theory. For the present case, an internalist could simply add the requirement that S must be aware, or capable of becoming aware, that

she has an intention-in-action of ϕ -ing. Awareness of the intention does provide her with some evidence that she is ϕ -ing, since *ceteris paribus* she would not have that impression if she were not actually ϕ -ing. However, in order to really appreciate the evidence, she would have to reflect that intentions-in-action of ϕ -ing are usually accompanied by actions of ϕ -ing. This level of reflection might only be appropriate if the agent wishes to achieve a theoretical understanding of her knowledge. For the purpose of just having a rational belief that one is ϕ -ing, awareness of the intention of ϕ -ing may be enough.

For the future case, an internalist could require that the subject be aware, or capable of becoming aware, that whenever she has had intentions of the ϕ -type in type of context C, those intentions generally result in successful actions of ϕ -ing in C. Her past success in implementing ϕ -intentions provides a kind of *general* evidence that she will ϕ in a similar situation in the future. What she lacks, but does not need, is *specific* evidence that she is now ϕ -ing, which could indeed be provided by observation or perceptual feedback from her action.¹¹ So the account can be fleshed out so as to be made acceptable to those who advocate an internalist, evidentialist view.

It is worth pointing out, too, that the causal theory even in its externalist form still imposes a considerable warrant requirement on agent's knowledge.¹² It thus differs from the radical view that one can have weak knowledge that one is going to ϕ (or that one is ϕ -ing) merely in virtue of having a true belief.¹³ What seems wrong with such a view is that there is

now nothing relevant to the agent's knowledge -- neither evidence available to the agent nor some external factor in the agent's situation -- to raise the likelihood that the agent's belief is true. In such a case, we have a hard time differentiating the agent's true belief from one that is merely luckily true. And most epistemologists -- with some notable exceptions -- think that one cannot have knowledge simply because one is lucky enough to acquire a true belief.¹⁴

Our second objection notes that it is clearly an assumption of the causal theory that mental states such as intentions are causally efficacious. The theory therefore requires a belief in the reality of mental causation. Philosophers who suspect that mental states are epiphenomenal will conclude that the necessary conditions for possessing agent's knowledge according to the causal theory are never in fact met.

Such scepticism about mental causation may take the form of scepticism about the role of conscious intention in causing action. The work of Libet (1985) is sometimes cited in this connection. In Libet's experiment, subjects were told to move whenever they felt the urge to so after the experiment started. They were told to note the time when they felt an "urge", intention, or will (W) to move by noting the position of a moving spot on a clock dial. Their brains' physical preparation for movement was measured using a readiness potential (RP). The time at which they actually moved was recorded using an electromyogram (EMG) designed to measure the activation of a particular muscle. The time of actual movement provided a zero reference point. Correlating the measurements, Libet claimed that whereas W preceded the time of actual movement by about 200 ms, the RPs preceded the time of actual movement by much more, by some 300-500 ms.

One of the conclusions that a sceptic about causation by intention might draw from Libet's experiments is that they demonstrate that conscious, introspectively accessible, intentions to act cannot be what causally determine agents to perform voluntary actions. The unconscious cerebral processes detected by RP must be the true causes of the action, because they occur before conscious awareness of the intention.

Such a conclusion can be challenged on several grounds. First, it is not clear that Libet's measurement procedure should be trusted. It is possible that making subjects attend to the position of a dot -- giving them an additional mental task beyond simply attending to their intention -- has the effect of increasing cognitive load and delaying when subjects can note their conscious intention. Second, it is not obvious that the sorts of movements studied by Libet -- movements that one performs whenever one feels like it -- are comparable to the kinds of actions for which agent's knowledge should be claimed. Perhaps proponents of agent's knowledge should claim that such knowledge only applies to deliberate actions preceded by conscious prior intentions. This would be a significant concession to Libet. A less concessive response would simply deny that Libet has measured intentional actions at all. Perhaps the instructions to "move whenever one feels like it" do not produce intentional actions, but mere sporadic, random movements.¹⁵

The third objection is that the causal theory is in no way sceptic-proof. The causal theory requires that an agent actually be ϕ -ing at some point in order to know that she is ϕ -ing. But the agent will not be in a position herself to know for certain that she is ϕ -ing. It is

always possible for someone to have the intention of ϕ -ing, form the belief that she is ϕ -ing, and yet fail to actually be performing the action in question.

This does not seem to be a decisive objection to the theory, any more than it is a decisive objection to a causal theory of perceptual knowledge. Any theory of knowledge that incorporates a robust, objective conception of truth and that allows that subjects do not have total control over their beliefs and experiences is going to have to acknowledge the possibility that knowledge may elude subjects. That is, if the truthmakers for beliefs are beyond the control of the subject's mind, then in aiming to have true beliefs the subject can simply miss the target. In the case of agent's knowledge, the truthmaker for the beliefs about one's action is the action itself. Although the agent exerts some control over whether or not she manages to perform the action, and her mental state (of intending to act) is usually causally efficacious, agents do not always have total control over their actions. Indeed, these considerations suggest that, on the causal approach, the objects of agent's knowledge should plausibly be restricted to those actions over which agents have control -- so-called "basic actions". I will return to the issue of sceptical objections to agent's knowledge in §4 below, after considering one more model of agent's knowledge.

3.1. The reasons-based model

I turn now to a different model of agent's knowledge, one based on knowledge of practical reasons for acting. The model is much more closely suggested by Anscombe's remarks in *Intention* than the efficient causal model.¹⁶ It is based on a teleological model of action

explanation. The causal model portrays intentional actions as the effects of certain efficient psychological causes. The teleological model sees actions as performed for the sake of achieving a certain purpose, end, or goal. The purpose of the action, if you like, exerts an attraction on the agent who performs it.

On Anscombe's view, intentional actions are those that are done for some reason. They are actions for which the question "Why?" has application (Anscombe 2000: 9). Such actions are rationally intelligible to the agent as moving her towards her goals. Anscombe's special term for an agent's knowledge of her action is "practical knowledge". And, Anscombe says, we cannot understand the nature of practical knowledge without knowing about practical reasoning (*ibid.*: 57).

Practical reasoning consists in means-end calculation about what to do in order to achieve a goal. Thus, for example, someone may have the overall goal of building a sand castle. In order to build the castle, some intermediate steps are necessary: one has to pile up sand, shape it into towers, connect the towers with walls, and so on. Let the general action aimed at be "A" and each of the intermediate steps needed to reach "A" be "X, Y, Z". Then if someone asks the agent "Why are you Xing?", she can answer "In order to A". The sense in which she knows her reason for acting is just that she knows her goal is to A and understands how all of her sub-actions X, Y, and Z are directed towards the goal of accomplishing A. The explanation is easily adapted to the future case. In the future case, someone may ask "Why are you going to X?", and the answer is as before "In order to A".

According to the reasons-based model, an agent's knowledge of what she is doing is bound up with her "practical knowledge", that is, her knowledge of her practical reasoning. The idea is that having practical knowledge of a certain sort is *constitutive* of performing that very intentional action. Without the practical knowledge, one would not be acting in the light of certain reasons. One's action would not actually have the same intentional, purposive character without being backed up by such practical reasoning. So practical knowledge determines *what* it is that one is doing. This suggests that practical knowledge exerts a backwards pull. Whereas on the causal model, an intention to ϕ gave rise to knowledge of ϕ -ing, on the teleological model, the knowledge that one is aiming at ϕ -ing makes it the case that one is intentionally ϕ -ing.

It is in light of this reasons-based model that we should understand one of Anscombe's most important passages (2000: 87) on agent's knowledge:

Surprising as it may seem the failure to execute intentions is necessarily the rare exception ... What is necessarily the rare exception is for a man's performance in its more immediate descriptions not to be what he supposes. Further it is the agent's knowledge of what he is doing that gives the descriptions under which what is going on is the execution of an intention.

This passage would be badly misunderstood if it were read simply as supporting the idea that there is an efficient causal, reliable link between an intention to ϕ and ϕ -ing. Rather, what makes it the case that an action is intentional under the description ϕ -ing (rather than under some other description) is the agent's knowledge that she is ϕ -ing. So an agent's knowledge of what she is doing is *constitutive* of the very identity of her intentional action.¹⁷

How can this consideration about intentional actions yield knowledge? Anscombe writes (*ibid.*):

If we put these considerations together, we can say that where (a) the description of an event is of a type to be formally the description of an executed intention (b) the event is actually the execution of an intention ... then the account given by Aquinas of the nature of practical knowledge holds: Practical knowledge is 'the cause of what it understands', unlike speculative knowledge, which is 'derived from the objects known'.

There is much compressed in this passage. In the first part, Anscombe lays out the requirements for an agent to know what she is doing. For example, in order for an agent to know that she is ϕ -ing, it must be the case that

a) the agent takes herself to be ϕ -ing (where her intention is to be ϕ -ing); and b) the agent's action does fall under the description of ϕ -ing.

In the second part, Anscombe claims that when these conditions hold, we should think of the agent's knowledge of what she is doing as the cause (in a sense to be explained) of the agent's intentional action itself. The kind of cause that would be appropriate in this context is not an efficient cause, but something like a formal and final cause.¹⁸ The knowledge that one is ϕ -ing gives shape and form to the intentional action of ϕ -ing. In the final part, Anscombe states that an agent's knowledge of her action, *qua* practical knowledge, differs from observational, speculative knowledge in not being derived from the objects known. That is, agent's knowledge is not based on detecting the action performed and then inferring one is performing the action. Rather, such practical knowledge is spontaneously generated by the agent herself and makes the intentional action what it is.

It's clear from Anscombe's remarks here and elsewhere (2000: sec. 32), that she intends the difference between practical knowledge and speculative knowledge to be very radical. She notoriously claims that practical knowledge differs from speculative knowledge in having a reverse direction of fit: whereas in speculative knowledge, the belief is true if it matches the world, in practical knowledge, the belief is "practically true" just in case the world matches the belief.¹⁹ Many theorists find this part of Anscombe's theory obscure or downright false. I think such theorists would be right to do so, on the grounds that all

knowledge involves the concept of truth, where propositions are true just in case they fit with how the world is. The fact that a particular proposition concerns an agent's action does not exempt it from this basic characteristic of true propositions.

Even if we reject Anscombe's idiosyncratic conception of practical knowledge, though, we can appreciate the idea that an agent's knowledge could differ from observational knowledge in being based on knowing one's practical reasoning and plans for action. So understood the theory of agent's knowledge resembles the medieval doctrine of "maker's knowledge".²⁰ According to this doctrine, makers occupy a superior epistemic position with respect to what they produce or make themselves, when compared with mere users of a product. The difference is that only makers, but not users, have direct introspective access to their intentions or plans for making a certain product. In the same way, the theory of agent's knowledge implies that agents have a superior knowledge of their actions compared with mere observers, because agents know their plans for acting in a relatively *a priori* fashion.

3.2. Mind the gap

On the broadly Anscombean model developed above, an agent has knowledge of what she is doing "from the inside", by knowing her intentions (reasons, plans) whereas the observer does not. However, this can make it look as though Anscombe slides from a plausible view -- that one can have non-observational knowledge of one's intention for acting -- to the implausible view that this same knowledge is knowledge of one's intentional action. It is true that the overarching intention -- which is specified in the practical reasoning -- is indispensable for the

intentional action. But it's not clear why practical knowledge should yield knowledge of the intentional action, not just the intention. There seems to be a gap -- epistemic (if not also metaphysical) -- between the intention and the action.

In fact the objection applies not just to the teleological model, but also to the modified causal view. (The modified causal view is just the causal theory made more acceptable to internalists.) On the modified causal theory:

S knows that she is ϕ -ing just in case: i) S is ϕ -ing; ii) S believes she is ϕ -ing; iii) S's belief that she is ϕ -ing is caused *and justified* by her intention-in-action of ϕ -ing.

Here the immediate basis for S's belief that she is ϕ -ing is not direct awareness of the action itself, but awareness of her intention-in-action of ϕ -ing. So as with the teleological model, there is an epistemic gap between the subject's intention and her intentional action. Again it is not clear why knowledge of the former suffices -- non-inferentially, no less -- for knowledge of the later.

Both models face the problem that there appears to be an epistemic gap between having an intention and performing a certain intentional action. In each case, we have no difficulty with the idea that agents know their intentions without observation (by introspection), but the conclusion falls short of (AK), the claim that agents know their intentional actions without observation.

I shall first consider how the objection is to be met using the resources and sticking to the confines of Anscombe's approach to intentional action. I shall then discuss the objection in the context of a causal approach and show in section §4 how the objection can be met using fallibilism. The fallibilist approach provides welcome relief from the objection regardless of which model of agent's knowledge (causal or teleological) we adopt.

Anscombe's approach to the problem of the gap is simply to deny that the gap exists. She denies that the gap exists in every sense: both metaphysically and epistemically. More specifically, because there is no real metaphysical gap, there is no epistemic gap either. There are two ways in which a metaphysical gap could interpose between an intention and an action. First, the gap could be temporal and causal, as when a prior intention fails to eventuate in the corresponding action. Second, the gap might consist in the fact that the intention itself is conceived of as something that falls short of the whole intentional action. Here the intention would stand to an action as a sense-datum does to an object of perception. Anscombe denies both sorts of metaphysical gaps.

First, as we saw, Anscombe says that the failure to execute intentions to perform basic intentional actions is "necessarily rare". Consider once again this (2000: 87) passage:

What is necessarily the rare exception is for a man's performance in its more immediate descriptions not to be what he supposes. Further, it is the agent's knowledge of what he is doing that gives the descriptions under which what is going on is the execution of an intention.

One important point made in this passage is a point that we are now inclined to recognise as a Davidsonian point about the unintelligibility of massive errors (in this case massive errors about what we are doing). Our identification of our actions as intentional under some description depends on there being a reliable link between intending and doing. In the absence of such a reliable link between the intention to ϕ and the action of ϕ -ing, we would cease to be able to identify actions as intentional under that description (ϕ) at all. This means that the link between intentions and actions is not merely reliable, but constitutive.

The consideration appears powerful for simple cases. Consider what would be the case if a neuroscientist (with a perverse sense of humour) interfered with someone's brain so that whenever she intended to build a sand castle, she wound up eating sand instead. Repeated attempts to build the castle resulted only in more sand-eating. The subject's behaviour would be very puzzling to her. She would claim: "That's not what I meant to do at all." Eventually perhaps she would adjust her expectations, so that whenever she felt like she used to feel when intending to build a sand castle, she expected that she would eat sand. Anscombe's point would be that we begin to lose a grasp on the content of the original intention or plan apart from its manifestation in action. The intelligibility of the action (to the agent) depends on seeing it as the expression of intention; likewise, the identification of the intention reciprocally depends on viewing a certain action as its fulfilment. Still, one might worry that this sort of claim has to allow for occasional exceptions. And an occasional exception is all the sceptic needs to reinstitute the gap.

Second, Anscombe denies that a gap can open up between an intention-in-action and the intentional action itself. In (2000: sec. 29), Anscombe considers the temptation to split an intentional action into a mental component (such an intention, volition) and a physical component (the movement). It is then tempting to say that whereas the mental component is known without observation (by introspection), the physical component can only be known by observation. Her arguments against such a dual component view of action, however, are very compressed and not entirely convincing. She dismisses the view that only the mental component of the action is known non-observationally as “a mad account”, on the grounds that it is hard to see what this mental component amounts to apart from its physical effect. However, a behaviourist fear of unmanifested mental states is no argument against them. More obscurely, but more plausibly, she dismisses the view that all we ever do intentionally is what we *think* we are doing as “nonsense” (ibid.: 52). Presumably this is because this view leads to scepticism about whether anyone ever acts intentionally. Though such scepticism runs counter to common sense, it is not incoherent and Anscombe owes an argument against it. Anscombe’s arguments against a dual component view of action are suggestive, but not sufficient.

4. Fallibilism and the gap

The objection raised against (AK) based on the gap between intention and action instantiates a pattern of argument wholly familiar from certain sceptical arguments against our knowledge of the external world. The external world sceptic (EW-sceptic) might argue:

- (1) To know there really are material, physical objects, I need to know that my perceptions as of objects are veridical.
- (2) I don't know that my perceptions are veridical, since I cannot certainly distinguish them from non-veridical perceptions.
- (3) So, I don't know that there really are material objects -- at least not by perception.

Two standard moves made in response to EW-scepticism are: (A) to allow that one can in fact distinguish veridical and illusory perceptions, albeit only fallibly, and (B) to espouse a form of direct realism about perception on which perceptions reach all the way out to the objects themselves.²¹ I am going to advocate (A) in preference to (B). But I will briefly and programmatically indicate why I don't think (B) is the most promising response to the sceptical argument.

Option (B) is often advocated as an answer to scepticism, particularly that motivated by the argument from illusion. But such an approach is actually powerless to answer the sceptical argument mounted above, which we might call "the argument from the subjective indistinguishability of veridical and non-veridical perceptions". It is wholly compatible with direct realism that a subject not be able to introspectively distinguish the content of her veridical and non-veridical perceptions. Given externalism about mental content, the content of these perceptions will in fact be different. But this difference need not be transparent to the subject.²² So this kind of direct realist approach can at best persuade us not to take up the kind

of Cartesian standpoint from which an argument like the one above is raised. So the direct realist approach does not offer a straight response to that sceptical argument.

A straight response is possible by attacking the sceptic's underlying epistemological assumptions as in (A). The quest for certainty must be given up if knowledge is possible. Though I cannot know with certainty that my perceptions are veridical, I can know with less than certainty -- fallibly -- that my perceptions are veridical when they are. So I prefer a fallibilist motivation for rejecting (2).

To be sure, the concept of fallible knowledge is not without its detractors and a full development of a theory of fallible knowledge remains in its infancy.²³ For some theorists, the intractability of the Gettier problem is taken to indicate that fallible warrant is incompatible with knowledge. There is promising work that opposes this claim, so I shall not attempt a rebuttal here.²⁴ For other theorists, the idea of fallible knowledge just seems intuitively wrong. For example, Lewis says that "To speak of fallible knowledge, of knowledge despite uneliminated possibilities of error, just sounds contradictory" (1996: 549). Lewis may be right about how odd, to Cartesian trained ears, attributions of fallible knowledge sound. But the perceived oddness might be the result of a confusion. Once we recognise the confusion for what it is, we have reason to be optimistic about the coherence of the concept of fallible knowledge. It is an axiom of Cartesian epistemology that:

(*) If S knows that p, then S cannot be mistaken that p.

In one sense, this is trivially true: S knows that p entails that p is true. So if S knows that p then S cannot have gone wrong in picking up a false belief. Fallibilists of course deny (*). But they are not committed to denying the truism that knowledge requires truth. Fallibilists point out that S's knowing p does not logically require there is no way in which S *could have been* mistaken about p. All that is required in this respect for S to know p is that S is not *actually mistaken* about p. There are plenty of ways in which S could have gone wrong, but thankfully in the case where S knows that p, S didn't go wrong. The perceived oddness of the claim that S's knowing that p is compatible with the possibility of S's having been wrong about p must stem, I think, from a tendency of those who hear the claim to reduce possibilities to actualities. On the reductionist view, every possibility must be actual at some point in time (the Aristotelian view) or realised in some actual world (the Lewisian view). That is why a possibility of error is such a threat to knowledge for reductionists. For if there is a possibility of error, then at some point, in some world, there is an error and knowledge is lacking. But of course the fact that knowledge is lacking in that other world does not mean that it is lacking in our actual world. Reductionism, by not allowing mere (unrealised) possibilities, encourages a focus just on the actual. That reductionist tendency could lead those to read claim (*) wrongly as just the claim:

(**) If S knows p, then S is not in fact actually wrong about p.

It would be contradictory to deny this later claim.²⁵ However, as we've seen, fallibilists do not deny (**). So fallible knowledge is free from the charge of incoherence. I suspect that many

objections to fallible knowledge will be seen ultimately to rest on such confusion. So I conclude the way is wide open to affirm fallible knowledge.

Allowing fallible knowledge, far from encouraging scepticism, enables an elegant and mature response to scepticism. This type of response is available, moreover, not just in the case of sceptical threats to perceptual knowledge of the external world. Fallibilism can be used to defuse sceptical objections to the claim that agent's know what they are doing by intention. One sceptical argument against agent's knowledge has the following structure:

(A1) To know what she is doing, S must know that her intention-in-action is actually being executed.

(A2) S doesn't ever know that her intention-in-action is actually being executed, because S cannot distinguish between (i) the case where the intention-in-action corresponds to the actual action, and (ii) the case where S has an illusion of acting with that intention -- a mere illusion of having an intention-in-action.

(A3) So, S doesn't ever know that she is performing a given intentional action -- at least not by awareness of her intention.

As with the argument for EW scepticism, one could deny the second premise in two different ways: (A*) by adopting a fallibilist approach to knowledge of actions, or (B*) by adopting a direct approach to knowledge of actions.

This argument focuses on intentions-in-action. But since intentions-in-action are always embodied in actions, there is some optimism for thinking that one cannot have the relevant intention-in-action without performing the corresponding intentional action. This means that in this case approach (B*) is viable: one's intention-in-action does reach all the way out to include one's action. But again, it's not clear what help this closure of the metaphysical gap between intention and action is in meeting the sceptical argument. For it is possible that the intentions with which subjects act are not transparent to themselves. This is a possibility that Anscombe did not see.²⁶

As before, though, there is an alternative fallibilist motivation for rejecting (A2). S does fallibly know that her intention is being executed. Her awareness of her intention (and her lack of awareness that anything is wrong) provides good fallible reason for thinking that she is acting as described by her intention. In those cases where S is suffering an illusion or delusion about the efficacy of her intentions, S might falsely claim to have knowledge. But the fact that S might go wrong in that case does not detract from S's right to claim knowledge in the normal case, where her intention is actually being executed.

5. Conclusion

I have presented two distinct models of agent's knowledge: a causal-reliabilist model and a teleological, reasons-based account. Both approaches face a parallel difficulty about how the gap between an intention and an action gets crossed. In each case it might seem that closing the epistemic gap requires closing the metaphysical gap between intention and action. On a

causal theory the metaphysical gap gets closed only if (i) prior intentions perfectly reliably cause their corresponding intentional actions, or (ii) intentions-in-action are always embodied in the corresponding intentional actions. In each case, knowledge of an intention would suffice (on an externalist, reliabilist theory) for knowledge of the corresponding intentional action. However, the claim that prior intentions perfectly reliably give rise to actions needs further defence. So it looks as though the causal theory cannot deliver the result that agent's knowledge is anticipatory, being derived from knowledge of prior intentions.

On the teleological approach, what is required is a defence of why knowledge of a plan yields knowledge of what one is actually doing. As we saw, on Anscombe's theory of intentional action, the metaphysical gap is closed by making knowledge of one's plan for acting constitutive of one's intentional action. But again the closure of the gap needs more motivation.

I have suggested that we could learn to live with a metaphysical gap between intentions and intentional actions by adopting a more fallibilist conception of agent's knowledge. On the causal model, awareness of one's intention-in-action provides fallible evidence for thinking that one's intention is being executed. On the teleological model, awareness of one's plan for acting and awareness that one has started to act on that plan provide fallible evidence for thinking that one is acting as planned. If this is right, then progress in epistemology -- in the form of an improved understanding of fallible knowledge -- will lead to progress in the theory of action.

References

- Anscombe, G. E. M. (2000 [1958]). *Intention*. Cambridge, Mass.: Harvard University Press.
- . (1983). The causation of action. In C. Ginet and S. Shoemaker (eds.), *Knowledge and Mind: Philosophical Essays*, New York: Oxford University Press, 174-190.
- Feldman, R. (1981). Fallibilism and Knowing that One Knows. *Philosophical Review* 90, 77-93.
- .(2003). *Epistemology*. Upper Saddle River, NJ: Prentice Hall.
- Goldman, A. (1967). A Causal Theory of Knowing. *Journal of Philosophy* 64, 357-72.
- . (1976). *A Theory of Human Action*. Princeton: Princeton University Press.
- Gopnik, A. (1993). How We Know Our Own Minds: The Illusion of First-Person Knowledge of Intentionality. *Behavioral and Brain Sciences* 16, 1-14.
- Hetherington, S. (2001). *Good Knowledge, Bad Knowledge: On Two Dogmas of Epistemology*. Oxford: Clarendon Press.
- . (2005). Fallibilism. *The Internet Encyclopedia of Philosophy*.
<http://www.iep.utm.edu/f/fallibilism.htm>
- Hintikka, J. (1974). Practical vs. theoretical reason -- an ambiguous legacy. In his *Knowledge and the Known: Historical Perspectives in Epistemology*. Dordrecht: Reidel, 80-97.
- Hornsby, J. (1980). *Actions*. London: Routledge & Kegan Paul.
- Howard-Synder, D., Howard-Sydney, F., and Feit, N. (2003). Infallibilism and Gettier's Legacy. *Philosophical and Phenomenological Research* 66, 304-27.

- Hursthouse, R. (2000). Intention. In R. Teichman (ed.), *Logic, Cause and Action: Essays in Honour of Elizabeth Anscombe*. Royal Institute of Philosophy Supplement 46, Cambridge: Cambridge University Press, 83-105.
- Lewis, D. (1996). Elusive Knowledge. *Australasian Journal of Philosophy* 74, 549-67.
- Libet, B. (1985). Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action. *Behavioral and Brain Sciences* 8, 529-66.
- McDowell, J. (1986). Singular thought and the extent of inner space. In P. Pettit and J. McDowell (eds.), *Subject, Thought, and Content*. Oxford: Clarendon Press, 137-68.
- Moran, R. (2004). Anscombe on 'practical knowledge'. In J. Hyman and H. Steward (eds.), *Agency and Action*, London: Royal Institute of Philosophy, 43-69.
- Newton-Smith, W. (1980). *The Structure of Time*. London: Routledge & Kegan Paul.
- Pickard, H. (2004). Knowledge of Action without Observation. *Proceedings of the Aristotelian Society* 104, 205-30.
- Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Trevarthen, C. (1984). How control of movement develops. In H. A. Whiting (ed.), *Human Motor Actions -- Bernstein Reassessed*. North Holland: Elsevier Science Publications, 223-61.
- Velleman, J. D. (1989). Epistemic Freedom. *Pacific Philosophical Quarterly* 70, 73-97.
- Wittgenstein, L. (1958). *The Blue and Brown Books*. Oxford: Blackwell.
- Zagzebski, L. (1994). The Inescapability of Gettier Problems. *Philosophical Quarterly* 44, 65-73.

Notes

¹ Anscombe's term for such knowledge in *Intention* is "practical knowledge". However, I have avoided that term as I wish to defend the idea that agents have a distinctive non-observational knowledge of their actions without being committed to her claims about the existence of a kind of knowledge with a reverse direction of fit from ordinary knowledge.

² The idea originates in English speaking philosophy with Wittgenstein's qualified claim that "in a large class of cases it is the impossibility of taking an observant attitude towards a certain action that characterizes it as a voluntary one" (1958: 153). Anscombe makes a stronger claim with regard to intentional actions, that "the class of things known without observation is of great interest because of the class of intentional actions is a sub-class of it" (Anscombe 2000: 14).

³ This should not be misinterpreted as the stronger, false claim that agents cannot in principle observe what they are doing. The point is just that an agent's awareness of her action is qualitatively different from (and certainly more first personal in character) than that of a mere observer.

⁴ I am indebted to Stephen Hetherington for helping me to see how an epistemologist's toolbox (and especially fallibilism) might be fruitfully brought to bear on an issue in action theory.

⁵ The basic description of the action should present the action as occurring early in the causal genesis of the action, but it is also important that it not be a recondite description of the action in terms of inner mechanisms. Rather, a "suitably basic" description of the action will describe the action in terms recognisable by the agent as something that she knows how to do directly without having to do anything else first (e.g. "raising my arm"). This may mean that a suitably basic description is "teleologically basic" but not "causally basic". On the distinction, see Hornsby (1980: chs. 5-6).

⁶ That is, we generally assume that effects are temporally posterior to their causes. This assumption has been challenged by believers in immanent causation. For example, a cloth may bleed its dye into some boiling water, with the bleeding of the cloth being the immanent cause of the discoloration of the water. For a full discussion, see Newton-Smith (1980). Even if causes and effects can be simultaneous, however, the anticipatory character of agent's knowledge would not be

restored.

⁷ The estimation of 50-250 ms is from Trevarthen (1984: 224). Could such anticipatory knowledge be based on a prediction about what will happen, using a feed-forward model? Yes, it could and this account would be compatible with the one offered here.

⁸ On the distinction between prior intentions and intentions-in-action, see Searle (1983). The existence of actions without prior intentions is one reason Anscombe advances for rejecting the causal analysis of intentional action in Anscombe (1983). The problem of causally deviant chains is another problem for the causal theory of action, but see Goldman's (1976) solution.

⁹ The assumption that the link between the intention-in-action and the action is causal will bother some theorists, since efficient causes precede their effects. It may well be that the nature of the link is not that of an efficient causal one. This difficulty might persuade some to abandon a purely causal approach in favour of a reliable one, on which intentions-in-action are reliably correlated with (but need not cause) corresponding actions.

¹⁰ Here for simplicity I have assumed that there is a prior intention causing the action. I think generally there will be a prior intention causing the action, whether conscious or not.

¹¹ On the distinction between general and specific evidence, see Velleman (1989).

¹² I follow standard usage in using "warrant" as a neutral term for whatever turns a true belief into knowledge.

¹³ If this requirement is lifted, then it is indeed the case that "agent's knowledge" would be a kind of true belief not at all based on evidence. Though some, such as Goldman (1999) and Hetherington (2001), would countenance such true belief as a weak kind of knowledge, most epistemologists would want to maintain some justification requirement on knowledge.

¹⁴ For an argument against this common view, see Hetherington (2001: ch. 4).

¹⁵ I am indebted to a conversation with David Coutts here.

¹⁶ It is also indebted to some recent Anscombe exegesis by Richard Moran (2004) and Rosalind Hursthouse (2001). However, it requires, I think, a correspondingly greater commitment to Anscombe's peculiar views than the other model.

¹⁷ This point is emphasized in Moran (2004).

¹⁸ That the notion of cause is not efficient is a point that should be obvious given Anscombe's reference to Aquinas, but Hursthouse (2001) should have credit for pointing it out. There is evidence of opposition to the causal analysis of intentional action in *Intention* (2000: secs. 11-19), and the rejection of the causal analysis was a lifelong theme, appearing again in Anscombe (1983).

¹⁹ "Direction of fit" is Searle's term in Searle (1983), but the idea (as he acknowledges) derives from Anscombe (2000 [1958]).

²⁰ The observation is due to Hintikka (1974).

²¹ For developments, see McDowell (1986).

²² McDowell (1986) is of course well aware of this point.

²³ For a comprehensive introduction to fallibilism, see Hetherington (2005). For the application of fallibilism to sceptical problems, see Feldman (2003: 122-9). For more specific considerations relating the KK principle and fallibilism, see Feldman (1981).

²⁴ See Zagzebski (1994). For arguments against the idea, see Howard-Synder *et al.* (2003).

²⁵ We need not convict Lewis of such confusion: he is just reporting the perceived oddness of claiming that S knows p despite uneliminated possibilities. However, it is then puzzling why, if he is not confused, Lewis is not more sympathetic to fallibilism. It is an interesting question for further research to what extent Lewis's strong modal realism -- his belief that all worlds are actual -- influences his rejection of fallibilism.

²⁶ The approach will not, of course, work either if the sceptical argument is recast in terms of a subject wondering whether her prior intention to ϕ will give rise to an action of ϕ -ing. Here there is an undeniable metaphysical gap between intention and action, so that knowledge of intention won't yield knowledge of the action immediately.