As the road to hell teaches us, intending and doing are not the same thing. That is, at least not for fallible human agents like ourselves. For a divine agent, perhaps, there is no gap between intending and doing. If God intends to change the world in some way, that fact is presumably sufficient for the change to be brought about; there is nothing further He must do. But for us to bring the world into conformance with our intentions, we must often take means—means that may fail. And while we might assume that God never procrastinates or forgets to do what He intends, we sublunary creatures cannot say the same. Quite simply, for us, merely having an intention falls far short of ensuring that we will get it done.

God's omnipotence has also been said to afford him a special way of knowing about the world, dubbed "practical knowledge" by Thomas Aquinas: He can know how the world is, or how it will be, just by forming an idea of it in His mind and bringing the world to conform to that idea.¹ He is not limited to investigating the world and only then forming beliefs about what He finds there; He can obtain knowledge simply by constituting the truth in a particular way. This notion of practical knowledge is opposed to "speculative" knowledge, which is attained by bringing one's mind into conformance with the facts. If there is an error in speculative knowledge, the fault is in the judgment; there is a fact out there that has not been properly apprehended. A failure of practical knowledge, in contrast, is a failure to bring the facts into conformance with one's idea of them. As Aquinas conceived of it, practical knowledge is "the cause of what it understands," and as such, it enables the knower to remain a step ahead of the truth, rather than a step behind—a desirable power to have, indeed.

In spite of our patent lack of omnipotence, it has been thought by some that through our intentional actions, we human agents can also have a special kind of knowledge of the world. The root of this idea is a provocative claim of Elizabeth Anscombe's: that when we act intentionally, we seem to know what we are doing without observation.² Anscombe initially suggests that when we act intentionally, we

---

necessarily know what we are doing; if one is not aware that one is \( \phi \)-ing, this suffices to show that one is not \( \phi \)-ing intentionally.\(^3\) But more than this, she claims that this knowledge is had without the need to observe what it is we are bringing about. Whereas an agent might be able to discover that he is inadvertently drumming his fingers on the table only by hearing the noise he is making, or by noticing the annoyed expressions on his colleagues’ faces, he will not need to appeal to sense evidence to know that he is drumming intentionally (in order to annoy them). He will be able to report a correct description of his intentional actions without looking to see, and often at a level of description that far outstrips the observational evidence provided by his bodily movements and their immediate results.\(^4\) For instance, our drumming agent might know that what he is doing in drumming his fingers is preventing his rival from receiving a job offer by sabotaging the upcoming vote, a claim there is simply not sufficient observational evidence to support. Most intriguingly, Anscombe insists that as long as things go well, what one knows without observation is *what happens* — not merely what one is trying to do, but what is actually happening out in the world.\(^5\)

Anscombe makes the further suggestion that this knowledge of our intentional actions is practical rather than speculative, in the sense that a discrepancy between belief and fact is a failure of performance, not of belief — it is a failure to perform the action one had in mind, not a failure to have the correct action in mind.\(^6\) She cites Aquinas in characterizing this knowledge as “the cause of what it understands,” in that “without it, what happens does not come under the description — execution of intentions — whose characteristics we have been investigating.”\(^7\) It is important to note, however, that she does not claim that this knowledge is the *efficient* cause of the action it represents.

The claim that our special knowledge of our actions is practical in Aquinas’s sense and the claim that it is non-observational are customarily taken together. However, the two ideas are clearly not the same; at least, a bit of knowledge might be non-observational without in any sense being the cause of what it represents.\(^8\) Thus, to gain clarity on the alleged connection between action and knowledge, I think it is vital to keep these two claims distinct. My chief concern will be with the latter, which I will call Anscombe’s “Non-Observational Knowledge Thesis”: that if an agent is acting intentionally, he will know without observation what he is doing, under the description(s) on which the action is intentional. The notion of practical knowledge will also play a significant role in the discussion, however, and a second object will be to get clear on the relationship of non-observational knowledge to practical knowledge.

There are two major sorts of reactions Anscombe’s Non-Observational Knowledge Thesis has provoked in the philosophical literature. One response is a kind of skepticism, motivated by the fact that Anscombe’s formulation of the thesis is overly strong and easily defeated by counterexamples. Donald Davidson took this skeptical line, holding that it is simply a mistake to suppose that acting intentionally entails acting knowingly — non-observationally or otherwise. He pointed out that one may aim at some outcome, and end up bringing it about in the way one is trying to — so doing it intentionally — while remaining in doubt all the while that one is doing it.\(^9\) His well-known example concerns a man who presses down heavily on a stack of carbon-paper with the aim of making ten carbon copies at once, in the face of substantial doubt that his efforts will succeed. But if they do, he will

\(^3\) Anscombe (1963), 11.

\(^4\) This is not to say that there is no observational input to intentional action whatever, but rather that what we know about what we are doing often greatly exceeds these inputs.

\(^5\) She writes, “I myself formerly, in considering these problems, came up with the formula: I do what happens. That is to say, when the description of what happens is the very thing which I should say I was doing, then there is no distinction between my doing and the thing’s happening” (Anscombe 1963, 52–53, emphasis in the original).

\(^6\) Anscombe (1965), 56–57.

\(^7\) Anscombe (1963), 88.

\(^8\) For instance, knowledge of the answer to a math problem.

\(^9\) Davidson (1978), 91–92.
intentionally have made ten carbon copies at once, though without actually believing in the process that he was doing so. Such examples, easily multiplied, convincingly demonstrate that non-observational knowledge is not as coincidental with intentional action as Anscombe would have it. Davidson concluded that a special kind of knowledge is not a distinguishing feature of intentional action after all, and his causal-psychological approach to understanding the domain of action theory became dominant in Anglo-American philosophy in the decades following the publication of *Intention*.

More recently, however, there has been a resurgence of interest in non-observational knowledge as an important and revealing phenomenon associated with agency. This trend is associated with the second sort of reaction to Anscombe’s Non-Observational Knowledge Thesis — a brand of “cognitivism” about intention. The Cognitivist thought is that if we take the connection between intentional action and non-observational knowledge sufficiently seriously, we should be led to the conclusion that intending must be constitutively tied to belief. Most prominently, J. David Velleman and Kieran Setiya have each appealed to versions of what I will call “Strong Cognitivist” theories of intention to explain the special knowledge we have of what we do intentionally. This kind of theory identifies intending with a sort of believing, claiming that an intention to φ is some special kind of belief that you are φ-ing or will φ. I call this type of view “Strong” Cognitivism to distinguish it from views on which intending is a composite attitude that involves the belief that you will do what you intend, but that also involves some further, separable conative component; the latter I call “Weak” Cognitivism. I will explain this distinction further and say more about Weak Cognitivism in Section 6, but for now the focus will be on Strong Cognitivism — the identification of intention with a kind of belief — since both Velleman and Setiya appeal to this sort of view.

The Strong Cognitivist strategy for explaining Anscombe’s insight is to note that if intentions (in action) are beliefs about what you are doing, and all intentional action involves the presence of some intention, then all intentional action must involve the presence of a belief about what you are doing. As long as one is successful in one’s intention, this belief will be true (bracketing here the question of why it should count as knowledge). And since intentions are formed out of a preference for performing a given action, rather than on the basis of evidence that one will perform it, they will be beliefs that are not grounded in observation or evidence of any kind — as Velleman puts it, they will amount to “spontaneous” knowledge, or knowledge that the agent originates. Finally, according to Strong Cognitivism, non-observational knowledge of doing what we intend is indeed practical in Aquinas’s sense. Because the knowledge is embodied in the intention itself, which causally brings about the very action that is the representational content of the intention-belief, the knowledge turns out to be the cause of what it represents. In effect, the concern that Davidsonian skepticism is an inadequate response to Anscombe’s puzzle has led both Velleman and Setiya to cash out the practical-knowledge idea by arguing that intentions are a kind of belief.

My view is that both the skeptical and the Strong Cognitivist reactions to the Non-Observational Knowledge Thesis are extreme. Davidson focuses on the overly ambitious formulation of the claim and ignores the kernel of truth in the insight that we usually do know at least something about what we are doing intentionally, and know it without having to observe ourselves. The Strong Cognitivists rightly take this kernel of truth seriously, but go too far in concluding from it that intentions must be a kind of belief. I will argue that there are reasons to doubt the truth of Strong Cognitivism, as well as its success in accounting well for the full extent of non-observational knowledge.

---

10. It should be noted that although Bratman (1991) also employs the term “cognitivism” in the context of discussing Velleman’s theory of intention, his use of the term is slightly different from mine. Bratman uses the term to refer to a particular strategy of explaining some of the norms of practical reasoning by appeal to the claim that intention involves belief, whereas I am using it here to refer to belief theories of intention themselves. Neither of these uses should be confused with the way the term is deployed in the context of metaethics.


of agency — enough to prompt us to try to avoid this route. To do this, a middle course must be found between these two extremes, neither dismissing the significance of non-observational knowledge of intentional action nor embracing a Strong Cognitivist theory of intention. My aim here is to show how this middle course can be steered.

In this, I will to some extent be following a path suggested by another contemporary of Anscombe and Davidson — H.P. Grice. In his 1971 British Academy Lecture, Grice sketches the outlines of a theory of the relationship between intention and belief that has a structure some have called “inferential.” There are aspects of Grice’s account that are problematic, but I think the general inferential schematic can be extracted and developed in a way that occupies the middle ground I have described. The object of this paper is to show how such an inferential, neo-Gricean story about non-observational knowledge of action might go, and to argue for why we should be moved in this direction.

The basic strategy is to push any groundless component of our knowledge of our own intentional actions back into the head, while still capturing the sense in which that knowledge is non-observational. The view I will propose acknowledges that we normally know non-observationally what we are doing, while holding that this knowledge is not spontaneous, but evidentially based on our knowledge of what we intend to be doing. With respect to the latter, I will sketch my own preferred account of how we know what we intend, but it is important to note that the Inferential Theory of how we know what we’re doing does not depend on accepting this account; the theory is compatible with different views regarding our knowledge of our intentions. To really nail down the details of the theory, more space would be required than I have here, but I think the contours will be enough to reveal its advantage over the aforementioned extremes. What I will argue, in a way, is that we can account for non-observational knowledge of action without abandoning a commonsense view of intention, by rejecting the idea that this knowledge is the efficient cause of what it understands.

1. Intentional action and non-observational knowledge

Though Davidson correctly points out that Anscombe’s formulation of the Non-Observational Knowledge Thesis cannot be right, there is clearly some important connection between intentional action and knowledge that has made her observations so compelling to many. But how to capture this kernel of truth? It cannot be the case that one knows non-observationally of everything one ends up doing intentionally, but even in carbon-copy cases, there is usually something the agent knows he is doing. Setiya attempts to accommodate Davidson’s point by proposing a more modest formulation of the thesis: that when an agent is acting intentionally, there will be some description of the action he does in the belief that he is doing it, even if he does not believe he is accomplishing his ultimate aim. The carbon-copier does not believe he is making ten carbon copies at once, for instance, but he does believe he is pressing down very hard on the pile of carbon paper. Setiya’s formulation is weaker than Anscombe’s, but it still posits a necessary connection between intentional action and belief (again preferring to leave aside the question of whether the belief counts as knowledge), in that an action will not count as intentional unless it involves some non-observational belief about the action.

I am in agreement with Anscombe and Setiya that non-observational knowledge (or at least belief) is a pervasive phenomenon among actions that are intentional. It is worth questioning, however, whether the adverb ‘intentionally’ really works in a way that will support the


15. Op. cit., 25–26. In his formulation of the Belief condition here, he does not explicitly characterize the belief as non-observational, but the preceding discussion on page 24 makes it clear that what is at issue is Anscombe’s non-observational knowledge thesis.

16. Again, I am bracketing the complicated question of when (if ever) this belief amounts to knowledge. We do commonly want to speak of knowledge in straightforward cases where the world cooperates with the agent’s efforts,
claim of a necessary connection between intentional action and any kind of knowledge — non-observational or otherwise. The significance of this question is that the answer may bear on the kind of explanation that can be given. As Setiya has argued, if breakdowns between acting intentionally and believing one is so acting are impossible, this might be thought to rule out any explanation on which intentional action and belief are only contingently tied together. A merely causal, inferential connection would be bound to break down now and again, allowing for the possibility of acting intentionally while having no relevant belief (let alone knowledge) about what one is doing.

But is this impossible? One way to establish the necessity of the connection would be to demonstrate that an action counts as intentional only if the agent has the non-observational belief that he is performing it, or believes non-observationally that he is doing something with that aim. It does not seem to me clear that this is so. We classify actions as intentional for a variety of purposes, and I think the adverb can be sensitive to multiple properties of behavior. For one, we use it to capture the fact that an episode of behavior had a goal-directed, means-end structure and exhibited responsiveness to reasons, and I think we sometimes classify such behavior as intentional even if the agent has no related non-observational belief about what he is doing. In the classic example of driving home on autopilot, for instance, the driver takes a certain route because it is the best way home, uses his turn signal in the appropriate places for reasons of safety, raises the garage door in order to park his car, and so forth. To express the fact that his turning and signaling are purposive, have a means-end structure, and are performed in response to his reasons, it is natural to say that he turns and signals intentionally. If his driving is habitual enough, however, he may have no belief that he is doing these things, and might only be able to discover that he is doing them by observation (e.g., noticing that the turn signal is on). Or likewise, an expert cabinetmaker may be completely immersed in his craft, and respond without thought to what is to be done without forming a belief at each step about what he is doing. This is no clear reason to deny that the craftsperson intentionally executed the steps involved in making a cabinet.

I do not take these examples to demonstrate conclusively that intentional action need not involve the presence of some relevant non-observational belief. My aim is merely to cast doubt on the ability of the adverb ‘intentionally’ to support the claim that intentional action is necessarily connected with non-observational belief, and thereby on the constraint that our explanation of the Non-Observational Knowledge Thesis must account for all instances of action that can be characterized as intentional. One might respond by conceding that the adverbial characterization of action as intentional cannot carry that much weight, but maintaining that there is a sense of “acting for a reason” that does require that one believe one is so acting. This may be right, but it does not help to establish that it is anything deeper than a mere classificatory convention. That is, though the driver on autopilot flips his turn signal in response to a reason, it may be true that we do not count him as using it for a reason unless he believes he is using it. But such a classificatory decision would not necessitate a non-inferential explanation of the source of the belief in question, in the way Setiya would have it. It is consistent with supposing that breakdowns occur between acting in response to reasons and believing one is doing so, but that we choose not to call such behavior action for a reason.

Rather than argue over the limits of the concept of intentional action, I suggest that we bracket questions of classification as intentional or not and focus on the underlying relationships between intention, belief, and action. Anscombe’s insight is compelling because in paradigm cases of action we can all agree are intentional, the agent usually

---

18. Joshua Knobe’s research on the application of the adverb ‘intentionally’ is relevant here and bears this claim out, I think. See for instance Knobe (2003).
19. Thanks to Setiya for pointing out this possible move.
does have beliefs about what he is doing, and not on the basis of observing himself in action. I will leave it open for now whether this non-observational knowledge is “the cause of what it represents,” though I will return to this question later. The pervasive presence of non-observational belief in intentional action is the kernel of truth that Davidson ignores, and our theory of action should account for this phenomenon.

2. Strong Cognitivist theories of intention

Once we acknowledge the need to explain the pervasiveness of non-observational belief in cases of intentional action, should we be led to a Strong Cognitivist theory of intention? Velleman and Setiya have both argued that in order to account for why, as Setiya puts it, “when an agent acts for reasons, there must be something she does in the belief that she is doing it,” we should conclude that intentions themselves must be a kind of belief.20 Recall that the basic idea behind this strategy is that if all intentional action involves an intention, and an intention (in action) is a kind of belief about what one is doing, then (successful) intentional action involves fulfilling one’s belief about what one is doing — thus providing oneself with a special kind of knowledge of one’s actions. And because this knowledge is embodied in one’s intention, which in turn is the cause of the action, it clearly does not depend upon observation of that action.

But let us be a bit more precise about the views on the table. According to Velleman, intentions are spontaneous, self-fulfilling expectations of action, formed in part in order to have advance knowledge of what we are doing and why in the instant we move to act.21 Intentions are ordinary beliefs on his account, although they are formed out of a preference for a given action rather than on the basis of antecedent evidence for their truth, thus meriting the status of “spontaneous.” Setiya also holds that intentions are a kind of belief, but for him they are a distinctive kind of “desire-like belief”; they are self-referential, motivating beliefs that you are hereby φ-ing going to φ.22 The two agree, however, that the explanation of the Non-Observational Knowledge Thesis is that intentions are cognitive attitudes that can both embody knowledge and bring about what they represent.23

An initial thing to emphasize about this kind of theory is the peculiarity of beliefs that are standardly formed out of a preference for their coming true, and not in response to sufficient prior evidence.24 Indeed, they are conclusions that are often leapt to in the face of evidence rendering the occurrence of the represented action highly improbable, since prior to forming the intention-belief, the agent is quite unlikely to end up performing that particular action. In Velleman’s own words, intending is a kind of “licensed wishful thinking” — licensed by the agent’s knowledge that he is able to φ if he intends to, together with his knowledge that intention-beliefs tend to be self-fulfilling.25 Setiya provides an alternative account of the epistemic justification of intention-belief formation, arguing that knowledge how to φ entitles one to form the intention to φ.26 On both accounts, however, the fact remains that of all the actions open to the agent, each of which he knows he is able to do and knows how to do, he must jump to the conclusion that he will do a particular one of them, out of a preference for performing that action.

23. As I argue in more detail in Paul (2009).
24. See especially Velleman (2004) and Setiya (2004, 2007). Although Velleman’s and Setiya’s views differ with respect to how the agent is motivated to act on an intention-belief — Velleman holds that intentions are ordinary predictions that the agent is motivated to fulfill by a powerful independent desire for self-understanding, while Setiya holds that intentions are intrinsically motivating states — I think it is acceptable to treat them in kind for my purposes here. In this paper, I do not wish to challenge either of their views with respect to how intention-beliefs are motivating once they are in place. Their explanations of the Non-Observational Knowledge Thesis are sufficiently similar to address jointly, and I take the worries I will raise regarding intentions as spontaneously formed doxastic commitments to apply equally to both views.
Now, it is a matter of controversy whether jumping to wishful, self-fulfilling conclusions may be justified in some cases, but the practice certainly deviates from the paradigm rules for rational believing. If there are situations in which we are licensed to form beliefs without sufficient prior evidence, we would expect such cases to be exceptional, rather than the basis of our entire practical lives. Omnipotent agents may experience no breakdowns between intending and doing, but surely we human agents need sufficient grounds to believe we will accomplish our aim in any given case, beyond some general knowledge of ability and know-how. This is not a knock-down argument against the view, but I think the fact that Strong Cognitivism makes intending out to be a kind of wishful jumping-to-conclusions gives us prima facie grounds to hesitate—especially given that we are talking about an attitude that is supposed to amount to knowledge.

Strong Cognitivism does appear to provide an elegant explanation of the Non-Observational Knowledge Thesis. However, this explanation is only as plausible as the claim that intentions are beliefs, and there are good reasons to doubt that this is so. Quite simply, as many people have pointed out, it seems that we can intend things we do not believe we will do. This is a different point than Davidson’s carbon-copier example, which demonstrates only that an action can be intentional without the agent believing he is performing it. It is potentially consistent with that point that intending is believing, if actions can be intentional without being intended.

But even intended actions seem not necessarily to involve the belief that you will do what you intend. One might intend to stop and get gas on the way home, but think there is a significant possibility that one will forget and drive straight home without stopping. Or perhaps one has a complicated action plan one believes will possibly succeed—a six-year plan to get into graduate school and get one’s Ph.D.—but that one is not sufficiently certain about to count as believing one will achieve the intended aim. The aspiring student recognizes that there are many ways her plan could fail, and yet we still want to say that she intends to get her Ph.D., in light of the role that goal plays in guiding her actions over the years, generating means-end reasoning, and her disposition to take alternative means if the original plan isn’t working—transferring to another institution if she cannot find anyone to advise her, for instance. But she may not fully believe she will get her Ph.D.; she would not bet a great deal of money on it.

A possible response to these examples is to insist that if one has doubts about succeeding in one’s ultimate aim, then one does not actually intend it. Perhaps what one intends is merely to try to stop for gas or to get one’s Ph.D. This is a tempting move, but it is important to recognize that it is potentially in tension with the strategy of appealing to Strong Cognitivism to explain non-observational knowledge. After all, if all one intends is to try to get gas, then the Strong Cognitivist would have to say that the knowledge embodied in one’s intention is limited to the fact that one is trying; one would have no special knowledge of what is actually happening. But Anscne’s claim was that when an action is intentional, the agent will have a special kind of knowledge of what happens, and not merely of trying.

Setiya’s strategy is to settle for a weak vindication of that claim, maintaining that if the agent does not believe he is bringing about

27. William James (1897) notoriously argued that it is.
28. To be clear, Velleman does not agree that theoretical rationality requires us to have sufficient evidence for our beliefs prior to forming them, as long as the balance of evidence supports the belief once it is in place. However, the fact remains that prior to forming an intention, there are many actions available to the agent, each of which he would do if he intended it, and he must jump to the as-yet-unsupported conclusion that he will do a particular one of them. Against this claim, see Langton (2003). For a more detailed discussion of Setiya’s argument, see Paul (2009).
29. This scenario and the following one are adaptations of examples provided by Bratman (1987).
30. Setiya himself points out that to account for the Non-Observational Knowledge Thesis, intending to try is not enough: “Despite what Davidson suggests, it is not enough that the carbon-copier is intentionally trying to make ten copies, in the paradigm sense of ‘intentional action’ that involves belief. He is and must be doing specific things—for instance, pressing hard on the paper—in that paradigm sense” (2007, 25).
his aim, he does not intend it — he intends only to try — but he must fully intend something he is doing in pursuit of that aim. At the least, he will intend the basic actions involved in his action plan, and he will perform those basic actions in the belief that he is doing so. However, even this more qualified claim seems to have its counterexamples. Imagine someone who has had their arm temporarily paralyzed, and who was told by the doctor that the paralytic would wear off at time $t$. Having good reason at $t$ to believe it is possible for him to raise his arm, it seems that he can form the intention to raise his arm at $t$. It also seems possible, though, that he does not fully believe his arm will go up — indeed, it would plausibly be epistemically irresponsible for him to believe this, given his only tepid confidence that he will be able to.\footnote{Setiya (2007), 5–56.} I think this type of example provides good reason to think that an agent can intend an action $\phi$, and even perform it intentionally, without doing anything he fully believes he will do. And this gives us good reason to worry about the truth of Strong Cognitivism.

This last issue arises because if intentions are beliefs about what you will do, the Strong Cognitivist must find a way to connect intentions up to further aims the agent pursues intentionally but does not fully believe he will accomplish. A related challenge faced by the cognitivist account is that it has no obvious way to explain any non-observational knowledge of intentionally bringing about anything you are not actually aiming at. But many think we do things intentionally that we do not intend.\footnote{Bratman presented an example of this form in commenting on an earlier draft of Setiya (2008) at the 2006 Wake Forest Conference on Agency and Action. Examples with a comparable structure have been employed by Robert Audi (1993) and by Velleman (2004), but in both cases to make different points than I am making here.} For instance, it is common to think of some of the expected side effects of our intended actions as being brought about intentionally, though they are not themselves intended. If I know I am going to scratch the car by trying to park it in a small space, and I go ahead and try to park it, I intentionally bring it about that the car is scratched.\footnote{Harman (1976) offers an instructive discussion of this issue; for research on the characterization of some foreseen side effects as intentional in ordinary language, see Knobe (2003).} And on the face of it, it seems to me that we often have non-observational knowledge of bringing about the intentional but unintended side effects of our intended actions, knowledge that appears quite similar to that we have of doing what we intend.\footnote{To borrow an example from Harman.}

Take the well-known example of the Strategic Bomber who intends to bomb the enemy’s munitions factory, in order to undermine the enemy’s war effort. Upon reflection, he realizes that bombing the factory will also damage a nearby school and kill some schoolchildren. He takes this fact seriously in his practical deliberation about what to do, and deeply regrets the outcome of killing the children, but ultimately decides to proceed with dropping the bombs. His does not have the intention to kill the children; he is no terrorist, and would not track the children with his bombs if they happened to move somewhere else. To him, this is an unfortunate side effect of pursuing his intended aim. Nonetheless, as he flies his plane toward the factory and the school, what he has non-observational knowledge of is not limited to the fact that he is on his way to bomb the factory; it seems to me that he knows in a similar way that he is en route to cause damage to the school and kill some children.\footnote{Anscombe (1963, 42) actually denies that the bringing about of expected but unintended side effects are intentional actions, because they fail her criterion that an intentional action must be the kind of thing for which practical reasons could be proffered in answer to the question “Why?” She does not consider the question of whether the agent has non-observational knowledge of them, however.} What to say about this? Some might want to deny that foreseen side effects count as being brought about intentionally, but again, the labeling issue is beside the point; whatever we call these deeds, the non-observational knowledge we have of bringing them about demands an explanation. This poses a special challenge to the Strong Cognitivist.
for if, as he supposes, non-observational knowledge is simply embodied in one’s intention – to bomb the factory – then such knowledge of anything beyond what is intended remains unaccounted for.\(^{37}\) One possible move would be for them to deny that the side effects are unintended; that is, they could pack all the foreseen side effects into what is intended. This would explain how non-observational knowledge of foreseen side effects is possible, but it would be a steep price to pay for the theory of intention. It is highly counterintuitive to say that the Strategic Bomber intends to kill the children, given that this outcome is neither his goal nor a means to his goal, and that he lacks any disposition to track that outcome with his efforts or means-end reasoning. The natural way to capture the difference between the Strategic Bomber and a Terror Bomber, part of whose reason for bombing the factory is in order to terrorize the enemy by killing some children, is to say that the Terror Bomber intends to bomb the factory \textit{and} to kill the children, while the Strategic Bomber intends to bomb the factory but does not intend to kill the children. Packing all the foreseen side effects into what is intended would have the effect of divorcing intention from the agent’s goals, which seems to me to revise the concept of intention to the point of talking about a different attitude altogether. Moreover, it

\(^{37}\) We must not be distracted by the fact that killing the children is a causal upshot of dropping the bombs, which requires that the agent engage in cause-and-effect reasoning to know of this upshot. The same point holds with respect to side effects that are not causally related to the intended action, but rather are related in virtue of an abstract conventional structure. Consider a second example: I am intentionally signing my name to a piece of paper that I know to be a letter of intent to attend graduate school at Stanford, and I am signing my name with the intention of bringing it about that I am a Stanford graduate student. Intuitively, I know without observation that what I am doing is signing my name and thereby making myself a Stanford student. However, I also know on the basis of background knowledge that Stanford students are Cardinals. But I do not \textit{intend} to be making myself a Cardinal with my signature; perhaps it is quite important to me that I am a die-hard Berkeley Bear, and if I could achieve my objective without becoming a Cardinal, I would. But nevertheless, my act of signing the letter is also an act of making myself a Cardinal, and just as I know without observation that what I am doing is bringing it about that I am a Stanford student, I know without observation that I am bringing it about that I am a Cardinal, though the latter is merely a side effect of what I intend.

\[ How \text{ We Know What We’re Doing } \]

The central idea behind the Inferential Theory is that our beliefs about our intentional actions are not spontaneous, but evidence-based – based largely, I will argue, on our knowledge of what we \textit{intend} to be doing, where intentions are not themselves beliefs. This idea has commonly been dismissed out of hand on the grounds of another claim of Anscombe’s: that the special knowledge we have of our intentional actions is self-evidently “immediate.”\(^{38}\) That is to say, the agent’s belief about what he is doing not only seems not to be inferred

\(^{38}\) Anscombe (1963, 12) compares our special knowledge of what we do intentionally to our knowledge of the position of our limbs (problematically, it

\[ \text{PHILOSOPHERS’ IMPRINT} \]

\[ \text{- 9 -} \]

\[ \text{VOL. 9, NO. 4 (OCTOBER 2009)} \]
from observational evidence, but not to be based on prior evidence of any kind. After all, if the agent believes he will \( \phi \) on the basis of sufficient evidence that he is going to \( \phi \), we would no longer think of him as intending to \( \phi \) but rather as predicting that he will \( \phi \).40 Furthermore, the knowledge feels to us to be available “directly,” without the need for any sort of reflection. Anscombe goes so far as to say that non-observational knowledge is a special kind of knowledge that requires no grounds whatsoever.41 If these claims are true, it would rule out the possibility of the non-observational knowledge being grounded in anything prior to itself.

This line of reasoning is unconvincing, however. First, merely noting that our knowledge of what we are doing seems to be directly accessible upon reflection does nothing to demonstrate that it is non-inferential. Evidence-based information processing can take place rapidly and automatically at a non-conscious level, without the mindful entertaining of premises or feeling of drawing a conclusion. We are quite comfortable with this idea in the contexts of information-processing subsystems like visual perception or language processing. Some may balk at calling such non-conscious information-processing an inference, but that is again a labeling issue; the essential point is that the beliefs resulting from such a process are derived from prior states that (ideally) evidentially support those beliefs. Recognizing that evidenced-based processing of this type need not present itself to consciousness as such clears the way for the suggestion that non-observational knowledge of what we do intentionally is based upon—but not equivalent to—knowledge of what we intend to be doing.42

Second, if the evidential basis for the non-observational belief is the agent’s intention, the objection that his \( \phi \)-ing could not be up to him does not get off the ground. We must distinguish between basing one’s intention on sufficient evidence that its content will come about and basing one’s belief on such evidence. The former is incoherent, but the latter need not be. As long as we are not supposing that the intention is the belief, then the intention can be based on practical reasons in the ordinary way (in the normal case, at least), while the belief can be based (in part) on the evidence provided by the intention. Thus, neither sense of “immediacy” commonly attributed to non-observational knowledge of action precludes it from being evidentially based on the agent’s intention in acting.

A stronger objection to inferential theories is the aforementioned claim that acting intentionally necessarily involves some non-observational belief about what he is doing. The charge was that an inferential process could not function so infallibly as to preserve such a necessary connection, and that the explanation must therefore be a conceptual one. However, I argued in Section 1 that it is not clear that all actions that can be appropriately characterized as intentional do involve a non-observational belief—the adverb is not sufficiently tied to the metaphysics of agency to guarantee this connection. Even if one is inclined to think that there is a reading of intentional action or acting “for a reason” that does require non-observational belief, this does not show that the belief could not be inferential; the presence of such a belief may simply be a condition on our classification of the action as intentional or done for a reason. To rule this out, we need non-question-begging grounds for thinking there is a relevant difference between the zoned-out driver or zoned-in cabinetmaker responding to their reasons, and their equivalents who do these things for reasons, other than the lack of belief in the former cases. Our answer to this will

\[ \phi \]

is generally agreed, in that “nothing shows (the agent) the position of his limbs; it is not as though he were going by a tingle in his knee … .”


40. Anscombe (1965), 50.

41. Anscombe considers and summarily rejects a theory in this vein— that what is known to the agent non-observationally is what he wills, while knowledge of the result is based on evidence in the ordinary way. She calls this a “mad account”, claiming to be able to make no sense of the notion of “willing” that is independent of what the agent actually does. But in fact, a natural way to make sense of the notion of willing is to think of it as the agent’s intention in acting. A great deal of work has been done on intention since Anscombe wrote this, making the possibility of drawing upon these resources to explain non-observational knowledge of intentional action much more promising than she envisioned.

---

Sarah K. Paul

How We Know What We’re Doing

Vol. 9, No. 4 (October 2009)
depend on the theory of action we ultimately accept, and this is part of what is at issue in here. I conclude that there is no clear basis on which to rule out in advance that the connection between intentional action and non-observational belief could not be inferential.

With these objections out of the way, let us return to Grice. Grice’s formulation of the puzzle of non-observational knowledge is slightly different from Anscombe’s, Velleman’s, or Setiya’s. Like Velleman and Setiya, Grice held that intending to \( \phi \) involves believing that you will \( \phi \), but he was a Weak Cognitivist — he did not think intentions just are beliefs, but he thought that believing one will \( \phi \) is part of being in the state of intending to \( \phi \). His chief reason for having this view was not the need to explain non-observational knowledge, however, but because he thought the assertion that one intends to \( \phi \), made strictly and honestly, implies that one believes one will \( \phi \), and not in a way that is possible to cancel. Saying “I intend to go to the theater on Tuesday, but I don’t believe I will go” merely sounds confused, rather than serving to cancel an inaccurate implicature. But as Davidson once pointed out, I think rightly, this appeal to linguistic practice is not a good reason to conclude that intending implies belief; it could simply be the asserting that implies believing, and not the intending.\(^{42}\) I agree with Davidson that Grice’s reason for being a Weak Cognitivist is not a good one, and for some of the reasons cited in Section 2, I think even Weak Cognitivism is better avoided. But we can still agree with Grice that an intention to \( \phi \) is usually accompanied by a doxastic commitment to one’s \( \phi \)-ing.

What is distinctive about Grice is that he was not persuaded that this doxastic commitment must be of a special groundless variety, and so set out to discover what the justification for it could be. His solution is a kind of dual-component theory of intention, in which intentions are the combination of a belief and a technical notion he dubs “willing.” As I understand the notion, willing is something like a conclusion about what would be the best thing to do, where assuming that this thing is in the agent’s power to do, he will be motivated to do that thing and nothing else incompatible with it.\(^{43}\) Grice proposed that an agent intends to \( \phi \) just in case he wills that he \( \phi \) and believes that he will \( \phi \) as a result of his willing it. The theory is what I call inferential because the agent’s belief that he will \( \phi \) is evidentially based on his knowledge of his will, plus the background knowledge provided by experience of what kinds of things he will do if he wills himself to, and his grounds for thinking nothing will prevent him from \( \phi \)-ing on this occasion. His knowledge of his will, on the other hand, is presumably not an evidential matter (though Grice did not attempt to explain what kind of matter it is). This dual-component theory of intention is designed to allow that the agent’s belief that he will \( \phi \) is justified in the ordinary way by his experiential evidence that he can \( \phi \) at will, while preserving the intuition that whether or not he \( \phi \)’s depends upon whether he decides to (wills that he) \( \phi \), and not merely upon theoretical evidence that he is going to end up \( \phi \)-ing.

The neo-Gricean theory I shall propose dispenses with the technical notion of “willing.” There is insufficient reason to make this rather artificial move if we simply reject Weak Cognitivism instead.\(^{44}\) If we think of intentions as distinctively practical, conative commitments that do not constitutively involve the belief that one will do what one intends, then we can replace the role of willing in the Gricean structure with intending. I want to remain neutral between specific versions of a Distinctive Practical Attitude (DPA) view of intention; the Inferential Theory should be more or less equally available to Davidson,\(^{45}\) Bratman,\(^{46}\) Alfred Mele (1992), and any other such view. For present purposes, we

\(^{42}\) Davidson (1978), 91.

\(^{43}\) Grice (1971), 277–278.

\(^{44}\) Note that it is natural for Weak Cognitivists to have an inferential story about non-observational knowledge, but inferentialists about non-observational knowledge need not be Weak Cognitivists.

\(^{45}\) According to Davidson (1978, especially 98–102), intentions are a kind of pro-attitude; they are an all-out judgment that a certain kind of future action would be desirable, given the rest of what the agent believes about the future.

\(^{46}\) Bratman (1987) thinks of intentions as practical attitudes defined by their functional role as elements in plans of action, aiding in intra- and inter-
can think of intentions quite generally in terms of the functional role they play as settled objectives that shape further practical thinking, initiate and sustain action, and at least usually engage certain norms of practical rationality. What is common to DPA views is the denial that the attitude that plays this role is itself a belief, or even entails the belief that one will do what one intends. But to succeed in bringing us to achieve our goals and to aid in intra- and inter-personal coordination at a time and over time, our intentions must be intimately tied up with our beliefs about ourselves and about the world. My hypothesis is that we achieve more or less as Grice suggested, by tending to believe, in the absence of evidence to the contrary, that we are doing what we intend to be doing.

The basic structure of the proposal is quasi-Gricean in that what the agent knows non-evidentially is what he intends to be doing, while insofar as he has a belief about what he is actually doing, it is evidentially based on his knowledge of what he intends, plus his evidence for thinking he is doing what he intends. I shall explain each of these components in turn, and clarify why an evidence-based belief can still count as non-observational. At the end of the paper, I will discuss a way of extending the basic picture, but first I will focus on explaining the core strategy.

4. Knowing what we intend

A challenge for the Inferential Theory is to establish that in core personal coordination over time, and subject to the norms that structure practical reasoning.

Talk of settled objectives can be traced back to Hugh McCann (1991). The term is involved in a debate between McCann and Bratman over whether there is a useful distinction between the state of having a settled objective and the state of intending. Bratman thinks, roughly, that intending is having a settled objective plus being subject to the rational constraints of consistency and means-end coherence, while McCann thinks that having a settled objective is already just intending, even if that means that it sometimes makes sense to violate those rational constraints with one’s intentions. In saying that intending “at least usually engages rational constraints,” I mean to avoid coming down on either side of that issue here.

Cases of intentional action, the evidential basis for non-observational knowledge will be in place, in the form of knowledge of what we intend to be doing. This is where Grice leaves off; he does not attempt to explain knowledge of one’s will. While I suggested earlier that breakdowns of belief are possible in cases we might describe as intentional, non-observational knowledge is nonetheless a strikingly pervasive phenomenon; breakdowns are the exception and not the rule. If the agent’s knowledge of what he is doing is inferential, we must be sure that the basis of the inference will dependably be available in the relevant circumstances. But our knowledge of our own minds is notoriously fallible and prone to error. We routinely misidentify the motives behind our decisions, agonize over trying to figure out what it is we want, spend years not knowing that we’re really in love, and so on. Why should we expect our knowledge of our intentions to be dependable and accurate enough to serve as the basis for non-observational knowledge of action?

The Inferential Theory is compatible with different positive answers to this question. One may insert one’s favored theory of how we know what we intend here, provided it does the work of establishing that we generally know what we intend to be doing when we go to act intentionally. Those who are willing to take for granted that we usually know what we intend will not see this as a problem, but others will find that assumption unsatisfying. I will sketch my own preferred view of the connection here, but the overall structure of the Inferential Theory does not depend upon accepting my view of why we generally should not fail to know what we intend to be doing when we act intentionally.

On my view, the key idea is that intentional action and knowledge of what one intends are connected, by the mental act of making a decision about what to do. Most cases of action that count as intentional will involve something the agent decided on doing, and self-ascribing an intention on the basis of making a decision about what to do is a

---

47. Talk of settled objectives can be traced back to Hugh McCann (1991). The term is involved in a debate between McCann and Bratman over whether there is a useful distinction between the state of having a settled objective and the state of intending. Bratman thinks, roughly, that intending is having a settled objective plus being subject to the rational constraints of consistency and means-end coherence, while McCann thinks that having a settled objective is already just intending, even if that means that it sometimes makes sense to violate those rational constraints with one’s intentions. In saying that intending “at least usually engages rational constraints,” I mean to avoid coming down on either side of that issue here.

fundamental way of knowing what we intend. Neither claim is one I can fully argue for here. With respect to the former claim, I think the relationship between intentional action and decision is fairly intuitive. If the agent never decided to perform a certain action, or to do something which he foresaw would lead to or constitute that action, that fact would weigh heavily against its counting as an intentional action of his (though again, this may not be a decisive consideration). This is certainly an oversimplification that ignores borderline cases, but this much seems plausible to me: the core cases of full-blooded intentional action we are interested in, in which the agent knows full well what he is doing, will be ones in which some relevant decision to act has been made.

The second idea is that we can know what we intend on the basis of knowing what we have decided to do. This is because making a decision to φ is ordinarily sufficient for coming to intend to φ. This claim is also not especially controversial, but it is not obvious. Decisions are conscious, relatively discrete mental acts, whereas intentions are complex, diachronic, heavily dispositional mental states. Nevertheless, I submit that we usually come to be in the state of intending by “making up our minds” — by making a decision about what to do — in a way we cannot not simply make up our minds with respect to what we want or feel. We perform the mental act of making a decision in order to settle the question for ourselves of what to do, by committing ourselves to planning on and around that action, deciding what means to take, deciding on nothing else incompatible with that action, and initiating the action when the time comes. In other words, in deciding to φ, we undertake to shape our thought in action in just those ways characteristic of being in the state of intending to φ.

Not that we can simply decide on any action at all and count as intending it. As Harry Frankfurt (1999, 101) has pointed out, an agent can call into the “vasty deeps” of himself for his will to be a certain way, but it is not always the case that calling will make it so “when the chips are down.” There are constraints on what one can intend by deciding on it; one must believe it is in one’s power to perform the action if one decides to, and one’s motivations must be at least somewhat in line with so acting. On the other hand, I do not think it is a general constraint that we must decide what to do in light of what we take ourselves to have sufficient practical reason to do. We commonly decide to do things we know to be at odds with our best reasons, or when our reasons are incommensurable or indeterminate, in light of no reason whatever. As I am thinking of it, then, the ability to decide what to do is not to be identified with practical judgment.

Because making a decision about what to do is a way of forming an intention — I venture to say the fundamental way — we can know at any point what we intend either by making a decision about what to do, or by recalling a previous decision, and self-ascribing an intention on the basis of that decision. We take ourselves to intend the actions we have decided on, and we are entitled to do so. We are entitled because making a decision to φ is normally sufficient for intending to φ, and because this fact is part of the very concepts of intention and decision. A thinker who possesses these concepts, however implicitly, will be disposed and willing to ascribe to himself an intention to φ on the basis of having decided to φ, when the question of what he intends arises. Someone who knew he had decided to φ, but considered it a serious open question whether he intended to φ, would be misapplying.  

This is in contrast with, e.g., Nishi Shah’s (2008) argument claiming that the explicit deliberative question of whether to φ must be settled by answering the normative question of whether one ought to φ.

I am taking for granted here that decisions are conscious mental acts, and that one necessarily knows at the time of deciding what it is one has decided on. There may be further questions about how exactly one knows of a mental act that it is a decision, but I suggest that knowledge of a conscious mental act that happens at a time is much less problematic than knowledge that one is in a diachronic state defined by complex dispositional and normative roles.

I have been influenced here by Christopher Peacocke’s (2003) account of consciously-based self-ascription of belief.

49. I do argue for these claims in my (ms.).
50. Here I make a broad appeal to analyses of the functional role of intention along the lines of Bratman (1987), although I do not mean anything to depend on the precise details of this role.
51. 52. This is in contrast with, e.g., Nishi Shah’s (2008) argument claiming that the explicit deliberative question of whether to φ must be settled by answering the normative question of whether one ought to φ.
53. I am taking for granted here that decisions are conscious mental acts, and that one necessarily knows at the time of deciding what it is one has decided on. There may be further questions about how exactly one knows of a mental act that it is a decision, but I suggest that knowledge of a conscious mental act that happens at a time is much less problematic than knowledge that one is in a diachronic state defined by complex dispositional and normative roles.
53. I have been influenced here by Christopher Peacocke’s (2003) account of consciously-based self-ascription of belief.
ing the concepts of decision and intention— at least in the absence of evidence of an abnormal psychological breakdown between deciding and putting an intention in place.

I have not really argued for any of these claims, though I do elsewhere. My object here is merely to sketch a picture of how one’s intentional actions could be connected with knowledge of one’s intentions in such a way as render uncommon a situation in which an agent is acting intentionally but does not know what he intends to be doing. On the picture I have offered, this situation will be rare because of the role of decision in intermediating between intentional action and knowledge of one’s intentions. If the agent is acting intentionally, he will usually have decided at some point to act, and he can know on the basis of that decision what he intends to be doing. It is not that we must always refer back to the original decision in order to know of our intentions; we employ a variety of shortcuts in keeping track of our goals. Most importantly, we often simply keep track of our reasons for action, or of habitual patterns we have, without referring to anything mental. My suggestion is not that we must always keep track of our decisions in order to know what we intend, but rather that the option is there when other shortcuts fail. And if one does not know what one intends, one may always simply decide anew.

This certainly leaves room for a variety of breakdowns—in knowing whether you have decided, between deciding and self-ascripting an intention, and between deciding and knowing what you intend when you go to act. And I think this is in fact what we find; we do find ourselves uncertain whether we have really decided to φ, or making decisions that leave no mark. But when we fail to know what we have decided, it seems to me that we also lack knowledge of what we intend, and insofar as we find ourselves acting, we don’t know what we are doing. The point is not these connections are perfect, but rather that the elements of decision, intention, self-knowledge, and action are interdependent. But I do think the connection is ordinarily sufficiently tight to explain the pervasiveness of non-observational knowledge of action.

5. From knowledge of intention to knowledge of action

Whether or not one accepts the foregoing picture, let us now simply take for granted that in most cases of full-blooded intentional action, we know what we intend to be doing. This is merely knowledge of the content of one’s own mind, and does not yet amount to knowledge of what one is actually doing. The question now at hand concerns what, if anything, the agent knows beyond his own mind when he is acting intentionally. The claim of the neo-Gricean inferential model I am proposing is that the agent will tend to believe, partly on the basis of his intention to φ, that he is in fact φ-ing. I say “partly” because insofar as he is rational, his belief that he is φ-ing will also depend upon several other epistemic factors. The following is not an exhaustive list of these factors, but I will suggest some plausible candidates. I will then explain why these evidence-based beliefs should nonetheless count as non-observational in the relevant sense.

55. In this section, I will mostly talk about the agent’s forming an all-out belief about what he is doing. This is an over-simplification; a more sophisticated picture would talk in terms of degrees of confidence, allowing that we will sometimes experience a non-observational increase in confidence that we are φ-ing when we intend to be, without fully believing that we are φ-ing.

56. I should emphasize that what I am describing here will not necessarily be the agent’s actual thought process in any given instance of non-observational belief formation. There are many pieces of background knowledge the agent must have in order for his inference from intention to action to be justified. This does not mean that he actually rehearses all of these premises in a given case; these resources must simply be available to him to cite if pressed. We must therefore distinguish between the project of characterizing the subterranean structure of the inference, and the project of describing the agent’s actual thought pattern. My focus here is primarily on the former project, since the epistemic status of the agent’s belief depends upon this deep structure and not just upon the often abbreviated explicit reasoning he goes through. Thus, if the picture I am presenting appears excessively baroque, recall that the agent will not necessarily go through all the steps that constitute the structure of the inference.
One relevant factor will be the agent’s background belief in his ability to perform the action(s) he intends. It would surely be an unlicensed leap of faith for him to believe that he is φ-ing just because he intends to, if he doesn’t also have good reason to believe he can φ if he intends to. This will be a matter of his experience as an agent, coming to know what his capabilities are. As Grice puts the point, the agent will know from experience that he can raise his arm at will but cannot raise his hair at will.\(^{57}\) For most basic actions, the agent will know whether or not he can perform them at will — though perhaps if it is a movement he has not made for years, he may not know whether he can still perform it if he intends to. But most basic actions are performed for the sake of achieving some further end. Take an agent raising his arm in order to aim his rifle at the target and shoot the bulls-eye. He will most likely believe that he is raising his arm and firing the rifle, but whether or not he believes non-observationally that he is hitting the bulls-eye will depend upon his confidence that he can consistently hit it at will. In general, the agent’s confidence that he is doing what he intends will vary under different descriptions of the action, and will usually decrease the further away an action description gets from his bodily movements and the more it depends upon a causal chain working out in the way he plans.

A second factor operant in the background of the belief formation will be the agent’s background knowledge of his circumstances as being conducive to his φ-ing — or at least, the absence of reasons to believe his φ-ing will be obstructed. He will not be justified in believing he is shooting the bulls-eye if he has reason to doubt that his gun is in working order, or that he has an unobstructed shot, and so forth. Where an action-type counts as such in virtue of a convention, he must know that the relevant convention applies in these circumstances, and if his action-plan depends upon a causal chain, his confidence that this causal chain will succeed should be reflected in his belief. If his intention is to hit the bulls-eye by firing the gun at it, he must understand what to do to cause the bullet to hit the target, and not have reason to believe that he is likely to be prevented. Insofar as one is rational, most of this background knowledge of ability and circumstance will be brought to bear in the formation of the intention or plan, and so should already be in place as the intention is executed.

Foreknowledge of what one will do in the future adds additional complication. To have it, one must have reason to believe one will follow through on one’s intention — that one will not change one’s mind. This will depend, for one thing, on one’s history as an agent; a highly stable agent can confidently predict that he will follow through on this occasion, whereas a flaky agent may not be able to do so. The strength of the agent’s reasons as he sees them may also play a role. Even if one tends in general to be quite unreliable, one might take oneself to have very strong reason in this case to perform the intended action, and so rationally predict on that basis that one will follow through on this intention.\(^{58}\) In contrast, if one takes oneself to have only very weak reason to perform the action, and no especially strong record of following through on intentions when the stakes are low, one has much weaker grounds to predict one will act on that intention. The kind of action may also be relevant. One might be very reliable when it comes to paper deadlines, but very unreliable in meeting social obligations. Knowing this, such an agent will be in a position to have intention-based foreknowledge that he will submit a paper to a conference next Tuesday, while not in a position to know that he will be attending a party later that night, though he knows he now intends both of these things. Again, this list is not exhaustive, but it is indicative of the considerations relevant specifically to foreknowledge of action.

But knowledge of ability, conduciveness of circumstances, and one’s history as an agent are things one knows from experience and observation. It is tempting to conclude from this that all we know non-observationally is what we intend to be doing, and that any knowledge of the results of our actions is based on observation. This view

\(^{57}\) Grice (1971), 276.

would be what Kevin Falvey calls the “two-factor thesis.” Although the two-factor thesis avoids the mysterious claim that we can know what is happening in the world without looking, it is also a disappointing deflation of Anscombe’s thesis. She contends that “when the description of what happens is the very thing which I should say I was doing, then there is no distinction between my doing and the thing’s happening.”

Strong Cognitivism purports to capture the force of Anscombe’s fascinating but cryptic claim, and for the Inferential Theory to do the same, rather than merely explaining it away, it needs to clarify why it is not simply the two-factor thesis. In what sense does what the agent knows without observation extend beyond his own mind to what actually happens?

Part of the answer is simply that the experience and observation on which the agent’s belief is partly based is not experience or observation of the particular action in question. He has background knowledge from experience that he can hit the target at will, but his belief does not depend on waiting to see where the bullet goes on this occasion. Much more interestingly, though, there is a certain feature of the kind of descriptions under which we expect the agent to have non-observational knowledge of his actions, such that knowing what you intend to be doing also largely amounts to knowing what you are doing, under that description. That is, many action descriptions function in such a way as to apply truly in large part because that description is what the agent intends.

The act-descriptions that have this characteristic are (in English at least) a subset of those that take the progressive form — “is walking,” “is writing a paper,” “is eating,” and so forth. Descriptions of this form denote actions that are in progress and that take place over some amount of time. Here, following Comrie (1976), we can further distinguish between “telic” and “atelic” actions. Telic actions have built into them a terminal point, at which the action is complete; this is the case of writing a paper, which has built into it the endpoint of completing the paper. Atelic actions have no such endemic endpoint and can be continued indefinitely or broken off at any point; this is the case of walking and eating.

The relevance of this distinction to this discussion is that telic and atelic progressives differ with respect to the logical deductions they license. Atelic actions described in the progressive tense entail the truth of statements about the same action in the past-tense, perfective aspect. That is, if it is true that Jen is or has been walking, then it is also true that Jen walked/has walked. Progressive-tense descriptions of telic actions, in contrast, do not entail the truth of the past-tense perfective description. The truth of the statement ‘Jen is crossing the street’ does not entail that Jen crossed the street; she might turn around and go back, or she might get run over by a truck, without its ceasing to be true that she was crossing the street.

I suggest that both of these kinds of progressive-tense act-descriptions help us see why the agent needn’t observe himself in action to know what he is doing, under these descriptions. In the case of atelic progressives, if the agent is or has been doing them at all, it is true that he has done them; there is no further goal he must reach in order to complete the action apart from initiating it. And in the case of telic progressives, the agent can count as so acting even if the goal of the activity is never successfully achieved. As Anscombe puts the point, a man can be doing something which he nevertheless does not do (as when Jen gets killed halfway through crossing the street). And given that the agent need not ever achieve some terminal endpoint in order to count as doing a telic action, he also need not have observational evidence that he is succeeding to know he is doing it.

Rather, I suggest, the central condition for him to count as doing

\[62. \text{§2.2} \]

\[63. \text{Telic and atelic act-descriptions are both opposed to stative descriptions, which do not take the progressive form at all—e.g., ‘know’, ‘love’, ‘believe’}. \]

\[64. \text{Anscombe (1963), 39.} \]

---

59. Falvey (2000), 21
60. Anscombe (1973), 52–53.
61. This point illustrates the distinction between non-inferential knowledge and non-observational knowledge, since I want to say that the agent has inferential but non-observational knowledge of what he is doing, in the paradigm case.
an action of this sort is that it is what he intends to be doing. For it to be true that Jen is walking to the philosophy department (and not to the library), the central condition is that she is walking with the intention of getting to the department. There are some other minimal conditions that are difficult to specify precisely; perhaps she must have already taken a few steps and believe she knows how to get to the department. On the other hand, it can be true that she is walking to the department even if she is simply standing at a red light at the time, or if she is actually walking in the wrong direction, as long as she is disposed to correct herself when she discovers her error. All observable evidence available might fail to support the claim that Jen is walking to the department, and yet — perhaps within some limits of plausibility of success — it can nonetheless be true that she is, under the progressive description, if she intends to be doing so.65

These features of many progressive-tense act-descriptions make it possible for the agent’s knowledge of his intention to amount to knowledge of what is actually happening, given some minimal further conditions that mark the boundaries of possibility of success. Insofar as the agent’s belief about his action takes this kind of description, that belief can be based on little more than his knowledge of his intention in acting and the aforementioned background knowledge of ability, for he need not have evidence that he is actually succeeding in his intended aim — unless it is an aletic act, in which case he has

Falvey (2000) makes a similar point about the “broadness” and “openness” of the progressive tense. To be clear, this is a point about the truth-conditions of act-descriptions and not directly about the concept of intention. However, one might worry that in our practice of deferring to the agent’s intention (defeasibly) in applying progressive-tense act-descriptions to an episode of behavior, we are thinking of intention in a certain way, and perhaps a way that is incompatible with what I have said here. I see no reason to think this is so. I have suggested that 1) intentions are distinctively practical attitudes: they are not beliefs, nor do they entail the belief that one will do what one intends; 2) that insofar as an agent has a non-observational belief about what he is doing, that belief is evidentially based on his knowledge of what he intends to be doing; and 3) intentions tend to cause the initiation of the intended action. As far as I can see, none of these claims are incompatible with the role I have suggested intentions play as a defeasible truthmaker of progressive-tense act-descriptions.

How We Know What We’re Doing

succeeded as soon as he has begun. And in fact, the knowledge we expect agents to have of their actions takes precisely this progressive form, and concerns these kinds of actions. The examples given in the literature standardly concern the agent’s knowledge of what he is now doing — painting the wall yellow, writing the word ‘action’, pumping water, and so forth.66

This is no coincidence, for we would be much less inclined to take seriously an agent whose non-observational belief takes a description that does entail success, without checking to see that he has really pulled it off. We grant that the agent can know he is writing the word ‘action’ on the board at time $t$ without looking to see that it is really happening, because it can be true that he is writing it at $t$ even if his marker fails at $t$. Contrast this with the agent’s claim that he knows without looking, say, that he is teaching his students how to spell the word ‘action’. We would be much less willing to grant that he knows non-observationally that he is doing this, for he cannot have been teaching his students to spell ‘action’ unless he succeeds in getting it done.67 However, as long as what the agent believes he is doing is described in such a way that he need not succeed in order to be doing it, he need have little or no evidence that he is succeeding. This, then, is a further respect in which the Inferential Theory can maintain that the agent has non-observational knowledge of what happens and not just of what he is aiming at.68

In summary, we have the contours of a model on which, when the

65. The first and third are from Anscombe, and the second from Davidson (1978).

66. Again, this is not to say that the agent’s non-observational knowledge never extends beyond this progressive-tense description; if he has sufficient evidence (the evidence of experience that he consistently succeeds in teaching his students to spell using this method, perhaps), he is licensed to infer that this is indeed what he is doing.

67. Of course, all I have said here is also available to the Strong Cognitivist and is not meant to distinguish one theory from the other. The overall point is simply that both stories can acknowledge that our non-observational knowledge extends beyond our own minds to what actually happens — though only the Inferential Theory can maintain that this knowledge is based on sufficient prior evidence in the ordinary way.
agent knows that he intends now to $\phi$, he will tend to believe he is $\phi$-ing, and not on the basis of observation of the action in question. Among the factors upon which this belief does depend, I have suggested, are: a) the agent’s confidence that he has the abilities required by his plan; b) his knowledge of his circumstances, including his understanding of the applicable conventions, the obtaining of the conditions necessary for his action to count as $\phi$-ing, and his understanding of the cause-and-effect relationships his plan will be exploiting; c) for foreknowledge of action, his history as an agent who reliably follows through on his intentions, in general and with respect to actions of the type in question; and d) his understanding of the way some action descriptions apply partly in virtue of his intention in acting. The last consideration in particular makes it possible for him to know what is actually happening, under a certain description, in knowing what he intends to be doing. 69

69. One further thing the agent might be required to know in order to have non-observational knowledge of what he is doing is that his intention is what is moving him to act. For instance, Daniel Wegner (2002) argues that the feeling of consciously willing is not the direct cause of our actions. If the agent does not believe his intentions will be effective in moving him, his knowledge of what he intends to be doing cannot justify a belief about what he is actually doing. This is a point that Setiya and Falvey do not address at all. Velleman does have a story about it, involving the agent’s understanding of being motivated by the desire for self-understanding to fulfill his expectations of acting, once he has formed them. As I see it, this question is complicated by the need to remain neutral about issues that are properly the domain of empirical science—issues about whether intentions are physical states in the brain that cause behavior, and so forth. It is important that our ability to have non-observational knowledge of our intentional actions does not depend upon the science turning out in any particular way. What I think is that the presupposition that your intentions will be effective in moving you to act is built into the very concept of a present-directed intention, such that if the agent ascribes to himself an intention now to $\phi$, he thereby presupposes that this intention will move him to $\phi$. Someone who entertained the thought “I intend to raise my arm and vote yes in just a second, but I don’t know if I will even try to raise my arm” would simply not be deploying the concept of intention in the correct way. It seems to me that we have a default epistemic warrant to presuppose that our intentions will move us to act, but I cannot argue for this here.

6. Three threads

In this section, I want to pick up on three threads that were earlier left dangling. I left promissory notes about 1) whether non-observational knowledge of intentional action is in fact the cause of what it represents; 2) the place of Weak Cognitivist theories of intention in the logical space of the debate; and 3) the puzzle about non-observational knowledge of foreseen but unintended side effects of intended actions. I will discharge each of these notes in turn.

Non-observational knowledge — the cause of what it understands?

A consequence of the Inferential Theory of non-observational knowledge is the rejection of the interpretation of practical knowledge as being the efficient cause of what it represents. On this account, the belief itself does not bring about its object; it is evidentially based on the agent’s intention in acting, which is (presumably) what brings about the believed-in action. To be more precise, there are at least two roles the intention plays with respect to the truth of the agent’s belief about what he is doing. First, it is the intention (we presume) that brings the agent to initiate the intended action and guides him through to its completion, in the successful case. Secondly, because of the nature of performance-verbs expressed in the progressive tense, the agent’s intention is a powerful determining factor in whether his behavior indeed comes under the intended description (given other background conditions, as I have argued). Thus, according to the Inferential Theory, the agent’s non-observational belief about what he is doing is based on the cause of its truth, but is not itself the efficient cause. This is a major difference from the Strong Cognitivist views of Velleman and Setiya.

There is a way in which it is still a special, practical kind of knowledge, however. It is not merely speculative, formed in response to antecedent evidence about how the world is; rather, it is tracking the effect of one’s own mind upon the world. The truth of the belief depends on the agent’s choice to make it true, by carrying out the intention on which the belief is based, and not on its being an accurate reflection
of the way things already stand. The belief is not the cause of what it represents, but the agent is, and I think in this sense it still deserves to be characterized as “practical”. Since I have reserved the term ‘practical knowledge’ for Aquinas’s notion of knowledge that itself causes what it represents, however, we might instead call our intention-based knowledge “agential” knowledge — knowledge we attain by bringing the world to conform to an idea we have of it.

The advantage of denying that non-observational beliefs about our actions cause the truth of what they represent is that the Inferential Theory, unlike Strong Cognitivism, can thereby avoid the claim that in forming intentions, we must jump to conclusions that are insufficiently supported by prior evidence. As I have emphasized, at least some may find it unsettling to think that such doxastic leaps lie at the heart of human agency. We are not divine agents; there is a gap for us between intending and doing, and our beliefs ought to be sensitive to this gap without thereby preventing us from intending at all. By holding that agential knowledge is evidentially based on one’s intention, together with one’s reasons to think the intention will be successful, the Inferential Theory precludes the need to posit a widespread exception in the case of intention for justified believing. We can explain agential knowledge with a simple, familiar epistemic structure, without the need to suppose that the practical and the theoretical must be fused in surprising ways.

Weak Cognitivism

In Section 2, I drew a distinction between Strong and Weak Cognitivism about intention. I defined Weak Cognitivism as the claim that intending to φ partly involves believing you will φ, but also consists in some other separable component. Grice is the paradigmatic Weak Cognitivist, since on his view, intending to φ is a combination of two attitudes: willing that φ, and believing that one will φ as a result of willing it, where willing that φ can occur without any such belief.\(^70\) This

is as opposed to the Strong Cognitivist idea that intentions just are a kind of belief — in Setiya’s case, a belief that has special motivational properties, but nonetheless an attitude that cannot be broken down into distinct conative and cognitive components. I have focused on Strong Cognitivism because both Velleman and Setiya appeal to this kind of view as a way of explaining Anscombe’s Non-Observational Knowledge Thesis, but this may strike us as an artifact; although these two prefer Strong Cognitivist theories, wouldn’t Weak Cognitivism do just as well? It might seem that all we need for this kind of explanation to work is that believing is a part of intending, and not necessarily the entirety of it.

There is a sense in which this line of thought is on the right track, although I think ultimately misguided. Exploring this idea turns out to be revealing of the relationship between the different kinds of views on the table. Weak Cognitivist views claim that intentions are “beliefs plus” — believing you will φ plus some other component, presumably conative. Once we have this dual-component structure, there are two possibilities for this type of view with respect to non-observational knowledge of intentional action: an inferential version and a non-inferential one. Either the belief component is evidentially based on the conative component, at least in part, or it is not.

Now, the inferential version of Weak Cognitivism is precisely what Grice proposes. On his view, the conative component is willing, and the belief component is evidentially based on one’s knowledge of one’s will. I have argued here that this kind of inferentialism is indeed a viable strategy for explaining non-observational knowledge of intentional action, but that the appeal to willing is unnecessary; intending can play the role Grice accords to willing, and the belief can be evidentially based on the agent’s intention without purporting to be a part of the intention itself. As just noted, this kind of theory gives up on Anscombe’s idea that the belief is the cause of its own truth, but it has the

\(^70\) Wayne Davis (1984) also has a Weak Cognitivist view of intending, on which intending that φ consists in desiring that φ and believing that one will φ for the reason that his desire that φ will motivate him to act in such a way that φ.
advantage of maintaining that it is justified in the ordinary way, rather than an exception to the standard rules of rational belief formation.

There are no prominent defenders of a non-inferential brand of Weak Cognitivism, and I think this is no coincidence. Such a theory would hold that intentions are “beliefs plus,” but would deny that the conative component plays any justifying role in the agent’s belief that he will φ. This would be quite an odd view, since the belief would be neither based on evidence nor self-fulfilling. That is, presuming that the conative component is necessary for the agent’s φ-ing to occur, the belief component would not be independently sufficient to bring the φ-ing about, and yet it also would not be evidentially based on a necessary condition for its truth. This would leave the belief quite ill-supported indeed. Essentially, what this shows is that once we make the step of positing a belief component of intending that does not require sufficient antecedent evidence for its truth, we are strongly pushed to go all the way to Strong Cognitivism and say that the belief is the intention, so that its justification can be claimed to lie in its self-fulfilling nature.

Thus, the middle ground Weak Cognitivism may appear to offer between Strong Cognitivism and an inferential model turns out to be an unhappy one. We are either pushed in the direction of Grice’s dual-component inferentialism, which is subject to the same counterexamples against intention involving belief by which Strong Cognivist theories are challenged, or we are pushed all the way back to Strong Cognitivism itself. I have argued here that an Inferential Theory that avoids cognitivism about intention altogether, strong or weak, is preferable to either of these options.

Foreseen but unintended side effects

The third strand I wish to resume here is the puzzle concerning non-observational knowledge of bringing about the expected but unintended side effects of one’s intended actions. As discussed, we seem to have a similar kind of non-observational knowledge of bringing about these unintended side effects to that we have of performing the intended action itself. It is not clear what the Strong Cognivist theorist could or would want to say about this, since according to them, what the agent knows non-observationally is limited to what is embodied in the content of his intention.

I argued in Section 2 that expanding intentions to include all of the foreseen side effects would be too consequental a move for a theory of intention to make merely in order to account for the full extent of non-observational knowledge. The second option I see available to the Strong Cognivist would be to insist that any knowledge we have of bringing about unintended side effects is a fundamentally different kind of knowledge than that we have of our intended actions — that it is theoretical or “speculative” rather than practical. That is, the Strong Cognivist might gesture to the “foreseen” part of “foreseen but unintended consequences” and rest easy with pointing out that the Strategic Bomber may not intend to bring about the deaths of the children, but he already believes he will do so if he drops the bombs on the factory. He knows practically that he is bombing the factory, but he reasons his way theoretically to the conclusion that he is thereby bringing about the deaths of some children.

This solution entails that knowledge of bringing about expected side effects of our intended actions has precisely the same inferential structure I am proposing all of non-observational knowledge of action has, in that the chief grounds for that knowledge will be the agent’s knowledge of his intention. After all, merely believing that children die if they are hit with bombs is not the same as believing that one is bringing about the deaths of some children — it is not a belief about

\[^{71}\text{This does not rule out the belief component of the intention being a cause of the intended action, in conjunction with the practical component. Such a view could not appeal to the belief as being self-justifying, however, since the practical component would presumably also be necessary for the belief to come true.}\]

\[^{72}\text{Although Velleman may be independently committed to this claim. Bratman (1991) criticizes Velleman’s theory as being “promiscuous,” in that it appears to treat anything the agent foresees he will bring about as part of what he intends to do.}\]
what one’s own agency is amounting to. To reason to the latter belief, the Strategic Bomber’s grounds must include his intention to bomb the factory, for otherwise he would not have sufficient reason to believe that he is acting in a way that will cause the deaths. This proposal would thus end up claiming that non-observational knowledge of intended actions is spontaneous, while any non-observational knowledge of bringing about unintended consequences is inferential. This would be a divided, two-part theory of non-observational knowledge of agency.

The Inferential Theory, on the other hand, has the resources to provide a unified account of agential knowledge of intentional but unintended side effects. To do this we must extend the basic theory somewhat. In Section 4, I suggested that we often know what we intend to do because we have deliberated, reflected on reasons for various actions, and resolved in favor of a given course of action. In particular, this kind of deliberation and self-ascription of the resolved-upon action as one’s intention often occurs in the core cases of full-blooded agency in which we expect the agent to have non-observational knowledge. But normally, if the agent is a responsible deliberator, he will not consider his aim in isolation; he will also reflect on what he takes to be the likely side effects of the courses of action he is contemplating. Before deciding to bomb the factory, if the Strategic Bomber is not a rash and heedless agent, he will weigh in his deliberation the fact that in so acting, he is also likely to bring about damage to the school and kill some children.73 In the process of deliberating and reflecting on various courses of action, the agent will therefore be led to form a set of beliefs about the likely side effects of each of the plans he considers.

Following Bratman (1987), we can think of the combination of the potential action and the anticipated side-effects as a “scenario,” and note that the agent’s choice of a certain action is in some sense a choice in favor of the whole scenario he foresees.74 This does not entail that he intends the whole package — the bomber does not intend to kill the children, because he is not tracking this outcome with his efforts or means-end reasoning — but it does mean that the side-effects he foresees are included in the conclusion of his deliberation. I will call this whole package what the agent is “practically committed” to, where this includes both what he actually intends and the foreseen consequences he seriously considers in his deliberation.

The significance of this whole-package view of the conclusion of practical deliberation is that the evidential basis of the agent’s non-observational belief is actually his entire practical commitment, and not merely the part he intends. He will be disposed to believe that he is bringing about the entire scenario he plumped for, and not only that part at which he is actually aiming (though he may consider some parts of the envisioned scenario more likely than other parts). Therefore, in cases where the agent has considered a side effect in his deliberation, and made a practical commitment in favor of the scenario as a whole, the same inferential process I have been describing should also lead him to believe he is bringing about the side effects. The Strategic Bomber will infer from his practical commitment not only that he is on his way to bomb the factory, but also that he is on his way to bring about the deaths of some children.

In fact, I think this closer look at the agent’s deliberative process reveals why some are inclined to consider knowledge concerning side effects different from knowledge of intended actions. Deliberation about what to aim at is predominantly a matter of practical reasoning — weighing reasons for and against courses of action — while figuring out the likely side effects of those courses of action will be a matter of theoretical reasoning. This is an important difference, but it is a difference relevant to the deliberative construction of a scenario, not to one’s knowledge of bringing that scenario about. Once a course of action is chosen, one’s knowledge of bringing about that scenario is uniformly agential, in the sense I articulated above: it is based on one’s choice to make it true that these are the things one is doing, and not on antecedent evidence for the truth or probability of these events.

The Inferential Theory can thus be extended in a natural way to

---

73. As discussed by Bratman (1987), § 10.
provide a unified account of this special knowledge of our actions, whether intended or merely foreseen. The most plausible option available to the Strong Cognitivist, on the other hand, is the two-part account on which knowledge of unintended side effects is inferential, while knowledge of intended actions is spontaneous. It is my view that once it is admitted that one kind of non-observational knowledge is evidence-based, there is theoretical pressure to move toward the unified structure proposed by the Inferential Theory, on which it is all of this kind.

7. Conclusion

Many more details could be added to this inferential story of non-observational knowledge of intentional action, but I think enough has been said here to indicate some of the advantages of the theory. Most significantly, it neutralizes one of the major motivations for having a Strong Cognitivist theory of intention. The purported ability of Strong Cognitivism to explain non-observational knowledge has been touted as an advantage of this kind of view, while the DPA camp has said very little about it (perhaps out of Davidsonian skepticism). The Inferential Theory provides DPA-type views with a story to tell about non-observational knowledge of action, showing how they can take the connection between intentional action and non-observational knowledge seriously while retaining the flexibility to allow for breakdowns in belief at the margins of intentional action.

Second, more than merely leveling the playing field between Strong Cognitivism and DPA theories with respect to non-observational knowledge, I have argued that the Inferential Theory is better equipped to account for non-observational knowledge of the bringing about of unintended side effects. Unlike the options I can see for Strong Cognitivism, there is a natural and straightforward way to extend the Inferential Theory to account for non-observational knowledge of bringing about effects you foresee but are not aiming at. Since our knowledge of bringing about these effects is agential — grounded in our choice to make it true that we are bringing them about, rather than a post hoc apprehension of the facts — it seems to me that a complete theory of non-observational knowledge of what we do intentionally ought to account for it.

This point highlights a significant difference between my view and those of Anscombe, Velleman, and Setiya, however. Not only am I proposing that non-observational knowledge of intentional action is inferential rather than spontaneous or immediate, but I am also denying that it is the cause of what it represents in any straightforward sense. This might be a divergence from Anscombe’s original characterization of the phenomenon, but the advantage is that the Inferential Theory is thereby able to avoid making an exception in the rational requirements on belief formation for beliefs about your own intentional actions. The Inferential Theory is a demystification of the epistemic structure of non-observational knowledge; it maintains that this knowledge is based on sufficient prior evidence in the way ordinarily required by theoretical rationality.

Finally, the Inferential Theory can still acknowledge and incorporate one of the fundamental insights of its Strong Cognitivist rivals: that we are creatures who are deeply motivated to know what we are doing. This is, of course, the guiding hypothesis behind Velleman’s well-known theory of action. I do not mean here to endorse Velleman’s claim that the desire for self-understanding “plays the role of the agent,”75 but I do think that it is highly plausible that we have a strong disposition to maintain awareness of what we are doing, in a way that anticipates the lagging deliveries of sense evidence. This insight helps the Inferential Theory explain why we would so consistently and automatically form beliefs about our actions, even in advance of observational confirmation. We desire to know what we are about, and I hope to have motivated the idea that we achieve this in part by inferring from our intentions.76

75. (2000b), 138.

76. Versions of this article were presented and insightfully discussed at Stanford University, UC Berkeley, and the University of Southern California. I am especially indebted to Michael Bratman for many invaluable conversations and comments on earlier drafts. My thanks also to Krista Lawlor, Nadeem Hussain, John Perry, Lanier Anderson, Kieran Setiya, Luca Ferrero, and two anonymous referees at Philosophers’ Imprint for extremely helpful comments.
References


___________ (ms.). “How We Know What We Intend.”


“Symposium on J. David Velleman’s The Possibility of Practical Reason.”
