

## INTENTION AND ACTION

*We've gone on holiday by mistake!*  
—Withnail

Some significant phenomena:

- *Prospective intention*: S intends to A later
- *Intention-in-action*: S intends to continue (or finish) A-ing
- *Intention-to-act* or *intention*: S intends to A [neutral as to when]
- *Intention-with-which*: In A-ing, S intended to B
- *Intentional action*: S is A-ing intentionally
- *Action*: S is A-ing
- *Unintentional action*: S is A-ing and not A-ing intentionally
- *Behaviour*: S is “A-ing” [performing the motions that go with A-ing]
- *Mere behaviour*: S is “A-ing” and not A-ing

The major players in our discussion are underlined.

Some examples:

It is now 3 o'clock; Joan gets off work at 5. She plans to detour to shop for groceries on the way home after work. At 5:15 she is taking a slight detour down Prill Avenue from the most direct route home because the grocery is down that street.

- a. Joan's prospective intention at 3 is to detour to shop for groceries on the way home after work
- b. Joan's intention-in-action at 5:15 is to continue detouring to shop for groceries on the way home from work
- c. Each of these is among Joan's intentions
- d. The intention-with-which Joan is walking down Prill Avenue at 5:15 is to *detour to shop for groceries on the way home from work*: “in walking down Prill Avenue at 5:15, she intended to ...”
- e. Joan's intentional actions at 5:15 include *intentionally detouring to shop for groceries on the way home from work* and *walking down Prill Avenue*
- f. Joan's actions at 5:15 include *detouring to shop for groceries on the way home from work* and *walking down Prill Avenue*
- g. Nothing in the case suggests that Joan is doing anything in particular unintentionally
- h. Joan's behaviour at 5:15 includes “*detouring to shop for groceries on the way home from work*” and “*walking down Prill Avenue*”
- i. Joan's mere behaviour at 5:15 doesn't include either of these, but does include things like maybe “flattening bugs on the sidewalk of Prill Avenue”

### I. Questions in the philosophy of action

1. *The unity of intention*  
Anscombe:

it is implausible to say that the word is equivocal as it occurs in these different cases ... [if] we are tempted to speak of “different senses” of a word which is clearly not equivocal, we may infer that we are pretty much in the dark about the character of the concept which it represents

What is the underlying unity of (c), (d), and (e)?

## 2. *The nature of action*

- \* Wittgenstein: what is added to the claim “my arm goes up” by “I raised my arm”? What is the difference between an action and a nonaction, a piece of mere behaviour?/What is it for o to be an action (as opposed, say, to mere behaviour)?
- \* Action theory typically understands Wittgenstein’s question to concern *intentional* action. But should it? Perhaps the question concerns *action*, and we can also ask what is added to the claim “I raised my arm” by “I raised my arm intentionally”: after all, *going on holiday* is perhaps an action, whether by mistake or on purpose. Unless otherwise indicated, I will be neutral on the possibility of unintentional action.

## 3. *The epistemic dimension of intention*

Setiya:

If I have no idea that in humming Beethoven’s Ninth I am driving my wife crazy, I simply cannot be driving her crazy *intentionally*—at least not so far as my humming goes.

Accordingly, it seems as if doing something intentionally carries with it an exceptional degree of first-person access. One possibility is that one thereby *knows* what one is doing, although in Davidson’s famous case:

I am attempting to make ten carbon copies. I press with all my might, hoping to do so.

Unbeknownst to me, I succeed; I thereby am making ten carbon copies intentionally.

One question is whether the example succeeds. One reaction: practical psychological talk can be a bit loose so we might want to pat theory into shape to accommodate the occasional hangnail.

Setiya thinks the case works, replacing it with the following:

When someone is acting intentionally, there must be something he/she is doing intentionally, not merely trying to do, in the [at least implicit] belief that he/she is doing it

Either way, we wonder what it is about intention that brings with it this degree of first-person access.

## II. The inner and the outer

A basic contrast in approaches to resolving these questions:

- *Internalism*: intention (c) is more basic than action (e/f)
- *Externalism*: action (e/f) is more basic than intention (c)

(Nobody thinks “intention with which” is more basic than either.)

After presenting the rough pictures and the answers they give to our questions, I will explore ways in which they might be brought together and then explain why I have a dog in this fight.

### A. Internalism

- An intention to A is an inner representational occurrence, representing the behaviour type *A-ing*.

- ii. An intention typically causes itself to persist into the future, and causes behavioural processes of the type that it represents (when the type in question is highly future-oriented, as in a prospective intention, this component will be negligible until the future becomes present); if the behavioural process in question is of a telic type, the intention will self-destruct once the subject comes to believe that the telos has been attained.
- iii. An intention has as a part a “de se” (essentially self-referential) belief that one “performs its object”: if the object is *continuing A-ing*, the content of the belief is that one continues A-ing; if the object is *A-ing at future time t*, the content of the belief is that one A’s at future time t.
- iv. A behavioural process of “A-ing” is an act of A-ing intentionally just when it is caused “in the right way” by an intention to continue A-ing.
- v. When one A’s, intending to B, this is because one’s behavioural process of A-ing is caused “in the right way” by an intention to continue B-ing.

Comments:

- The right-causation requirement is a notorious issue. Setiya’s simplification of Davidson’s famous worry: if I intend to shake in order to signal my comrade, but this makes me so nervous that I start shaking, my shaking behaviour is *caused* by my intention to shake but it seems to be *mere* behaviour. So the viability of the account relies on the existence of some distinctive variety of causation which holds between an intention and a course of behaviour just when the latter is an action. We can see that in trying to explain the nature of this causation, Gettierology is right around the corner.
- A significant internal dispute in this camp is what the precise nature of an intention is. I have just sketched the crudest contours of an answer here. Setiya’s subtle view is that an intention to A is a desire-like belief that one is hereby A-ing. The ultimate answer here hangs on a mammothly complex set of considerations pertaining to the nature of reasons, rationality, planning, motivation, ...
- I said that an intention is an occurrence. The orthodox view here is that intention is a state, but it can also look somewhat activity- or accomplishment-like: intentions cause themselves to persist into the future, rather like such ordinary biological processes as respiration. If an intention causes itself to persist into the future *essentially* rather than by mere disposition, we should deny that it is a state. It is a nice question how one might argue that the internalist should accept this view.
- Internalists typically think that the objects of intentions are types of *behavioural process*, rather than types of *action*. It is not entirely clear what the motivation for this view is. The anti-dualist motivation here is fairly clear, but it is a nice question why we would think this if we were not anti-dualists.
- That said, if the reductivist internalist wishes to acknowledge the possibility of unintentional action, his “right-causation” troubles increase: actions in general are *right-caused* behaviours, while *intentional* actions are *super-right-caused* behaviours.
- *Davidson’s proto-view*: o is an intentional A-ing iff o is behaviour performed with the intention to A iff o is right-caused by a desire for A-ings and the belief that o is of this kind. Davidson threw over the desire element for the orthodox view on the grounds that there is no straightforward way to accommodate prospective intention.

## B. Externalism

- i. An act of A-ing is a process with the distinctive irreducible property *being an A-ing*.

- ii. When such an act is an activity, it will cause itself to continue until either stamped out by some competing process or terminated by some governing activity; when an accomplishment there is also the possibility of stopping when attaining its telos.
- iii. Sometimes an act of A-ing can be “liminal”: it hasn’t gotten far enough along yet for anybody to really notice that it is happening.
- iv. Intending to A is a kind of act *defined by reference to A-ing* (in a sense to be explored further).
- v. When one A’s because one is B-ing, one’s B-ing is the *ontic ground* of one’s A-ing; when one A’s, intending to B, one’s act *defined by reference to B-ing* is the ontic ground of one’s A-ing.
- vi. A-ing and “A-ing” are related in that the former is something like a guiding cause of the latter.
- vii. What we should say about the relation between A-ing and A-ing intentionally, and about the epistemic dimension of intention, will be explored further.

Comments:

- The key issue here is the notion of being defined by reference to A-ing.
- As we have seen, according to Thompson, this notion is *identity: intending to A = A-ing*.
- This is somewhat problematic:
  - If A-ing unintentionally involves the absence of an intention to A, there can be no such thing as A-ing unintentionally
  - It can seem as though an intention to A can persist once A-ing has stopped, as in Setiya’s example of Fred, who has unknowingly recently failed in his attempt to walk home by the shortest route possible and is now walking home by a very long route.
  - The epistemic dimension of intention is not straightforwardly explained. Thompson’s response to this sort of concern (102–3) seems not to the point: maybe he is right that sometimes first-person access is missing, but if it is the normal case we wish to know why this is. What distinguishes action from other biological processes, such as metabolism of alcohol by the liver, in which I *never* have first-person access?

C. How the views could be brought closer together

Central differences:

- a. The difference in priority
- b. The internalist’s insistence that intention is a state
- c. The externalist’s insistence that A-ing and intending to A are always found together

If there were a view which (a) treated intending to A and A-ing as metaphysically distinct (though perhaps somehow conceptually interdefined) types; (b) accepted the externalist’s view that intention is a process; (c) accepted that these phenomena can be brought apart—this view would count as a sort of intermediary between the internalist and externalist views.

D. What’s a consciousness guy doing “rolling with” action people?

Simple answer. As we will see, my view is that ordinary narrative directly illuminates the varieties of consciousness in its claims about our actions: “I went for a stroll” states that among the types of my earlier experiences was *strolling*.

But according to the internalist, action types are *behavioural types*: whether a B-ing is an action is not a matter of its intrinsic nature, but is rather due to how it is caused by an inner occurrence. The “strollings” and the strollings belong to the same kind of event.

If this were correct, it would be very bad for me. Our ordinary narratives would at best “gesture at” the natures of our experiences. A midway salvage position would be that what is gestured at are inner episodes of intending: instead of hanging on literal interpretation, I could hang on “the most straightforward interpretation” or some such thing.

But that position is rather weak dialectically: the phenomenal state person could then quibble about why only this amount of interpretation rather than the slightly more expansive amount it would take to get to his view, which is that these narratives gesture at phenomenal states.

At that point, I would be in the position of fighting out a technical dispute against a strongly entrenched majority, rather than of standing with the common person against a rogue philosophical faction. Not such a great position!

### III. Epistemic externalism

I think that we can generate the intermediate position, saving the view from the threat of internalism. I am going to borrow from Mike Martin’s stuff in the philosophy of perception here.

#### A. Indiscriminability

- Discrimination is, as Williamson has taught us, *knowledge of distinctness*:
  - To discriminate a and b is to know that a is distinct from b ( $a \neq b$ );
  - To discriminate a from an F is to know that a is distinct from an F ( $\neg Fa$ ).

*Complication*: knowledge is under a mode of presentation, and I might discriminate a from b under one mode of presentation but not another (Williamson’s example: I discriminate The False from The True when presented as such, but not when the latter is presented as the truth-value of Fermat’s Last Theorem [ok, ok, the example is a bit old, you get the idea]). Typically we will leave this consideration tacit.
- *Discriminability* is *knowability* of distinctness, or the possibility of knowledge of distinctness.
  - Talk of possibilities is almost always hedged around by tacit restrictions—who, when, how. That’s a good thing, as it enables us to save lots of words.
- *Indiscriminability* is *unknowability* of distinctness, or the *impossibility* of knowledge of distinctness.
- Example: presented with two unidentified glasses of red fluid, I sip, sniff, swirl, slurp ... all to no end. I have no idea whether *this* kind of fluid = *this* kind of fluid. So for me, Red Bull and 1990 Barolos are indiscriminable—great way to save money if I weren’t such a snob. More precisely: indiscriminable *by ordinary methods of beverage appreciation*—eventually I throw up my hands and the secret is revealed, so they were discriminable for me *by asking*. Also, stunned with embarrassment, I enrol in beverage appreciation 101, where they tell me on the first day how to detect the HFCS-y notes in the Red Bull and absent in the Barolo: aha! *now* they are discriminable to me.
- Example: visiting New Zealand, I’m presented with Butch, a friendly shaggy fellow. “A fine example of one of your famous sheep!” I announce with pride in my swift mastery of the local fauna. Sadly, Butch is a *sheepdog*. Butch wasn’t a sheep, but I couldn’t know this: for me, Butch was therefore then indiscriminable by looking from a sheep.
- Points of **VAST IMPORTANCE**:
  - One cannot know, what is not so (Johnnie Cochran, eat your heart out). If a *is* b, I cannot discriminate a and b: this would be to know that a was not b, but it is, so I can’t. If a *is* F, I cannot discriminate a from an F. It does not matter how much weight Tweedledee

loses between the first and second times I meet him: I cannot discriminate the tubby brat from the svelte rocker. Nor does it matter how convincingly Butch is made up for Halloween to resemble a cow: I still cannot discriminate him from a dog. In this case I will sometimes say that the indiscriminability is *trivial*.

- Indiscriminability does not imply anything about seeming the same! a and b are indiscriminable even though b seems to me like a pebble and a is in fact a boulder, when a is off in another county and does not seem any way to me at all. Or: a and b *are* in fact very different and indeed they *seem* very different—however, b seems the way it does due to illusions that corrupt my sense of how it is. I am thus in no position to know anything about it.
- A condition stronger than indiscriminability is inability to know *whether* there is distinctness. The colors of two paint chips are in fact exactly the same and in fact the way they seem is exactly the same; still, my perception is infected with noise, and is not reliable about how things are *exactly*. Here I can't know that the colors are the same; and of course they are trivially indiscriminable. Conversely, though they are ever-so-slightly different, whether the way they seem is exactly the same or ever-so-slightly different, I may still be unable to know *whether* they are distinct.

## B. Introspection

This will be regarded as our garden variety notion of “reflecting on consciousness”. This notion is intricate and pesky and raises complications which we will put off for several weeks.

## C. The view

The view I wish to push basically borrows most of Thompson's apparatus, with one big difference.

Recall the externalist's view that intending to A is “defined in terms of” A-ing. For Thompson, that was understood as *identity*.

### 1. The rough view

I propose rather to say something in line with the following:

- For Joan to intend to A is for Joan to be doing something which is reflectively indiscriminable from an A-ing.

Unpacking the RHS, the thought is that it can't be known by reflecting on consciousness that what Joan is doing is not A-ing.

One way for this to happen is for Joan to be A-ing: in that case, she is trivially doing something indiscriminable from A-ing.

Another is for Joan to be doing something which seems for all the world like an A-ing. This is good: unlike on the Thompson view, one can intend to A even though one is not A-ing. This resolves Setiya's puzzle of Fred.

Note that at this point, we have primitive action types which are distinct from behavioural types; and we have the possibility of intending to do something which one is not doing.

However, we have not yet generated the converse, doing something which one does not intend to do. After all, whatever one is doing, it is trivially indiscriminable from an instance of its own kind; hence, by the crude view, one intends to do it. So Fred both intends to walk home by the shortest route possible and intends to walk home by a very long route. That is not plausible!

Moreover, we do not yet have a characterization of what it is to A intentionally. I want to suggest that to do so is to be A-ing, knowably by introspection. This would be nice, because it would explain the epistemic role of intention—why one is typically in a privileged position to know what one is doing intentionally.

A further question arises: what about action makes such knowability the ordinary case? I have an answer here: action types are types of experience, and ordinarily one is in a privileged position to know what type of experience one is having. This is a point on which Thompson is neutral, of course—a big gap in his account.

We might additionally want to know why one is typically in a position to know what one intends to do: there is a bit of techiness about this but the rough idea is that there will usually be a lot of things that one *is* in a position to know one isn't doing; one's intention will thus stick out like a sore thumb.

However, on this proposal, Fred plausibly is not intentionally walking home by a very long route; and yet he is doing so, and he intends to do so; and (since the action just is the intention in this case) the action is as lashed to the intention as you could like. In what sense is he not doing so intentionally, then?

Fortunately, we can fix this problem.

## 2. A Martinesque view

On MGF Martin's treatment of "perceptual experience as of seeing a white picket fence", one has such experience just if one is in some state which is indiscriminable from *seeing a white picket fence*. More generally, if F is a *fancy type of perceptual state*, one has perceptual experience as of F just if one is in a state indiscriminable from an F. Here the fancy types of perceptual state are the cases of seeing things; whatever type of perceptual state is involved in the case of hallucination is not a possible object of perceptual-experience-as-of. Accordingly, it is unclear whether we have any grasp on this type at all.

- If A is a "fancy" action, one intends to A just if one is doing something which is indiscriminable from A-ing;

Non-fancy actions can't be the objects of intentions.

The thought then is that we similarly cut the types of action in half, into the "normals" and the "goof-ups", with normals being fancy. Goof-ups aren't possible objects of intention, and indeed it is unclear whether we have any grasp on their nature at all.

The anxiety here is that we might want to say about Fred that he *intended* to walk home by the shortest route possible but was *actually* walking home by a very long route, where this is a genuine action rather than mere zombie-walking. More generally, we could say that someone was acting unintentionally but could never say that they were A-ing unintentionally for any particular A.

## 3. In the good case

We can play a trick which builds fanciness into the object of discrimination but removes it from the action type, as follows:

- One intends to A just if one is doing something which is indiscriminable from *A-ing in circumstances C*.

Circumstances C are thus rather fancy, and when one A's in C, one A's rather fancily, but one might also A not quite so fancily outside of circumstances C.

To be explicit about the scope: just if one is doing something which *both* can't be known not to be an A-ing *and* can't be known not to be taking place in circumstances C. This strengthens the condition on intending to A, thereby making it more difficult to intend to A, as desired.

Considered in the abstract, this has the power to resolve the difficulty for the initial proposal that one intends to do everything one in fact does. If one is A-ing, one is doing something trivially indiscriminable from A-ing; but if the circumstances C are distinct from one's actual circumstances, what one is doing is *not* trivially indiscriminable from something taking place in circumstances C.

On the flipflop, however, by making intending to A more difficult, a new threat opens up, namely that certain paradigm straightforward cases of intending to A might be ruled out. In particular, there is the worry that cases of A-ing intentionally might fail to be cases in which one intends to A. Now, this might not be the worst thing in the world (it seems as if when A is odious people will often ascribe intentional A-ing but not an intention to A). Still, it is clear that we must position C carefully so as to get reasonable answers to when there is intention and when not.

Suppose we have in hand a distinction between the "good case" and the "bad case": roughly, cases in which the agent is not/is significantly deluded. If so, we could let circumstances C be the good case. If so, then one intends to A just if one is doing something indiscriminable from A-ing in a case in which one is not deluded. Modulo the vagueness and relativity of 'significant', this gets the key paradigms right.

- In a normal case, Joan is not significantly deluded in regard to her actions. Accordingly, the 'no significant delusion' condition is trivially indiscriminable from the truth, and it follows that whatever she is doing is what she trivially intends to do.
- Although he is significantly deluded, Fred nevertheless can't discriminate his action from getting home in the quickest way possible in a case in which he is not significantly deluded (if he could he would know he is not significantly deluded, but he is, so, of course, he can't: that's the nature of delusion). More importantly however, he can discriminate his situation from walking home in a long way in a case in which he is not significantly deluded. Roughly, under such circumstances, he would have evidence that he was walking home in a long way. Plausibly his case doesn't prevent him from recognizing his lack of such evidence; so he has what he needs to establish that he is not in such a case. This is great, because it allows us to accept that Fred is taking a very long route unintentionally and that he does intend to do so. This allows us to push in good conscience the view that A-ing intentionally is A-ing that can be known as such by introspection.

All sorts of technical questions arise at this point: how is the good case determined? Is it absolute, or relative to the action type, or relative to something else in the attributee's circumstances? Or is it rather pegged to certain of our expectations about the attributee and/or his or her type of action? And once we have settled on just the most general framework questions, we must then go on to fill in the details.

However, perhaps a lot of this debate can be avoided: I'm willing to say up front that it is *our* judgements that set the standard, and we make them by going by the seat of our pants. This would be a problem if I needed to say that intention has a real essence, but fortunately this is something I am quite comfortable denying!

#### D. The promised mixed view

Recall the central differences between the internalist view and Thompson's externalism:

- a. The difference in priority (action or intention)
- b. The internalist's insistence that intention is a state
- c. The externalist's insistence that A-ing and intending to A are always found together

If there were a view which (a) treated intending to A and A-ing as metaphysically distinct (though nevertheless still conceptually interdefined) types; (b) accepted the externalist's view that intention is a process; (c) accepted that these phenomena can be brought apart—this view would count as a sort of intermediary between the internalist and externalist views.

The epistemic externalist view does this:

- a. Intending to A is *sometimes* A-ing and *other times* doing something other than A-ing (as when one intends to A but does not do so).
- b. Intending to A is always doing something, hence a process. (I could be concessive here and grant that in many cases this process involves a representation of A-ing, but I don't think I could say that it does so essentially—not quite sure on this point)
- c. They can be brought apart, as per (a).