

Lecture 2: 'Intention, Belief and Desire'

I. Recap

The puzzle of intention:

- (1) **Future-directed intention:** I have an intention to take this course.
- (2) **Intention-with-which:** In taking this course I intend to learn about the will.
- (3) **Intentional action:** I am intentionally taking this course.

The question: **Is there a deep unity to these three uses of intention?** If so, how should we understand this unity?

Last week, we looked at early accounts which sought to reduce future-directed intention (1) and (2) to (3). We saw how, following Davidson, these 'intention as doing' accounts face various problems - such as how they seem to struggle to account for 'pure intentions' which might never lead to any kind of action.

Our question for the next two weeks: If one is convinced, like Davidson, that intentions must be a kind or attitude or mental state after all, what *kind* of mental states are intentions?

II. Cognitive and Conative Attitudes

A standard way to motivate the distinction between beliefs and desires as mental attitudes is to invoke the idea of 'direction of fit'. Take the following famous example from Anscombe:



Conative attitudes

- Have a '**world to mind**' direction of fit
- They represent some states of affairs as to-be-made true, or to-be-realised
- Examples: desiring, hoping etc.



Cognitive Attitudes

- Have a '**mind to world**' direction of fit
- They represent their contents as true or 'as being the case'
- Examples: believing, imagining, assuming

In Anscombe's case the shopper's list aims to bring the world into conformity with itself (and in this way is analogous to desire). The detective's list is the other way around: their list aims to conform itself to whatever is in the cart (and thus is analogous to belief).

This gives us a way of framing our question viz. are intentions more like conative attitudes, more like cognitive attitudes, both, or neither?

III. Intentions as Predominant Desires

Intentions clearly seem to have a world-to-mind' direction of fit: they don't represent their contents as already being the case, but rather as to-be done (yes they seem to be constrained by reality, but so do desires e.g. desiring the impossible?).

This has led some philosophers to think that we can identify intention with predominant desire:

Intention as predominant desire: An agent intends to do A just in case she desires to do A more than she desires to do anything incompatible with A-ing. (the notion of predominant desire here is capturing the fact that we don't - luckily - intend everything we desire).

These simple predominant desire views face obvious objections relating to the extensional adequacy of the account of intention provided by desire theorists.

- (1) Relating to the element of **commitment** that many have thought is constitutive of intention. It seems possible - and indeed, quite common - to predominantly desire to do something you do not intend to do, or even that you intend not to do.
- e.g. an addict who has decided (and formed the intention) to quit drinking, despite an overwhelming desire to drink.

So predominant desire does not appear sufficient for an intention.

- (2) Conversely, it seems we can easily form intentions to A **without a predominant desire for A**.
- e.g. Choosing between two equally good options at lunch.

So merely predominantly desiring something does not appear necessary for having an intention either.

IV. A Simple Belief-Desire Model of Intention

Many philosophers have picked up on the fact that the objection in (1) seems to reveal something important about intention viz. that when we intend to A, this typically involves a commitment or '**settledness**' about what we will do, which appears to distinguish desire from intention.

Compare 'Cecily desires to go to a party in London this weekend' and 'Cecily intends to go to a party in London this weekend'.

One suggestion about how to account for this feature of intentions is to supplement our ontological account of intention in terms of desire to also include belief. For a while, this was viewed as the 'standard model' of intention in the philosophical literature.

A simple belief-desire model of intention: An agent intends to do A just in case she (i) desires to do A and (ii) believes that one will A.

- This seems to make sense of cases where we make up our mind before intending e.g. choosing an option at lunch, going to the party.
- But can such accounts make sense of the addict who forms an addiction? Do they have a *belief* that they will stop drinking?

Objection I: Belief about Doing

Belief-Desire models of Intention appear committed to the idea that if one intends to A, one has the belief that one will A, or will succeed at doing A.

But its not always clear this is the case:

- Davidson's carbon copier example.
- Someone who is recovering from paralysis and intends to clench their fist. Do they believe they will succeed?

Relatedly, consider *Belief*: if A is doing ϕ intentionally, A believes she is doing it.

Setiya (2007;33): "An agent can be motivated by the desire for an end, and the belief that doing x is a more or less effective means to that end, without having the belief that he is doing x".

"If, when I am clenching my fist intentionally, the belief that I am doing so had to be inferred from the premise that I intend to be clenching my first and from the empirical knowledge of the conditional that if I intend to do so, I will, it would be possible for the inference not to take place. I might simply fail to put two and two together".

If a belief-desire-model of intention is correct, why don't we, to use Setiya's phrase, 'put two and two together?'

Objection II: Reason-Choosing and Acting for the Wrong Reasons

Many philosophers hold that we can choose which reason to act on — that our reasons for acting are in an important sense 'up to us':

Setiya (2007):

"There are many reasons for which I might decide to write a book: personal satisfaction, a fragment of immortality, professional ambition. I am not passive in the face of this: even if I believe that books give their authors a kind of immortality, and even if I think that this is a reason — a good reason — to write a book, it may not be my reason for doing so. That is up to me." (39-40 in Sinhababu 2013 686).

- If belief-desire views of intention commit to the idea that motivating reasons are to be equated with beliefs and desire pairs (questionable - we'll come back to this next week) there seems to be a problem if one accepts the idea that we can choose our reasons.
- This is because the majority philosophers of mind accept the idea that we can't choose our beliefs and desires in this way, just by choosing them. But we seem able to choose our reasons, even if the desires motivating other reasons are stronger.

- According to Setiya, we ought to reject the desire-belief model of intention on the basis that it cannot account for the 'active and selective character of this attitude we have toward our reasons for action'.

Objection III: The Distinguishing Features of Intention

Bratman:

- (a) **Commitment** (as above): intentions as "conduct-controlling dispositions". How can belief-desire models of intention account for, or explain, this salient feature of intention?
- (b) **Stability**: In having an intention to act, agent A is disposed not to deliberate any further about whether or not to act, unless new and relevant information comes to light. If desires are understood as 'merely potential influencers of action' how can the belief-desire model account for this?

"even if I now have a predominant desire to go to [the faculty building]... I might still not see the issue as settled. I might be disposed to continue to give serious consideration to the possibility of taking the afternoon off and going to a concert. **But if I were to intend to go [to the faculty building], I would not be disposed to continue to deliberate in this way: this is what is involved in the resistance to reconsideration characteristic of intention** (Bratman 1987:18-19).

- (c) **Means-end coherence** :A is disposed to deliberate about intended means or preliminary steps to act, as well as about more specific intentions, as when one reasons from an intention to take a bus to the library and information on a bus schedule to the intention to take a particular bus. Again, it's not obvious how desires and beliefs alone could explain this feature of intention.

V. Interlude: Intention as Evaluative Judgement

Once convinced by the idea that intentions are real attitude, Davidson went on to identify intention not with desire but with a special kind of belief: the belief that some available option is best or most choice worthy.

Intention as evaluative judgement: to intend to perform action A is to judge that A is better than any other alternative to A-ing that one is considering.

- Fits with a conception of **agency as rational activity** - on the evaluative judgement conception agency is essentially guided by reason: a rational agent takes whatever she takes to be the best thing to do, regardless of whether it is what she most wants or desires.

Problems for this account:

- Cases in which **one intends to A even if you believe that some other available action B would be better** (can addiction be understood in these terms? We'll come back to this in week 4)
- Even if we think addiction should not be understood in these terms, there are cases where even fully rationally agents encounter non-addictive situations in which they can have intentions to act in the absence of a belief that the intended action is best: incomparable options cases.
- Cases often discussed in the literature: **Burdian cases. Sartre's young man.**
- These have been thought to put tension on Davidson's evaluative judgement account of intention.

VI. Sinhababu's Revised Belief-Desire Model of Intention

Back to Objections I-III. These and related objections are often proposed as damning objections to belief-desire accounts of intention, but is there a version of the belief-desire model of intention which can accommodate these features of intention?

Sinhababu thinks so. In his paper '*The Desire-Belief Account of Intention Explains Everything*' he puts forward a revised account of the simple view we've just considered which purports to avoid the consequences of the simple belief-desire model. He argues that critics of desire-belief models of intention have failed to pay sufficient attention to relevant features of desire which explain the features of intention they appeal to.

Revised Belief-Desire Model: A intends that ϕ if A has a desire and belief such that for some behaviour B and situation S:

(1) A desires that ϕ .

(2) A believes that S will obtain, and that A's B-ing in S would make ϕ more likely.

(3) If A were to believe that S obtained, the desire and the belief would, with out further practical reasoning, produce **motivational force** causing A to initiate B-ing.

- Strictly speaking, Sinhababu presents this as an account of intention to ϕ as a "an appropriately situated desire" to ϕ — but we can also view this as a compositional account which takes (1)-(3) as constitutive of what it is to intend to ϕ .
- Situation S in (2) accommodates those both in the present and future, thus providing a **unified account of intention** in line with the puzzle of intention we find in Anscombe.
- Doesn't cash out the belief involved in intention in the same way as Davidson does, and so arguably avoids the problems we just discussed.
- This is an account of the **sufficient conditions** for intending — "an account of how intention is psychologically realised in human beings" (682). (Intentions could be realised in different ways in artificial agents).

Overcoming Objection I: Belief about Doing

"If, when I am clenching my fist intentionally, the belief that I am doing so had to be inferred from the premise that I intend to be clenching my first and from the empirical knowledge of the conditional that if I intend to do so, I will, it would be possible for the inference not to take place. I might simply fail to put two and two together".

Sinhababu argues we can explain why agents often reason from beliefs about behaviour as a means for achieving an end to the belief that one is acting (having formed an intention) by appealing to **desire's ability to direct attention on things, and reason in ways, which are conducive to its satisfaction**. When acting on an intention, the desires motivating the action drive reasoning about how their object might be attained.

"Someone may believe all the premises from which he could discover how to enter his house without a key, but not attend to this matter and draw the conclusion until he's locked out and desires to get in" (684).

- This doesn't explain the supposed *necessity* of belief about doing, but accounts for it as a contingent psychological fact about human desire.
- This is continuous with the idea that in fact, belief one is acting is not robust (e.g. driving home on autopilot).

Overcoming Objection II: Reason-Choosing

According to Setiya, we ought to reject the desire-belief model of intention on the basis that it cannot account for the 'active and selective character of this attitude we have toward our reasons for action'.

How is it possible that we have two reasons for action, one of which comes from a stronger desire, and one which follows from a weaker desire, and we can choose to act on the weaker of the two?

Answer: The desire-belief view allows for reason-choosing by allowing a **second-order volition** to block the strongest desire supporting some action from motivating it, so that a weaker desire is the sole motivator.

"While I don't know the reason Setiya chose for writing his book, I can explain how he chose. He had a number of desires which pushed in favour of writing the book, and because of which he saw a number of reasons for writing. He desired to write the book for some reasons rather than others. If this desire about which reason to act on was strong enough, it could exert motivational force blocking any decision to write which he believed arose from the wrong reasons".

- How often do we actually choose our own reasons?

Overcoming Objection III: The Distinguishing Features of Intention

(a) **Commitment:** Sinhababu argues (a) is easily explained by appeal to desire's ability to motivate action when combined with means-end belief.

(b) **Stability:**

Sinhababu distinguishes two features of intention that stability, in Bratman's sense, picks out:

(i) First, we don't constantly **reconsider** our intentions.

- He thinks this doesn't require a robust explanation. In fact, we can reconsider intentions without any new information (where vivid images intensify pre-existing desires).
- e.g. we don't often reconsider our beliefs without new information either

(ii) Second, we don't usually form new intentions that conflict with existing ones (**'exclusion'**).

- Sinhababu thinks we can explain this by appealing to our beliefs about the future
- All we need for a future desire to give an option deliberative stability here is a belief about our future actions, and that this explains how and why intentions are often easily revised:

"Suppose that at 8am Mike believes that at noon he'll intend to go to [lecture] and that he'll be able to get there. Even if he has no desire to spend noon that way at present, his belief about his future intention will prevent him from seriously deliberating now about whether to do things at noon he can't do [if he's at the lecture]". (691).

- (c) **Means-end coherence:** ‘desires themselves are inputs in deliberation’. ‘As we deliberate, desires direct our attention, pointing out potential means and combining with means-end beliefs to form new intentions’:

“The intention to go to [to lecture] is made of a desire to go to [lecture]. Riding an appropriately timed bus focuses the agent’s attention on the causes and effects of going to lecture” (693).

VIII. Looking forward

- Next week, we’ll look at Bratman’s arguments (relating to some of the considerations we’ve considered today) for the view that we need to posit a sui-generis mental category ‘intention’ in addition to belief and desire, to accommodate our talk of intention.
- According to Sinhababu: ‘the desire-belief view’s explanations thus use conceptual resources that both sides must admit, while leaving out any unexplained properties of intention. Thus, it allows us to simplify our total psychological theory’.
- We’ll also look at the **methodological question** underlying this debate: when do we have grounds for positing an additional psychological category in our theorising about the mind? What should it take for us to be convinced of arguments like Bratman’s to extend our psychological repertoire to include a category of intention? Is explanation — in Sinhababu’s words — the measure of success here? If so, what are the facts we are trying to explain?