

Timothy Williamson, “Knowledge and belief”

Excerpts from Timothy Williamson, “Knowledge and belief”, in *The Norton Introduction to Philosophy*, edited by Gideon Rosen, Alex Byrne, Joshua Cohen, and Seana Shiffrin (Norton, 2015).¹

Williamson starts by investigating the relationship between belief, truth, and knowledge. First, he notes that although you can *believe* falsely, you cannot *know* falsely: if you know *p*, then *p* is true.

The most striking difference between knowledge and belief is that although there is false belief there cannot be false knowledge. People once believed that the earth was flat. They believed falsely, because the earth was not flat. They did not *know* that the earth was flat, because knowing that the earth was flat would have required the earth to be flat. They *believed* that they knew that the earth was flat, but that was another of their false beliefs.

We can make the same point about a disagreement without even taking sides. Suppose that Mary believes that there is life on other planets while John believes that there is no life on other planets. We do not know which of them is right, but we know that there are only two possibilities. Either there *is* life on other planets, in which case Mary believes truly while John believes falsely, so John has belief without knowledge, or there is *no* life on other planets, in which case Mary believes falsely while John believes truly, so Mary has belief without knowledge. Either way, one of them falsely believes that something is the case without knowing that it is the case, even if we cannot tell which of them it is.

You can believe *p* without knowing *p*: belief does not imply knowledge. What about the other way round? Can you know *p* without believing *p*?

It seems obvious that you could not know that the earth is round without believing that the earth is round. However, there are some tricky cases. Suppose that many years ago Kerry read a good history of China, but has forgotten all about doing so. Now she enters a quiz. Some of the questions turn out to be on Chinese history, of which Kerry believes herself to be totally ignorant. Nevertheless, answers pop into her head. She regards them as random guesses, but nevertheless tries them out, since she has nothing better. They are all correct. In fact, her answers were caused by memory traces derived from the book. One hypothesis is that Kerry unconsciously

¹ Full text at http://media.wnorton.com/cms/excerpts/NINTPHIL_Ch3.pdf, pp. 124-30.

knows those truths about Chinese history, because she remembers them, although she does not believe them. If so, knowledge does not imply belief. But that hypothesis has problems. Consider Terry, who also read a history of China many years ago and has forgotten all about doing so. When Terry enters the quiz, answers about Chinese history pop into his head too, and he tries them out too, for want of anything better, despite regarding them as random guesses. In fact, Terry's answers too are caused by memory traces derived from a book. However, all Terry's answers are wrong, because his book was a bad one, full of mistakes. The hypothesis that Terry unconsciously knows those falsehoods about Chinese history does not work, since false knowledge is impossible. If Kerry unconsciously knows the right answers, Terry unconsciously *believes* the wrong answers. But then, since Kerry is no less sincere than Terry, Kerry also unconsciously believes the right answers. Thus Kerry is not a convincing example of knowledge without belief. She may instead be an example of unconscious knowledge and belief without conscious knowledge or belief. That knowledge implies belief is a good working hypothesis.

It is reasonable, then, to take knowledge to imply belief—moreover, *true* belief.

Williamson now gives two examples to show that true belief does not imply knowledge: you can truly believe *p* without knowing *p*.

If Larry believes that the name of the capital of California starts with 'S', he believes truly, since the capital is Sacramento. But if that belief rests only on his irrational belief that the capital is San Francisco, Larry does not *know* that the name of the capital begins with 'S'. Similarly, although either John or Mary has a true belief as to whether there is life on other planets, perhaps neither of them *knows* whether there is life on other planets, because neither of them has sufficient evidence for their belief.

As Williamson says "Many philosophers have reacted to such examples by asking: what must be added to true belief to get knowledge?". And a popular answer to that question used to be:

justification, in the sense of blameless belief. The idea was that Larry's true belief that the name of the capital begins with 'S' does not amount to knowledge because he deserves blame for irrationally believing that the name of the capital begins with 'S'; his belief, although it happens to be true, is not justified. However, we can imagine a slightly different story, in which Barry is the victim of a massive hoax, so that he has strong misleading

evidence that San Francisco is the capital. For example, that is what his high school teacher tells the class, everyone whom he asks confirms that it is, his classmates hack into his computer so that he cannot access websites that say differently, and so on. Barry's beliefs that San Francisco is the capital and that the name of the capital begins with 'S' are blameless, and in that sense justified. Thus Barry has a justified true belief that the name of the capital begins with 'S', but he still does not *know* that the name of the capital begins with 'S'. For he does not know that San Francisco is the capital, since that is false, and beliefs based on ignorance do not constitute knowledge. In a famous article, "Is Justified True Belief Knowledge?", the philosopher Edmund Gettier used such examples to make just this point, that justified true belief is not always knowledge.

Gettier's 1963 article acted as a challenge to philosophers to find the 'missing ingredient' that added to true belief would make knowledge. Many proposals have their supporters, but in each case they are greatly outnumbered by opponents. In effect, the aim is to find a solution to the equation:

$$\text{knowledge} = \text{true belief} + X$$

Typically, when someone proposes such an X, other philosophers soon find examples of knowledge without true belief + X, or of true belief + X without knowledge, either of which suffices to refute the equation. Although no argument refutes all such proposals at one shot, their track record looks increasingly poor. Rather than examine in detail various attempts to solve the equation, let us take a step back and consider the presupposition that it has a solution.

An analogy: Crimson is a specific type of red. Just as all knowledge is true belief but not all true belief is knowledge, so all crimson is red but not all red is crimson. Now consider the equation:

$$\text{crimson} = \text{red} + Y$$

We have no reason to expect this equation to have a useful solution. It asks for a property Y such that the crimson things are exactly those red things that have Y. The only natural suggestion is: $Y = \text{crimson}$. Crimson is indeed equivalent to red that is crimson, but as an account of crimson that is blatantly circular (all it tells us is that crimson implies red). Similarly, knowledge is indeed equivalent to true belief that is knowledge, but as an

account of knowledge that is blatantly circular (all it tells us is that knowledge implies true belief). The attempt to analyse crimson as red plus other elements is wrong-headed. Why should the attempt to analyse knowledge as true belief plus other elements do better? Why should we try to explain knowledge in terms of belief rather than belief in terms of knowledge? What should we take as our starting point? In philosophy, as in the rest of life, where you start makes a big difference to where you end up.

Why have philosophers “regarded belief as ‘simpler’ or ‘more basic’ than knowledge, and therefore as a better starting-point for explanation”?

One reason is that, until recently, the dominant conception of mind was an *internalist* one. According to internalism, what mental states you are in is completely determined by what is going on internally to you, which for present purposes we can understand as: inside your head. Although an event outside your head can *cause* you to be in a specific mental state, as when a glass breaking causes you to have a corresponding experience, it does so by causing other events to occur in your head, and internalists say that the events in your head completely determine that you are having the experience, irrespective of what is going on outside your head. For them, any difference between two situations in your mental states implies a difference in what is going on in your head. Belief seems to fit this account much better than knowledge does. In one situation, a pilot knows that he is flying above the Atlantic. In another situation, without realizing it the pilot was put in a perfect flight simulator back at the airport and falsely believes that he is flying above the Atlantic; therefore the pilot does not know that he is flying above the Atlantic. Thus the two situations differ in what the pilot knows. They do not seem to differ in what he believes. In both situations, he *believes* that he is flying above the Atlantic. By hypothesis, what is going on in the pilot’s head is also the same in the two situations, in the sense that exactly the same microscopic descriptions apply. Consequently, his knowledge violates the internalist principle ‘No difference in mental state without a difference in the head’, while his belief seems not to. The internalist diagnosis is that knowledge, unlike belief, is not a ‘pure’ mental state. Rather, for internalists, knowing that one is flying above the Atlantic is a *mixture* of mental states such as believing that one is flying above the Atlantic with non-mental conditions, typically on the external environment, such as that one really is flying above the Atlantic. On that view, it is very natural to try to analyse knowledge into components such as belief and truth and perverse to try to analyse belief into components such as knowledge.

However, further reflection suggests that not even belief really fits the internalist model. Imagine a third situation, a perfect duplicate of the first except for being on a different planet, exactly like Earth but billions of miles from it. The Atlantic is not on that other planet—it is on Earth. Rather, they have another ocean exactly like the Atlantic. They even spell its name ‘Atlantic’, but that is *their* name for it. When we use the name ‘Atlantic’, we refer to the ocean on Earth, not to the one on the other planet. Does the extraterrestrial pilot believe that he is flying above the Atlantic? If so, his belief is false, because he is not flying above the Atlantic; he is flying above another ocean billions of miles from the Atlantic. But he is no more mistaken about his position by billions of miles than the terrestrial pilot is (the one who really is flying above the Atlantic). Both of them know where they are. Thus the extraterrestrial pilot does *not* believe that he is flying above the Atlantic. In fact, neither pilot has any beliefs about the other’s ocean at all, because he has no idea that there is any such ocean. Thus they differ in their beliefs, even though what is going on in their heads is exactly similar. More specifically, they differ in the *content* of their beliefs: the content of the terrestrial pilot’s belief is that he is flying above the Atlantic; the extraterrestrial pilot also has a belief, but its content is different, because it is about a different ocean. More generally, the contents of mental states are *world-involving* in the sense that they essentially involve relations to things out there in the world, such as oceans.

These relations to the external environment are, Williamson says, “the point of the mental”; they are not “impurities”:

With minds, we can get what we need by adjusting our behaviour to what we know of a complex, changing environment. We perceive our surroundings and intentionally act on them. Thinking mediates between perception and action. Emotions too involve relations to the external environment. To treat the person whom you love or hate as inessential to your emotion is to forget that love and hate are essentially relations, not undirected qualities of feeling. Since mental states have this sort of world-involving function, no wonder they have world-involving contents. To abstract away from relations to the world in search of pure mind is like peeling layer after layer away in search of pure onion.

Belief is world-involving in its content. Knowledge is world-involving not only in its content but also in the way in which the knower is related to that content. Whether the pilot knows or merely believes that he is flying above

the Atlantic depends in part on whether he *is* flying above the Atlantic. Given what was just said about the nature of mind, this extra dimension of world-involvingness in knowledge may make it more central to mind than belief is, not less.

When things go as they should with our cognitive faculties, such as perception and memory, we get knowledge. When something goes wrong, we get mere belief. 'Knowledge' is a success term; 'belief' is neutral between success and failure. The relation between believing and knowing resembles that between trying to be something and being it by intention. If you believe that you are popular, you may or may not *be* popular. Similarly, if you try to be popular, you may or may not *be* popular. But if you know that you are popular, you *are* popular. Similarly, if you are popular by intention, you *are* popular. Cases in which you believe truly that you are popular without knowing that you are popular correspond to cases in which you try to be popular and are popular, but not by intention—for example, you may be popular *despite* your embarrassing attempts to be popular. Just as it would be perverse to investigate the phenomenon of trying to be something without special reference to the phenomenon of being something by intention (the case when action goes well), so it is perverse to investigate the phenomenon of believing something without special reference to the phenomenon of knowing something (the case when cognition goes well). Malfunctioning must be understood in relation to good functioning. Misremembering must be understood in relation to remembering, misperceiving in relation to perceiving, and so on. All this suggests a knowledge-first methodology.

Williamson considers a reply on behalf of “defenders of a belief-first methodology”, namely that:

once we start giving detailed causal explanations, success terms like 'knowledge' are no longer useful, because they are irrelevant to a step-by-step analysis of a causal process. In explaining how an automobile engine works, at some point you have to specify the actual physical processes involved, and their effects do not depend on whether they are classified as functioning or malfunctioning. Similarly, they say, whether you drink from the glass does not depend on whether you *know* that it contains water; you will drink from it as long as you *believe* that it contains water (and desire water), whether or not your belief constitutes knowledge or is true. However, this simple picture faces several problems.

First, explanations of action in terms of mental states typically involve a time lag between the mental states and the completion of the action, during which feedback can occur. For example, a reporter decides to interview a politician involved in a scandal; she drives to his house and knocks on the door. The mental states 'immediately behind' an action at a given instant, such as moving her hand a fraction closer to the door, are typically just those concerned with the execution of that stage of the action plan. The connection with the original reasoning that gave the action its point—'I want more embarrassing details for this story, and he can supply them, so I'll interview him'—is less direct. Once you have worked out an action plan, you need not keep referring back to the reasons for adopting it in the first place. When we seek to explain human action, our aim is typically to understand it in terms of the earlier reasoning that gave the action its point, so there is a time lag between the reasoning and the completion of the action. That allows for the difference between knowledge and mere belief to make a causal difference to whether the action is completed. For example, how the reporter reacts if her knock at the door is not answered may depend on whether she started with knowledge or mere true belief that the politician was at home. If she knew he was at home, she is likely to be more persistent, and so more likely to get the interview. If she merely had a true belief, she is likely to give up more easily.

Second, what *reasons* are available to you to act on depends on what you know, not on what you believe. If you know that the glass contains water, you may drink from it *because it contains water*. Your knowledge makes the fact that the glass contains water available to you as a reason to act on. If you believed falsely that the glass contained water, my explanation 'You drank from the glass because it contained water' is automatically false. In cases of mere belief, we might say 'You drank from the glass because *you believed that* it contained water'. However, such a fact about your beliefs is not normally a reason on which you act, in the way in which you act on the fact that the glass contains water when you know. For the premise of your reasoning is normally something like 'The glass contains water', not 'I believe that the glass contains water'. You are thirsty, so you think about the water, not about your beliefs. Moreover, the fact that *you believed that* the glass contained water is not what made drinking from it a good thing to do; what made it a good thing to do is the fact that the glass *did* contain water. Water quenches thirst; beliefs do not. A reason for drinking from the glass is a fact that makes drinking from it a good thing to do. But to act on a fact you must

be aware of that fact, which is to know the fact. You need knowledge; not even blameless true belief is enough. For example, if your blameless true belief that the glass contained water were based on your blameless false belief that you could see the water, when it was a trick opaque glass with water in it, the fact that the glass contained water would be outside your awareness, since you lacked knowledge. Thus in order to act on a reason, you must *know* the fact that is the reason. For acting on reasons, what matters is knowledge, not belief.

If knowledge can explain action, how do we explain the actions of an agent who merely believes but does not know?

The agent who merely believes acts as *if* on known facts. They are in a state that resembles knowledge in its immediate effect on action. If you did not know that the glass contained water, you were not in a position to act on the fact that it contained water (even if there was such a fact), but you could act as if on the fact that it contained water, if you believed that it contained water. Thus the central case is reason-giving explanation, in which we explain why an agent did something by citing facts known to the agent that made it a good thing to do, but the central case is surrounded by a mass of somewhat similar cases that deviate from it more or less because things went more or less wrong, in one way or another. That the agent has mere belief rather than knowledge is one common deviation. Another is that the agent merely tried to do something, but did not succeed. These defective cases do not fit the original pattern, but we can nevertheless understand them as deviations from it.

Mere belief is to be understood as a deviation from knowledge. To believe is to be in a mental state similar to knowing in its immediate effects on action, but which differs from knowing in other respects. To work with such an account is to understand belief in terms of knowledge, rather than knowledge in terms of belief.

Mental life is a bewildering complex of interacting processes. The key to understanding the nature of these processes is to focus on what happens when things go right. For that, we need the notions of knowing and doing. Having seen the point of these processes, we must then go on to understand all the ways in which things can go more or less wrong. For that, we need the notions of believing and trying.