Knowledge and Belief: Preliminary Notes

Timothy Williamson, University of Oxford

These notes are not intended to be fully self-explanatory, but they may be helpful in giving participants in the class a better clue as to what will be discussed.

Meeting 1: Knowledge, belief and conceptual analysis.

Knowledge as the precondition for intelligent life.

Knowledge, belief and truth: some basic interrelations

Knowledge entails truth (knowledge is *factive*): always, if S knows that P, then P.
There is no false knowledge. People didn't know that the earth was flat; they merely believed (falsely) that they knew that the earth was flat.

Belief does not entail truth (belief is *non-factive*): sometimes S believes that P, but not P.
There is false belief. People really did believe (falsely) that the earth was flat.

Consequence: belief does not entail knowledge: sometimes S believes that P without knowing that P.

Knowledge entails belief: always, if S knows that P, then S believes that P.
[?] Qualm: The unconfident examinee who reliably reproduces correct information which he has forgotten ever having been taught, under the impression that he is just guessing. But consider the variant unconfident examinee who reliably reproduces incorrect (mis)information which he has forgotten ever having been taught, under the impression that he is just guessing.

Consequence: knowledge entails true belief: always, if S knows that P, then S believes truly that P.

True belief does not entail knowledge, sometimes, S believes truly that P without knowing that P. Example: S believes that the surname of the Prime Minister on 18.1.2006 began with a 'B' because S believes that Gordon Brown was Prime Minister on 18.1.2006.

Truth does not entail knowledge: sometimes P although nobody ever knows that P.
The difficulty of giving an example of an unknown truth: To know that it is an unknown truth that P, one must know that it is a truth that P (since knowledge of a conjunction implies knowledge of its conjuncts), in which case it is not an *unknown* truth that P, so one does not *know* that it is an unknown truth that P after all (since knowledge is factive). Thus one cannot know that it is an unknown truth that P.

However, we can still know that *there are* unknown truths. For instance, we can know that either it is an unknown truth that E or it is an unknown truth that O — but we cannot know which:
E: The number of books in TW's room on 1.1.2000 was even.
O: The number of books in TW's room on 1.1.2000 was odd.
In fact, the previous arguments show that we can know that that there are unknow*able* truths. For instance, we can know that either it is an unknowable truth that E* or it is an unknowable truth that O* — but we cannot know which.
E*: It is an unknown truth that the number of books in TW's room on 1.1.2000 was even.
O*: It is an unknown truth that the number of books in TW's room on 1.1.2000 was odd.

Truth does not entail belief: sometimes P although nobody ever believes that P.
Just as we cannot give an example of an unknown truth, we cannot give an example of an unbelieved truth (if knowledge entails belief, any unbelieved truth is an unknown truth). Nevertheless, we know that there are unbelieved truths concerning matters on which nobody ever forms beliefs at all (for example, concerning microscopic events millions of years ago).

Knowledge and Analysis

Since true belief is necessary but not sufficient for knowledge, can we say that there is something (X), which must be added to true belief to make knowledge? If so:
Knowledge = true belief + X

X = justification?
Gettier cases: In the example above of true belief without knowledge, we can imagine an elaborate hoax which makes S's belief that Gordon Brown was Prime Minister on 18.1.2006 justified. A real life Gettier case.

Analyses of the concept *K* vs. analyses of K:
"Water = H2O" states an analysis of water, not of the concept *water*.
"A bachelor is an unmarried man" states an analysis of the concept *bachelor*, not of a bachelor.
But Gettier cases refute both JTB analyses of knowledge and JTB analyses of the concept *knowledge*.

Other values of "X"? (justification with no false premises; causal connection between belief and truth; ….)

Being red = being coloured + Y
What is Y? The problem of circularity (also as applied to justification).

Two starting-points for explanation:
Success-neutral (e.g. belief, structure)
Success-oriented (e.g. knowledge, function)

Meeting 2: The Problem of Scepticism

Scepticism arises from the generalization of intellectual habits which we have and which seem valuable in some ordinary cases. Example: The elimination of prejudices about race, gender, sexual orientation. We test beliefs by assessing them on the basis of independent evidence. If we cannot find independent evidence for them, shouldn't we try to drop them? What happens when we apply this method to belief in an external world? Or to the belief that there can be good reasons for belief? Thus we can't simply ignore scepticism, because arguments for it seem to be implicit in our own ordinary ways of thinking about the world and our knowledge of it. Even if we are sure that scepticism is wrong, we need to know where those arguments go wrong: what false assumptions or invalid methods of argument are built into our own ordinary ways of thinking?

Sceptical arguments are sometimes thought simply to establish *fallibilism* ("Nothing is certain" — presumably they don't establish it with certainty).

A more exact statement of fallibilism:
The only 100% probable propositions on one's evidence are logical truths.

Notes:
1. There are many definitions of 'fallibilism', as of most philosophical terms. This one is convenient for our purposes.
2. We have to make an exception of logical truths (such as "If I have hands then I have hands") because the axioms of mathematical probability theory require them to be 100% probable.

Fallibilism seems easy to live with because it is consistent with assigning a probability of 99.9999% to common sense claims such as "I have hands".

Good case: Things are as they appear to me; I appear to have hands and I do have hands.
Bad case: Things are not as they appear to me; I appear to have hands but really I'm a handless brain in a vat beings electronically stimulated so that things appear to me just as they do in the good case.

The sceptic argues that our evidence is neutral between the good case and the bad case; we acquire the same evidence in the two cases. Thus it is sheer prejudice to assign a higher probability to "I am in the good case" than to "I am in the bad case". Since the two cases are mutually incompatible, their probabilities sum to at most 100%. If the probabilities are equal, each of them is therefore at most 50% (otherwise the probability that I am in either the good case or the bad case is more than 100%, which is impossible). Roughly: it is no more likely on my evidence that I have hands than that I'm a BIV.
(It may get worse: consider different bad cases in which the brain is floating in differently coloured liquids. Complications arise in assigning probabilities when scenarios are not maximally specific. But they don't seem to help the anti-sceptic.)

Thus sceptical arguments lead to conclusions far more radical than fallibilism. Presumably scepticism entails fallibilism but not *vice versa*.

An objection to fallibilism:
Consider one's evidence itself. It does not just consist of logical truths (not all sceptical hypotheses are on a par). But trivially one's evidence is 100% probable on itself. Thus evidence propositions are counterexamples to fallibilism and therefore to scepticism.

The objection does not refute fallibilism and scepticism as restricted to some specific class of propositions (e.g. propositions about the external world) provided that the evidence for propositions in the class does not itself consist of propositions in the class.

What sort of propositions can be evidence? Propositions about one's own present experiences? But are such propositions really so certain? Why can't our evidence include propositions about the external world, e.g. "I see that I have hands" [which entails "I have hands"] not just "I appear to myself to be seeing that I have hands".

"I see that I have hands" isn't part of one's evidence in the bad case, since the BIV doesn't have hands. But why shouldn't "I see that I have hands" be part of one's evidence in the good case? If so, our evidence in the good case is not neutral after all between the two cases, and the sceptical argument fails. Our evidence does not consist only of appearances. Call this the *realist view of evidence*.

The sceptic may object to the realist view of evidence that it implies that the BIV in the bad case will not be in a position to recognize that "I see that I have hands" isn't part of its evidence. That is an objection only on the assumption that we must always be in a position to distinguish what is part of our evidence from what isn't. Initially, that assumption seems plausible. What use is evidence if we can't tell whether we have it?

However, consider a gradual process of change from time T0 to time Tn, where the interval between times Ti and Ti+1 is very short:
T0 T1 T2 T3 T4 ……Tn-1 Tn
Suppose that for each i from 0 to n-1, your evidence is so similar at Ti and Ti+1 that for all you can know it is exactly the same [in some relevant respect]. If your evidence at Ti is *exactly* the same [in that respect] as your evidence at Ti+1 for each i from 0 to n-1, then your evidence at T0 is exactly the same as your evidence at Tn [in that respect] (exact sameness is a transitive relation). But we can choose the process to be one in which your evidence at Tn is massively different from your evidence at T0. Thus there must be at least one i (in fact, many) for which your evidence at Ti+1 is different from your evidence at Ti, even though by hypothesis you can't know that there is a difference. Thus *whatever* evidence is, it has aspects which we cannot know. Hence the sceptical objection to the realist view of evidence depends on an unsatisfiable conception of evidence — ironically, one that assumes more knowledge of evidence than we can have.

Meeting 3: The nature of justification

Epistemic and non-epistemic notions of justification

Non-epistemic justification applies to all sorts of actions as well as to beliefs. Unlike non-epistemic justification, epistemic justification applies only to what has a truth-evaluable content, such as beliefs (for most actions the question of truth or falsity does not arise). For example, someone may be non-epistemically justified in believing that they will recover from a disease (because that belief maximizes their chance of recovery) without being epistemically justified in believing that they will recover from the disease (because all their evidence indicates that they are very unlikely to recover). Similarly, might one be epistemically but not non-epistemically justified in believing that one will not recover.

Is theistic belief a case in point? Pascal's Wager: The expected utility of believing that God exists is higher than the expected utility of not believing that God exists, because if God exists the utility of believing is infinitely higher than the utility of not believing, whereas if God does not exist there is at most a finite advantage in not believing, and there is at least a small positive probability that God exists. Thus (?) one is non-epistemically but not epistemically justified in believing that God exists.
Problem: Belief is not under one's voluntary control. Solution (?): Hang out enough with a religious crowd and you will eventually acquire a genuine belief that God exists.
You can make yourself believe that you are scratching your nose by scratching your nose.
Even if beliefs are not under voluntary control, we can make some normative distinctions between rational and irrational beliefs.

For the rest of the lecture, 'justification' will refer to epistemic justification.

Two notions of justification:

S is dialectically justified in believing *p* ↔ S has good reasons for believing *p*.

S is normatively justified in believing *p* ↔ S is above reproach in believing *p*.

Both 'reasons' and 'reproach' should be understood here as qualified by 'epistemic'.
The notion of *a* justification seems to fit dialectical justification better than it does normative justification.

How could one have good reasons without in any sense being justified in believing those reasons? Let us try Hypothesis 1:
Having good reasons entails being dialectically justified in believing those reasons.

By Hypothesis 1, if one has good reasons, one must have good reasons for believing those good reasons, and so on …. It does not seem psychologically possible to have infinitely many reasons. Even if one could, why would they all count as good reasons

rather than bad reasons? Consider someone who for each natural number *n* believes that there are at least *n* goblins by deduction from the premise that there are at least *n* + 1 goblins. If one has only finitely many reasons, one's chains of reasons must go in a circle somewhere (in the extreme case, a circle with only one member), but circular reasons do not seem to be good reasons. Even if circles of reasons have a sort of coherence (the parts all support each other), that seems to be too weak to justify belief — too many incompatible circles are possible. Thus Hypothesis 1 seems to make it impossible to have good reasons. This motivates rejecting Hypothesis 1 in favour of Hypothesis 2:
Having good reasons entails being normatively justified in believing those reasons.

If Hypothesis 2 is true and Hypothesis 1 is false, then being normatively justified does not entail being dialectically justified (is it a professional deformation of philosophers to over-emphasize dialectical justification, since they make their living with it?) Hypothesis 2 seems to generate no regress. But what kinds of beliefs are normatively justified without being dialectically justified?
Non-inferential perceptual and memory beliefs are obvious candidates.

Internalism about justification: intrinsic duplicates are normatively justified in believing exactly the same propositions. E.g. your unlucky twin in a sceptical scenario is justified in believing exactly what you are justified in believing, even though in your case the beliefs constitute knowledge and in your twin's case they do not.

Internalism about justification seems to be motivated by the following claim:

(*) Differences epistemically inaccessible to the subject make no normative difference.

The idea is that any difference between your case and your unlucky twin's is epistemically inaccessible and therefore makes no normative difference.

Problems for the motivation for internalism about justification:
The differences are inaccessible to your twin, but are they to you if scepticism fails? On the view of evidence proposed last time, you have more evidence than your twin, which should make a difference to your justification.
Principle * collapses normative distinctions, because lots of small epistemically inaccessible differences can add up to a big epistemically accessible difference.
Other normative distinctions are sensitive to extrinsic differences ('moral luck').

A form of externalism about justification: One is normatively justified in believing *p* if and only if one knows *p*. If you know *p*, you are above reproach in believing *p*. If you don't know *p*, you are not above reproach in believing *p* (but a BIV has a good excuse). Degrees of justification are, roughly, degrees to which one comes close to knowledge (just as the question 'How full is this glass?' is a question about how close the glass comes to being full).

Meeting 4: Semantic internalism and externalism, and their implications for justification

A state is narrow if and only if, necessarily, every intrinsic duplicate of something which is in the state is also in the state.
A state is broad if and only if it is not narrow.

Being round is a narrow physical state; being a football and being British are broad states
Being in pain may be a narrow mental state; loving Griselda is a broad state.

Internalism about (core) mental states: all (core) mental states are narrow.
Externalism about (core) mental states: not all (core) mental states are narrow.

Non-core mental states are hybrids of core mental states with environmental conditions (perhaps on the external causes of the core mental states).

On a traditional internalist picture, believing that one is holding a glass of water is a narrow state (e.g. one is in it whether one is in a suitable everyday scenario or a brain in a vat), whereas knowing that one is holding a glass of water is a broad state (one is in it in the former scenario but not the latter). According to internalism about justification, being justified in believing that one is holding a glass of water is also a narrow state, which may be counted as a core mental state. Consequently, knowing that one is holding a glass of water is at best a non-core mental state, a hybrid of core mental states such as believing and being justified with external conditions such as that one *is* holding a glass of water. This picture provided further motivation for the project of attempting to analyze knowing into components such as believing, being justified and truth.

For a Putnam-inspired externalist argument that believing that one is holding a glass of water is *not* a narrow mental state, see overleaf. Hard-line internalists about core mental states *either* try to wriggle out of the Putnamian argument *or* treat believing that one is holding a glass of water as itself a non-core mental state, a hybrid of core mental states with external conditions. But what are those core mental states?

If believing that one is holding a glass of water is a core mental state, why not knowing that one is holding a glass of water too? Putnam's argument is for externalism about the *content* of (core) mental states, but we can also consider externalism about the *attitudes* to the contents, even granting internalism concerning the contents themselves (is knowing that there are other minds an example?). Such externalism undermines the motivation for the attempt to analyse knowing in terms of believing.

Genuine core mental states play a significant role in the causal explanation of action. Do states of knowing play such a role? Not much in explaining short-term effects (explaining a walk step by step) but a significant one in explaining long-term effects (explaining how someone reached home in terms of their knowing the way).

An adaptation of Putnam's Twin-Earth thought experiment (in 'The meaning of "meaning") to belief rather than meaning

(1)	Believing that one is holding a glass of water is a narrow state.	Assumption to be reduced to absurdity
(2)	Oscar is holding a glass of H20.	Assumption
(3)	Twin-Oscar is not holding a glass of H2O.	Assumption
(4)	Oscar and Twin-Oscar are in the same narrow states.	Assumption
(5)	Oscar believes that he [Oscar] is holding a glass of water.	Assumption
(6)	Water = H2O.	Fact
(7)	Oscar is holding a glass of water.	(1), (6)
(8)	Twin-Oscar is not holding a glass of water.	(2), (6)
(9)	Twin-Oscar believes that he [Twin-Oscar] is holding a glass of water.	(1), (4), (5)
(T)	If S believes that P, S believes truly that P iff P.	Axiom
(F)	If S believes that P, S believes falsely that P iff not P.	Axiom
(10)	If Oscar believes that he [Oscar] is holding a glass of water, Oscar believes truly that he [Oscar] is holding a glass of water iff he [Oscar] is holding a glass of water.	(T)
(11)	If Twin-Oscar believes that he [Twin-Oscar] is holding a glass of water, Twin-Oscar believes falsely that he [Twin-Oscar] is holding a glass of water iff he [Twin-Oscar] is not holding a glass of water.	(F)
(12)	Oscar believes truly that he [Oscar] is holding a glass of water.	(5), (7), (10)
(13)	Twin-Oscar believes falsely that he [Twin-Oscar] is holding a glass of water.	(9), (8), (11)

Is semantic externalism compatible with internalism about justification?

Let ‘twater’ be a natural kind term for the XYZ-based substance just as ‘water’ is for the H2O-based substance.

Oscar believes that there are pools of water.
Twin-Oscar does not believe that there are pools of water.
Twin-Oscar does not believe that there are no pools of water.
Twin Oscar believes that there are pools of twater.
Oscar does not believe that there are pools of twater.
Oscar does not believe that there are no pools of twater.

<u>Externalism about what justified beliefs one has</u>
Oscar has a justified belief that there are pools of water
[he is swimming in one in normal conditions of observation].
Twin-Oscar does not have a justified belief that there are pools of water
[he does not have a belief that there are pools of water].
Twin-Oscar has a justified belief that there are pools of twater
[he is swimming in one in normal conditions of observation].
Oscar does not have a justified belief that there are pools of twater
[he does not have a belief that there are pools of twater].

<u>Internalism about what beliefs one is justified in having</u>
Suppose that ‘S is justified in believing that P’ does not entail ‘S believes that P’.
Could Oscar and Twin-Oscar be the same in what beliefs they are justified in having, even though they differ in what justified beliefs they have (as Audi suggests)? Thus:
Oscar is justified in believing that there are pools of water, so
Twin-Oscar is justified in believing that there are pools of water.
Twin-Oscar is justified in believing that there are pools of twater, so
Oscar is justified in believing that there are pools of twater.

Problem for internalism about what beliefs one is justified in having:
In Oscar’s world there are no pools of twater, and in Twin-Oscar’s world there are no pools of water. Thus the beliefs which Oscar and Twin-Oscar are supposedly justified in having are false in their respective worlds. When one is justified in believing falsehoods, one’s evidence is in some way misleading. But although Oscar and Twin-Oscar’s evidence is incomplete, it is not relevantly misleading.

Internalist patch: Given semantic externalism, Oscar is conceptually incapable of believing that there are pools of twater, and Twin-Oscar is conceptually incapable of believing that there are pools of water. So we might try to avoid the above problem by restricting internalism about what beliefs one is justified in having to beliefs which one is conceptually capable of having.

Externalist reply: Suppose that a traveller once showed Oscar a small bottle containing a few drops of twater and told him that it was a rare liquid, existing only in droplet form,

superficially like water but of a very different underlying nature. Similarly, a traveller once showed Twin-Oscar a small bottle containing a few drops of water and told him that it was a rare liquid, existing only in droplet form, superficially like twater but of a very different underlying nature. Now Oscar is conceptually capable of believing that there are pools of twater, but he is still not justified in believing that there are pools of twater. Similarly, Twin-Oscar is conceptually capable of believing that there are pools of water, but he is still not justified in believing that there are pools of water. Thus semantic externalism effectively forces externalism about what beliefs one is justified in having.

Semantic externalism and sceptical scenarios

According to Hilary Putnam, BIVs lack the causal connections to brains and vats required for thinking that they are not BIVs. The belief (if any) that a BIV expresses by tokening 'I am not a BIV' is not the false belief that it is not a BIV but a different, true belief. Thus one cannot falsely believe that one is not a BIV.

Note that this relies on a much stronger form of semantic externalism (a causal theory of reference) than anything that Twin-Earth scenarios establish.

Moreover, Putnam's argument does not work for a recently envatted BIV, since it has the requisite causal connections to brains and vats; by 'I am not a BIV' it does mean that it is not a BIV. Thus Putnam's argument is not a generally effective anti-sceptical strategy.

Singular thoughts: Nevertheless, even a scenario of recent envatment makes some difference to content. I believe that *that* [looking at a passing fly] is circling. The corresponding BIV does not believe that *that* is circling, because it is not thinking of *that*; if it means anything by '*That* is circling', it means something different. I am justified in believing that *that* is circling; the BIV is not justified in believing that *that* is circling (whether or not it encountered *that* before it was envatted, and so is conceptually capable of thinking thoughts about it. Thus what I am justified in believing differs from what even the recently envatted BIV is justified in believing.

Externalism and internalism about perceptual content

Internalism about perceptual content is the view that how things perceptually appear to be is the same for intrinsic duplicates. On this view, since it does not visually appear to the BIV that *that* is circling, it does not visually appear to me that *that* is circling. Thus perceptual appearances are neutral as to the identity of the objects perceived (if any). But I have perceptual knowledge that *that* is circling. If perceptual appearances are neutral as to which thing is circling, how do I know which thing is circling? Do I know *a priori* that if exactly one fly is circling, *that* is circling? How can I rule out *a priori* the possibility that what is circling is some other fly? According to (some) externalists about perceptual content, it visually appears to me that *that* is circling, even though it does not so visually appear to the BIV: I do not need to infer which fly is circling.