PHL232 Handout 4: Externalism and Internalism

§1 Gettier Cases
Recall the JTB analysis of knowledge: S knows that p iff
  i.   p is true
  ii.   S believes that p
  iii. S's belief that p is justified
Supposedly this was the standard view of knowledge from Plato to the mid-20th Century. But then Edmund Gettier published a set of counterexamples. These were scenarios in which (i)-(iii) appear satisfied yet S's belief that p does not count as knowledge (instead it counts as merely *luckily* true belief).

Gettier's cases force epistemologists to choose one of four responses:
    1. Retain the JTB analysis, but revise our account of justification to avoid Gettier cases
    2. Extend the JTB analysis with conditions that supplement (i)-(iii)
    3. Reject the JTB analysis, and put something else in place of justification
    4. Deny that knowledge admits of analysis
We've already seen Nozick's attempt to pursue 3. To save our sanity, we won't investigate 2. 4 has become popular since Williamson (2000), but generates special complications. Today's topic is an influential attempt to pursue 1: Goldman's *process reliabilism*.

§2 Externalism and Scepticism
Goldman's view of justification commits him to externalism about justification:

> *Externalism about Justification*: S's belief that p can be justified even if she does not know that it is justified.

This view entails that the features responsible for a subject's justificatory status need not be *accessible* to the subject (cf. p. 99). Combined with the claim that justification is necessary for knowledge (cf. p. 90), externalism forces us to reject the *KK-Principle*: S knows that p only if S knows that she knows that p. Yet to reject externalism is to adopt some version of the following:

> *Internalism about Justification*: S's belief that p is justified only if she knows that it is justified.

And this is a difficult view to defend (though see the recommended Smithies paper for a valiant attempt).

Externalists argue that knowing that p requires much less than knowing that we know that p. By decoupling the two, they believe a case can be made that we know that we are not brains in vats – we just can't *know that we know* that we are not brains in vats:

> Our temptation to agree with the sceptic arises from the fact that we can't know that we know that we're not brains in vats. But I *can* know that I am not a brain in a vat: any belief to the effect that I am not a brain in a vat will be caused by processes that reliably track whether or not I am a brain in a vat (e.g. perception).

§3 Process Reliabilism
We must distinguish reliabilism about justification (Goldman) from reliabilism about knowledge (cf. Kornblith)
1. *Reliabilism about Justification (RJ)*: a subject's belief that p is justified iff it is caused by a reliable method (i.e. a method that reliably generates true beliefs).
2. *Reliabilism about Knowledge (RK)*: a subject knows that p iff p is true, S believes that p, and S's belief was caused by a reliable method.

RJ usually leads to RK, but the converse doesn't hold. Proponents of RK might just eschew justification altogether. RJ entails externalism about justification: to have a belief caused by a reliable method does not entail that we know whether the belief was caused by a reliable method.

Goldman adopts a particular version of RJ, namely *Process Reliabilism* (cf. p. 95):

> *Process Reliabilism:* S's belief that p is justified iff her belief was caused by a process that reliably generates true beliefs

Apart from the general virtues of the account, Goldman does provide a somewhat direct argument for the view in his paper:

> *The Argument* (p. 95):
> 1. Gettier cases (along with other examples) show that whether a belief counts as justified depends upon how it was caused
> 2. So only those beliefs caused in the right way will count as justified (from 1)
> 3. Process reliabilism best explains why justification is so sensitive to the causal mechanism responsible for belief-formation
>   a. *Further evidence*: our judgments about comparative justification track our beliefs about the comparative reliability of the processes that give rise to the justified beliefs (cf. pp. 95-96)
> 4. So process reliabilism is true (from 2 and 3)

Further Question: One virtue Goldman claims for *Process Reliabilism* concerns its credentials as a naturalistic and reductionist account of justification (cf. p. 90). But why ought we to assume that a right account of justification must be given in 'non-epistemic terms'?

> *Note*: Goldman's process reliabilism has much in common with the virtue-theoretic conception of justification we saw in Sosa. Consider which account you find more attractive, and why. Are Goldman's arguments available to Sosa, or *vice versa*?

Foundationalism and Reliabilism: on pp. 98-99 Goldman notes that we might call his account a kind of 'diachronic foundationalism' (in contrast, Sosa rejected both coherentism and foundationalism). Try to compare Goldman's version of foundationalism with the sort of classic foundationalism discussed by Sosa.

Counterfactual Conditions and Reliabilism: process reliabilism satisfies Nozick's counterfactual conditions for knowledge, as well as the less demanding 'safety condition' (roughly: S knows that p only if p would be the case were S to believe that p). So we should ask: is process reliabilism correct because it satisfies these conditions, or are (some) of these conditions correct because they track process reliability?

§6 Objections

Generality Problem
Across how many scenarios must our belief-forming mechanism prove reliable? Too many, and reliabilism will make justification too scarce; too few, and the theory will make justification too easy to acquire.

Individuation of Methods
We can re-describe putative methods of belief acquisition such that on some descriptions the process comes out as reliable, in others it does not.

> Recall Russell's example of the broken clock that points to 2 o'clock. Here are two ways we might describe the process by which the subject's belief was formed:
> 1. Believe that it is time t if the clock indicates that it is t
> 2. At 2 p.m., believe that it is time t if the clock indicates that it is t
>
> In Russell's case, a belief formed according to 1 would count as unreliable (since beliefs formed in accordance with this process a minute before or after 2 p.m. will be false). In contrast, a belief formed in accordance with 2 will be reliable (since the clock points to 2 o'clock, it will always give the right time at 2 p.m.).

Clearly reliabilists must provide an account of the individuation of belief-forming processes that blocks such re-description (*Note*: Nozick's counterfactual analysis runs into a similar problem, since the counterfactuals are evaluated relative to a method of belief-formation).

> He treats belief-forming processes as information processing done by an organism's cognitive system.
>
> > *Question*: Why should we restrict belief-forming processes to our cognitive systems? Surely the first stages in a belief-forming process might be external to the creature (e.g. turning one's head to explore the environment)? So perhaps a belief-forming process must include some processing from our cognitive systems, yet needn't be exhausted by this processing. But if external factors are allowed into belief-forming processes, the problem of the individuation of methods becomes particularly acute.

Knowledge vs. Reliable Superstition
If process reliabilism is correct, then any belief formed on the basis of a reliable process will count as justified (and thus, if true, as a case of knowledge). Yet this generates some counterintuitive results.

> *Reliable Superstition*: Imagine a woman whose knee aches every time the Dalai Lama flies in a plane. And, as it happens, this knee ache always causes her to believe that the Dalai Lama is on a flight somewhere (perhaps she is bizarrely superstitious). She has no reason to think that these beliefs are reliable, and in fact thinks her tendency to form these beliefs is somewhat perverse. [Cf. Bonjour]

Do these beliefs count as knowledge? Surely not. Yet process reliabilists must say that they do. Furthermore, we don't have to look to fancy made-up cases to generate strange consequences for the reliabilist:

> *Blindsight*: Some unfortunate victims of brain damage suffer from a condition known as *blindsight*. When asked, patients with blindsight report that they are blind to the goings-on in a portion of their visual field. Yet – and this is bizarre – once forced to answer these patients are very reliable with respect to objects that occupy the 'blind' portions of their visual fields. Should we say that these subjects know facts about objects that, as far as they are concerned, lie outside what they see?

Goldman replies to these sorts of cases on pp. 100-101. Check out what he says, and consider whether you find his responses convincing.

Moore's Paradox

G. E. Moore observed that there is something irrational or absurd about someone who asserts 'It is raining, but I don't believe that it is raining'. Yet the assertion doesn't seem inconsistent: that p is true does not entail that anyone believes that p (nor does belief that p entail that p – belief isn't factive). This tension between the absurdity of the assertion and its apparent consistency is *Moore's Paradox*.

A solution to Moore's Paradox would explain why the relevant assertions are irrational (and why this irrationality does not apply to sentences such as 'It is raining, but *you* don't believe that it is raining'). The literature on this topic is vast, and growing.

Unfortunately, Moore's Paradox does not appear confined to belief. Smithies (2012) observes that it would also be irrational to assert 'I have justification to believe that p yet I don't have justification to believe that I have justification to believe that p' (shorthand: 'Jp and not-JJp'). He argues that parallel cases crop up whenever we deny either of the following two biconditionals (*note*: here 'J' is meant to track what Smithies calls 'propositional justification' and Goldman calls '*ex ante* justification'):
   1. Jp if and only if JJp
   2. Not-Jp if and only if J(not-Jp)
But these biconditionals are inconsistent with reliablism. A subject might be reliable with respect to p without being reliable with respect to whether they are reliable with respect to p.

> Smithies argues that only internalism about justification can explain the irrationality present in the justification versions of Moore's paradox.

Mathematical Knowledge

The problem (briefly): given that our best account of mathematical truth requires the existence of causally inert mathematical objects, how can we accommodate mathematical knowledge? Many take this question to force a choice between a plausible metaphysics and a plausible epistemology. This forced choice has become known as *Benacerraf's Dilemma* (cf. Benacerraf 'Mathematical Truth').

> Benacerraf's dilemma was originally formulated in terms of the causal theory of knowledge – the view that S knows that p only if she bears an appropriate causal relation to what makes p true. Goldman endorses the causal theory on p. 99.