Chapter 4
Human Knowledge and Reflection
Is reflection essential to knowledge?

- One long-standing view is that knowledge requires the human capacity for reflection on our beliefs.
- Socrates claimed that “the unexamined life is not worth living.” The naturalistic antidote might be Daniel Dennett’s quip that “the over examined life is nothing to write home about either.”
- Kornblith notes that some philosophers, like Ernest Sosa, who seem to accept the existence of animal knowledge make such a strong distinction between animal knowledge and human knowledge, that it effectively makes human knowledge the only thing worth taking seriously in epistemology:
  - ‘no human being blessed with reasons has merely animal knowledge of the sort attainable by beasts’ for a ‘reason-endowed being automatically monitors his background information and his sensory input for contrary evidence and automatically opts for the most coherent hypothesis even when he responds most directly to sensory stimuli’ (KPN, p.105)
Descartes on the role of reflection

- Descartes is the modern philosopher best known for emphasizing the necessity of reflection for establishing knowledge. And Kornblith returns to Descartes, perhaps because he ultimately accomplished so little by his methods.
- In the *Meditations* Descartes claims that reflection both revealed that his current epistemic situation is problematic (i.e., he had simply accepted the majority of his beliefs without ever reflecting on their foundation) and also was needed to remedy this situation.
- As Kornblith points out, Descartes accepted an essentially *historical* theory of justification: whether a belief is justified depends on how it got into your head in the first place. For Descartes, the only way a belief can be justified is if you get it through a reflective process.
Impatient Cartesians

- So as not to drag the reader through the nonsense of rejecting all of ones current beliefs and then re-adopting only those that can be established with absolute certainty, Kornblith attempts to capture what might be useful in the Cartesian approach, namely reflection on our current habits of belief acquisition with the aim of ridding ourselves of those that are not reliable.

- Kornblith then identifies two contemporary epistemologist whom he believes may be fairly characterized as impatient Cartesians: Roderick Chisholm, and Lawrence Bonjour.

- Chisholm is a contemporary foundationalist, which means that while he understands that there is no such thing as an absolutely reliable foundation for our beliefs, still we only justified in our beliefs if we reflectively trace our beliefs to their epistemic foundation, and then use our a priori intuitions to determine that the foundation consist only of propositions whose truth is directly evident.

- Bonjour, by contrast, is (was, actually) a contemporary coherentist, which means that he rejects the existence of a foundation and claims that justification is a matter of establishing that our belief system is internally coherent. (The minimum condition of coherence is logical consistency. Other criteria like explanatory power, simplicity, unification, fecundity etc. are relevant as well.)
Although foundationalism and coherentism are at odds with respect to the nature of justification, they agree on the role of introspection.

This is the point that you need to dial in: From a naturalistic point of view it would seem that whether your belief system is well-founded or coherent should be an objective fact about you. In other words, it would seem one may have a well-founded or coherent belief system whether or not you have reflected on this fact.

But according to Chisholm and Bonjour you do not actually have knowledge until you have engaged in this reflective process. So Descartes would be proud.

We should note here that Kornblith begins using the term ‘internalism’ which contrasts with ‘externalism”, though he does not explicitly define either one. So, roughly speaking:

- An internalist is someone who believes that one has knowledge only if one has established the existence of certain epistemic relations internal to one’s system of beliefs.
- An externalist is someone who believes that knowledge is an objective property that does not depend on the knower’s cognitive access to internal epistemic relations.

For the most part, philosophers who believe that justification is necessary for knowledge are internalists, and philosophers who believe that reliability is necessary for knowledge are externalists.
Kornblith’s critique of introspectionist, internalist epistemology begins with empirical data on the unreliability of introspection. The point here is that there is an abundance of data showing that people are shockingly introspectively unaware of the basis of their beliefs. As evidence of this he notes

- the “position-effect” which causes people to make differential claims about the quality of a product simply by changing its position in a display.
- The baseline and framing effects, which causes people to have very different intuitions about the correct answer to a question depending on how it is posed. For example, people give opposite answers concerning the desirability of a particular procedure or policy depending on whether it is framed in terms of how many lives it will save or how many people live it will fail to save.

In general it is an utterly trivial matter to create situations in which people express great introspective certainty in these judgments they have in fact been manipulated by external means.

The methodological point is that in these cases no amount of introspection is sufficient to rectify the error.
Kornblith also notes that there is considerable empirical evidence that introspection does not reliably reveal the reasons for our beliefs.

- Subjects who read an emotionally significant story were then to rate its emotional appeal and then identify the aspects of the story that contributed to this appeal. But when those aspects were removed from the story other subjects rate the story as having the same level of emotional appeal.

- People on a plane believe that they are detecting changes in the plane’s motion and orientation by visual means when in fact they are doing it kinesthetically.
The standard response to this kind of data is that it demonstrates only that introspection is fallible. But everything is fallible, so what?

Kornblith’s claim, however, is that introspection is so fallible, that it does not meet basic reasonable standards of reliability.

The basic problem is that introspection simply does not provide an adequate prophylactic against very basic and robust biases in hypothesis testing. Kornblith discusses several well-known examples of confirmation bias, the upshot of which are that virtually all people show an introspectively undiscoverable tendency to dismiss or discount evidence against a view that they currently hold, and to embrace or find plausible evidence for a view that they currently hold.

Kornblith quote Ross and Nisbett on the results of such a study where subjects where subjects were asked to read articles providing evidence for the deterrent value of capital punishment:

- ‘One might expect that...if the evidence presented were mixed that it would serve to moderate opposing views. Instead, the effect of introducing mixed evidence may be to polarize public opinion.’

In other words, people often seem to feel the greatest introspective certainty for views for which the least evidence. (Something one routinely observe in political debates.)
How we discover mistakes

Kornblith’s basic point then is that we do not typically discover our mistakes through introspection. In fact, we seem to be protected from discovering our mistakes in this way. Our reflective capacities seem to have the primary function of rationalizing what we currently believe, rather than calling it into question.

The way we discover the mistakes due to inherent confirmational biases is to have them pointed out by others, and this requires acknowledging the legitimacy of a fundamentally externalist perspective that focuses on the actual reliability of ones belief formation processes rather than their degree of internal justification.
The Cartesian theater

- As descendants of the Cartesian tradition we still grant high intuitive plausibility to Descartes’ view of the mind as a kind of private theater:
  - Only you really know what is going on inside.
  - Other people can’t know what you are feeling.
  - Other people can’t know what you believe.

- Of course, we needn’t deny that introspection provides a kind of access to our own beliefs and feelings that others do not possess, but that is a far cry from saying that introspection is the only truly reliable method of knowing this:

- We’ve already seen that introspection is highly fallible. The reason for this is that introspection really does not give us direct access to our cognitive processes at all. There are two reasons for this, that Descartes could never have appreciated.
Why introspection is fallible 1

- **Reason 1:** The vast majority of our thinking is subcognitive. This is most easily grasped by considering the fact that processing visual information is a highly inductive process that we do without any conscious reasoning at all. But this applies well beyond vision. In the blink of an eye your eyes can provide you with the belief that a car has swerved into your lane. It is only afterwards that anything like conscious reasoning can take place. Reasoning with consciously held beliefs is an extremely slow and cognitively expensive process. It is has specific purposes, but it is not suited to the vast majority of thinking that is required of us.

- It’s useful to think a little bit about the history of science in this context. The process of human enlightenment begins with an awareness of some phenomenon that has high value to human beings and we naturally assume that it will be explained by principles that have quite general application.
  - The earth, for example, has extraordinary value to humans, and so we naturally think that it has extraordinary importance in the universe as a whole. Of course, the recognition that the earth is a tiny planet revolving around an ordinary star in an ordinary galaxy, of which there are probably about 100 billion others tends to diminish this feeling somewhat.
  - Humans have extraordinary value to human beings, so we naturally explain their genesis as the primary purpose of the universe.

- Similarly then, with conscious thought. When philosophers first started thinking about thoughts, of course they naturally assumed that conscious thoughts were the big Kahuna, directing everything that went on in the mind. But once we grasped that the mind is the activity of an actual physical organ, and gained a glimmering of how it actually works, we slowly began to realize that conscious thoughts are just the tip of the iceberg.
Why introspection if fallible 2

- When we try to recall the reasons for our beliefs (as in the product preference experiments) we do not actually have direct introspective access to those reasons. This is partly because we were never actually aware of those reasons in the first place. But it is also because memory is a complicated and highly fallible process.

- Instead, what we seem to do is construct a **theory** about how a rational person would form a belief in such a case, and then attribute that belief formation process to ourselves. This is the process Kornblith calls “rational reconstruction.”

- As Kornblith notes, since the guiding assumption in this process is that our beliefs were formed rationally, it is exceedingly difficult to use this process to improve upon our epistemic practices. Improvement depends on an external perspective that uses controlled experimental conditions to point out systematic errors in reasoning.
The irrelevance card

As often occurs in philosophical disputes, much of the foregoing is regarded as interesting science but ultimately irrelevant to the basic aims of epistemology.

Those who play the irrelevance card will typically argue something like this: The point of epistemology is to establish standards that must be met to achieve knowledge. If ordinary people engaging in ordinary introspection are not able to meet these standards, that’s unfortunate. But we should simply conclude that such people lack knowledge, not that the standards are incorrect.

The basic problem with this response is that it ignores the fact that there is an empirical assumption at the root of internalist standards, namely that introspective practice can, at least in principle, cause one to become justified in one’s beliefs. It can not be irrelevant if this practice can not reliably produce such a state.
Internalism requires that the properties that constitute epistemic justification be cognitively accessible to the agent. Hence, the question for coherence theorists like Bonjour is whether coherence is cognitively accessible. Is the coherence of our belief system something that can be established by reflection?

If what we require is establishing the coherence of an entire belief system, then the answer to this is obviously no. It is just too big a job.

Kornblith raises the ante here by noting that even establishing something like logical consistency (which is only one coherence criterion) for a very small number of beliefs can be incredibly difficult.

The coherence theory gains some plausibility from examples of obviously inconsistent sets of beliefs. For example:
- Bob is my friend
- My friends are good people.
- Good people don’t sell drugs to children.
- Bob sells drugs to children.

But it can also be extremely difficult. As Kornblith notes, Frege, one of the smartest philosophers in history was unable to see that a very small number of set-theoretic axioms implied a contradiction. (Russell pointed it out to him in a letter and it pretty much destroyed his life’s work.)
A digression into truth

- It’s important to realize that we are talking here about the coherence theory of justification, not the coherence theory of truth. But it’s worth noting that a certain misconception about coherence theory actually stems from it’s consideration as a theory of truth.

- The two traditionally acknowledge theories of truth are:
  1. the correspondence theory: a statement is true if accurately reflects the facts.
  2. the coherence theory: a statement is true if it coheres with a background of beliefs.

- Both of these theories are wrong, but it’s interesting to note that one alleged defect of the correspondence theory is that we have no direct cognitive access to the facts- think Kant’s noumenon- so correspondence is something we are never in a position to ascertain.

- This criticism is generally regarded as favoring the coherence theory, but as Kornblith points out the coherence of our beliefs is really no more cognitively accessible than correspondence.
Think of it this way. If the coherence of our belief system is an objective property of that system, then we inquire into whether it has the property just like we inquire into anything else. On the basis of our inquiry we may develop the belief that our belief system is coherent, but from a correspondence perspective this is just the assertion that:

- The statement “My belief system is coherent” accurately reflects the fact that my belief system is coherent.

In other words, to the extent that coherence is a real property, we are no more directly cognitively connected to it than we are to correspondence.
Bonjour on Reliabilism

- Kornblith makes this point by considering Bonjoures critique of externalism/reliabilism:
- Bonjour points out that it is a simple matter to define conditions under which a belief has been produced reliably, but in fact is not justified. From this fact he distills a general criticism of externalism:
  - “If the acceptance of a belief is seriously unreasonable from an observer’s own standpoint, then the mere fact that unbeknownst to him its existence in those circumstances lawfully guarantees its truth will not suffice to render the belief epistemically justified”
- Kornblith has a tu quoque response to this: Coherence is no different.
- We should note, that Bonjour’s point is a good one, but if it applies equally well to internalist as externalist views, then it constitutes an argument that knowledge is not about justification at all.
The combinatorial explosion

- The final nail in the coffin of coherence theory is that it is literally impossible to check the consistency for even a very small fraction of the beliefs that humans have.
- This point is easily appreciated when you consider what it actually means to check for consistency, namely to examine lines in a truth-table. For example, we can show that the following set of beliefs is inconsistent:
  - \( P; P \rightarrow Q ; \sim Q \)
- By showing that every possible interpretation, i.e., every line on its truth table is false.
- But as Cherniak points out, to determine the consistency of a belief system with 100 atomic proposition, which has a truth table with \( 2^{100} \) this would take a super computer more than 20 billion years.
Idealized knowledge

- Bonjour’s response to this point is interesting. He claims that it proves only that human agents are incapable of the epistemic ideal. He is content with this conclusion, because he is content to think of our definition of knowledge as an ideal standard, rather than something that is achieved in practice.

- This actually makes a great deal of sense, even in the context of naturalism. In science we do employ idealized standards some which are even in some sense incoherent (frictionless surfaces, instantaneous velocity). We do this because they provide a standard that we can approximate in the lab, or which facilitates accurate calculation.

- Kornblith does not think this gambit is open to Bonjour, however, simply because the human mind falls so incredibly short of this standard. To adopt it leads inevitably to skepticism.
Some remarks on skepticism

- It’s worth stopping here to note how most philosophers think about skepticism today.

- Skepticism, as a basic cognitive attitude toward unsupported claims is still the defining characteristic of a philosophical mind set. However, skepticism about the possibility of knowledge itself is no longer seen as a coherent position. This is not because philosophers have proved that we have knowledge. Rather, it is because the demand for proof is incoherent.

- Hume pointed this out quite a while ago in his critique of the Cartesian project: If we don’t assume that we know something, then by what principles would we ever establish that we do? In other words, any proof that we know something is going to be based on principles that are themselves subject to doubt. If one were to claim that certain principles are not subject to doubt, one would still have to demonstrate that on the basis of other principles, ad infinitum.

- As we say noted earlier in the semester, Kant and Carnap saw universal skepticism as inherently confused. It makes sense to ask whether we know specific things about the world, but the question whether we know anything at all is not properly formed. It results from the attempt to answer transcendental/external questions from an empirical/internal perspective.
Kornblith’s critique of Bonjour

- So, Kornblith’s point is that on Bonjour’s view, humans are never even approximately justified in their views, which means that humans rarely if ever have knowledge, which makes his view self-refuting.
Foundationalism a fortiori

- Foundationalist theories of justification require that our beliefs be derived from epistemically basic beliefs and principles. It is problematic because there do not seem to be any such things, but Kornblith’s point here is that foundationalists have the same problem as coherentists because they accept coherentist criteria of reflection. The foundationalist argument has always been that reflective awareness of coherence is necessary, but not sufficient for justification.

- So Kornblith concludes that the making reflective awareness a criterion of knowledge results in putting knowledge out of reach of normal human beings.