











**1. What is Knowledge?**

**1.1. Defining Knowledge**

What is knowledge? Can it be defined, and how could a definition be useful? To approach this issue, we will first distinguish between different types of knowledge.

**1.1.1 Types of knowledge**

To understand what ‘knowledge’ is, we first need to think about what kind of knowledge we are trying to understand. Importantly, we can distinguish the kind of knowledge involved in skills and abilities from knowledge about the world. The first kind, ‘ability knowledge’, is knowing how to do something. For example, I know how to ride a bike. The second kind of knowledge involves being ‘in cognitive contact with reality’, as Zagzebski puts it in her article ‘What is knowledge?’.

We can sub-divide this second kind of knowledge into two further kinds. The first is ‘acquaintance knowledge’. This is knowledge that involves direct contact with something in experience, e.g. a person, a place, or one’s own thoughts and feelings. For example, I know Oxford, I know my wife, and I know what I’m thinking. In acquaintance knowledge, the word ‘know’ is (normally) followed by a noun (or pronoun) – ‘Oxford’, ‘my wife’, ‘what I’m thinking’.

The second kind of ‘cognitive contact with reality’ is ‘propositional knowledge’. Propositional knowledge is knowledge about some part of reality, which I may or may not have experienced myself. It is knowledge that some claim – a proposition – is true or false. A proposition is a declarative statement, or more accurately, what is expressed by a declarative statement, e.g. ‘eagles are birds’. Propositions can go after the phrases ‘I believe that . . .’ and ‘I know that . . .’. I know that eagles are birds, that 2 + 2 = 4, and that Oxford is a city in England.

**1.1.2. Propositional knowledge**

While ability knowledge or acquaintance knowledge are interesting and important, we will focus on just propositional knowledge. One reason that philosophers have written more about propositional knowledge is that it is the form of knowledge that we can pass on to each other. To know a person, you need to ‘get to know’ them by meeting them, talking with them, and so on. You can’t know them just from other people’s reports. Indeed, it’s hard for us to say (to each other) just what we ‘know’ when we know someone. From now on, our question is ‘What is propositional knowledge?’. When I talk of ‘knowledge’, I mean ‘propositional knowledge’.

**1.1.3. Can propositional knowledge be defined?**

Can knowledge be defined in terms of necessary and sufficient conditions? If so, will we have a real definition that tells us the true nature of knowledge? Philosophers have aimed at such a definition. But we might doubt its success in advance. For example, different societies have meant different things by KNOWLEDGE. Or again, sometimes we are quite generous with the word, e.g. I know that Descartes wrote the Meditations, but other times, we make more stringent demands, e.g. do I really know that my experience isn’t illusory, that I’m not in The Matrix or The Truman Show (and if I don’t know that, do I really know that Descartes wrote the Meditations)? Or again, the kind of knowledge we gain by, say, seeing something in front of us may be quite different from the kind of knowledge we gain by abstract reasoning, e.g. a young child can have the first kind of knowledge but not the second. And so on. Could there really be an abstract characterization of what knowledge is that holds true in all these cases?

Zagzebski suggests that we should adopt the aim of providing a real definition of knowledge at least until we can show that we have failed to find one. The only way we will know that we cannot give necessary and sufficient conditions for knowledge is by trying and failing to succeed.

**1.2. The Tripartite Definition of Knowledge**

In this section, we discuss the claim, deriving from Plato’s dialogue *Theaetetus*, that knowledge is a belief that is both true and justified. The tripartite definition of knowledge claims that knowledge is justified, true belief. It claims that you know some proposition, p, if and only if

1. the proposition p is true;

2. you believe that p;

3. your belief that p is justified.

The tripartite definition aims to provide a complete analysis of the concept and nature of propositional knowledge. Its three conditions, taken together, are intended to be equivalent to knowledge, to be the same thing as knowledge. So, first, if you fulfil those conditions, then you know the proposition. If all the three conditions it lists are satisfied – if you have a justified true belief that p – then you know that p. You don’t need anything else for knowledge; the three conditions, together, are sufficient. Second, if you know some proposition, you fulfil exactly those three conditions. If you know that p, then you have a justified true belief that p. There is no other way to know that p, no other analysis of knowledge. So, it claims, each of the three conditions is necessary. If p is false, or you don’t believe that p, or your belief that p is not justified, then you don’t know that p. The conditions are necessary and sufficient conditions for knowledge that p.

The definition puts forward two conditionals: if all three conditions are satisfied, then you know that p; and if you know that p, then all three conditions are satisfied. This is what is meant by the phrase ‘if and only if’ – that the conditions that follow are both necessary and sufficient. We may thus conclude that knowledge and justified true belief are the same thing. Justified true belief is necessary for knowledge (you can’t have knowledge without it), but it is also sufficient for knowledge (you don’t need anything else).

# 1.2.1. Why ‘Justified True Belief?

Why accept the tripartite view and adopt these three conditions for knowledge? In her article, ‘What is knowledge?’, Linda Zagzebski describes knowledge as a form of cognitive contact with reality. Reality is described or comprised by what is true, not what is false; what is false is precisely what reality isn’t. As a result, we can only know what is true. This is a reason to adopt the first condition, that p is true.

The idea of ‘cognitive contact’ also motivates the second condition. Propositional knowledge is a relation between the person who has knowledge and the proposition that is known. The relation involves the person taking some proposition to be true. Taking a proposition to be true is to believe it. If I take ‘eagles are birds’ to be true, if I assent to it, then I believe it.

But what is it for a belief is ‘justified’ and why think knowledge must involve justified belief? One’s ‘justification’ for a belief is what one offers as a reason or evidence to accept it. To understand the importance of justification, we first need to understand that a belief can be true and yet not justified. For example, someone on a jury might think that the person on trial is guilty just from the way they dress. Their belief, that the person is guilty, might be true; but how someone dresses isn’t evidence for whether they are a criminal! True beliefs can be formed or held on irrational grounds, for no good reason. Or again true beliefs can just be lucky. For example, there is a lot of evidence that astrology does not make accurate predictions, and my horoscope has often been wrong. Suppose on one occasion, I read my horoscope and believe a prediction, although I know there is evidence against thinking it is right. And then this prediction turns out true!

Zagzebski notes that we think knowledge is good; it is desirable and perhaps it is praiseworthy in some sense. Knowledge is undoubtedly good for helping us satisfy our needs and desires (from knowing where the closest supermarket is to finding a cure for cancer); many people have thought that it is also good in itself, irrespective of whether we can use knowledge (e.g. knowing about the origin of the universe). Whatever the reason why knowledge is good, we seek out knowledge for ourselves and support others who do so. We understand that knowledge can be difficult to acquire, requiring motivation or special skills, and we value these.

The examples show that lucky or irrational true beliefs are not good in the way knowledge is. They certainly aren’t praiseworthy. The tripartite theory explains this in terms of justification. Justification is what someone takes as their reason or evidence (or other basis) for their belief. In both examples, it is counter-intuitive to say that the belief counts as knowledge, because the person has no reason, no evidence, no justification, for their belief. We ought not to form our beliefs in this way, even if they sometimes turn out true. When we form a belief, we should do so rationally, on the basis of reasons and evidence. If we do, then our belief will be justified. And this belief, if it is also true, says the tripartite theory, will amount to knowledge.

## 1.2.2. A Note on Certainty

Some philosophers have thought that another difference between knowledge and belief is certainty. Knowledge must be certain; beliefs don’t have to be. If a belief isn’t certain, then it can’t count as knowledge. We can only really know something if we can be certain of it.

But how ‘certain’ does certainty have to be? The difficulty with defining knowledge as believing a true proposition that it is impossible to doubt is that we end up with very little knowledge indeed. Since we tend to think that we do know all sorts of things that it is possible to doubt, this is clearly not how we usually think about knowledge.

The tripartite theory does not claim that a belief must be certain to be knowledge. To say a belief is justified is not to say that it is certain. We have good reasons to believe many things that it remains possible to doubt.

Of course, some of these beliefs may be false. We are fallible. It is possible to have a false justified belief – many scientific theories that we have now discarded are false, but the evidence available at the time was strong. For example, before Galileo invented the telescope, there was good reason to think that the planets circle the Earth and little evidence that they didn’t. But this is why, according to the tripartite view, we must say that knowledge is justified true belief, and not simply justified belief. If some belief that we take to be true turns out to be false, then it is not knowledge. If we discover that it is false, e.g. by uncovering new evidence, then we should give up the claim to know it.

# 1.2.3. Are the Conditions Individually Necessary?

We can raise two kinds of objection to the tripartite definition of knowledge by searching for counterexamples. First, it may be that one of the conditions is not necessary for knowledge – can we have knowledge without justified true belief? Second, it may be that all of the conditions together are still not sufficient for knowledge – can we have justified true belief without knowledge? We discuss only the first question in this section. For discussion of the second question, see section 3: ‘Gettier’s objection to the tripartite definition of knowledge’.

## 1.2.3.1 Justification is not a necessary condition of knowledge

Is justification necessary for knowledge, or could knowledge be simply ‘true belief’? We can object that sometimes we use the word ‘know’ just to mean ‘believe truly’, without worrying about justification. If I ask, ‘Do you know who wrote the *Meditations*?’, I’m only interested in whether you have the true belief that it was Descartes.

We can understand this in terms of the practical purpose of knowledge. If you can reliably inform me in answer to my query, perhaps that’s enough for practical purposes to talk of knowledge. But this won’t do as a definition for theoretical purposes. In particular, as we saw above, it fails to capture what is good about knowledge, since true belief can be formed and held in both good ways and bad. If you don’t have a good reason for believing that Descartes wrote the *Meditations*, then the mere fact that your belief is true doesn’t make it knowledge.

However, even if true belief is not sufficient for knowledge, that doesn’t mean that justification is a necessary condition. There may be some other condition that turns true belief into knowledge.

## 1.2.3.2. Truth is not a necessary condition of knowledge

We connected the idea that knowledge involves truth to Zagzebski’s claim that knowledge is cognitive contact with reality. What ‘reality’ is, is an issue in metaphysics. And the question of what we should mean by ‘truth’ can become quite a challenging one. Nevertheless, we can make some important points on whether a definition of knowledge should include mention of truth, whatever truth turns out to be.

Could knowledge be simply justified belief? In an everyday sense, it is difficult to see how. Justified beliefs can be true or false. People can believe propositions that aren’t true. For example, someone may claim that flamingos are grey, and think that they know this. They could even be justified, e.g. their science teacher told them, and they saw a grey picture of a flamingo in a textbook. But they are mistaken: flamingos are not grey, but pink. Of course, they believe that flamingos are grey, they may even be certain that flamingos are grey. But given the idea that knowledge involves cognitive contact with reality, a false belief is not knowledge. You can’t know something false, or so it seems.

### 1.2.3.2.1. Relativism about truth

What if many people, perhaps a whole society, share a particular false belief and have good reasons for doing so? For instance, almost everybody used to believe that the Earth is flat. It does, after all, look that way. Should we say that people used to know that the Earth is flat? Or should we say that they didn’t know it, they only believed it, because their belief was false?

One response to this line of thought is to adopt some form of relativism. We reject talking about ‘truth’ without qualification, and talk instead about what is ‘true for’ someone or some society. Knowledge could still be justified true belief, but because what is ‘true’ is relative to someone or some society, knowledge is also relative.

Let us assume that the belief that the Earth is flat was justified. Is there any sense in which we can say that this belief was ‘true for’ people in the past? To say it was ‘true for them’ must be to say more than simply that they believed it. We all agree they believed it, but to believe that some proposition is true is not the same as the proposition being true. If all it takes to make something true is to believe, then the best cure for cancer is simply to believe that one doesn’t have cancer! If there is no difference between a belief and a true belief, then how does anyone get less than 100% on any exam? To make sense of our lives, we must allow that beliefs can be true or false. So for a belief to be ‘true for’ someone, this can’t simply mean that they believe it. So what does it mean?

A second difficulty arises with the idea of ‘true for’: If we say that it was true, for people in the past, that Earth is flat, and it is true for us now that the Earth is a sphere, the question arises how both of these claims about the Earth can be true. Did the Earth miraculously change from being flat to being a sphere? Did a change in people’s beliefs change the shape of the Earth? No one believes that.

Perhaps we can defend relativism by giving up all talk of truth (and perhaps all talk of ‘reality’), and restrict ourselves to talking about what people believe. Knowledge is simply justified belief. We cannot ask ‘what shape is the Earth (truly)?’, we can only ask ‘what shape is the Earth (‘for us’)?’. There is no ‘objective truth’ about the shape of the Earth.

However, it seems hard to resist the claim that the modern view of the shape of the Earth is closer to the truth than the ancient theory. For instance, we have more evidence, e.g. photos from space, than they did. To make the claim that our beliefs are true while those of some other culture are false is not to say that their beliefs were unjustified or irrational or unintelligent. We are discussing truth, not justification, and the evidence available to people changes over time. Nor is it to say that our beliefs are certain or infallible.

The response that no one knows the shape of the Earth because we could be mistaken is irrelevant – it retreats from relativism to scepticism, a completely different view. To allow that we could be mistaken assumes that there is some non-relative truth! To claim that ‘there is a truth that we don’t know’ is very different from claiming that ‘there is no objective truth and what we know is relative to society’.

## 1.2.3.3. Belief is not a necessary condition of knowledge

There are two strengths of the objection that belief is not necessary for knowledge. The weak objection is that sometimes it is possible to know something without believing it. The strong objection is that knowledge is never a form of belief.

The weak objection: suppose John is sitting an exam, but he’s very nervous and has no confidence in his answers. Suppose when answering ‘Which philosopher wrote the Meditations?’, he writes ‘Descartes’. He’s right, and the answer isn’t a lucky guess – he has remembered what he learned. So it is plausible to say that John knows the answer, he knows more than he thinks – he’s just unconfident. But because he’s unconfident, we should say that John doesn’t believe that the answer is Descartes. So he knows the answer without believing it.

We can defend the tripartite definition by replying in one of two ways. We could say that John doesn’t know the answer. Although there is a sense in which he remembers the answer, because he doesn’t ‘commit’ to the answer that occurs to him, he doesn’t believe what he remembers. Alternatively, we could say that John does know that the answer is ‘Descartes’, because he believes this, although this belief is unconscious or ‘tacit’. This unconscious belief amounts to knowledge.

In the *Republic*, Plato presents arguments for the claim that knowledge is never belief. What is a matter of belief is not known, and what is matter of knowledge is not believed. Instead, belief and knowledge involve different ‘faculties’ and take different ‘objects’. He appeals to the connection between knowledge, truth and reality to make the case. First, knowledge is infallible, because you cannot know what is false. But beliefs, however, can be mistaken. Belief and knowledge have different powers. So belief cannot be knowledge. Second, knowledge is only of what is real. We cannot have knowledge of what is not real or does not exist. Knowledge is ‘about’ what is real. By contrast, ignorance relates to what is not real, what does not exist, i.e. ‘nothing’. If you are completely ignorant of something, you don’t think of it at all; if you don’t understand it, you can’t form an opinion about its reality. If there is something between what is real and what is not real (e.g. what is constantly changing from one thing to another), then there must be something between knowledge and ignorance. This is belief – neither knowledge nor ignorance. Belief and knowledge have different objects.

Plato argues that we may divide reality into the realm of the ‘sensible’ – what we detect through our senses – and the realm of the ‘intelligible’ – what we discover using the intellect. Belief relates to the former, knowledge to the latter. So we form beliefs about the changeable, natural world; but we gain knowledge – using reasoning in mathematics and philosophy – of abstract things, like numbers and concepts.

As Zagzebski notes, almost everyone now agrees that Plato is wrong to distinguish belief and knowledge as he does. First, knowledge and belief need not be different faculties even if knowledge is always true and belief is not. This difference isn’t a result of different ‘powers’, but because knowledge is always true and justified belief, whereas belief in general can be true or false, justified or unjustified.

Second, belief and knowledge do not need to be about different things. Because what is known is always true, you cannot know something that goes from true to false. Plato seems to try to explain this by saying that we can only know about things that cannot change. But a different explanation is to separate the truth that is known from the object by talking about truth at a time or in a context. For example, I can know that a particular object of sense experience – this book – has a particular property, e.g. it is a certain size. Yet its size can change, for example if you burn it. What I know is that the book is this size now (at a specific moment in time), and this truth won’t change even if the size of the book changes. Plato seems to have confused a property about knowledge (the truth of the proposition known doesn’t change) with a property about the object of knowledge (it doesn’t change).

1.3. Gettier’s objection to the tripartite definition of knowledge

In this section, we discuss a famous objection to the tripartite definition of knowledge and two responses to it. For more on the tripartite definition, see the section titled ‘The tripartite definition of knowledge’ starting on page \*\*\*.

In his article ‘Is justified true belief knowledge?’, Edmund Gettier famously presented cases in which we want to say that someone has justified, true belief but not knowledge. They show that the three conditions of the tripartite definition are not sufficient for knowledge.

Gettier starts by claiming, uncontroversially, that deductive argument preserves justification. Suppose you are justified in believing that p (say, that Socrates is a man and that all men are mortal), and p entails another proposition, q (that Socrates is mortal). If you deduce q from p, you are also justified in believing that q. Your reasons for believing the premises are also reasons for believing the conclusion, because if the premises are true, the conclusion must be true.

He then gives two counterexamples to the tripartite definition. In the first, Smith and Jones are applying for the same job. Smith has excellent reason to believe that Jones will get the job, e.g. Smith has been told this by the employer. Smith also has excellent reason to believe that Jones has ten coins in his pocket, e.g. Smith has just counted them. Therefore, both of these beliefs are justified. Smith then puts the two beliefs together and deduces that the man who will get the job has ten coins in his pocket. This belief is justified, because it is inferred deductively from justified beliefs. However, it turns out that Jones doesn’t get the job, Smith does. It also so happens that, unknown to him, Smith also has ten coins in his pocket. So Smith’s belief that the man who will get the job has ten coins in his pocket happens to be true.

Smith’s belief is both true and justified, but we shouldn’t say that Smith knows that the man who will get the job has ten coins in his pocket. Smith inferred his belief from a false belief, namely that Jones would get the job. So the reason Smith has for his belief is false. What makes his belief true (Smith, who has ten coins in his pocket, gets the job) has come apart from what justifies his belief (the evidence that Jones, who has ten coins in his pocket, will get the job). There is no connection between what justifies his belief and his belief’s being true. We might say that it is only by luck that his belief is true.

In the second counterexample, Smith believes that Jones owns a Ford. Smith remembers that for as long as he has known him, Jones has always had a Ford, and Jones has just offered Smith a lift while driving a Ford. So Smith’s belief is justified. Smith then thinks about another acquaintance of his, Brown, and wonders where he is. He has no idea, and chooses a place name at random, Barcelona, say. He then deduces the belief (quite oddly, but don’t worry about that) that either Jones owns a Ford or Brown is in Barcelona.

An either/or claim is called a ‘disjunction’. A disjunction is true if either (or both) of the two ‘disjuncts’ are true. So ‘Either Jones owns a Ford or Brown is in Barcelona’ is true if

1) Jones owns a Ford; or

2) Brown is in Barcelona; or

3) Both (1) and (2) are true.

Now, from ‘Jones owns a Ford’, Smith can deduce that ‘Either Jones owns a Ford or Brown is in Barcelona’, because it is impossible for the premise ‘Jones owns a Ford’ to be true and the conclusion ‘Either Jones owns a Ford or Brown is in Barcelona’ to be false. His belief that either Jones owns a Ford or Brown is in Barcelona is also justified, because he has deduced it from his justified belief that Jones owns a Ford. However, unknown to Smith, Jones has just sold his Ford and is currently driving a rented Ford, so he doesn’t own a Ford. But by complete coincidence, and unknown to Smith, Brown is in Barcelona. So the proposition ‘Either Jones owns a Ford or Brown is in Barcelona’ is true – not because (1) Jones owns a Ford, but because (2) Brown is in Barcelona.

So Smith’s belief is both true and justified, but we shouldn’t say that Smith knows that either Jones owns a Ford or Brown is in Barcelona. Once again, Smith inferred his belief from a false belief, namely that Jones owns a Ford. So the reason Smith has for his belief is false. What makes his belief true (that Brown is in Barcelona) has come apart from what justifies his belief (that Jones owns a Ford). There is, again, no connection between what justifies his belief and his belief’s being true.

Thought experiments are a philosophical method designed to test a hypothesis or philosophical claim through imagining a hypothetical situation, and coming to a judgment. We can use thought experiments, like the examples Gettier gives, to test definitions of knowledge. Examples of justified true belief without knowledge became known as ‘Gettier cases’. They all describe situations in which we have justified true belief, but not knowledge, because the belief is only true by chance, given the evidence that justifies it. So justified true belief is not sufficient for knowledge.

This shows that justified true belief is not the same as knowledge. If A is the same thing as B, then you cannot have A without B or B without A. A and B are just one thing – you haven’t it or you don’t. For example, you can’t have water without H2O. If you could, that would show that water is not the same thing as H2O. So if knowledge isn’t justified true belief in Gettier cases, then knowledge and justified true belief can come apart. That shows that they are not the same thing, and so the definition of knowledge as justified true belief is false.

# 1.4. Add A ‘No False Lemmas’ Condition

Gettier’s argument poses a serious challenge to the tripartite definition of knowledge. In the resulting debate over what knowledge is, most attention has focused on the claim about justification. One response is to *strengthen* what we mean by justification in the case of knowledge. Another, that we don’t discuss further in this section, is that we need to *replace* the justification condition with something else. But we start with a simpler idea.

Smith doesn’t know that the man who will get the job has ten coins in his pocket, we said, because he inferred this belief from a false belief, namely that Jones will get the job. Similarly, he inferred his belief that either Jones owns a Ford or Brown is in Barcelona from his belief that Jones owns a Ford. To deal with these two cases, all we need to do is to add an extra condition to the definition of knowledge. You know that *p* if and only if

1*. p* is true

2. you believe that *p*

3. your belief that *p* is justified

4. you did not infer that *p* from a false belief.

Condition (4) is called the ‘no false lemmas’ condition. A lemma is a claim part way through an argument. For example, Smith concluded that Jones will get the job from being told by the employer; and he then used that information to conclude that the man with ten coins in his pocket will get the job. So ‘Jones will get the job’ is a lemma.

Condition (4) certainly deals with Gettier’s two examples. But it doesn’t deal with the underlying worry about truth and justification ‘coming apart’. There are Gettier cases that satisfy condition (4) without the person having knowledge.

In her article ‘What is knowledge?’, Linda Zagzebski describes such a case based on induction. Dr Jones has very good evidence that her patient, Smith, is suffering from virus *X*, e.g. the symptoms and the lab tests are all consistent with Smith having this virus and no other known virus produces these results. Jones therefore believes that Smith has virus *X*, and this belief is justified. However, Smith’s symptoms and lab results are caused by Smith having the unknown virus *Y*. But, by chance, Smith has *just* caught virus *X*, so recently that it has not caused any symptoms nor does it show up in lab tests. So Dr Jones’ belief that Smith has virus *X* is true. So her belief is both true and justified. But she does not know that Smith has virus *X* because the evidence from which she infers her belief has nothing to do with the fact that Smith has virus *X* as it is all caused by virus *Y*.

This example satisfied condition (4). All the evidence is *true* – Smith does manifest the symptoms he does and the lab reports are accurate. Only the diagnosis is false. So Dr Jones did not infer her true and justified belief that Smith has virus *X* from any false belief, yet it is still not knowledge.

# 1.5. Infallibilism

Gettier has assumed that Smith’s beliefs – that Jones will get the job, that he has ten coins in his pocket, that Jones owns a Ford – are justified. Because they are justified, his deductions, that the man who will get the job has ten coins in his pocket and that either Jones owns a Ford or Brown is in Barcelona, are justified. But we can challenge Gettier’s assumption. While Smith has good reasons for the beliefs from which he starts, but perhaps these reasons aren’t enough for justification. If so, then his conclusions will not be justified enough either.

The tripartite definition of knowledge does not tell us what it is for a belief to be justified, although it does not usually require certainty. Infallibilism argues that knowledge is *certain*. One way of combining this with the tripartite view is to say that justification requires certainty. We can say that if a belief is not certain, then it is not justified, or at least, it is not sufficiently justified to count as knowledge. The implication that we should draw from Gettier cases is not that knowledge is not justified true belief, but that our beliefs are rarely sufficiently justified to count as knowledge.

Zagzebski notes that Gettier cases arise because of the gap between truth and justification. The challenge to the tripartite theory is how to bridge that gap. The suggestion of infallibilism is to make justification so strong that it is impossible for a justified belief to be false. Justification as certainty somehow guarantees the truth of the belief, so truth and justification can’t come apart, and Gettier cases become impossible.

What kind of certainty is relevant to knowledge? The *feeling* of certainty won’t help. That could vary from one person to another, e.g. you might *feel certain* that God exists or that your friends will never betray you. This kind of certainty is subjective and psychological, a feeling of conviction. But we can all make mistakes, and be certain of something even though what we are certain of is not true. So a feeling of certainty can’t guarantee truth. Instead, the kind of certainty involved must relate to the belief being *infallible* in some way. For whatever reason, it is impossible that we could be making a mistake.

With this in mind, here is an argument for infallibilism, the view that we only know a proposition if our belief is infallible:

P1. No one can know what is false.

C1. Therefore, if I know that *p*, then I can’t be mistaken about *p*.

C2. Therefore, for justification to secure knowledge, justification must guarantee truth.

C3. Therefore, if I am justified in believing that *p*, I *can’t possibly* be mistaken.

C4. Therefore, if it is possible that I am mistaken, then I can’t be justified in believing that *p*.

C5. Therefore, infallibilism is true.

Infallibilism defends the tripartite view of knowledge and rules out Gettier cases, because in these cases I do not have *justified* true belief.

But is this a good definition of knowledge? It is rare that the justification of our beliefs rules out the *possibility* of error. Infallibilism entails that we have very little knowledge (even if we still have many beliefs that are very probably true). It also identifies what is *good* or praiseworthy or desirable about knowledge as its infallibility. But is that right? Perhaps it would be better to find a definition of knowledge that allows us more of it and recognises something good about knowledge without seeking immunity to error.

## 1.5.1 Rejecting the argument for infallibilism

The argument for infallibilism rests on a logical error. (C1) ‘If I know that *p*, then I can’t be mistaken about *p*’, has more than one meaning, depending on how one understands ‘can’t’:

C1a. It can’t be the case that if I know that *p*, I *am* mistaken that *p*.

We should agree with this, because of (P1) ‘No one can know what is false’.

C1b. If I know that *p*, I *can’t possibly* be mistaken that *p*.

This is what infallibilism assumes in moving from (C1) through (C2) to (C3). It is a much stronger claim than (C1a), because it says that not only am I *not* mistaken, but I *can’t possibly be* mistaken that *p*. Obviously, there are many cases of perception or memory in which I *could* be mistaken that *p*, but in fact I am not, and my true belief rests on evidence, so there are good reasons why I am not mistaken.

The argument for infallibilism slips from (C1a), inferred from (P1), to (C1b), used to support (C3). But this is a mistake, confusing one claim for another. The two claims are distinct, since one is a claim about whether I *am* mistaken, and the other is a claim about whether I *could be* mistaken. So the argument fails. To accept infallibilism, we need some other, independent reason to believe (C1b).

**1.6. Reliabilism**

Reliabilism rejects the claim that we need justification for knowledge. It claims instead that you know that p if and only if

1. p is true;

2. you believe that p;

3. your belief is produced by a reliable cognitive process.

A reliable cognitive process is just one that produces a high percentage of true beliefs. Examples include perception, memory and testimony. True beliefs caused by such processes count as knowledge. (Of course, if these processes cause a false belief – if you misperceive or misremember or someone lies to you – then your belief isn’t knowledge, but that’s because it is false.)

One advantage of reliabilism is that it allows young children and animals to have knowledge. It is odd to say, of many animals, that they have reasons or evidence for their beliefs – they don’t have that kind of sophisticated psychology. But they get around the world very well indeed, so it is also odd to deny that they have knowledge. Reliabilism explains both points. Children and animals have knowledge because their true beliefs are caused by reliable processes; whether they have a justification for their beliefs is irrelevant.

In her article ‘What is knowledge?’, Linda Zagzebski notes that we think knowledge is good; it is desirable and perhaps it is praiseworthy in some sense. Knowledge is undoubtedly good for helping us satisfy our needs and desires (from knowing where the closest supermarket is to finding a cure for cancer); many people have thought that it is also good in itself, irrespective of whether we can use knowledge (e.g. knowing about the origin of the universe). Whatever the reason why knowledge is good, we seek out knowledge for ourselves and support others who do so. We understand that knowledge can be difficult to acquire, requiring motivation or special skills, and we value these.

Lucky or irrational true beliefs are not good in the way knowledge is. They certainly aren’t praiseworthy. So knowledge must be more than simply true belief. The tripartite theory explains the goodness of knowledge in terms of justification. Reliabilism, Zagzebski notes, understands knowledge as a ‘natural’ good, like strength or beauty. It isn’t something that is praiseworthy, e.g. we don’t deserve credit for having good eyesight, but it is something desirable. Perhaps, in particular, having reliably true beliefs confers a significant benefit on a creature, e.g. it can act in a way that satisfies its desires. This contrasts with the justified true belief theory: Believing something only on good evidence or with good reason is something that we do, something that we can be praised for doing, while believing irrational or without evidence is something that we can be criticised for doing.

# 1.6.1. A Gettier Case for Reliabilism

However, reliabilism doesn’t solve Gettier’s challenge. Here’s another Gettier case, from Alvin Goldman’s article ‘Discrimination and Perceptual Knowledge’. Henry is driving through the countryside. He doesn’t know it, but in this part of the country – call it ‘Barn County’ – there are lots of fake barns, mere barn facades. But they have been built so that they look just like real barns when seen from the road. As he drives along, Henry often thinks ‘There’s a barn’, or ‘Hey, there’s another barn’. These beliefs don’t count as knowledge because they are false. But just once, Henry thinks ‘There’s a barn’ when he is looking at the one and only real barn in the area. This belief is true. And it is produced by a very reliable process, namely vision. But it is not knowledge, because – as in other Gettier cases – it is only a matter of luck that Henry’s belief is true in this one instance. But reliabilism has to say Henry does know there’s a barn. His belief is true and produced by a reliable process. The problem is that in Barn County, this reliable process has produced a true belief in circumstances in which the belief still seems only accidentally true.

Zagzebski provides another example. Dr Jones has very good evidence that her patient, Smith, is suffering from virus *X*, e.g. the symptoms and the lab tests are all consistent with Smith having this virus and no other known virus produces these results. Diagnosing whether someone suffers from virus X by looking at their symptoms and the results of their lab tests is a reliable process (the diagnosis is correct in a high percentage of cases). Jones therefore believes that Smith has virus *X*, and this belief is produced by a reliable process. However, Smith’s symptoms and lab results are caused by Smith having the unknown virus *Y*. But, by chance, Smith has *just* caught virus *X*, so recently that it has not caused any symptoms nor does it show up in lab tests. So Dr Jones’ belief that Smith has virus *X* is true. So her belief is both true and produced by a reliable process. But she does not know that Smith has virus *X* because the reliable process which causes her belief has nothing to do with the fact that Smith has virus *X* as it is all caused by virus *Y*. The reliable process has produced a true belief that is nevertheless not knowledge.

# 1.6.2. Truth and the Third Condition

Both the tripartite view and reliabilism share a common approach to the analysis of knowledge. They claim that knowledge is true belief + some third condition that is closely connected to truth but independent from it – justification or being the product of reliable cognitive processes. As Gettier cases show, the independence of this third condition allows the possibility that it comes apart from truth – that the truth of the belief is lucky, not the result of the third condition.

Zagzebski argues that as long as the third condition is independent of truth like this, no matter how we add to the conditions for knowledge, we will always be able to construct Gettier cases that show that the proposed definition of knowledge is incorrect.

Why does she think this? Because there is a ‘recipe’ for making up such cases, no matter what the additional conditions are – as long as the additional conditions are independent of truth. Let’s call the additional conditions Q (justification, justification + no false lemmas, reliable cognitive processes, etc.), so the proposed definition is that knowledge = true belief that is Q.

1. Start with a belief that is Q but false as a result of ‘bad luck’, e.g. Henry believes, of a façade in Barn County, that it is a barn.
2. Now add some ‘good luck’ so that the belief is true after all, e.g. Henry happens to be looking at the one real barn in Barn County.
3. This true belief will be Q, since it is exactly like the false belief that is Q, but happens, by luck, to be true. But it will not be knowledge, because it is true by luck.
4. So for any theory, where Q is independent of truth, knowledge is not true belief + Q.

What if we say knowledge is true belief that is not accidentally true? First, as a definition, this is terrible – it is vague, negative and no help. What is it for a belief to be ‘non-accidentally true’? Second, the third condition – that the belief is non-accidentally true – is not independent of truth, so it is no objection to Zagzebski’s argument.

The force of her argument is that we need a definition of knowledge that demonstrates both how and why truth and the third condition are connected, and not merely added together. As with justification and reliability, this connection will have something to do with why knowledge is good in a way that mere true belief is not.

**1.7. Virtue Epistemology**

Zagzebski’s definition of propositional knowledge is ‘belief arising out of acts of intellectual virtue’. In other words, you know that p if and only if

1. you believe that p
2. your belief that p arises from an act (or acts) of intellectual virtue.

We associate virtues more with ethics, but here we are concerned with intellectual virtues. However, Zagzebski notes that we think knowledge is good; it is desirable and perhaps it is praiseworthy in some sense. Knowledge is undoubtedly good for helping us satisfy our needs and desires (from knowing where the closest supermarket is to finding a cure for cancer); many people have thought that it is also good in itself, irrespective of whether we can use knowledge (e.g. knowing about the origin of the universe). Whatever the reason why knowledge is good, we seek out knowledge for ourselves and support others who do so. We understand that knowledge can be difficult to acquire, requiring motivation or special skills, and we value these.

Zagzebski develops her account of an intellectual virtue from Aristotle’s theory of virtues. A virtue is a state of a person that is good by way of helping the person achieve some good purpose or goal. Moral virtues, such as generosity or kindness, aim at moral goods, such as well-being. Intellectual virtues aim at intellectual goods, especially truth. Zagzebski defines a virtue as having two components.

1. A virtue motivates us to pursue what is good; in the case of intellectual virtues, we are motivated to discover the truth – so we care about believing what is true, not false.
2. A virtue involves a component that enables us to be successful – it gives us the ability to be reliable in forming true beliefs.

For example, being open-minded is an intellectual virtue that disposes us to care about carefully considering views that conflict with our own and it enables us to do this successfully.

Now, someone who has an intellectual virtue will reliably believe what is true, but not always. Everyone makes mistakes. So while a belief that is the product of a person’s exercising their virtues is epistemically good, it is still not completely good if it is false. On the other hand, a belief that is true is good, but it is not completely good if it is only accidentally true. For example, someone on a jury might think that the person on trial is guilty just from the way they dress. Their belief, that the person is guilty, might be true; but how someone dresses isn’t evidence for whether they are a criminal! True beliefs can be formed or held on irrational grounds, for no good reason. Or again true beliefs can just be lucky. For example, there is a lot of evidence that astrology does not make accurate predictions, and my horoscope has often been wrong. Suppose on one occasion, I read my horoscope and believe a prediction, although I know there is evidence against thinking it is right. And then this prediction turns out true!

We can find parallels in ethics. Helping someone is good – but if one does so accidentally, then the act is not completely good. Aiming to help someone is good – but if one fails, then again the act is not completely good. And aiming to help, messing it up, but through sheer good luck, ending up helping after all is still not completely good!

The morally best action will be one that aims to help, succeeds, and succeeds as a result of acting in the way a virtuous person would act. Call this an ‘act of virtue’. People with the relevant virtue, e.g. kindness, will be disposed to help and will reliably succeed, so they will reliably perform acts of kindness. However, Zagzebski notes that someone could perform such an action without having the virtue of kindness (the disposition to help reliably on many occasions) as long as they act in the way a virtuous person would act on this occasion.

Similarly, someone may form a belief on the basis of an act of intellectual virtue, e.g. being by open-minded on this occasion, without being open-minded generally. As long as, on this occasion, the person is motivated to find the truth (as a virtuous person would be) and does what a virtuous person would typically do (e.g. carefully considering a view that conflicts with their own), and the person succeeds in forming a true belief as a result, then the person performs an act of intellectual virtue.

We can now say that knowledge is belief arising out of acts of intellectual virtue. We don’t need to mention that it is true belief, because we have defined ‘an act of intellectual virtue’ as entailing that the belief formed is true. Virtues dispose us to succeed reliably, acts of virtue are ones in which we do succeed, and we succeed because we do what a person with the relevant virtues would do. That is why Zagzebski’s analysis of knowledge has just two conditions: that p is true is entailed by the second condition, that the belief that p arises out of acts of intellectual virtue.

# 1.7.1 Evaluating the Account

Our first question must be, Does the account avoid Gettier counter-examples? It should: it avoids the gap between truth and the ‘third condition’ that allows Gettier cases to be constructed. On Zagzebski’s definition, not only must the belief be true and be produced by acts of intellectual virtue, but its truth must be the result of such acts.

Zagzebski provides a Gettier case to test definitions of knowledge. Dr Jones has very good evidence that her patient, Smith, is suffering from virus *X*, e.g. the symptoms and the lab tests are all consistent with Smith having this virus and no other known virus produces these results. This seems a good way to discover whether someone is suffering from virus X. Jones therefore believes that Smith has virus *X*, and it would seem that Jones has exercised her intellectual virtues in coming to this belief. However, Smith’s symptoms and lab results are caused by Smith having the unknown virus *Y*. But, by chance, Smith has *just* caught virus *X*, so recently that it has not caused any symptoms nor does it show up in lab tests. So Dr Jones’ belief that Smith has virus *X* is true. So her belief is both true and she has exercised her intellectual virtues in coming to the belief. But she does not know that Smith has virus *X* because the way in which she acquired the belief has nothing to do with the fact that Smith has virus *X* as it is all caused by virus *Y*.

Zagzebski responds that while Dr Jones reached her belief that her patient has virus X through various intellectually virtuous activities, these didn’t lead her to the truth about her patient (that his symptoms etc. are caused by virus Y and he has only just acquired virus X.) So while Dr Jones’ belief is true and formed through intellectually virtuous activities, it isn’t true because Dr Jones’ performed acts of intellectual virtue. This demonstrates the importance of saying that knowledge is belief acquired *because* of acts of intellectual virtue.

However, here’s another Gettier case, from Alvin Goldman’s article ‘Discrimination and Perceptual Knowledge’. Henry is driving through the countryside. He doesn’t know it, but in this part of the country – call it ‘Barn County’ – there are lots of fake barns, mere barn facades. But they have been built so that they look just like real barns when seen from the road. As he drives along, Henry often thinks ‘There’s a barn’, or ‘Hey, there’s another barn’. These beliefs don’t count as knowledge because they are false. But just once, Henry thinks ‘There’s a barn’ when he is looking at the one and only real barn in the area. This belief is true. We should say that it is not knowledge, because it is only a matter of luck that Henry’s belief is true in this one instance.

This challenges Zagzebski’s analysis as follows. Normally, of course, when Henry sees and recognises a barn, he believes it is a barn because he sees and recognises it by paying attention to his environment. So normally, his belief arises from acts of intellectual virtue. In Barn County, he performs the same acts to acquire the true belief ‘there’s a barn’. Should we say that because he is in Barn County, his true belief is true not because of Henry’s intellectually virtuous activities but because of luck (and so it isn’t knowledge)? Or should we say that his true belief is the result of his acts of intellectual virtue (and so it is knowledge) since he does reach the truth that there is a barn, even if he doesn’t reach the truth that this is the only barn in the area? We need to know more about what it is for a belief to be true because it arises from acts of intellectual virtue before we can reach a verdict in cases like this. As it stands, the analysis is too vague.

A second issue is this. We noted above that knowledge is good – certainly desirable and perhaps also praiseworthy. Zagzebski’s theory explains the goodness of knowledge in terms of intellectual virtues. Does this restrict knowledge to adult human beings? Or can children and animals perform ‘acts of intellectual virtue’? If we define ‘virtue’ broadly enough, Zagzebski says, acts of intellectual virtue can include not only intellectually demanding actions, such as engaging in complex reasoning, but also relatively automatic, unconscious ones, such as looking or remembering. In these latter cases, the motivation to ‘find the truth’ doesn’t need to be obvious – one simply wants to know what is in one’s environment or what happened yesterday. And the relevant virtues, e.g. being attentive, need not demand much – an attentive person need only pay as much attention as needed to reach the truth. So young children, at least, can have knowledge just as soon as they can tell the difference between truth and falsehood and are motivated to find the truth.

A third issue concerns the place of virtue in knowledge. There are two objections here:

1. Suppose someone generally believes whatever they read on the internet. They don’t exercise caution or check other sources of evidence. So they don’t have some important intellectual virtues. Now suppose that on one occasion, they come across some strange and interesting claim and really want to know if it is true, so they do spend enough time finding other evidence, and so reach the truth. They have performed an act of intellectual virtue. But this is completely out of character. Do we really want to say that this person has knowledge, given that for lots of relevantly similar claims, they simply can’t be bothered to find out what is true? Isn’t their true belief still only ‘accidentally’ true even though it arose from acts of intellectual virtue? Should we grant knowledge only to people who have intellectual virtues, at least to some significant degree?
2. Why does someone’s motive make a difference to whether they have knowledge or not? Take animals: they simply acquire true beliefs in a very reliable way. Do their motives really matter? We can extend the case to people as well. As long as one discovers the truth reliably, and on this occasion, one belief is true because of one’s ability to discover the truth reliably, what should motives matter? So do we need virtues, rather than reliable processes, in the analysis of knowledge?

2. Perception as a Source of Knowledge

**2.1 Direct Realism**

How do we gain knowledge? If, as Linda Zagzebski suggests, knowledge involves being in ‘cognitive contact’ with reality, what means of being in contact with reality do we have? I don’t mean how we can know what is *inside* our minds, but how we gain knowledge of what is outside our minds.

The most obvious and immediate answer to the question ‘how do we gain knowledge of what is outside our minds?’ is ‘sense experience’ - awareness of physical objects through our senses. Sense experiences are those experiences given to us by our senses – sight, hearing, smell, taste, and touch, as well as bodily sensations. We can use our senses to perceive the world outside our minds. But *how* does perception by sense experience tell us about the world, and *what* do we learn about the nature of the world using perception? To answer these questions, we will need to think carefully about what sense perception involves.

Philosophers of perception divide into realists and idealists. Realists claim that what we perceive are physical objects, which exist independent of our minds and of our perceptions. Idealists argue that physical objects are not, in fact, independent of our minds. What they are, and so what we perceive, are mental things – ideas of some kind. In this section, we will look at just one theory of perception: direct realism.

It is common sense to say that we perceive physical objects, and these exist independently of our minds. ‘Physical objects’ include tables, books, our own bodies, plants, mountains. Cosmology and the theory of evolution suggest that physical objects, such as stars and planets, existed for billions of years before minds existed to experience them. And it is part of our idea of physical objects that they continue to exist when we don’t perceive them. When I leave my study, all the physical objects – the desk, the chairs, the books, and so on – remain just as they are. Physical objects exist objectively in space and time.

Direct realism is the natural starting point for thinking about perception. According to direct realism, what we perceive through our senses are just these very things, physical objects, together with their various properties. When I perceive my desk, for example, I perceive its size, shape, colour, smell and texture (I’ve never experienced its taste, but I could, I suppose!). So, direct realism claims that what we perceive are mind-independent physical objects and their properties. Another way of putting this is to say that the ‘immediate object’ of perception is the physical object itself. There isn’t something else, e.g. a mental image, that we perceive in perceiving physical objects.

Direct realists explain that we can gain knowledge through perception because perception is a form of ‘openness’ to the world. What perception gives us is a direct awareness of mind-independent objects. Importantly, our awareness of these objects is sensitive to how the objects are – differences in the properties of the objects we perceive will be detected by differences in our perceptual experience of them.

2.1.1. Objections

There are four objections to direct realism that are covered in the specification and these are outlined in this section. You should be aware that each of these serves both as a criticism of direct realism, and an argument in favour of indirect realism which is covered in section 2.2. starting on page \*\*\*.

# 2.1.1.1. Objection: The Argument from Perceptual Variation

A little reflection suggests that what we perceive isn’t quite the same as what is ‘out there’. In *The Problems of Philosophy*, Bertrand Russell gives an example of looking at a shiny, brown table. We say it is brown, but it doesn’t actually look an even brown colour all over: depending how the light falls, some parts are lighter than others, and some are even white from the shininess. So Russell objects that saying the table is brown means no more than that it looks brown ‘to a normal spectator from an ordinary point of view under usual conditions of light’ – but why think that this colour is more real, more a property of the table, than any of the other colours that you experience? Just what colour any part of the table looks to you depends on where you stand. If you and someone else look at the table together, you will see different patterns of colour. Suppose a shiny spot on the table looks light brown to you but white to the other person. The table can’t *be* both brown and white in the same spot at one time.

Russell then runs the same argument, appealing to variations in our perceptual experience, for the properties of texture and shape. The table might be smooth to touch, but at a microscopic level, there are all kinds of bumps and dips – so should we say that when we touch the table, the smoothness we feel is a property of the table? And the shape that something appears to have, like its colour, varies with the angle from which you view it. A rectangular table, from every angle except 90 degrees, does not look perfectly rectangular.

These examples draw our attention to a distinction between appearance and reality. Obviously, much of the time, we talk as though things are just as they seem. But, clearly, we also distinguish between appearance and reality – and Russell remarks that having any skill as a painter requires that one does.

All this perceptual variation causes a real problem for the direct realist. The direct realist says I perceive physical objects and their properties, in this case the table, ‘directly’, as they are. The argument from perceptual variation runs like this:

P1. There are variations in perception.

P2. Our perception varies without corresponding changes in the physical object we perceive. (For instance, the table remains rectangular, even as the way it looks to me changes as I look at it from different angles.)

C1. Therefore, the properties physical objects have and the properties they appear to have are not identical.

C2. Therefore, what we are immediately aware of in perception is not exactly the same as what exists independently of our minds.

C3. Therefore, we do not perceive physical objects directly.

## 2.1.1.1.1. Indirect Realist Solution: Sense-data

We now need a name for talking about what we are immediately aware of in perception, e.g. the colour and shape of the table as I see it now. Russell calls these ‘sense-data’ (singular: ‘sense-datum’). When I look at the table, I have a (visual) sensation – I am immediately aware of something. The ‘content’ of my sensation – what I am immediately aware of – is sense-data (on Russell’s view). We can also think of sense-data as appearances (how things appear to us to be).

Sense-data are distinct from the table. The table exists independently of my perception of it, while sense-data are defined as what it is that I perceive – so they depend on my perception. If I close my eyes, the colour and shape of the table as seen by me, cease to exist. And the colour and shape of the table as seen by me varies from where I look at it, while we don’t want to say that the table itself varies in this way. We can summarise the argument so far by saying that perceptual variation shows that what we directly perceive are not physical objects, but sense-data.

See section 2.2. on page \*\*\* for more detail.

## 2.1.1.1.2. Direct Realist Response: Relational Properties

We can challenge Russell’s claim that there is no good reason to say that one of the colours we experience the table as having is more real than the others. As he notes, what we *mean* by the colour of an object is the colour that it appears to have when seen by normal observers under normal conditions. That we don’t *always* see this colour – that our perception of its colour varies – doesn’t show that direct realism is false: we can still say that we see the table, and its colour, under normal conditions. After all, we do all see it as some shade of *brown* (shading to white), rather than some of us seeing it as brown, others as red, others as blue. So, in seeing its colour (as some variant of brown), we see the table and its properties.

With shape, we have an even better reason to privilege the claim that the table is rectangular, rather than obtuse – we can use its shape to perform various actions, like getting it through a narrow doorway, which will only succeed if it *is* rectangular and not obtuse.

But the argument from perceptual variation does show that direct realism needs a more sophisticated account of what it is to see the table and its properties. To develop this, we need to introduce the idea of a ‘relational property’. A relational property is a property that something has only in relation to something else (and, in some cases, only in some circumstances). For example, ‘being to the north of’ is a relational property; Manchester is to the north of London. Another example is ‘being in love with’; Jack is in love with Joan. Notice that in these examples, it is Manchester and Jack that have the properties; but we can only say what properties these are by mentioning other things – London and Joan.

In perception, we can be aware of a range of properties, some of which the object has independent of our minds, and some of which it has in relation to being perceived. For instance, a rectangular table has the property of ‘looking obtuse’. The property of ‘looking obtuse’ is a distinct property from ‘being obtuse’ – so a table can *be* rectangular and *look* obtuse. The property of ‘looking obtuse’ is a relationalproperty, in this case, a property the table has in relation to being *seen*. ‘Looking obtuse’ is a property *the table* has, claims direct realism, not the property of a sense-datum. And we can even explain why the table has the property of looking obtuse (to us) in terms of its being rectangular plus facts about light and vision.

Direct realism can claim that in perceiving physical objects, some of the properties we perceive are relational properties while others are not. It doesn’t have to claim that *all* the properties of physical objects, as we perceive them, are mind-independent as long as there is a clear sense in which we are directly aware of physical objects themselves. This response challenges the inference from (C2) to (C3) above.

# 2.1.1.2. The Argument from Illusion

We have seen that the appearance/reality distinction challenges direct realism. We can appeal to illusions to press the case. If you look at an oar half-submerged in water, it looks crooked; but it isn’t. We see a crooked oar, but the oar isn’t crooked. However, *just* from what you experience, you can’t tell whether you are seeing an illusion or not. Someone who doesn’t know about the illusion thinks they are seeing a crooked oar. It *looks* just like a crooked oar. The point applies generally to illusions. From *just* what we see in an illusion, without other background knowledge, we cannot tell whether what we are seeing is an illusion or not. Illusions can be ‘subjectively indistinguishable’ from veridical perception. This provides an argument against direct realism.

P1. We perceive something having some property *F* (e.g. an oar that is crooked).

P2. When we perceive something as having some property *F*, then there is something that is *F*. (*Something* we see is F.)

P3. In an illusion, the physical object does not have the property *F* (the oar is not crooked).

C1. Therefore, in illusions, what has the property *F* is something mental, a sense-datum.

C2. Therefore, in illusions, we see sense-data, and not physical objects, immediately.

P4. Illusions can be ‘subjectively indistinguishable’ from veridical perception.

C3. Therefore, we see the same thing, namely sense-data, in both illusions and veridical perception.

C4. Therefore, in all cases, we see sense-data, and not physical objects, immediately.

C5. Therefore, direct realism is false.

## 2.1.1.2.1. Direct Realist Response: Relational Properties Again

Direct realism can give the same reply as before. When the oar in water looks crooked, there is nothing that *is* crooked; (P2) is wrong. Instead, the stick has the property of *looking crooked* when half-submerged in water. There is a difference between the property ‘being crooked’ and the (relational) property ‘looking crooked’. Usually, of course, something looks crooked when it is crooked. But the two properties can come apart, and something can look crooked when it is straight. So, in illusions, we perceive the ‘looks’ properties of physical objects, and these ‘looks’ properties don’t match the ‘is’ properties of the object. But we still directly perceive physical objects and their properties.

# 2.1.1.3. The Argument from Hallucination

We have seen that direct realism can explain the difference between how things are and how they appear to us, in cases of perceptual variation and illusion, by appealing to the ‘looks’ properties of physical objects. But how can direct realism respond to the challenge of hallucinations? We can experience perceptual hallucinations – not just visual ones, but auditory and olfactory hallucinations as well. As with illusions, hallucinations can be subjectively indistinguishable from veridical perception. But here we can’t say that what is seen is how some physical object looks, because no physical object is seen at all! So direct realism’s reply to the previous arguments won’t work here.

P1. In a hallucination, we perceive something having some property *x*.

P2. When we perceive something as having some property *F*, then there is something that is *F*.

P3. In a hallucination, we don’t perceive a physical object at all.

C1. Therefore, what we perceive must be mental – sense-data.

P4. Hallucinations can be experiences that are ‘subjectively indistinguishable’ from veridical perceptions.

C2. Therefore, we see the same thing, namely sense-data, in both hallucinations and veridical perception.

C3. Therefore, in all cases, we see sense-data, and not physical objects, immediately.

C4. Therefore, direct realism is false.

## 2.1.1.3.1. Direct Realist Response: The Disjunctive Theory of Perception

There is another way that direct realism can challenge (P2). If something looks a certain way, then one of *two quite different things* is going on: *either* I directly perceive a mind-independent physical object that is *F or* (as in the case of hallucination) it appears to me just *as if* there is something that is *F*, but there is nothing that *is F*.

According to the disjunctive theory of perception, hallucinations and veridical perception are two completely different kinds of mental state, because in hallucination, the person isn’t connected up to the world. They can *seem* exactly the same, but that doesn’t prove that they *are* the same. The fact that hallucinations are subjectively indistinguishable from veridical perception tells us nothing significant about what *perception* is. In hallucination, we don’t *perceive* anything, we *imagine* it. To imagine something is not to perceive something mental, such as sense-data, but not to perceive anything at all. Perception is a relation of the subject to the world, a form of ‘cognitive contact’. Hallucination is not.

We can use this to challenge (C2). And so the argument from hallucination doesn’t show that in veridical perception, we perceive sense-data instead of physical objects.

# 2.1.1.4. The Time-Lag Argument

It takes time for light waves, or sound waves, or smells, to get from physical objects to our sense organs. For example, it takes 8 minutes for light from the sun to reach the earth. If you look at the sun (not a good idea!), you are actually seeing it as it was 8 minutes ago. If it blew up, you would see it normally for 8 minutes after it had blown up – it wouldn’t even exist anymore, and you’d still see it! Therefore, we could argue, you aren’t seeing it directly.

However, it would be a mistake to think that this shows that what you perceive is a sense-datum of the sun. The ‘image’ you see is not mental but *physical*, carried in light waves. The light waves exist during those 8 minutes. So *if* you see the sun indirectly, then it is because you see light waves directly. But then what we perceive immediately is not the sun, but the light from the sun. We can generalise: what we perceive is the physical medium by which we detect physical objects (light waves, sound waves, chemicals for smell and taste). So, we don’t perceive (ordinary) physical objects directly.

Direct realism can reply that this is a confusion between *how* we perceive and *what* we perceive. Compare these two pairs of questions:

1. ‘Can you see the lake?’ and ‘Can you see the light reflecting off the lake?’

2. ‘Can you see the paper?’ and ‘Can you see the light reflecting from the paper?’

In (1), we can turn our attention from the lake to the light reflecting off it. So we can talk, literally, about seeing the light. But in (2) there is no difference in *what* one is supposed to see. To ‘see’ the light that the paper reflects is just to see the paper. In fact, you cannot *see* the light itself – only the paper. So, direct realism can argue, except in special conditions, we don’t perceive light waves directly and physical objects indirectly. Light waves are part of the story of how we see physical objects.

The time lag involved in how we perceive means we see the physical object as it was a moment before, not as it is now. This means that we literally see (into) the past. We always experience the world as it was a moment ago, or in astronomy, when we look at distant stars and galaxies, we look into the distant past.

# 2.1.2. Strength of Direct Realism: Openness

We said earlier that direct realism understands perception as ‘openness’ to the world, a direct awareness of mind-independent objects. We can get a fuller sense of this by trying to describe what we see. We would usually do so by referring to physical objects: ‘I see a desk, covered with pens and paper, and a plant’. If we perceive the world via sense-data, the immediate ‘content’ of what we perceive is mental. So try to describe your experience in terms of sense-data, without referring to any physical objects. You could talk about ‘coloured patches’ standing in spatial relations (above, below, left, right, etc.) to each other. But this is very awkward, and it is virtually impossible for any normal scene. What shape is that green patch on the left? – well, ‘plant-shaped’! But ‘plant’ refers to a physical object. So our way of describing sense-data is dependent on concepts of physical objects. We can’t give an account of what we experience without referring to physical objects, even if we try.

What this shows is that our perceptual experience presents what we perceive *as* mind-independent objects. That doesn’t *prove* that we perceive mind-independent objects, but it does make such a claim highly intuitive. Only direct realism holds onto this basic intuition of ‘openness’. It is very counter-intuitive to think, then, that what we perceive are sense-data. Any theory that claims that we perceive sense-data has to say that perception is not what it seems to be. It has to say that it *seems* that we immediately perceive mind-independent objects, but we don’t. We need very strong reasons to accept that perception is misleading in this way.

**2.2. Indirect Realism**

Indirect realism claims that we perceive physical objects which are mind-independent, but we do so via, or in virtue of, perceiving mind-dependent sense-data that are caused by and represent physical objects. We perceive sense-data immediately, and physical objects indirectly, mediated by our perception of sense-data.

Sense-data (singular: ‘sense-datum’), according to indirect realism, are what we are immediately aware of in perception, e.g. the colour and shape of the table as I see it now. When I look at the table, I have a (visual) sensation – I am immediately aware of something. The ‘content’ of my sensation – what I am immediately aware of – is sense-data. We can also think of sense-data as appearances (how things appear to us to be).

Sense-data are distinct from the table. The table exists independently of my perception of it, while sense-data are defined as what it is that I perceive – so they depend on my perception. If I close my eyes, the colour and shape of the table as seen by me, cease to exist. And the colour and shape of the table as seen by me varies from where I look at it, while we don’t want to say that the table itself varies in this way. Sense-data (if they exist at all) exist as part of the mind.

So, assuming realism about physical objects, we can draw the following contrasts:

1. Sense-data are mental things which are the way we perceive them to be. They are appearances, and so are exactly as they seem. There is no further reality to an appearance than how it appears. By contrast, physical objects can appear differently from how they really are.
2. Sense-data only exist while they are being experienced. An experience must be experienced by someone to exist at all. Physical objects can exist when no one experiences them.
3. Sense-data are ‘private’. No one else can experience your sense-data. They are the particular sense-data they are, by definition, as part of your consciousness. Physical objects are ‘public’. One and the same object can be experienced by different people.

# 2.2.1. Indirect Realist (Bertrand Russell) Solution to Perceptual Variation and Illusions

A little reflection suggests that what we perceive isn’t quite the same as what is ‘out there’. In *The Problems of Philosophy*, Bertrand Russell gives an example of looking at a shiny, brown table. We say it is brown, but it doesn’t actually look an even brown colour all over: depending how the light falls, some parts are lighter than others, and some are even white from the shininess. So Russell objects that saying the table is brown means no more than that it looks brown ‘to a normal spectator from an ordinary point of view under usual conditions of light’ – but why think that this colour is more real, more a property of the table, than any of the other colours that you experience? Just what colour any part of the table looks to you depends on where you stand. If you and someone else look at the table together, you will see different patterns of colour. Suppose a shiny spot on the table looks light brown to you but white to the other person. The table can’t *be* both brown and white in the same spot at one time.

Russell then runs the same argument, appealing to variations in our perceptual experience, for the properties of texture and shape. The table might be smooth to touch, but at a microscopic level, there are all kinds of bumps and dips – so should we say that when we touch the table, the smoothness we feel is a property of the table? And the shape that something appears to have, like its colour, varies with the angle from which you view it. A rectangular table, from every angle except 90 degrees, does not look perfectly rectangular.

We can press the case by appealing to illusions. If you look at an oar half-submerged in water, it looks crooked; but it isn’t. We see a crooked oar, but the oar isn’t crooked. However, *just* from what you experience, you can’t tell whether you are seeing an illusion or not. Someone who doesn’t know about the illusion thinks they are seeing a crooked oar. It *looks* just like a crooked oar. The point applies generally to illusions. From *just* what we see in an illusion, without other background knowledge, we cannot tell whether what we are seeing is an illusion or not. Illusions can be ‘subjectively indistinguishable’ from veridical perception.

We can combine these reflections on perceptual variation and illusion to make the following argument:

P1. There are many perceptual experiences in which what we experience are not the properties of physical objects (an oblong table, a crooked oar…).

P2. When we perceive something as having some property *x*, then there is something that is *x*.

C1. In such cases, given that what we perceive is not the way the world is, what we perceive are sense-data.

P3. Such cases are subjectively indistinguishable from veridical perception.

P4. When two perceptual experiences are subjectively indistinguishable, they are perceptual experiences of the same thing. (This claim is the best hypothesis, given (P3).)

C2. Therefore, we always perceive sense-data (not just in cases in which what we perceive is not the way the world is).

P5. Nevertheless, except in hallucinations, it still makes sense to say we perceive the world. In cases of both veridical perception and illusion, the sense-data we perceive are caused by and represent physical objects. This representation can be accurate or inaccurate in certain ways – physical objects may be as they appear to us, or they may differ in certain ways.

C3. Therefore, we perceive physical objects indirectly, via sense-data.

Direct realists, of course, have responses to the challenging examples that indirect realists raise. But indirect realists, of course, are unpersuaded by these replies. (P4) records this: direct realism does not satisfactorily explain what two subjectively indistinguishable perceptual experiences have in common. A better explanation is to accept that we perceive sense-data in all cases. (P5) states the commitment to realism, and introduces two new ideas about the relation between sense-data and physical objects, viz. that sense-data are caused by physical objects and that they represent them.

# 2.2.2. Locke’s Distinction Between Primary and Secondary Qualities

We can develop indirect realism further by look at an important distinction in the qualities that we perceive physical objects to have.

In *An Essay concerning Human Understanding*, John Locke defended a form of indirect realism. As part of his theory, he argued that we can distinguish the qualities that we perceive – such things as shape, size, colour, warmth, and so on – into two kinds. He uses the distinction to provide an additional argument for rejecting direct realism in favour of indirect realism. While the distinction between ‘primary’ and ‘secondary’ qualities is most famously associated with Locke, many other philosophers and scientists working at the same time (the seventeenth century) also made the distinction in some form.

Locke’s argument begins in Bk 2, Ch. 8, §8. A ‘quality’ is a ‘power’ that a physical object has ‘to produce an idea in our mind’. So a snowball has the powers – the qualities – to produce in us the ideas of ‘white’, ‘cold’ and ‘round’.

Locke then argues that qualities are of two different kinds. Primary qualities are qualities that are ‘utterly inseparable’ from the object whatever changes it goes through, e.g. even if it is divided into smaller and smaller pieces. The object has these properties ‘in and of itself’. The primary qualities are extension (Locke also talks of size), shape, motion, number and solidity. Secondary qualities are qualities that physical objects have that are ‘nothing but powers to produce various sensations in us’. Locke lists ‘colours, sounds, tastes, and so on’, later adding smells and temperature.

The important phrase here is ‘nothing but’. Primary qualities, of course, also produce sensations in us – both the roundness (primary quality) and the whiteness (secondary quality) of the snowball cause sensations in us. But shape is a quality that the snowball has irrespective of whether we perceive it or not. Colour, by contrast, has to be understood in terms of how the snowball affects us. By definition, colour is something that is experienced in vision. So it is a quality that an object can have only in relation to its being seen by some one. And similarly for sound, taste and the other secondary qualities. By contrast, primary qualities are those properties of an object that are not related by definition to perceivers.

The distinction between primary and secondary qualities is a distinction between qualities that physical objects have ‘in themselves’, and qualities they have that are related to how they are perceived.

## 2.2.2.1. Locke on primary qualities

Why does Locke pick out extension, shape, motion, number and solidity as primary qualities? He says that these qualities cannot be separated from a physical object. For example, physical objects must always have some size and shape, they must always be at rest or in motion of some kind, they can be counted. By contrast, physical objects don’t have to have the secondary qualities of colour or smell, e.g. odourless, clear glass.

Is Locke’s list right? He believed that when you break up physical objects, you get smaller objects which also have all the primary qualities. But physics has moved on, and sub-atomic particles aren’t like physical objects that we know in lots of ways. Many of them have some form of electrical charge and many of them can behave as much like packets of energy as like small bits of matter. We may want to change Locke’s definition of primary qualities to those qualities that physics tells us physical objects have ‘in and of themselves’.

In Bk 2, Ch. 4, Locke explains what he means by ‘solidity’. He does not mean to contrast being ‘solid’ with being liquid or gas. Rather, solidity is the quality of a physical object whereby it takes up space and excludes other physical objects from occupying exactly the same space. This is just as true of liquids and gases; if I fill an empty glass with water, the liquid drives the air out. If I add air to a balloon, the balloon inflates, pushing air outside the balloon away from the space now filled by the balloon (and the air inside it). Just as anything physical has to have some size and shape, thinks Locke, it must also take up space.

## 2.2.2.2. Qualities and resemblance

In Bk 2, Ch. 8, §15, Locke argues that our perceptual experiences of primary qualities ‘resemble’ the primary qualities of the object. Physical objects have shape, extension and so on just as we perceive them. By contrast, our perceptual experiences of secondary qualities don’t resemble the object as it is in itself at all. Secondary qualities are the result of the primary qualities of the object’s ‘imperceptible parts’ (§15) – or as we would now put it, its atomic and molecular structure. Light, by which we perceive colour, can be explained in terms of the effects and activity of subatomic particles, smell in terms of chemical compounds, and so on. Physics and chemistry deal only with primary qualities – the size, shape, motion and so on of tiny bits of matter. So secondary qualities as we perceive them are nothing like what they are in the object, viz. macroscopic effects of the primary properties of atoms and molecules.

Locke uses the distinction to argue for indirect realism. The world as we experience it through our senses and the world as it is ‘in itself’, as science describes it, are quite different. We experience all the wonderful secondary properties of the senses; the world as described by science is ‘particles in motion’ and empty space. It must be, then, that we don’t perceive physical objects directly.

**2.2.3. Objections to Indirect Realism**

Indirect realism claims that we perceive physical objects which are mind-independent, but we do so via, or in virtue of, perceiving mind-dependent sense-data that are caused by and represent physical objects. We perceive sense-data immediately, and physical objects indirectly, mediated by our perception of sense-data. In this section, we raise objections to this theory of perception.

# 2.2.3.1. Objection 1: Scepticism about the Existence of Mind-Independent Objects

In *The Problems of Philosophy*, Russell raises a puzzle for indirect realism. If what we perceive directly are sense-data, then all we know about are sense-data. We believe that ‘behind’ the sense-data there are real physical objects, that physical objects cause our sense-data. But how can we know this? To know that physical objects cause sense-data, we first have to know that physical objects exist. But the only access we have to physical objects is through our sense-data.

Russell’s line of thought forms an objection to indirect realism. Because we directly perceive sense-data, we cannot know that a world of physical objects – a world external to and independent of our minds – exists. Scepticism is the view that we cannot know a particular claim, in this case the claim that physical objects exist. Indirect realism leads to scepticism about the existence of mind-independent objects. And if we can’t that physical objects exist, we can’t know that sense data are caused by physical objects. But this is a claim that indirect realism itself makes! So if indirect realism is true, we can’t know that it is true.

## 2.2.3.1.1. Russell’s Response: The Existence of the External World is the Best Hypothesis

Russell offers two responses, both appealing to how we should explain what we do know. The first is this:

P1. The fact that sense-data are private means that no two people actually ever perceive the same thing, unless we can say that there are physical objects that they both perceive (indirectly).

P2. People have very similar sense-data if they are at the same place and time.

P3. The best explanation of this is that there are physical objects causing their sense-data: they both perceive the same physical object.

C1. So physical objects exist.

Russell rejects this argument because it assumes something that we can’t know: that there are other people, that they have sense-data, and that their sense-data is similar to mine. To assume that there are other people is to assume that there are physical objects, since people are physical objects. But the question was how, from my sense-data, do I know that there are physical objects? In answering that question, I can’t assume that there are physical objects (such as other people) – that’s begging the question!

So Russell offers a second argument.

P1. Either physical objects exist and cause my sense-data or physical objects do not exist nor cause my sense-data.

P2. I can’t prove either claim is true or false.

C1. Therefore, I have to treat them as hypotheses. (A hypothesis is a proposal that needs to be confirmed or rejected by reasoning or experience.)

P3. The hypothesis that physical objects exist and cause my sense-data is better.

C2. Therefore, physical objects exist and cause my sense-data.

What is Russell’s argument for (P3)? One way to test a hypothesis is to see whether it explains why my experience is the way it is. If I see a cat first in a corner of the room and then later on the sofa, then if the cat is a physical object, it travelled from the corner to the sofa when I wasn’t looking. If there is no cat apart from what I see in my sense-data, then the cat does not exist when I don’t see it. It springs into existence first in the corner, and then later on the sofa. Nothing connects my two perceptions. But that’s incredibly puzzling – indeed, it is no explanation at all of why my sense-data are the way they are! So the hypothesis that there is a physical object, the cat, that causes what I see is the best explanation of my sense-data.

(Russell runs the same argument for supposing that other people have minds. When I perceive how people behave, e.g. when talking to me, the best explanation of my experience is that it is caused by what they say (a physical event) and what they say is caused by their thoughts.)

## 2.2.3.1.2. Two Responses from Locke that Support Russell

In *An Essay concerning Human Understanding*, John Locke offers two arguments that supports Russell’s line of thought here. First, he notes that in perception, I cannot avoid having certain sense-data ‘produced’ in my mind. By contrast, if I turn from perception to memory or imagination, e.g. by shutting my eyes, I find that I can choose what I experience. Perceptual experiences – which ‘I have whether I want them or not – must be produced in my mind by some exterior cause’ – physical objects.

Second, Locke presents an argument from comparing perceptual experiences from different senses. He notes that our different senses ‘confirm’ the information that each supplies. If I see a fire and doubt whether it is real, I can confirm its reality by touching it. Another example of this kind was given by Catherine Trotter Cockburn in her ‘A letter from an anonymous writer to the author of *The Minute Philosopher*’. She notes that one and the same object causes perceptual experiences through different senses. The experiences themselves are very different, e.g. seeing an object and hearing the sound it makes. But we learn which visual experiences go with which auditory experiences, e.g. seeing a dog and hearing a dog’s bark. With the association made, we can accurately infer from one experience to the other; just from seeing a dog, we know what sound it will make. And if we experience a change in vision, e.g. a train moving from near to far, we can infer the change in sound, e.g. its horn becoming quieter. Why would we be able either to confirm our experiences using different sense, or be able to predict them, unless there is something which both senses perceive but that is independent of being perceived by any particular sense?

Locke brings the two arguments together in an extended example. I know from experience that I can change how a piece of paper looks by writing on it. (This connects sight and proprioception – my sense of my hand moving.) I can plan what to write, and I know in advance what the paper will look like. But I cannot bring about the sense-data of seeing the paper with words on it just by imagination; I have to actually write. And once I have written something, I can’t change the words I see. This shows that sense-data aren’t ‘merely playthings of my imagination’. Finally, if someone else reads those words aloud, what I hear corresponds to what I intended to write. And this ‘leaves little reason for doubt’ that the words as written on the paper exist independent of my mind.

## 2.2.3.1.2.1. Evaluation of Locke

##

Locke claims to have shown that mind-independent objects exist, that there ‘must’ be some external cause of sense-data. But this is overstating the case, and Trotter Cockburn is more accurate in talking of inferring such a cause. Both are presenting the same argument as Russell, strengthening it by adding further features of our experience that need explaining. If physical objects don’t exist, we can’t explain

1. why sense-data aren’t under our control but imagination and memory are;

2. why we should get the same information from different senses;

3. why we can infer from perceptual experiences of one sense, e.g. vision, what perceptual experiences we would have in another sense, e.g. audition;

4. the very complex interaction between our actions and our perceptions.

So we have very good reason to claim that physical objects exist and cause our sense data.

## 2.2.3.1.3. The existence of mind-independent objects is not a hypothesis

If indirect realism is correct, then it seems the existence of physical objects remains a hypothesis, something we have to infer. We can argue that this is a significant weakness. First, perhaps some other hypothesis that explains our sense-data is just as good, but we just don’t know it. Second, it is very counterintuitive to think that perception doesn’t put us in direct touch with physical objects. But this is entailed by the claim that we have to infer the existence of physical objects.

Indirect realism can respond to these objections by rejecting the theories of Russell, Locke, and Trotter Cockburn. They write as if sense-data ‘come between’ us and the world, with physical objects merely being the cause of sense-data, so that in perceiving sense-data, we aren’t also perceiving physical objects. But instead, we should say that we perceive physical objects via sense-data. Sense-data don’t get in the way of perceiving physical objects. They are how we perceive physical objects. They don’t block our access to the external world, they mediate it. The existence of the external world is not a hypothesis. It is something that we experience in perception.

But what of the fact that sense-data differ from the physical objects they represent, e.g. in perceptual variation and illusions? Doesn’t this show that sense-data come between us and the world? No, this is all explicable in terms of physical objects and their effects on us, and only in these terms. The best explanation of illusions and perceptual variation needs both sense-data and physical objects. We can develop this thought by appealing to how sense-data represent physical objects.

# 2.2.3.2. OBJECTION 2: Representation, Resemblance and the Nature of Physical Objects

Indirect realism maintains that sense-data are not only caused by mind-independent objects, they also represent them. One way of understanding this is that mind-independent physical objects are like our experiences of them in many respects. Locke says that physical objects ‘resemble’ our sense-data. For example, physical objects have shape and size, and so resemble my experience of their shape and size. At the very least, we can say that there are systematic correlations between what we experience and the nature of the world. My experience of the shape and size of an object is (at the very least) systematically correlated with what shape and how large the physical object is. Perception is the source of our knowledge of not only the existence of mind-independent objects, but also something of their nature. This then raises the question, what are physical objects really like? Of all the properties I experience physical objects having, which ones do they have? How should we draw the distinction between the appearance of physical objects and their reality? Locke answers this question by talking about the primary and secondary qualities of physical objects. he argued that we can distinguish the qualities that we perceive – such things as shape, size, colour, warmth, and so on – into two kinds. Primary qualities are qualities that are ‘utterly inseparable’ from the object whatever changes it goes through, e.g. even if it is divided into smaller and smaller pieces. The object has these properties ‘in and of itself’. The primary qualities are extension (Locke also talks of size), shape, motion, number and solidity. Secondary qualities are qualities that physical objects have that are ‘nothing but powers to produce various sensations in us’. Locke lists ‘colours, sounds, tastes, and so on’, later adding smells and temperature. Secondary qualities as we experience them are nothing like how physical objects are in themselves. However, our experience shows us that physical objects ‘in themselves’ have primary qualities which we then experience.

But how do we know that physical objects have primary qualities that resemble our experience of them? One difficulty is that, in general, we can’t tell what a cause is like just from its effects. Consider: if all you knew was smoke, would you be able to work out that its cause was fire? Fire is very different from smoke. Experience shows that the world is full of surprising causal relationships. So, if all we experience are sense-data, how can we know whether the world is similar to how it appears to us in sense-data, say in having the primary qualities we experience, or whether it is very different? Can we rule out the claim that physical objects cause our experiences of primary and secondary qualities, but don’t resemble these experiences at all?

## 2.2.3.3. OBJECTION 3: Berkeley’s Argument That Mind-Dependent Ideas Cannot Be Like Mind-Independent Objects

As we saw, Locke claims that primary qualities in the object resemble our experience of them. For example, the squareness of a physical object resembles the squareness we see. In his *Three Dialogues between Hylas and Philonous*, George Berkeley questioned whether this makes sense.

Our experience of the distinction between appearance and reality shows our sense-data change depending on the conditions of perception but the physical object does not. This applies just as much to primary qualities as secondary qualities. For instance, a rectangular table remains rectangular, even as the way it looks to me changes as I look at it from different angles. This is true of the size and shape of almost all the physical objects we commonly perceive. Furthermore, we constantly flit our eyes from one thing to another, so what we experience at any moment changes. Again, we don’t take these changes in our sense-data to be the result of changes in the physical objects we are looking at.

This forms the basis of an objection: how can our sense-data, which are ‘perpetually fleeting and variable’ be ‘like’ or ‘resemble’ a physical object is ‘fixed and constant’? For instance, how can circular sense-data and oval sense-data both resemble something that has just one shape? If you want to say that one of these appearances resembles the object, while all the others do not, then how do we distinguish which is the ‘true copy’ – the true size or shape?

Second, physical objects are themselves not something we experience (directly) – we only experience the sense-data that they cause. But how can something that we don’t experience (a mind-independent physical object) be like something that is experienced (mind-dependent sense-data)? What can we mean when we say that the shape of the table ‘resembles’ the shape we see? How can squareness (as it is in the object) resemble the idea of squareness? Our ideas of size, shape, motion and so on, derive from our perceptual experience. The only idea of shape we have is the one we see (or feel). We can only make sense of the table’s squareness in terms of our experience of squareness. There is nothing like a perceptual experience except another perceptual experience. So we can’t meaningfully say that our sense-data are like or resemble physical objects. But in that case, Berkeley concludes, if indirect realism is true, then we can’t know the nature of mind-independent objects.

Indirect realists have generally agreed that Locke’s idea of ‘resemblance’ between sense-data and physical objects is problematic. But, they say, we can still argue that sense-data represent physical objects (just not by resembling them). The pattern of causal relations between the external world and our sense-data is very detailed and systematic. We can explain how sense-data represent physical objects in terms of this complex causation.

But does this answer the objection? Because causes and effects can be very different, if we understand representation just in terms of detailed and systematic causal relation, it seems that we still won’t know what mind-independent objects are like in themselves.

**2.3. Berkeley’s Idealism**

Berkeley rejects our usual understanding of physical objects as mind-independent. He claims that reality is dependent on minds. The ordinary objects of perception – tables, chairs, trees and other physical objects – must be perceived in order to exist at all. The only things that exist are minds (that perceive) and what minds perceive. The claim that nothing exists that is independent of mind is idealism. Does it make sense, and why does Berkeley argue for it?

# 2.3.1. Berkeley on Primary and Secondary Qualities

John Locke argued that we can distinguish the qualities that we perceive – such things as shape, size, colour, warmth, and so on – into two kinds: primary and secondary qualities (see section 2.2.2. on page \*\*\*).

Berkeley begins his *Three Dialogues between Hylas and Philonous* by arguing that secondary qualities are mind-dependent. But he then uses similar arguments to show that primary qualities are also mind-dependent.

Berkeley, in the character of Philonous, begins by arguing that ‘sensible things’, i.e. things perceived by the senses, must be perceived immediately by the senses. The causes of our perceptions – the reality behind appearances – if they are not immediately perceived, we must infer. Whatever is inferred is not perceived: because what we perceive is immediately apparent to us, if we need to infer something then we aren’t perceiving it. So if we need to infer the causes of our perceptions, we should not say that such causes are themselves perceived.

Philonous then argues that what we immediately perceive are the qualities of things, and nothing more. Through vision, we perceive colours, shapes, size, etc.; through hearing, sounds; through smell, odours – and so on. Each sense perceives particular types of qualities. There is nothing we perceive in addition to these qualities.

Do these qualities exist independently of being perceived? The character Hylas, who plays the role of the realist, starts off as a very simple direct realist. He claims that whatever we perceive exists independently of our minds, and in the form in which we perceive it. Thus heat, as we feel it, exists in the object.

Philonous points out that intense heat, or indeed intense cold, light, sound, pressure, can all be experienced as pain, but pain obviously doesn’t exist ‘in’ physical objects. If we say these secondary qualities are in the object, then we have to say that at some point, it also has the quality of pain. Hylas responds that the heat (light, etc.) isn’t itself pain, but causes pain. Philonous objects that we feel just one unified sensation of painful heat (painful light, etc.). In fact, all our perceptions of secondary qualities are accompanied by some form of pleasure or pain. Since this is an empirical argument, we can challenge it on empirical grounds. Is Berkeley right to say that we can’t distinguish between the sensation of heat and that of pain? If he is right for some sensations, is he right for all sensations? Can we not, for instance, distinguish between the sensation of sweetness and the sensation of pleasure we associate with it?

Berkeley’s second argument for thinking secondary qualities are mind-dependent is an argument from perceptual variation. He repeats Locke’s example of placing a hot and a cold hand in a bowl of tepid water. The water feels hot to the cold hand and cold to the hot hand, but the water cannot be both hot and cold. He later develops the argument in relation to colours:

P1. A cloud from a distance looks pink, but up close, it loses its colour (or appears grey).

P2. A solid physical object, viewed through a microscope, appears to have different colours than those it has when viewed normally.

P3. Different animals perceive the colours of objects differently.

P4. If colours really existed in physical objects, then to change the colour, it would be necessary to change the object itself. But, of course, different kinds of light – daylight, candlelight, etc. – change the colour of an object without changing the object.

C1. Therefore, all colours are appearances, not properties of physical objects.

Suppose we insist that secondary qualities ‘really’ exist in the object as physical particles in motion (light waves, sound waves, the chemicals of smell and taste). Berkeley points out that if we say that sound is a vibration of the air, then we can’t hear sound, since vibrations are something perceived by sight or touch. If we say that colour is tiny particles of matter in motion (photons with a particular energy, perhaps), then we can’t see ‘real’ colour, since we cannot see these tiny particles moving. And that is very counter-intuitive.

## 2.3.1.1. Berkeley’s attack on the primary/secondary quality distinction

Having persuaded Hylas to agree that secondary qualities are mind-dependent, Philonous (Berkeley) argues that the argument from perceptual variation applies equally well to primary qualities.

P1. What looks small to me may look huge to a small animal.

P2. What looks small from a distance looks large when viewed close up.

P3. What looks smooth to the naked eye appears craggy and uneven under a microscope.

P4. If you look at a circle straight on, it looks circular. But if I’m looking at it from an angle, it looks elliptical. We see it differently, but it doesn’t change.

P5. Even motion isn’t constant. We measure the speed of motion by how quickly our minds work – to a creature that thinks much faster than us, e.g. a housefly, our fastest movements appear leisurely.

P6. In the case of colour, when an object appears to have many colours, depending on how it is perceived, we can’t say that it has one real colour which is independent of how we perceive it.

C1. Therefore, P1–P5 show that we can’t say that an object has one real shape or size or motion, independent of how it is perceived.

C2. Therefore, the primary qualities of objects are just as mind-dependent as secondary qualities.

Primary qualities vary just as much as secondary qualities, and so we have no reason to attribute either kind of property to something that exists independent of our perceptual experiences.

# 2.3.2. The Immediate Objects of Perception are Mind-Dependent Objects

We have seen that Berkeley argues that both primary and secondary qualities are mind-dependent. He has also argued that when we perceive physical objects, we don’t perceive anything in addition to its primary and secondary qualities. This gives us the following simple argument:

P1. Everything we perceive is either a primary or a secondary quality.

P2. Both primary and secondary qualities are mind-dependent.

C1. Therefore, nothing that we perceive exists independently of the mind: the objects of perception are entirely mind-dependent.

This doesn’t show that there aren’t any mind-independent physical objects, because they could exist unperceived. But as we shall see, Berkeley goes on to argue that the idea of a physical object as something that exists independently of our perception of it is an idea so problematic that we should reject it entirely.

Once we grant Berkeley’s claim that all we perceive are primary and secondary qualities, it becomes more difficult to reject his later arguments for idealism. One way to challenge his idealism, therefore, is to argue that we can be said to perceive physical objects themselves, and not just their qualities.

## 2.3.2.1 Three arguments against mind-independent objects

### 2.3.2.1.1. ‘A material substratum’

Suppose we perceive only primary and secondary qualities. If there were no mind-independent objects, what would ‘have’ the qualities we experience? Hylas argues that we need the idea of ‘a material substratum’ – the stuff or substance that possesses primary and secondary qualities and holds them together to make one thing, one physical object. This ‘material substratum’ can exist unperceived.

Berkeley responds that this idea is, in fact, incoherent. The ‘material substratum’ is never perceived, since it is distinct from its primary and secondary qualities, and we have said that all we perceive are primary and secondary qualities. So what can we say about it? Once you list all the qualities of a table, what is left of the table? For instance, size is a quality – if the matter of the table is distinct from its qualities, then in itself, it has no size! If substance exists unperceived, it exists without any qualities at all.

Locke saw the point, and accepted that the idea of substance was the idea of something unknown. A realist view of physical objects involves a mystery. Worse, Berkeley argues, physical substance is quite literally inconceivable – we can say nothing about how it exists at all.

### 2.3.2.1.2. ‘We perceive physical objects’

Don’t we just see that physical objects exist? Berkeley argues that neither our senses nor reason supports such a claim.

P1. As argued previously, all we perceive are primary and secondary qualities, not mind-independent physical objects.

C1. Therefore, our experience cannot verify the hypothesis that there is a mind-independent physical world.

P2. Worse still, the hypothesis of ‘physical substance’ is not one that is even suggested by experience.

C2. So close attention to experience supports the claim that all there is (all we can say there is) is what we can experience.

P3. What we experience are ideas.

C3. Therefore, our experience supports idealism, not realism.

### 2.3.2.1.3. Scepticism

Berkeley’s next argument is that supposing that the objects of perception can and do exist independently of being perceived leads to scepticism. How can we connect up our experiences to something ‘beyond’ them – which, following the objection just made, we can’t even describe or understand? How we can know that ideas really do represent (and represent accurately) something that exists completely independently of them? If there were mind-independent physical objects, we can know nothing about them. By contrast, if there is no mind-independent reality, then what we perceive is what there is, viz. ideas, and so perception can give us knowledge of what there is.

## 2.3.2.2 Berkeley’s ‘master’ argument

Berkeley provides another argument against the possibility of the objects of perception being mind-independent. It has come to be known as his ‘master’ argument, since he appears to set great weight upon it. Thus, Philonous says, ‘I am willing to let our whole debate be settled as follows: If you can conceive it to be possible for any mixture or combination of qualities, or any sensible object whatever, to exist outside the mind, then I will grant it actually to be so’. Hylas responds that he is thinking of a tree existing unperceived by anyone. Philonous objects that what Hylas is thinking depends on his mind. He isn’t actually thinking of a tree that exists independently of any mind; he is imagining a tree standing ‘in some solitary place’ where no one perceives it. But all the time, he is thinking of such a tree. We cannot think of a tree that is neither perceived nor conceived of. We can think of the idea of a tree, but not of a tree that exists mind-independently.

However, Berkeley seems to have confused a thought with what the thought is about.

P1. Thoughts cannot exist outside the mind – thoughts are psychological events or states.

C1. Therefore, my thinking of a tree is not mind-independent. It is impossible (inconceivable) that there is a thought of a tree when no one is thinking of a tree.

P2. But what a thought is about, e.g. a tree, is not the same thing as the thought itself.

C2. Therefore, just because my thinking of a tree is mind-dependent, it does not follow that what I am thinking of is also mind-dependent. It is not impossible (inconceivable) to think that a tree may exist when no one is thinking of it.

# 2.3.3. Idealism and God

Berkeley has argued that it makes no sense to think of physical objects as mind-independent. His arguments entail that what we think of as physical objects – indeed, what we must mean by ‘physical object’ if the term is to be coherent – are bundles of ideas, the perceptions we have of physical objects. Physical objects exist as mind-dependent things. But without mind-independent physical objects causing our perceptions, what explains why we perceive what we do? Answering this question will complete the outline of Berkeley’s idealist theory.

P1. As (the ideas that comprise) physical objects are mind-dependent, there are three possible causes of my perceptions: ideas, my mind, and another mind.

P2. Ideas (including the ideas that comprise physical objects) don’t cause anything.

P3. If (the ideas that comprise) physical objects depended on my mind, then I would be able to control what I perceive.

P4. But I can’t. Perception is quite different to imagining; in perception, we are more passive – the sensations just occur to us, and we can’t control them. Imagination is voluntary, but perception is involuntary.

C1. Therefore, (the ideas that comprise) physical objects don’t depend on my mind.

C2. Therefore, (the ideas that comprise) physical objects must exist in another mind, which then wills that I perceive them.

C3. Given the complexity and systematicity of our perceptions, that mind must be God.

## 2.3.4. Idealism and the cause of our perceptions

Why does Berkeley claim (P2) above, that ideas can’t be causes? Think about the functioning of the mind. Should we say that thoughts think themselves, or that the mind thinks thoughts? Should we say that our perceptions cause our experience of them or that the mind perceives? Berkeley thinks the latter option is clearly better. Ideas themselves are passive; we do things with ideas in thought or we receive them in perception. How could an idea ‘make’ you think it or perceive it? It is the mind, and only the mind, that can cause, that is active – the mind that perceives, thinks, wills, and so on.

Berkeley uses this thought not only to support idealism in the argument above. He also gives two arguments for the claim that ‘matter’ can’t be the cause of our perception. Hylas suggests that ‘matter’ is simply whatever is the cause of our perceptions. Berkeley responds:

P1. Matter in the normal sense of the word, i.e. as mind-independent and possessing primary qualities, cannot exist (as argued previously).

C1. Therefore, to talk any sense about matter, we must think of it as our perceptions of it.

P2. What we perceive – primary and secondary qualities – are ideas.

C2. Given that ideas are passive, whatever causes our perceptions must be a mind, not ‘matter’.

In a second argument, Berkeley points to our ignorance of how mind and matter can interact. Realism requires that our minds are causally affected by mind-independent physical objects. Suppose physical objects causally affect our sense organs, which then affect our brains. Philosophers and scientists have struggled with the next step – how does what happens in our brains causally affect our conscious perception? How can something physical and mind-independent possibly cause an idea in a mind? How could nerve signals in the brain produce sensations of sound and colour? 300 years after Berkeley made the objection, the puzzle still remains unsolved.

Berkeley is aware that his idealism is counter-intuitive. But, he argues, it follows from his previous arguments, and there is nothing impossible about his conclusion. We know from our own experience that minds can give rise to thoughts. Berkeley adds that many metaphysical puzzles can be solved by adopting idealism: for example, we can establish the existence of God and dissolve problems about the ultimate nature of matter, how matter can cause ideas in a mind, and how matter could ever produce mind. The rest of his defence of idealism amounts to answering possible objections and correcting misunderstandings.

**2.3.5. Objections to Idealism**

In this section, we discuss three objections to Berkeley’s theory and his responses. The objections relate to illusions, solipsism and the role of God.

# 2.3.5.1. Problems with the Role Played by God in Berkeley’s Idealism

However persuasive one finds Berkeley’s arguments regarding perception, one may object to his appeal to God. It is important to note, however, that Berkeley does not assume that God exists, and then wheel him in to resolve philosophical difficulties in his theory. Rather, the existence of God is an inference, supported by the arguments. The cause of our perceptions is a mind, because we can only conceive of minds being active: ‘I have no notion of any action other than volition, and I can’t conceive of volition as being anywhere but in a spirit’. The ‘variety, order, and manner’ of what I perceive shows that the mind that produces these ideas is ‘wise, powerful, and good, beyond anything I can comprehend’. I derive the idea of God from my knowledge of my own mind, ‘heightening its powers and removing its imperfections’.

But the exact relationship between (the mind of) God and what we perceive is puzzling. Berkeley infers that our perceptions are caused by God. Physical objects don’t depend on my mind; but as ideas, they depend on some other mind. So, Berkeley says that they exist in the mind of God. But does this make sense? There are three reasons to think that ideas we are caused to have, whether perceptions or sensations, can’t be part of God’s mind:

P1. My perceptions and sensations are part of my mind. What I perceive and feel is in my mind, not God’s mind.

P2. God can’t have the sorts of perceptual experiences I have – God doesn’t perceive as I do, and does not undergo sensations, such as pain.

P3. The ordinary objects of my perception change and go out of existence, but God’s mind is said to be unchanging and eternal.

C1. Therefore, what I perceive and feel can’t be part of God’s mind.

Berkeley clarifies his theory by responding:

1. What I perceive is a copy of the idea in God’s mind.

2. The ideas of physical objects exist in God’s mind not as perceptions, but as part of God’s understanding. The same is true of sensations. So while God doesn’t perceive the colour red or feel pain, he knows what it is for us to undergo these experiences.

3. What I perceive, which changes, is what God wills me to perceive. The whole of creation exists in God’s understanding, eternally; ‘things . . . may properly be said to begin their existence . . . when God decreed they should become perceptible to intelligent creatures’.

We will see that understanding the role of God in Berkeley’s idealism enables us to understand his responses to the other objections that can be raised.

# 2.3.5.2. Arguments from Illusion and Hallucination

In the *Three Dialogues*, Hylas asks how idealism can explain illusions. Since we perceive ideas, there must be an idea that corresponds to the illusion. But we don’t want to say that the physical object is as it looks in the illusion. If we see an oar half-submerged in water, it looks crooked, but it isn’t. But the oar is just what we see; and what we see is crooked, not straight.

Berkeley’s response is that we aren’t misperceiving – what we perceive in the case of the half-submerged oar is crooked. However, this is misleading if we infer that the oar would feel crooked if we touched it or would look crooked when pulled out of the water. So illusions mislead us regarding the ideas we might associate with what we perceive.

This entails that the oar is crooked when half-submerged. Because Berkeley argues that reality is the ideas we perceive; there is no appearance–reality distinction. But to say the oar is crooked is very odd indeed – it just sounds false!

In *Three Dialogues*, Berkeley doesn’t consider or respond to this objection. Elsewhere in his writings, however, he replies that the problem here is with language. He agrees that we shouldn’t say ‘The oar is crooked’, since what we understand that to mean is that it would look crooked under normal conditions. And this is false. So to avoid this implication, we should say ‘The oar looks crooked’ – and this is correct.

What about hallucinations? Berkeley discusses these, in the form of dreams. Hallucinations are products of imagination. Normally, imagination is voluntary, and perception is not. But hallucinations are involuntary, like perceptions, so Berkeley provides two other criteria that mark off hallucinations from perception. First, they are ‘dim, irregular, and confused’. Second, even if they were as ‘vivid and clear’ as perceptions, they are not coherently connected with the rest of our perceptual experience.

To this, we might object that these criteria mark a difference of degree – perceptual experiences can be more or less clear or dim, more or less coherently connected with other experiences. But surely the difference between hallucination and perception is a difference in kind. In perception, you experience something that exists outside your mind, in hallucination, you don’t. In response, perhaps Berkeley could agree – the ideas you perceive originate in God, but in hallucination they don’t. His criteria are only supposed to indicate how we can tell.

**2.3.5.3. Idealism Leads to Solipsism**

Solipsism is the view that only oneself, one’s mind, exists. There are no mind-independent physical objects and there are no other minds either. We can object that Berkeley’s idealism leads to the conclusion that all that exists is my own experience. Or at least, experience gives me no reason to believe that anything apart from my experience exists (or can exist). If all I perceive are ideas, what reason do I have to think that other minds exist? For that matter, what reason do I have to think that minds (including God) exist? After all, I do not perceive minds.

Berkeley doesn’t discuss this objection from solipsism explicitly, though Hylas expresses a version of it, and Berkeley makes a number of remarks we can draw upon. He accepts that ‘strictly speaking’, I have no idea of a mind. But because I am a mind – a ‘thinking substance’ – I know I exist.

P1. The mind is that which (actively) perceives, thinks and wills, while ideas are passive.

P2. I am aware of myself as capable of this activity.

C1. Therefore, I am not my ideas, but a mind.

P3. Being a mind myself, I have a ‘notion’ of what a mind is.

C2. Therefore, it is possible that other minds exist.

P4. My perceptions don’t originate in my mind.

C3. Therefore, they are caused by some other mind.

C4. The complexity, regularity, etc., of my experience indicates that this mind is God.

As for other finite minds – other people – Berkeley doesn’t spend much time on the matter, but indicates that there is evidence in my experience that they exist. Their existence is a matter of inference.

**3. Reason as a Source of Knowledge**

3.1. Introducing key terms and concepts

What kinds of knowledge are there, and how do we gain knowledge? If, as Linda Zagzebski suggests, knowledge involves being in ‘cognitive contact’ with reality, what means of being in contact with reality do we have? One obvious and immediate answer to the question ‘how do we gain knowledge?’ is ‘perception’ or ‘sense experience’ - awareness of physical objects through our senses. But does all our knowledge, directly or indirectly, come from perception?

An alternative is that reason provides us with an independent source of knowledge. This question goes to the very heart of epistemology in reflecting on how human beings are ‘hooked up’ to the world. It makes a central contribution to our understanding of the nature and possibilities of human thought. The debate over the origin (and nature) of our knowledge coincided with the scientific revolution in Europe. The debate led to the first proposals about how modern science works and the type of knowledge it can give us. And because philosophy is form of thinking that relies (more) on reasoning than on an empirical investigation of the world, this debate also has implications for the nature and scope of philosophy itself.

We can identify three positions on the origin of our knowledge. Very roughly, empiricism claims that all our knowledge comes from sense experience, rationalism claims that we can gain further knowledge by pure reasoning, while innatism claims that our minds are innately predisposed to know certain truths. To help us think about the debate between these views, we need to make a number of distinctions, the first between two types of knowledge, and the second and third between ways in which propositions can be true. This section presents these distinctions and their relations to the three views about the origin of knowledge.

# 3.1.1. A Priori/A Posteriori Knowledge

We may draw a distinction between two types of knowledge, based on how we know whether a proposition is true:

*A priori*: We have a priori knowledge of a proposition if we do not require sense experience to know it to be true. An example is ‘Bachelors are unmarried’. If you understand what the proposition means, then you can see straightaway that it must be true. You don’t need to find bachelors and ask them if they are married or not. Another example is ‘537 + 654 = 1191’. You can figure out whether this is true just by thinking about it.

*A posteriori*: Propositions that can only be established through sense experience are known a posteriori. An example is ‘There are more than 6 billion people on the Earth’.

The a priori/a posteriori distinction rests on how we check or establish knowledge of a proposition. How we come to understand the proposition is irrelevant. To learn what a proposition means, to acquire the concepts or words involved, we may well always need sense experience. For instance, to understand ‘Bachelors are unmarried’, we will first need to learn English, and that requires sense experience. But how we learn to understand a proposition is a different issue from how, once we understand it, we check if it is true.

(Philosophers sometimes also talk about a priori and a posteriori concepts. An a posteriori concept is one that is derived from experience. An a priori concept is one that cannot be derived from experience.)

# 3.1.2. Analytic/Synthetic Propositions

The contrast between ‘analytic’ and ‘synthetic’ is a contrast between types of proposition:

*Analytic*: A proposition is analytic if it is true or false just in virtue of the meanings of the words. Many analytic truths, such as ‘squares have four sides’, are obvious, but some are not, e.g. ‘In five days’ time, it will have been a week since the day which was tomorrow three days ago’ (think about it!).

*Synthetic*: A proposition is synthetic if it is not analytic, i.e. it is true or false not just in virtue of the meanings of the words, but in virtue of the way the world is, e.g. ‘ripe tomatoes are red’.

You may have already noticed a similarity between the examples for a priori knowledge and analytic propositions. ‘Bachelors are unmarried’ – an example of a priori knowledge – is also an analytic proposition. ‘Squares have four sides’ – an example of an analytic proposition – is also an example of a priori knowledge. So is all a priori knowledge just knowledge of analytic propositions? It is a question we return to below. But first, the third distinction.

# 3.1.3.Necessary/Contingent Truth

The distinction between ‘necessary’ and ‘contingent’ draws a different contrast in how propositions can be true:

*Contingent*: A proposition is contingently true (or false) if it is possible that it could be true or false. Of course, it will be either true or false, but the world could have been different. It is true that you are reading this book; but you could have been doing something else – it could have been false. So it is contingently true. It is contingently true that there are more types of insect than there are of any other animal. This wasn’t always true, and one day it might be false again.

*Necessary*: A proposition is necessarily true if it must be true (or necessarily false if it must be false). Mathematical propositions are necessarily true (or false): 2 + 2 must equal 4; it is not possible (logically or perhaps mathematically possible) for 2 + 2 to equal any other number. Likewise, analytic truths are necessary: if a proposition is true by definition, then it must be true. If a square is, by definition, a closed two-dimensional figure with four sides, it is impossible for there to be a square with three sides. It simply wouldn’t be a square.

(Of course, it is possible that the figure ‘2’ could have been used to mean the number 3 or the word ‘square’ used to mean triangle. But then ‘2 + 2’ wouldn’t mean 2 + 2; it would mean 3 + 3. To test whether a proposition is true or false, in all cases, you have to keep the meanings of the words the same. If ‘2’ means 2, and ‘4’ means 4, then 2 + 2 must equal 4.)

# 3.1.4. Defining Rationalism, Empiricism and Innatism

We can now return to our question of whether reason provides us with knowledge. A posteriori knowledge is knowledge from sense experience, so if reason is a source of knowledge at all, then it provides us with a priori knowledge (of either analytic or synthetic propositions). But how would ‘reason’ provide such knowledge? In fact, philosophers have proposed two distinct theories. The first, innatism, is that such knowledge is ‘innate’, built into that part of the mind with which we think about and understand the world. The second theory is that we can gain knowledge using rational insight and reasoning.

Historically, philosophers who defended the claim that reason is a source of knowledge, such as Plato, Descartes, and Leibniz, connected these two theories and defended both of them, while philosophers who rejected the claim, such as Locke and Hume, rejected both theories. The first group of philosophers were called ‘rationalists’, because they defended reason as a source of knowledge; the second group were called ‘empiricists’, because they argued that all our knowledge derives from sense experience. The two groups also disagreed on how we acquire concepts, and so produced ‘rationalist’ and ‘empiricist’ theories of this as well. To mark their historical nature, let’s call these two families of theories ‘classical rationalism’ and ‘classical empiricism’.

We should now note that the claim that we have innate knowledge can be separated from the claim that we have knowledge through rational intuition and reasoning, and one could be true without the other. It is more useful, therefore, to split the classical debate into two separate debates. In particular, let us split ‘classical rationalism’ into the following two claims, and re-use ‘rationalism’ for just one of these claims:

*Innatism*: Innatism (about knowledge) claims that we have some innate knowledge.

*Rationalism*: Rationalism claims that we have some a priori knowledge from rational insight and reasoning.

For now, we can keep ‘empiricism’ to mean:

*Empiricism*: Empiricism (about knowledge) claims that there is no a priori knowledge which is either innate or gained from rational insight and reasoning.

We should clarify immediately that we are excluding knowledge of our own minds from the debate. We can each know such truths as ‘I feel sad’ or ‘I am thinking about unicorns’. How? Not obviously through sense experience nor reason (and certainly not innately). We don’t need to worry about this. The argument is about knowledge of things other than our own minds.

We can develop our understanding of the debates between classical rationalism and empiricism using the distinctions we drew earlier. We noticed in passing that the distinction between a priori/a posteriori and analytic/synthetic might line up. This would mean that our knowledge of true analytic propositions is always a priori; and our knowledge of true synthetic propositions is always a posteriori. Is this right?

1. Everyone agrees that analytic propositions are known a priori. However, what is the source of this knowledge? Is such knowledge innate? Is it gained by reasoning? Or is it, as empiricists argue, something else again, e.g. a form of conceptual knowledge?
2. Are all synthetic propositions known a posteriori? Empiricists argue that they are. But could we know some synthetic propositions a priori, either innately or through reasoning?

How do these questions connect to the third distinction above, between necessary and contingent truth? Historically, philosophers agreed that knowledge of propositions that are necessarily true is a priori knowledge while knowledge of propositions that are contingently true is a posteriori. Why? Because a posteriori knowledge is knowledge of how the world is gained through our senses, and surely the world as we experience it could always have been a different way – so all propositions about the world could have been true or false. But it is hard to see how necessary truths could be established a posteriori. Take the claim that ‘2 + 2 = 4’ or ‘Squares have four sides’. In his *New Essays on Human Understanding*, Leibniz points out that our sense experience only provides us with information about particular instances – that these two apples and these two apples make four apples; that this square has four sides; and so on. But ‘however many instances confirm a general truth, they aren’t enough to establish its universal necessity’. Our experience tells us how things are, but not how things must be. If we reject this, and argue that ‘2 + 2 = 4’ is just a generalisation of our experience so far, then we are saying that it is possible, one day, that 2 + 2 will equal some other number. But this is inconceivable.

All necessary truths tell us how things must be. Because experience doesn’t tell us how things must be, it seems that all knowledge of necessary truths must be a priori. If our knowledge of necessary truths is innate or gained through rational insight and reasoning, then this will show that empiricism is false.

However, empiricists can argue that necessary truths are analytic propositions, and that knowledge of analytic propositions isn’t innate or gained through rational insight, but a form of conceptual knowledge. To know an analytic truth, one simply needs to understand the concepts involved. As long as the concepts are learned from experience, knowledge of analytic truths is no threat to the empiricist claim that all knowledge ultimately derives from experience.

The real debate between empiricism and rationalism, then, concerns whether we can have a priori of any synthetic propositions that don’t concern our own mental states. We can use this to sharpen our definitions of rationalism and empiricism:

*Rationalism (2.0)*: Rationalism claims that we have some a priori knowledge of synthetic propositions about the world external to our minds.

*Empiricism (2.0)*: Empiricism (about knowledge) claims that there is no a priori knowledge of synthetic propositions about the world external to our minds (whether this is innate or gained from rational intuition and deduction).

Rationalism

|  |  |  |
| --- | --- | --- |
|  | Analytic | Synthetic |
| A Priori |  |  |
| A Posteriori |  |  |

Empiricists

|  |  |  |
| --- | --- | --- |
|  | Analytic | Synthetic |
| A Priori |  |  |
| A Posteriori |  |  |

# 3.1.5 ‘Intuition’ And ‘Deduction’

We will see that the form of reasoning Descartes presents is intended to be deductive. So if the premises are true, then the conclusion must be true if the argument is valid. In this case, we say that the conclusion is entailed by the premises. If the premises are true, but the conclusion could be false, then the deduction has failed. The advantage of using deduction is that it is a way of proving the conclusion from the premises. If we can be confident that the premises are true, and the inference is correct, we can be confident that the conclusion is true.

What is ‘intuition’ in the context of this debate, sometimes also referred to as ‘insight’? This doesn’t mean a ‘gut feeling’ or ‘instinct’. It refers to rational intuition. For example, when you consider a deductive argument, do you understand why, if the premises are true, then the conclusion must be true? Take the example, ‘Socrates is a man and all men are mortal. Therefore, Socrates is mortal.’ How is it that you can ‘see’ the conclusion follows – that it must be true if the premises are? This grasping of rational truths takes us towards the idea of ‘rational intuition’, though it covers much more than deductive reasoning. Another example is necessary truth. How do you understand that 2 + 2 not only equals 4, but must equal 4? Or that it is impossible for the same thing to be and not to be? At the heart of rational intuition is discovering the truth of a claim just by thinking about it. Very often, what we discover in rational intuition is that the claim is true because it must be true.

Descartes puts the two methods together: we know a number of claims by rational intuition, and we can use these as the premises in deductive arguments to gain knowledge of further claims. He argues that using these methods, we can gain knowledge of our own existence as mental substances, of the existence of God, and of the existence and nature of physical objects. In this section, we will just look at the first of these – his cogito.

# 3.1.6. Empiricists on Knowledge

Whether we have rational intuition in this sense is disputed by empiricists, who claim that all knowledge comes from experience. They can agree with rationalists that we can know analytic propositions a priori (just by thinking about them). However, empiricists can argue that knowledge of analytic propositions isn’t a function of ‘rational intuition’, but a form of conceptual knowledge. To know an analytic truth, one simply needs to understand the concepts involved. As long as the concepts are learned from experience, knowledge of analytic truths is no threat to the empiricist claim that all knowledge ultimately derives from experience.

Empiricists also allow that we can, of course, know our own mental states ‘just by thinking about them’. The origin of this knowledge isn’t rational intuition either, but ‘impressions of reflection’ – experience of our own minds.

The real debate between empiricism and rationalism, then, concerns whether we can have a priori of any synthetic propositions that don’t concern our own mental states. Rationalism claims that we have some a priori knowledge of synthetic propositions about the world external to our minds. Empiricism (about knowledge) claims that there is no a priori knowledge of synthetic propositions about the world external to our minds.

Given this, for any claim that rationalists offer as an example of knowledge through rational intuition and deduction, empiricists have four possible responses:

1. that the proposition is analytic, not synthetic;
2. that the proposition is about our own minds, known from impressions of reflection;
3. that knowledge of the proposition is a posteriori, not a priori; or
4. that we can’t know the proposition at all.

## 3.1.6.1 Hume’s fork

In *An Enquiry concerning Human Understanding*, David Hume argues that we can have knowledge of just two sorts of claim: the relations between ideas and matters of fact. He uses two related criteria to make the distinction, though it is easier to grasp what he means by taking them in a different order:

1. Relations of ideas ‘can be discovered purely by thinking, with no need to attend to anything that actually exists anywhere in the universe’. Matters of fact, by contrast, are ‘propositions about what exists and what is the case’.
2. Relations of ideas are statements that are ‘either intuitively or demonstratively certain’ (p. 11). Hume gives the example of 3 × 5 = 30/2 – a statement about the relations of numbers. Relations of ideas that are demonstratively certain are known by deduction. Matters of fact, by contrast, are not known by deduction, because they are statements that can be denied without contradiction.

### 3.1.6.1.1. Relations of ideas

The second point needs explanation. First, a contradiction both asserts and denies something. For example, a true analytic proposition cannot be denied without contradiction. To say that vixens are not foxes is a contradiction in terms; it is to say that female foxes are not foxes. Second, in a deductive argument, if you assert the premises, but deny the conclusion, then you contradict yourself, e.g. ‘3 × 5 = 15 and 30 ÷ 2 = 15, but 3 × 5 ≠ 30 ÷ 2’. Hume is claiming that we gain knowledge of relations of ideas through merely understanding concepts and through deductive inference from such understanding. To deny any claims we know this way would involve a contradiction.

(We can also use deduction to infer matters of fact from other matters of fact, e.g. Socrates is a man and all men are mortal, so Socrates is mortal. But ‘Socrates is mortal’ isn’t known by deduction in Hume’s sense, since the premises rely on sense experience.)

We can now connect the two criteria. What we know that is intuitively or demonstratively certain is also what can be discovered purely by thinking – relations of ideas. On the other hand, propositions about what exists – matters of fact – we cannot know by a priori reasoning. Hume goes on to argue that we can know them through experience.

The history of philosophy is full of debate about what qualifies as relations of ideas in Hume’s sense. As we shall see, Descartes argues that a great deal can be known through rational intuition and demonstration, while Hume rejects many of Descartes’ claims. So we need to interpret Hume in line with empiricism, as saying that a priori knowledge (relations of ideas) is either analytic (and what can be deduced from analytic truths) or only about my own mind, while all knowledge of synthetic propositions about the world beyond my mind (matters of fact) is a posteriori.

### 3.1.6.1.2. Matters of fact

While the main focus of debate between rationalism and empiricism concerns a priori knowledge, it is worth briefly describing Hume’s theory of our a posteriori knowledge of matters of fact. The foundation of knowledge of matters of fact, Hume argues, is what we experience here and now, or can remember. We gain it by using observation and employing induction and reasoning about probability. All knowledge that goes beyond what is present to our senses or memory rests on causal inference. We take our experience to be an effect of whatever fact we infer. If I go out in the morning and all the streets are wet when they were dry yesterday evening, I’ll infer that it rained in the night. I do this because I think that rain causes the streets to become wet, and if the whole area is wet, not just small part, I’ll believe the cause is rain (rather than liquid spilling or some other explanation).

And how do I know all this? How do I know what causes what? Not by a priori reasoning or deduction. If you encounter some object that you’ve never experienced before, you cannot work out what effects it will have just by examining it. Just by examining a magnet – having never experienced one before – could you deduce what effect it will have on metal? Just by examining bread, could you work out that it doesn’t nourish tigers? Just by seeing a billiard ball roll towards another billiard ball, could you conclude that the second one will move away? Even if you imagine that this is what will happen, that’s arbitrary, groundless.

It is only our experience of what causes what that enables us to make causal inferences in particular cases. It is only our experience that enables us to infer from the existence of some cause to its effect, or from some effect to its cause. I have experienced rain wetting the streets around me and spillages wetting smaller areas. Reason can impose some order on the particular causal relations we discover through experience. For example, reason can simplify our causal principles, for instance by identifying different instances (the movements of billiard balls and the vibrations of molecules, say) as examples of the same kind of thing (kinetic energy). But reason can do no more than this.

# 3.2 Descartes’ Theory of Rational Intuition

Descartes’ *Meditations* provide an extended study in establishing knowledge through rational intuition and deduction. We will look at his ‘cogito’ and his elaboration of the idea of rational intuition through his concept of ‘clear and distinct’ ideas.

## 3.2.1. The cogito

At the start of Meditation II, we find Descartes supposing that all that he perceives and remembers is an illusion; that he has no body or senses at all; that in believing anything else, he is being deceived by a ‘supremely powerful and cunning deceiver’, an ‘evil demon’. How did he get into this state?!

Descartes is seeking to find out what he can know as true. To achieve this, he has decided to avoid believing anything that is not ‘completely certain and indubitable’. He then argues that he can doubt his senses, his memory and even that he has a body (note that these are all a posteriori claims we would use perception to establish). The demon could make it seem that he sees a tree when he doesn’t, that he has a body when he doesn’t, and so on. The question now, at the start of Meditation II, is whether he can know anything at all.

Descartes begins by arguing that, even if the evil demon is deceiving him about his senses and so on, ‘he will never bring it about that I am nothing while I think I am something’. Why not? Descartes cannot doubt that he exists: if he were to doubt that he exists, that would prove he does exist – as something that thinks (doubting is a kind of thinking). He cannot be deceived that he thinks. So he knows that he exists as something that thinks. The cogito, Latin for ‘I think’, is Descartes’ first stepping stone to knowledge.

However, Descartes can’t know that he exists as a body – his sense perception of his body, and of bodies in general, could be something he is deceived about. The demon could make it seem that he has a body when in fact he does not. Could he nevertheless be a body, without knowing it? Descartes can’t say, but at least his knowledge of what he is can’t depend on his being a body, since he knows he exists whether or not he has a body. What he is is a thinking thing, ‘a thing that doubts, understands, affirms, denies, wants, refuses, and also imagines and senses’. Furthermore, he knows which type of thought he is engaging in: he can’t mistakenly think that he is imagining when he’s conceiving, can’t think he’s doubting when he’s willing and so on.

The last activity of the mind that Descartes lists is ‘senses’. But doesn’t sense perception involve having a body? So doesn’t the fact that he senses establish the existence of physical objects? No, because, Descartes notes, he has sensory experiences in his dreams as well, when he is not seeing or hearing at all. ‘Sensing’ is just having sensory experiences. Understood like this, independent of their cause, these experiences are nothing more than a form of thinking, and so don’t depend on having a body.

## 3.2.2. Clear and distinct ideas

At the start of Meditation III, Descartes reflects on the cogito. He finds that his certainty in it rests on how the idea presents itself to his mind. So he argues

P1. ‘In this first item of knowledge there is simply a clear and distinct perception of what I am asserting.’

P2. If clarity and distinctness do not guarantee truth, then I cannot know that I exist.

P3. I do know that I exist.

C1. Therefore, ‘as a general rule . . . whatever I perceive very clearly and distinctly is true’.

This argument lays the foundations for Descartes’ theory of rational intuition. Descartes has defended the cogito as a claim that he knows to be true just by thinking about it. He knows because it is an idea that is ‘clear and distinct’.

What does this mean? Descartes doesn’t say in the *Meditations*, but gives this definition in his *Principles of Philosophy*: an idea is clear ‘when it is present and accessible to the attentive mind – just as we say that we see something clearly when it is present to the eye’s gaze and stimulates it with a sufficient degree of strength and accessibility’. An idea is distinct if it is clear and ‘it is so sharply separated from all other ideas that every part of it is clear’. In the Meditations, again drawing on an analogy with vision, Descartes connects clear and distinct ideas to what he calls ‘the natural light’: ‘Things that are revealed by the natural light – for example, that if I am doubting then I exist – are not open to any doubt, because no other faculty that might show them to be false could be as trustworthy as the natural light’. So, for Descartes, rational intuition is the ‘natural light’, our ability to know that clear and distinct ideas are true.

In what sense are clear and distinct ideas ‘indubitable’? Just saying ‘I can’t doubt it, so it must be true’ is clearly not good enough. The fact that you can’t doubt something may just be a psychological fact about you (cp. ‘I’m sure he told the truth. I can’t believe he would lie to me’ – and yet he did…). Things that we cannot doubt in this sense are not yet a good guide to the truth.

But this subjective sense of ‘indubitable’, a feeling of certainty, is not what Descartes means. He means that when I, as a rational thinker, using my best, most careful judgment, consider a proposition, I judge that it is impossible that it should be false. It is necessarily true that when I think of the proposition, it is true. When I think the thought ‘I think’, then that thought, ‘I think’, must be true. The indubitability of the proposition is an epistemological fact about the proposition, not a psychological fact about me.

# 3.2.3. An Empiricist Response to the Cogito

What does it mean to say ‘I exist’ or ‘I think’? Descartes claims that he is a thinking thing. He is the same thing from one thought to another. But can Descartes know this? The evil demon may deceive him: perhaps there is only a succession of thoughts, nothing that persists between thoughts which is a single thing.

In his *Treatise of Human Nature*, Hume develops the argument as follows: we don’t experience a continuing mental substance over time, we only experience a continually changing array of thoughts and feelings. So what is the basis for thinking that there is a thing that thinks? In coming up with the idea of a ‘thinking thing’ – a mental substance – we confuse similarity for identity. We’ve confused our experience of the similarity of our thoughts and feelings from one moment to the next with the idea that there is one identical ‘thing’ persisting through such thoughts and to which they belong.

Descartes’ response to this objection, in an appendix to the *Meditations* called ‘Objections and Replies’, is to say that thoughts logically require a thinker. But is this something Descartes could be deceived about?

Perhaps it is true that there can’t be a thought unless something thinks it. But that doesn’t entail that the ‘thinker’ is a subject that persists from one thought to another. Hume argues that even if we experience thinking as active in this way, how does our experience enable us to move to the claim that I am one and the same active substance, persisting through time and different thoughts? As soon as Descartes says that to be a thinker is to doubt, will, imagine, and so on, he assumes we can say these activities belong to the same subject, that he (the same thinker) does all this. But perhaps the evil demon is simply creating a series of false thoughts, among which is the thought that a thinker, a substance, an ‘I’, exists. Descartes’ claims about what he is could be false.

Is the cogito an example of knowledge by rational intuition? Descartes will argue that it is, because it is a clear and distinct idea. But first, we have just questioned whether it is clear and distinct that I am a mental substance. Second, Hume argues that we can know immediately about our minds through impressions of reflection. This will be a priori knowledge and intuitively certain, but impressions of reflection don’t provide us with knowledge of our existence as a mental substance.

**3.3. Innate Knowledge**

The claim that there is at least some innate knowledge is sometimes called ‘innatism’. Exactly what ‘innate’ means in this context is disputed. But the claim is that some knowledge is part of the mind, already ‘in’ the mind from birth, rather than gained from experience. If there is any innate knowledge, it cannot be a posteriori, but must be a priori. If we want to say that ‘reason’ is the source of this knowledge, then we can say that the knowledge is built into the ‘faculty’ of reason, that part of the mind with which we think about and understand the world.

The debate over innatism is whether there is any innate *propositional* knowledge. Everyone can agree that there is innate ability knowledge. Of course babies are born knowing how to breath, how to see (and apparently, how to hold their breath under water!). We can also agree that they have certain psychological abilities, such as memory and the disposition to learn a language. But is there any innate propositional knowledge?

In this section, we look at three arguments for innatism – two historical and one contemporary – and consider how an empiricist, who denies innatism, might reply.

# 3.3.1. Plato’s Slave Boy Argument

Plato’s dialogue *Meno* is mostly about virtue. But it includes an extended example and discussion of innate knowledge. Our interest begins with Socrates saying ‘You argue that man cannot enquire either about that which he knows, or about that which he does not know; for if he knows, he has no need to enquire; and if not, he cannot; for he does not know the very subject about which he is to enquire’. This is ‘Meno’s Paradox’. Put another way, it says that it is impossible to learn anything because, for anything you might learn, either you already know about it or you don’t know about it. If you already know about it, learning is unnecessary; if you don’t know about it, you won’t know how to go about learning it.

Plato’s solution to this puzzle is to say that learning is a form of remembering. He demonstrates this by Socrates asking Meno’s slave boy a series of questions about a theorem in geometry.

Socrates draws a square in the ground that is 2 feet × 2 feet. Its total area is therefore 4 square feet. How long are the sides of a square with a total area of 8 square feet? The slave boy has not been taught geometry, and yet is able to work out the right answer in response to Socrates only asking questions. The boy first guesses that the sides will each be 4 feet long, but when asked what 4 feet × 4 feet is, he realises that the area of this square is 16 square feet, not 8 square feet. The answer must be between 2 feet and 4 feet – he guesses 3 feet. But again, when asked what 3 feet × 3 feet is, he realises this square would be 9 square feet, not 8 square feet.

Socrates then draws three more squares of 2 feet × 2 feet, arranging them with touching sides to make one big square of 4 feet × 4 feet. He then draws a diagonal line across each small square, dividing them into triangles. The four diagonals are arranged to form a (square) diamond in the middle of the big square. Through questioning, he gets the slave boy to agree that each triangle is half of 4 square feet, i.e. 2 square feet. There are four such triangles making up the diamond, which is therefore 8 square feet. The sides of the diamond are the diagonals of the original 2 foot × 2 foot squares. So a square with an area of 8 square feet has sides the length of the diagonal of a square that is 4 square feet.

The boy wasn’t taught any geometry, yet he correctly answers each stage of the proof (or realises his mistake). How? He didn’t gain the knowledge from experience, so he must have recovered the answers from within his mind, i.e. the knowledge must be innate. The argument for innate knowledge is that we have knowledge that we can’t have gained from experience. Plato’s example is supposed to show that all we need to recover our innate knowledge is the right ‘prompts’ from experience (in this case, Socrates’ questions).

(Socrates goes on to argue that the mind must exist from before birth, to have gained this knowledge in a previous form of existence. Socrates’ questions triggered the knowledge he had from before birth, but had forgotten – just as memories can be triggered by some event or question. However, we don’t have to draw this conclusion about the pre-existence of the mind. Other explanations of innate knowledge are possible, as we see in the following pages.)

# 3.3.2. Leibniz on Knowledge of Necessary Truths

In his *New Essays on Human Understanding*, Leibniz argues that knowledge of necessary truths is not derived from experience. Experience only teaches us how things are on any occasion; it cannot teach us how things must be. Surely the world as we experience it could always have been a different way – so all propositions about the world could have been true or false, i.e. they are contingently true. But it is hard to see how necessary truths could be established a posteriori. Take the claim that ‘2 + 2 = 4’ or ‘Squares have four sides’. Leibniz points out that our sense experience only provides us with information about particular instances – that these two apples and these two apples make four apples; that this square has four sides; and so on. But ‘however many instances confirm a general truth, they aren’t enough to establish its universal necessity’. Our experience tells us how things are, but not how things must be. If we reject this, and argue that ‘2 + 2 = 4’ is just a generalisation of our experience so far, then we are saying that it is possible, one day, that 2 + 2 will equal some other number. But this is inconceivable.

In sum, all necessary truths tell us how things must be. Because experience doesn’t tell us how things must be, it seems that all knowledge of necessary truths must be a priori.

Leibniz then argues that we should regard such a priori knowledge of necessary truths as innate. We discover their truth in a priori reasoning by ‘attending carefully and methodically to what is already in our minds’. This is what Plato’s example of the slave boy shows. (In fact, in a broad sense of ‘innate’, all the knowledge we gain by a priori reasoning from ‘basic’ innate knowledge can also be called innate.) An example of such innate, a priori knowledge is ‘It is impossible for the same thing to be and not to be’. We can know the general truth of this claim by reflecting on it; but sense experience can’t teach us this, for the reasons already give above.

Importantly, saying that this knowledge is innate doesn’t mean that we can discover our innate knowledge without any sense experience. We need sense experience in order to form abstract thoughts; we rely on words, letters, sounds, which we learn from experience. That makes sense experience necessary but not sufficient for our knowledge of necessary truths. If sense experience isn’t sufficient, then the knowledge must already be part of our minds.

# 3.3.3. Empiricist Response: Experience Triggers Innate Knowledge

Philosophers who defend innate knowledge argue for it as knowledge *which cannot be gained from experience*, e.g. geometry (Plato) and other necessary truths (Leibniz). Since we are not consciously aware of this knowledge from birth, there is some point at which we first come to be aware of it. And so innatists argue that experience *enables our awareness* of the knowledge.

How is experience ‘enabling’ knowledge different from simple *learning* from experience? We have already said that, with innate knowledge, experience is necessary but not sufficient. But can we say more than this? One modern version of the theory talks of experience ‘triggering’ knowledge.

The idea of triggering is often used in the study of animal behaviour. For example, in some species of bird, a baby bird need only hear *a little bit* of the bird song of its species before being able to sing the *whole* song itself. There has been far too little experience of hearing the song sung by other birds for the baby bird to learn from experience; rather the experience has triggered its innately given song.

In *Human Knowledge and Human Nature*, Peter Carruthers notes that there are many developments in our cognitive *capacities* that are genetically determined. For example, infants cannot see further than approximately 12 inches when first born. Within 8 weeks, they can see much further. This development of the eye is genetically encoded. The same could be true for certain types of *knowledge*. At a certain genetically determined point in development, children begin to think in a particular way for the first time, but that way of thinking has not been learned from experience. For example, around 3–4 months, babies *quickly* shift from thinking of objects as existing only while they experience them to thinking of objects as something that can exist outside their experience. So, for example, they begin looking for things they have dropped. Or again, babies *very quickly* relate to other people as having minds – beliefs, desires, intentions, emotions, etc. In both cases, they couldn’t have learned this knowledge (that objects exist independent of experience, that other people have minds) from experience. So the knowledge is innate.

Again, this is not to say that experience has no role. A child must be exposed to the relevant stimuli – interactions with objects and people – for the knowledge to emerge. What shows that the knowledge is innate is that it cannot be learned from experience.

The claim is not simply that we have the *capacity* to gain this knowledge. Rather, the claim is that our capacities are ‘preshaped’ or ‘predisposed’ towards thinking truly about the world in some ways rather than others. So experience merely triggers our knowledge, rather than being the source of the knowledge.

# 3.3.4. Alternative Empiricist Accounts

Many recent philosophers, including Carruthers, have argued that innate knowledge is compatible with spirit of empiricism and the claim that *ultimately* all our knowledge derives from sense experience. We can provide an empirical explanation of innate knowledge in terms of evolution. Knowledge is innate in the sense of it being encoded genetically that we will develop and use the knowledge at a certain point in cognitive development under certain conditions. Evolution has prepared our minds to form an understanding of the world in terms of mind-independent physical objects and the existence of other minds with beliefs and desires. We can argue that these beliefs constitute knowledge because they are reliable.

It is worth noting, though, that claims about physical objects and other minds are contingently true. What can an empiricist say about the kinds of necessary truths Leibniz discusses, such as ‘2 + 2 = 4’ and ‘It is impossible for the same thing to be and not to be’? How could evolution give us knowledge of necessary truths if Leibniz is right that necessary truths cannot be established through experience?

Here an empiricist can provide an alternative account of how we know them. We don’t know necessary truths innately; instead, necessary truths are *analytic*. We acquire the concepts involved from experience, and then in understanding the concept, we come to know the necessary (analytic) truths. The knowledge is conceptual, not innate.

This alternative explanation will only be successful on two conditions. First, the empiricist has to show that the relevant concepts are acquired from experience. If the concepts are innate, then the knowledge will count as innate as well. Second, the empiricist must show that necessary truths are, in fact, analytic.

**3.4. Locke’s Arguments Against Innate Knowledge**

One issue that the question raises is whether we can really make sense of the idea of innate knowledge. What are we really saying when we say that knowledge is ‘innate’? If we get clear on that, does any knowledge qualify? This is the line of argument with which John Locke begins his attack on innatism in *An Essay concerning Human Understanding*.

Locke argues that we have no innate knowledge. He begins by asking how we acquire our ideas. By ‘idea’, he means ‘whatever it is that the mind can be employed about in thinking’. Or again, an ‘idea’ is any ‘immediate object of perception, thought, or understanding’. So he uses the word to cover a very wide range of mental phenomena. An idea can be

1. a complete thought, taking the form of a proposition, e.g. ‘bananas are yellow’;

2. a sensation or sensory experience, e.g. a visual sensation of yellow; or

3. a concept, e.g. ‘yellow’.

Our focus here is on propositions, as these are what we can know or not know.

Locke understands innate ideas as ‘thoughts printed on to the soul at the point of existence, which it brings into the world with it’. As examples of potential innate knowledge, taken from the debate at the time, he offers ‘Whatever is, is’ and ‘It is impossible for the same thing to be and not to be’. He assumes that innate knowledge must be universal – every human being has it (§3). However, the converse is not true: just because some claim is universally accepted, that doesn’t mean it is innate – it may be that we could explain in some other way why everyone agrees.

Locke then argues:

P1. If there is innate knowledge, it is universal.

P2. For an idea to be part of the mind, the mind (the person) must know or be conscious of it: ‘it seems to me nearly a contradiction to say that there are truths imprinted on the soul that it doesn’t perceive or understand. No proposition can be said to be in the mind which it has never known or been conscious of’.

C1. Therefore, innate knowledge is knowledge that every human being is or has been conscious of.

P3. Children and ‘idiots’ (people with severe learning disabilities) do not know theorems in geometry or ‘It is impossible for the same thing to be and not to be’. (They do not know these claims, because they do not understand them.)

C2. Therefore, these claims are not innate.

P4. There are *no* claims that are universally accepted, including by children and ‘idiots’.

C3. Therefore, there is no innate knowledge.

We can undermine Locke’s argument if we can reject C1. Is there something wrong with Locke’s conception of innate knowledge? After all, innatists such as Plato and Leibniz don’t talk of innate knowledge as *conscious*. But if innate knowledge isn’t conscious as Locke says, then what can it be? Locke anticipates and objects to four alternative definitions of innate knowledge.

If we define as ‘innate’ *any* knowledge that we can gain, Locke objects that this is a misuse of the term – everything we come to know, including through sense experience, will be innate! What we should say is that the *capacity* for knowledge is innate. This is true – we are born with the ability to know things – but it doesn’t mean that there is innate *knowledge*. Compare: the capacity to see (vision) is innate, but that doesn’t mean that *what* we see is innate as well!

What if we define innate knowledge as what everyone knows and agrees to when they gain the use of reason? After all, both Plato and Leibniz emphasise the role of reason in innate knowledge. But, Locke presses, why think that what we can discover by reasoning is *innate*? If the knowledge is innate, and so we already have it, why do we need to ‘discover’ it? Anyway, even if we grant the definition, there is still no innate knowledge, because children can reason *before* they understand mathematical and logical truths.

To take account of this, we could say that innate knowledge is gained at some point after the use of reason. This is hopeless – it doesn’t mark off innate knowledge from all kinds of other knowledge, including what we learn from sense experience.

Finally, what about defining innate knowledge as truths that are assented to promptly as soon as they are understood? Innate knowledge is ‘self-evident’. But there are many such claims that rely on sense experience, e.g. ‘white is not black’. So they can’t be innate.

Locke concludes that there is no satisfactory definition of ‘innate’ that can be used to defend the claim that there is innate knowledge.

# 3.4.1. Leibniz’s Response to Locke

Leibniz wrote his *New Essays on Human Understanding* as a commentary on and response to Locke. Leibniz argues that Locke has not understood the sense in which knowledge can be innate. Locke’s theory that ideas must be conscious has misled him. We can know things without being conscious of them. Locke is wrong to claim (P2) that an idea can only be in the mind if we are conscious of it. Innate knowledge exists as ‘a disposition, an aptitude, a preformation’ in the mind towards developing, understanding and knowing certain thoughts. In other words, according to Leibniz, *none* of Locke’s definitions of ‘innate’ are quite right. We have innate knowledge in a sense not envisaged by Locke.

## 3.4.1.1. Unconscious knowledge

Leibniz picks up the example of ‘It is impossible for the same thing to be and not to be’, and rejects Locke’s claim that this is not universally accepted. Everyone uses this knowledge all the time, but ‘without explicitly attending to it’. Indeed, we can’t really think without it, since it is needed to distinguish the concept of one thing from the concept of something different. ‘General principles [such as the example given] enter into our thoughts, serving as their inner core and as their mortar. Even if we give no thought to them, they are necessary for thought. The mind relies on these principles constantly’.

We can see this with Locke’s example of ‘white is not black’. Leibniz accepts that claims like ‘white is not black’ aren’t innate. But they are applications of a necessary truth that *is* innate, namely ‘It is impossible for the same thing to be and not be at the same time’, to particular cases and concepts acquired from sense experience. Locke might object that the particular cases, such as ‘white is not black’, are known *before* the abstract principle. Leibniz responds that in the particular cases, we *unconsciously deploy* our knowledge of the abstract principle that something can’t both be and not be at the same time.

Leibniz’s theory entails that knowledge can be unconscious. But this shouldn’t be controversial. Memory ‘stores’ ideas and usually, but not always, retrieves them when we need them. This shows two things: we can know things without being conscious of them; and retrieving this knowledge can need assistance. So even Locke, who says that an idea can only be part of the mind if it is something the person can be conscious of, must accept that there is nothing impossible about unconscious knowledge.

Locke can reply that this is true, but irrelevant to the question of innate knowledge, because in memory, we are recalling what has been conscious. But, says Leibniz, why accept that what is unconscious must always have once been conscious or gained from experience? Why think that we can know everything about our minds straightaway?

## 3.4.1.2. Innate knowledge as a disposition

Leibniz comments on Locke’s contrast between ‘innate knowledge’ as knowledge we can acquire and the innate capacity for knowledge. The contrast restricts the options. While innate knowledge does not exist ‘fully formed’ or explicitly in our minds, it is more than mere capacity. In gaining knowledge of necessary truths, the mind needs to actively engage with itself, albeit at the prompting of sense experience. Thus, Leibniz says, ‘The actual knowledge of [necessary truths] isn’t innate. What is innate is what might be called the potential knowledge of them, as the veins of the marble outline a shape that is in the marble before they are uncovered by the sculptor’. It takes work to uncover what is within us, but what we uncover, we have not learned from sense experience.

**3.5. Locke’s Arguments Against Innate Concepts**

The claim that there are innate concepts means that not all concepts are learned from experience; some concepts are somehow part of the structure of the mind.

If some propositional knowledge is innate, then some concepts must be innate, because propositional knowledge is formulated in terms of concepts. Conversely, if we can show that there are no innate concepts, we will have shown that there is no innate knowledge. In *An Essay concerning Human Understanding*, John Locke provides this very argument against innate knowledge. To have innate knowledge requires that one has the concepts involved in the proposition one knows. If we first had to acquire the concepts, then the knowledge can’t be innate. But there are no innate concepts.

# 3.5.1. Locke’s Objection to Innate Concepts

It is an important part of Locke’s argument that whatever concepts we have, we are conscious of. Furthermore, he assumes (and everyone in the debate agrees) that innate concepts must be universal – every human being has them. If we put these two thoughts together, an innate concept must be one that every human being is or has been conscious of.

Locke gives three main reasons for rejecting the existence of innate concepts, given his definition of what they are:

1. If we observe new-born babies, we have no reason at all to think that they have any concepts beyond, perhaps, ones deriving from their experience in the womb, such as WARMTH and PAIN. Certainly, we can’t think that such advanced concepts as IDENTITY or IMPOSSIBILITY are concepts babies are familiar with and conscious of. But these concepts are necessary for the knowledge that ‘It is impossible for the same thing to be, and not to be’ (an example of supposed innate knowledge from the debate at the time).

2. Another favourite of innatism at the time of Locke was the concept of GOD. But not only is this not a concept that babies have, it is not a concept that all human beings have – whole societies, historically, have been atheist. The concept of GOD is not innate, but learned by children from their teachers.

3. The only way a concept can be part of the mind without the mind being conscious of it is if it is lodged in memory. To remember something is to have been conscious of it in the past. If you aren’t remembering a concept, then it is new to your mind – arising from some impression of sensation or reflection. Innate ideas would have to be neither remembered nor new. How could there be such a thing?

# 3.5.2. Rejecting Locke’s Definition Of ‘Innate Concept’

Defenders of innate knowledge disagree with Locke’s definition of innate concepts. They reject his claim that it is impossible for concepts to exist ‘in the mind’ unless we are or have been conscious of them. Innate concepts are concepts which cannot be gained from experience, and arguments defending innatism try to show that experience cannot explain how we have or use the concept. Experience is necessary to trigger our development of the concept, but it is not sufficient to explain our having the concept.

The idea of experience ‘triggering’ the concept needs to be understood carefully. The claim is not that we simply have the capacity to form the concept. Rather, we are predisposed to form just this concept, which we cannot form on the basis of experience alone.

On this understanding of innate concepts, it is no objection that babies don’t have the relevant concept of GOD or IDENTITY – it needs to be triggered by experience before it develops.

# 3.5.3. Leibniz’s Defence of Innate Concepts

In his *New Essays on Human Understanding*, Leibniz accepts Locke’s claim that innate knowledge requires innate concepts. Therefore, if we want to say that ‘It is impossible for the same thing to be and not be’ is innate knowledge, we will have to say that concepts such as IDENTITY and IMPOSSIBILITY are innate. But, to answer Locke’s first objection, this means that we have, from birth, the disposition to form these concepts. Indeed, they are essential to all thought, even though it takes time for us to make them explicit in our thinking.

In answer to Locke’s second objection, Leibniz points out that to lack the word for God is not to lack the concept of GOD. Some societies have no word for ‘being’, but that doesn’t mean they don’t have thoughts that use the concept. Again, it may take considerable work of reflection to develop the concept of GOD. We are disposed, from our experience of nature, to develop the idea of a higher power. But this isn’t yet the full concept of GOD as we have it. Our experience enables a concept that goes beyond what we can learn from experience; our minds are ‘receptive’ to the idea of God.

In answer to Locke’s third objection, Leibniz claims that innate knowledge and concepts exist as dispositions in the mind. While innate concepts and knowledge do not exist ‘fully formed’ or explicitly in our minds, they are more than mere capacity certain concepts or knowledge. Thus, Leibniz says, ‘What is innate is what might be called the potential knowledge of them, as the veins of the marble outline a shape that is in the marble before they are uncovered by the sculptor’. It takes work to uncover what is within us, but what we uncover, we have not learned from sense experience.

**3.6. Locke: The Mind as *‘Tabula Rasa’***

In *An Essay concerning Human Understanding*, John Locke argues that at birth – or more accurately, since there can be consciousness and thought before birth, prior to any experience – the mind is a ‘tabula rasa’. ‘Tabula rasa’ is Latin for ‘blank slate’. The phrase recalls the time when children would have slates (or tablets (tabula)), like small blackboards, to write on. Until the teacher told them to write something, the slates would be blank. The mind at birth, says Locke, contains no ideas – no thoughts or concepts. If you observe new-born babies, says Locke, you’ll find no reason to disagree. All our ideas, then, derive from one of two sources:

1. Sensation: our experience of objects outside the mind, perceived through the senses. This gives us ideas of ‘sensible qualities’.
2. Reflection: our experience of ‘the internal operations of our minds’, gained through introspection or an awareness of what the mind is doing. This provides the ideas of perception, thinking, willing, and so on. These ideas may well arrive later in childhood.

# 3.6.1. Hume on Impressions and Ideas

Locke uses the term ‘idea’ to cover sensations and concepts (and propositional thoughts!). This is very confusing. The sensation of yellow isn’t the same thing as the concept YELLOW. When we see something yellow, this perceptual experience is quite different from the role YELLOW plays in the thought ‘If it is yellow, it is coloured’. David Hume’s terminology is a little clearer, though it still doesn’t quite match everyday meanings.

According to Hume in *An Enquiry concerning Human Understanding*, what we are immediately and directly aware of are ‘perceptions’. ‘Perceptions’ are divided into ‘impressions’ and ‘ideas’. Although he doesn’t say so explicitly here, Hume, following Locke, divides impressions into those of ‘sensation’ and those of ‘reflection’. Impressions of sensation derive from our senses, impressions of reflection derive from our experience of our mind, including emotions.

Hume distinguishes between impressions and ideas on three grounds. First, there is a difference between the two marked by a difference of ‘forcefulness’ and ‘vivacity’ or ‘liveliness’; impressions relate roughly to ‘feeling’ (or ‘sensing’) and ideas to ‘thinking’. Think what it is like to see a scene or hear a tune; now, what it is like to imagine or remember that scene or tune. The latter is weaker, fainter. (Thinking, for Hume, works with ideas as images in the same way as imagination and memory.) However, Hume immediately qualifies this claim – disease or madness can make ideas as lively and forceful as impressions. So, second, Hume argues that ideas are ‘copies’ of impressions. Hume later provides a third distinction between ideas and impressions: we are liable to confuse and make mistakes about ideas, but this is more difficult with impressions.

Just as there are impressions of sensation and reflection, so there are ideas of sensation (e.g. RED) and ideas of reflection (e.g. THINKING). What Hume means by ‘idea’ here, we can refer to as concepts. So his theory of how we acquire ideas is a theory of how we acquire concepts. His claim is that we copy them from impressions.

So Locke and Hume have slightly different versions of how we first acquire ideas with which we can think (concepts). We start with sense experiences of the physical world and experiences of our own minds; for Locke, this gives us ideas; but this makes it sound as if the experiences themselves are the ideas with which we think. Hume corrects this: it is copies of sensory impressions that we use in thinking.

Why think that all ideas derive from impressions? Hume gives two arguments. First, without having a particular type of experience, a person lacks the ability to form an idea of that experience. Thus, a blind man does not know what colour is and a mild man cannot comprehend the motive of revenge. We’ll return to this claim below. Hume’s second argument relates to ‘simple’ and ‘complex’ ideas.

# 3.6.2. Simple and Complex Concepts

Locke argues that the basic building blocks of all thought are simple ideas, or more precisely, in Hume’s terminology, simple impressions – single colours, single shapes, single smells and so on. For each, there is a corresponding simple idea (for clarity, I shall talk, from now on, about concepts). A simple impression or simple concept ‘contains nothing but one uniform appearance or conception in the mind, and is not distinguishable into different ideas’. Of course, we experience many such simple impressions at once, e.g. we hold a toy car that is at once both cold and hard. But there is no confusing the sensation of cold with the sensation of hardness – they are quite distinct.

As the building blocks of thought, simple concepts can be used to construct complex concepts.

1. We can unite or combine the impressions of the qualities we perceive into the concept of a single object – we identify one and the same thing, a dog, say, as having a particular colour, shape, smell. So we can think of ‘that thing’, where the concept of ‘that thing’ is made up of many concepts of colour, shape, smell.
2. We can also form complex concepts by abstraction, e.g. the concept DOG doesn’t correspond to any one particular dog. When we abstract, we ignore certain specific features and concentrate on others; so to develop the concept DOG, we ignore the different colours and sizes of dogs and pick out features they have in common, such as four legs, tail, bark, hairy.
3. We can put together simple concepts in an original way. While many of us have seen a picture of a unicorn, someone had to invent the concept without seeing a picture. They did it by putting together concepts of HORSE and HORN and WHITENESS.

Hume agrees with Locke’s claim that all concepts are either simple concepts or complex concepts that have been constructed out of simple concepts. He claims, like Locke, that all concepts can be analysed into simple concepts which each correspond to an impression. (This is his second argument for thinking that all concepts derive from impressions.) Therefore, all concepts ultimately derive from experience.

For example, in direct opposition to Descartes, Hume claims that the concept GOD, based on concepts of PERFECTION and INFINITY, is extrapolated from concepts of IMPERFECTION and FINITUDE: ‘The idea of God – meaning an infinitely intelligent, wise, and good Being – comes from extending beyond all limits the qualities of goodness and wisdom that we find in our own minds’.

# 3.6.3. Issues with the Empiricist Theory of Concepts

Hume and Locke argue that no concept, no matter how abstract or complex, is more than a putting together, altering, or abstracting from simple concepts, and that all simple concepts derive from impressions. We can show that this theory is false if we can find a counterexample, either a concept that does not derive from an impression or a complex concept that cannot be analysed into simple concepts. We start with a possible example of a concept that does not derive from an impression.

## 3.6.3.1. Challenging the copy principle

Is it true that without a specific experience, we can’t form the relevant concept? Hume notes that there is an exception to this claim. If you present someone with a spectrum of shades of blue with one shade missing, then using their imagination, they will be able to form an idea of that shade. This idea has not been copied from an impression. Hume dismisses the example as unimportant, but it is not. If it is possible that we can form an idea of a shade of blue without deriving it from an impression, is it possible that we could form other ideas without copying them from impressions?

The question is important because Hume uses his ‘copy principle’ repeatedly in his philosophy. For example, he says that in metaphysics, we become confused because the ideas we work with, e.g. SUBSTANCE, are ‘faint and obscure’, so we don’t understand them well. Because ideas derive from impressions, we can solve metaphysical debates by asking, of the words used, ‘From what impression is that supposed idea derived?’ If we can’t find the associated impression, we can conclude that the word is used without a proper meaning, and reject the debate.

However, if we can form ideas without copying them from impressions, then we can’t use Hume’s copy principle to cut through metaphysical debates as he suggests. So can the copy principle be defended against the counterexample of the missing shade of blue?

There are two possible solutions that allow for the case of the shade of blue while maintaining a strong link between ideas and impressions.

We can weaken the copy principle to say ‘Any ideas that are not (ultimately) copied from impressions are only meaningful if they could be copied from impressions’. In other words, what the idea is an idea of is something we can encounter in experience. The missing shade of blue clearly meets this condition, but perhaps many metaphysical ideas will not.

We can keep the copy principle as it is – ‘all ideas are (ultimately) copied from impressions’ – but explain how and why the missing shade of blue is an ‘exception’. The simple impressions of different shades of blue are related to each other, as they can be arranged according to how they resemble each other (from dark to light, say). From the arrangement, we can form the idea of the missing shade drawing on other similar impressions we already have. This only works when impressions are structured by resemblance like this. If we have no relevantly similar impressions which strongly resemble the missing impression, we cannot form the missing idea. This is the same reason that a blind man cannot form an idea of colour, and so it fits well with Hume’s theory.

## 3.6.3.2. Leibniz on ‘intellectual ideas’

In his *New Essays on Human Understanding*, Leibniz gives a number of examples of concepts that he claims are not derived from experience as Locke and Hume claim, but are innate. He comments on Locke’s division of concepts into those that originate in sensation and those that originate in reflection, which Leibniz calls ‘intellectual ideas’. He comments, ‘to reflect is simply to attend to what is within us, and something that we carry with us already is not something that came from the senses! So it can’t be denied that there is a great deal that is innate in our minds’. Thus, he says the concepts of BEING, UNITY, SUBSTANCE, DURATION, CHANGE, ACTION, PERCEPTION and PLEASURE are all innate, because we are ourselves beings, unities, substances, that we endure through time, change, act, perceive and experience pleasure. In fact, all the concepts we acquire through reflection can be called ‘innate’.

Locke can rightly respond that reflection upon what I am does not establish innate concepts. My existence and my ability to perceive are innate, but that doesn’t mean that the concepts of SUBSTANCE and PERCEPTION are innate. Locke argues that we must first experience our own mind and its activities (in reflection) to develop the concepts – hence they are not innate. It is a confusion to argue that because we derive the concepts from our mental activities that we do not therefore derive them from experience.

## 3.6.3.3. The concept of substance

The response we gave on Locke’s behalf is too quick. Locke allows that I am a substance, and of course my existence is innate, but this doesn’t mean that the concept SUBSTANCE is also innate. His position is that we gain the concept of substance from our experience – perhaps our experience of ourselves in reflection. A substance is something that continues to exist as one and the same thing through time, that possesses properties which can change even as it remains the same thing. But, even if we are substances, do we experience ourselves in reflection *as a substance*? To defend his claim that we acquire the concept SUBSTANCE from experience, Locke will need to show that we do. We have two particular concepts of substance, namely PHYSICAL SUBSTANCE (physical objects) and MENTAL SUBSTANCE (minds or selves). Do these concepts come from experience?

### 3.6.3.3.1. Berkeley on substance

In *Three Dialogues between Hylas and Philonous*, George Berkeley, who was an empiricist, argues that the concept of MENTAL SUBSTANCE or MIND can be derived from our experience of ourselves, but that the concept of mind-independent PHYSICAL SUBSTANCE is incoherent. We start with the first. He claims that I am not only aware of my mental activities, but aware of my mind as that which is active in thinking, perceiving and willing. So I am aware that my mind is not reducible to the activities themselves. So, Berkeley argues, we can derive the concept of MENTAL SUBSTANCE from our own minds, but the concept is not innate, as it is derived from our experience of ourselves.

However, Berkeley argues that we can gain no idea of PHYSICAL SUBSTANCE from sense experience. We do not experience physical substance, only the primary and secondary qualities of physical objects, and both are mind-dependent. That anything exists beyond these changeable properties is not an idea that sense experience supports. But rather than conclude that the concept is innate, we should conclude that it is confused.

Berkeley’s arguments illustrate the two ways empiricists can respond to proposed counterexamples to their theory of acquiring concepts. First, they can argue that the concept is, in fact, derived from experience. Second, they can argue that the concept is incoherent, the result of some kind of mental error. This would explain its origin as neither derived from experience nor innate.

### 3.6.3.3.2. Hume on substance

In *A Treatise of Human Nature*, Hume develops this last objection further, and adds a further argument to those of Berkeley against the concept of PHYSICAL SUBSTANCE.

The concept of a PHYSICAL SUBSTANCE is the concept of something independent of experience existing in three-dimensional space. But how can experience show us that something exists independently of experience? I see my desk; a few moments later, I see it again. If my two experiences are of one and the same desk, then the desk existed when I wasn’t looking at it. But I don’t experience the desk existing when I’m not looking at it. So how do I arrive at the idea that it is one and the same desk, which has persisted through time even when I wasn’t experiencing it?

Hume’s diagnosis is this. My experience only provides the information that my two experiences of the desk are very similar. The desk as I first experience it is very similar, perhaps exactly similar, to the desk as I experience it the second time. But similarity, even exact similarity, is not quantitative identity. Being qualitatively identical is not the same as being numerically identical. (For example, two people can sit comfortably on identical chairs, but they can’t sit comfortably on one and the same chair.) My sense experience can only provide the concept of a physical object that is numerically identical (with itself) while I am experiencing it.

Hume applies the same argument to the concept of MENTAL SUBSTANCE.

He disagrees with Berkeley (and Descartes): we don’t experience a continuing mental substance (self) over time, we only experience a continually changing array of thoughts and feelings. Even if we experienced thought as active, as Berkeley maintains, this experience doesn’t support the claim that I am one and the same active substance, persisting through time and different thoughts.

So far, Hume has argued that we cannot derive the concepts of MENTAL or PHYSICAL SUBSTANCE from our experience. If he is right, then we could argue that both concepts must be innate. After all, we do have the concept of SUBSTANCE as something that persists through change, and we have the concepts of PHYSICAL SUBSTANCE and MENTAL SUBSTANCE. If we can’t learn them from experience, they must be innate.

But Hume takes his argument to show that both concepts of SUBSTANCE are confused rather than innate. In coming up with the concept of a PHYSICAL SUBSTANCE that exists independently of my experiences, I have confused similarity with identity. How does this happen? Our perceptions of physical objects exhibit constancy: if I look at my desk and then shut my eyes and open them again, the desk looks exactly as it did before. On the basis of this similarity, the mind simply has a tendency to imagine that what I see after I opened my eyes is not just similar but identical to what I saw before I closed my eyes. The origin of the idea that the two experiences are of something identical – something that exists between and independent of perceptions – is the imagination. The imagination creates the idea of identity from similarity and unity (the idea of an individual thing, being ‘one’), both of which we can derive from experience. But there is nothing in experience that matches the concept of PHYSICAL SUBSTANCE.

A similar story applies in the case of MENTAL SUBSTANCE. We’ve confused the similarity of our thoughts and feelings from one moment to the next with the identity of a ‘thing’ to which such mental states belong. The concept is not innate, it is confused.

We can object that Hume’s theory makes our common-sense idea of the world wrong. If we are to avoid scepticism, we must either find a way to derive these concepts from experience or accept that they are innate.

# 3.6.4. Final Thoughts on Innate Concepts

We began our discussion of the empiricist theory of concepts by saying that we could show the theory to be false if we can find a counterexample – a simple concept that is either not copied from an impression or a complex concept that we cannot analyse into simple concepts that are copied from impressions. If the concept of SUBSTANCE isn’t a genuine counterexample to empiricism, then perhaps some other concept will be. For instance, we might argue that attempts to analyse philosophical concepts like KNOWLEDGE, TRUTH and BEAUTY into their simple constituents have all failed to produce agreement. A good explanation for this is that they don’t have this structure, and Locke and Hume’s theory of the origin of concepts is wrong.

3.7. Descartes’ Trademark Argument

In the Trademark argument, Descartes tries to prove the existence of God from just the idea of GOD as a being that is, among other things, supremely powerful and supremely perfect. He argues that the concept of GOD is innate, like a ‘trademark’ that our creator has stamped on our minds.

Descartes begins by identifying three possible sources of any idea:

1. the idea derives from something outside my mind, such as I experience in sense perception;
2. I have invented it;
3. it is innate. (Descartes explains this as ‘it derives from my own nature’, but he also uses the usual innatist argument that it can’t be learned from experience (or invention).)

We cannot in general be certain which of the three types of cause an idea has. Which is the source of the concept GOD?

Before answering that question, Descartes embarks on a long defence of the claim that a cause must have at least as much ‘reality’ as its effect, and that the cause of an idea must have as much reality as what the idea is an idea *of*. Both the claim and the argument are very puzzling, so we set them aside to explore further below.

For now, here is a common-sense example from Bernard Williams: if we discover a picture of a sophisticated machine, even though it’s just a *picture*, we think it must be the product of an advanced society or a highly fertile imagination. It is what it is a picture *of* that makes us think the cause is sophisticated. Where could the ‘sophistication’ of the machine in the picture come from except a mind that is itself just as sophisticated? The cause must have as much ‘reality’ as the machine in the picture.

With this in place, Descartes argues:

P1. I have the concept GOD.

P2. The concept GOD is a concept of something infinite and perfect.

P3. As a mind, a thinking substance, I can think up (create) many ideas, including ideas of people and physical objects.

P4. But I am finite, while the concept GOD is a concept of something infinite.

C1. Therefore, it is a concept of something with more reality than my own mind.

P5. The cause of the concept GOD must have as much reality as what the concept is of.

C2. Therefore, my mind could not have created it.

P6. The only possible cause is God.

C3. Therefore, God exists.

Descartes considers and rejects an objection to (P4), namely that I have all the perfections I attribute to God, and so could invent the concept. But given that I am in doubt, I clearly do not have infinite knowledge – I am not infinite, but finite.

How could we have acquired the concept GOD? Descartes has argued that he couldn’t have invented it. He then adds that it does not derive from sense experience (it isn’t something that arises ‘unexpectedly’ as do other ideas of sense). So by elimination, the concept GOD must be innate, built into the structure of our minds by God.

# 3.7.1. Degrees of Reality

Descartes’ argument rests on claims (C1) and (P5), but these are difficult and highly controversial. The idea of ‘degrees of reality’ is strange to us, but was a standard part of medieval metaphysics.

1. A ‘substance’ is defined as something that can exist independently, such as the mind, God and physical objects.

2. An ‘attribute’ is a property of a substance – the attribute of mind is thought, while extension (having spatial dimensions) is an attribute of physical objects.

3. A ‘mode’ is a particular determination of a property. So ideas are modes of the mind – specific ways of thinking. Being specific sizes or shapes are modes of physical objects.

A substance has more reality than an attribute, because a property cannot exist without a substance, and so is dependent on it. There can be no thoughts without a thinker. Modes, therefore, also have less reality than substances. Ideas are modes of the attribute ‘thought’, which is possessed by thinking substances.

Descartes applies these thoughts to cause and effect. He simply takes it to be a clear and distinct idea that the cause of something must contain at least as much reality as its effect. From this, he derives the claim that something can’t come from nothing. But in fact, it is easier for us to understand this the other way around – something can’t come from nothing, and so whatever is part of the effect must have originated in the cause. For instance, a stone can only be created by or from something that contains the qualities of the stone (what is needed to make a stone). Or again, something hot can’t derive its heat from something cold.

Ideas are more complicated. As modes of thought, the ‘intrinsic reality’ of all ideas is the same, and less than the reality of my mind, which is a substance. But ideas also represent something, e.g. an object, a size, a tune, a mind, God. Some of these things – object, mind, God – are substances; others – a size, a tune – are modes. The degree of reality of the thing that the thought is about determines the idea’s ‘representative reality’. Just as we need to able to explain where the heat in something hot comes from, so we need to be able to explain the representative reality of an idea. Just as heat comes from something hot, so an idea with a certain representative reality must come from something with at least as much intrinsic reality. So ideas of substance can only be caused by substances; ideas of modes can be caused by either modes or substances.

We can now apply this to the concept GOD. As a concept, it is a mode of thought, and so it seems my mind – a substance – could cause it, just as my mind causes many other ideas. But the special features of what GOD is a concept *of*, namely something infinite and perfect, mean that it has a representative reality *greater* than the intrinsic reality of my mind. If I invented the concept, GOD would contain things – infinity and perfection – that are not in its cause, because I am imperfect and finite. But this is impossible – there must be as much reality in the cause as in the effect. So only God, being perfect and infinite, could create a concept of something perfect and infinite.

# 3.7.2. Empiricist Responses to the Trademark Argument

Empiricists will reject Descartes’ Trademark argument as it claims both that the idea of GOD is innate, and that we can prove the existence of God using rational intuition and deduction. We noted above that claims (C1) and (P5) are particularly problematic, and it is these claims that empiricist objections focus upon.

## 3.7.2.1. Knowledge of causes

Descartes assumes that all ideas have a cause. But Hume argues in *A Treatise of Human Nature*, that this is not something we can know.

The claims ‘everything has a cause’ and ‘something cannot come out of nothing’ are not analytically true. ‘Some things do not have a cause’ is a not a *contradiction in terms* like ‘Some bachelors are married’ is. Of course, from our experience, we have good reason to think that everything has a cause, but this is still only a contingent truth; it may be false. We cannot show that it holds without exception.

Second, if we can’t know that it is impossible for something to come out of nothing, then we can’t know, either, that a cause must contain at least as much ‘reality’ as its effect. What causes what is something we must discover from experience; we cannot know it by a priori reason.

## 3.7.2.2. Is the concept of GOD innate?

David Hume rejects the claim that the concept GOD cannot be created by our minds. We can form this concept by starting from ideas of finitude: ‘The idea of God – meaning an infinitely intelligent, wise, and good Being – comes from extending beyond all limits the qualities of goodness and wisdom that we find in our own minds.’ In ‘extending beyond all limits’ the ideas of finite goodness and wisdom we have from experience, we create an abstract negation of what is finite. Thus, we create ideas of what is NOT-FINITE (INFINITE) and NOT-IMPERFECT (PERFECT).

However, Descartes considers and rejects just this proposal. The idea of imperfection or lack depends upon an idea of perfection; we can’t recognise that we are imperfect *unless* we have an idea of perfection with which to compare ourselves.

This argument seems to work in other cases, e.g. REAL and REALITY. It is intuitively plausible that our concept REAL is not an abstraction from NOT-UNREAL – how could we first have experiences of what is unreal on which UNREAL is based? Our experiences are fundamentally of what is real, so REAL is the primary concept. But this is not as clearly true for the cases of PERFECTION and INFINITY – we could first experience limits and then create a new concept UNLIMITED and then use this concept to create the concepts PERFECTION and INFINITY.

Furthermore, PERFECTION and INFINITY – if they mean more than ‘not imperfect’ and ‘not finite’ – are arguably challenging and unclear concepts. What is it, exactly, to think not merely of the *absence of limits*, but of something for which there could be no limits? Yet Descartes claims that we have a very powerful – clear and distinct – positive idea of God as perfect and infinite, and not some hazy notion of something indefinitely great. Yet he also accepts that, as a finite mind, he cannot ‘grasp’ this thought, but he merely ‘understands’ it. With this admission, his claim that the concept of GOD is both clear and distinct and involves a positive conception of God’s infinity and perfection is unpersuasive.

3.8. Descartes’ Cosmological Argument

Descartes offers another argument for knowledge, through intuition and deduction, of God’s existence in the form of a cosmological argument. Cosmological arguments for God’s existence start from (some version of) the question ‘Why does anything exist?’.

At this point in the *Meditations*, the only thing that Descartes knows to exist is himself. Why so? Descartes is seeking to find out what he can know as true. To achieve this, he has decided to avoid believing anything that is not ‘completely certain and indubitable’. He then argues that he can doubt his senses, his memory and even that he has a body (note that these are all a posteriori claims we would use perception to establish). Descartes supposes that all that he perceives and remembers is an illusion; that he has no body or senses at all; that in believing anything else, he is being deceived by a ‘supremely powerful and cunning deceiver’, an ‘evil demon’. The demon could make it seem that he sees a tree when he doesn’t, that he has a body when he doesn’t, and so on. But he has managed to establish his own existence as a ‘thinker’, and is now meditating on this and the implications of the concepts that he finds in his mind.

So, Descartes asks what causes his existence. As the argument is long and complicated, I have divided it into sections.

P1. If I cause my own existence, I would give myself all perfections (omnipotence, omniscience, etc.).

P2. I do not have all perfections.

C1. Therefore, I am not the cause of my existence.

P3. A lifespan is composed of independent parts, such that my existing at one time does not entail or cause my existing later.

P4. My existence is not uncaused.

C2. Therefore, some cause is needed to keep me in existence.

P5. I do not have the power to cause my continued existence through time.

C3. Therefore, I depend on something else to exist.

P6. I am a thinking thing and I have the idea of God.

P7. There must be as much reality in the cause as in the effect. (See the section ‘Descartes’ Trademark Argument’ for discussion of this claim.)

C4. Therefore, what causes my existence must be a thinking thing and have the idea of God.

P8. Either what causes me is the cause of its own existence or its existence is caused by another cause.

P9. If its existence is caused by another cause, then the point repeats: this second cause is in turn either the cause of its own existence or its existence is caused by another cause.

P10. There cannot be an infinite sequence of causes.

C5. Therefore, some cause must be the cause of its own existence.

P11. What is the cause of its own existence (and so, directly or indirectly, the cause of my existence) is God.

C6. Therefore, God exists.

Descartes adds a further argument, picking up (P3) and (C2).

C2. Some cause is needed to keep me in existence.

P12. There cannot be an infinite chain of causes because what caused my existence also causes my continued existence in the present.

P13. My parents, or any other supposed cause of my existence, do not keep me in existence.

P14. The only cause that could keep me in existence is God.

C7. Therefore, God exists.

## 3.8.1. The cause of continued existence

Why does Descartes say that not only the start of his existence, but his continued existence through time, needs to be caused (C2)? For instance, we might object that my continued existence doesn’t require a cause, because nothing changes – I simply continue to exist. If I cease to exist, that requires a cause.

But this misunderstands both causation and continued existence. I am sitting on a chair – nothing is changing. But there is a cause of this continued state of affairs, namely gravity and the rigidity of the chair. Should either of those standing conditions change, then I would no longer be sitting on the chair. I’d either be floating (no gravity) or sitting on the ground (collapsed chair). That people don’t die at any given instant is the result of whatever it is that keeps them alive. Therefore, we should accept that my continued existence does require a cause. It is worth noting that what causes my continued existence must itself continue to exist – it can’t be a cause in the past, since my continued existence must be caused from moment to moment (just as my sitting on a chair is).

We might object, however, that my continued existence is simply dependent on the immediately preceding state of affairs, and so we don’t need to say that what caused me to exist in the first place also keeps me in existence. For instance, my bodily processes keep me alive at any moment, but they didn’t give me life.

But, first, this forgets that Descartes is talking about his self, which is his mind, not his body. Descartes has argued that he, his mind, is an entirely separate substance from the body. So what keeps a mind in existence through time? If it was something in his mind itself, he would know, he claims (C1). If he could cause his own existence at the next moment, he would give himself all perfections (P1). And it can’t be his parents – they only gave existence to him originally, but don’t keep him in existence. Second, even if we allowed that our bodily processes keep us alive from moment to moment, what are they causally dependent on? This line of thought triggers the argument from (P7). Bodily processes aren’t the cause of their own continuation. If Descartes’ existence is causally dependent on something else, and an infinite regress of causal dependency is impossible, then, Descartes argues that something must exist that is not causally dependent on anything else for its existence. This is God.

# 3.8.2. Empiricist Responses to Descartes’ Cosmological Argument

Descartes assumes that his existence has a cause (P4). He also assumes that an infinite series of causes is impossible. We can raises doubts about both claims.

## 3.8.2.1. Hume on the causal principle

The causal principle is the claim that everything has a cause. But is it true? Must everything be sustained in existence by causal dependency on something else? Hume argues in *A Treatise of Human Nature*, that we cannot know whether everything has a cause.

The claims ‘everything has a cause’ and ‘something cannot come out of nothing’ are not analytically true. ‘Some things do not have a cause’ is a not a *contradiction in terms* like ‘Some bachelors are married’ is. Of course, from our experience, we have good reason to think that everything has a cause, but this is still only a contingent truth; it may be false. We cannot show that it holds without exception. Thus (P4) and (C2), we may object that it is possible that Descartes’ existence is uncaused; we cannot show otherwise by rational intuition and deduction. What causes what is a matter of fact, and this can only be established by experience.

Second, if we can’t know that it is impossible for something to come out of nothing, then we can’t know, either, that a cause must contain at least as much ‘reality’ as its effect. Rather, ‘anything may produce anything’. What causes what is something we must discover from experience; we cannot know it by a priori reason. Thus to (P7), we can object that we cannot know that a cause must have as much ‘reality’ as its effect. What the cause of a ‘thinking thing’ is we must discover through experience, and cannot know not a priori. There is no a priori reason to think that matter cannot produce thought, and experience would indicate that matter does indeed produce thought. So we cannot infer that either the first cause or what sustains Descartes’ continued existence as a mind must itself be a mind, let alone one that has the perfections attributed to God.

## 3.8.2.2. The possibility of an infinite series

Descartes claims that there cannot be an infinite series of causes. Before going further with this thought, can’t we just cut it short by invoking science? We don’t need to show that an infinite series of causes is impossible, because cosmology shows that the universe started with the Big Bang, just under 14 billion years ago.

However, there are two problems with this response. First, it deals with a sequence of causes in time, while Descartes is interested in what keeps him in existence now. Second, it doesn’t get rid of the problem of an infinite series of causes. The universe isn’t the kind of thing that is self-sustaining – not itself causally dependent on anything. We can ask what caused or causes the universe? At this point, the possibility of an infinite series arises afresh. Even if *this* universe has a cause, perhaps it was caused by a previous (or another) universe, and so on, *infinitely*. Current speculation in physics suggests several different ways in which universes might be related to each other, including the idea that our universe is just one aspect of an infinite ‘multiverse’. But could there be an infinite series of causes in this sense?

An infinite series is not a very long series. Infinity is not a very large number. It is not a number at all. An infinite series of causes, quite literally, has no beginning. Because the universe exists, to claim that it is part of an infinite series of causes is to claim that an actual infinity – something that is in fact infinite – exists. This is quite different from talking about the idea of infinity. The idea of infinity makes sense; but does it make sense to think that something infinite actually exists?

Here’s a popular example. Suppose there is a hotel with infinite rooms. Even when the hotel is completely full, it can still take more people! You cannot add any number to infinity and get a bigger number: ∞ + 1 = ∞. Suppose, when the hotel is full, infinitely more people show up. They can all be accommodated! ∞ + ∞ = ∞. But it is impossible for the hotel to be full and still have room for more guests. So there cannot be an ‘actual’ infinity.

We can apply the point to an infinite series of causes. Each thing that begins to exist in the universe – stars, planets, people – is caused to exist by something before it, and whatever caused each thing is itself caused by something before it. But if there is an infinite chain of causes, that series of causes never has a starting point. The process never gets started, because it has always been going on. So each new cause doesn’t add one more cause to the series, since ∞ + 1 = ∞. But

surely each cause is one more cause. And we would never have reached the point in the series of causes at which we are now if it were an infinite series. How could anything exist if there were an infinite number of levels of sustaining cause below it? So we have good reason to think that an infinity of causes is impossible.

In response, we may appeal again to Hume. The claim ‘there cannot be an infinite series of causes’ is not an analytic truth, nor can we have experience of this matter. It seems conceivable, therefore, that something has always existed, and each thing has in turn caused the next. We cannot infer, then, that something that is its own cause – God – exists and is the cause of everything else.

But this is too quick. An actual infinity (of causes or hotel rooms or whatever) leads to paradoxes. If these paradoxes cannot be resolved, then they are genuine self-contradictions (e.g. that each new cause adds to the number of causes and that it does not). Anything that entails a contradiction must be false. So, if we cannot solve the paradoxes, Hume is wrong: we can deduce that there cannot be an infinite series of causes. We do not need experience to establish the claim.

But perhaps the paradoxes are the result of limitations on how we are thinking about infinity. Mathematicians (following Georg Cantor) argue that we are mistaken to apply intuitions about finite numbers to infinity, and new ways of thinking are needed (e.g. about different ‘sizes’ of infinity).

3.9. Descartes’ Ontological Argument

Ontological arguments claim that we can deduce the existence of God from the concept of God. They argue that once we understand the concept GOD, we understand that a being corresponding to this concept, God, must exist.

Descartes’ argument relies heavily on his doctrine of clear and distinct ideas (for more on this doctrine, see the section on ‘Reason, intuition and knowledge’). He opens *Meditation* V by explaining how we can explore our concepts in thought to gain knowledge. For example, you may think that there can be triangles whose internal angles don’t add up to 180 degrees, but reflection proves this impossible. Our thought is constrained in this way. The ideas we have determine certain truths, at least when our ideas are clear and distinct. Once you make the idea of a triangle (the concept TRIANGLE) clear and distinct, you understand that the internal angles of a triangle add up to 180 degrees, and this shows that this is, in fact, true.

We can now apply this method to the concept of GOD. Descartes’ argument is very brief:

The idea of God (that is, of a supremely perfect being) is certainly one that I find within me . . .; and I understand from this idea that it belongs to God’s nature that he always exists.

We can understand this passage either in terms of rational intuition of the clear and distinct idea of GOD or as a very short deduction from such a clear and distinct idea. Understood the first way, Descartes is arguing that careful reflection on the concept of GOD reveals that to think that God does not exist is a contradiction in terms, because it is part of the concept of a supremely perfect being that such a being has existence. Thus, we can know that it is true that God exists.

In fact, it shows that God must exist. A contradiction in terms does not just happen to be false, it must be false. So to say ‘God does not exist’ must be false; so ‘God exists’ must be true.

As in the case of the triangle, it is not our thinking it that makes the claim true. Just as the concept triangle forces me to acknowledge that the internal angles of a triangle add up to 180 degrees, so the concept GOD forces me to acknowledge that God exists.

Furthermore, I cannot simply change the concept in either case; I can’t decide that triangles will have two sides nor that it is no part of the concept of a supremely perfect being that such a being exists. I haven’t invented the concept of GOD.

One striking puzzle is why Descartes thinks that the concept of a supremely perfect being includes the thought that such a being exists. Spelling this out (P4 below) gives us a short deductive argument:

P1. I have the idea of God.

P2. The idea of God is the idea of a supremely perfect being.

P3. A supremely perfect being does not lack any perfection.

P4. Existence is a perfection.

C1. Therefore, God exists.

But why should we accept (P4)? In the main body of the *Meditations*, Descartes doesn’t say. However, in an appendix to the *Meditations*, known as ‘Objections and Replies’, Descartes explains that God’s existence is entailed by the other perfections of God. For example, a supremely perfect being is omnipotent, possessing all power it is logically possible to possess. An omnipotent being cannot depend on any other being for its existence, since then it would lack a power, viz. the power to cause its own existence. An omnipotent being has this power and so depends on nothing else to exist. Such a being exists eternally, never coming into being or going out of being. As a supremely perfect being, God is omnipotent by definition, and so God must exist.

God is the only concept that supports this inference to existence, because only the concept of God (as supremely perfect) includes the concept of existence (as a perfection). We can’t infer the existence of anything else this way.

# 3.9.1. An Empiricist Response to Descartes’ Ontological Argument

Empiricists claim that nothing can be shown to exist by a priori reasoning. It is not self-contradictory to say that God does not exist. Hume provides an example of this response. Hume’s fork separates what we can know a priori – ‘relations of ideas’ – from claims about what exists – ‘matters of fact’. Matters of fact can’t be established by a priori reasoning, but require experience. So anything that can be established by a priori reasoning must be a relation of ideas.

In his *Dialogues concerning Natural Religion*, David Hume provides an objection to Descartes’ argument:

P1. Nothing that is distinctly conceivable implies a contradiction.

P2. Whatever we conceive as existent, we can also conceive as non-existent.

C1. Therefore, there is no being whose non-existence implies a contradiction.

We can put the argument another way: If ‘God does not exist’ is a contradiction, then ‘God exists’ is an analytic truth. But this can’t be right, because claims about what exists are matters of fact, synthetic propositions.

Descartes could respond in either of two ways. He could claim that ‘God exists’ is a synthetic truth, but one that can be known by a priori reflection. Or he could claim that ‘God exists’ is an analytic truth, though not an obvious one. Because he doesn’t have the concepts ‘analytic’ and ‘synthetic’ (they were invented 150 years later, by Kant), he doesn’t, of course, say either. Instead, he defends his claim as the product of rational intuition (and perhaps deduction).

On these grounds, he rejects Hume’s (P2). Because our minds are finite, we normally think of the divine perfections, such as omnipotence and necessary existence, separately and so we don’t notice that they entail one another. But if we reflect carefully, we shall discover that we cannot conceive of one while excluding the other. It is a contradiction to deny that God exists.

3.10. Descartes’ ‘Proof’ of the External World

In the *Meditations*, Descartes is seeking to find out what he can know as true. To achieve this, he has decided to avoid believing anything that is not ‘completely certain and indubitable’. He then argues that he can doubt his senses, his memory and even that he has a body (note that these are all a posteriori claims we would use perception to establish). He supposes that all that he perceives and remembers is an illusion; that he has no body or senses at all; that in believing anything else, he is being deceived by a ‘supremely powerful and cunning deceiver’, an ‘evil demon’. The demon could make it seem that he sees a tree when he doesn’t, that he has a body when he doesn’t, and so on. One conclusion of these sceptical arguments is that we cannot know from perception that physical objects exist. While we can know that we have sensory experiences, those experiences don’t give us the knowledge that their causes are physical objects.

Descartes then argues that he can know that he exists, as something that thinks. He can also reflect on his thoughts and concepts. One of these concepts is that of a physical object. If perception doesn’t show that physical objects exist, then in order to prove that they exist, we need to undertake a number of preliminary steps.

1. We need to understand our concept of a physical object – what is it that we think exists?
2. We need to show that this is a coherent concept, not something self-contradictory (like the concept of a round square).
3. We need to show that it is possible that physical objects exist.

With all that in place, we can then argue that

1. Physical objects do, in fact, exist, and we can know this.

We start with (1) and (2).

# 3.10.1. The Concept of a Physical Object

Descartes discusses the concept of physical object when discussing the nature of his mind in *Meditation* II. He has argued that ‘sensing’ is just having sensory experiences – whether physical objects are the cause of these experiences is not clear and distinct. This is puzzling, so he considers perceptual experiences further, focusing on the example of perceiving a piece of wax. His question is, ‘exactly what is it that I think a piece of wax, as a physical object, is?’ (In the argument that follows, ‘imagination’ is the faculty that deals with images, including those derived from sense experiences.)

P1. When I melt a piece of wax, it loses all of its original sensory qualities (the particular taste, smell, feel and shape it has).

P2. Yet I believe it is the same wax.

C1. Therefore, what I think of as the wax is not its sensory qualities.

P3. What I think is the wax is what remains through the changes of its sensory qualities.

P4. This is a body, something that is extended – i.e. has size and shape and takes up space – and changeable, i.e. its sensory and spatial properties can change.

P5. I know that the wax can undergo far more possible changes, including changes in its extension, than I can imagine.

C2. Therefore, my concept of the wax as extended and changeable does not derive from my imagination (and therefore it does not derive from perceptual experiences).

C3. Therefore, I comprehend the wax as what it is (as opposed to its sensory qualities) by my mind alone.

C4. Only this thought of the wax, and not the perceptual experience of it, is clear and distinct.

Descartes finishes by commenting that the wax he comprehends by his understanding is the same wax that is presented by images from the senses. Although we say we ‘see’ the wax (through vision), in fact we judge (through understanding) that it is present from what we see.

Descartes’ question is not about the wax itself, but about his experience, knowledge and concept of it. This is shown by his comment, on p. 8, that ‘[w]hat I see might not really be the wax; perhaps I don’t even have eyes with which to see anything’. He doesn’t, in Meditation II, know that there are physical objects. But he knows he has experiences of them. And it is this – his concept of what he experiences – that he is exploring.

Descartes turns to the question of whether anything corresponds to our concept of PHYSICAL OBJECT in Meditation V. He argued, in Meditation III, that whatever is clearly and distinctly perceived is true. His concept of PHYSICAL OBJECT, refined by the wax argument to mean a body that is extended and changeable, is clear and distinct. Therefore, it is a coherent concept and if physical objects exist, then they are indeed extended and changeable.

# 3.10.3 The Existence of Physical Objects

Having established the coherence of our concept PHYSICAL OBJECT, in Meditation VI Descartes turns his attention to whether physical objects are possible and exist. His argument that they are possible is straightforward:

P1. I have a clear and distinct idea of what a physical object is.

P2. (God exists and is supremely powerful.)

P3. The only reason for thinking that God cannot make something is that the concept of it is contradictory.

C1. Therefore, God can make physical objects.

C2. Therefore, (if God exists) it is possible that physical objects exist.

To prove that physical objects in fact exist, Descartes first considers two arguments that aim to show that the existence of the external world is the best hypothesis to explain our experience. But he is dissatisfied because neither of them gives us certainty, which he thinks is necessary for knowledge.

The first argument is from imagination. Descartes begins by showing that the faculty of imagination is different from the faculty of understanding.

P1. The imagination uses images, e.g. imagining a triangle. But the understanding does not. We cannot imagine a chiliagon, a two-dimensional figure with 1,000 sides. But we can work mathematically with it, e.g. working out its internal angles.

P2. Imagining takes more effort than understanding.

C1. Therefore, imagination and understanding are different.

P3. Imagination is not essential to me, while understanding is. I cannot be me (a thinking thing) without understanding, but I can be me without imagination.

P4. The best explanation for all these differences is that imagination depends upon having a body. (Imagination draws its ideas from the body, which makes its ideas sensory images and difficult to work with, and makes imagination not essential to a thinking thing. Being purely mental, understanding draws its ideas from itself, making them non-imagistic and easy to work with, and understanding is essential to a thinking thing.)

C2. Therefore, it is probable that the body (a physical object) exists.

It is, however, only probable, so the argument doesn’t give us knowledge of the existence of physical objects.

Descartes’ second argument is from perception. It is natural to think that we know that physical objects exist because we perceive them. Our perceptions are both involuntary and ‘much more lively and vivid’ than imagination or memory. One explanation is that they are caused by physical objects that exist independent of our minds. But Descartes reminds us of his sceptical arguments: perception does not give us knowledge of the causes of our perceptual experiences. The mere fact that perceptual experiences are vivid and involuntary isn’t enough to show that they are caused by mind-independent physical objects.

It does, however, provide the starting point for his next argument. I have added in missing premises in brackets, some of which Descartes assumes because he has argued for them previously.

P1. I have involuntary perceptual experiences of physical objects.

P2. (These experiences are caused by some substance.)

P3. If the cause of my perceptual experiences is my own mind, my perceptual experiences are voluntary.

P4. Because I know my mind, I would know if my perceptual experiences are voluntary.

C1. Therefore, because I know that my perceptual experiences are involuntary, I know that the cause of my perceptual experiences is not my own mind.

C2. Therefore, the cause must be some substance outside me – either God or physical objects.

P5. If the cause is God, then God has created me with a very strong tendency to have a false belief (that physical objects exist) that I cannot correct.

P6. If God has created me with such a tendency, then God is a deceiver.

P7. (God is perfect by definition.)

C3. (Therefore,) God is not a deceiver.

C4. (Therefore, God did not create me with a tendency to have false beliefs that I cannot correct.)

C5. (Therefore, if God exists, I do not have such a tendency.)

C6. Therefore, if God exists, the cause of my perceptual experiences of physical objects is the existence of physical objects.

P8. (God exists.)

C7. Therefore, there is an external world of physical objects that causes our perceptual experiences.

This argument is one of the best examples of the use of rational intuition and deduction. It is surprising to think that we cannot know from sense experience that physical objects exist. It is even more surprising to be told that we can nevertheless know that physical objects exist using a priori reasoning.

# 3.10.4. Empiricist Responses to Descartes’ Proof of the External World

Descartes’ argument for the existence of physical objects depends on his arguments for the existence of God. If these fail, then he hasn’t shown that physical objects exist. For discussion of objections to his arguments for God’s existence, see the sections ‘Descartes’ Trademark argument’, ‘Descartes’ cosmological argument’ and ‘Descartes’ ontological argument’.

In (P2), Descartes assumes that his perceptions have a cause. If this is because he thinks that everything has a cause, then Hume argues in *A Treatise of Human Nature*, that this is not something we can know. The claims ‘everything has a cause’ and ‘something cannot come out of nothing’ are not analytically true. ‘Some things do not have a cause’ is a not a *contradiction in terms* like ‘Some bachelors are married’ is. Of course, from our experience, we have good reason to think that everything has a cause, but this is still only a contingent truth; it may be false. We cannot show that it holds without exception. So it is possible that our perceptual experiences of physical objects have no cause. This is not to argue that ‘for all we know’, there is no cause of our perceptual experiences. Rather, Hume’s argument claims that we cannot know a priori that they have a cause. It is no contradiction (though it is very strange!) to suppose that they are uncaused.

If we don’t know a priori that there are physical objects that cause our experiences of them, how do we know that they exist? Bertrand Russell provides an inductive argument for the claim that the existence of physical objects is the best explanation of our perceptual experience.

P1. Either physical objects exist and cause my sense-data or physical objects do not exist nor cause my sense-data.

P2. I can’t prove either claim is true or false.

C1. Therefore, I have to treat them as hypotheses. (A hypothesis is a proposal that needs to be confirmed or rejected by reasoning or experience.)

P3. The hypothesis that physical objects exist and cause my sense-data is better.

C2. Therefore, physical objects exist and cause my sense-data.

What is Russell’s argument for (P3)? One way to test a hypothesis is to see whether it explains why my experience is the way it is. If I see a cat first in a corner of the room and then later on the sofa, then if the cat is a physical object, it travelled from the corner to the sofa when I wasn’t looking. If there is no cat apart from what I see in my sense-data, then the cat does not exist when I don’t see it. It springs into existence first in the corner, and then later on the sofa. Nothing connects my two perceptions. But that’s incredibly puzzling – indeed, it is no explanation at all of why my sense-data are the way they are! So the hypothesis that there is a physical object, the cat, that causes what I see is the best explanation of my sense-data.

As we have seen, Descartes rejects such arguments as unsatisfactory as the conclusion lacks certainty. But perhaps inductive arguments are as good as we can get, as no deduction of the existence of physical objects is possible. (George Berkeley argues that the very concept of a mind-independent physical object is incoherent. See the section ‘Idealism’ for his argument.)

**4. Limits of Knowledge**

**4.1. Scepticism**

There is a distinction between belief, even true belief, and knowledge. According to many philosophers, knowledge requires a justification. For example, someone on a jury might think that the person on trial is guilty just from the way they dress. Their belief, that the person is guilty, might be true; but how someone dresses isn’t evidence for whether they are a criminal! True beliefs can be formed or held on irrational grounds, for no good reason. Or again true beliefs can just be lucky. For example, there is a lot of evidence that astrology does not make accurate predictions, and my horoscope has often been wrong. Suppose on one occasion, I read my horoscope and believe a prediction, although I know there is evidence against thinking it is right. And then this prediction turns out true! When we form a belief, we should do so rationally, on the basis of reasons and evidence. If we do, then our belief will be justified. And this belief, if it is also true will amount to knowledge.

Scepticism is the view that our usual justifications for claiming our beliefs amount to knowledge are inadequate, so we do not in fact have knowledge. Scepticism can target knowledge from any source, including perception and reason. And so it challenges both empiricism and rationalism.

But before saying more about the nature of philosophical scepticism, let’s look at a famous example.

## 4.1.1. Am I a brain in a vat?

Thought experiments are a philosophical method designed to test a hypothesis or philosophical claim through imagining a hypothetical situation, and coming to a judgment. Here is a thought experiment that tests whether we have any knowledge at all.

Suppose that I am not a walking, talking human being, but simply a brain in a vat. Connected to my brain is a supercomputer that feeds in just the right impulses to generate the illusion of reality as it is. All of my sensory experiences are being produced in my brain by electrical signals from the supercomputer. I’m living in a virtual reality. Since I think that the reality I experience is one of physical objects and other people, I’m being deceived.

Here is the sceptical challenge: I cannot know that I am not a brain in a vat. If I were, things would seem exactly the same as if I am a walking, talking person. If I were a brain in a vat, my experiences would be qualitatively indistinguishable from the experiences I have if I am not. So I can’t any evidence that I am not a brain in a vat. So I can’t know, therefore, whether I am, in fact, a brain in a vat or not.

But if I am a brain in a vat, all my beliefs about what I experience are false; I have no body, I’m not sitting at a computer, I’m not hearing the sounds of keys clicking, etc. More importantly, even if I’m not a brain in a vat, and reality is as I think it is, my true belief lacks justification. I don’t have any reason to believe that reality is as I think rather than to believe that I am a brain in a vat. But if my belief that reality is as I think is not justified, then it isn’t knowledge. I don’t know that I’m not just a brain in a vat.

Let’s extend the thought experiment. The supercomputer feeds me not only sense experiences, but also ‘memories’. So I cannot trust my memories, because the computer could create ‘memories’ of things that never happened. So I cannot know anything about the past, including whether it happened at all. Perhaps I only just came into existence, and all my memories are false.

Let’s take it one step further. Perhaps even my thoughts are being fed to me by a super-computer. Isn’t it possible that every time I think ‘2 + 2 = ?’, the computer makes me think ‘4’ when the answer is actually 5? Can I know that this isn’t happening? How? If I can’t, then my belief that 2 + 2 = 4 isn’t justified. And so I can’t know that 2 + 2 = 4.

## 4.1.2. The distinction between philosophical scepticism and normal incredulity

Reflecting on the thought experiment of being a brain in a vat helps us understand some peculiar features of philosophical scepticism and how it differs from normal incredulity, our normal everyday doubts about whether some claim is true or not.

Philosophical scepticism can get started by reflecting on how we know what we think we know. Take the belief that I have two hands. I think I know that I have two hands. But how do I know this? Well, I can feel them, I can see them. But, says the sceptic, couldn’t my experience – what I feel and what I see – be just the same if I were a brain in a vat? If I don’t know I’m not a brain in a vat, do I really know that I have two hands? In fact, do I know that a world exists outside my mind at all? How do I know that appearance is a reliable guide to reality?

Philosophical doubts are peculiar. They don’t make sense in everyday circumstances. Of course, if I’ve just been in an accident, and can’t feel my left arm, doubting whether I have two hands does make sense! But the sceptic is not interested in these propositions when we have an ‘everyday reason’ to doubt them. The sceptic’s reason for doubting them does not arise from a particular context – it is a general doubt about their justification. The sceptic admits that there is no everyday reason to doubt whether I have two hands or whether there is an external world. But that doesn’t mean there isn’t any reason to doubt these things.

Is this sort of sceptical doubt doubt? It has no practical consequences, and a philosophical sceptic is not a very cautious person! Yet the sceptic insists that sceptical doubts are relevant – we should know that we are not a brain in a vat wired up to a super-computer if we are to know that ‘This is a hand.’

# 4.1.3. The Role/Function of Scepticism Within Epistemology

The effect of philosophical scepticism is not ‘We can’t be certain of our everyday judgements, although they are probably true.’ It is to put the whole idea of our usual justifications into question. If these sceptical possibilities were true, we would have absolutely no reason to hold on to our usual beliefs. If I was wired up to a supercomputer, things seem exactly the same, but the reality is completely different. Sceptical arguments aim to completely undercut our usual justifications.

We might think that it is ‘unreasonable’ to have such doubts. But this misunderstands the role or purpose of doubt. While some philosophers have understood philosophical scepticism as a kind of theory, it is better to understand it as a kind of challenge. The sceptic doesn’t suggest that there is any reason to believe in sceptical possibilities of thought experiments, but requests that we rule them out as possibilities. In trying to meet the sceptic’s challenge, we can discover what we know and how we know it.

Scepticism is sometimes taken as the claim that nothing is known. But this is not a good definition of philosophical scepticism, for it must then defend the claim that we can know nothing, which is trivially self-defeating anyway (because then we would know that we know nothing – so there is something we know). Likewise, scepticism is not claim that our beliefs are all false. For this is not logically coherent. For instance, my beliefs that ‘I am not at the South Pole’ and that ‘I am not at the North Pole’ can’t both be false (obviously, both can be true).

Scepticism is best understood as the claim that our usual justification for claiming our beliefs amount to knowledge is inadequate. Doubt based on challenging us to rule out the possibility of very unlikely situations is called ‘hyperbolic’ doubt. And the purpose of this doubt is to help us find what we can know, if anything.

# 4.1.4. Local and Global Scepticism

Local scepticism is scepticism about some specific claim, or more commonly, about some area/branch of supposed knowledge. We might doubt whether we can know how many planets exist in the universe (without doubting astronomy in general). Or more broadly, we might doubt whether there can be any knowledge about God’s existence and nature (without doubting, say, scientific knowledge). Our normal incredulity is always local – we have specific reasons for doubting specific claims, and in our philosophical ‘moods’, we may doubt knowledge about religion, ethics, and so on.

Global scepticism extends doubt without limit. The brain in a vat provides an example. If we can’t know whether or not we are brains in vats, and cannot even trust our reasoning, then it seems all our knowledge comes into question. Global scepticism has focused especially on having no knowledge of an external world of physical objects. If we can secure knowledge of such a mind-independent world, we will have defeated global scepticism, whatever conclusions we reach about the other branches of knowledge.

4.2. Descartes’ Sceptical Arguments

Descartes begins his *Meditations* by presenting three arguments that support scepticism, sometimes called the ‘three waves of doubt’. In the first, he presents an argument from illusion to begin to question his knowledge from sense experience. In the second, he questions whether he can know whether he is awake or asleep. This challenges empirical knowledge more robustly. In the third, he presents his version of the ‘brain in a vat’ thought experiment: can he know that he is not being deceived in all his experience and thought by a powerful, malicious demon?

Descartes begins *Meditation* I by declaring that he has known for a long time that in order to establish anything ‘in the sciences that was stable and likely to last’, he would have to start from the foundations. He does not need to reject as false everything he thinks he knows, but he needs to avoid believing things ‘that are not completely certain and indubitable’. To establish this certainty, he seeks to test his beliefs by doubting them. He adopts philosophical scepticism as his starting point. As he tries to call his beliefs into question, he repeatedly asks how he can know they are true. So he understands knowledge in terms of what is ‘completely certain and indubitable’. If we can doubt a belief, then it is not certain, and so it is not knowledge. This procedure for establishing what we can know to be true is Descartes’ ‘method of doubt’.

If Descartes doubted each belief in turn, this would take forever. So he decides to question the principles on which his beliefs are based. We can understand this as his calling into question the general justifications we offer for our beliefs.

# 4.2.1. First Wave of Doubt: An Argument from Illusion

So what can we doubt? Descartes begins by presenting an argument from illusion as many of his beliefs are based on his sense experience. He notes that he has, in the past, been deceived by his senses – things have looked a way that they are not. Things in the distance look small, for instance. Or again, an oar half-submerged in water looks crooked.

But, Descartes remarks, such examples from unusual perceptual conditions give us no reason to doubt all perceptions, such as that I am looking at a piece of paper with writing on it. More generally, we might say that perceptual illusions are special cases (and ones we can frequently explain). Otherwise we wouldn’t be able to talk about them as illusions. So they don’t undermine perception generally.

# 4.2.2. Second Wave: The Argument from Dreaming

Descartes then doubts whether he knows he is awake. Sometimes when we dream, we represent to ourselves all sorts of crazy things. But sometimes we dream the most mundane things. I could be dreaming that I’m looking at a piece of paper. I could even have the thought, while I’m dreaming, that I’m not dreaming! There is no reliable way to tell whether I’m awake or asleep.

This argument attacks all sense perception, even the most mundane and most certain. I cannot know that I see a piece of paper because I cannot know that I am not dreaming of seeing a piece of paper. It questions whether we can tell what reality is like from what we experience, since those experiences could be no more than a dream.

We can object that there are reliable ways of distinguishing waking perception from dreaming, such as the far greater coherence of perception. But what Descartes means is that I cannot know, of my perception now, whether I am awake or asleep. The objection assumes that I can rely on my memory of what I have experienced to compare it with my dream. But what if I’m dreaming that I remember this?

Descartes then claims that even if he were dreaming, and may be imagining particular physical objects, dreams are constructed out of basic ideas and these must correspond to something real – ideas of body, extension, shape, quantity, size, motion and time. And so the truths of geometry seem secure, as do truths of arithmetic, such as ‘2 + 3 = 5’. Even if he is dreaming, this seems impossible to doubt.

# 4.2.3. Third Wave: The Evil Demon

But Descartes then casts doubt on even these claims of mathematics by questioning whether God may have deceived him. Is it possible that he could go wrong in adding two and three? To the objection that God is good and wouldn’t deceive Descartes like this, Descartes introduces a further doubt. Suppose that God does not exist. Suppose, worse, that all my experiences are produced in me by an evil demon who wants to deceive me. If this were true, I wouldn’t know, because my experiences would be exactly the same (just as with the brain in the vat thought experiment). So I cannot know that I am not being deceived by an evil demon.

Descartes uses the evil demon supposition to make sure that he doesn’t believe anything he can’t know. It seems that he can’t know anything – that there is an external, physical world or even the basic truths of mathematics. Unless he can rule out the possibility that he is being deceived by an evil demon, then he can’t be certain of anything. He has reached the point of global scepticism. He describes himself as: “drowning in doubt.”

# 4.3. Descartes’ Response to Scepticism

Descartes’ response to the global scepticism at the end of *Meditation* I is to use rational intuition and deduction. He begins with the *cogito*. He argues that even if an evil demon is deceiving him, Descartes can know that he exists and that he thinks. He can know about his own mind and the thoughts he has. He then builds on this by developing his theory of clear and distinct ideas. (For more on the cogito and clear and distinct ideas, see the handout ‘Reason, intuition and knowledge’.) This enables him to argue that the truths of mathematics are indubitable, and so he can know these as well.

However, he still has no certainty about the causes of sense perception, and so the existence of a world of mind-independent physical objects. He has sensory experiences in his dreams as well, when he is not seeing or hearing at all. However, by exploring his ideas, he is able to know that his concept of a physical object is a concept of something extended in space, but that is all for now. At this point, Descartes cannot move beyond idealism.

Descartes’ next move in reconstructing what he can know is to attempt to prove the existence of God relying just on the concept of GOD and other clear and distinct ideas, e.g. concerning causation. (See the handouts ‘Descartes’ Trademark argument’, ‘Descartes’ cosmological argument’ and ‘Descartes’ ontological argument.) Not only can God’s existence be known from the concept of GOD, but also a number of other truths about God’s nature. Among these, one of the most important for Descartes’ project is that God is not a deceiver and God is omnipotent. It is on this basis that Descartes meets the challenge of scepticism head on.

Because God is not a deceiver, then God would not allow incorrigible errors, but has given me the ability to form true beliefs (see the handout ‘The Cartesian circle’). And so I can dismiss the possibility of the evil demon – if an evil demon were deceiving me, I would have no way of correcting my beliefs about the world. Among these beliefs is that physical objects exist and are extended in space. Because God can bring about anything that corresponds to a clear and distinct idea, and is not a deceiver, I can know that there are such physical objects, which I experience in perception. (For the full argument, see the handout ‘Descartes’ proof of the external world’.)

But how do I know that I am not dreaming? At the very end of the *Meditations*, Descartes also uses God’s not being a deceiver to solve this objection. He accepts that we can tell the difference between dreaming and being awake, because memory connects up perceptions coherently, but not dreams, and because we can confirm our perceptions using different senses. This response is only available now (and not in *Meditation* I) because he has established that God is not a deceiver. Without that, we couldn’t rely on memory in this way.

We must nevertheless be careful about what we claim to know through perception. Our sense perceptions are not ‘reliable guides to the essential nature of the bodies located outside me, for on that topic they give only very obscure and confused information’. We can and do continue to make mistakes about what we perceive. Our individual perceptual judgments will not qualify as certain in the way clear and distinct ideas do. In particular, we shouldn’t think that our perceptions of physical objects as having properties of colour, smell, taste, temperature, and so on, resemble the objects themselves. The essential nature of physical objects is given not through sense experience, but through an a priori analysis of our concept of PHYSICAL OBJECT. Thus, Descartes ends up adopting indirect realism, drawing a distinction between primary and secondary qualities and defending the claim that we can know that physical objects have primary qualities but we cannot know that they have secondary qualities.

**4.4. Empiricist responses to scepticism**

Empiricists deny that there is any a priori knowledge of synthetic propositions. This can be understood as a form of local scepticism. They restrict the knowledge that we have to

1. (a priori) knowledge of analytic propositions and what can be deduced from them;
2. (a posteriori) knowledge of synthetic propositions about the world outside one’s minds; and
3. knowledge of our own minds, derived from impressions of reflection.

Any claim that does not fall under one of these three categories, we do not know.

One objection to empiricism is that this restriction on knowledge is too severe. It leads to local scepticism across many branches of supposed knowledge. For example, can we know that God exists? If ‘God exists’ is a synthetic claim, according to empiricism, we could only know that God exists from sense experience. But can we? Hume argued that we can’t – the reasoning involves claims that sense experience cannot establish. So we can’t know whether God exists.

Another example is morality. Moral claims, such as ‘Murder is wrong’, don’t appear to be analytic. But could we know them through sense experience? Which of our senses pick up on ‘wrongness’, and how? If empiricists can’t show that moral claims are either analytic or a posteriori, then they will be forced to conclude that there is no moral knowledge either.

We put these local scepticisms aside to ask how empiricists respond to global scepticism and the attack on knowledge from sense experience. As indirect realists, Locke, Russell and Trotter Cockburn argue that we only immediately perceive sense-data, and we must infer the cause of our sense experiences. They defend the claim that an external world of physical objects is that cause on the grounds that it is the best explanation of our experience, in particular, that it is involuntary, coherent between different sense modalities (vision, hearing etc.), and that it is systematic.

However, there are two objections that may be raised to this argument understood as a response to scepticism. First, Descartes argues that the existence of the external world remains a hypothesis – we cannot know with certainty that physical objects exist. But we can respond that Descartes sets the standard for knowledge too high – knowledge does not require certainty. If the existence of physical objects is the best explanation of our sense experience, this is sufficient justification for us to know that they exist (if the belief is also true). Rather than respond to scepticism by appealing to rational intuition and deduction to meet the demand for certainty, this response rejects the demand for certainty.

The second objection challenges the claim that the existence of physical objects is the best explanation for our sense experience. If we are brains in vats, then our sense experience is exactly as it is now – involuntary, coherent, systematic – and yet the world is nothing like we experience it. (Ok, so a physical object, the supercomputer, exists, but this is hardly the result Russell and others were after!) Our experience provides no reason to prefer the hypothesis of physical objects over the hypothesis of being a brain in a vat. Both explanations of our experience are equally good. So inference to the best explanation cannot show that we can rule out the possibility of being brains in vats. So it doesn’t meet the challenge of global scepticism.

Berkeley takes a completely different approach. He removes the challenge of scepticism by rejecting the distinction between appearance and reality. What we experience is reality. Idealism has no need to discover how our perceptions of physical objects relate to reality. In experiencing ideas, we are experiencing what exists.

One interesting point about Berkeley’s argument is that his understanding of the relation between our minds and God as the source of our perceptions is rather similar to the relation between our brains and the supercomputer in the brain in a vat thought experiment. Berkeley infers that something outside my mind must cause my perceptions, and given the complexity of my experience, that something must be God. But could it be a supercomputer instead?! Berkeley would, of course, respond that a supercomputer is a physical object (as is a brain), and we have no reason to think that physical objects can exist independent of minds. The thought experiment is incoherent.

The responses we have discussed in this section are from classical empiricists (and Russell). But empiricism may be combined with reliabilism (as well as other theories of knowledge). So the response that follows next can also count as empiricist as it defends our ability to gain knowledge from sense experience.

# 4.4.1. Reliabilism as a Response to Scepticism

Reliabilism disputes the theory of knowledge that scepticism assumes and denies that we need the condition of justification. Reliabilism counters that our true beliefs do not need to be justified to count as knowledge, they only need to be produced by a reliable cognitive process, that is, a process that produces a high percentage of true beliefs.

Suppose I am not a brain in a vat. Then perception is a reliable process – most of the beliefs I form on the basis of sense experience are, in fact, true. That is enough to say that I gain knowledge from sense experience – my beliefs are true and produced by a reliable cognitive process. I do not need, in addition, to justify my beliefs, e.g. to have evidence that perception is reliable. In particular, I do not need to know that I am not a brain in a vat. If I am not a brain in a vat, then, because perception is a reliable process, I know that there are physical objects.

This can sound very counterintuitive. If I am a brain in a vat, then I don’t know that there are physical objects. Even the reliabilist agrees with that: if I am a brain in a vat, how I form beliefs is very unreliable, and most of them are false. So surely I need to know that I am not a brain in a vat, and that my beliefs are caused by a reliable process, in order to know that there are physical objects.

But as an objection to reliabilism, this point is confused. First, reliabilism can respond that there is a difference between knowing that p and knowing that I know that p. Suppose that I don’t know whether I am a brain in a vat. If that’s true, then I don’t know whether perception is reliable or not. So since I don’t know whether perception is reliable, I can’t know that I know that there are physical objects. This is all true, but why does any of this matter? If I’m not a brain in a vat, then – whether I know it or not – perception is reliable, and so I know that there are physical objects.

Why should we think that in order to know that p, I must also know that I know that p? It could be that the cognitive process that produces beliefs about p is reliable, but the cognitive process that produces beliefs about what I know is unreliable. In such a case, I can know that p without knowing that I know that p. For instance, animals can gain knowledge through perception, but they can’t even think about whether they are brains in vats, let alone having reliable thoughts about such matters. Reliabilists reject the sceptical claim that we cannot know about physical objects through perception unless we know that we are not brains in vats.