PHILOSOPHY PREPARATORY WORK

CP, MP, and PP version

OVERVIEW

During the autumn term—or Michaelmas term, in Oxford-speak—Computer Science & Philosophy (CP), Maths & Philosophy (MP), and Physics & Philosophy (PP) first years at Balliol study *Introduction to Logic* and some *General Philosophy*. The notes and exercises below are to help you prepare for this. Read them carefully, and submit completed exercises to me by email by Friday, 2nd October. We will discuss your solutions to them in classes and tutorials during the term.

A note on buying books. Library provision in Oxford is excellent, but you will want to buy your own copies of certain books: *The Logic Manual* and the appropriate set text (which for CP students is Charles Petzold's *The Annotated Turing*, for MP students is Frege's *Foundations of Arithmetic*, and for PP students is the Leibniz-Clarke Correspondence). These can often be picked up for cheap second-hand.

Jonny McIntosh, July 2026

jonathon.mcintosh@philosophy.ox.ac.uk

INTRODUCTION TO LOGIC

CP, MP, and PP students have eight weeks of *Introduction to Logic* in Michaelmas. (They go on to study a follow-up course, *Elements of Deductive Logic*, in the spring term—or Hilary, in Oxford speak.) Each week there will be a set-up class or lecture on the Monday or Tuesday, followed by a back-up class on the Thursday or Friday, in which you will be discussing solutions to exercises that you will have submitted beforehand. (The Faculty also offers weekly lectures, which are usually posted online.)

The *Introduction to Logic* course is based around Volker Halbach (2011) *The Logic Manual* (Oxford University Press). You may want to buy a copy of this before you arrive. However, while we won't discourage you from reading it now, it is not essential that you do so. It will be enough to read the material below, and work through the following **LOGIC EXERCISES**.

LOGIC EXERCISES

MODAL and LOGICAL VALIDITY

We define an **argument** to be a (possibly empty) set of declarative sentences, called the **premises**, together with a declarative sentence designated as a **conclusion**. For example:

> *The ball is red. If the ball is red then the ball is coloured. So the ball is coloured.*

This is an argument. We have a set of sentences ('The ball is red' and 'If the ball is red then the ball is coloured') and a sentence designated (by 'so') as a conclusion ('The ball is coloured').

> *The ball is coloured. If the ball is red then the ball is coloured. So the ball is red.*

This is also an argument. Again we have a set of sentences ('The ball is coloured' and 'If the ball is red then the ball is coloured') and a sentence designated as a conclusion ('The ball is red').

There is an intuitive sense in which the first argument is a good argument and the second one is not. Logicians try to be more precise about what we might mean by a **good** argument here.

By definition, an argument is **modally valid** if and only if there is no possible circumstance in which all of its premises are true and its conclusion is false. An argument is, by contrast, **logically valid** if and only if there is no interpretation of its subject-specific words (i.e. no way of assigning meanings to them) under which all its premises are true and its conclusion is false.

The first argument above is both modally valid and logically valid. To see that it is modally valid, consider any possible circumstance in which all of its premises are true. That would be a circumstance in which the ball is red and in which, if it is red, it is coloured. So it would be a circumstance in which the ball is coloured. But that means it is a circumstance in which the conclusion of the argument is true. In short, any possible circumstance in which all of its premises are true is also a circumstance in which its conclusion is also true, and so not false. So there is no possible circumstance in which all of its premises are true and its conclusion is false.

To see that the first argument is also logically valid, consider the result of replacing each of its subject-specific words with schematic letters: *The F is G. If the F is G then the F is H. So the F is H*. By virtue of your grasp of English, you are able to see that there is no way of replacing these letters so that the result is an argument in which the premises are all true and the conclusion is false. So there is no interpretation of the subject-specific words of the argument (i.e. no way of assigning them meanings) under which all of its premises are true and its conclusion is false.

By contrast, the second argument is neither modally valid nor logically valid. To see that it is not modally valid, consider a possible circumstance in which the ball is green, and so coloured, but not red. That is a circumstance in which the the ball is coloured and in which if it is red then it is coloured. But it is not a circumstance in which it is red. So it is a circumstance in which all the premises of the argument are true but the conclusion is false. But that means there is a possible circumstance in which all the premises of the argument are true and the conclusion is false.

To see that the second argument is also not logically valid, consider the result of replacing each of its subject-specific words with schematic letters: *The F is H. If the F is G then the F is H. So the F is G.* It is easy to see that there is a way of replacing the letters so that the result is an argument in which the premises are all true and the conclusion is false. Consider, for instance:

> *The capital of the United Kingdom is in England. If the capital of the United Kingdom is in Cornwall then the capital of the United Kingdom is in England. So the capital of the United Kingdom is in Cornwall.*

So there is an interpretation of the subject-specific words of the second argument (i.e. a way of assigning meanings to them) under which all of its premises are true and its conclusion is false.

Q1. For each of the following arguments, say whether they are (i) modally valid, (ii) logically valid. State your answers, using complete sentences (e.g. "This argument is modally valid") and providing in each case an *explanation* of your answer, using my explanations above as a model.

> a. *All unicorns are beautiful. Orcs hate anything beautiful. So orcs hate all unicorns.*

> b. *Diamond is hard. So diamond is not soft.*

> c. *8 is a prime number. Therefore, all frogs are pink.*

> d. *All Teddy Hall students are clever. Annie is a Teddy Hall student. So Annie is clever.*

> e. *Everything is coloured red all over. So nothing is coloured blue all over.*

> f. *All cows are green. All cows are not green. So all pigs are purple.*

Logicians are interested in logical validity, rather than modal validity, and for the remaining questions, wherever you see the word 'valid', you should take it to mean 'logically valid'.

Let me also give you a couple more definitions. First, we say that a set of sentences is (logically) **inconsistent** if and only if there is no interpretation of its subject-specific words under which every sentence in that set is true, and second, that a sentence is (logically) **contradictory** if and only if there is no interpretation of its subject-specific words under which it is true. (Inconsistency is a property of *sets* of sentences, and contradictoriness a property of sentences themselves.)

Q2. Are there any valid arguments with the following features? In each case, either provide an example of such an argument or explain why there are no arguments with the relevant feature.

> a. false premises and a true conclusion.

> b. false premises and a false conclusion.

> c. some true premises, some false premises and a false conclusion.

> d. some true premises, some false premises and a true conclusion.

> e. an inconsistent set of premises

> f. a contradictory conclusion.

You may have found some of your answers to **Q2.** surprising. If so, it may help to offer one last definition: an argument is **sound** if and only if it is both valid and has all true premises. An argument can be valid, and so good in one sense, without being sound, and so good in another!

Q3. Suppose that P. J. Mangler of Magdalen argues as below in his logic work. What *exactly* is it that has Mangler has got wrong? State your answer as briefly as possible, but no briefer.

> David is a philosopher, and not a werewolf. So the argument which goes "David is a philosopher. All philosophers are werewolves. Therefore David is a werewolf" is invalid in this possible circumstance. But there is a possible circumstance in which all philosophers, including David, are werewolves. In that possible circumstance, the argument is valid.

MATHEMATICS for LOGIC

Logicians, as I say above, are interested in *logical* validity. One reason for this is that it is especially amenable to mathematical investigation. The mathematics involved is very simple, and we will go over it carefully and thoroughly. But it is useful to have a bit of a headstart.

One central notion is that of a **set**. Put roughly, sets are collections of things, called their **members** or **elements**. When a thing, x, is a member of a set, S, we write this as: $x \in S$.

Each set is *determined* by its members. So if the members of the set S are the same as the members of the set S' then S and S' are the same set—using the symbol for set membership: if, for anything x, $x \in S$ if and only if $x \in S'$, then $S = S'$. This means that, to specify a set, all we need to do is specify its members. This can be done using curly brackets, '{' and '}'. For example, we can specify the set of countries that make up the United Kingdom as: {England, Northern Ireland, Scotland, Wales}. In some cases, it is possible and useful to specify a set by means of a condition satisfied by all and only its members. For example, we can specify the set of even numbers as: {x : x is an even number}. (Read as: *the set of xs such that x is an even number*.)

Note that it also follows from the fact that each set is determined by its members that the *way* in which we specify the members of the set is irrelevant. In particular, the *names* we use to specify them doesn't matter. Neither does the *order* in which they are specified. Thus, we can equally well specify the set of countries that make up the United Kingdom as: {England, Scotland, Wales, Northern Ireland}. Thus {x : x is a country that makes up the United Kingdom} = {England, Scotland, Wales, Northern Ireland} = {England, Northern Ireland, Scotland, Wales}. Lastly, it doesn't matter how many *times* a member is specified: it only counts once. Thus {England, Scotland, Wales, Northern Ireland} = {England, Northern Ireland, Scotland, Wales, Wales}.

A special case is the set called the **empty set**, or Ø. This is the set containing *no* members.

Sometimes we are interested in particular orderings of things. For example, we might want to talk about England, Northern Ireland, Scotland, Wales *in that order*. To do that, we talk, not about sets, but about **ordered tuples**: **ordered pairs**, **ordered triples**, etc. This is done using angled brackets, '<' and '>'. Thus, the ordered tuple (in fact, ordered **quadruple**) containing England, Northern Ireland, Scotland, Wales, in that order, can be specified as follows: <England, Northern Ireland, Scotland, Wales>. Here, of course, order *does* matter. In other words, <England, Scotland, Wales, Northern Ireland> ≠ <England, Northern Ireland, Scotland, Wales>.

All sorts of things can be members of sets and ordered tuples—including sets and ordered tuples! Sets of ordered tuples (sets whose members are all tuples with the same number of members) are called **relations**. Logicians are

particularly interested in sets of ordered pairs, or **binary relations**, such as {<England, Scotland>, <England, Wales>, <Scotland, Scotland>}.

Note that the empty set is again a special case. It contains no members. As a result, it is a set whose members are all ordered pairs—all *none* of them! So it is, perhaps surprisingly, a binary relation. (Similarly, it is a **ternary relation**, a set each member of which is an ordered triple, a **quaternary relation**, a set each member of which is an ordered quadruple, and so on.)

Different binary relations have different properties. Given a binary relation R and set S, we say that R is **reflexive** on S if and only if, for each element d of S, <d, d> is an element of R. Thus if R is the binary relation {<England, England>, <Wales, Wales>, <Scotland, Scotland>} then R is reflexive on the set {England, Wales, Scotland} but is *not* reflexive on the set {England, Wales, Scotland, Northern Ireland}. R is **symmetric** on S if and only if, for any elements d and e of S, <e, d> is a member of R whenever <d, e> is a member of R. Thus if R is the binary relation {<x, y> : x and y are countries in the United Kingdom that share a land-border} then R is symmetric on the set {England, Wales, Scotland, Northern Ireland}. Indeed, R is then symmetric on *all* sets:

> PROOF:
>
> Let R be the binary relation {<x, y> : x and y are countries in the United Kingdom that share a land-border} and S be any set. Suppose that d and e are members of S and that <d, e> is a member of R. Then <d, e> ∈ {<England, Scotland>, <Scotland, England>, <England, Wales>, <Wales, England>}. But if so then also <e, d> ∈ {<England, Scotland>, <Scotland, England>, <England, Wales>, <Wales, England>}. That is, <e, d> is also a member of R. R is therefore symmetric on S. But S was any set. R is therefore symmetric on all sets.

There are various other properties of binary relations. Let me mention just two more. First, given a binary relation R and set S, R is **transitive** on S if and only if, for any elements d, e, and f of S, if <d, e> is a member of R and <e, f> is a member of R then <d, f> is also a member of R. An example here is the binary relation {<x, y> : x is an ancestor of y}. This is transitive on all sets, and in particular on any set of people, past, present, or future. For if one person is an ancestor of a second, and the second is an ancestor of the third, the first is also an ancestor of the third.

Lastly, a binary relation R is said to be a **equivalence relation** on a set S if and only if it is **reflexive**, **symmetric**, and **transitive** on S. An example is the binary relation R = {<x, y> : x and y are at the same college}, which is an equivalence relation on the set S of Oxford undergraduates. R is clearly reflexive on S: if d is a member of S, and so an Oxford undergraduate, then d is at the same college as him- or herself. It is also both symmetric and transitive on S. For suppose that d, e, and f are members of S, <d, e> is a member of R, and <e, f> is a member of R. Then d, e, and f are all Oxford undergraduates, with d and e at the same college and e and f at the same college. It follows that e and d (in that order) are at the same college, and so that <e, d> is a member of R. That means R is symmetric on S. And since, as a matter of fact, each Oxford undergraduate is at one and only one college, it also follows that d and f are at the same college, and so that <d, f> is a member of R. That means R is transitive on S. R is thus reflexive, symmetric, and transitive on S, and so an equivalence relation on S.

Q4. If R is a binary relation that is symmetric on a set S and also transitive on S, does it follow that R is an equivalence relation on S? Explain your answer by providing either a proof (as above) or a counter-example (where a counter-example is an example of a binary relation R and a set S such that R is symmetric on S and transitive on S but not an equivalence relation on S).

GENERAL PHILOSOPHY

CP and MP students also study *General Philosophy*. This component of the course looks at eight topics, mainly in epistemology (i.e. the theory of knowledge) and metaphysics: Scepticism, Knowledge, Mind & Body, Personal Identity, Perception, Induction, Free Will, and God & Evil. In Michaelmas, you will have tutorials on some of these topics. There are also weekly Faculty lectures for *General Philosophy* in Michaelmas, usually on Wednesdays. You will have classes and/or tutorials on more topics in Hilary.

The reading for *General Philosophy* comprises various articles and book chapters. Rather than trying to read these now, read the notes below, working through the various **GENERAL PHILOSOPHY EXERCISES**. You might also want to take a look at Simon Blackburn (1999) *Think* (Oxford University Press), which is an excellent introduction to many of the problems and issues that we will be exploring. Note that some of the **GENERAL PHILOSOPHY EXERCISES** presume familiarity with the concept of (logical) validity introduced in the **LOGIC EXERCISES** above. You should therefore complete those before tackling these.

GENERAL PHILOSOPHY EXERCISES

1. SCEPTICISM

One of the central questions in epistemology, the theory of knowledge, is *what, if anything, can we know?* According to **sceptics**, the answer is: *not very much*. One of the *General Philosophy* topics, **SCEPTICISM**, looks at arguments for **scepticism about the external world**, the view that we do not, and cannot, know anything about the external world around us. Some of these arguments are presented in the First Meditation of Descartes's *Meditations on First Philosophy*, in which Descartes (or rather the meditator) raises a series of sceptical **hypotheses**, concerning the veracity of the senses, the indistinguishability of waking experience from dreams, and the possibility of a deceiving God or evil demon. Contemporary epistemology often concerns itself with a variant of the last hypothesis, concerning the possibility that you are nothing more than a brain in a vat of nutrients, whose nerve endings are connected up to a supercomputer controlled by an evil scientist. While it seems to you as if you are going about your everyday life, interacting and experiencing ordinary objects, in fact (according to the hypothesis) you are not. You are merely receiving the electrical signals that you (or your brain) would have received had all been as it seems to be.

In broad outline, the sceptical argument inspired by this hypothesis proceeds as follows:

1. You do not know that you are not a brain in a vat
2. If you do not know that you are not a brain in a vat, you do not know you have hands
3. So you do not know that you have hands

The argument is valid. (Use the methods above to check this.) Moreover, it also seems to be sound: intuitively, at least, both premises seem to be true. But if the argument is not just valid, but also sound, then the conclusion is true: you do not know that you have hands! Worse, it seems that similar arguments can be given to show that you don't know pretty much anything about the external world. For whatever sentence concerning the external world that we substitute for 'you have hands', the resulting argument is no less valid, and seems to be no less sound.

In light of this troubling result, the obvious thing to do is to examine the premises more closely. While they might *seem* true, might one or the other of them in fact be false? In order to examine this, we need to get a better understanding of *why* each premise seems true. To that end...

Q1. Present a *valid* argument for the first premise of the argument, i.e. the claim that you do not know that you are not a brain in vat. Try to present the most convincing argument you can, and briefly explain (in no more than one

paragraph) the motivation behind *its* premises.

Hint: it seems plausible and important that, whether you are having the experience you seem to be having (the so-called **good** case) or that of the brain in the vat (the so-called **bad** case), the way that things experientially seem to you is the same. For example, if it experientially seems to that you are seeing a blue circle in the good case, it will seem that way in the bad case too.

Q2. Present a *valid* argument for second premise, i.e. the claim that if you do not know that you are not a brain in a vat then you do not know that you have hands. Again, present the most convincing argument you can, and explain (in a paragraph) the motivation behind its premises.

Hint: regardless of whether or not you know that you have hands, it is plausible that you at least know that *if* you have hands, you are not a brain in a vat. Suppose you do know that, and suppose you also knew that you have hands, would it follow that you know (or can know) that you are not a brain in a vat? If not, what else is required for before that conclusion will follow?

2. KNOWLEDGE

In order to get clearer on how, if at all, we might respond to sceptical arguments, it is helpful to think more about what knowledge *is*. Another of the *General Philosophy* topics, **KNOWLEDGE**, examines accounts of **propositional knowledge**, the sort of knowledge that is reported by sentences of the form 'S knows that P', where 'S' is replaced by the name of a person and 'P' is replaced by a declarative sentence—sentences like 'Javahn knows that it is raining' and 'Priya knows that the capital of Kiribati is Tarawa'. (There are other sorts of knowledge: **know how**, reported by sentences of the form 'S knows how to X', where 'X' is replaced by verb phrase describing a kind of act, and **acquaintance knowledge**, reported by sentences of the form 'S knows O', where 'O' is replaced by the name of an object or individual.)

Many philosophers try to provide what's called an **analysis** of knowledge. (Here and from now on, whenever you see the word 'knowledge', you should take it to mean 'propositional knowledge'.) What's meant by an analysis is perhaps best approached via an example, the so-called **Justified True Belief** or **JTB** analysis, which holds that S knows that P if and only if:

1. It is true that P;
2. S believes that P; and
3. S is justified in believing that P.

This says that three conditions—known as the truth, belief, and justification conditions[1]—are individually **necessary** for S knowing that P (S *doesn't* know that P if *any* of them *doesn't* obtain) and are jointly **sufficient** for it (S *does* know that P if *all* of them *do* obtain). Plausibly, the analysis is also **non-circular**: each of the three analysing conditions can be understood independently of (and antecedently to) the analysed condition and the concept of knowledge.

We'll be looking at different analyses of knowledge, starting with the JTB analysis itself, which is subject to various sorts of challenge. The following exercises help you to think about these.

Q3. Briefly explain and assess the threat to the JTB analysis posed by the following examples.

> a. The ancient Greeks knew that the earth was the centre of the universe

> b. I know Newtonian mechanics, but Newtonian mechanics is false.

> c. I don't *believe* that Boris Johnson has resigned; I *know* that he has!

> d. I dreamt that the lottery numbers will be such and such, and they were: I knew it!

Q4. Consider the following example, taken from a classic article of epistemology, Edmund Gettier (1963) 'Is Justified True Belief Knowledge?' in *Analysis* **23**(6), pp. 121-3:

> Smith and Jones have applied for the same job. Smith has been told by the boss that Jones will get the job, and has just watched Jones count ten coins and put them in his pocket. On this basis, he forms the belief that Jones will get the job and has ten coins in his pocket, and infers that the person who will get the job has ten coins in his pocket. In fact, it is Smith himself who will get the job. Smith also happens to have ten coins in his pocket.

To which of the following claims is this a potential counter-example? Briefly explain your answer.

> a. S knows that P only if S is justified in believing that P.

> b. Having a justified true belief that P is a necessary condition for knowing that P.

> c. Having a justified true belief that p is a sufficient condition for knowing that P.

3. MIND & BODY

Another *General Philosophy* topic, **MIND & BODY**, returns to issues discussed in Descartes's *Meditations on First Philosophy*. Having raised a series of sceptical doubts in the First Meditation, Descartes's meditator goes on in the Second Meditation to establish that she can nevertheless be certain that, when she is thinking, she herself exists, and that she is a thinking thing. We will be looking at Descartes's views on the relationship between thinking things, **minds**, and their bodies.

Descartes holds that mind and body are **substances**. Now, the word 'substance' here is a philosophical term of art. In ordinary usage, it denotes a kind of matter with uniform properties, as when we talk of the wax-like substance (cutin) that is found on the upper surface of the leaves of a plant. But when philosophers like Descartes use it, they have something different in mind. The core idea is that a **substance** is an object or thing in a narrow sense, contrasting with both a **property** (such as the property of being wooden) and an **event** (such as the UK's formal exit from the EU at 11pm GMT on the 31st of January 2020). The idea is spelled out in different ways by different philosophers. For Descartes, substances are *capable of independent existence*, so that something is a substance if and only if there could be a world in which only it exists. It seems plausible, for example, that my desk is a substance on this view: there could be, it seems, a world containing nothing but my desk. The shape of my desk, however, is not a substance: any world containing the shape of my desk would also have to contain something *with* that shape.[2]

Descartes doesn't just hold that mind and body are substances. He also argues that they are *distinct* substances—as he puts it, that there is a "real distinction" between mind and body. One argument for this that is sometimes attributed to him is the following **argument from doubt**:

1. I can doubt that my (and any other) body exists (e.g. by appeal to the evil demon doubt)
2. But I cannot doubt that I, i.e. my mind, exists (for even if I am being deceived, I exist)
3. So I (my mind) am distinct from my (and any other) body

This argument is generally regarded to be fallacious. The main reason for this is that it seems subject to **parody**, a parody of an argument A being a structurally similar argument A*, the premises of which are at least as plausible as those of A but the conclusion of which is absurd.

Q5. Can you think of a parody of the argument from doubt? Hint: start by finding a conclusion that is structurally similar to the claim that I am distinct from my body but also clearly false.

Inasmuch as the argument from doubt is attributed to Descartes on the basis of anything he says in the *Meditations*, it is attributed on the basis of certain remarks in the Second Meditation. Seen in their proper context, however, it is fairly clear that Descartes was not actually offering the argument from doubt—not in the *Meditations*, at any rate: he seems to offer it in earlier work. In the *Meditations*, his arguments for the claim that mind and body are distinct substances are rather to be found in the Sixth Meditation. He presents his main argument there as follows:

> First, I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God. The question of what kind of power is required to bring about such a separation does not affect the judgement that the two things are distinct. Thus, simply by knowing that I exist and seeing at the same time that absolutely nothing else belongs to my nature or essence except that I am a thinking thing, I can infer correctly that my essence consists solely in the fact that I am a thinking thing. It is true that I may have (or, to anticipate, that I certainly have) a body that is very closely joined to me. But, nevertheless, on the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-extended thing; and on the other hand I have a distinct idea of body, in so far as this is simply an extended, non-thinking thing. And accordingly, it is certain that I am really distinct from my body, and can exist without it. (Descartes, *Meditations on First Philosophy with Selections from the Objections and Replies*, ed. by John Cottingham (Cambridge UP, 1996), p. 54.)

Q6. Present Descartes' argument as a *valid* argument with *two* premises. Hint: one of your premises should be a general claim that is argued for in the first half of the passage above.

4. PERSONAL IDENTITY

Many of the issues concerning a fourth *General Philosophy* topic, **PERSONAL IDENTITY**, were first raised by John Locke, a philosopher working a little after Descartes, in his *An Essay Concerning Human Understanding*, Bk. II, Ch. XXVII, 'Of Identity and Diversity'.

To better appreciate these issues, consider something that might at first seem unrelated, an apparent contrast between (clay) statues, like Rodin's thinker, and the lumps of clay from which they are made. On the face of it, these have different **criteria of identity**: what makes some statues—the statue in room 2 on Tuesday and the statue in room 4 on Wednesday, say—the *same* statue is not the same as what makes the lumps of clay from which they are made the same lumps of clay. For on the one hand, there seem to be changes that lumps of clay can survive and statues cannot. Suppose, for example, that just after it was made and before it had a chance to dry, Rodin's thinker was squished into a ball and refashioned into something else—a statue of the leaning tower of Pisa, perhaps. Intuitively, the lump of clay that it was made out of survives—it has been reshaped but not destroyed—but the original statue does not. On the other hand, there may also be changes that statues can survive and lumps of clay cannot. Suppose, for example, that after the thinker was completed, an arm fell off, and was replaced by a new arm, made out of an entirely different piece of clay. In this case, arguably, the statue survives, albeit with a new arm, and the lump of clay out of which it was originally made does not.

If this is right—if statues and lumps of clay have different criteria of identity—then it seems to follow that nothing can be *both* a statue *and* a lump of clay. For if a statue and a lump of clay have different criteria of identity, they have different qualities: one is able to survive changes that the other cannot. According to what is known as **Leibniz's law**, however, it is necessarily the case that, for any things x and y, if x and y are identical, they have the same qualities. Or what amounts to the claim, only put contrapositively[3]: necessarily, for any things x and y, if x and y have *different* qualities, they are *distinct*. If a statue and a lump of clay have different criteria of identity, it therefore seems to follow from Leibniz's law that nothing can be identical to both.

(Leibniz's law connects two different notions of identity. Consider Tweedledum and Tweedledee from Lewis Carroll's *Through the Looking Glass*. On the one hand, there is a sense in which they are identical: they look and act alike, sharing all the same qualities. This is what philosophers call **qualitative identity**. On the other hand, there is a sense in which they are *not* identical: although they share all the same qualities, Tweedledum and Tweedledee are *two* in number, not one. By contrast, J. K. Rowling and Robert Galbraith are *one* in number, 'Robert Galbraith' being a pen name of Rowling's. This is **numerical identity**. Leibniz's law, also known as the **indiscernibility of identicals**, states that things that are numerically identical must be qualitatively identical. [4])

Q7. "The lump of clay out of which Rodin's thinker was made has different qualities to the lump of clay out of which the statue of the leaning tower of Pisa was made: they differ in shape. So they must be numerically distinct lumps of clay." Present this as a *valid* argument, making explicit any implicit premises. Can you explain why it might be thought to show that the foregoing argument that nothing can be a statue and a lump of clay fails? Briefly assess whether it does.

We have, then, an argument for the conclusion that nothing can be both a statue and a lump of clay. Locke offers similar arguments, designed to show both (a) that nothing can be both a living organism, such as a human being, and the 'parcel' of matter out of which that organism is composed and, more strikingly, (b) that nothing can be both a living organism and a **person**.

By a person, Locke means "a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places": more briefly, a thinking thing, capable of reasoning, and conscious of itself and its mental states. It also suggests Locke's answer to the further question concerning the criterion of identity for persons, the question of what makes *some* persons—the person in room 2 on Monday and the person in room 3 on Tuesday, say—the *same* person: that "as far as this consciousness can be extended backwards to any past Action or Thought, so far reaches the Identity of that *Person*" (II.xxvii.9).

Locke's views about what a person is and what makes some persons the same person are reflections of his insistence that 'person' is a **forensic** term. To be a person, as Locke understands it, is to be the sort of thing that can be held legally and morally responsible for its actions, and to be the *same* person as some person P is to be legally and morally responsible for P's actions. Locke interest in legal and moral responsibility was driven by his interest in questions arising out of the Christian doctrine of the resurrection of the body on the day of judgement, and in particular what it is for the person whose body stands before God on the day of judgement to be the person who is responsible for such and such sins.

Locke is generally taken to think that what makes some persons the same person, if at all, is the fact that one remembers doing and experiencing the actions and experiences of the other. The person in room 3 on Tuesday is thus, on Locke's account, identical with the person in room 2 on Monday if she can remember performing the actions of the person in room 2 on Monday and experiencing her experiences. If, however, she *cannot* remember performing those actions or experiencing those experiences then, on this account, she is *not* identical with that person. So understood, his account has been subject to various objections. One of these objections was raised by the Scottish philosopher Thomas Reid, and is based on the following example:

> Suppose a brave officer to have been flogged when a boy at school for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life; suppose, also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that, when made a general, he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging. (Reid, *Essays on the Intellectual Powers of Man*, p. 333).

Q8. Explain how Locke's theory of personal identity seems to give contradictory verdicts about Reid's example. Hint: you will want to exploit the fact the relation, *x is the same person as y* is a transitive relation (on any set) but the relation *x remembers experiencing y's experiences* is not.

OTHER SET TEXTS

CP, MP, and PP students have separate set texts, studied in Trinity. Since you don't study these until later on, it is less urgent that you read them now, but you may wish to do so anyway.

For CP students, the relevant texts are certain of Turing's papers on computability and intelligence. If you want to get a head start on these, the book to get is Charles Petzold (2008) *The Annotated Turing* (Wiley).

For MP students, the relevant text is Frege's *Foundations of Arithmetic*. If you want to get hold of a copy now, you should make sure it is Gottlob Frege (1980) *The Foundations of Arithmetic*, rev. 2nd edition, transl. and ed. by J. L. Austin (Blackwell). (A recent translation by Dale Jacquette is more easily obtained, but it is terrible.)

For PP students, the relevant text the Leibniz-Clarke Correspondence. If you want to get hold of a copy, the edition to get is R. G. Alexander, ed. (1977) *The Leibniz-Clarke Correspondence* (Manchester UP).

---

1. The justification condition is sometimes stated slightly differently, as follows: S's belief that P is justified. Could one version of the condition be satisfied without the other being satisfied too? ↩
2. The distinction between substances and their dependents lies behind Lewis Carroll's joke of the Cheshire cat's smile in *Alice in Wonderland*. There cannot be a smile without something that is smiling. In the story, however, the Cheshire cat's smile is the last part of the cat to disappear. ↩
3. The **contrapositive** of a conditional claim of the form, *if P then Q*, is the corresponding conditional claim of the form, *if not-Q then not-P*. So the contrapositive of the claim *if x and y are identical then x and y have the same qualities* is the claim *if x and y do not have the same qualities then x and y are not identical*. Conditional claims and their contrapositives are equivalent in the sense that, necessarily, if either is true, so too is the other. ↩
4. Leibniz also endorsed another law of identity, the so-called **identity of indiscernibles**. This is the **converse** of the indiscernibility of identicals, stating that things that are qualitatively identical must be numerically identical. Unlike the first law, which is accepted by almost everyone in some form or another, this second law is very controversial. If it is correct, Tweedledum and Tweedledee aren't really possible—or insofar as they are, they aren't really qualitatively identical: there is some qualitative difference between them. ↩