PHILOSOPHY PREPARATORY WORK

PPE, PPL, and PML version

OVERVIEW

During the autumn term—or Michaelmas term, in Oxford-speak—Philosophy first years at Teddy Hall study *Introduction to Logic* and some *General Philosophy*. PPE, PPL, and PML students also study some *Moral 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 September. 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 at least Halbach's *The Logic Manual* and Mill's *Utilitarianism*. These can often be picked up for cheap second-hand. Bear in mind also that Teddy Hall offers grants of up to £300 covering your essential course materials, up to £100 of which may be used to purchase books.

Jonny McIntosh, July 2026

jonathon.mcintosh@philosophy.ox.ac.uk

INTRODUCTION TO LOGIC

PPE, PPL, and PML students have eight weeks of *Introduction to Logic* in Michaelmas. Each week there will be a lecture presenting that week's material. This is followed by a back-up class later in the week, in which you will be discussing solutions to exercises that you will have submitted beforehand.

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

PPE, PPL, and PML students may also study some *General Philosophy* in Michaelmas. 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. You will have tutorials on some of these topics in Michaelmas and/or Hilary—Hilary being Oxford-speak for the spring term. There are also weekly Faculty lectures for *General Philosophy* in Michaelmas, usually on Wednesdays.

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.

MORAL PHILOSOPHY

In addition to *Introduction to Logic* and *General Philosophy*, first year Philosophy students also study a set text. For PPE, PPL, and PML students, this is Mill's *Utilitarianism*, studied for what is called the *Moral Philosophy* component of the first year course. You will have tutorials on this component of the course in Hilary and/or Trinity. There are also weekly Faculty lectures in Michaelmas for *Moral Philosophy*, usually on Fridays.

You should make sure you have your own copy of Mill's *Utilitarianism*, and read through it in its entirety before you arrive. There are various editions of it available. I particularly like the Oxford Philosophical Texts edition, edited by Roger Crisp, but as it doesn't need translating, any edition will do. You should also read through the notes below, working through the various **EXERCISES**. Again, some of the questions presume familiarity with the concept of (logical) validity introduced in the **LOGIC EXERCISES** above. You should therefore complete those before tackling these.

MORAL PHILOSOPHY EXERCISES

1. THEORIES OF WELL-BEING

Mill articulates and defends positions on various issues in *Utilitarianism*. One of these is the issue of **well-being**, of what makes one's life *go well* for one. Philosophers in the utilitarian tradition to which Mill belongs—Jeremy Bentham, in particular—typically defended **hedonism** about well-being, the view that what makes life go well is *pleasure*, and the absence of pain. Simplifying a little, and putting the view in comparative form:

> **Hedonism about Well-Being**
> Life X goes better than life Y if and only if, and because, X is more pleasurable Y.

At first sight, Mill also defends hedonism about well-being. The issue is complicated, however, by his response to an important objection, the so-called **doctrine of swine** objection. Roughly put, this is the objection that hedonism about well-being offers a conception of the good life that is degrading, recognising no higher purpose than that which achievable by animals, such as pigs.

The objection is nicely illustrated by a thought experiment offered by Oxford's Roger Crisp. Consider the contrast between the life of Haydn, the Austrian composer, and an oyster. Haydn lived for 77 years, experiencing both great highs and lows. The oyster, on the other hand, leads a fairly monotonous life, sitting in a flow of warm water, gently absorbing nutrients. Intuitively, no matter how long the oyster lives, its life does not go as well as Haydn's. However, hedonists about well-being seem to be committed to the view that, so long as it lives long enough, the oyster's life goes better than Haydn's. This is because they typically hold that how pleasurable a pleasurable experience is depends only on two factors, its intensity and its duration. Suppose, then, the oyster's experience of sitting in warm water, gently absorbing nutrients, is mildly pleasurable (i.e. pleasurable, but of a low intensity). Then it seems that, simply by extending the *duration* of that experience, we can make the oyster's life as pleasurable as we like. In particular, we can make it *more* pleasurable than Haydn's. If so, however, the hedonist about well-being is committed to the view that the oyster's life *goes better* than Haydn's. But surely it doesn't!

Mill's response to the doctrine of swine objection is to draw a distinction between what he calls **higher pleasures**, like writing a symphony, and **lower pleasures**, like sitting in a nice, warm bath. Higher pleasures, he suggests, are more pleasurable than lower pleasures, and are such that they remain more pleasurable than lower pleasures, no matter how long or how intense the latter are. Analogously, consider a card game in which some hands are worth more than others. How valuable a given hand is depends in part on the numerical values of the individual cards, with an ace being worth 1, a jack 11, a queen 12, and a king 13. Suppose also, however, that hands containing cards of a certain suit are more valuable than any hands that don't contain cards of that suit—

perhaps that hands containing hearts are more valuable than ones that don't. In such a game, a hand containing a card of the relevant suit is worth more than a hand that doesn't, no matter what cards are in the latter. So too, on Mill's view, a life containing a higher pleasure is more pleasurable than a life containing any amount of merely lower pleasures. As you might see in tutorials on this topic, this seems to enable Mill to avoid the doctrine of swine objection but, according to some philosophers, it does so only at the cost of forcing him to abandon hedonism about well-being.

A second objection to hedonism about well-being arises out of a thought experiment offered by Robert Nozick in his *Anarchy, State, and Utopia* (Blackwell, 1974). Nozick invites us to imagine an **experience machine**. It can be programmed in such a way that anyone plugged into it has an experience that is subjectively indistinguishable from any experience you might like. If you would like to experience what it is like to climb Mount Everest, for example, it can be programmed in such a way that, by plugging into it, you will have a "virtual" experience that will seem to you exactly as it would if you were to actually have climbed Mount Everest. Of course, by plugging into the machine, you *don't* actually climb Mount Everest. You merely have an experience which is **subjectively indistinguishable** from the one you would have if you did climb it: things seem the same to you when you have the plugged-in experience as they would do were you to have the genuine experience. (In this respect, experiences had inside the experience machine are the same as the experiences had by the brain in a vat.) But this is all that we need to get a problem for hedonism about well-being up and running. For it seems that hedonists about well-being have to say lives lived inside the machine go as well as lives outside.

Roger Crisp has provided a nice illustration of this worry, too. Consider Ahmed and Bina. Ahmed lives his live outside the machine. It's a nice life, full of various accomplishments, as well as the occasional disappointment. Bina, by contrast, lives the entirety of her life inside the machine, which has been carefully programmed to deliver an exact simulacrum of Ahmed's life: for every experience Ahmed has, Bina undergoes a corresponding experience which is subjectively indistinguishable from it. Intuitively, Ahmed's life goes better than Bina's. After all, he really gets to do the various things he seems to do. Bina, by contrast, merely seems to do them. However, hedonists about well-being seem to be committed to the view that Bina's life goes just as well as Ahmed's. For Bina's life is subjectively indistinguishable from Ahmed's. But if it is subjectively indistinguishable from it, doesn't it follow that it must just as pleasurable as Ahmed's life? We'll examine how hedonists about well-being can respond to this objection, as well as whether alternative theories are better placed to account for the differences between Ahmed and Bina.

Q1. Set out the objection inspired by Haydn and the oyster as a *valid* argument with three premises and the conclusion, *Hedonism about well-being is not true*, and identify which premise(s) of your argument Mill is rejecting in drawing his higher/lower pleasure distinction.

Hint: one of your premises should spell out what hedonists about well-being are committed to saying about the example of Haydn and the oyster, and take the form, *If hedonism about well-being is true then, if P, Q*. Another will be of the form, *P*, and the last the form, *Not-Q*.

Q2. Set out the objection inspired by Ahmed and Bina as *valid* argument with three premises and the conclusion, *Hedonism about well-being is not true*, and assess the prospects for defending hedonism by rejecting each of the premises of your argument in turn.

2. ACT- vs. RULE-UTILITARIANISM

It is not controversial—and will come as no surprise, given the title of the book—that Mill endorses **utilitarianism** in some form or other. However, what *is* controversial is the exact form his utilitarianism takes. Is he an **act-utilitarian**, holding that an act is right if and only if, and because, it, of the options available, would produce the greatest increase in well-being? Or is he rather a **rule-utilitarian**, holding that an act is right if and

only if, and because, it conforms to a set of rules, general conformity to which would produce the greatest increase in well-being?

For a long time it was assumed that Mill was an act-utilitarian, but in the middle of the twentieth century some philosophers began to argue that he was in fact a rule-utilitarian. We'll examine their arguments, but we'll also think about which of the two approaches is more appealing.

To get a feel for the issues here, consider an example taken from a famous paper by James McCloskey. You are the sheriff of a frontier town in the Wild West. A terrible murder has been committed, and the townsfolk are convinced that a particular man was the culprit. You have taken this man into custody and discovered that he is not, in fact, the culprit. Unfortunately, an angry mob has surrounded the jail, demanding that you hand the man to them so that they can exact their revenge. You have no way of convincing them that the man did not commit the murder, but know that if you do not hand them over, a riot will start and many innocent people will be killed in the violence. What should you do? Hand over the man, allowing one innocent man to be killed, or refuse, and allowing many other innocent people to be killed?

The sheriff example poses a problem for act-utilitarianism, but rule-utilitarianism faces problems of its own. Consider the charge that it leads to irrational **rule worship**: cases in which it irrationally enjoins one to follow a rule where more good would be achieved by breaking it. Suppose, for example, that the set of rules general conformity to which would produce the greatest increase in well-being includes a rule against telling lies. In many cases, however, it seems more good could be achieved by breaking this rule. Take the famous example of a murderer at the door. If you follow the rule, an innocent person will be killed. If you lie, however, they will be spared. The rule-utilitarian might reply that we have simply made a mistake in supposing that the rules general conformity to which would produce the greatest increase in well-being includes a rule against telling lies. But won't a similar problem arise whatever the rules?

Q3. Explain why the example of the sheriff is a problem for act-utilitarianism, setting out the objection that it inspires as a *valid* argument with the conclusion, *Act-utilitarianism is not true*. Is rule-utilitarianism better placed to account for the example? Briefly explain why, or why not.

Q4. Some act-utilitarians have argued that rule-utilitarianism can only avoid the charge of rule worship if rule-utilitarianism collapses into act-utilitarianism. Explain and assess this objection.

---

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? ↩