

## 6 Deontic Logic

### 6.1 Permission and obligation

Deontic logic studies formal properties of obligation, permission, prohibition, and related normative concepts. The box in deontic logic is usually written ‘O’ (for ‘obligation’ or ‘ought’), the diamond ‘P’ (for ‘permission’). If we read  $q$  as stating that you cook dinner, we might use  $Oq$  to express that you are obligated to cook dinner.

We assume that obligation and permission are duals. You are not obligated to cook dinner iff you are permitted to not cook dinner; you are not permitted to cook dinner iff you are obligated to not cook dinner.

There are many kinds of norms: legal norms, moral norms, prudential norms, social norms, and so on. There may also be overarching norms that combine some or all of the others. Deontic logic is applicable to norms of all kinds. We do not have to settle whether  $O$  expresses legal obligation or moral obligation or some other kind of obligation. It is important, however, that we don’t equivocate. If the law requires  $q$  and morality  $\neg q$ , we should not formalize this as  $Oq \wedge O\neg q$ . It would be better to use a multi-modal language with different operators for legal and moral obligation.

Obligations and permissions often vary from agent to agent. If it is your turn to cook dinner then you are obligated to cook dinner, but I am not. To capture this agent-relativity, we could add agent subscripts to the operators, as we did in epistemic logic. We could then express our different obligations as  $O_1q \wedge \neg O_2q$ . But what does the sentence letter  $q$  stand for? When I say that you are obligated to cook dinner, the object of the obligation appears to be a type of act: cooking dinner. In the language of modal propositional logic,  $O$  and  $P$  are sentence operators. Unless we want to say that verb phrases in English (like ‘cook dinner’) should be translated into sentences of  $\mathcal{L}_M$  – which is possible, but non-standard – we have to transform the acts that appear to be the true objects of obligation and permission into propositions.

Consider sentence (1), which is arguably equivalent to (2).

- (1) You ought to cook dinner.
- (2) You ought to see to it that you cook dinner.

In (2), the operator ‘you ought to see to it that’ attaches to a sentence, ‘you cook dinner’. So we can translate (1) via (2) as  $O_1 q$ , where  $q$  translates ‘you cook dinner’, and  $O_1$  corresponds to ‘you ought to see to it that’.

The subject (you) is mentioned twice in (2). A common assumption in deontic logic is that we can drop the agent subscripts from deontic operators, since the embedded proposition will tell us upon whom the obligation or permission falls. Informally, the idea is that (2) is equivalent to (3), with an impersonal ‘ought’.

- (3) It ought to be the case that you cook dinner.

The impersonal ‘ought’ also figures in statements like (4).

- (4) Nobody ought to die of hunger.

When I say (4), I don’t mean that nobody is obligated to die of hunger. Nor do I mean that everybody is obligated to not die of hunger. Rather, I mean that a certain state of affairs – that nobody dies of hunger – ought to be the case. Without further assumptions, this does not impose any obligations on anyone.

There are reasons to question the equivalence between agent-relative ‘ought’ statements like (2) and impersonal ‘ought’ statements like (3). Suppose Amy has promised to play with Betty. Then Amy is obligated to play with Betty. But Betty is not thereby obligated to play with Amy. Betty may even have promised not to play with Amy. It is hard to express these facts in terms of impersonal oughts. If we say that it ought to be the case that Amy plays with Betty, we’re missing the fact that the obligation falls on Amy, not on Betty (who might be under a contrary obligation). So perhaps it would be better to keep the agent subscripts after all.

It can also be useful to make the ‘see to it that’ component in statements like (2) explicit. That Amy ought to play with Betty could then be translated as  $O_a \text{ STIT } p$ , where STIT formalizes ‘sees to it that’. This allows us to distinguish between the following three claims.

- |                            |   |
|----------------------------|---|
| $O_a \text{ STIT } \neg p$ | Amy ought to see to it that she doesn’t play with Betty.                  |
| $O_a \neg \text{STIT } p$  | Amy ought to not see to it that she plays with Betty.                     |
| $\neg O_a \text{ STIT } p$ | It is not the case that Amy ought to see to it that she plays with Betty. |

The STIT operator has proved useful to represent different concepts of rights and duties. In what follows, we will nonetheless stick to the simplest (and oldest) approach, without a STIT operator and without agent subscripts. This approach is sufficient for many applications, but its limitations should be kept in mind.

**Exercise 6.1**

Translate the following sentences into the standard language of deontic logic (without STIT or agent subscripts).

- (a) You must not go into the garden.
- (b) You may not go into the garden.
- (c) Jones ought to help his neighbours.
- (d) If Jones is going to help his neighbours, then he ought to tell them he's coming.
- (e) If Jones isn't going to help his neighbours, then he ought to not tell them he's coming.

## 6.2 Standard deontic logic

Think of a possible world as a history of events. For any such history, and any system of norms, we can ask whether the history conforms to the norms. Let's call a world *acceptable* relative to some norms if everything that happens at the world conforms to the norms. That is, a world is acceptable if it contains no violation of any relevant norm.

By definition, whatever happens at an acceptable world is permitted, in the sense that it does not violate any (relevant) norms. The converse is plausible as well: whenever something is permitted then it is the case at some acceptable world. For example, if it is permitted that Amy plays with Betty, then there should be a complete history of events in which Amy plays with Betty and no norms are violated. If there were no such history, then Amy's playing with Betty would logically entail the violation of some norms; but if an act entails the violation of some norms, then it is hard to see how the act could be permitted relative to these norms.

So we have the following connection between permission and acceptable worlds, which amounts to a possible-worlds analysis of permission:

$A$  is permitted (relative to some norms) iff  $A$  is the case at some possible world that is acceptable (relative to these norms).

Given the duality of permission and obligation, we also get a possible-worlds analysis of obligation:

$A$  is obligatory (relative to some norms) iff  $A$  is the case at all worlds that are acceptable (relative to these norms).

In logic, we are not interested in who is in fact obligated to do what, but in whether a given deontic statement is logically valid, or whether it logically follows from other statements.

Validity means truth in every conceivable scenario under every interpretation of the non-logical vocabulary. A scenario for deontic logic has to specify the relevant norms. This can be done by specifying which worlds are acceptable relative to which other worlds.

A Kripke model represents a scenario of this type, together with an interpretation of the sentence letters. In this application, a world  $v$  in the model is accessible from a world  $w$  if  $v$  is acceptable relative to the norms at  $w$  – equivalently, if everything that ought to be the case at  $w$  is the case at  $v$ . Worlds that are accessible from  $w$  in this sense are called **ideal** relative to  $w$ .

Our possible-worlds analysis of obligation and permission is reflected in definition 3.2, which settles under what conditions a sentence is true at a world in a model. Writing the box as ‘O’ and the diamond as P’, clause (g) of the definition states that  $OA$  is true at a world  $w$  in a model  $M$  iff  $A$  is true at all worlds of  $M$  that are ideal relative to  $w$ . Clause (h) states that  $PA$  is true at  $w$  in  $M$  iff  $A$  is true at some world that is ideal relative to  $w$ .

A sentence is valid iff it is true at every world in every suitable model. If we count all Kripke models as suitable, the logic of obligation and permission will be the minimal normal modal logic K. We can get stronger logics by imposing constraints on the accessibility relation. Let’s have a look at a few options.

We might stipulate that the deontic accessibility relation is reflexive, so that every world can see itself. This would make all instances of the (T)-schema valid:

$$(T) \quad OA \rightarrow A$$

In deontic logic, the (T)-schema is highly implausible. The fact that something ought to be the case does not entail that it is the case. Semantically speaking, many worlds are not ideal relative to themselves. We will not assume reflexivity.

We might, however, impose the weaker condition of seriality – that each world can see some world. This would validate principle (D):

$$(D) \quad O A \rightarrow P A$$

Intuitively, (D) says that the norms are consistent: if you're obligated to do  $A$ , then you are not obligated to do not- $A$ . (Remember that  $P A$  is equivalent to  $\neg O \neg A$ .) Semantically, (D) corresponds to the assumption that there is always at least one world at which all the norms are satisfied.

Without seriality, we have to allow for worlds from which no world is accessible. At such a world, all sentences of the form  $O A$  are true, and all sentences of the form  $P A$  are false. Everything is obligatory, but nothing is allowed. It is hard to make sense of such a situation. If we use Kripke semantics for deontic logic, we should rule out inconsistent norms and accept (D) as valid.

Here it may be important to distinguish *prima facie* obligations from *actual*, or *all-things-considered* obligations. If you've promised to cook dinner, you are under a *prima facie* obligation to cook dinner. But the obligation can be overridden by intervening circumstances or contrary obligations. If your child has an accident and needs urgent medical care, the right thing to do may well be to not cook dinner and instead bring your child to the hospital. In a sense, you are under conflicting obligations: you ought to cook dinner, and you ought to look after your child (and not cook dinner). There is no world at which you meet both of these obligations. But that is not a counterexample to (D), if we understand  $O$  as all-things-considered obligation. You are *prima facie* obligated to cook dinner, but all things considered, you should not cook dinner.

Let's return to the non-reflexivity of the deontic accessibility relation. Many things that are not the case nonetheless ought to be the case. Some have argued that this is only true in non-ideal worlds. In an ideal world, everything that ought to be the case is the case. By this line of thought, if a world  $v$  is accessible from some world  $w$  – meaning that  $v$  is ideal relative to  $w$  – then  $v$  should be accessible from itself. This condition is sometimes called “shift reflexivity” and corresponds to the

following schema (U) (for “utopia”)

$$(U) \quad O(OA \rightarrow A)$$

In words: it ought to be the case that whatever ought to be the case is the case.

The (U) principle is entailed by an alternative way of formalizing obligation and permission that goes back to Leibniz. Let ‘N’ be a propositional constant whose intended meaning is that all norms are satisfied, no obligations violated. Suppose we add this expression to  $\mathcal{L}_M$ , and we interpret the box of  $\mathcal{L}_M$  as a suitable kind of circumstantial necessity. Leibniz’s idea was that  $OA$  is definable as  $\Box(N \rightarrow A)$ : it ought to be that  $A$  iff, necessarily,  $A$  is the case whenever all obligations are met. It is not hard to show that if the (T)-schema is valid for the circumstantial box, and  $OA$  is defined as  $\Box(N \rightarrow A)$ , then the (U)-schema is valid for  $O$ .

### Exercise 6.2

- (a) Translate the (U)-schema into the Leibnizian language just proposed.
- (b) Give a tree proof for the translated (U)-schema, using the T-rules for the box.

### Exercise 6.3

How could we define  $P$  in terms of  $\Box$  and  $N$ , so that  $P$  is the dual of  $O$ ?

Turning to more familiar schemas and frame conditions, what shall we say about transitivity and euclidity, and the corresponding schemas (4) and (5)?

$$(4) \quad OA \rightarrow OOA$$

$$(5) \quad PA \rightarrow OPA$$

If something ought to be the case, ought it to be the case that it ought to be the case? If something is permitted, is it obligatory that it is permitted? Iterations of deontic operators sound strange in ordinary language. But they have a well-defined meaning in our Kripke semantics. The validity of (4) would mean that whenever something is obligatory at a world, then it is also obligatory at all ideal alternatives to that world. (5) would mean that if something is permissible at a world, then it’s

also permissible at all ideal alternatives to that world. On the background of (D), these two assumptions together imply that for each world there is a class of ideal worlds all of which are ideal relative to one another.

To get a clearer grip on whether that is plausible, we need to clarify how obligations and permissions can vary from world to world.

One obvious sense in which norms can vary across worlds is that people subscribe to different norms at different worlds. In our world, UK traffic law requires driving on the left, and most people think it is morally wrong to torture animals for fun. At other worlds, the laws and attitudes are different.

Let  $v$  be a world at which the traffic laws require driving on the right, and at which everyone thinks it is fine to torture animals. Suppose Norman at  $v$  is torturing kittens, while driving on the right (in the UK). Is Norman doing something that's morally wrong? Is he doing something that violates the traffic laws? The answer depends on whether we evaluate Norman's acts relative to our norms – the norms at our world – or relative to the norms at Norman's world. Both perspectives are intelligible. They lead to different deontic logics.

On an **absolutist** conception, the basic norms do not vary from world to world. Whichever world we look at, we always assess it relative to the same set of norms. On this conception, it is natural to assume that the very same worlds are ideal relative to any world: a world will be accessible from any world just in case it contains no violation of the (fixed) norms. The resulting logic of obligation and permission is KD45.

#### Exercise 6.4

Explain why the deontic accessibility relation is transitive and euclidean if the same worlds are ideal relative to any world.

#### Exercise 6.5

Show that euclidity implies shift reflexivity.

On a **relativist** conception of norms, we evaluate the events at other worlds relative to the norms at these worlds. Transitivity and euclidity now become implausible, as does shift reflexivity. To see why, add another world  $u$  to the Norman scenario. The laws at  $u$  say that one must drive on the right. But the inhabitants of  $u$  are rebellious:

everyone at  $u$  drives on the left. Nothing that happens at  $u$ , we may assume, violates the traffic laws of our world. So  $u$  is deontically accessible from the actual world. But if we evaluate the events at  $u$  relative to the laws at  $u$ , then much of what happens at  $u$  violates the norms, so  $u$  is not deontically accessible from itself. Shift reflexivity fails.

#### Exercise 6.6

Explain why deontic accessibility is neither transitive nor euclidean, on the relativist conception.

The relativist conception is more common in deontic logic. So-called **standard deontic logic** assumes only that the accessibility relation is serial, making the system D the complete logic of obligation and permission.

The proposed logics of absolutism and relativism only disagree about sentences in which a deontic operator occurs in the scope of another deontic operator. Any sentence that does not contain an O or P operator embedded under another O or P operator is D-valid iff it is KD45-valid.

#### Exercise 6.7

Use the tree method to check which of the following sentences are D-valid and which are KD45-valid.

- (a)  $P(p \vee q) \rightarrow (Pp \wedge Pq)$
- (b)  $OPp \rightarrow Pp$
- (c)  $\neg P(p \vee q) \rightarrow (P\neg p \vee P\neg q)$
- (d)  $OPp \vee PO p$

#### Exercise 6.8

Consider a world in which there are no sentient beings, and nothing else that could introduce norms or laws. Since there are no norms at this world, one might hold that nothing is obligatory relative to the world's norms, and nothing is permitted. Explain why this casts doubt on the validity of (Dual1) and (Dual2) in the logic of relativist obligation and permission.



**Exercise 6.9**

Amy ought to have either promised to help Betty or to help Carla. She hasn't made either promise. If she had promised to help Betty, she would be obligated to help Betty. If she had promised to help Carla, she would be obligated to help Carla. So it ought to be the case that Amy is either obligated to help Betty or obligated to help Carla. In fact, since Amy made neither promise, she is neither obligated to help Betty nor to help Carla. Explain why this casts doubt on the assumption that deontic accessibility is euclidean.

### 6.3 Norms and circumstances

The possible-worlds analysis from the previous section assumes that something ought to be the case iff it is the case at all ideal worlds, where no norms are violated. Many ordinary statements about oughts and obligations do not fit this analysis.

Suppose you are walking past a drowning baby. You ought to save the baby. But are you saving the baby at every world at which no norms are violated? Clearly not. There are worlds at which the baby never fell into the pond, and others at which you are overseas and have no means to rescue the baby. These worlds need not involve any violations of norms.

Whether something ought to be the case depends not just on the norms but also on the circumstances. Under circumstances in which you have the opportunity to save a drowning baby, you ought to save it. Under other circumstances you do not.

We can account for the dependence of obligations on circumstances by changing our interpretation of the accessibility relation. Previously, we assumed that a world  $v$  is accessible from  $w$  iff all the norms at  $w$  are respected at  $v$ . On the new interpretation, we also require that the relevant circumstances at  $w$  are preserved at  $v$ . If  $w$  is a world at which you come across a drowning baby then any accessible world will also be a world at which you come across a drowning baby.

As a first stab, we might redefine deontic accessibility as follows:

A world  $v$  is deontically accessible from a world  $w$  iff (a) the relevant circumstances at  $w$  also obtain at  $v$ , and (b) no norms from  $w$  are violated at  $v$ .

I use ‘relevant circumstances’ as a placeholder for the circumstances we hold fixed when we consider what ought to be the case. Often we hold fixed everything that is *settled* in the sense we studied in section 1.5 – everything that can no longer be changed. If the baby has fallen into the pond at  $w$ , then there is nothing anyone can do to undo the falling; the falling is a “relevant circumstance” that takes place at every world accessible from  $w$ .

Clause (b) in the above definition assumes that no norms are violated at any accessible world. But if accessibility is restricted by circumstances, then this is implausible because the relevant circumstances will often involve violations of norms.

The problem is brought about by Arthur Prior’s “Samaritan Paradox”. Suppose someone has been injured in a robbery, and Jones has the opportunity to help. We want to say that Jones ought to help the victim. On the possible-worlds analysis of ‘ought’, this means that Jones helps the victim at all worlds accessible from the actual world. It follows that the robbery took place at all these worlds. (In a world without a robbery, there is no victim to help.) But then all the accessible worlds contain a violation of norms. In a truly ideal world, nobody would have been robbed and injured.

In the Samaritan Paradox, the robbery is settled; it has happened at all worlds that are compatible with the “relevant circumstances”. None of these worlds are ideal. Among these worlds, however, worlds at which Jones doesn’t help the victim are even *worse*, in terms of norm violations, than worlds at which he helps the victim. Both kinds of worlds are bad, because the victim got robbed. But our norms don’t just divide the possible worlds into good and bad; they allow for finer distinctions between bad worlds and even worse worlds. Jones ought to help the victim because that’s what he does in the *best* worlds among those he can bring about, even though none of these worlds are ideal.

So here is a second pass at the revised definition of deontic accessibility.

A world  $v$  is deontically accessible from a world  $w$  iff (a) the relevant circumstances at  $w$  are also the case at  $v$ , and (b)  $v$  is one of the best worlds, by the norms at  $w$ , among worlds at which the relevant circumstances from  $w$  are the case.

The revised accessibility relation combines circumstantial and purely deontic conditions. It can be useful to separate these two components. To this end, let’s first add a circumstantial accessibility relation to our models. In addition, a model needs to

specify which worlds are better than others, relative to the norms at any given world (which may be the norms at every world, on an absolutist approach).

Let ' $u <_w v$ ' mean that world  $u$  is better than world  $v$  relative to the norms at  $w$ . The symbol ' $<$ ' hints at the idea that  $u$  contains *fewer* violations of norms than  $v$ . We assume that for any world  $w$ , the relation  $<_w$  is transitive. We also assume that it is asymmetric, meaning that if  $u <_w v$  then it is not the case that  $v <_w u$ . Asymmetric and transitive relations are known as **strict partial orders**.

### Definition 6.1

A **deontic ordering model** consists of

- a non-empty set  $W$  (the worlds),
- a binary relation  $R$  on  $W$  (the circumstantial accessibility relation),
- for each world  $w \in W$ , a strict partial order  $<_w$  on  $W$  (the world-relative ranking of worlds as better or worse), and
- a function  $V$  that assigns to each sentence letter of  $\mathcal{L}_M$  a subset of  $W$ .

Now we need to say under what conditions a sentence of the form  $OA$  is true at a world in an ordering model. Informally,  $OA$  will be true at  $w$  iff  $A$  is true at the best worlds among those that are circumstantially accessible. Let's introduce one more piece of notation. For any set of worlds  $S$  and any partial order  $<$ , let  $Min^<(S)$  be the set of  $<$ -minimal members of  $S$ :

$$Min^<(S) =_{\text{def}} \{v : v \in S \wedge \neg \exists u (u \in S \wedge u < v)\}.$$

An expression of the form ' $\{x : \dots x \dots\}$ ' denotes the set of all things  $x$  that satisfy the condition  $\dots x \dots$ . So  $Min^<(S)$  is the set of all things  $v$  that are members of  $S$  and for which there are no members  $u$  of  $S$  for which  $u < v$ .

Here, then, are the truth-conditions for  $OA$  and  $PA$  in deontic ordering models:

### Definition 6.2: Ordering semantics

If  $M$  is a ordering model and  $w$  a world in  $M$ , then

$$M, w \models OA \quad \text{iff} \quad M, v \models A \text{ for all } v \in Min^{<_w}(\{u : wRu\})$$

$$M, w \models PA \text{ iff } M, v \models A \text{ for some } v \in \text{Min}^{<_w}(\{u : wRu\})$$

This is just a formal way of saying that  $OA$  is true at  $w$  iff  $A$  is true at the best worlds (by the norms at  $w$ ) among the worlds that are circumstantially accessible at  $w$ .

If we want the (D)-schema to be valid, we have to assume that there is always at least one best world among the circumstantially accessible worlds, so that  $\text{Min}^{<_w}(\{u : wRu\})$  is never empty. Let's make this assumption.

The logic of obligation and permission now depends on formal properties of the circumstantial accessibility relation  $R$  and the deontic orderings  $<_w$ . In section 1.5, I argued that the logic of historical necessity (of what is settled and open) is S5. This suggests that in normal contexts,  $R$  is an equivalence relation. If we adopt an absolutist approach, on which the orderings  $<_w$  are the same for every world  $w$ , we then still get KD45. If we allow the orderings to vary from world to world, we still get D, unless we impose further restrictions on the orderings.

#### Exercise 6.10

Suppose fatalism is true and the only world that is open (circumstantially accessible) relative to any world  $w$  is  $w$  itself. Can you describe the resulting deontic logic (on either an absolutist or a relativist approach)?

Ordering models prove useful when we want to formalize statements with modal operators and if-clauses, like (1)–(3).

- (1) If you smoke then you must smoke outside.
- (2) If you miss the deadline for tax returns then you must pay a fine.
- (3) If you have promised to call your parents then you must call them.

How would you translate these into our language  $\mathcal{L}_M$ ? We seem to face a choice between (W) and (N).

(W)  $O(p \rightarrow q)$

(N)  $p \rightarrow Oq$

In (W), the operator  $O$  is said to have **wide scope** because it applies to the entire conditional  $p \rightarrow q$ . In (N), the operator has **narrow scope** because it only applies to the consequent  $q$ .

On reflection, neither translation is satisfactory. Starting with (N), note that  $p \rightarrow Oq$  and  $\neg Oq$  together entail  $\neg p$ . But from (1), together with the assumption that you are not required to smoke ( $\neg Oq$ ), we surely can't infer that you do not in fact smoke.

(W) is not much better. For one, in our Kripke-style semantics,  $O(p \rightarrow q)$  is entailed by  $O(\neg p)$ . But it is easy to imagine a scenario in which you must not smoke, or you must submit your tax return before the deadline, but in which (1) and (2) are false.

Another problem with both (N) and (W) is that they would license a problematic form of “strengthening the antecedent”. For example, they both suggest that (3) entails (4).

- (4) If you have promised to call your parents and you know that someone has attached a bomb to your parents' phone that will go off if you call, then you must call them.

#### Exercise 6.11

Give tree proofs with the K-rules to show that  $p \rightarrow Or$  entails  $(p \wedge q) \rightarrow Or$ , and that  $O(p \rightarrow r)$  entails  $O((p \wedge q) \rightarrow r)$ .

Let's think about what is expressed by statements like (1)–(4). Intuitively, when we ask what must be done if  $p$  is the case, we are limiting our attention to situations in which  $p$  is the case, and consider which of *these* situations best conform to the relevant norms. It is irrelevant whether  $p$  is in fact the case or whether it ought to be the case. (1) says – roughly – that among worlds where you smoke, the “best” worlds are worlds where you smoke outside. Worlds where you smoke inside are worse than worlds where you smoke outside. Similarly for (2). A world at which you miss the deadline for tax returns and pay the fine contains only one violation of the tax rules. Worlds at which you miss the deadline and don't pay the fine contain two. The “best” worlds among those at which you miss the deadline are worlds at which you pay the fine. Likewise for (3). Among worlds at which you have promised to call your parents, the “best” are worlds at which you keep the promise and call them.

The if-clause in sentences like (1)–(3) therefore seems to *restrict* the worlds over which the modal operator quantifies. Whereas ‘ought  $q$ ’ alone says that  $q$  is true at the best of the open worlds, ‘if  $p$  then ought  $q$ ’ says that  $q$  is true at the best of the open worlds *at which  $p$  is true*.

There is no way to express these truth-conditions with the resources of  $\mathcal{L}_M$ . But we can introduce a new, binary operator for **conditional obligation**. The operator is often written ‘ $O(\cdot/\cdot)$ ’, with a slash separating the two argument places. Intuitively,  $O(B/A)$  means that  $B$  ought to be the case if  $A$  is the case.

The formal truth-conditions for  $O(B/A)$  are much like those for  $O B$ , except that we add the assumption  $A$  to the circumstances that are held fixed:

**Definition 6.3: Ordering semantics for conditional obligation**

If  $M$  is a ordering model and  $w$  a world in  $M$ , then

$M, w \models O(B/A)$  iff  $M, v \models B$  for all  $v \in \text{Min}^{<w}(\{u : wRu \text{ and } M, u \models A\})$ .

Here,  $\{u : wRu \text{ and } M, u \models A\}$  is the set of worlds  $u$  that are circumstantially accessible from  $w$  and at which  $A$  is true.  $\text{Min}^{<w}(\{u : wRu \text{ and } M, u \models A\})$  is the set that comprises the best of these worlds. So  $O(B/A)$  is true at  $w$  iff  $B$  is true at all of the best  $A$ -worlds that are accessible at  $w$ .

**Exercise 6.12**

“Deontic detachment” is the inference from  $O A$  and  $O(B/A)$  to  $O B$ . “Factual detachment” is the inference from  $A$  and  $O(B/A)$  to  $O B$ . Which of these are valid on the present semantics?

**Exercise 6.13**

In exercise 6.1, you were asked to translate the following statements.

- (c) Jones ought to help his neighbours.
- (d) If Jones is going to help his neighbours, then he ought to tell them he’s coming.
- (e) If Jones isn’t going to help his neighbours, then he ought to not tell them he’s coming.

Let's add a fourth statement:

(f) Jones is not going to help his neighbours.

Intuitively, none of these four statements is entailed by one of the others. Moreover, they don't impose contradictory requirements on Jones: it is easy to think of a scenario in which they are all true and Jones is not obligated to perform some act and also obligated to not perform the act. This shows that your translations in exercise 6.1 were incorrect. Explain. (This puzzle is due to Roderick Chisholm.)

#### Exercise 6.14

The dual of conditional obligation is conditional permission. Spell out truth-conditions for  $P(B/A)$  that parallel the truth-conditions I have given for  $O(B/A)$ , so that  $P(B/A)$  is equivalent to  $\neg O(\neg B/A)$ .

## 6.4 Further challenges

Many apparent problems for standard deontic logic arise from the dependence of obligations on circumstances. We can avoid these problems by using deontic ordering models and formalizing conditional obligation statements with the binary  $O(\cdot/\cdot)$  operator. There are, however, other problems and "paradoxes" for which this move doesn't help. I will mention three.

First, we already saw that standard deontic logic does not allow for conflicting obligations. Suppose you have promised your family to be home for dinner and your friends to join them at the pub. You are under conflicting *prima facie* obligations. It is not clear that one of them overrides the other. Legal systems can also contain contradictory rules, without any higher-level rules for how to resolve such contradictions.

We can, of course, drop principle (D). But even in the minimal logic K,  $O p$  and  $O \neg p$  entail  $O A$ , for any sentence  $A$ . Intuitively, however, the fact that you have given incompatible promises does not entail that you are obligated to, say, kill the Prime Minister.

Another family of problems arises from the fact that in any logic defined in terms of Kripke models,  $O$  is closed under logical consequence, meaning that if  $O A$  is

true and  $A$  entails  $B$ , then  $O B$  is true. Since logical truths are logically entailed by everything, it follows that all logical truths come out as obligatory. (This is easy to see semantically. A logical truth is true at all worlds; so it is true at all deontically accessible worlds.) But ought it to be the case that it either rains or doesn't rain?

In response, one might argue that the relevant statements sound wrong not because they are false, but because their utterance would violate a pragmatic norm of cooperative communication. A basic norm of pragmatics is that utterances should make a helpful contribution to the relevant conversation. In a normal conversational context, it would be pointless to say that something ought (or ought not) to be the case if it is logically guaranteed to be the case anyway. An utterance of 'it ought to be that  $p$ ' is pragmatically appropriate only if  $p$  could be false. This might explain why it sounds wrong to say that it ought to either rain or not rain.

Note also that by duality,  $\neg O(p \vee \neg p)$  entails  $P\neg(p \vee \neg p)$ . If we deny that it ought to either rain or not rain, and we accept the duality of obligation and permission, we have to say that it is permissible that it neither rains nor doesn't rain. That sounds even worse.

The problem of closure under entailment has special bite when obligation statements are restricted by circumstances. Return to the Samaritan puzzle. Suppose the victim is bleeding, and Jones ought to stop the blood flow. It is logically impossible to stop a blood flow if no blood is flowing. In all the deontic logics we have so far considered, the claim that Jones ought to stop the victim's blood flow therefore entails that the victim ought to be bleeding. But wouldn't it be better if the victim weren't bleeding?

Here, too, one might appeal to a pragmatic explanation. When we say that Jones ought to stop the blood flow, we take for granted that the victim is bleeding. We are interested in what should be done *given* the state in which Jones found the victim. Worlds where the victim isn't injured are set aside; they are not circumstantially accessible. But circumstantial accessibility can shift with conversational context. The claim that the victim ought to be bleeding is pointless if we hold fixed the victim's state of injury. So when we evaluate *this* claim, we naturally assume that the relevant circumstantial accessibility relation does not hold fixed the injuries. Intuitively, we are no longer considering what should be done given the state in which Jones found the victim, but whether that state itself should have obtained. Worlds in which the state doesn't obtain become circumstantially accessible.

A third family of problems arises from disjunctive statements of permission and



obligation. Consider (1).

- (1) You ought to either mail the letter or burn it.

Intuitively, (1) suggests that both mailing the letter and burning it are permitted. In standard deontic logic, however,  $O(A \vee B)$  does not entail  $PA \wedge PB$ . (This puzzle was first noticed by Alf Ross and is known as “Ross’s Paradox”.)

A similar puzzle arises for permissions. (This one is known as the “Paradox of Free Choice”.)

- (2) You may have beer or wine.

Intuitively, (2) implies that beer and wine are both permitted. But in standard deontic logic,  $P(A \vee B)$  does not entail  $PA \wedge PB$ .

We could add the missing principles.

$$(R) \quad O(A \vee B) \rightarrow (PA \wedge PB)$$

$$(FC) \quad P(A \vee B) \rightarrow (PA \wedge PB)$$

But both of these have unacceptable consequences when added to the minimal modal logic K. With the help of (R), we could show that  $OA$  entails  $PB$ :  $OA$  entails  $O(A \vee B)$ , which by (R) entails  $PA \wedge PB$ . But clearly ‘you ought to mail the letter’ does not entail ‘you may burn the letter’. Similarly for (FC). In K,  $PA$  entails  $P(A \vee B)$ ; by (FC),  $P(A \vee B)$  entails  $PB$ . But ‘you may have beer’ does not entail ‘you may have wine’.

#### Exercise 6.15

Analogous puzzles to those raised by Ross’s Paradox and the Paradox of Free Choice arise for epistemic ‘must’ and ‘might’. Can you give examples?

## 6.5 Neighbourhood semantics

In reaction to apparent problems for standard deontic logic, some have argued that we should not interpret obligation and permission in terms of quantification over possible worlds. If we give up this core tenet of Kripke semantics, we can define “non-normal” logics weaker than K. (A **normal** modal logic is a modal logic that can be defined in terms of classes of Kripke frames.)

A popular alternative to Kripke semantic is **neighbourhood semantics**, also known as Scott-Montague semantics, after its inventors Dana Scott and Richard Montague.

Models in neighbourhood semantics still involve possible worlds. Validity is still defined as truth at all worlds in all (suitable) models. But the box and the diamond are no longer interpreted as quantifiers over accessible worlds. Instead, we simply assume that at every world, some propositions are “necessary” and others are not.  $\Box A$  is true at a world if  $A$  expresses one of the necessary propositions at that world.

Formally, the accessibility relation in Kripke models is replaced by a **neighbourhood function**  $N$  that associates each world in a model with the propositions that are necessary relative to  $w$ . Propositions are identified with sets of possible worlds. Thus  $N(w)$  is a set of sets of worlds. Each set of world in  $N(w)$  is necessary at  $w$ .

#### Definition 6.4

A **neighbourhood model** consists of

- a non-empty set  $W$ ,
- a function  $N$  that assigns to each member of  $W$  a set of subsets of  $W$ , and
- a function  $V$  that assigns to each sentence letter of  $\mathcal{L}_M$  a subset of  $W$ .

The interpretation of non-modal sentences at neighbourhood models works just as in Kripke semantics (definition 3.2). To state the semantics for modal sentences, let  $[A]^M$  be the set of worlds in model  $M$  at which  $A$  is true. This is our proxy for the proposition expressed by  $A$ . Then:

$$M, w \models \Box A \text{ iff } [A]^M \text{ is in } N(w).$$

$$M, w \models \Diamond A \text{ iff } [\neg A]^M \text{ is not in } N(w).$$

Intuitively, the clause for the box says that  $\Box A$  is true at  $w$  iff the proposition expressed by  $A$  is one of those that are necessary at  $w$ . The clause for the diamond ensures that the box and the diamond are duals.

In neighbourhood semantics, the modal operators are not closed under logical consequence. The neighbourhood function  $N$  can easily make  $p$  necessary at a world without making  $p \vee q$  necessary, even though  $p$  entails  $p \vee q$ . If we interpret  $O$  and  $P$  as the box and the diamond in neighbourhood semantics, we can therefore say that Jones ought to tend to the victim’s injuries even though it is not the case that

someone ought to be injured.

We can also allow for conflicting obligations. If the laws at  $w$  require both  $p$  and  $\neg p$ , we simply have  $[p]^M \in N(w)$  and  $[\neg p]^M \in N(w)$ . It longer follows that any proposition whatsoever is obligatory.

We may further hope to escape the problems from section 6.3 that led us to introduce a primitive conditional obligation operator. I argued that the wide-scope translation  $O(A \rightarrow B)$  of conditional obligation sentences is problematic because  $O(A \rightarrow B)$  is entailed by  $O(\neg A)$ . In neighbourhood semantics, this entailment fails.

Bare neighbourhood semantics determines a very weak logic called **E**. It is axiomatized by (Dual), (CPL), and a rule (called “RN”) that allows inferring  $\Box A \leftrightarrow \Box B$  from  $A \leftrightarrow B$ . We can get stronger logics, with more validities, by imposing conditions on the neighbourhood function  $N$ .

For example, suppose we want to maintain that if something is logically guaranteed to be true, then it can’t be forbidden. Equivalently, any logically necessary truth should be permitted. By the neighbourhood semantics for **P**,  $A$  is permitted at a world  $w$  in a model  $M$  iff  $[\neg A]^M$  is not in  $N(w)$ . If  $A$  is a logical truth, then  $A$  is true at all worlds; in that case,  $\neg A$  is true at no worlds, and  $[\neg A]^M$  is the empty set. If we want logical truths to be permitted, we therefore have to stipulate that  $N(w)$  never contains the empty set.

In Kripke semantics, the assumption that logically necessary truths are permitted is equivalent to the assumption that (every instance of) the (D)-schema  $O A \rightarrow P A$  is valid. Both assumptions correspond to seriality of the accessibility relation. In neighbourhood semantics, we can distinguish between the two assumptions. While the permissibility of logical truths requires that  $N(w)$  doesn’t contain the empty set, the validity of  $O A \rightarrow P A$  requires that  $N(w)$  doesn’t contains contradictory propositions  $[A]^M$  and  $[\neg A]^M$ .

If we assume that the neighbourhood function is closed under intersection, in the sense that whenever two sets  $X$  and  $Y$  are in  $N(w)$  then so is their intersection  $X \cap Y$ , then  $(\Box A \wedge \Box B) \rightarrow \Box(A \wedge B)$  becomes valid. If we also require the converse, that whenever  $X \cap Y \in N(w)$  then  $X \in N(w)$  and  $Y \in N(w)$ , and in addition that  $W \in N(w)$ , we get back the minimal normal logic **K**.

**Exercise 6.16**

Can you find a condition on the neighbourhood function that renders the (T)-schema valid?

For some purposes, even the minimal logic of neighbourhood semantics is too strong. Return to the intuitive “Free Choice” principle from the previous section:

$$(FC) \quad P(A \vee B) \rightarrow (P A \wedge P B)$$

We have seen that this principle is untenable in Kripke semantics. It is still untenable in neighbourhood semantics.

To see why, note first that whenever two sentences  $A$  and  $B$  are logically equivalent, then in neighbourhood semantics  $P A$  and  $P B$  are also equivalent. The reason is that the modal operators in neighbourhood semantics operate on the set of worlds at which the embedded sentence is true. If  $A$  and  $B$  are logically equivalent, then in any model  $M$ , the set  $[A]^M$  is the same set as  $[B]^M$ , and so  $[A]^M$  is in  $N(w)$  iff  $[B]^M$  is in  $N(w)$ . Likewise,  $[\neg A]^M$  is in  $N(w)$  iff  $[\neg B]^M$  is in  $N(w)$ .

Now any sentence  $A$  is logically equivalent to  $(A \wedge B) \vee (A \wedge \neg B)$ , for any  $B$ . In the logic E,  $P A$  therefore entails  $P((A \wedge B) \vee (A \wedge \neg B))$ . By (FC),  $P((A \wedge B) \vee (A \wedge \neg B))$  entails  $P(A \wedge B)$ . We could still reason from ‘you may have a cookie’ to ‘you may have a cookie and burn down the house’.

**Exercise 6.17**

Rational beliefs come in degrees, which are often assumed to satisfy the formal rules of probability. Suppose we say that someone believes  $A$  iff their degree of belief in  $A$  is above a certain threshold – say, 0.9. Explain why one can’t give a Kripke semantics for this concept of belief. (Although one can give a neighbourhood semantics.) *Hint:* One rule of probability says that if  $p$  and  $q$  are independent propositions, then the probability of their conjunction  $p \wedge q$  is the product of their individual probabilities.