EXERCISES, DISCUSSION PROMPTS AND WANDERINGS ON DEONTIC LOGIC

1 Semantic Characterization of Standard Deontic Logic (SDL)

SDL is the logic that results from restricting the class of relevant model in the definition of entailment to models based on a *serial* accessibility relation, where:

R is **serial** iff $\forall w \exists v(wRv)$

1a. Give an example of a frame with a serial accessibility relation featuring three worlds.

1b. Give an example of a frame with a serial accessibility relation with no loops and no worlds accessing themselves.

1c. Show that if a frame is *reflexive*, then it is serial.

2. Show that $\models_{SDL} \bigcirc A \rightarrow \mathbf{P}A$

2 Greatest hits

3a. Show that $\bigcirc A \wedge \bigcirc B \models_{SDL} \bigcirc(A \wedge B)$.

3b. Now give a derivation of the same argument in the axiomatic system—i.e. establish $\bigcirc A \wedge \bigcirc B \vdash_{SDL} \bigcirc(A \wedge B)$.

4. Show that $\mathbf{P}A \wedge \mathbf{P}B \not\models_{SDL} \mathbf{P}(A \wedge B)$.

5a. Show that $\bigcirc(A \wedge B) \models_{SDL} \bigcirc(A \vee C)$.

5b. Now give a derivation of the same argument in the axiomatic system.

6a. Show that $\bigcirc A \wedge \bigcirc \neg A \models_{SDL} \bot$

6b. Your argument in 4a should make an important appeal to the seriality of the accessibility relation R. Would the argument be valid if we could look at models in which R was allowed to be non-serial?

6c. Can you see a connection between $\bigcirc A \wedge \bigcirc \neg A \models_{SDL} \bot$ and $\models_{SDL} \bigcirc A \rightarrow \mathbf{P}A$? (Sorry for the deliberately vague question!)

6d. The set $\{\bigcirc A, \bigcirc \neg A\}$ is a special kind of set in which one's obligations are incompatible. Can you describe another pair of incompatible obligation that does not have this syntactic structure (that is, another set of the form $\{\bigcirc B, \bigcirc C\}$ in which C is not the negation of B).

6e. Does SDL have the general consequence that *whenever B and C are incompatible*, $\bigcirc B \wedge \bigcirc C \models_{SDL} \bot$?

7a. SDL does not validate the principle $\bigcirc A \rightarrow A$, i.e. $\not\models_{SDL} \bigcirc A \rightarrow A$. Show this by giving a model.

7b. Speaking informally, and on the basis of the characterization of the deontic accessibility relation, what would you say is the reason why it's appropriate that $\bigcirc A \rightarrow A$ fails in SDL?

7c. Give an example of a model that is based on a non-reflexive frame but such that $\bigcirc A \rightarrow A$ is nonetheless true everywhere in the model.

3 Deep cuts: shift-reflexivity

Shift-reflexivity is the condition that $\forall w \forall v(wRv \rightarrow vRv)$.
Informally: whenever a world v is accessed by some world w, then v accesses itself.

8a. Show that $\bigcirc(\bigcirc A \rightarrow A)$ is not SDL-valid.

8b. Show that $\bigcirc(\bigcirc A \rightarrow A)$ is valid in the class of shift-reflexive frames.

9. Give an example of a model that is not based on a shift-reflexive frame such that $\bigcirc(\bigcirc A \rightarrow A)$ is nonetheless true.

10. Find an equivalent way of expressing $\bigcirc(\bigcirc A \rightarrow A)$ using only the prohibition operator **F**.

11a. How might we justify requiring shift-reflexivity in deontic logic?

11b. What are some reasons why we might *not* want to require it?

4 Deep cuts: multi-agent relational models

Suppose we wanted to keep track separately of the obligations of multiple agents. Perhaps, we want to have two operators $\bigcirc_a$ and $\bigcirc_b$ tracking what each agent—a or b—ought to do. Call the joint logic of $\bigcirc_a$ and $\bigcirc_b$ $MADL$ for *Multi-Agent Deontic Logic*.

12. How would you modify the definition of relational models to characterize models for MADL?

13a. Intuitively — without thinking specifically about the models you designed in the previous exercise — do you expect the following entailment to hold?

$$\bigcirc_a(A) \wedge \bigcirc_b(A \rightarrow B) \models_{MADL} \bigcirc_a(B)$$

14b. If you think the entailment fails in the model theoretic analysis, give a counter-model to it based on your account of models for MADL. If you think it doesn't, give a proof of validity.

Now suppose you wanted to account for the *universal obligations* of a and b — maybe your language includes another operator $\bigcirc_{ab}$ for those things that *both* a and b are obliged to do. From now on, $MADL$ refers to the logic for this extended language.

15. To accommodate $\bigcirc_{ab}$, do you need to change the definition of models for $MADL$ or is it enough to just modify the semantics?

16. Does $\bigcirc_{ab}(A) \wedge \bigcirc_{ab}(A \rightarrow B) \models_{MADL} \bigcirc_{ab}(B)$?

17. Can you characterize an accessibility relation for $\bigcirc_{ab}$ based on the accessibility relations for $\bigcirc_a$ and $\bigcirc_b$?

> *Hint:* It is very helpful to think about this with an example. Suppose that a and b are museum guards.
>
> a's obligations are to see to it that (i) no one touches the Warhol paintings and (ii) that no one touches the Rothko paintings.
>
> b's obligations are to see to it that (i) no one touches the Warhol paintings and (ii) that no one touches the Lichtenstein paintings.
>
> Then their universal obligation is to see to it that no one touches the Warhol paintings. However, the universal obligation do *not* include that no one touches the Rothko paintings or the Lichtenstein paintings.
>
> This means that universally speaking, the ab-accessible worlds must *all* be worlds in which the Warhols are untouched, but may disagree on whether the Rothko/Lichtenstein paintings are untouched.

Let us add one last operator. Informally, we intended to have $\bigcirc_{ab}(A)$ true when A was an obligation for *both* a and b. Now, instead, we want to consider an operator $\bigcirc_{a+b}$ such that $\bigcirc_{a+b}(A)$ would be true if A was an obligation for either a or b. Informally, $\bigcirc_{a+b}$ tracks what the world would have to be like if everyone acted as they ought to.

18. Can you characterize an accessibility relation for $\bigcirc_{a+b}$ based on the accessibility relations for $\bigcirc_a$ and $\bigcirc_b$?

19. Is the accessibility relation for $\bigcirc_{a+b}$ guaranteed to be serial? (Whatever you say here, think about the conceptual significance of this!)

20a. Does $\bigcirc_{ab}(A) \models \bigcirc_{a+b}(A)$?

20b. Does $\bigcirc_{a+b}(A) \models \bigcirc_{ab}(A)$?

5 Derivations

Feel free to use derived rules: RK, PL, Substitution.

21. Provide a derivation for $\mathbf{K} \vdash \neg\Diamond\bot$.

22. Provide derivations for $\mathbf{KD} \vdash \bigcirc\neg A \leftrightarrow \neg\mathbf{P}B$ and $\mathbf{KD} \vdash \neg\bigcirc A \leftrightarrow \mathbf{P}\neg B$. Then feel free to use these substitutions in later derivations.

23. Provide a derivation of $\mathbf{KD} \vdash \bigcirc(A \to B) \to (\mathbf{P}A \to \mathbf{P}B)$

24. Provide a derivation of $\mathbf{KD} \vdash \bigcirc(A \wedge B) \to (\bigcirc A \wedge \bigcirc B)$

25. Provide a derivation of $\mathbf{KD} \vdash \mathbf{P}(A \vee B) \to (\mathbf{P}A \vee \mathbf{P}B)$.