Deontic

Bjørn Kjos-Hanssen

February 28, 2026

Carmo and Jones in 2022 [1] proposed certain axioms 5(a)–(g) for a relation ob that holds between sets of possible worlds X and Y if X is obligatory in the context Y. It was the latest iteration in a sequence of systems [2, 3, 4].

We will exhibit a paradox therein. Our paradox will be a weak form of conditional deontic explosion: given that something is somewhat desirable (passing a course with a grade of C, say) and given that the most desirable outcome (the grade of A) is unavailable, the somewhat desirable outcome becomes obligatory.

We then show that despite this paradox, the systems of Carmo and Jones have interesting mathematical content. For the strongest system we provide a full classification of its models; for weaker versions we characterize the least models (under inclusion) satisfying the axioms and basic contrary-to-duty assumptions.

Chapter 1

The paradox

Let W be a finite set of possible worlds of a given model, and let $\mathcal{P}$ denote the power set operation.

Combining the systems from several papers, the full list of Carmo and Jones' conditions on a function $ob : \mathcal{P}(W) \to \mathcal{P}(\mathcal{P}(W))$ is as follows.

5(a) Axiom 5(a): $\emptyset \notin \mathrm{ob}(X)$. (Obligations cannot be impossible to fulfil.)

5(b) Axiom 5(b): If $Y \cap X = Z \cap X$ then $Y \in \mathrm{ob}(X)$ iff $Z \in \mathrm{ob}(X)$. (Whether Z is obligatory in context X depends only on Z through $Z \cap X$.)

5(c) Axiom 5(c$^-$) (2002, page 319; finite version of 5(c*), 2013, which is called 5(c) in 2022): If $Y \in \mathrm{ob}(X)$ and $Z \in \mathrm{ob}(X)$ and $X \cap Y \cap Z \neq \emptyset$, then $Y \cap Z \in \mathrm{ob}(X)$. (Closure under intersection.)

5(c) Axiom 5(c) (1997 and 2002, page 287): If $Y \in \mathrm{ob}(X)$ and $Z \in \mathrm{ob}(X)$ then $Y \cap Z \in \mathrm{ob}(X)$. (Closure under intersection, even for disjoint sets.)

5(d) Axiom 5(d): If $Y \subseteq X$ and $Y \in \mathrm{ob}(X)$ and $X \subseteq Z$, then $(Z \setminus X) \cup Y \in \mathrm{ob}(Z)$. (If Y is obligatory in context X then a form of the material conditional "$X \to Y$" holds in a wider context.)

5(e) Axiom 5(e): If $Y \subseteq X$ and $Z \in \mathrm{ob}(X)$ and $Y \cap Z \neq \emptyset$, then $Z \in \mathrm{ob}(Y)$. (If Z is obligatory in the context X then it remains obligatory in any subcontext in which it is still possible.)

5(f) Axiom 5(f): If $X \in \mathrm{ob}(Y)$ and $X \in \mathrm{ob}(Z)$ then $X \in \mathrm{ob}(Y \cup Z)$. (If X is obligatory in both of the contexts Y and Z, then it is obligatory in their union.)

5(g) Axiom 5(g): If $Y \in \mathrm{ob}(X)$ and $Z \in \mathrm{ob}(Y)$ and $X \cap Y \cap Z \neq \emptyset$, then $Y \cap Z \in \mathrm{ob}(X)$. (A form of transitivity for obligations.)

Carmo and Jones gave several axioms on their function ob.

Definition 1. Axiom 5(a) for a function $\mathrm{ob} : \mathcal{P}(U) \to \mathcal{P}(\mathcal{P}(U))$ says that $\emptyset \notin \mathrm{ob}(X)$ for all $X \in \mathcal{P}(U)$.

Definition 2. Axiom 5(b) for a function $\mathrm{ob} : \mathcal{P}(U) \to \mathcal{P}(\mathcal{P}(U))$ says that

$$\forall X\, Y\, Z, Z \cap X = Y \cap X \to (Z \in \mathrm{ob}\, X \leftrightarrow Y \in \mathrm{ob}\, X)$$

Definition 3. The weak, but potentially infinite, version of Axiom 5(c) for a function ob : $\mathcal{P}(U) \to \mathcal{P}(\mathcal{P}(U))$ says that

$$\forall X, \forall \beta \subseteq \text{ob}\, X, \beta \neq \emptyset \to \bigcap \beta \cap X \neq \emptyset \to \bigcap \beta \in \text{ob}\, X.$$

Definition 4. The strong version of Axiom 5(c) for a function ob : $\mathcal{P}(U) \to \mathcal{P}(\mathcal{P}(U))$ says that If $Y \in \text{ob}(X)$ and $Z \in \text{ob}(X)$ then $Y \cap Z \in \text{ob}(X)$.

Definition 5. Axiom 5(d) for a function ob : $\mathcal{P}(U) \to \mathcal{P}(\mathcal{P}(U))$ says that

$$\forall (X, Y, Z \subseteq U), Y \subseteq X \to Y \in obX \to X \subseteq Z \to Z\ X \cup Y \in obZ$$

Lemma 6. *If* $A \in \text{ob}(X)$ *whenever* $A \subseteq X$ *and* $a_1 \neq a_2$ *then* $A \cup \{a_1\} \in \text{ob}(A \cup \{a_1, a_2\})$.

Definition 7. Axiom 5(e) for a function ob : $\mathcal{P}(U) \to \mathcal{P}(\mathcal{P}(U))$ says that If $Y \subseteq X$ and $Z \in \text{ob}(X)$ and $Y \cap Z \neq \emptyset$, then $Z \in \text{ob}(Y)$.

The argument. Axioms 5(b)(e)(f) will now be used to derive a paradox.

Suppose that James is taking an exam on which the possible grades, in decreasing order of quality, are A, B, C, D and F.

If seems reasonable to assume that given that James' grade is A or B, it ought to be A.

Moreover, given that the grade is C or D, it ought to be C.

Then by axiom 5(b), given that the grade is C or D, it ought to be A or C. Moreover, also by 5(b), given that the grade is A or B, it ought to be A or C. In other words, James ought to make sure the proposition "the grade is A or C" is true (of course, ideally by getting an A).

It also follows that the grade ought to be A or C given that it is A, B, C, or D, this time using axiom 5(f) applied to the two previous statements. (If this is getting hard to follow, the reader may rest assured that a formalization is available, see 2.1.)

So far this is not entirely unintuitive. But now we use 5(e), and conclude that since A or C was obligatory in the context A, B, C, or D it must remain obligatory in a more restrictive context in which it is still possible, namely B, C, or D.

Finally, using 5(b) again, we conclude that given that the grade is B, C, or D, it ought to be C. While there is such a concept as "gentleman's C", surely this is a paradox.[1]

A literary perspective. Let us consider this paradox above in another context, beside students and their grades.

Suppose James has the following options.

A Marriage to Alice, his favorite.

B Marriage to Alice's sister Beatrice, or one of several other women he also quite likes.

C Remaining a lonely bachelor.

D Marriage to Deirdre, whom he despises.

Then our conclusion is like that of the Bee Gees in 1977:

If I can't have you, I don't want nobody.

[1]Traditionally a *Gentleman's C* is given to someone who deserves D or F but for extraneous reasons is deemed worthy of a C. Here this is reversed, as a C is deemed obligatory even when a B is available.

From despondency to mathematics. Hage [5] thinks the task Carmo and Jones and others have set themselves is impossible.

From Hage's book review, in the context of discussing a man who (i) should visit his neighbor, (ii) if he does visit, should call to say that he is coming, and (iii) is not in fact coming:

> [...] Actually, A does not assist his neighbours and therefore should not call them. The two intuitions presuppose a different role for deontic logic, namely, reasoning about what is ideally the case, and reasoning about what ought to be done in the real world. These two roles are hard to reconcile in one logic, unless the logic combines the two kinds of reasoning in different parts. The paper by Carmo and Jones in the present volume provides such a combination logic. Nevertheless, many have attempted to do what is, in my opinion, **undoable**[2] and this has lead to many modifications of the Standard System of Deontic Logic (Hilpinen, 1971, p. 13f.; Chellas, 1980, p. 190f.). These new systems have lead to new paradoxes that deal with CTD obligations.

Given Hage's sentiment and the paradox we have presented, one might temporarily become despondent. If deontic systems always have flaws, why pursue them? However, in the course of uncovering this paradox I also discovered interesting *mathematical structure* in the axioms 5(a)–5(g).

Definition 8. Given a set W and a family $\mathcal{F}$ of functions $f : \mathcal{P}(W) \to \mathcal{P}(\mathcal{P}(W))$, we say that $f_0 \in \mathcal{F}$ is the *least* element of $\mathcal{F}$ if for all $f \in \mathcal{F}$ and all $X \subseteq W$, $f_0(X) \subseteq f(X)$.

The *least model* of a set of axioms $\mathcal{A}$ concerning a variable function $\mathrm{ob} : \mathcal{P}(W) \to \mathcal{P}(\mathcal{P}(W))$ is the least element of the collection $\mathcal{F}$ of all functions $\mathrm{ob} : \mathcal{P}(W) \to \mathcal{P}(\mathcal{P}(W))$ satisfying all axioms in $\mathcal{A}$.

"Axiom" here is used to mean simply a condition on ob, although we may note that all the axioms 5(a)–(g) may be formulated in first-order set theory.

The least model of 5(b) given certain "oughts" has a very natural characterization, as does the least model of 5(b), 5(d), and 5(f).

To be specific, let us write $\mathrm{Ought}(A \mid B)$ to mean that for each $X \subseteq B$, if $A \cap X \neq \emptyset$ then $A \in \mathrm{ob}\, X$. This is the semantic condition used by Carmo and Jones for the conditional obligation operator $O(A \mid B)$. A pair of oughts $(\mathrm{Ought}(A \mid W), \mathrm{Ought}(B \mid W \setminus A))$ forms a basic contrary-to-duty obligation of B given that our duty A has failed to be observed.

Definitions 9 and 10 introduce canon_2 and $\mathrm{canon}_2\mathrm{II}$, "canonical" models of some of Carmo and Jones' axioms with two parameters: a set of most desirable worlds, A, and a set of worlds that are desirable when A is not available, B.

Definition 9. The set canon_2 A B X consists of all contexts that are obligatory in the context X in the model canon_2 A B.

It is defined to be: if $X \cap B = \emptyset$ then $\emptyset$, else: if $X \cap A = \emptyset$ then $\{T \mid X \cap B \subseteq T\}$, else $\{T \mid X \cap A \subseteq T\}$.

Definition 10. The set $\mathrm{canon}_2\mathrm{II}$ A B X is defined by: if $X \cap B = \emptyset$ then $\emptyset$, else: if $X \cap A = \emptyset$ then $\{Y \mid X \cap B = X \cap Y\}$ else $\{Y \mid X \cap A = X \cap Y\}$.

In other words, if $Y \in \mathrm{ob}(X)$ under a canon _II model then generically $X \cap Y$ consists exactly of the most desirable worlds in X.

Theorem 11. *The least model* ob *(under inclusion) of* $\mathrm{Ought}(A \mid W)$, $\mathrm{Ought}(B \mid W \setminus A)$, *and axiom 5(b) is canon _II A B.*

[2] [emphasis ours]

Even though the model arises from assuming 5(b) only, it also satisfies axioms 5(a), 5(c), 5(e) and 5(g). This indicates a certain robustness of our definitions.

Theorem 12. *The least model of axioms 5(b), 5(d), 5(f) and the two "oughts" in 11 is canon A B.*

In other words, $X \cap Y$ contains at least all the most desirable worlds in X. Since axiom 5(f) follows from 5(a)(b)(c)(d), as shown by Carmo and Jones, the model can alternatively be characterized as the least one satisfying the latter four axioms and the two specified oughts.

The two families of models in 11 and 12 represent two alternative approaches to contrary-to-duty obligations as discussed in [7] (called I and II there). They do not exhaust the interesting models by any means: for instance, we may have conflicting obligations. This may lead us to prefer 5(c) to 5(c)*, in particular. For a concrete example, suppose that James has received acceptances on separate marriage proposals to both Alice and Beatrice, but cannot marry them both.

Classification. In mathematics, classification theorems are fairly common. For example, the finite abelian groups have a straightforward characterization and the finite simple groups a complicated one. Vector spaces over a fixed field $\mathbb{F}$ of finite dimension are characterized by their dimension d, hence the single parameter d determines the space up to isomorphism.

In deontic logic, where axioms and rules are added based on moral intuitions, we should perhaps not expect structures that are mathematically natural enough to be classifiable.

However, for the full theory of Carmo and Jones's axioms 5(a)–5(e) with the strong version 5(c)* in which we do not impose nondisjointness, we can characterize its models completely. The only nontrivial ones basically say that there is just one bad world and the only obligation is to avoid it:

Theorem 13. *Let W be a finite set of possible worlds and let $\mathrm{ob} : \mathcal{P}(W) \to \mathcal{P}(\mathcal{P}(W))$. The following are equivalent:*

1. ob *satisfies the full system suggested in [2]: 5(a), (b), (c)*, (d) and (e).*
2. *One of the following three holds:*
   - *(a)* $\mathrm{ob}(X) = \emptyset$ *for all* X.
   - *(b)* $\mathrm{ob}(X) = \{Y \mid \emptyset \neq X \subseteq Y\}$ *for all* X.
   - *(c) There is a distinguished possible world a (the "forbidden" world) such that for all X,* $\mathrm{ob}(X) = \{Y \mid X \cap Y \neq \emptyset$ *and* $X \setminus \{a\} \subseteq Y\}$.

Details on the proof of Theorem 13 can be found in the next section.

Chapter 2

Technical details of the characterization of CJ97

Definition 14. The theory BDE consists of the axioms 5(b)(d)(e).

Definition 15. The theory ADE consists of the axioms 5(a)(d)(e).

Definition 16. Conditional explosion for ob is the statement that (A B C : Finset U), A ob C → B A C → B ob (A C).

Theorem 17. *If ob satisfies axioms 5(a)(b)(d)(e) then ob satisfies conditional explosion.*

Lemma 18. *If* $a_1 \neq a_2$*,* $a_1 \notin A$*,* $a_2 \notin A$*,* $\{a_1, a_2\} \in \mathrm{ob}(\{a_1, a_2\})$*, and* $A \in \mathrm{ob}(X)$ *whenever* $A \subseteq X$*, then* $\{a_1\} \in \mathrm{ob}\{a_1, a_2\}$ *and* $\{a_2\} \in \mathrm{ob}\{a_1, a_2\}$*.*

Lemma 19. *If* $a_1 \neq a_2$*,* $a_1 \notin A$*,* $a_2 \notin A$*,* $\{a_1, a_2\} \in \mathrm{ob}(\{a_1, a_2\})$*, then it cannot be that* $A \in \mathrm{ob}(X)$ *whenever* $A \subseteq X$*.*

Lemma 20. *If* $a_1 \notin A$ *and* $a_2 \notin A$ *and* $a_1 \neq a_2$ *then it cannot be that* $A \in \mathrm{ob}(X)$ *whenever* $A \subseteq X$*.*

Definition 21. A world a is bad if (X : Finset (Fin n)), a X $X \setminus \{a\} \in \mathrm{ob}\, X$.

Definition 22. The world a is quasibad if (X : Finset (Fin n)) (Y : Finset (Fin n)), a X \ Y Y ob X.

Thus, a world a is bad if in some context there is an obligation to simply avoid a. For example, if there is an obligation "do not go to war" then the world representing "going to war with Syria" is *quasibad*, but it is not *bad* unless there is also the specific obligation "do not go to war with Syria". (In "reasonable" systems this distinction would perhaps not need to be made, but here we are in the process of proving that a certain system 5(abcde) is not reasonable.)

Lemma 23. *If* a *is bad and* $Y \setminus \{a\} \neq \emptyset$ *then* $Y \setminus \{a\} \in \mathrm{ob}\, Y$*.*

Lemma 23 says that badness of the world does not depend on context.

Lemma 24. *If* $X \in \mathrm{ob}\, X$ *and* $Y \neq \emptyset$ *then* $Y \in obY$*.*

Lemma 24 says that if any context is obligatory relative to itself, then they all are.

Lemma 25.
If $\emptyset \neq X \setminus \{a\} \in obX$ *then* $X \in \mathrm{ob}\, X$.

Lemma 25 says that if there is a bad world then the corresponding context is self-obligatory.

Lemma 26. *Assume axioms 5(abde). If a is* ob-*bad then* $\{a\} \in \mathrm{ob}\{a\}$.

Lemma 26 is a technicality: even if a is bad, it is still self-obligatory.

Lemma 27. *If* $\mathrm{univ} \setminus \{a\} \in \mathrm{ob}\, \mathrm{univ}$ *then* $\mathrm{univ} \in \mathrm{ob}\, \mathrm{univ}$.

Lemma 27 is another technicality: if a is bad in the global context then the global context is self-obligatory.

Lemma 28. *Suppose that for all contexts A, if A is obligatory in all larger contexts $X \supseteq A$, then A is a cosubsingleton, i.e., missing at most one element (from the global context). Then for all B and C, if $B \subseteq C$ is obligatory relative to C then $C \setminus B$ is a cosubsingleton.*

Lemma 28 is a "global-to-local" principle allowing us to conclude a fact about an arbitrary context C from a fact about the global context.

The antecedent of Lemma 28 is provided by Lemma 29 and hence the consequent is provided by Lemma 30.

Lemma 29. *For all contexts A, if A is obligatory in all larger contexts $X \supseteq A$, then A is a cosubsingleton, i.e., missing at most one element (from the global context).*

Lemma 30.
For all B and C, if $B \subseteq C$ is obligatory relative to C then $C \setminus B$ is a cosubsingleton.

Definition 31. The model *stayAlive* is defined by $\mathrm{stayAlive\ e\ X} = \{Y | X \cap Y \neq \emptyset \wedge X \setminus \{e\} \subseteq X \cap Y\}$.

Definition 32. The model *alive* is defined by $\mathrm{alive\ n\ X} = \{Y \mid X \neq \emptyset \wedge Y \supseteq X\}$.

Definition 33. The model *noObligations* is defined by noObligations X = .

We think of *alive* as a computer game like "Snake" where the objective is to stay alive, with the surprising twist that it is not possible to die. In contrast, in the model *noObligations* there are no obligations at all, and in the model *stayAlive* the objective is standard: stay alive.

Thus, someone playing Snake under the noObligations model can relax completely, whereas someone playing under the *alive* model may worry that perhaps there is a way to die that they just have not seen yet. In fact, *alive* is a reduct of *stayAlive* where we remove the one bad world.

We prove several technical lemmas, culminating in Theorem 42:

Lemma 34. *If ob noObligations and a,* $\neg$ *quasibad ob a, then ob = alive k.*

Lemma 35. *If bad ob a and bad ob b then* $a = b$.

Lemma 36. *If bad ob a and* $X \cap Y \in obX$ *then* $X \cap Y = X$ *or* $X \cap Y = X \setminus \{a\}$.

Lemma 37. *If bad ob a and X a and X then X ob X.*

Lemma 38. *If bad ob a and* $X \setminus \{a\} \neq \emptyset$ *then* $X \in \mathrm{ob}\, X$.

Lemma 39. *If bad ob a then Y, ob Y stayAlive a Y.*

Note that Lemma 40 does not require any form of C5.

Lemma 40. *If bad ob a then Y, stayAlive a Y ob Y.*

Lemma 41. *If bad ob a then ob = stayAlive a.*

Theorem 42. *Every model of axioms 5(abcde) is either stayAlive a for some bad world a, alive, or noObligations.*

2.1 Acknowledgments

All mathematical claims above are verified in the proof assistant Lean, see [8].

The main argument for our paradox was discovered by carefully analyzing some output from a script in the computer mathematics system Maple by Maplesoft [6, Appendix].

This work was partially supported by a grant from the Simons Foundation (#704836 to Bjørn Kjos-Hanssen).

Bibliography

[1] José Carmo and Andrew J. I. Jones. Carmo and Jones' logic for contrary-to-duty obligations revised. *J. Logic Comput.*, 32(7):1352–1364, 2022.

[2] José M. C. L. M. Carmo and Andrew J. I. Jones. A new approach to contrary-to-duty obligations. In D. Nute, editor, *Defeasible Deontic Logic*, volume 263 of *Synthese Library*, pages 317–344. Kluwer Academic Publishers, 1997.

[3] José M. C. L. M. Carmo and Andrew J. I. Jones. Deontic logic and contrary-to-duties. In D.M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 8, pages 265–343. Kluwer Academic Publishers, 2nd edition, 2002.

[4] José M. C. L. M. Carmo and Andrew J. I. Jones. Completeness and decidability results for a logic of contrary-to-duty conditionals. *J. Logic Comput.*, 23(3):585–626, 2013.

[5] Jaap Hage. Donald Nute (Ed.), Defeasible Deontic Logic. *Artificial Intelligence and Law*, 8(1):75–91, 2000.

[6] Bjørn Kjos-Hanssen. Models of the Chisholm set. 1996. Philosophy term paper, University of Oslo. arxiv 1607.02189.

[7] Bjørn Kjos-Hanssen. A conflict between some semantic conditions of Carmo and Jones for contrary-to-duty obligations. *Studia Logica*, 105(1):173–178, 2017.

[8] Bjørn Kjos-Hanssen. Deontic, 2025. Lean 4 github repository at https://github.com/bjoernkjoshanssen/deontic.