Obligations in a context

Simplified semantics for the deontic logic of contrary-to-duties

Bjørn Kjos-Hanssen

University of Hawaii at Manoa

May 23, 2016

Neither Reagan nor Gorbachev must be told the secret.

But if one of them is told, the other must be told as well.

- Neither Gorbachev should be told the secret.
- But if one of them is told the secret, then the other one should be told as well.

- Neither Gorbachev should be told the secret.
- But if one of them is told the secret, then the other one should be told as well. A *contrary-to-duty obligation.*

Conditional obligation

CTD obligations do not formalize well if $O(B \mid A)$ is taken as $A \to O(B)$ or $O(A \to B)$. A separate dyadic operator $O(B \mid A)$ is needed.
Carmo and Jones (1997, 2002, 2013) base it on a function

$$\mathrm{ob} : 2^W \to 2^{2^W}$$

with

$$\models_\alpha O(A \mid B) \iff \|A\| \in \mathrm{ob}(\|B\|),$$

where W is the set of worlds.

Here we propose a simpler function $F : 2^W \to 2^W$ and

$$\mathrm{ob}(X) = \{Y : Y \supseteq F(X)\}.$$

Because of the idea of a given context as a set of worlds, it seems we cannot further simplify to just a relation on worlds $R \subseteq W \times W$.
An alternative notion is obtained by

$$Y \in \mathrm{ob}(X) \iff Y \cap X = F(X).$$

Carmo and Jones' conditions on ob

5(a) $\mathrm{ob}(X) \neq \emptyset$.

5(b) If $Y \cap X = Z \cap X$ then $Y \in \mathrm{ob}(X)$ iff $Z \in \mathrm{ob}(X)$.

5(c) If $Y \in \mathrm{ob}(X)$ and $Z \in \mathrm{ob}(X)$ then $Y \cap Z \in \mathrm{ob}(X)$.

5(d) If $Y \subseteq X$ and $Y \in \mathrm{ob}(X)$ and $X \subseteq Z$, then $(Z \setminus X) \cup Y \in \mathrm{ob}(Z)$.

5(e) If $Y \subseteq X$ and $Z \in \mathrm{ob}(X)$ and $Y \cap Z \neq \emptyset$, then $Z \in \mathrm{ob}(Y)$.

Carmo and Jones' conditions on ob

5(a) $\mathrm{ob}(X) \neq \emptyset$.

Carmo and Jones' conditions on ob

5(a) $\mathrm{ob}(X) \neq \emptyset$. We only require this for $X \neq \emptyset$. In any case $\mathrm{ob}(\emptyset)$ is not that interesting.

5(b) If $Y \cap X = Z \cap X$ then $Y \in \mathrm{ob}(X)$ iff $Z \in \mathrm{ob}(X)$.

5(b) If $Y \cap X = Z \cap X$ then $Y \in \mathrm{ob}(X)$ iff $Z \in \mathrm{ob}(X)$. Follows from $F(X) \subseteq X$.

5(c) If $Y \in \mathrm{ob}(X)$ and $Z \in \mathrm{ob}(X)$ then $Y \cap Z \in \mathrm{ob}(X)$.

5(c) If $Y \in \mathrm{ob}(X)$ and $Z \in \mathrm{ob}(X)$ then $Y \cap Z \in \mathrm{ob}(X)$. Follows from $\mathrm{ob}(X) = \{Y : Y \supseteq F(X)\}$.

5(d) If $Y \subseteq X$ and $Y \in \mathrm{ob}(X)$ and $X \subseteq Z$, then $(Z \setminus X) \cup Y \in \mathrm{ob}(Z)$.

5(d) If $Y \subseteq X$ and $Y \in \mathrm{ob}(X)$ and $X \subseteq Z$, then $(Z \setminus X) \cup Y \in \mathrm{ob}(Z)$. This becomes the condition $F(X \cap Y) \supseteq F(X) \cap Y$: our standards of perfection can only be relaxed, not strengthened, when restricting the context. If we take $Y \in \mathrm{ob}(X)$ to mean $Y \cap X = F(X)$ then 5(d) is just wrong.

5(e) If $Y \subseteq X$ and $Z \in \mathrm{ob}(X)$ and $Y \cap Z \neq \emptyset$, then $Z \in \mathrm{ob}(Y)$.

5(e) If $Y \subseteq X$ and $Z \in \mathrm{ob}(X)$ and $Y \cap Z \neq \emptyset$, then $Z \in \mathrm{ob}(Y)$. Must be weakened. We use: $F(X \cap Y) = F(X) \cap Y$ whenever $F(X) \cap Y \neq \emptyset$: standards of perfection should only be relaxed when absolutely necessary. However, when $Z \in \mathrm{ob}(X)$ is taken as not $Z \supseteq F(X)$ but $Z \cap X = F(X)$, so Z consists of only ideal worlds, then it is okay.

My main argument

Suppose $W = \{a, b, c, d, e\}$ where a is mandatory: $\{a\} \in \mathrm{ob}(W)$.

- $\{a, b, c\} \in \mathrm{ob}(W)$ by 5(d) since it is "if $\{a, d, e\}$ then $\{a\}$" (which is equivalent to "not d or e"). And then
- $\{a, b, c\} \in \mathrm{ob}(\{b, c, d, e\})$ by 5(e). Similarly,
- $\{a, b, d\} \in \mathrm{ob}(\{b, c, d, e\})$. So
- $\{b, c\} \in \mathrm{ob}(\{b, c, d, e\})$ and $\{b, d\} \in \mathrm{ob}(\{b, c, d, e\})$ by 5(b).

CJ (2002) defined by 5(c$^-$):
If $Y \in \mathrm{ob}(X)$ and $Z \in \mathrm{ob}(X)$ and $X \cap Y \cap Z \neq \emptyset$, then $Y \cap Z \in \mathrm{ob}(X)$.
By 5(c$^-$), $\{b\} \in \mathrm{ob}(\{b, c, d, e\})$. But this was obtained without using any desirability property of b.

More generally, in "general position", if

- $\|A\| \in \mathrm{ob}(W)$, then
- $\|A\| \in \mathrm{ob}(\|A \vee B\|)$ by 5(e). Then
- $\|A \vee \neg(A \vee B)\| = \|A \vee \neg B\| \in \mathrm{ob}(W)$ by 5(d) (slide 6). Then
- $\|A \vee \neg B\| \in \mathrm{ob}(\|\neg A\|)$ by 5(e). So
- $\|\neg B\| \in \mathrm{ob}(\|\neg A\|)$ by 5(b).

But this is absurd, as B was fairly arbitrary.

Prisoner's dilemma

D_i = player i defects.
We adopt player 1's point of view.

$$O(D_1 \wedge \neg D_2)$$

$$O(\neg D_1 \mid D_1 \leftrightarrow D_2).$$

This is another way to try to defeat condition 5(e).
CJ might argue that $D_1 \wedge \neg D_2$ is the "true" obligation, whereas D_1 and $\neg D_2$ are not.
They might argue that the true obligation is to minimize your own number of years in jail. Then, by first-order logic, it follows that $D_1 \wedge \neg D_2$ should hold.

Model for prisoner's dilemma

Let $\mathrm{ob}(X)$ consist of all sets containing the most favorable element of X. That is, we have a valuation $v : W \to \mathbb{N}$ on worlds and we let

$$F(X) = \{u \in X : (\forall x \in X)(v(x) \leq v(u)\}.$$

We can recover an ordering by

$$a \leq b \iff (\forall X)(a \in F_X \to b \in F_X).$$

A weak version of 5(e), 5(e$^-$):

If $Y \subseteq X$ and $Z \in \mathrm{ob}(X)$ and $Y \cap Z \neq \emptyset$, and $\overline{Z} \notin \mathrm{ob}(Y)$, then $Z \in \mathrm{ob}(Y)$.

will help if *the worlds are (strictly) linearly ordered in value*.

Let us consider a model in which they are not. Let's say we want to maximize our profits and minimize the variance in our profits. Then some pairs (μ_1, σ_1^2) are better than others (μ_2, σ_2^2) but for some pairs it is hard to declare a preference. So let's say world a is best overall, b and c are incomparable, and d is worst. Then

- $\{a\} \in \mathrm{ob}(W)$, so
- $\{a\} \in \mathrm{ob}(\{a, b\})$ by 5(e$^-$), so
- $\{a, c, d\} \in \mathrm{ob}(W)$ by 5(d), so
- $\{a, c, d\} \in \mathrm{ob}(\{b, c, d\})$ by 5(e$^-$), since $\{a, c, d\}$ is still possible and in fact not forbidden in this new context.

And so we have a preference for c over b, which is bad.

History of condition 5(e)

- Proposed by Carmo and Jones (published 1997) with the motivation $Y = \mathrm{av}(w)$, the set of actually possible versions of w, and $X = \mathrm{pv}(X)$, the set of potentially or ideally possible versions of X.

History of condition 5(e)

- Proposed by Carmo and Jones (published 1997) with the motivation $Y = \mathrm{av}(w)$, the set of actually possible versions of w, and $X = \mathrm{pv}(X)$, the set of potentially or ideally possible versions of X.
- K-H gives a counterargument (term paper, 1996)

History of condition 5(e)

- Proposed by Carmo and Jones (published 1997) with the motivation $Y = \mathrm{av}(w)$, the set of actually possible versions of w, and $X = \mathrm{pv}(X)$, the set of potentially or ideally possible versions of X.
- K-H gives a counterargument (term paper, 1996)
- Carmo and Jones discuss the counterargument and defeat it by weakening condition 5(c) (*Deontic logic and contrary-to-duties*, 2002)

History of condition 5(e)

- Proposed by Carmo and Jones (published 1997) with the motivation $Y = \mathrm{av}(w)$, the set of actually possible versions of w, and $X = \mathrm{pv}(X)$, the set of potentially or ideally possible versions of X.
- K-H gives a counterargument (term paper, 1996)
- Carmo and Jones discuss the counterargument and defeat it by weakening condition 5(c) (*Deontic logic and contrary-to-duties*, 2002)
- Completeness results published (2013)
- K-H gives a stronger version of the same counterargument (2016)

Fence scenario

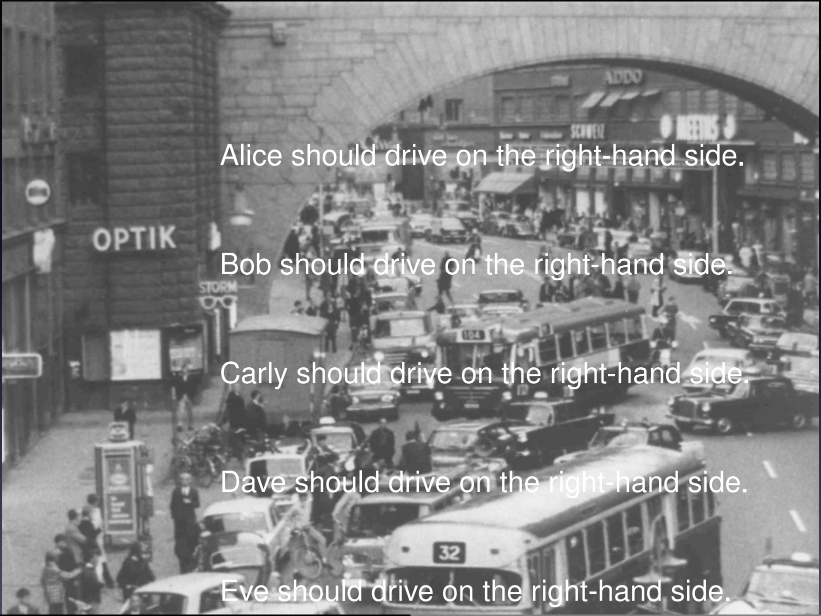
- There ought to be no dog.
- If there is dog, there ought to be a frontyard fence and a backyard fence.
- No point in having a frontyard fence if there is no backyard fence.

may appear to defeat 5(e): let Y be "there is no frontyard fence" and let Z be "there is a backyard fence", and let $X = W$ (no restriction).

Fix Use 5(e$^-$): require $W \setminus Y \notin \mathrm{ob}(X)$ in the antecedent.

CJ No fix is needed! "There ought to be a backyard fence" is not a *true* obligation, only the conjunction of backyard and frontyard is.

Obligation satisfying 5(e)



Alice should drive on the right-hand side.

Bob should drive on the right-hand side.

Carly should drive on the right-hand side.

Dave should drive on the right-hand side.

Eve should drive on the right-hand side.

Traffic scenario

Two cars are driving down the same street in opposite directions.

- A is the proposition that car A is driving on the right side of the street.
- B is the proposition that car B is driving on the right side of the street.

1. There is a primary obligation that $A \leftrightarrow B$. This one is implied by the laws of all countries.
2. Then there is a secondary obligation (which is easier to administrate, and which implies the primary one) that $A \wedge B$. However, some countries instead use $\neg A \wedge \neg B$.

Note however that because of the primary obligation, if $\neg A$ is a fixed fact then $\neg B$ becomes an obligation.

Standard conditional models

We may note that our new $O(\cdot \mid \cdot)$ definition and semantics makes it a normal conditional logic (i.e., extending *CK*), and its models are standard conditional models, in the sense of Chellas 1980, if we define $A \Rightarrow B$ as $O(B \mid A)$.
(In fact they fit a special case of standard conditional models in which the truth of $A \Rightarrow B$ does not depend on the current world.)
Chellas does not consider this option. He considers to define $O(B \mid A)$ as $A \Rightarrow O(B)$, and he considers using *minimal* models (like CJ do) rather than the simpler standard models.

Credits

- Thanks to Andrew J.I. Jones for advice in 1996.
- This work was partially supported by a grant from the Simons Foundation (#315188 to Bjørn Kjos-Hanssen).