Obligation as Counterfactual Reasoning: A Solution to Zvolenszky's Puzzle

WooJin Chung (New York University)

This paper decomposes the semantics of obligation into a counterfactual conditional and a proposition conveying deontically best worlds. We suggest that ***must p*** translates to ***if***$_{cf}$ ***not p, then not good***, where ***if***$_{cf}$ is understood as a counterfactual and ***good*** signifies deontically best worlds. More specifically, we propose the semantics in (1) for *must*. We import Portner's (2009) BEST operator which takes a modal base *M* and an ordering source *O*, and returns the set of best worlds in *M* with respect to *O*. The formula asserts that given the prejacent p, all of the closest $\neg p$-worlds are not deontically best worlds.

(1) $[\![\text{must}]\!]^{f,c,d} = \lambda p \lambda w . \forall w' \in f(w, \neg p) : \neg\text{BEST}_{d(w)}(\cap c(w))(w') = 1$

where f(w, ¬p) returns the ¬p-worlds most similar to w,

c is a circumstantial modal base, and d is a deontic ordering source

Our proposal is inspired by the morphosyntax of Japanese obligation. In Japanese, deontic concepts are not expressed via a modal auxiliary but rather in the form of a conditional. As shown in (2), *John must eat* literally translates to *If John doesn't eat, then not good*. The conditional morpheme *-(ke)reba* 'if' is not only used in deontic contexts but also in run-of-the-mill conditional constructions and receives either epistemic or counterfactual reading. We assume that it is interpreted counterfactually in deontic contexts, because (2) can be uttered even when the prejacent 'John eats' is true.

(2) John-wa tabe-**na-kereba ike-nai**. (Japanese)
John-TOP eat-**NEG-if good-NEG**
'(Lit.) If John were not to/doesn't eat, then not good.' ☞ 'John must eat.'

Under the assumption that English and Japanese speakers have the same conception of obligation, we take the conditional formulation of Japanese obligation to be a transparent version of English *must*.

Comparison with deontic necessity Our semantics can be derived from the Kratzerian view of obligation (Kratzer 1991b) by making two adjustments. The formula in (3) is Portner's analysis of *must*, which is a simplified variant of Kratzer's original account. We can derive (1) from (3) by calculating the contrapositive of the latter and replacing the strict implication with a counterfactual conditional.

(3) $[\![\text{must}_{\text{Kratzer/Portner}}]\!]^{c,d} = \lambda p . \lambda w . \forall w' : \text{BEST}_{d(w)}(\cap c(w))(w') = 1 \rightarrow p(w') = 1$

Although the changes are relatively minor, our account has an advantage over the Kratzerian view in that it provides a solution to the Zvolenszky's puzzle (Zvolenszky 2002).

The Zvolenszky's puzzle Zvolenszky claims that possible worlds semantics, paired with the Kratzerian account of obligation, fails to make the right prediction for (4). Suppose that Britney Spears has a contract with Pepsi, requiring that she does not drink non-Pepsi cola (e.g., Coke) in public. In this scenario, (4) is intuitively false.

(4) If Britney Spears drinks Coke in public, then she must drink Coke in public.

However, Kratzer's modal semantics predicts that any sentence of the form ***if p, then must p*** is vacuously true. According to Kratzer (1991a), an *if*-conditional is not itself a modal operator but rather restricts the modal base of an independently supplied modal operator. Accordingly, (4) receives the following interpretation: *Given the set of circumstantially accessible worlds in which Britney Spears drinks Coke in public, the best worlds are selected based on the deontic ordering source, and in those best worlds the prejacent 'she drinks Coke in public' is true.* This is vacuously true because the modal only quantifies over the worlds in which Britney drinks Coke in public (henceforth **coke**-worlds).

One way of ruling (4) out is to adopt the *double modalization strategy*. In this view, an indicative *if*-conditional imports its own epistemic necessity operator. Accordingly, (4) is reinterpreted as in (5).

(5) $[\![(4)]\!]^{e,c,d} = \lambda w . \forall w' \in e(w, \textbf{coke}) : [\![\text{must}_{\text{Kratzer/Portner}}]\!]^{e,c,d}(\textbf{coke})(w') = 1$
$= \lambda w . \forall w' \in e(w, \textbf{coke}) : (\forall w'' : \text{BEST}_{d(w')}(\cap c(w'))(w'') = 1 \rightarrow \textbf{coke}(w'') = 1)$

where e(w, ***coke****) returns the epistemically accessible* ***coke****-worlds from w,*

c is a circumstantial modal base, and d is a deontic ordering source

The formula reads as follows: *In all of the epistemically accessible* ***coke****-worlds, the consequent she must drink Coke is true.* In order to avoid the vacuous truth of (4), the double modalization strategy needs to stipulate that the antecedent of the indicative conditional (i.e., **coke**) does not affect the evaluation of the consequent, failing to restrict the circumstantial modal base of *must*. Only under this assumption can the modal domain of *must* exclude the **coke**-worlds. This would let us interpret (4) as false.

However, Zvolenszky shows that we cannot make this stipulation unconditionally, because (6) needs to be true under the same contract-with-Pepsi scenario. If the antecedent proposition is disregarded in evaluating the modal *must* of the consequent, the modal domain of *must* can contain some of the worlds in which Britney does not drink cola in public at all. Then the consequent *she must drink Pepsi* is predicted to be false because in a world where Britney did not drink cola at all, she did not drink Pepsi.

(6) If Britney Spears drinks cola in public, then she must drink Pepsi.

To summarize, the problem is that there is no systematic way of determining whether the modal in the consequent should disregard the antecedent proposition (ex (4)) or take it into consideration (ex (6)).

An alternative solution is Frank's (1997) *Expansion* strategy. Expansion stipulates that whenever we evaluate ***must p***, the modal base should not contain the proposition ***p***. However, Zvolenszky argues that it is a stipulation invented solely to circumvent the Zvolenszky's puzzle, hence is not well-motivated.

Analysis We suggest that our counterfactual-based analysis of obligation motivates Frank's Expansion strategy. More specifically, we claim that whenever a modal sentence of the form ***if***$_{cf}$ ***not p, then not good*** (ex (1)) is uttered, the negation of the counterfactual antecedent, ***p***, is disregarded in evaluating the circumstantial modal base of the BEST operator. The reasoning is that counterfactuals have an effect of removing the piece of information that contradicts their antecedent proposition.

Our analysis of (4) is fleshed out in (7). The formula can be informally read as follows: *In all of the epistemically accessible worlds where Britney Spears drinks coke in public, suppose that she didn't drink coke in public. Then those worlds are not deontically best worlds.* Under the contract-with-Pepsi scenario, this is false because Britney never had an obligation to drink coke in public, and there is no reason to completely rule out the possibility that some ¬**coke**-worlds are deontically best worlds. What is important here is that the antecedent of the indicative conditional (**coke**) contradicts the antecedent of the counterfactual (¬**coke**). Thus, the proposition **coke** is disregarded when evaluating the circumstantial modal base of the BEST operator, and **coke** is not necessarily true in the deontically best worlds.

(7) $[\![\text{(4)}]\!]^{e,f,c,d} = \lambda w.\forall w' \in e(w, \textbf{coke}) : [\![\text{must}]\!]^{e,f,c,d}(\textbf{coke})(w') = 1$

$= \lambda w.\forall w' \in e(w, \underline{\textbf{coke}}) : \forall w'' \in f(w', \underline{\neg\textbf{coke}}) : \neg\text{BEST}_{d(w')}(\cap c(w'))(w'') = 1$

where e(w, ***coke****) returns the epistemically accessible* ***coke****-worlds from w,*
*f(w', ¬****coke****) returns the ¬****coke****-worlds most similar to w',*
c is a circumstantial modal base, and d is a deontic ordering source

We also make the correct prediction for (6). The formula in (8) can be read as follows: *In all of the epistemically accessible worlds where Britney Spears drinks cola in public, suppose that she didn't drink Pepsi. Then those worlds are not deontically best worlds.* This is predicted to be true under the contract-with-Pepsi scenario: In all of the deontically best worlds, Britney either drinks Pepsi or does not drink cola at all. Thus, no world in which **cola** and ¬**pepsi** are both true is a deontically best world.

(8) $[\![\text{(6)}]\!]^{e,f,c,d} = \lambda w.\forall w' \in e(w, \underline{\textbf{cola}}) : \forall w'' \in f(w', \underline{\neg\textbf{pepsi}}) : \neg\text{BEST}_{d(w')}(\cap c(w'))(w'') = 1$

How (8) crucially differs from (7) is that in the former, the antecedent of the indicative conditional (**cola**) does not contradict that of the counterfactual (¬**pepsi**). Therefore, the proposition **cola** remains to be a circumstantially relevant proposition in evaluating *must*. By selectively filtering out the antecedent proposition of the indicative conditional, we have a principled account of the Zvolenszky's puzzle.

SELECTED REFERENCES **Frank 1997**, Context Dependence in Modal Constructions. **Kratzer 1991a**, Conditionals. **Kratzer 1991b**, Modality. **Portner 2009**, *Modality*. **Zvolenszky 2002**, *SALT* 12.