

On the Dynamics of Professor Procrastinate

WooJin Chung (New York University)

Procrastinate in a static context Professor Procrastinate receives an invitation to review a book. He is the best person to do the review, and the best thing that can happen is that he says yes, and then writes the review. However, suppose it is further the case that were Procrastinate to say yes, he would not in fact get around to writing the review because he would keep on putting the task off. Thus, what would in fact happen were he to say yes is that he would not write the review. Moreover, this is the worst that can happen. It would lead to the book not being reviewed at all. In this scenario, both ‘Procrastinate ought to accept and write’ (i.e., ‘*ought* **accept** \wedge **write**’) and ‘Procrastinate ought not accept’ (i.e., ‘*ought* \neg **accept**’) are intuitively true. But the standard account of modality predicts that if ‘*ought* **accept** \wedge **write**’ is true in every deontically best world, ‘*ought* **accept**’ will also be true. Therefore, ‘*ought* \neg **accept**’ is predicted to be false. This led Lassiter (2011) to develop a non-monotonic treatment of *ought*.

Procrastinate in a dynamic context von Fintel (2012) claims that monotonicity is in fact a desired property. He notes that ‘*ought* \neg **accept**’ is true because we are led to believe that Professor Procrastinate’s act of writing the review is not a possible course of action. But once we consider it to be a possibility, we can no longer claim that Procrastinate ought not accept (ex (1)). I would like to add that reversing the order of utterance improves the sentence (ex (2)).

- (1) # Procrastinate ought to accept and write, but he ought not accept.
- (2) Procrastinate ought not accept, though he ought to accept and write.

The question remains, as there is no principled reason not to entertain the possibility of Procrastinate writing the review when interpreting ‘*ought* \neg **accept**’ but do so when interpreting ‘*ought* **accept** \wedge **write**’. The standard assumption is that the quantificational domain of *ought* is jointly determined by a modal base and an ordering source supplied by context. Given that we have a single context, what motivates handcrafting the modal base to not include **write**-worlds when interpreting ‘*ought* \neg **accept**’ but leave it as a possibility when interpreting ‘*ought* **accept** \wedge **write**’? We are effectively peeking at the prejacent to decide which worlds are relevant and which worlds are not. My hypothesis is that this built-in sensitivity to the prejacent is a mere side effect of counterfactual revisions underlying the semantics of *ought*.

Counterfactuals in a dynamic context I would like to note that the debate is reminiscent of the one concerning the (non-)monotonic property of counterfactuals. Lewis (1973) proposes that counterfactuals are non-monotonic, but von Fintel (2001) points out that the non-monotonic property is due to the shift of context. The logic is that counterfactuals are sensitive to the order of utterance as shown below, so they cannot be non-monotonic simpliciter:

- (3) If the USA threw its weapons into the sea tomorrow, there would be war; but if all nuclear powers threw their weapons into the sea tomorrow, there would be peace.
- (4) ?? If all the nuclear powers threw their weapons into the sea tomorrow, there would be peace; but if the USA threw its weapons into the sea tomorrow, there would be war.

von Fintel introduces a contextual parameter dubbed the modal horizon to capture the dynamic aspect of counterfactuals. Each time a counterfactual is uttered, the closest antecedent-worlds are added to the modal horizon, and the counterfactual is a strict implication over the updated modal horizon. In this view, (4) is degraded because the second counterfactual is interpreted with respect to the modal horizon that contains the closest worlds where all the nuclear powers abandoned their weapons. Having already entertained this possibility, it cannot be the case that every world in which the USA abandoned its weapons is a world where there is war, because in some of those worlds, all other nuclear powers have abandoned their weapons too.

Proposal Inspired by Korean where ‘must/ought ϕ ’ literally translates to ‘only if (it were) ϕ , good’, I propose that *ought* assesses the goodness of the hypothetical prejacent-worlds and its alternative-worlds. I assume along with Herburger (2015) that an *if* conditional scoping below *only* conveys existential quantification. But unlike Herburger, I assume that the conditional existentially quantifies over antecedent-worlds rather than events. Concerning the semantics of *good*, I suggest that it denotes the set of deontically best worlds among the relevant worlds. The relevant worlds for *good* in the context ‘only if (it were) ϕ , good’ consist of worlds that are relevant to the interpretation of the counterfactual, because *good* appears as the counterfactual consequent. Adopting von Fintel’s (2001) view, the modal horizon serves as the set of relevant worlds, and accordingly *good* denotes the set of deontically best worlds within the modal horizon. Putting the pieces together, ‘ought ϕ ’ first updates the modal horizon with respect to the closest ϕ -worlds and the closest alternative-to- ϕ -worlds. Then it conveys that there exists a ϕ -world within the updated modal horizon that is good (deontically best within the updated modal horizon), and no alternative-to- ϕ -world within the updated modal horizon is good.

Analysis The first *ought* of (1) updates the modal horizon h with respect to the closest **accept** \wedge **write**-worlds, **accept** \wedge \neg **write**-worlds, and \neg **accept**-worlds, yielding h' . Then it conveys that there exists an **accept** \wedge **write**-world within h' which is good, and no **accept** \wedge \neg **write**-world or \neg **accept**-world within h' is good. The update of h' by the second *ought* is redundant because the closest **accept**-worlds (which are **accept** \wedge \neg **write**-worlds) and the closest \neg **accept**-worlds are already in h' . It asserts that no **accept**-world within h' is good, but this is false because the best worlds within h' are **accept** \wedge **write**-worlds, which are **accept**-worlds.

$$\begin{aligned}
(5) \quad & \llbracket \text{Procrastinate ought to accept and write...} \rrbracket^{w,h} \\
& = (\exists w' \in h' : (\mathbf{accept} \wedge \mathbf{write})(w') = 1 \wedge \underbrace{\text{BEST}_{d(w)}(h')(w') = 1}_{\text{good worlds of } w}) \\
& \wedge \neg \exists w' \in h' : (\mathbf{accept} \wedge \neg \mathbf{write})(w') \wedge \underbrace{\text{BEST}_{d(w)}(h')(w') = 1}_{\text{good worlds of } w}) \\
& \wedge \neg \exists w' \in h' : \neg \mathbf{accept}(w') \wedge \underbrace{\text{BEST}_{d(w)}(h')(w') = 1}_{\text{good worlds of } w}), \text{ where}
\end{aligned}$$

$h' = h[\mathbf{accept} \wedge \mathbf{write}]^{\leq}[\mathbf{accept} \wedge \neg \mathbf{write}]^{\leq}[\neg \mathbf{accept}]^{\leq}$ and d is a deontic ordering source

$$\begin{aligned}
(6) \quad & \llbracket \dots(\text{but}) \text{ he ought not accept} \rrbracket^{w,h} \\
& = (\exists w' \in h' : (\neg \mathbf{accept})(w') = 1 \wedge \underbrace{\text{BEST}_{d(w)}(h')(w') = 1}_{\text{good worlds of } w}) \\
& \wedge \neg \exists w' \in h' : \mathbf{accept}(w') \wedge \underbrace{\text{BEST}_{d(w)}(h')(w') = 1}_{\text{good worlds of } w}), \text{ where}
\end{aligned}$$

$h' = h[\mathbf{accept} \wedge \mathbf{write}]^{\leq}[\mathbf{accept} \wedge \neg \mathbf{write}]^{\leq}[\neg \mathbf{accept}]^{\leq}$ and d is a deontic ordering source

Example (2) is acceptable on the other hand. In the first *ought* statement, among the closest \neg **accept**-worlds and the closest **accept**-worlds (which are **accept** \wedge \neg **write**-worlds), the deontically best ones are in the former, so there exists a \neg **accept**-world which is good and no **accept**-world is good. The second *ought* adds the closest **accept** \wedge **write**-worlds to the modal horizon, and the best worlds are now **accept** \wedge **write**-worlds. Thus, it is true that there exists an **accept** \wedge **write**-world which is good and no **accept** \wedge \neg **write**-world or \neg **accept**-world is good.

SELECTED REFERENCES von Fintel 2001. Counterfactuals in a dynamic context. von Fintel 2012. The best we can (expect to) get? Challenges to the classic semantics for deontic modals. Herburger 2015, *SuB* 19. Lassiter 2011, PhD thesis. Lewis 1973, *Counterfactuals*.