Expected Value of Conditionals and Expected Utility:
A Probabilistic Account of Conditional Evaluative Constructions

WooJin Chung (New York University)

Introduction This paper proposes that the Korean construction that is used to convey obligation (cf. English *must*, *ought*) suggests to maximize expected utility, as proposed in the decision theory literature (Gibbard and Harper 1978, among others). Korean and Japanese do not make use of an auxiliary or a verb to express deontic concepts. Rather, they are expressed in terms of a conditional and an evaluative predicate, and thus they have been dubbed conditional evaluative constructions (CECs; M. Kaufmann 2018). I will focus on the Korean construction which utilizes an *only if* conditional as exemplified in (1), but the analysis to be offered can easily be extended to the Japanese one. In Korean, context determines whether *-(e)ya* 'only if' is interpreted indicatively or counterfactually. I focus on the counterfactual interpretation for reasons of space, but I would like to note that the indicative interpretation of (1) is in accordance with evidential decision theory.

(1) ne-nun aleppo-ey ka-**ya** **toy**-n-ta.
you-TOP Aleppo-to go-**only.if** **EVAL**-PRES-DECL
'You ought to go to Aleppo / (Lit.) Only if you (were to) go to Aleppo, good/EVAL.'

I make the following two assumptions and compositionally derive the result: (i) adopt a version of S. Kaufmann's (2005) probabilistic account of conditionals and (ii) interpret the evaluative predicate *toy* 'EVAL' as a function of worlds that returns the utility value of the world argument.

Conditionals and probability S. Kaufmann introduces a probabilistic account of conditionals, where sentences are interpreted w.r.t. a valuation function V which is a function from propositions/conditionals to real values. I will present a version of the analysis that omits the time-related parameters and does not assume Lewis's (1973) strong centering. The valuation function (i) returns 0/1 for false/true propositions, and (ii) returns the degree of support for the consequent (i.e., expected value) given the antecedent for conditionals. The interpretation of counterfactuals additionally conditions on facts that are *causally independent* of the antecedent (cf. Pearl 2000). Causal independence is interpreted with respect to (i) Φ: a set of causally relevant propositions singled out from the set of all propositions and (ii) $\prec$: a strict partial order where $X \prec X'$ reads as "the probability of X' depends on whether or not X occurs". The pair $\langle \Phi, \prec \rangle$ uniquely determines a causal graph that characterizes the causes and effects. The gist of the proposal is provided below:

(2) Causal independence
Given a causal structure $\langle \Phi, \prec \rangle$, for all $X, X' \in \Phi$: X' is *causally independent* of X iff $X \nprec X'$

(3) $V(\phi \mathrel{\Box\!\!\rightarrow} \psi)(w) = \mathbb{E}[V(\psi) \mid \phi, X_1, ..., X_i] = \sum_j V(\psi)(w_j) * Pr(\{w_j\})$
where X_1, ..., X_i are facts of w that are causally independent of ϕ
and $w_j \in \cap\{V(\phi), V(X_1), ..., V(X_i)\}$

Extending the domain of the valuation function In S. Kaufmann's framework, the valuation function V is defined over propositions and conditionals. I extend its domain and let it additionally take evaluative predicates. As for the interpretation of the evaluative predicate *toy* 'EVAL', $V(\text{EVAL})(w)$ returns the utility value of w. This is reminiscent of how Lassiter's (2017) predicates goodness of worlds, although he eventually lifts the domain of assessment from worlds to propositions. What will be shown is that the compositional semantics of (1) naturally lifts the domain of assessment from worlds to propositions and thus motivates Lassiter's stipulation.

Expected value of conditionals and expected utility As exemplified in (1), 'ought ϕ' effectively translates to 'only if ϕ, good/EVAL' in Korean. I will first leave out the exhaustifier 'only' and interpret 'if ϕ, good/EVAL' in a probabilistic framework. The formula in (4) is derived from (3) by simply replacing the consequent ψ with the evaluative predicate EVAL. The upshot is that the valuation function returns the utility value of a given world, instead of the truth value (0 or 1) of a proposition. The value of '$\phi \mathrel{\Box\!\!\rightarrow}$ EVAL' is the expected utility of the worlds conditioned on the antecedent and facts that are causally independent of the antecedent.

(4) Interpretation of 'if ϕ, good/EVAL' (counterfactual reading)

$$V(\phi \mathrel{\Box\!\!\rightarrow} \text{EVAL})(w) = \mathbb{E}[V(\text{EVAL}) \mid \phi, X_1, ..., X_i] = \sum_j V(\text{EVAL})(w_j) * Pr(\{w_j\})$$

where X_1, ..., X_i are facts of w that are causally independent of ϕ and $w_j \in \cap\{V(\phi), V(X_1), ..., V(X_i)\}$

Converting real numbers to a bivalent representation The value of a counterfactual, irrespective of whether its consequent is a proposition or an evaluative predicate, can be converted to a bivalent representation (true (1) or false (0)) by invoking the thresholding operation (Lassiter 2017). If the value of a counterfactual is greater than a contextually determined threshold, we can map the value to true (1); if the value is less than or equal to the threshold, we can map the value to false (0). Intuitively, if the expected value of the consequent ψ given the antecedent ϕ is sufficiently high, '$\phi \mathrel{\Box\!\!\rightarrow} \psi$' can be rendered true. Likewise, if the expected utility of the counterfactual ϕ-worlds is sufficiently high, '$\phi \mathrel{\Box\!\!\rightarrow}$ EVAL' is rendered true.

Exhaustifying the counterfactual A full analysis of (1) is given by applying the thresholding operation to (4) and exhaustifying the outcome. What is additionally conveyed due to 'only' is that for every alternative γ to ϕ, '$V(\gamma \mathrel{\Box\!\!\rightarrow} \text{EVAL})(w)$' returns a value that is less than or equal to the threshold (i.e., not sufficiently high). The upshot is that the expected utility of the counterfactual γ-worlds is not sufficiently high, whereas that of the counterfactual ϕ-worlds is sufficiently high.

Connection to decision theory Causal decision theory partitions the set of worlds into act-independent states s_i and calculates the expected utility of a choice ϕ by summing over the product of (i) the probability that s_i would obtain if ϕ were the case (i.e., $Pr(\phi \mathrel{\Box\!\!\rightarrow} s_i)$) and (ii) the utility value of the outcome jointly determined by the act ϕ and the state s_i (i.e., $o[\phi, s_i]$). This amounts to calculating the expected utility of the counterfactual ϕ-worlds.

(5) $EU_{CDT}(\phi) = \Sigma_i Pr(\phi \mathrel{\Box\!\!\rightarrow} s_i) * u(o[\phi, s_i])$

After calculating the expected utility of every available choice, the choice with the best expected utility is recommended, typically via an *ought*-statement. My analysis of (1) effectively reproduces causal decision theory, as it conveys that the counterfactual prejacent-worlds have the best expected utility (only the counterfactual ϕ-worlds have the expected utility higher than the threshold). In the presentation, I further point out that my analysis improves on causal decision theory as it can additionally resolve the paradox of supererogation (Heyd 1982, Lassiter 2017), which questions whether it is possible to systematically distinguish duties from supererogatory acts while maintaining the cherished relation between *ought* and the conception of goodness.

SELECTED REFERENCES **Gibbard & Harper 1978**. Counterfactuals and two kinds of expected utility. **Heyd 1982**. *Supererogation*. **S. Kaufmann 2005**. Conditional Predictions. **M. Kaufmann 2018**. What 'may' and 'must' may be in Japanese. **Kratzer 1991**. Modality. **Lassiter 2017**. *Graded Modality*. **Lewis 1973**. *Counterfactuals*. **Pearl 2000**. *Causality*.