Analyzing *intentionally* with local listening and would-be preventers

Bridget Copley and Clémentine Raffy, SFL (CNRS/Paris 8)

The Knobe effect: What does it mean to do something intentionally? (1a) is often reported false but (1b) is often reported true, in contexts where where helping/harming the environment is a side-effect of the chairman's action, even though the chairman says in both cases she doesn't care about helping/harming the environment (Knobe 2003; the "Knobe effect").

(1) a. Help condition: "The chairman intentionally helped the environment."
b. Harm condition: "The chairman intentionally harmed the environment."

Our contribution here is twofold. We propose that causal models with ***local listening***, where each edge (arrow) is associated with its own dependency function on truth values (Copley, to appear), can shed light on the Knobe effect. The idea is that the only functions that can be associated with arrows are those where the influenced truth value actually depends on the influencing truth value; otherwise there would be no influence (this is "listening" as in Pearl 2000; here it is "local" to each arrow). Another contribution is the proposal that the meaning of *intentionally* relies on the agent being a ***would-be preventer*** (McGrath 2005).

New data: In support of this idea, note that French *laisser* 'let' requires would-be preventer subjects (Raffy 2021). In this it contrasts with English *let*, which does not have this requirement. Given the chairman scenario, (2a), corresponding to the Help condition, is odd, while (2b), corresponding to the Harm condition, is felicitous.

(2) a. ??Le PDG a laissé les employés améliorer l'environnement.
the chairman AUX let the employees better the-environment
'The chairman let the employees help the environment.'
b. Le PDG a laissé les employés nuire à l'environnement.
the chairman AUX let the employees harm to the-environment
'The chairman let the employees harm the environment.'

Local listening: Classically in causal models (e.g. (3a)) the value of an endogenous variable Y is given by a function on all the variables that Y depends on, as shown in (3b). Sloman et al. (2012) use causal models fruitfully in an analysis of the Knobe effect. However, they use probabilities as the values of the variables, which is not useful for (most) formal semantic approaches. An approach using truth values can, however, have similar flexibility: Following Copley 2021, we alter the framework such that each arrow corresponds to its own function, representing the dependency that occurs if all other nodes are erased ("if all else is equal"). Where conflicts arise, an otherwise expected influence can be blocked from determining the value of the endogenous variable, which we notate using a double-barred arrow: $X \nrightarrow Y$.

(3) a. $X \rightarrow Y \leftarrow Z$ b. $F(X, Z) = Y$ c. $F(X) = Y$ and/or $F'(Z) = Y$

Indifference and disjunctive values We assume a third truth value "indiff" representing indifference, for nodes representing desires such as those of the CEO. We also allow for the returned value of an arrow function to be a disjunction between two truth values; such a disjunction licenses either of its values for the node in question.

(4) a. **Meaning of *intentionally*:** Let $D_{@p}$ represent an desire toward either p or $\neg p$. *x intentionally p* presupposes the model in (5), and is true iff $D_{@p}$ is a would-be-preventer for p.
b. $D_{@p}$ is a would-be-preventer for p iff there is a path via arrows from $D_{@p}$ to R_p with all values licensed, and there is a possible value of $D_{@p}$ that licenses $R_p = 0$.

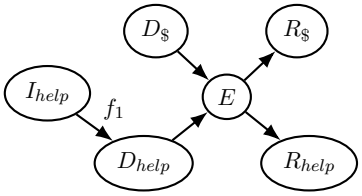
(5)
$$\underbrace{\underset{\text{whether desire-@p}(x)}{D_{@p}} \xrightarrow[influences]{}}_{\text{contributed by } \textit{intentionally}} \quad \underbrace{\underset{\text{whether } \exists e : \text{agent}(x, e)}{E} \xrightarrow[influences]{} \underset{\text{whether } \exists e' : p(e)}{R_p}}_{\text{contributed by rest of sentence}}$$

In words: whether the result occurs depends causally on whether the agent's action occurs (this is the not-at-issue meaning), and whether the agent's action occurs depends causally on whether the agent is a has an intention about the result and could have an intention to prevent the result (this is the at-issue meaning).

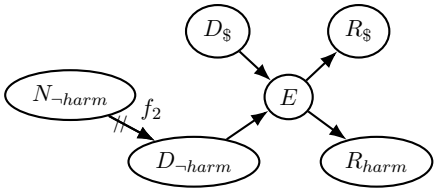
Models for the Help and Harm conditions: Along the lines of the above discussion, we argue for the functions below for ideals (e.g. I_{help}) and for norms (e.g. $N_{\neg harm}$). Having an ideal ($I_{help} = 1$) doesn't require you to want to realize it, while holding to a norm ($D_{\neg harm} = 1$) does.

(6)

a. Help condition



b. Harm condition



Function f_1 associated with $I_{help} \rightarrow D_{help}$ in (6a)

I_{help}	D_{help}
1	1∨indiff
0	0

Function f_2 associated with $N_{\neg harm} \rightarrow D_{\neg harm}$ in (6b):

$N_{\neg harm}$	$D_{\neg harm}$
1	1
0	$1 \vee 0$

What decides the judgments, according to (5), is whether there is a licensed line of the truth table from the desire $D_{help/\neg harm}$ to E such that $E = 0$. If D is not influenced by anything and can freely choose, then the CEO is a would-be preventer. However, if D is influenced by another node, it may not allow for such a line in the table. We assume that if the context doesn't make us block the D to E influence, it remains in the model. Because the actual value of I_{help} licenses the actual value of D_{help} (namely, indiff; see f_1), we don't block the arrow between those nodes, and the actual value of I_{help} does not permit $D_{help} = 0$, so the CEO cannot be a would-be preventer, and (1a) is false. But because $N_{\neg harm}$ does not allow $D_{\neg harm} =$ indiff (see f_2), we have to block the $D_{\neg harm}$ to E influence. This blocking allows $D_{\neg harm}$ to counterfactually have the value 0 and thereby make $E = 0$, making the CEO a would-be preventer and (1b) true.

We will further show how this analysis works for Machery's (2008) "Smoothie" scenario. The agent is not a would-be preventer in (7a), in the version where the action is less typically judged intentional, but is one in (7b), in the version where the action is more typically judged intentional.

(7) a. ??Le client a laissé l'employé lui donner une tasse commémorative.
the customer AUX let the-employee him give a cup commemorative
'The customer let the employee give him a commemorative cup.'

b. Le client a laissé l'employé lui faire payer 1 dollar de plus.
The customer AUX let the-employee him make pay 1 dollar of more
'The customer let the employee charge him a dollar extra.'

References: Copley, B., to appear. Reconciling causal and modal representations for two Salish out of control forms. WCCFL 39. **Knobe, J. 2003.** Intentional action and side effects in ordinary language. *Analysis*, 63, 190–93. **Machery, E., 2008.** The folk concept of intentional action: Philosophical and experimental issues. *Mind & Language,* 23(2), 165-189. **McGrath, S., 2005.** Causation by omission: A dilemma. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition,* 123(1/2), 125-148. **Raffy, C. 2021.** Letting *in Romance.* PhD diss., Univ. Paris 8 & Univ. zu Köln. **Sloman, S.A., Fernbach, P.M. and Ewing, S., 2012.** A causal model of intentionality judgment. *Mind & Language,* 27(2), 154-180.