Forbid is not order not

Bridget Copley & Alda Mari

Structures formelles du langage (Centre National de la Recherche Scientifique/Université Paris 8), Institut Jean Nicod (Centre National de la Recherche Scientifique/Ecole Normale Supérieure/Ecole des Hautes Études en Sciences Sociales/Paris Sciences et Lettres)

1. Introduction

Since the seminal work of Hintikka (1962) on belief, attitudes have been understood as introducing a universal quantification over a modal base, as exemplified for believe in (1):

\[
[[\alpha \text{ believe } p]]^{i,Dox} = \lambda w \lambda p [\forall w' \in Dox(i,w) p(i)]
\]

Modal bases are lexically determined, and, in the most sophisticated refinements of the Hintikka initial view, authors have proposed that attitudes can feature several modal bases and introduce quantification over subsets of those. In this line of thought, Heim (1982) was the first extending this treatment to bouletics (and see for subsequent proposals Anand and Hacquard (2008) and Portner and Rubinstein (2020)). Heim proposed that bouletics establish an ordering over doxastic alternatives. Giannakidou and Mari (2016, 2021) and Mari and Portner (2021) propose a unitary treatment for all attitudes including epistemics, as exemplified for want in (2):

\[
[[\alpha \text{ want } p]]^{i,Dox,Boul} = \lambda w \lambda p [\forall w' \in \text{Best}_{Boul}(Dox(i,w)) p(i)]
\]

The notional categories of orders and prohibitions have received much attention in the literature on imperatives on the one hand (e.g., Condoravdi and Lauer (2012); Kaufmann (2011)) and in the literature on attitudes and most specifically expletive negation on the other (e.g., Mari and Tahar (2020)). Across these works, order would be treated as a universal preferential attitude operating over the common ground CG, where Order would be a subtype of Boul, as shown in (3):

\[
\]

1

\footnote{We would like to thank the NELS organizers, audience, and editors.}

\footnote{See in particular Kaufmann (2011) on imperatives as attitudes.}
Among the questions that this literature has left open is the one of the lexical semantics of forbid, in contrast and comparison with the lexical semantics of order. This contrast will be instrumental for us to show that the modal approach does not account for the meaning of these two verbs, which instead, we argue, should be analyzed along causal lines. In section 2 we discuss our arguments against what we take to be the only viable possible worlds-based analysis for forbid, namely, that forbid is equivalent to order not. In section 3 we propose a causal model-based account that, we argue, accounts rather naturally for data the order not analysis cannot. Notably, in our proposal, forbid is not equivalent to order not. Section 4 concludes.

2. The modal treatment of order and forbid and its shortcomings

We may start from the intuition that forbid is, in some sense to be made explicit, a kind of negative order. In a possible worlds approach, there are two ways to cash out this intuition, only one of which ends up being a viable contender. The first (and non-viable) way would be to scope negation high, over the ordering source, as in (4). We assume for the sake of the argument a modal semantics à la Kratzer (1991) for attitudes, where Order is an modal base that complies with the orders of the attitude holder α in the actual world w.

(4) a. \[\alpha \text{ order } p \text{ } i.CG.Boul = \lambda w \lambda p[\forall w' \in \text{BestOrder}(CG(i, w)) p(i)]\]
   (3)
   b. \[\alpha \text{ forbid } p \text{ } i.CG.Boul = \lambda w \lambda p[\forall w' \in \text{Best}\neg\text{Order}(CG(i, w)) \neg p(i)]\]

According to (4b), p is true in all worlds that comply with what α does not order. There are several ways to refine this idea; e.g., replacing Order with INT, an intentional modal base that encodes a larger notion of intention/volition rather than orders, along the lines proposed by Grano (2017). A problem for any analysis of this kind, however, is that it predicts that (5a) should be equivalent to (5b). However, the sentence in (5a) conveys that not all the non-desire worlds are ones where you enter, and this is not equivalent to what the sentence in (5b) conveys, namely, that there are desire worlds where you enter.

(5) a. You are not forbidden to enter. On (4b) analysis: \(\forall w' \neg\text{ORD}(w, \alpha) P(w')\)
   b. You are permitted to enter. On (4b) analysis: \(\exists w' \text{ ORD}(w, \alpha) P(w')\)

Thus, the ‘high negation’ analysis is not viable. A more promising modal solution is to scope negation low, under the ordering source, such that forbid means ‘order that not p’, as in (6b):

(6) a. \[\alpha \text{ order } p \text{ } i.CG.Boul = \lambda w \lambda p[\forall w' \in \text{BestOrder}(CG(i, w)) p(i)]\]
   (3)
   b. \[\alpha \text{ forbid } p \text{ } i.CG.Boul = \lambda w \lambda p[\forall w' \in \text{Best}\neg\text{Order}(CG(i, w)) \neg p(i)]\]

\(^2\)In this it would be one of the ‘negative’ attitudes of apprehension and frustratives (see Mari and Tahar (2020) and references therein).
Forbid is not order not

This ‘low negation’ analysis, where forbid means order not, does better than the ‘high negation’ analysis in (4b), as it does not predict equivalence of (5a) and (5b). And indeed, if \( \alpha \) orders that \( \neg p \) and all else is equal, \( p \) ends up false; similarly, if \( \alpha \) forbids that \( p \), and all else is equal, \( p \) ends up false. However, the ‘low negation’ modal analysis does not correctly capture the meaning of these verbs, because it misses out on a series of observations which pertain to two differences and one similarity between forbid and order not:

(i) Aspect: In English, simple present eventives with an ongoing reading are not possible in ordinary conversational (i.e., non-narrative) contexts, as shown in (7a, b), and eventives under must lack epistemic readings, as shown in (8a,b):

(7) a. #Mary walks.  
    b. Mary knows French.

(8) a. Mary must walk.  
    b. Mary must know French

By these tests we can see that order not has only an eventive reading, while forbid can have a stative reading, as shown in (9) and (10); note that while both (10a) and (10b) can have an epistemic reading, in (9a) there must be an event of ordering not to go every time he wants to go to the park, while in (9b) there need only be a single event of forbidding. We take this to mean that the stative prejacent in (9b) has to be a habitual, while the stative prejacent in (9b) refers to a true state or stance of forbidding (initiated, perhaps, by a single speech event). If (6b) is adopted as the analysis, all this must be stipulated.

(9) a. #John’s parents order him not to go to the park after dark.  
    b. John’s parents forbide him to go to the park after dark.

(10) a. John’s parents must order him not to go to the park after dark.  
    b. John’s parents must forbid him to go to the park after dark.

(ii) Animacy: Forbid, but not order, admits inanimate subjects, as shown in (11). The example in (12) confirms that forbid also has an eventive reading, given that statives lack agents and an agent is required for deliberately. These data can be described, again, by stating that order is always eventive, and forbid can either be eventive or stative. However, again, if (6b) is adopted, all this must be stipulated.

(11) a. #The sign ordered us not to enter.  
    b. The sign forbade us to enter.

\(^3\)The policeman must forbid John to go to the park after dark does not have an epistemic reading, contrary to (10b). We think this means that the state in stative forbid sentences is individual-level, rather than stage-level (Carlson 1977), since, perhaps, we wouldn’t expect the policeman to have an individual-level forbidding stance about John’s going to the park after dark, but we would expect John’s parents to have such a stance. For more on stativity and causality and the individual/stage-level distinction, see Jayez and Mari (2005) and Mari (2005).
(12) The policeman deliberately forbade me to enter.

(iii) Authority: The matrix agent is also presupposed to have authority over the lower agent, as seen in (13a) where the speaker’s toddler does not have authority over the speaker, and (13b), where the speaker accepts that whatever the sign says with respect to $p$, it influences their behavior. But in (6b) as with any deontic modal analysis, the fact of authority has to be stipulated, and it is not immediately obvious how to represent authority relations.

(13) a. #My toddler didn’t order/forbid me to go.
   b. The sign didn’t forbid me to enter.

We hypothesize that *forbid* and *order* not differ in how they represent the not bringing about that $p$. A proper understanding how falsity of $p$ comes about, and explanations for the facts presented above, is achieved by considering the causal representations in which the meanings are grounded.

3. Formal tools

To explain the aspectual, animacy, and authority facts given above, we propose to analyze *order* and *forbid* using causal models, augmented by a few other tools and assumptions.

3.1 Basics of causal models

A *causal model* (Pearl (2000), or see a gentle introduction to causal models in Chapter 1 of Pearl and Mackenzie (2018)) is a directed acyclic graph representing the structure that causal relations give to the world (or more precisely, to our conceptual model of the world).

The vertices, or nodes, of such a graph are *variables*, which can have various values. The variables are connected by a set of directed edges, or arrows. The causal model conveys that the values of some variables influence the values of other variables, according to both the arrows of the graph (which indicate the fact and direction of causation) and the functions associated with the arrows (which give more information about which values go together). For instance, consider the causal model whose variables are $A$ and $B$, with an arrow pointing from $A$ to $B$, as in (14).

(14) $A \rightarrow B$

For the purpose of providing denotations for *order* and *forbid*, we will not need to deal with models any more complex than this. The structure in (14) represents a state of affairs where the value of $B$ is dependent in some way on the value of $A$, without specifying which values of $A$ and $B$ go together. Information about which values go together is something that comes from world knowledge; the value of $B$ can be expressed via a function on the value of $A$. 
3.2 Flavors of influence and arrow functions

At the heart of our proposal is the idea that order (not) and forbid differ in a dimension that cannot be represented in any modal analysis. While ordering is, broadly, causing or getting someone to do something, forbidding is preventing someone from doing something. We can think of this distinction as a distinction between stimulatory and inhibitory ‘flavors of influence’.

We will model flavors of influence using arrow functions. Given a simple causal model like the one above in (14), we can think of the arrow itself as being associated with an arrow function representing which values of $A$ go with which values of $B$. Crucially, the arrow function must be such that the value of $\overline{B}$ must ‘listen’ to the value of $\overline{A}$ (Pearl 2000); that is, a difference in the value of $\overline{A}$ must make a difference in the value of $\overline{B}$.

We use the set of truth values as both the domain and the range of arrow functions. There are only four possible one-place functions from $\{0,1\}$ to $\{0,1\}$, shown in Table 1.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$x$</th>
<th>$y$</th>
<th>$x$</th>
<th>$y$</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

But only the first two functions in Table 1 can serve as arrow functions. The reason is that, in order to represent influence, whatever the flavor, the value of the effect must ‘listen’ to the value of the cause. This means that any difference in the value of the cause (here represented by $x$) must make a difference in the value of the effect (here represented by $y$). This is the case, for instance, in the first function in Table 1: When $x = 1$, $y = 1$; and if we were to change the value of $x$ to 0, the value of $y$ would also change, namely, to 0. Likewise, in the second function, changing the value of $x$ leads to a change to the value of $y$. However, in the last two functions, changing the value of $x$ does not lead to a change in the value of $y$. This being the case, only the first two functions in Table 1 can serve as arrow functions. These two functions are familiar: the first is the identity function and the second is negation. However, here they will be playing a different role due to arrow functions’ association with causation.

In our simple causal models either of these functions can be associated with the arrow from $\overline{A}$ to $\overline{B}$, and this is how they correspond to the two flavors of influence we are using here, namely stimulation and inhibition.

<table>
<thead>
<tr>
<th>Stimulatory arrow function ($f^+$)</th>
<th>Inhibitory arrow function ($f^-$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{A} \xrightarrow{f^+} \overline{B}$</td>
<td>$\overline{A} \xrightarrow{f^-} \overline{B}$</td>
</tr>
<tr>
<td>1 1</td>
<td>1 0</td>
</tr>
<tr>
<td>0 0</td>
<td>0 1</td>
</tr>
</tbody>
</table>
What we are doing is reifying two flavors of influence—stimulatory and inhibitory—using a distinct arrow function for each one. Being able to do such a reification is precisely why causal models are useful for our purposes. The denotation of order, we argue, includes a stimulatory influence, while that of forbid includes an inhibitory influence. Because these are reified, we can attribute the aspectual and animacy differences in judgments of order not and forbid sentences to the difference between these two flavors of influences. Moreover, as we will see, not only the aspectual and animacy differences, but also the authority similarity, in fact fit very well with our proposal of these two flavors of influence.

By contrast, the analysis of forbid as order not as in (6b) is an explanatory dead end: Since it provides no difference between order not and forbid, it provides no way to explain the differences in aspectual and animacy judgments between order not and forbid. Furthermore, it provides no obvious path to characterizing the authority relation shared by order not and forbid. We must prefer, then, our proposal reifying flavors of influence.

3.3 The Common Causal Model

As its name suggests, the Common Causal Model is a common ground, a ‘field on which a language game is played’ (Stalnaker 2002:720). For our purposes this field takes the form of a causal model, so this is the Common Causal Model. Generally, utterances operate on the Common Causal Model. We can think of such operations as falling into two broad categories: accommodations, which are operations that reflect not-at-issue commitments, and moves, which are operations that reflect at-issue commitments. Here we are interested in just one kind of accommodation, which we will express with a GIVEN operator that takes a causal model as its argument and accommodates its presence as (part of) the Common Causal Model. We will also be interested in just one kind of move, namely to set the value of a variable; this is Pearl’s (2000) do operator.

We emphasize that the Common Causal Model is not a to-do list in the sense of e.g., Portner (2004). It is simply a formal object (specifically a causal model) which is identified as the Common Causal Model and which is operated on by utterances. A further point to underline, the importance of which we will see shortly, is that it's possible to have under-specified causal models as the Common Causal Model, in particular causal models that are underspecified for the values of the nodes. Where a two-node causal model has no values but does have an arrow function, what is represented in this model is exactly the fact of a certain flavor of influence. That is, we know what arrow function is being used, but it is left open which line of the arrow function table might reflect reality.

3.4 Intentions

In our framework, intentions participate as nodes in causal models. This assumption may be contentious in philosophy, but we take it as a closed issue in semantics that intentions are causal, given that causation is the organizing principle of (English) verbs and that, e.g., intentions and physical dispositions are frequently treated the same in causal structures (Copley 2018). We will assume that an intentional node can have either the value 1 (the entity in question holds the intention in question) or 0 (the entity in question does not hold
the intention). We will not notate the entity who holds the intention—here it is always the subject of the verb—but we will notate the proposition to which the intention is directed in the complex name of the node. In *order not* and *forbid*, the intention is always toward \( \neg P \), but we find it convenient to remind ourselves of this. So we will write the label of the intentional influencing node as \( I^{\neg P} \) for both *order not* and *forbid*.

We intend any theory of intention to be compatible with this framework, as any theory of intention will have a holder of the intention and a proposition to which it is directed. The reader can use their favorite theory to explicate in a more comprehensive manner the individuation and use criteria of intentional nodes.

4. Proposal

Our proposal for *order not* and *forbid* is given in (15a,b). Both *order not* and *forbid* involve an intention toward \( \neg P \). Recall that the GIVEN operator accommodates its causal model argument (here, a causal model underspecified for variable values) into the Common Causal Model, and the do operator sets the value of a variable.

\[
\begin{align*}
\text{(15) a.} & \quad \text{order not } P: \text{ GIVEN } (I^{\neg P}) \xrightarrow{\text{do}} (\neg P), \text{ do } (I^{\neg P} = 1) \\
\text{b.} & \quad \text{forbid } P: \text{ GIVEN } (I^{\neg P}) \xrightarrow{\text{do}} (P), \text{ do } (I^{\neg P} = 1)
\end{align*}
\]

We see that *forbid P* makes \( P = 0 \) and *order not P* makes \( \neg P = 1 \). This corresponds to the ‘low negation’ modal analysis in (6b), as far as it goes, and they are similarly truth-conditionally equivalent, as desired. However, there is also a difference between (15a) and (15b) in the arrow function (representing the flavor of influence). Because we explicitly, rather than implicitly, represent flavors of influence using arrow functions, we can derive, rather than stipulate, the aspectual and animacy differences between *order not* and *forbid*.

(i) Aspect: Recall that *order* sentences can only be eventive, while *forbid* sentences can be either eventive or stative. Let’s consider events to involve inputs of energy (Copley and Harley 2015) while states do not. We will not formally model the difference between stative and eventive nodes here, but we propose that there is a connection between flavor of influence and eventivity:

\[
\begin{align*}
\text{(16) a.} & \quad \text{Stimulation (as in *order*) requires energy } \Rightarrow \text{ requires an eventive influencing node} \\
\text{b.} & \quad \text{Inhibition (as in *forbid*) does not require energy } \Rightarrow \text{ does not require an eventive influencing node}
\end{align*}
\]

\[\text{4The idea is essentially that to block something from happening, no energy is needed. The picture must be more complex than this, since there are stative causatives that appear to be stimulatory (e.g., *The flowers brighten the room*), but for these data, where any event is constrained to be a speech event, the generalization seems to hold.}\]
An influencing intentional event for *order* or *forbid* would have to be an ordering or forbidding speech event. So what (16) says is that it’s possible to refer to an influencing intentional state or stance in the case of inhibition, but not in the case of stimulation.

(ii) Animacy: The proposal in (16) is also why inanimate subjects are possible with *forbid* but not with *order*, as in (11). Only entities that can perform speech events can be the agent of a stimulatory influence. Inhibitions, however, support either agents or causers when their influencing node is stative (i.e., a forbidding stance).

(iii) Authority: Recall that *order* (not)* and *forbid* both presuppose that the referent of the matrix subject has authority over the referent of the embedded subject. If *forbid* meant *order not* as the ‘low negation’ analysis in (6b) would have it, the fact of authority would have to be stipulated, and there would not even be an obvious way to formally represent the authority relation. For us, however, authority for both *order* and *forbid* is modeled directly by the fact that the value of ¬P or P depends on the value of the intentional node, whatever it is. This analysis represents authority as a function: in other words, we claim that having authority means having an influence, and that influence cannot be represented uniquely by the correlation of two truth values. Influence means that the value of a variable (whatever it may be) has an effect on the value of a different variable (whatever it may be). These functions can be of different types, as we saw, depending on whether they are stimulatory or inhibitory. This corresponds to authority when the influencing node is an intention; whatever the boss wants, the boss gets.

One might object to the representation of *order* and *forbid* with stimulatory and inhibitory arrow functions on these very grounds, noting that the referent of the embedded subject can disobey.\(^5\) Thus, the objection goes, just because there is an intention toward ¬P does not mean that ¬P will end up true; therefore, the objection goes, our account does not explain the data. However, the fact that disobedient subordinates exist is not a problem for us, as long as we assume a defeasible ‘closed-world’ assumption. This assumption is that all relevant influences are included in the causal model.\(^6\) If the subordinate is disobedient, that means the value of their intention toward proposition in question is now a relevant influence on the value of P, so we must include it in the Common Causal Model. How exactly to do this is more than we can show in this space, but see Copley (to appear) and Copley & Kagan (to appear) for details.

5. Discussion

The idea that the influencing node of a stimulation can only be energetic/eventive while the influencing node of an inhibition can also be non-energetic/stative could well shed light on the reason why e.g., *Wearing a shirt* is not a possible sign one might see but *No smoking* is (Iatridou 2021). We assume that *Wearing a shirt* involves a stimulation and *No smoking* involves an inhibition. Since signs are not animate they cannot perform speech events. *Wearing a shirt* on a sign would thus require a stative influencing node for a stimulation,

\(^5\)In Giannakidou and Mari (2021) these cases are analyzed using non-veridical attitudes.

\(^6\)Formally this assumption does much the same job as partitions do in possible-world semantics.
Forbid is not order not

which is not possible; but there is no such problem with the inhibition in *No smoking*, which licenses statives.

We further suspect that the reason why *forbid* (as well as similar predicates such as *prevent, keep, prohibit, and inhibit*), can also take *from Ving* as a complement, is precisely that this class of verbs adds an inhibitory influence function to the Common Causal Model, that this function is available for the purposes of verbal selection, and that *from* can only be selected by inhibitory verbs.

On a broader note, our proposal draws out a key distinction between actions and propositions that has been historically obscured by our common modal logic heritage. We have taken inspiration from works such as Farkas (1988), Portner (2004), Condoravdi and Lauer (2016),7 and Grano (2017) which represent tremendous advances in seeing actions for what they are, even while hampered by the difficulty of representing actions in a system that was set up to only represent propositions. What we propose here is two steps further: first, to give up on using propositions for these verbs, outside of the intentional node label; and second, to treat actions as a subclass of influences, where the latter may or may not see their intended effect occur.

Effectively, on our account, *order* and *forbid* do not describe a mental state toward a proposition or even a preference. They describe an influence that the attitude holder intends to see achieved by the addressee. That is, *ceteris paribus*, the attitude-holder’s influence has an *impact* on the addressee—if all else is equal—in such a way that the addressee brings about the truth of the intended proposition. These influences are represented directly by the arrow functions, and come in two ‘flavors of influence’: stimulation and inhibition. In the case of a disobedient subordinate, all else is *not* equal. A different causal model must be used which represents the subordinate’s intention, and it is this intention that determines the outcome.

References


7If it is said to be effective, a preference structure is consistent and realistic with respect to an information state, i.e., the mental space of the relevant individual. These features are designed as minimal requirements to describe action-relative preference structures: consistency ensures that an agent ranks conflicting propositions; realism just rules out propositions that are believed not to lead to action, see Condoravdi and Lauer (2012), Condoravdi and Lauer (2016).


Bridget Copley, Alda Mari
bridget.copley@cnrs.fr, alda.mari@ens.fr