

Root and Epistemic Modals: Causality in Two Worlds

Author(s): Eve E. Sweetser

*Proceedings of the Eighth Annual Meeting of the Berkeley Linguistics Society* (1982), pp. 484-507

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via <http://linguistics.berkeley.edu/bls/>.

---

*The Annual Proceedings of the Berkeley Linguistics Society* is published online via [eLanguage](#), the Linguistic Society of America's digital publishing platform.

Root and Epistemic Modals:  
Causality in Two Worlds

Eve E. Sweetser  
University of California, Berkeley

0. Introduction. Modality has always been a tantalizingly elusive area for linguistic analysis, but perhaps its most elusive aspect is the connection between the "root" (or deontic)<sup>1</sup> and "epistemic" senses of modal expressions. Linguists have characterized as root those meanings which denote real-world obligation, permission, or ability (as in example 1); and as epistemic those which denote necessity, probability, or possibility in reasoning (as in 2).

- (1) John must be home by ten: Mother won't let him stay out later.  
(2) John must be home already: I see his coat.

So far, however, no plausible manner of unifying or linking the two senses of modal verbs like must has been proposed, despite a number of attempts which I shall discuss later. Indeed, much recent linguistic work seems to treat English modal verbs as essentially cases of homonymy rather than ambiguity, tacitly assuming that epistemic and root modality are unrelated. (cf. Lyons 1977, R. Lakoff 1972a) Root modal meanings are often treated as lexical predicates involving force or obligation, while epistemic readings are treated as combinations of logical operators. But if English root modals share lexical form with their epistemic counterparts by (historical) chance, this chance homonymy is widespread. There is a strong crosslinguistic tendency for lexical items to be ambiguous between these two semantic domains; many unrelated languages<sup>2</sup> are alike in having some set of predicates (often including a relatively small morphosyntactically distinct set) which carry both the root and epistemic modal readings, as English modal verbs do.

The present study will argue for a unified semantic analysis of root and epistemic modality. I shall suggest that root modal meanings are extended to the epistemic domain precisely because we view our reasoning processes as being subject to compulsions, obligations, and other modalities, just as our real-world actions are subject to modalities of the same sort. Nor is modality the only area where we treat the epistemic world as analogous to the sociophysical world: the root-epistemic modal contrast is only one example of our linguistic treatment of the causality of reasoning processes in terms of the causality of events and actions. An examination of our use of speech-

act verbs, of adverbial elements such as anyway, and especially of causal conjunctions, shows that all of these can be applied to the epistemic world as well as to the real world.

There is strong historical, sociolinguistic, and psycholinguistic evidence for viewing the epistemic use of the modals as an extension of the root meaning (rather than root as extension of epistemic, or both as subsets of some superordinate). Historically, the English modals developed from non-modal meanings (such as physical strength or force) to "deontic" modal meanings, and later still broadened to include the epistemic readings as well. (cf. Shepherd 1981, Ehrman 1966). Shepherd's work on Antigua Creole gives some evidence that creoles first develop their expression of root modality before going on to extend that expression fully to the epistemic domain. And studies of child language (Kuczaj 1979, Shepherd 1981) have revealed that children acquire the deontic senses of modal verbs earlier than the epistemic senses.

I shall begin, therefore, by putting forward an analysis of root modality which I have chosen because it is readily extendable from the sociophysical to the epistemic domain. Given our understanding of mental "forces" in terms of real-world forces, this analysis of modal semantics can apply unmodified to the epistemic world. And finally, I shall expand the analysis from the area of modality to propose a single semantics of causality for the sociophysical and epistemic worlds.

### I. The Root Modals in English

One of the main obstacles to the evolution of a unified understanding of modality has been the fact that semantic analyses of root modality were not systematically relatable to logical necessity or probability. So we must choose our root modal analysis with care, if we hope to make it mesh with epistemic modality. Talmy (1982) has suggested that the semantics of root modality is best understood in terms of force dynamics, that is in terms of our linguistic treatment of forces and barriers in general. Thus, for example, permitting (e.g. may, let, and allow) is an instance of taking away (or keeping away) a potential barrier of some kind. With let or allow, that barrier may be a physical one (as in 3) or a social one (as in 4); may seems more restricted to social permission.

(3) The crack in the stone let the water flow through.

(4) I begged Mary to let me have another cookie.

Adopting Talmy's basic idea of viewing modality in terms of forces and barriers, I shall offer tentative force-dynamic analyses of all the root modals. My primary

object will be to subsequently demonstrate that such analyses are possible and readily extendable to the epistemic domain, rather than to argue strongly for this specific set of analyses as they stand. It should be understood that I do not explicitly take my analyses from Talmy, except in the case of may, nor does he necessarily agree with my unification of root and epistemic modality. Further, he takes the purely physical level of force dynamics (e.g. a stone resisting water) as the most basic of all, while I prefer to view modality as basically referring to intentional, directed forces and barriers. Within the domain of intentional causality, I do feel (as Talmy does) that direct physical manipulation of the environment is more prototypical causality (and hence more prototypical modality) than is indirect or purely social manipulation (cf. Talmy 1976). But this paper will not attempt to deal with the relative basicness of different kinds of real-world forces in our understanding of causation; rather, I shall simply propose a force-dynamic analysis of modality, with the understanding that I am referring to generalized sociophysical concepts of forces and barriers.

May and must are perhaps the most clearly force-dynamic of the modals. Talmy's understanding of may in terms of a potential but absent barrier seems to me very reasonable, and can be viewed as a restatement of the standard analysis (e.g. "not require not") in terms of the more general concepts of forces and barriers. Must is equally readily understood as a compelling force directing the subject towards an act. Talmy would like to view must as a barrier restricting one's domain of action to a certain single act; and the physical result of force or constraint would be the same. But must has the force of an order to do something, a positive compulsion rather than a negative restriction. When I say "You must be home by ten," I indeed restrict my interlocutor's actions (or try to do so); but I do so by compelling the choice of some specific alternative. My intention is fixed not on the excluded alternatives but on the realization of the chosen alternative.

Can is far more difficult to pin down than may or must. Talmy analyzes it as parallel to may in structure, but with less tendency for the absent barrier to return to its position. This solution would of course explain the frequent overlap of can and may's semantic territories, but I think the overlap is equally explicable in terms of a more intuitively satisfying definition of can. Can denotes positive ability on the part of the doer; may denotes lack of restriction on the part of someone else. The closest physical analogy to can would be potential force or energy (note the Latin potential, referring to ability) - and perhaps the best force-dynamic definition I can give for

ability is to say that it is the human physical and social modality in terms of which we view potential energy in physics.

If we can permit ourselves an excursion into the simple physical domain for a moment, perhaps it will become clearer why can and may have such a tendency towards overlap. Let us view can as being the equivalent of a full gas-tank in a car, and may as the equivalent of an open garage-door. These two factors will exert certain similar influences on the situation: neither factor forces the car (or the driver) to travel a given path, and yet if either factor were reversed, then travel would be correspondingly restricted. The full tank is a positive enablement, while the open door is a negated restriction; yet the results are similar enough to allow a good deal of overlap in the larger force-dynamic schemata surrounding the two modalities. Thus it is not surprising to find can used to give permission: the remover of a barrier may even feel that in some sense this removal counts as an act of enablement. And of course, it is also politer to (cooperatively) enable than to invoke your restrictive powers by overtly refraining from exercising them.

We now come to ought, have to, and need to, which resemble must in denoting obligation or necessity: the difference is largely in the kind of obligation. Ought seems to be less strong than the others, and to have moral overtones, or at least to indicate that the obligation is one socially agreed upon between the imposer and the doer. Have to (as Talmy observes) has more of a meaning of being obliged by extrinsically imposed authority. And need implies that the obligation is imposed by something internal to the doer:

- (5) I have } to stay home, or Mom will get mad at me.  
?I need }
- (6) You have } to stay home, because I say so.  
?You need }
- (7) I need } to stay home tonight to study for the test.  
I have }

Either need or have to can be used in (7) because the obligation to study is an externally imposed one in one sense, and an internally imposed one in another ( the student is free to neglect studying, though at the risk of failing the test). Talmy would prefer to analyze have to, need to, and ought as barriers; I have once again some doubts about this viewpoint. Ought especially seems to me to indicate a positive compulsion; but need also refers to the necessity for some specific action or object, rather than to restrictions on other possible actions. My own analysis of

must, ought, have to, and need to is that they are different kinds of forces. Must has connotations of a directly applied and irresistible force, while have to, ought and need are resistible forces different with respect to their domains (social, moral) and/or sources of imposition (internal/external), as discussed above. Regarding the question of resistibility, note the contrasts in (8).

- (8) ?? I must } get this paper in, but I guess I'll  
           I have to } go to the movies instead.  
           I need to }  
           I ought to }

The basic point here is that within the limits of the meaning of each modal, anything that counts as a force can impose the relevant modality. Thus any internally rooted desire, lack, or compulsion can impose the modality need; and any social force which the subject participates in can count as conferring the obligation expressed in ought.

Finally, we come to the borderline modals will and shall (their distal forms, should and would, are highly modal). Shall and will can express simple futurity; but (as Palmer remarks with some surprise, after examining a large corpus) they don't usually do so in usage, despite grammar books. R. Lakoff (1972a) prefers to regard them as the strongest modals, on the grounds that the very strongest obligation or necessity is certainty of future action. Certainly the will in examples such as (9)-(10) seems volitional rather than future pure and simple.

- (9) All right, I'll do it; shake, mister.  
 (10) See if John will help you out. (=is he willing?)

Shall in my dialect (also in many of Palmer's examples) indicates the speaker or imposer (rather than the subject of the action) making him/herself responsible for the carrying out of the action. Thus (11)-(12) have a sense that the speaker undertakes to see to it or to command that the action be done; while in (13) the law is viewed as doing this.

- (11) You shall go, I insist on it.  
 (12) If Mr. Jones wants tickets for our concert, he shall have them.  
 (13) (The law decrees that) all citizens shall constantly carry violet parasols from 3/9/83 on.

The forces involved in (9)-(13) are those of volition and responsibility.

The purely future reading of will (shall has none

in my dialect) seems to indicate not some force or barrier, but a completed path to an action or intention. How this fits into a force-dynamic analysis (if at all) is a difficult question. The one mistake which I can clearly identify in some past analyses is the idea that future will is always epistemic, and concerns future truth-value. Like all the modals except present-tense shall, will has both a root and an epistemic reading - contrast the real futurity in (14a) with the epistemic futurity of knowledge in (14b).

(14a) He will be home in three hours.

b) He will be home by now; I just saw the lights go on.

In (14b) the person is or is not at home, in the present; the will is of future discovery or verification - "If we check, we will find out that he is home." When an action is in the future, of course its occurrence is automatically only knowable or verifiable in the future. But the epistemic use of will is an extension from the will of actual futurity to purely epistemic futurity: the actual event is not in the future, but only its verification. Note that so long as verification is future, the event can be past as easily as present - "future perfect" forms are thus ambiguous between a root will (perfectivity in the future) and an epistemic will (future verification of perfectivity):

(15a) He will have completed his requirements by the end of this term.

b) He will have completed his requirements long ago, of course - I don't know why I'm bothering to check the records.

The distal<sup>6</sup> forms of the root modals express past or conditional modality; distance in either a temporal or a causal sequence is thus marked identically. Could expresses past or conditional ability, and might (in those dialects where it has a root sense) a past or conditional absence of a barrier. Ought to and must have no morphologically distinct past forms: both of them can act as either present or past with respect to tense-sequencing in dependent clauses (e.g. He thinks he can/ought to vs He thought he could/ought to), but neither of them has an independent past or conditional form. Should has filled part of the distal slot for ought to; since shall is relatively rare, its distal form was perhaps freed to shift as needed within the modal system. It was a natural choice for this slot, since whatever a speaker is willing to assume responsibility for (should) is also something the speaker might conditionally agree was morally appropriate or obligatory

(ought). The pure past of ought, however, is usually represented by the periphrastic "be supposed to" form. Must is so specifically an expression of direct force that it seems natural for it to lack a distal form; when a past form is required, had to is used, but its meaning is not quite a distal must. Have to and need to have past forms; but like all conjugated English verbs, their past forms are not conditionals in main clauses - would have to and would need to are the conditionals, except in if-clauses. Finally, would expresses the distal form of both the future will and volitional-force will. In general, whatever modal forces or barriers the present form of a modal verb expresses, the distal form of the verb will express those forces conditionally or in the past.

## II. Epistemic modality as an extension of root modality.

A. Past unified analyses of modality. Given the tentative beginnings of a general analysis of root modality in terms of sociophysical forces, barriers, and paths of different kinds, let us now explore the results of transferring this view to the epistemic domain. We would like to achieve a unified analysis of modality. One direction taken by past "unified" analyses (e.g. Kratzer 1977) has been essentially to subsume the root meanings of the modals under very general epistemic readings; thus root can comes to refer to logical compatibility between a person's (or the world's) state and some event, while root must refers to logical necessity of the occurrence of some event, given the state of the world. Even if analyses such as Kratzer's did not have the drawback of ignoring intentionality entirely, they would still have the problem of explaining why the historical and developmental progression is from root to epistemic, rather than the other direction. A slightly more promising direction is that suggested in passing by Lyons (1977), namely that epistemic uses of the modals result from our understanding the logical necessity of a proposition in terms of the forces which give rise to the sociophysical necessity of the corresponding event in the real world. But this too falls down when closely examined: when (16) is uttered, the speaker does not really mean that somehow the proposition must be true because some real-world causes have brought about the relevant state of affairs, but rather that (s)he must conclude that it is true because the available informational premises cause him or her to reason thus.

(16) (looks at nametag) "You must be Seth Sweetser's sister."

Nonetheless, Lyons' idea is a more useful starting point than any of the analyses which assume the existence of



a superordinate modality that has deontic and epistemic subclasses. Ehrman's(1966) attempt to find superordinate "core meanings" for the modals resulted in some hopelessly vague analyses, and still left her with two separate meanings for may.

Boyd and Thorne(1969) and Tregidgo(1982)<sup>8</sup> in different ways propose analyses which allow epistemic modals to get readings referring to the necessity or permissibility of the act of stating, while root modals refer to necessity or possibility of the event described in the statement. This is getting warmer, but is still not quite correct, since in fact epistemic modals don't apply to our acts of stating, but to our acts of induction or deduction. Thus (16) does not express the speaker's compulsion to state that the addressee has a certain identity, but his compulsion to conclude that this is the case. Phrases like "I must say" or "I must tell you," which genuinely express modality applied to the act of speaking, have a completely different meaning from epistemic modals.

Finally, Antinucci and Parisi(1971) have suggested that belief figures in the semantics of epistemic modals. Thus they propose that must has two readings analyzable as in (17)-(18):

(17) You must come home. (deontic)

CAUSE ( $\{X$   
{Speaker}) (BIND (YOU COME HOME))

(18) You must have been home last night. (epistemic)

CAUSE (X) (BIND (BELIEVE (SPEAKER)(YOU BE HOME)))

Restated in English, this analysis proposes that epistemic modality binds the speaker to believe the proposition, while deontic modality binds the subject to do the action expressed in the proposition. Antinucci and Parisi are clearly on the right track. I would prefer to talk about conclusions rather than beliefs, since conclusions are precisely that class of beliefs which we are bound to adopt or not to adopt by our reasoning processes. Also, we shall see (in the next two sections of this paper) that an analysis of modality need not have separate semantic structures for root and epistemic modals; we need not view must as semantically ambiguous between CAUSE (BIND()) and CAUSE(BIND(BELIEVE())). Nor is it necessary for the imposer and impos-ee of the modality to be present in the semantic structure. (If they are present in semantics, then modals are ambiguous between potentially infinite numbers of structures; but in fact, these participants are pragmatically identified - see the end of this section.)

But the important gap in Antinucci and Parisi's argu-

ment is precisely the semantics of the general predicate bind: what does it mean (other than must), and why should it happen to apply equally well to real events and to reasoning processes? (There is some tacit assumption here that events and conclusions can be treated alike.) I trust that the rudimentary analysis of root modality in the preceding section has given some idea of the elements of my proposed general analysis of modality; in the next section, I shall attempt to explore and then motivate the linkup between real-world modality and epistemic modality.

B. Root modality applied to the epistemic world. If I view root modality as referring specifically to permission-giving or to social duty, for example, I would appear to have no hope of extending such an analysis to epistemic modality. The may of permission-granting and the may of possibility seem unconnected, since there is no permission-granter in the world of reasoning. But given that the epistemic world is understood in terms of the sociophysical world, we can see why permission should be the sociophysical modality chosen as analogous to possibility in the world of reasoning. May is an absent potential barrier in the sociophysical world, and the epistemic may is the force-dynamically parallel case in the world of reasoning. The meaning of epistemic may would thus be that there is no barrier to the speaker's process of reasoning from the available premises to the conclusion expressed in the sentence qualified by may. My claim, then, is that an epistemic modality is metaphorically viewed as the real-world modality which is its closest parallel in force-dynamic structure.

Let us set forth some similar analyses for the other modals' epistemic uses, attempting to apply our root modal analyses from section I to the speaker's reasoning process rather than to the subject's actions. We must now recast forces and barriers as premises in the mental world, since no other kinds of obstruction/force exist in that world. As we shall see, this will make some of the modals look rather more similar than in their real-world readings. The majority of the root modals refer to various forces, which is reasonable since we recognize many different varieties of force in the sociophysical world. In the epistemic domain, we have no contrast between internal forces (as in real-world need) and external forces (as in have to). Nor can we differentiate between kinds of authority or obligation; should and ought cannot refer to moral force (as opposed to threats, for example) in a world where no morality exists. In all of the following examples, I shall contrast the use of a modal in its real-world sense (a) with its corresponding usage in the epistemic domain (b).

May (19a) John may go.

"John is not barred by (my or some other) authority from going."

b) That may be true.

"I am not barred by my premises from the conclusion that that is true."

Must (20a) You must come home by ten. (Mom said so.)

"The direct force (of Mom's authority) compels you to come home by ten."

b) You must have been home last night.

"The available (direct) evidence compels me to the conclusion that you were home."

This epistemic analysis takes the premises in the speaker's mind as parallel to the force of authority in (20a). Note that the usual reluctance which is assumed to exist in the compelled person in (20a) has no counterpart in (20b). Such a contrast is a natural consequence of the differences between the sociophysical world and the epistemic world. In the real world, we don't usually use force unless we need to overcome reluctance on the part of the person we are forcing. But we do not view our mental processes as being affected by such reluctance, or by anything other than the available premises. Furthermore, in the real world force is usually resented by the victim because freedom is valued. But in the world of reasoning, we wish to have our conclusions forced or restricted because this gives us more certainties within our belief system, and knowledge is valued.

Can (21a) I can lift fifty pounds.

"Some potentiality enables me to lift 50 lbs."

b) You can't have lifted fifty pounds.

"Some set of premises dis-enables me from concluding that you lifted 50 lbs."

Positive can is almost unusable in an epistemic sense.<sup>9</sup> But its negative and interrogative forms are quite acceptable (cf. Can that be true?) and have the reading of questioned or negated epistemic enablement on the part of the speaker.

Ought to (22a) You ought to go.

"Certain forces (of moral obligation) influence you towards going."

b) That ought to be the right answer.

"The available set of premises (mental obligations or forces) influence me to conclude that that is the right answer."

- Have to (23a) He has to be home by ten.  
 "Some force of authority compels him to be home by ten."
- b) He has to be a New Yorker, with that accent.  
 "The available premises, including his accent, compel me to conclude he's from NY."
- Need to (24a) He needs to go to the grocery store.  
 "Some internal forces (e.g. wanting to eat tonight) compel him to go to the store."
- b) No, he needn't be a New Yorker - he could just have lived there a long time, or imitate accents well.  
 "The available premises do not force me to conclude that he's a New Yorker - they could also lead to other conclusions."

Once again, these analyses show the parallelism between the root and epistemic uses of modals. Sociophysical forces acting on the subject are taken as analogous to the logical "force" of premises acting on the speaker's reasoning processes. Note that need (like can) is epistemic only in its negative and interrogative forms.

Will (shall is not epistemic<sup>9</sup>)  
 (25a) John will come.

- "The present state of affairs will proceed to the future event of John's arrival."
- b) (hearing phone ring) That will be John.  
 "My present theory that that is John will proceed to future verification/confirmation."

Distal forms used epistemically (cf. discussion of root distals, sect. I)

These distal forms express past or conditional epistemic modalities.

- Might (26) He might go. (conditional)  
 "If some conditions were fulfilled, then my premises would not bar me from concluding that he will go."
- (27) I thought he might go. (past or conditional)  
 (The past (deontic) reading is simply the past of root may; the conditional (epistemic) reading is as above(26))

Note that conditionals with no expressed if-clause often have conditions so general as to become simply dubitatives; but this is a general crosslinguistic fact about conditional forms.

- Would (28) The folks you saw with John would be his  
 parents. (conditional)  
 "If some conditions (like having full data)  
 were fulfilled, my theory that those folks  
 were his parents would proceed to future  
 verification."
- (29) I knew it would be John. (past)  
 (past tense of the epistemic will in (25))
- Could (30) That could be the right choice for the living-  
 room curtains. (conditional)  
 "If some unspecified conditions were fulfilled,  
 then the available data would enable me to  
 conclude that that's the right choice for the  
 curtains."
- (31) I was dumbfounded: it simply couldn't be true.  
 (past tense of epistemic can as in (21))
- Should (32) John should be easy to talk to.

As previously mentioned, should is an odd distal form. Perhaps because of its dissociation from its rare present form shall, it has become only minimally conditional (there is no contrast with a non-conditional form any more). The relevant condition appears to be something very general like "if all goes right" or "if all goes as expected." Thus the epistemic should in (32) is a barely conditional expression of epistemic obligation, verging on synonymy with ought. Since shall has no epistemic reading, it may only be by association with ought that should has developed such an interpretation.

Non-auxiliary (conjugated) modals like have to, need to, as previously mentioned, have past but no independent conditional forms. Their (regular) past tense epistemic uses do not require discussion here.

The preceding description of epistemic modality has been nothing but a transfer of my proposed root-modal semantic analyses to the epistemic domain. I do not propose that epistemic modals have complex generative-semantic predicate structures to differentiate them from their root counterparts. Rather I propose that the root modal meanings can apply in two worlds, the "real" (sociophysical) world and the epistemic world. In the real world, the must in a sentence such as "John must go to all the department parties" is taken as indicating a real-world force imposed by the speaker (and/or by some other agent) which compels the subject of the sentence (or someone else) to do the action (or bring about its doing) expressed in the sentence. In the epistemic world the same sentence could be read as meaning "I must conclude that it is John's habit to go to the department parties, (because I see his

name on the signup sheet every time, and he's always out on those nights)." Here must is taken as indicating an epistemic force applied by some body of premises (the only thing that can apply epistemic force), which compels the speaker (or folks in general) to reach the conclusion embodied in the sentence.

Pragmatic factors will determine which world the modal is taken as operating in: for example, I swayed the interpretation of "John must go to all the department parties" towards an epistemic reading by adding a clause expressing a reason for reaching a conclusion. If instead I had added a clause expressing a real-world cause (such as "because he agreed to be bar-tender"), then the weight would have been towards a root reading. Past-tense sentences are strongly weighted towards an epistemic reading because real-world causality or modality can no longer influence frozen past events - I cannot inform you (except jokingly) that you are hereby put under an obligation, or given permission, to have done something yesterday. Conversely, modals in sentences concerning future actions are weighted towards a root reading, although an epistemic reading is not excluded.

Any sentence can be viewed under two aspects: as a description of a real-world situation or event, and as a self-contained part of our belief system (e.g. a conclusion or a premise). As descriptions, sentences describe real-world events and the causal forces leading up to those events; as conclusions, they are themselves understood as being the result of the epistemic forces which cause the train of reasoning leading to a conclusion. Modality is a specification of the force-dynamic environment of a sentence in either of these two worlds.

### C. Pragmatic interpretation of modal semantics in two worlds.

If a modal verb simply expresses the application of some particular modality towards the event or action described in a sentence, pragmatic factors will determine what appropriate entity is understood as imposing the modality, and upon what entity it is imposed. Thus root modals have a reading in which the speaker is taken as imposing the modality by stating it, and another reading in which some other entity (which may be elsewhere specified in the discourse) is the source of the modality. This contrast has interesting parallels with Searle's (1979) assertion/declaration distinction; modals are an area of language where speakers can either simply describe or actually mold by describing. However (as pointed out in Lakoff 1972a), there is a tendency for the describer or reporter of modality to be taken as sympathetic to the imposer, especially with the monomorphemic modals such as must (as opposed to have to).



ity (doing by describing) is natural, while it is impossible for the epistemic modalities. Epistemic modal sentences thus lack the multiple ambiguities inherent in the pragmatic interpretation of real-world modality: there is no possible doubt as to the nature of the mental modality's imposer and imposee.

This section has presented an analysis of epistemic modality not as a semantically distinct kind of modality, but as an essentially metaphorical application of our sociophysical modal concepts to the epistemic world. We have seen that such a unified viewpoint is possible if we analyze modality in terms of general forces and barriers evidently these are the basic sociophysical concepts in terms of which we understand our mental processes. In fact, I have argued that with the proper appeal to our pragmatic interpretation processes, there is no need to differentiate the semantic structure of root and epistemic modals at all. The next section will further motivate the application of the same linguistic modalities to the real and epistemic worlds, by setting modality in the larger context of a unified model of linguistic causality.

### III. Causality.

I have argued that our reason for applying the same modal verbs to the real world and the epistemic world is that we view the epistemic world as having a force-dynamic structure parallel to that of the sociophysical world (allowing for differences in the actual nature of the forces and barriers involved). If this is so, one might expect other parts of the English lexicon to manifest a similar tendency towards ambiguity between real-world force and epistemic force. And indeed several classes of lexical items can be applied to causal forces equally well in both worlds. Although all of these classes have been recognized as ambiguous, so far as I know they have not been analyzed as parallel to the modal case before (except for Tregidgo's brief mention of the insist/suggest verb-class). Together they constitute a very solid argument in favor of the kind of unified analysis of force and causality which would have to underlie my proposed unified modality.

A. Verbs. The root/ epistemic "ambiguity" of modal verbs is paralleled by the two possible uses of a number of other English verbs, mostly speech-act verbs. Some of the following examples are in fact ambiguous, but there is a possible deontic/epistemic contrast between the (a) and (b) sentences:

(33a) I insist {that you go to London.  
                  {on your going to London.

b) I insist that you did go to London. (though you may deny it.)



- (34a) I suggest that you leave the room now.  
 b) I suggest that you left the room to avoid being seen.
- (35a) I expect him to be there. (ambiguous)  
 b) I expect that he's there.

Tregidgo, who does not actually analyze these verbs but cites them as an example of some broader deontic/epistemic contrast, mentions that even the verb agree is ambiguous between agreement to (do something) and agreement that (something is true). Given the understanding that any sentence can be treated as an expression of some state of affairs in the real world, or as a conclusion in our world of reasoning, it is reasonable that a verb such as insist could express insistence on either the real-world doing of the action expressed in its complement, or the epistemic concluding of the proposition constituted by the complement.

These verbs are, then, not merely an argument for forces (such as insistence) being generalized over both the real and epistemic worlds, but also for our taking a sentence as an entity which exists in both those worlds simultaneously.

**B. Causal conjunction.** Ross(1970) has observed the ambiguity of English causal conjunction. Reasons for concluding something are generally treated like causes in the real world. Since we frequently reason from real-world effect to real-world cause, this can produce apparent inversions of causality such as (36)-(37):

- (36a) He heard me screaming, so he came.  
 b) He came because he heard me screaming.

(In both of these examples, the real-world hearing caused the real-world arrival.)

- (37a) (You say he's deaf, but-) He came, so he heard me screaming.  
 b) ( " " " " " ) He heard me screaming, because he came.

(In both of these examples, the knowledge of the arrival (a premise) causes the conclusion that he heard. The causality is in the epistemic world.)

Linguistically, reasons are treated as a subclass of causes. Although logical necessity itself is not a kind of causality, we view logical premises as causing us to draw conclusions. Causal conjunction refers to these (metaphorically viewed) epistemic forces, just as it refers to the more basic kinds of sociophysical causal forces. Only pragmatic factors will allow the hearer to decide

whether the causal connection expressed is between the two real-world events expressed in the clauses, or between the premise (expressed in one clause) and the conclusion which it causes in the speaker's mind (expressed in the other clause). In fact, (37a-b) are ambiguous between these two readings if we remove the suggested discourse-context. Just as certain propositional contents and pragmatic contexts tend to force either a root or an epistemic reading of modal verbs, the same is true of causal expressions. Thus (38) has almost inevitably an epistemic reading, while (39) is strongly weighted towards the root reading.

- (38) He loves me, because he wouldn't have proofread  
my thesis if he didn't.
- (39) He loves me because I remind him of his first love. <sup>10</sup>

This ambiguity is a general fact about causal conjunction in English. Some further examples are below (a=root, b=epistemic).

- (40a) The rules cannot be broken, therefore I will  
have to sentence you to two hours of trash  
collection.
- b) The rules cannot be broken, therefore the dean  
knew some way around them that allowed him  
to hire John.
- (41a) Since John isn't here, we'll just leave a note.
- b) Since John isn't here, he has (evidently) gone  
home for the day.
- (42a) Although he didn't hear the screams, he came  
(by chance) and saved her.
- b) Although he came and saved her, he didn't hear  
the screams; he came by chance.
- (43a) Despite their threats, she kept right on doing  
her job.
- b) Despite {the fact that she never wavered,  
her courage,}
- (we now know that) she was being threatened  
the whole time.

Ross uses examples like (37 a-b) and the subsequent (b) sentences as arguments for the presence of an abstract performative verb in every English sentence; he and Davison(1973) assume that the causality is between a premise and an act of assertion. In fact, as with epistemic modality, it is conclusions rather than assertions which are in question in the above cases; the (b) sentences do not express the speaker's reason for (or against) asserting the main clause to the hearer, but rather his/her reason

for or against concluding that it is true.

However, there are cases of causal conjunction which appear genuinely to connect premises with the current speech act, rather than linking premises to conclusions. I refer to cases like the following:

(44) Since {we're on the subject, } when was George  
          {you're so smart } Washington born?

(I ask you because we're on the subject or  
          because you're so smart.)

(45) Although I sympathize with your problems, get  
          that paper in by tomorrow.

(I command you despite my sympathy.)

(46) The answer is on page 242, since you'll never  
          find it out for yourself.

(I tell you because you'll never find out otherwise.)

Cases like (44)-(46) would seem to indicate that causal conjunction operates on sentences in three worlds,<sup>11,12</sup> as opposed to the two in which modal verbs apply. When sentences are joined with a causal conjunction, this can be understood as:

- (a) a conjunction of content; real-world cause related to its effect.
- (b) a conjunction of two epistemic entities - the premise and conclusion related in the reasoning chain.
- (c) a conjunction at the speech-act level; the causes (of whatever nature) are related to the resultant speech act.

As with the modals, only pragmatic factors will decide on which level the hearer should interpret the sentences as being causally conjoined.

C. Anyway. Anyway, the "despitive" adverbial which is closely related to the causal conjunctions, also is used at more than one level.

(47) He came on time anyway. (despite all the obstacles)

(48a) He came on time, anyway. } {though he did a bad job.  
  b) Anyway, he came on time. } {though you say he's  
  irresponsible.

(47) indicates real-world conflict between the event described and some surrounding forces. (48a-b) indicate either epistemic conflict between some piece of knowledge and some surrounding internal judgments or beliefs; or conflict at the speech act level (I persist in asserting this although you have said something which conflicts with it). In its speech-act sense, anyway can be used

as a marker of discourse-structure: I could use the first clause of (48b) to mean something like, "Despite the long digression I've gone through about the reasons I asked him to come, I return to the question you asked me and I assert in response to it that he came." The feeling is that the anyway-clause expresses the on-the-track topic of discourse, which we are returning to despite divergence from it.

The use of anyway is thus three-leveled like the use of causal conjunctions. Other despititive adverbials such as nonetheless seem to share at least some of anyway's ambiguity, though I shall not give examples here.

Our ideas of real-world causality are thus extended into the worlds of reasoning and speech acts, with the result that English causal lexemes are ambiguous among three levels of application. It is not clear why modals should be used on only two of these levels, though intuitively it seems to me that the two uses of the modals cause more real ambiguity for speakers than do the three uses of the causal conjunctions; so perhaps it is fortunate that modals don't operate at the speech-act level. It should be noted that modality can be expressed towards a speech act (or towards a conclusion) by putting that act overtly into the real world, so that modality can be applied to the act as a real-world event:

(49) I must tell you that your father wants you home, (although I would rather not).

(50) (Since all the evidence points that way), I must conclude that English and Tokharian are related.

If causal forces and barriers are viewed as generalized from the content (real-world) domain to the domain of propositions as epistemic objects or as speech acts, then it is scarcely surprising that modality (composed, like causality, of intentional forces) manifests a similar extension from its real-world application to application in an epistemic world. Besides the lexical classes mentioned above, there is a large body of general linguistic evidence that propositions and premises are thought of as causal forces, which bring about conclusions. Thus for example we talk about strong arguments, which have force, and weak ones, which don't. We ask someone their authority for believing or concluding that something is the case. None of these phrases is random; we have a coherent metaphorical treatment of epistemic forces in terms of sociophysical forces. (cf. Lakoff and Johnson 1980).

Conclusions. This paper has set forth an analysis of linguistic modality and causality as being generalized or extended from the real-world domain to the domains of reasoning and (in the case of causality) speech acts. The advantage of such an approach is that it allows us to give single semantic analyses to the modal verbs, causal conjunctions, and other "ambiguous" lexical items such as insist or anyway. Such words are not ambiguous between root and epistemic senses, but between their basic application and an extended application to the epistemic domain. My proposed analysis is also coherent with the historical and developmental linguistic evidence which suggest that an extension from the sociophysical world to the epistemic world would be more reasonable than an extension in the opposite direction.

Talmy's approach to deontic modality and causality in terms of forces and barriers has given us a way to look at modality which can be extended to the epistemic world as well. Attempts to find single superordinate analyses which include both deontic and epistemic modal meanings have proven unsuccessful (cf. Ehrman 1966). The same would probably be true for causality: logicians would castigate as hopelessly confused any analysis which tried to subsume in a single category both real-world forces or causes and the kind of necessity imposed on a conclusion by a premise. But in both cases, the problems for semantic analysis are removed by taking into account our understanding of mental processes as analogous to sociophysical interactions. Without taking into account this background metaphor, trying to unify deontic and epistemic modal meaning is like trying to figure out what are the common features of optimism and pink sunglasses without basing our analysis on the knowledge that physical sight is a primary metaphor for world-view in the psychological domain. But given the priority of the real world and the structuring of the epistemic world in terms of that prior world, it then follows naturally that the same understanding of modality and causality will apply to both worlds.

The single semantic analysis of the modals which I have proposed is a very simple one. It would not extend so easily into the epistemic domain if it explicitly mentioned a complex set of possible identities for real-world imposers and targets (imposees) of modalities. Rather, it leaves these identities to pragmatic interpretation. I consider this to be a further advantage of my analysis, since the semantics of the modals appear to be indeterminate in this area. That is, the semantic structure of the modal verbs does not explicitly pick out either subject or object (or any specific syntactic or semantic role) as the person on whom the modality rests;

rather it is the pragmatic factors inherent in the speech-act setting, together with our understanding of utterances as multi-leveled objects, which easily account for the possible ambiguities of modals with respect to the origins and targets of forces.

An utterance is content, epistemic object, and speech act all at once. There are areas of meaning which are naturally circumscribed within one of the three worlds in which utterances exist. But our linguistic treatment of causal force, and of the closely allied concepts of different modal forces, can only be fully understood by examining their application to more than one of the three.

### Notes.

0. My advisors, Charles Fillmore, George Lakoff, and Paul Kay, first encouraged this project and have given me crucial feedback throughout. Leonard Talmy kindly discussed his own work on modals with me at length. Julian Boyd, Elizabeth Closs Traugott, and Robin Lakoff have been the sources of many invaluable comments and suggestions. Iskendir Savasir and Julie Gerhardt have maintained a dialogue about our joint interest in modals, which I at least have found very stimulating. And finally, Orin Gensler must be thanked for his longsuffering assistance as a critic (at the levels of content and style alike), friend, and even occasionally informant. As always, none of these acknowledgments should be taken as placing the blame for mistakes on anyone but the author.

1. I shall throughout the ensuing discussion refer to root modality, rather than using the term deontic. Not only is root a broader term (some might take deontic as indicative of purely social or moral obligation), but it also reflects my leaning towards an analysis of epistemic modal meaning as rooted in sociophysical (root) modality.

2. I personally have data showing that modal verbs have a root/epistemic ambiguity in both the IndoEuropean and Semitic language families at large, and also in Finnish and Tagalog. Tregidgo(1982) lists a much larger set of languages cited by Perkins (forthcoming). I have not obtained a copy of Perkins' paper, but the list is as follows: Basque, Classical Aztec, French, German, Italian, Kapampangan, Korean, Luiseno, Polish, Tamil, Thai, Tzeltal Welsh, and "many ancient IndoEuropean languages."

3. There is a large literature on the subject of more and less prototypical agentivity and causality, which

I cannot begin to discuss here. Shibatani(1976) is an appropriate general reference. The other comment I have on the subject of more and less basic causality is that one could easily take the let of sentence (3) as being metaphorical, and claim that we understand non-intentional forces and barriers (like water and stones) in terms of our perceptually more basic concept of intentional force. This is what I feel is going on.

4. Viewing the schema of may as including a barrier, while must involves a force, also seems coherent with their different negation-scopes. The negation of removing or holding back a barrier would be leaving it in place, hence may not becomes prohibition. Must not, on the other hand, is a very forceful prohibition; which is scarcely what one would expect if must is a barrier whose negation is an open path. Rather, the internal-negative reading of must not indicates an oppositely directed force, a force compelling that one not do whatever it is. Note that the external negation of a force would simply be the absence of the force, which is the reading we get for German muss nicht.
5. The commonest use of shall in English is perhaps in consent-requests for mutual action, like "Shall we dance?" In these questions, it is precisely our joint intent to undertake an action which is being queried; so my analysis seems to make some sense. Likewise, in singular equivalents like "Shall I marry her?" (note the contrast with "Will I marry her?"), my undertaking to do so is in question. The third-person equivalents of these questions ("Shall he marry her?") still question the speaker's undertaking, of course, rather than the subject's.
6. The term distal I have taken from Langacker(1978), which uses this term precisely to refer to a generic "distance" within either the temporal or the causal sequence.
7. Lyons at no point attempts to give a unified analysis based on this suggestion. The suggestion in fact appears at the end of his (separate) analyses of deontic and epistemic modality.
8. Boyd and Thorne, for example, analyze root must as "I state I (or some Pro) (Imp)," where Imp is an imperative predicate applied to the content of the sentence. Epistemic must, on the other hand, they take to be, "I state," applied to the content of the sentence. There is a feature (nec), "necessary," which is marked on the predicate Imp in root-modal must, but on the predicate state in the epistemic must.

Tregidgo contrasts deontic and epistemic must as follows: the deontic "a must b" translates as "X DEMAND Y-

Y CAUSE - ab," while the epistemic "a must b" will translate instead as "X DEMAND Y - Y STATE - ab".

9. I would love to be able to explain why some of the root modals transfer better into the epistemic domain than others. Shall seems so much tied to the speaker that it is perhaps reasonable for it to lack an epistemic sense (there is no entity "the speaker" inside the epistemic world). But even that is just a guess. And why can and need should be epistemically used only in negative or interrogative forms, while ought has a full epistemic usage - well, maybe the internality of can and need (while ought is social/external) makes them transfer less fully to epistemic use? But why do they transfer at all, then?

10. Note the different intonation-patterns in (38)-(39), as shown by the comma. We will see this contrast again in the examples of use of anyway (47-8), where the epistemic use of the adverbial element is again set off from the sentence by a pause.

11. Conjunction in general operates on more than one level. And, but, and or have at least a couple of uses if not more: compare the conjunction of speech acts in (a-b) with the content-conjunction in (c-d).

(a) What is that phone number? - but don't bother to look it up if it's too much trouble.

(b) The Yangtze River has good dim sum, or the Taiwan makes great red-fried eggplant, but King Tsin has excellent mu shu pork.

(c) replace the or and but in (b) with two ands.

(d) I would like that number, but I don't want you to take too much trouble to look it up.

12. This triple view of sentences as content, belief or proposition, and speech act has some interesting ramifications, one of which is the self-referentiality of language. The application of modality or causal conjunction in the epistemic or speech-act worlds is implicitly self-referential use of language, since the relevant intentional force is understood as applying to the very act in which it is expressed.

### Bibliography

Antinucci, F., and D. Parisi (1971) "On English Modal Verbs," in Papers from the Seventh Regional Meeting of the Chicago Linguistic Society.

Austin, J.L. (1961) "Ifs and Cans," in J.L. Austin, Philosophical Papers, ed. Urmson and Warnock, 3rd ed. 1979, Oxford University Press.



- Boyd, J. and J. Thorne (1969) "The Semantics of Modal Verbs," Journal of Linguistics 5.
- Davison, Alice (1973) Performative verbs, adverbs, and felicity conditions. U of Chicago Ph.D.dissertation.
- Ehrman, Madeline (1966) The Meanings of the Modals in Present-Day American English. (Mouton)
- Fleischman, Suzanne (1980) "Futures: where do they come from, where are they going to?" talk presented to the Linguistics Group at Univ. of Calif. at Berkeley.
- Horn, Laurence (1972) On the Semantic Properties of Logical Operators in English. UCLA Ph.D. dissertation.
- Huddleston, Rodney (1979) "Would have become: empty or modal will," Journal of Linguistics 15.
- Kratzer, Angelika (1977) "What 'Must' and 'Can' Must and Can Mean," Linguistics and Philosophy 1.
- Kuczaj, Stan A. and Mary J. Daly (1979) "The development of hypothetical reference in the speech of young children," Journal of Child Language 6.
- Lakoff, George and Mark Johnson (1980) Metaphors We Live By. (U. of Chicago Press)
- Lakoff, Robin T. (1972a) "The Pragmatics of Modality," in Papers from the Eighth Regional Meeting of the Chicago Linguistic Society.
- \_\_\_\_\_ (1972b) "Language in Context," Language 48.
- Langacker, Ronald (1978) "The Form and Meaning of the English Auxiliary," Language 54.
- Lyons, John (1977) Semantics. (Cambridge University Press)
- Palmer, F.R. (1979) Modality and the English Modals. (Longman)
- Ransom, E. (1977) "On the Representation of Modality," Linguistics and Philosophy 1.
- Ross, J.R. (1970) "On Declarative Sentences," in Readings in English Transformational Grammar, ed. Jacobs and Rosenbaum. (Ginn)
- Searle, John R. (1979) Expression and Meaning. (Cambridge UP)
- Shepherd, Susan (1981) Modals in Antigua Creole, Child Language Acquisition, and History. Stanford Ph.D. diss.
- Shibatani, M. (1976) ed. Syntax and Semantics vol. 6: The Grammar of Causative Constructions. (Academic Press)
- Talmy, Leonard (1976) "Semantic Causative Types," in Shibatani. (1982) ms. Force Dynamics in Language and Thought.
- Tregidgo, P.S. "Must and May: Demand and Permission," Lingua 56 (1982).