Toward a linguistically adequate *causal model*

Zhuosi Luo · Beijing Foreign Studies University

COCOA, 12/12/2025

1 Introduction

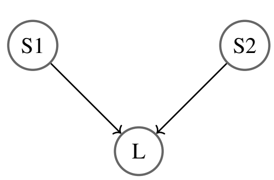
- In recent years, a growing body of research has applied *causal models*—a family of formal tools developed in statistics, computer science, and philosophy—to the study of natural language causation.
  - These models aim to offer "a formal representation of the structure that causal relations give to our conceptual model of the world" (Copley 2021).
- While these models have proven effective for core aspects of causal reasoning—such as counterfactual dependence and the distinction between causal sufficiency and necessity—it remains unclear to what extent they capture the full range of empirically attested phenomena that have been observed in natural language causation.
  - **Chomsky's three levels of adequacy (1965)**
    1. Observational Adequacy
    2. *Descriptive Adequacy*
       * Correct structural descriptions; generalization-capturing analyses
    3. Explanatory Adequacy
  - **The matching between *type-causal* representations and *token-causal* statements** (Baglini & Bar-Asher Siegal 2025)
    * "Type-causal statements outline broad causation principles and laws, while token-causal statements delve into specific causal event connections...The causal model can be understood as the repository of causal knowledge derived from empirical observations (Pearl, 2000), reflecting the transition from specific (token) causal instances to generalized (type) causal assertions." (Baglini & Bar-Asher Siegal 2025:673)
- In what follows, I take *descriptive adequacy* as a guiding benchmark and show that linguistic causation exhibits interpretive and structural distinctions that currently fall outside the representational capacity of the current causal model framework, which calls for further ontological enrichment of this approach.

- **This talk will present an emerging picture of how the causal models approach may be further developed within linguistics. Three directions will be the focus of this talk:**
  1. **Event-enriched structural equations** (Section 3)
     - Directness
       - The need for enriched temporal representation
       - The need for enriched spatial representation
       - The need for enriched participant configuration
     - Specifying causal manners / typed causal relations
  2. **Formalizing mental state attributions** (Section 4)
     - Mental states of event participants
     - Mental states of event observers
  3. **Integrating gradability / scalarity** (Section 5)
- Note: the proposals in this talk should not be interpreted as objections to the current theory of SCMs.
  - Instead, they constitute a principled way of aligning fine-grained linguistic distinctions with the standard causal model architecture.

2 Some background on *causal models*

- Originally developed within early twentieth-century statistics, *causal models* have since gained prominence in philosophy, psychology, and computer science, and have more recently been adopted in formal semantics to analyze causal meaning in natural language.
- The most widely used framework is the *Structural causal model* (SCM), introduced by Pearl (2000, 2009) and further elaborated in subsequent philosophical work (e.g. Woodward 2003; Halpern & Pearl 2005).
- In SCMs, causal relations are represented as directed acyclic graphs (DAGs), where nodes correspond to variables and edges encode asymmetric causal dependencies.[1]
  - A classic illustration involves a circuit with two switches and a light, where the light turns on if and only if the switches are in the same position (Lifschitz 1990; Schulz 2007; Nadathur & Lauer 2020):

(1) Graphic representation:



S1	S2	L
0	0	1
0	1	0
1	0	0
1	1	1

[1] For foundational discussions and formal techniques, see Pearl (2000, 2009), Halpern (2000), Halpern & Pearl (2005), Paul & Hall (2013), and Pearl & Mackenzie (2018).

- **Definition (Structural Causal Model)**
  - A Structural Causal Model (SCM) is a tuple

    $$M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F} \rangle,$$

    where:
    - $\mathcal{U}$ is a set of exogenous variables whose values are determined outside the model;
    - $\mathcal{V}$ is a set of endogenous variables whose values are determined within the model;
    - $\mathcal{F}$ is a set of structural equations, one for each $X \in \mathcal{V}$:

      $$X \coloneqq f_X(\mathrm{Pa}(X), U_X),$$

      where $\mathrm{Pa}(X) \subseteq \mathcal{V} \setminus \{X\}$ is the set of parents (direct causes) of X, and $U_X \in \mathcal{U}$ is the exogenous noise term associated with X.
    - Together, these equations determine the value of each endogenous variable as a function of its parents and the relevant exogenous factors.

- In linguistic applications, SCMs have primarily been used to model the truth-conditions of causal expressions, particularly in distinguishing causal sufficiency from necessity—that is, whether a (lexically) causal construction entails the actuality of its result.[2]
  - Key contributions to this line of inquiry include Nadathur (2019); Nadathur & Lauer (2020); Glass (2023); McHugh (2023); Copley (2024); Baglini & Bar-Asher Siegal (2025), among others.[3]
  - These studies converge on an important insight:
    - Linguistic causation cannot be uniformly captured by a primitive CAUSE operator (Dowty 1979), but instead involves a range of lexically and grammatically conditioned causal relations.[4]

3 Direction I: Event-enriched structural equations

- There has been some recent work integrating causal models into the analysis of complex event structure,
  - e.g., analyses of *grammatical aspect* (Nadathur & Bar-Asher Siegal 2025), *lexical aspect/telicity* (Nadathur & Everdell 2025), and *verb class* distinctions (Baglini & Bar-Asher Siegal 2025).
- However, the internal structure of an event chain still remains one of the most complex and theory-intensive domains.
- In what follows, I discuss several basic primitives of event structure and explore how they may be integrated into the causal models approach.

[2]For a recent comprehensive review of actuality entailments, see Nadathur (2025).

[3]See also Alonso-Ovalle & Hsieh (2021) on ability and involuntary action, and Kaufmann (2013) for early attempts to integrate causal and modal reasoning.

[4]For a detailed taxonomy of such causal combinatorics, see Appendices D and E in Luo (2024b).

3.1 Directness

- A central dimension of causality widely discussed in the linguistic literature is *directness*. Interpretive judgments of directness typically depend on the relational proximity between the *cause* and its *result*.[5]

  – The most frequently examined parameter is temporal contiguity, as defined by Nedjalkov & Silnitsky (1973) and Levin & Rapport Hovav (1999).

  – Another concerns participant structure, specifically whether an intermediary agent is involved in the causal chain (Masica 1976).

- Since events are generally understood as spatiotemporal entities—that is, concrete particulars situated in time and space (e.g., Davidson 1967)—a robust account of causal directness must integrate *temporal* and *spatial* relations, together with the *participant* configuration of the event.

3.1.1 The need for enriching temporal representation

- While the standard three-way distinction among *posteriority*, *simultaneity*, and *anteriority* (e.g., Grano 2015) provides a useful starting point, natural language frequently refers to events unfolding over extended temporal intervals rather than as atomic points.

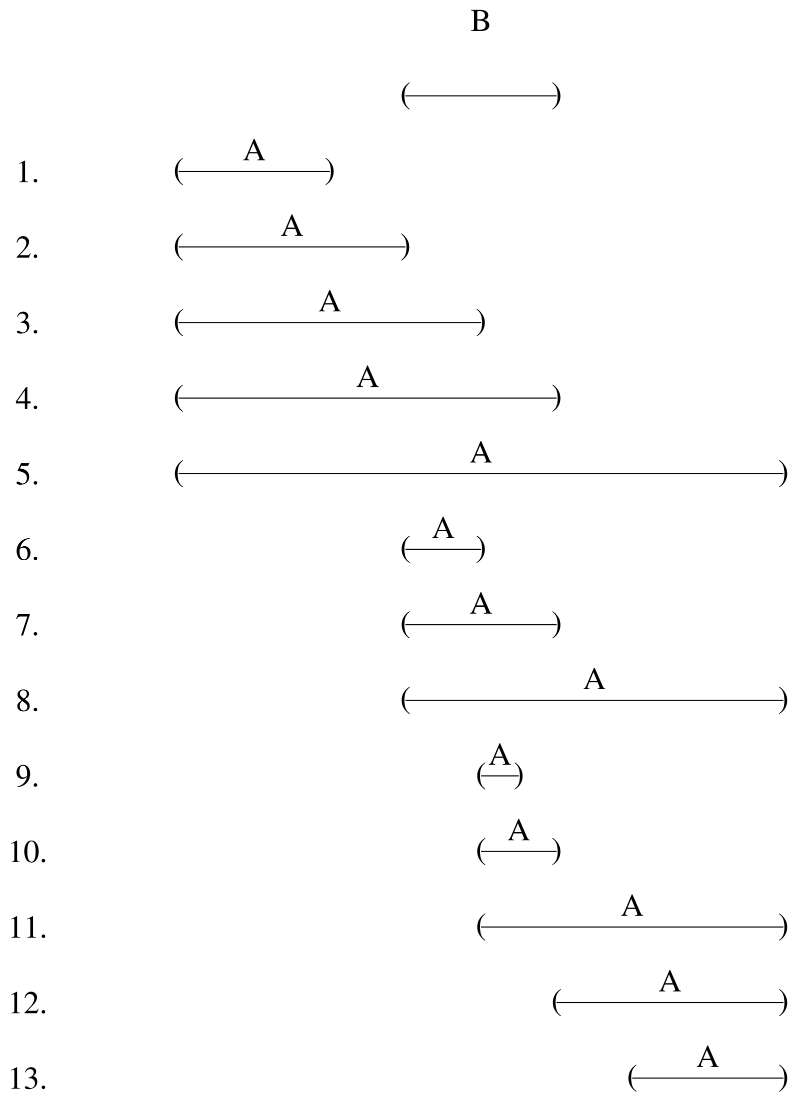
(2) Three logically possible temporal relations holding between two events, i.e., e_1 and e_2 (Grano 2015):

a. *Posteriority*: e_1 follows e_2
b. *Simultaneity*: e_1 overlaps with e_2
c. *Anteriority*: e_1 precedes e_2

- Kuhn & Portner (2002) models such intervals as continuous sequences of temporal points, yielding thirteen possible relations between two intervals. This fine-grained taxonomy is illustrated in (3).

[5] Structural notions of directness are often associated with the contrast between lexical and (morpho)syntactic causatives.

(3) Thirteen possible temporal relations between two events:



- Luo (2024b) shows that in Teochew (Southern Min), both lexical and periphrastic causatives instantiate at least seven of these thirteen temporal configurations—specifically, configurations 1–5 and 7–8 in (3)—where interval A corresponds to the causing event and B to the result.
  - This finding suggests that natural language is sensitive to fine-grained distinctions in causal temporality.
- To date, however, there is no established mature method for systematically encoding complex temporal relations within the architecture of standard causal models.
  - The most common (and arguably oversimplified) practice is to interpret the left-to-right ordering (or arrow direction) of variables as a reflection of causal order (Paul & Hall 2013).

- Note that such orderings in SCM diagrams reflect causal order, not necessarily temporal precedence. However, many linguistic applications implicitly identify the two, which leads to representational compression.

- Yet, as Copley (2021) observes, variables in causal models need not represent events; they may also denote propositions (e.g., Nadathur & Lauer 2020), situations (e.g., McHugh 2023), or other abstract entities.

- Building on Halpern & Pearl (2005), Copley (2021) and Cao et al. (2025) propose temporal indexing of variable values as an alternative means of encoding time, but this strategy remains largely underdeveloped in linguistic applications.

3.1.2 The need for enriching spatial representation

- A classic spatial contrast in event structure is the distinction between *proximal* and *distal* configurations.[6]

- Some languages even morphologically encode this distinction:

  - In Nez Perce, for instance, the cislocative suffix *-m* marks spatial proximity, while the translocative *-ki* indicates spatial distance (Deal 2009):

(4) Proximal spatial relation in Nez Perce:

a. *meet'u téemux 'e-wehye-**m**.*
but footprint 3.POSS-com-**CIS**
'But his footprints lead this way.'

b. *walíms sis 'inp'i-**m**.*
W mush take-**CIS**
'Walims, take mush from here!'

(5) Distal spatial relation in Nez Perce:

a. *kawo' heenek'e hi-q'uyim-cen-**ki**.*
the again 3.SUBJ-climb-IMPERF-**TRANS**
'He climbed farther up.'

b. *'iskit hi-ku-s'een-**ki**.*
trail S.SUBJ-go-IMPERF-**TRANS**
'The trail goes that way (away from the speaker).'

- In many causative constructions, the causing and resulting events need not occur in the same physical location.

  - For instance, the Japanese *sase* causative can support spatially disjoint interpretations—for example, a teacher assigns homework in the classroom while the student completes it at home:

[6] Unlike temporal relations, spatial relations between events are more difficult to define and compare due to vague or diffuse event boundaries. I thank Catherine Huang for raising this point.

(6) Japanese:

Sensei-wa gakusei-ni shukudai-o sase-ta.
teacher-TOP student-DAT homework-ACC do.CAUS-PST

'The teacher made the student do the homework.'

– In contrast, the Teochew *mue*-causative disallow spatial indirectness between the causing and caused events:

(7) Teochew:

#Nangy do bang lai mue Mimi do bang duakao tsao / u siokao / geʔsim.
Nangy at room inside make Mimi at room outside run / own wound / sad

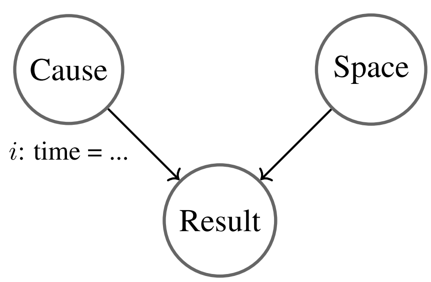
Intended: 'Inside the room, Nangy makes Mimi run/have wounds/be sad outside the room.'

- Despite the relevance of spatial distance for causal interpretation, current causal models provide no explicit mechanism for representing it.

- **Potential enrichment**:

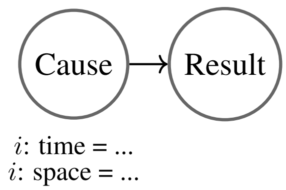
1. **Option 1** One possible strategy is to treat spatial proximity as (part of) a valued variable, which capture spatial relations between the causing and resulting events.
2. For the purpose of linguistic causation, I temporarily adopt the simplifying assumption that nodes correspond to event variables—not because SCM requires this, but because natural language data require event-level distinctions.
3. However, as shown in (8), this approach introduces an asymmetry:
   – whereas temporal information is often structurally encoded—either via causal order or through temporal indexing—spatial information, though equally central to event individuation (Davidson 1963), is relegated to the level of variable content.

(8)



4. **Option 2** A more uniform representation would index variables for spatial location in direct parallel to their temporal indexing, as shown in (9).

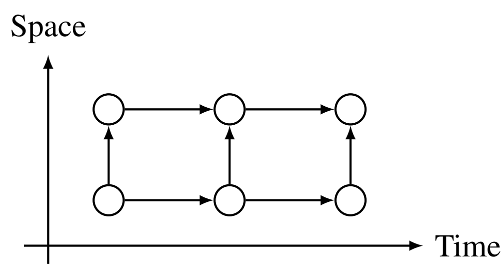
(9)



   – Note: Indexing should be understood as an annotation scheme rather than a primitive element of SCMs.

5. **Option 3** An even more ambitious possibility would be to encode spatial information structurally, in a manner analogous to temporal encoding via variable order, though doing so would require a multidimensional extension of the causal model architecture.

(10)



   – Such a multidimensional structure cannot be represented in classical DAG-based SCMs; it would require either hypergraphs, tensorized causal models, or intervention-sensitive dynamic models.

- Note: The goal of presenting these options is not to advocate a specific encoding, but to show that any linguistically adequate approach must make explicit commitments about how event substructure is represented in the SCM.

3.1.3 The need for enriching participant configuration

- A third dimension of causal directness concerns *the internal structure of participant roles*, most notably the presence or absence of an intermediary agent.

- As with *temporal* and *spatial* parameters, current causal models lack a systematic mechanism for representing intermediary participants.

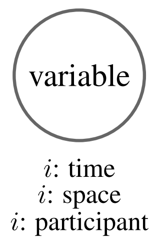
- **Potential enrichment:**

1. **Option 1** Because this role arises from the relational configuration between the causer and the causee, and depends on how the overall event structure is interpreted, it is most naturally encoded as an indexed feature on existing variables.

   – However, this raises an ontological concern: if *time* and *space* are treated as contextual parameters of the same type, *participant information* intuitively belongs to a different category.

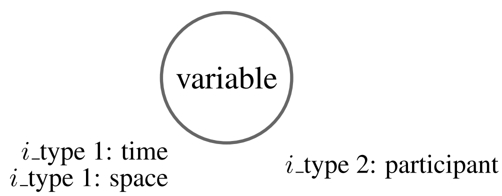
- As shown in (11), treating participant roles as a third type of index yields a relatively coarse-grained representation—particularly given that such roles are not standardly used to individuate events in formal semantics.
  - ∗ These participant indices are not part of the classical SCM ontology (my impression is that the SCM community usually think participants are values of variables); rather, they serve as linguistic metadata enabling finer-grained distinctions relevant to causative semantics.

(11)



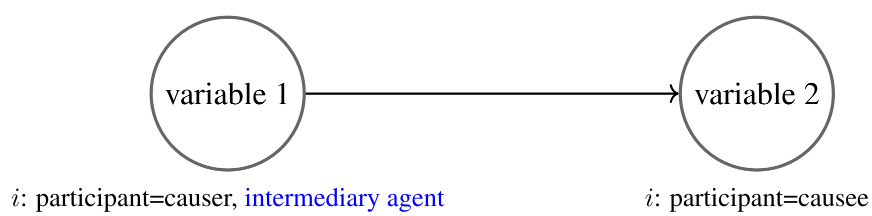
2. **Option 2** A more fine-grained approach would maintain a clear separation between temporal/spatial indices and participant-related indices, as illustrated in (12).[7]

(12)

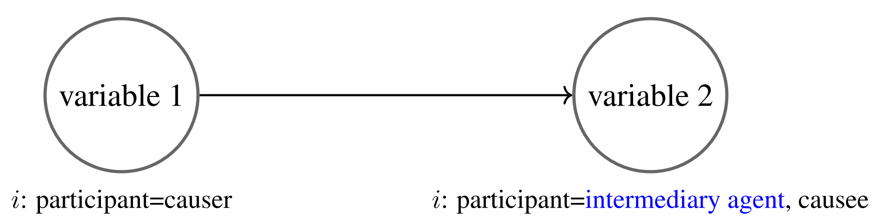


- This refinement, however, raises a further question:
  - Should information about the intermediary agent be indexed on the variable associated with the causer (13-a), on the one associated with the causee (13-b), or instead on a distinct intermediate variable (13-c)? I leave this issue for future investigation.

(13) a.

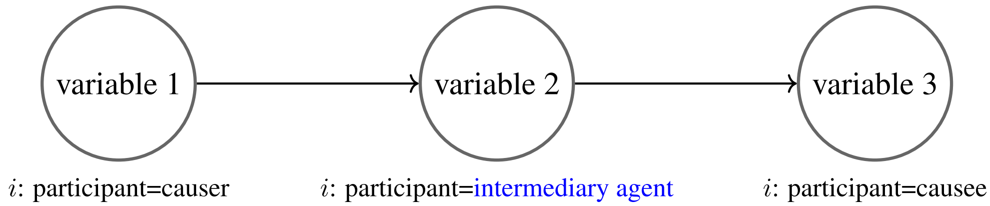


b.



[7] In some respects, this division parallels the distinction between interpretable and valued features in syntactic theory.

c.



- What makes the picture more complicated is that some causal relations—especially deontic ones such as permissive causation (Luo 2024c) and forbidding/prohibiting causation (Copley & Mari 2022)—take relations between event participants as part of their semantic premise.
  - For example, the Teochew permissive *bun*-causative. Evidence from sentence-final particles supports the presence of the social-status implication in this construction:
    - ∗ The clause-final emphatic yes/no-question marker *meh* can target only an event participant of higher social status in the discourse context, regardless of its syntactic position.

(14) <u>Teochew:</u>

a. *Tsiangbue lo gi haosegia meh*?
senior PROG meet junior $Q_{social-status}$
'Is the senior that is meeting with the junior?'
NOT 'Is the junior that the senior is meeting with?' (*meh* targeting the subject)

b. *Haosegia lo gi tsiangbue meh*?
junior PROG meet senior $Q_{social-status}$
'Is the senior that is meeting with the junior?'
NOT 'Is the junior that the senior is meeting with?' (*meh* targeting the object)

- ∗ In the permissive *bun*-causative (15-a), *meh* can target only the causer:

(15) <u>Teochew:</u>

a. *Nangy bun Mimi tsao meh*?
Nangy separate Mimi run $Q_{social-status}$
The only reading: 'Is it Nangy that causes Mimi to run?'

b. *Nangy hai Mimi tsao meh*?
Nangy hurt Mimi run $Q_{social-status}$
'Is it Nangy that causes Mimi to run (adversative)?'
Or 'Is it Mimi that Nangy causes to run (adversative)?' (compare *hai*-causative)

  - Incorporating such participant-based relations into the representational architecture of causal models is likewise a non-trivial challenge.

3.1.4 Strengthening the formal status of enriched structural equations

- While temporal, spatial, and participant indices enrich the representational space of SCMs, a linguistically adequate integration requires clarifying how such indices enter into the structural equations themselves.

- **Structured value ranges for event variables.** In classical SCMs, endogenous variables typically take atomic values (e.g., $0/1$). However, natural-language events require *structured* values:

$$v(e) \;=\; \langle \mathrm{time}(e),\; \mathrm{space}(e),\; \mathrm{participants}(e),\; \mathrm{core}(e) \rangle.$$

  - This expands the domain of the structural functions f_e and forces the model to compute event values componentwise.

- **Structural equations over structured objects.** Each event variable e must obey a structural equation of the form:

$$e \coloneqq f_e\big(\mathrm{Pa}(e), U_e\big),$$

  where f_e now maps *tuples of structured objects* to another structured object. Formally:

$$f_e : \prod_{p \in \mathrm{Pa}(e)} D_p \times D_{U_e} \;\longrightarrow\; D_e = T \times S \times P \times C,$$

  where T, S, P, C are the domains of temporal, spatial, participant-related, and event-core components, respectively.

- **Componentwise causal dependence.** The structural equation must specify *which subcomponents* of parent events determine which subcomponents of the result event. For example:

$$\mathrm{time}(e) = f_e^{\mathrm{time}}\big(\mathrm{time}(\mathrm{Pa}(e)), U_e^{\mathrm{time}}\big),$$

$$\mathrm{space}(e) = f_e^{\mathrm{space}}\big(\mathrm{space}(\mathrm{Pa}(e)), U_e^{\mathrm{space}}\big),$$

  illustrating that causal arrows no longer encode only token-level dependence but also determine how each event subcomponent is inherited or transformed.

- **Multi-dimensional intervention semantics.** Once event variables have structured values, interventions must be defined at the level of *subcomponents*. For example:

$$do\big(\mathrm{time}(e) = t^*\big)$$

  yields a different model from

$$do\big(\mathrm{participants}(e) = p^*\big),$$

  even when both are interventions on the same event variable e. This predicts distinct linguistic interpretations (e.g. delaying vs. relocating vs. reassigning an action).

- **Limitations of classical DAG architecture.** Enriching event variables in this way does not change the graph-theoretic structure of the SCM, but it expands the semantic content of each node.

- – The classical DAG imposes acyclicity only at the level of variable dependence, not at the level of temporal, spatial and participant subcomponents.
- – Thus, interval-based temporal relations or spatial/participant configurations cannot be recovered from arrow direction alone and must be encoded explicitly inside the structural functions.

3.2 Specifying causal manners/typed causal relations

3.2.1 Some background

- Causal meaning in natural language often encodes not only *whether* one event brings about another, but also *how* the causal relation is established. See Martin et al. (2025) for a recent discussion.

  - **In linguistics**, the *manner* vs. *result* contrast is one of the fundamental distinctions in the literature on transitive verbs (e.g., Levin & Rappaport Hovav 1991; Levin & Hovav 2013; Beavers & Koontz-Garboden 2012, 2020; Rapoport 2014; Rappaport Hovav 2016; Ausensi et al. 2024).
  - **In philosophy**, a related view is *Minimalism* (Anscombe 1971), which maintains that causation is not a unified concept but a cluster of relations expressible via specific verbs or constructions. This is illustrated in (16).

(16) *Minimalism*:

a. ‘*C* was a cause of *E*’ is true iff the relation between *C* and *E* can also be described using some member of set *S*, or can be described as a chain of relations each of which can be described using some member of *S*.

b. *S* is a set of causal verbs like *scrape*, *push*, *wet*, *carry*, *eat*, *burn*, *knock over*, *keep off*, *squash*, *make* and *hurt* and other linguistic formulas which represent ‘special causal concepts’ in Anscombe’s sense.

- I use the term *causal manner* to refer to distinctions such as coercion, permission, intervention, and enablement. These distinctions are widely attested cross-linguistically and are frequently encoded lexically.

  - In Mandarin, for instance, the permissive *rang* causative contrasts with the semantically underspecified *shi* causative (17).

(17) Mandarin:

a. *Laoshi rang Xiaoming jinlai bangongshi*.
Teacher RANG Xiaoming enter office
‘The teacher permits Xiaoming to enter the office.’

b. *Zhe pian lunwen shi Xiaoming dui zhe ge zhuti de neirong*
this CL paper RANG Xiaoming towards this CL topic POSS contents
gengjia qingchu le.
more clear PFV.
‘This paper causes Xiaoming to have a clearer understanding on this topic.’

- A similar contrast appears in Finnish, where the coercive *pakottaa* contrasts with the permissive/enabling *antaa*, both expressing causation (18).

(18) Finnish:

a. *Opettaja pakotti opiskelijan kirjoittamaan esseen.*
teacher PAKOTTI studentn-ACC write-INF essay-ACC
'The teacher forced the student to write the essay.'

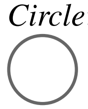
b. *Opettaja antoi opiskelijan kirjoittaa esseen.*
teacher allowed student-GEN write-INF essay-ACC
'The teacher let the student write the essay.'

- These differences are not merely paraphrastic; they involve differences in truth conditions, agency attribution, the presupposed autonomy of the causee, and pragmatic entailments, which have been extensively documented in the semantics literature on causation and control (e.g., Wolff & Song 2003; Wolff 2007; Sigurdsson & Wood 2021; Luo 2024b).
  - Such distinctions in causal manner are essential for understanding both the interpretation and distribution of causative constructions.
- However, current causal models fail to capture this level of granularity.
  - In the standard framework, causal dependencies are represented as untyped directed edges between variables.
  - While sufficient for modeling the *existence* of causal links, this architecture abstracts away from the *nature* of those links—whether coercive, permissive, instrumental, or otherwise mediated.
  - In effect, all forms of causation are formally collapsed into an undifferentiated graph-theoretic structure.
- To my knowledge, one of the few proposals to refine this architecture is Paul & Hall (2013):

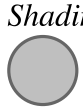
(19) Major annotations in Paul & Hall (2013):

a. *Circle:*

(i) *Circle*: a neuron

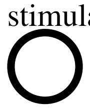


(ii) *Shading a circle*: a neuron fires



(iii) *Darkening a circle*: a neuron fires more intensely

(iv) *Circle with a thick border line*: a stubborn neuron, needing more than one stimulation in order to fire

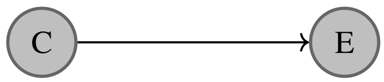


(v) *Circle with checkboard pattern*: a neuron acts as a kind of 'shunt'

(vi) *Striping a circle*: if and only if there are two incoming signals of the same kind but different intensities, this neuron emits an inhibitory signal along exactly one of the exit channels, equal in strength to the two incoming signals

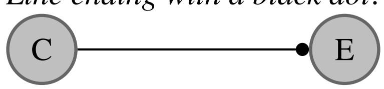
b. *Arrow and line:*

(i) *Arrow*: token-level stimulatory connections between neurons



(ii) *Fat and half-dark arrow*: the stimulatory signal is probabilistic, having a very small chance of dying out before reaching the next neuron

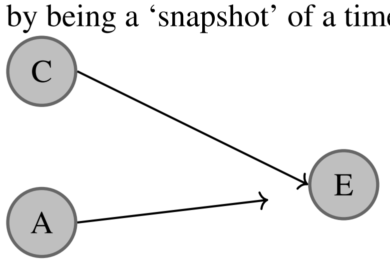
(iii) *Line ending with a black dot*: token-level inhibitory connections



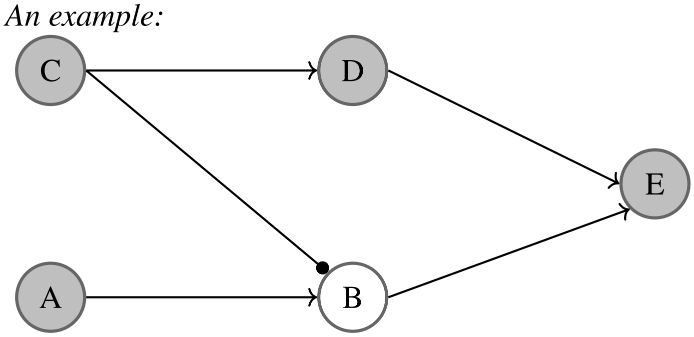
c. *Order:*

(i) The temporal order is represented by reading from left to right

(ii) The following graph shows that the stimulatory signal from C reaches E just before the stimulatory signal from A; the diagram represents this fact by being a 'snapshot' of a time



d. *An example:*



At time *0*, **C** and **A** both fire. **C** sends a stimulatory signal to **D**, which fires at time *1*. **A** sends a stimulatory signal to **B**, but the inhibitory signal from **C** (symbolized by the line with the blob on the end) blocks it, so **B** does not fire. **D** sends a stimulatory signal to **E**, which fires at time *2*.

  - Although these distinctions mark a step toward representing causal diversity, it remains far less fine-grained than what is needed to capture the typological and semantic richness of natural language causation.
  - This underspecification highlights a broader limitation of causal models:
    * While they can represent whether a causal relation exists, they are not equipped to represent *how* it is realized—precisely the aspect many natural languages encode morphosyntactically or lexically.

- Consequently, causal models in their current form fall short of achieving the descriptive adequacy required to analyze the full range of causative phenomena in natural language—a gap that future work must aim to close.

3.2.2 Encoding causal manner in structural equations

- Although standard SCMs represent causal dependencies via untyped edges, natural-language causatives require distinctions among coercive, permissive, enabling, and other manner-sensitive causal relations.

- I think a technically adequate treatment is better to operate at the level of *structural functions*. Specifically:

  - **Causal manner determines the class of structural functions.** In a standard SCM, an edge $C \to E$ only states that

    $$E \coloneqq f_E(C, U_E)$$

    for some function f_E. To capture causal manner, different natural-language causatives must correspond to *distinct subclasses of functions*:

    $$f_E^{\text{perm}}, \quad f_E^{\text{coerce}}, \quad f_E^{\text{enable}}, \ \dots$$

  - **Identical graph structures do *not* predict identical truth conditions.** Even when both causatives share the same graph:

    $$C \to E,$$

    the interpretation differs because:

    $$f_E^{\text{perm}} \neq f_E^{\text{coerce}} \neq f_E^{\text{enable}}.$$

    Hence the DAG underdetermines causal manner; only the structural equation can encode manner-sensitive semantic distinctions.

  - **Illustration.** To make the point concrete, different causal manners correspond to different structural equations for the same variable E:

    $$f_E^{\text{coerce}}(C, U_E) = 1 \quad \text{iff} \quad C \text{ overcomes resistance},$$

    $$f_E^{\text{perm}}(C, U_E) = 1 \quad \text{iff} \quad \text{barriers\_removed}(C).$$

    Although the graph contains only an arrow $C \to E$, the two equations implement distinct causal manners—coercive versus permissive—thus showing that manner is encoded in the functional form of f_E, not in the graph itself.

3.3 Interim summary

- To summarize:
  - Temporal intervals, spatial relations, intermediary participants, and typed causal manners jointly reveal that the current architecture of standard causal models undergenerates the event structure required by natural language.
- Event-enriched representations thus form a necessary foundation for any descriptively adequate causal model of linguistic meaning.
- Yet even an event-enriched model is not sufficient, because many causal interpretations depend not only on the structure of events, but also on the mental states of participants and observers—a dimension I turn to next.

4 Direction II: Formalize mental state attributions

- The close connection between mental states and causal interpretation has long been recognized (see, e.g., Beebee et al. 2009:and references therein).
- A substantial body of work in philosophy and linguistics shows that judgments about causation hinge not only on the external structure of events, but also on the internal mental states of both participants and observers.

4.1 Mental state of event participants

- **Causer** Fabienne Martin and collaborators (Martin 2015, 2016; Martin & Schäfer 2017) demonstrate that, across languages, the causal entailments of change-of-state predicates vary systematically with the agentivity of the subject.
  - More precisely, speakers find it easier to deny the resulting state when the causer is an agentive participant than when it is a non-agentive one. Consider the following contrast with the French verb *provoquer*:

(20) French:

a. *Pierre l'a provoquée, mais cela ne l'a pas touchée du tout.*
Pierre her has provoked but this NEG her has NEG touched at all
'Pierre provoked her, but this didn't touch her at all.' (agent subject)

b. *Cette remarque l'a provoquée, #mais cela ne l'a pas touchée du tout.*
this remark her has provoked but this NEG her has NEG touched at all
Intended: 'This remark provoked her, but this didn't touch her at all.' (causer/non-agent subject) (Slightly adapted from Martin & Schäfer (2017))

- **Causee** A large body of cross-linguistic research shows that the agency of the causee is often interpretively diminished due to the split distribution of agentive properties between the causer and the causee (e.g., Givón 1976; Lundin 2003; Pylkkänen 2008; Horvath & Siloni 2011; Key 2013; Legate 2014; Nash 2020; Sigurdsson & Wood 2021; Myler & Mali 2021; Akkuş 2022; Luo 2024b, 2025).

(21) Bemba:

a. *Naa-mu-fuund-ishya uku-laanda iciBemba ku-mufulo.*
1.SG-PAST-him-learn-CAUSE to-speak Bemba on-purpose
'In, on purpose, made him learn to speak Bemba.'
NOT 'I made him on purpose learn to speak Bemba.'

b. *Naa-butwiish-ya umuana ukwiitemenwa.*[8]
'I willingly made the boy run.' NOT 'I made the boy run willingly.' (Givón 1976)

(22) Finnish:

Ulla rakenn-utt-i Mati-lla uude-n toimistopöydä-n innokkaasti.
Ulla.NOM build-CAUSE-PAST Matti-AADESS new-ACC office.table-ACC enthusiastically

'Ulla, enthusiastically, had Matti build her a new office desk.' NOT 'Ulla had Matti, enthusiastically, build her a new office desk.' (Pylkkänen 2008)

(23) Hungarian:

a. *Az ügyvéd készség-gel / habozás n'elkül alá-ír-at-ta János-sal a szerzödést.*
the lawyer.NOM readiness-INSTR / hesitation without under-write-CAUS-PAST.DEF.DO János-INSTR the contract-ACC
The only reading: 'The Lawyer made [János sign the contract] readily/ without hesitation.'

b. ?*Az alá-ír-at-ta János-sal ügyvéd készség-gel / habozás n'elkül a szerzödést.*
the lawyer.NOM under-write-CAUS-PAST.DEF.DO János-INSTR readiness-INSTR / hesitation without the contract-ACC
The only reading: 'The Lawyer made [János sign the contract] readily/ without hesitation.' (Horvath & Siloni 2011)

(24) Turkish:

Tarkan Hasan-a Mehmet-i bil-erek döv-dür-dü.
Tarkan Hasan-DAT Mehmet-ACC know-PART beat-CAUS-PST

✓ 'Tarkan, on purpose, made Hasan beat Mehmet.'
× 'Tarkan made Hasan, on purpose, beat Mehmet.' (Key 2013)

[8] The glossing for this example is not provided in Givón (1976).

(25) <u>Acehnese:</u>

Bang geu-peu-koh ôk gobnyan bak lôn deungon singaja.
elder.brother 3Pol-CAUS-cut hair 3Pol at 1.SG withe purpose

'Brother made me cut his hair on purpose.'
✓ Brother did it on purpose.
× I did it on purpose. (Legate 2014)

(26) <u>Georgian:</u>

Keti-m gogo-s leks-i siamovnebit / ganzrax
Keti-ERG girl-DAT poen-NOM pleasure.with / intentionally
gada=a-targmn-in-a.
PREV=CAUS-translate-CAUS.AROR.3.SG

'Keti made the girl translate the poem with pleasure/intentionally.'
Keti did this with pleasure/intentionally
NOT the girl did this with pleasure/intentionally (Nash 2020)

(27) <u>isiXhosa:</u>

a. *uDallas$_i$ w-aphul-is-e uZoli$_j$ iglasi ngabom$_i$/$_{*j}$.*
1Dallas 1SBJ-break.TR-CAUS-PRF 1Zoli 9glass on.purpose
✓ 'Dallas [[made Zoli break the glass] on purpose].'
× 'Dallas [made [Zoli break the glass on purpose]].'

b. *uDallas$_i$ w-aphul-is-e ngo-Zoli$_j$ iglasi ngabom$_i$/$_{*j}$.*
1Dallas 1SBJ-break.TR-CAUS-PRF INS-Zoli 9glass on.purpose
✓ 'Dallas [[made Zoli break the glass] on purpose].'
× 'Dallas [made [Zoli break the glass on purpose]].' (Myler & Mali 2021)

(28) <u>Icelandic:</u>

a. *Ég lét byggja hús.*
I.NOM let.PST build.INF house.ACC
'I made (someone) build a house.'

b. *það á ekki að láta stjórna landinu af fjármálastofnunum.*
EXPL ought not to let.INF rule.INF country.the.DAT by financial.institutions
'...we ought not let the land be ruled by financial institutions...' (agentive *by*-phrase)

c. *Jón lét mála húsið með mjög litlum penslum.*
Jón let.PST paint.INF house.the.ACC with very small paintbrushes
'Jón had people paint the house with very small paintbrushes.' (instrumental phrase)

d. *þær létu byggja húsið (*af kappi).*
they.NOM let.PST build.INF house.the.ACC (*enthusiastically)
'They made (someone) build the house (*enthusiastically).' (agent-oriented adverb)

lét ϕ_j skoða thetta til thess að PRO$_{i/*j}$ fá meiri reynslu.
she.NOM let.PST insepct.INF this for it to get.INF more

experience.ACC
'She$_i$ had people$_j$ inspect this in order to PRO$_{i/*j}$ get more experience.' (rationale clause) (Sigurdsson & Wood 2021)

- The central point is that mental-state-sensitive interpretations of the participants in the causal event chain are not peripheral "noise", but are grammatically encoded in many languages.

- Yet only a small number of studies, to my knowledge, have attempted to encode the mental states of event participants within a causal model:

  1. One example is the "constraint on volitional action" in Nadathur & Lauer (2020):

     (29) *Constraint on volitional action*
     In the evaluation of a *make*-causative involving background situation s, causing event C, and caused event E, no proposition W_E representing the agent's intention to perform E can be such that $W_E = 0$ is sufficient for $E = 0$ relative to $s + (C = 1)$ and also determined by $s \setminus (C = 1)$.

  2. Another example is Copley (2024), where "subject-related influence" variables such as DESIRE and DISLIKE are modeled as nodes taking binary values (0 or 1):

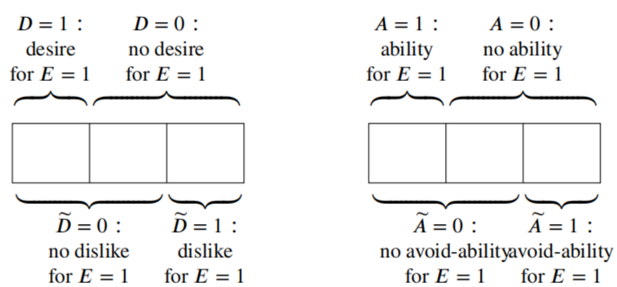


Figure 1: Desire, dislike, ability, and avoid-ability

Anti-efficacy readings	Subject-related influencing node: D desire	$\widetilde{D}$ dislike	A ability	$\widetilde{A}$ avoid-ability
actuality readings ($E = 1$)	$D \xrightarrow[\times]{f^+} E$ 0 1! unexpectedly	$\widetilde{D} \xrightarrow[\times]{f^-} E$ 1 1! unwillingly	$A \xrightarrow[\times]{f^+} E$ 0 1! *manage-to*	$\widetilde{A} \xrightarrow[\times]{f^-} E$ 1 1! accidentally
non-actuality readings ($E = 0$)	$D \xrightarrow[\times]{f^+} E$ 1 0! ?frustrative	$\widetilde{D} \xrightarrow[\times]{f^-} E$ 0 0! ?declined offer	$A \xrightarrow[\times]{f^+} E$ 1 0! ability	$\widetilde{A} \xrightarrow[\times]{f^-} E$ 0 0! ?near miss

Figure 2: The space of possible anti-efficacy readings on this proposal

From Copley (2024)

3. The third one is the "intention measure" proposed in Cao et al. (2025), a simplified implementation of the "degree of intention" framework originally developed by Halpern & Kleiman-Weiner (2018):

$$\mathsf{INT}(\mathcal{M}, \vec{a}, \vec{g}, \mathbf{u}') = \frac{\Pr((\mathcal{M}, \vec{u}) \models (A = \vec{a} \wedge G = \vec{g}))\mathbf{u}'(w_{\mathcal{M}}, A \leftarrow \vec{a}, \vec{u})}{\sum_{(\mathcal{M},\vec{u}) \in \theta : (\mathcal{M},\vec{u}) \models (A = \vec{a}' \wedge G = \vec{g})} \Pr(\mathcal{M}, \vec{u})\mathbf{u}'(w_{\mathcal{M}}, A \leftarrow \vec{a}', \vec{u})}$$

From Cao et al. (2025): "In prose, INT is the probability that an action performed in a state will result in the desired outcome, normalized by the probability of all alternative actions that would have resulted in the same outcome."

4.2 Mental state of event observers

- Beyond the mental states of event participants, speaker-oriented attitudes can also play a grammatical role in the interpretation of causative constructions.
- **Philosophical work** has long emphasized that causal judgments are deeply intertwined with moral, emotional, and evaluative considerations (e.g., Hume 1739; Kant 1785; Dewey 1939; Nagel 1979; Thomson 1985; Lagnado & Gerstenberg 2017).
- **Linguistic evidence** similarly demonstrates that speakers' evaluative stances may be encoded overtly.
  - A classic case is the Japanese lexical causative, which gives rise to an adversative interpretation—the nominative argument is understood not as the causer but as an affected party who bears a negative consequence related to the embedded event (Oehrle & Hiroko 1981; Miyagawa 1989; Harley 1996; Pylkkänen 2008):

    (30) Japanese:

    Taroo-ga musuko-o sin-ase-ta.
    Taro-NOM son-ACC die-CAUSE-PAST

    'Taro's son died on him.'

  - In Teochew, two distinct periphrastic causatives differ systematically in the speaker attitudes they license—one is compatible with positive evaluative stances and the other with negative ones, as revealed by their differing co-occurrence patterns with evaluative sentence-final particles (Luo 2024a):

    (31) Teochew:

    a. *Nangy hai Mimi tsao ku / *o.*
    Nangy hurt Mimi run PFV.NEG / *PGV.POS/NEU
    'Nangy caused Mimi to run (adversative).'

    b. *Nangy bun Mimi tsao o / *ku.*
    Nangy seperate Mimi run PGV.POS/NEU / *PFV.NEG
    'Nangy caused Mimi to run (benefactive).'

- Although one could argue that any causal graph implicitly reflects a subject's perspective, the notion of "perspective" is far too coarse to capture the fine-grained distinctions among different speaker attitudes.
  - **Option 1** One possible strategy is to introduce an attitude-holder or judge parameter into the values of variables (cf. Copley 2021).
  - **Option 2** Another is to manipulate the arrows themselves so that particular evaluative constraints are encoded in the graph (cf. Paul & Hall 2013).
  - **Option 3** A more radical option would be to extend the representational architecture beyond the standard nodes-and-arrows format.
    - Existing work in this direction is limited. Possible inspiration may come from *efficacy models* (Copley 2024), *intentional measure* (Cao et al. 2025), or Nadathur & Lauer (2020)'s *constraint on volitional action* mentioned above.
    - Yet these frameworks primarily formalize the internal states of event participants—not the stance of an external attitude-holder—leaving the observer-oriented dimension still unaccounted for.
    - Recall from Section 3.1.3 that causal relations also depend on the relational configuration among event participants—an aspect that itself relies on pragmatic factors shaped by the observer's evaluative judgment.
      - Capturing participant structure thus requires enriching the model so that it can represent participant roles and their interactions, for example by adding relational indices that encode those roles explicitly.
      - By contrast, encoding the observer's perspective calls for extending the model to represent judgments or attitudes toward those participants and events—potentially by introducing a dedicated evaluative layer that interacts with, but is distinct from, the event-structural layer.
- Taken together, these observations underscore the representational complexity that causal models must accommodate:
  - They require not only structural refinement, but also a principled mechanism for integrating pragmatic factors into causal interpretation.

4.3 Discussion: input vs. output variables?

- A foundational question for modeling mental-state-sensitive causation concerns how mental states should be represented in SCM: *as input variables, output variables, or both?*
- **Participant mental states as input variables**
  - In many linguistic contexts, event-participant-related properties such as *intention*, *willingness*, *control*, or *degree of effort* constrain the availability of certain causal interpretations.

- These mental states function as exogenous background conditions, shaping whether a predicate licenses actuality entailments or whether a causative construction supports coercive vs. permissive readings.
- In this sense, this type of mental state naturally bbehavesas *inputs* to the causal computation.

- **Observer mental states as output variables**
  - This type of mental state is an interpretive *result* of the causal configuration.
  - Speaker-oriented and observer-oriented evaluations—such as adversative readings in Japanese lexical causatives or judgments of responsibility, blame, or benefit—arise only after the event-level causal structure is fixed.
  - These are therefore better treated as *outputs* derived from the causal model.

4.4 Formal constraints on mental-state variables

- To integrate mental-state–sensitive interpretations of causation into a Structural Causal Model, additional constraints are required on how mental-state variables enter into the model. In particular:
  - **Mental-state variables as non-interventionable endogenous variables.** Unlike physical event variables, mental states (e.g., intention, willingness, desire) are not typically/easily manipulable by external interventions in a linguistically or cognitively plausible way. Formally:

    $$do(MS = m) \text{ is undefined for any mental-state variable } MS.$$

    This ensures that the semantics of volition, attitude, or agentive commitment is not distorted by operations that force a mental state to take an arbitrary value. Technically, this means mental-state nodes must be treated as *endogenous but non-interventionable*, paralleling preference structures in decision-theoretic causal models.
  - **Mental states derive functionally from event structure and contextual factors.** A mental-state variable must obey a structural equation of the form:

    $$MS \coloneqq f_{MS}(E_1, E_2, \text{Context}, U_{MS}),$$

    where E_1, E_2 are event variables, Context encodes socially or linguistically relevant information (e.g. hierarchy, discourse expectations), and U_{MS} captures background psychological factors. This reflects the empirical observation that observer-mental-state inferences—such as benefactive or adversative interpretations—arise only *after* the event-level causal configuration is fixed.
  - **Componentwise dependence on event substructure.** In an event-enriched model where events have temporal, spatial, and participant subcomponents, mental states must depend on specific aspects of event structure:

    $$MS \coloneqq f_{MS}\big(\text{participants}(E),\ \text{core}(E),\ \text{social-relations},\ U_{MS}\big).$$

This allows the model to capture the fact that mental-state readings—such as coercive vs. permissive interpretations or speaker evaluativity—are sensitive to relational and role-based configurations between event participants.

- **Mental-state variables may serve as inputs to causal computation.** Volition, intention, or willingness of the causee can constrain whether a causative predicate licenses actuality entailments. For example:

$$E \coloneqq f_E(C, MS_{\text{causee}}, U_E),$$

where MS_{causee} modulates whether the causer's action is sufficient to guarantee the realization of E. This aligns with cross-linguistic diagnostics showing that causee agentivity affects the availability of certain causative interpretations.

- **Observer-oriented mental states as output variables.** Speaker or observer evaluations—such as adversative readings in Japanese lexical causatives or benefactive readings in Teochew *bun*—are *interpretive outputs* determined by the causal configuration:

$$MS_{\text{observer}} \coloneqq f_{\text{eval}}(E_1, E_2, \text{ social-relations, norms, } U_{\text{eval}}).$$

This division ensures that grammatically encoded participant-related mental states (e.g. volition) function as inputs, while discourse-based evaluative judgments function as outputs of the causal computation.

5 Direction III: Integrating gradability/scalarity

5.1 Some background

- Traditional causal models treat causation as a binary relation:
  - *X causes Y* (yes/no)
- However, human judgments, natural language, and many linguistic constructions make causation appear *scalar* rather than binary:
  - *X slightly causes Y*
  - *X mostly causes Y*
  - *X strongly causes Y*
  - *X largely causes Y*
  - *X contributes more than Y*
  - *X is more of a cause than Y*
  - *X is more to blame than Y*
  - ....
- In philosophy, this topic is discussed under the label *graded causation*.

– "...if two companies jointly cause some environmental damage, we are often not only interested in whether they both contributed to the damage, but also in how large their contribution was" (cited from the project description of *Graded Causation* led by Vera Hoffmann-Kloss)[9]

– Related to our discussion of observer mental states, the scalarity of causation is also crucial for modeling *moral responsibility* (Hoffmann-Kolss & Rolffs 2025).

- **Possible enrichment:**

1. **Option 1** Paul & Hall (2013): a *fat and half-dark arrow* representing a probabilistic stimulatory signal with a small chance of dying out before reaching the next neuron.

2. **Option 2** Halpern & Hitchcock (2015): linking *graded causation* to *normality/typicality*, represented via ranking functions ≥ that order worlds according to the normality conditions relevant for causal judgment.

3. **Option 3** Cao et al. (2025): building on Pearl (2019), propose a *probability of sufficiency* measure indicating how likely Y would take value $\vec{y}$ given a counterfactual shift of X from $X \neq \vec{x}$ to $X = \vec{x}$.

$$\mathsf{SUF}(\mathcal{M}, \vec{x}, \vec{y}, \vec{u}) = \Pr(w_{\mathcal{M},\vec{u},Y=\vec{y}} \mid w_{\mathcal{M},\vec{u},X\leftarrow\vec{x}})$$

- **My preliminary thought: incorporating Degree Semantics** (e.g., Kennedy 1999, 2007)

– Gradable causation appears to require: (i.) a dimension, (ii.) an ordering relation/scale, and (iii.) a threshold—precisely the representational ingredients that degree semantics has long modeled.

– Modality was once considered a purely non-gradable quantificational category; recent work (Lassiter 2017; Portner & Rubinstein 2020) shows how to integrate gradability with modal semantics.

– A schematic representation might look like the following:

(32) CAUSE(e_1, e_2) ⇔ μ ($e_1 \rightarrow e_2$) ≥ θ, where:

a. μ is a measure of causal strength

b. θ is a contextually determined threshold

- **However, incorporating scalarity into causal models raises several foundational questions:**

– **What dimension is being measured?** Possible candidates include:

∗ strength of intervention

[9] https://www.philosophie.unibe.ch/research/projects/graded_causation/index_eng.html

    - probability of sufficiency / necessity
    - force or energy transfer
    - deviation from norms / typicality
    - degree of agentive control
    - contribution among multiple causes
  - **What is the scale type?**
    - ordinal vs. interval vs. ratio
    - bounded vs. unbounded
    - context-dependent vs. lexically fixed
  - **What is the relevant threshold?**
    - Is there an absolute threshold for *causing*?
    - Or is it context-sensitive like gradable adjectives such as *tall*?
    - Are we comparing degrees across events or within events?

5.2 How causal strength may be computed from enriched SCMs.

- A scalar notion of causation requires a principled method for extracting degrees from an enriched SCM.
- Given a model $M = \langle \mathcal{U}, \mathcal{V}, \mathcal{F}, P(\mathcal{U}) \rangle$[10] and two (event) variables $C, E \in \mathcal{V}$, we can define measures of causal strength as follows:
  - **Probability-based measures (sufficiency-oriented).** One family of measures links causal strength to differences in counterfactual probabilities. Let P_M be the probability distribution over endogenous variables induced by the SCM M and the exogenous distribution $P(\mathcal{U})$. A sufficiency-based measure is:

    $$\mu_{\text{prob}}(C, E) = P_M\big(E = 1 \mid do(C = 1)\big) - P_M\big(E = 1 \mid do(C = 0)\big).$$

    This compares the probability of E under an intervention enforcing $C = 1$ with a baseline where $C = 0$. In enriched SCMs, E may be non-Boolean; the condition $E = 1$ may be replaced by $E \in \mathcal{G}$ for some set of target outcomes.

[10] Here $P(\mathcal{U})$ is the joint probability distribution over all exogenous (background) variables, which represent factors not further analyzed within the model (e.g., contextual noise, default world states, or latent dispositions). This distribution propagates through the structural equations to induce a probability distribution over endogenous variables, allowing the model to evaluate expressions such as

$$P(E = 1 \mid do(C = 1))$$

which are required for probability-based measures of causal strength.

– **Force-dynamic measures (energy-/effort-based).** In an event-enriched model, each event value may be structured as:

$$v(e) = \langle \text{time}(e),\ \text{space}(e),\ \text{participants}(e),\ \text{core}(e) \rangle,$$

where $\text{core}(e)$ includes or maps to a force/energy parameter $\text{force}(e) \in F$.[11] A force-based measure is:

$$\mu_{\text{force}}(C, E) = g\big(\text{force}(C \Rightarrow E)\big),$$

where $g : F \to R$ is a monotone mapping. This measure distinguishes coercive causation (high force overcoming resistance) from permissive or enabling causation (low force, barrier removal).

– **Norm-based measures (normality-/typicality-based).** Let κ be a normality/ranking function on assignments to endogenous variables. Then a norm-based measure of causal strength is:

$$\mu_{\text{norm}}(C, E) = \kappa\big(\text{closest world where } C = 1,\ E = 1\big) - \kappa\big(\text{closest world where } C = 0,\ E = 1\big).$$

The more abnormal E would be without C, the stronger C counts as a cause. This ties graded causation to independently motivated structures of default reasoning.

– **A unified degree-based schema.** For any dimension $d \in \{\text{prob}, \text{force}, \text{norm}\}$:

$$\text{CAUSE}_d(C, E) \iff \mu_d(C, E) \geq \theta_d,$$

where:

* μ_d is the causal-strength measure computed along dimension d;
* θ_d is a contextually determined threshold, analogous to the POS operator in degree semantics.

This predicts:

$$C_1 \text{ is a stronger cause of } E \text{ than } C_2 \iff \mu_d(C_1, E) > \mu_d(C_2, E).$$

– **Scale type and cross-linguistic variation.** Different measures impose different scale requirements:

* probability-based measures: ratio scale on $[0, 1]$ (meaningful differences & ratios);
* force-based measures: interval or ratio scales depending on the physical/conceptual model;
* norm-based measures: typically ordinal (via κ), supporting comparatives but not always absolute thresholds.

• **Integration with event-enriched SCMs.** All three families of measures arise from the same enriched model:

[11] In force-dynamic semantics, events carry an abstract representation of effort, resistance, or enabling conditions, which can be mapped into a scalar domain.

  - probability-based measures evaluate how interventions on structured variables alter outcome distributions;
  - force-based measures extract force/energy from the event-core component;
  - norm-based measures derive from normality rankings informed by temporal, spatial, and participant-related substructure with typed causal relation.

- Thus, gradable causation is not external to SCMs but an operation on a richly articulated causal architecture.

6 Conclusion and discussion

6.1 Summary

- **This talk has outlined three major directions for developing a descriptively adequate integration of causal models into linguistic theory:**

  1. Event-enriched structural equations
     - Directness
       * The need to enrich temporal representation
       * The need to enrich spatial representation
       * The need to enrich participant configuration
     - Specifying causal manners/typed causal relations
  2. Formalize mental state attributions
     - Mental state of event participants
     - Mental state of event observer
  3. Integrating gradability/scalarity

- To make the empirical gaps more transparent, the table below summarizes how natural language systematically encodes distinctions that remain unrepresented or only partially represented in classic causal models.

Table 1: Where linguistic causation exceeds the representational capacity of current causal models

Dimension		Encoded in natural language	Encoded in standard SCMs	Resulting gaps
Event-enriched	**Temporal Structure**	multiple interval relations; lexicalized temporal sequencing	Often only causal order; temporal precedence not structurally represented	Undergeneration of linguistically attested temporal distinctions
	Spatial Structure	Proximal vs. distal causation; morphological marking	No structural representation of spatial location or distance	Complete absence of spatial constraints; loss of core distinctions
	Participant Configuration	Intermediary agent; social hierarchy	No representation of participant roles; participants treated as variable values	Undergeneration of role structure; difficulty modeling pragmatic participant relations
	Causal Manner / Type	Coercive / permissive / enabling / instrumental / ... causation; lexically encoded distinctions	Edges are untyped; causal dependencies treated uniformly	Flattening of causal types; inability to model manner-sensitive truth conditions
Mental States		Agentive control; intention; speaker evaluativity; benefactive/adversative readings	Only rare proposals include DESIRE / INTENTION-like variables	Critical undergeneration of mental-state–sensitive interpretations
Gradability / Scalarity		'More of a cause', slight / strong causation, contribution comparisons...	Limited to probabilistic edges or normality rankings in recent work	Lack of scale structure, measure functions, and thresholds

6.2 Discussion

- I do not intend to suggest that the causal models approach fails—such a claim would be inappropriate for a framework that has only recently entered linguistic research, and has already demonstrated considerable explanatory potential. Rather, I remain deeply optimistic about its future development.

- This talk is intended as an initial step toward integrating this promising framework with the fine-grained empirical distinctions uncovered by linguistic analysis.

  – Chomsky's notion of *descriptive adequacy* has served here as one guiding benchmark, and the ultimate aim is to contribute to a more systematic mapping between *type-causal* and *token-causal* representations (Baglini & Bar-Asher Siegal 2025).

- **A note on overgeneration and undergeneration**

  – Enriching the representational apparatus of causal models raises two complementary concerns: **undergeneration** and **overgeneration**. These concerns are familiar from linguistic theory, yet they surface in new ways when structural equations are adapted to natural language causation.

  – **Undergeneration** arises when the classical SCM framework fails to capture distinctions that natural language robustly encodes.

    * Examples include fine-grained temporal interval relations, spatially disjoint event configurations, intermediary-agency effects, and lexically encoded causal manners—all

of which are attested cross-linguistically but remain invisible to standard DAG-based representations.

  - **Overgeneration**, by contrast, becomes a risk when the model is enriched without sufficiently constraining the space of allowable representations.
    * For instance, unconstrained temporal or spatial indexing may predict event configurations that are never grammatically licensed.
    * Similarly, adding participant-related indices or observer-related layers encoding evaluativity risks inflating the ontology of the model in ways that exceed what linguistic evidence supports.
  - The goal, therefore, is neither to discard the classical SCM architecture nor to multiply representational resources indiscriminately.
  - Rather, the aim is to identify *exactly those dimensions*—temporal, spatial, participant-related, manner-sensitive, mental-state-dependent, or scalar—that are *empirically required* for achieving descriptive adequacy in linguistic causation.
  - This balance between expressive power and theoretical constraint will be crucial for any future development of a linguistically informed causal model.

- **Finally, I might be wrong about some/all of the discussion in this talk, and I very much look forward to discussing these ideas further with all the experts here. Thank you!**

References

Akkuş, Faruk. 2022. On causee in Sason Arabic. *Syntax* 25. 1–36. doi:10.1111/synt.12233.

Alonso-Ovalle, Luis & Henrison Hsieh. 2021. Causes and expectations: On the interpretation of the tagalog ability/involuntary action form. *Journal of Semantics* 38(3). 441–472. doi:10.1093/jos/ffab008.

Anscombe, Gertrude. 1971. *Causality and determination: An inaugural lecture*. London: Cambridge University Press.

Ausensi, Josep, Ryan Walter Smith & Jianrong Yu. 2024. On the expression of resultativity in English: The view from multiple resultatives. *Glossa* 9(1). doi:10.16995/glossa.11126.

Baglini, Rebekah & Elitzur A. Bar-Asher Siegal. 2025. Modeling linguistic causation. *Linguistics and Philosophy* 48(4). 647–691. doi:10.1007/s10988-025-09436-w.

Beavers, John & Andrew Koontz-Garboden. 2012. Manner and result in the roots of verbal meaning. *Linguistic Inquiry* 43(3). 331–369. doi:10.1162/LING_a_00093.

Beavers, John & Andrew Koontz-Garboden. 2020. *The roots of verbal meaning*. Oxford: Oxford University Press.

Beebee, Helen, Christopher Hitchcock & Peter Menzies (eds.). 2009. *The Oxford hanbook of causation*. Oxford: Oxford University Press.

Cao, Angela, Aaron White & Daniel Lassiter. 2025. Cause, make, and force as graded causatives. *Experiments in Linguistic Meaning* 3. 88–103. doi:10.3765/elm.3.5821.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.

Copley, Bridget. 2021. Causal model tutorial. Presented at Converging on causal ontology analyses.

Copley, Bridget. 2024. Reconciling causal and modal representations for two Salish out of control forms. In *Proceedings of the 39th West Coast Conference on Formal Linguistics*, 495–503.

Copley, Bridget & Alda Mari. 2022. "Forbid" is not "order not". In *Proceeding of 52th Annual Meeting of the North East Linguistic Society*, .

Davidson, Donald. 1963. Actions, Reasons, and Causes. *The Journal of Philosophy* 60. 685–700. doi:10.2307/2023177.

Davidson, Donald. 1967. The logic form of action sentences. In Nicholas Rescher (ed.), *The logic of decision and action*, 81–95. Pittsburgh: University of Pittsburgh Press.

Deal, Amy Rose. 2009. Events in space. In Tova Friedman & Satoshi Ito (eds.), *Proceedings of Semantics and Linguistic Theory 19*, 230–247.

Dewey, John. 1939. *Theory of valuation*, vol. 2. Chicago: The University Of Chicago Press.

Dowty, David. 1979. *Word meaning and montague grammar*. Dordrecht: Reidel.

Givón, Talmy. 1976. Some constraints on Bantu causativization. In Masayoshi Shibatani (ed.), *Grammar of causative constructions*, 325–351. New York: Academic Press.

Glass, Lelia. 2023. Using the Anna Karenina Principle to explain why cause favors negative-sentiment complements. *Semantics and Pragmatics* 16(6). 1–48. doi:10.3765/sp.16.6.

Grano, Thomas. 2015. *Control and restructuring*. Oxford: Oxford University Press.

Halpern, Joseph Y. 2000. Axiomatizing causal reasoning. *The Journal of Artificial Intelligence Research* 12. 317–337. doi:10.1613/jair.648.

Halpern, Joseph Y. & Christopher Hitchcock. 2015. Graded causation and defaults. *The British Journal for the Philosophy of Science* 66(2). doi:10.1093/bjps/axt050.

Halpern, Joseph Y. & Max Kleiman-Weiner. 2018. Towards Formal Definitions of Blameworthiness, Intention, and Moral Responsibility. Http://arxiv.org/abs/1810.05903.

Halpern, Joseph Y. & Judea Pearl. 2005. Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science* 56(4). 843–887. doi:10.1093/bjps/axi147.

Harley, Heidi. 1996. Sase bizarre: The Japanese causative and structural case. In *Proceedings of the 1995 Canadian Linguistics Society Meeting*, .

Hoffmann-Kolss, Vera & Matthias Rolffs. 2025. Graded causation and moral responsibility. *Erkenntnis* 90(6). 2219–2237. doi:10.1007/s10670-024-00797-5.

Horvath, Julia & Tal Siloni. 2011. Causatives across components. *Natural Language and Linguistic Theory* 29(3). 657–704. doi:10.1007/s11049-011-9135-3.

Hume, David. 1739. *A treatise of human nature*. London: John Noon.

Kant, Immanuel. 1785. *Groundwork of the metaphysics of morals*. Oxford.

Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive science* 37(6). 1136–1170. doi:10.1111/cogs.12063.

Kennedy, Christopher. 1999. *Projecting the adjective: The syntax and semantics of gradability and comparison*. New York: Garland Press.

Kennedy, Christopher. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30(1). 1–45. doi:10.1007/s10988-006-9008-0.

Key, Gregory. 2013. *The morphosyntax of the Turkish causative construction*. Tucson: The University of Arizona dissertation.

Kuhn, Steven & Paul Portner. 2002. Tense and time. In D.M. Gabbay & F. Guenthner (eds.), *Handbook of Philosophical Logic*, vol. 7, 277–346. Dordrecht: D. Reidel Publishing Company.

Lagnado, David A. & Tobias Gerstenberg. 2017. Causation in legal and moral reasoning. In Michael R. Waldmann (ed.), *The Oxford handbook of causal reasoning*, 565–601. New York: Oxford University Press.

Lassiter, Daniel. 2017. *Graded modality: Qualitative and quantitative*. Oxford: Oxford University Press.

Legate, Julie Anne. 2014. *Voice and v: Lessons from Acehnese*. Cambridge: MIT Press.

Levin, Beth & Malka Rappaport Hovav. 2013. Lexicalized meaning and manner/result complementarity. In Boban Arsenijević, Berit Gehrke & Rafael Marín (eds.), *Studies in the Composition and Decomposition of Event Predicates*, 49–70. Dordrecht: Springer Netherlands. doi: 10.1007/978-94-007-5983-1_3.

Levin, Beth & Malka Rappaport Hovav. 1991. Wiping the slate clean: A lexical semantic exploration. *Cognition* 41(1). 123–151. doi:10.1016/0010-0277(91)90034-2.

Levin, Beth & Malka Rapport Hovav. 1999. Two structures for compositionally derived events. In Tanya Matthews & Devon Strolovitch (eds.), *Proceedings of Semantics and Linguistic Theory 9*, 199–223.

Lifschitz, Vladimir. 1990. Frames in the space of situations. *Artificial Intelligence* 46(3). 365–376.

Lundin, Katarina. 2003. *Small clauses in Swedish: Towards a unified account*: Lund University dissertation.

Luo, Zhuosi. 2024a. Causality and modality: a case study on Teochew periphrastic causatives. In *Proceedings of the Linguistic Society of America*, vol. 9, 5650.

Luo, Zhuosi. 2024b. *Causality, modality and contextual argument interpretation: Lessons from Teochew*. Washington, DC: Georgetown University dissertation.

Luo, Zhuosi. 2024c. Interpreting causee in a 'permissive' causative: A case study on Teochew. In *Proceedings of the Linguistic Society of America*, vol. 9, 5654.

Luo, Zhuosi. 2025. Contextual causee interpretation: Lessons from Teochew ke-causative. In *Proceedings of the 41st West Coast Conference on Formal Linguistics*, 419–426.

Martin, Fabienne. 2015. Explaining the link between agentivity and non-culminating causation. In *Proceedings of SALT 25*, 246–266.

Martin, Fabienne. 2016. Atypical agents and non-culminating events. Presented at AG 3, DGfS 2016, Universität Konstanz.

Martin, Fabienne, David Rose & Shaun Nichols. 2025. Burning facts: Thick and thin causatives. Https://ling.auf.net/lingbuzz/009522.

Martin, Fabienne & Florian Schäfer. 2017. Sublexical modality in defeasible causative verbs. In Ana Arregui, María Luisa Rivero & Andrés Salanova (eds.), *Modality across syntactic categories*, 87–108. Oxford: Oxford University Press.

Masica, Colin. 1976. *Defining a linguistic area: South Asia*. Chicago: Chicago University Press.

McHugh, Dean. 2023. *Causation and Modality*. Amsterdam: University of Amsterdam PhD Thesis.

Miyagawa, Shigeru. 1989. *Structure and case making in Japanese*. San Diego: Academic Press.

Myler, Neil & Zoliswa O. Mali. 2021. Two places for causees in productive isiXhosa morphological causatives. *Syntax* 24(1). 1–43. doi:doi.org/10.1111/synt.12208.

Nadathur, Prerna. 2019. *Causality, aspect, and modality in actuality inferences*. Standford: Standford University dissertation.

Nadathur, Prerna. 2025. Actuality entailments. *Annual Review of Linguistics* 11. 275–297. doi:10.1146/annurev-linguistics-011724-121222.

Nadathur, Prerna & Elitzur A. Bar-Asher Siegal. 2025. Modeling progress: Causal models and the imperfective paradox. LingBuzz.

Nadathur, Prerna & Michael Everdell. 2025. Reanalyzing frustration: Event maximality and inertia in two O'dam frustratives. Presented at SALT 35.

Nadathur, Prerna & Sven Lauer. 2020. Causal necessity, causal sufficiency, and the implications of causative verbs. *Glossa* 5(1). doi:10.5334/gjgl.497.

Nagel, Thomas. 1979. *Mortal questions*. Cambridge: Cambridge University Press.

Nash, Lea. 2020. Causees are not agents. In Elitzur A. Bar-Asher Siegal & Nora Bohen (eds.), *Perspectives on causation*, 349–394. Dordrecht: Springer.

Nedjalkov, Vladimir & Georgij Silnitsky. 1973. The typology of morphological and lexical causatives. In F. Kiefer (ed.), *Trends in Soviet theoretical linguistics: Foundations of language*, 1–32. Dordrecht: Reidel.

Oehrle, Richard & Nishio Hiroko. 1981. Adversity. In A. K. Farmer & C. Kitagawa (eds.), *Proceedings of the Arizona Conference on Japanese Linguistics*, 163–187.

Paul, L. A. & Edward J. Hall. 2013. *Causation: A user's guide*. Oxford: Oxford University Press.

Pearl, Judea. 2000. *Causality: Models, reasoning and inference*. Cambridge: Cambridge University Press.

Pearl, Judea. 2009. *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press 2nd edn.

Pearl, Judea. 2019. Sufficient causes: On oxygen, matches, and fires. *Journal of Causal Inference* 7(2). doi:10.1515/jci-2019-0026.

Pearl, Judea & Dana Mackenzie. 2018. *The book of why: The new science of cause and effect*. New York: Basic Books.

Portner, Paul & Aynat Rubinstein. 2020. Desire, belief, and semantic composition: Variation in mood selection with desire predicates. *Natural Language Semantics* 28(4). 343–393. doi:10.1007/s11050-020-09167-7.

Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge: MIT Press.

Rapoport, Tova. 2014. Verb Meaning and Context: A Criticism of Manner-Result Complementarity. In Geneviève Girard-Gillet (ed.), *Autour du Verbe Anglais*, 121–137. Paris: Presses Sorbonne Nouvelle.

Rappaport Hovav, Malka. 2016. Grammatically relevant ontological categories underlie manner/result complementarity. In Noa Brandel (ed.), *Proceedings of IATL 2016*, 77–98. MIT Working Papers in Linguistics.

Schulz, K. 2007. *Minimal models in semantics and pragmatics: Free choice, exhaustivity, and conditionals*. Amsterdam: Universiteit van Amsterdam dissertation.

Sigurdsson, Einar Freyr & Jim Wood. 2021. On the implicit argument of Icelandic indirect causatives. *Linguistic Inquiry* 52(3). 579–626. doi:10.1162/ling_a_00384.

Thomson, Judith Jarvis. 1985. The trolley problem. *Yale Law Journal* (94). 1395–1415.

Wolff, Phillip. 2007. Representing causation. *Journal of Experimental Psychology: General* 136(1). 82–111. doi:10.1037/0096-3445.136.1.82.

Wolff, Phillip & Grace Song. 2003. Models of causation and the semantics of causal verbs. *Cognitive Psychology* 47(3). 276–332. doi:10.1016/S0010-0285(03)00036-7.

Woodward, James F. 2003. *Making things happen: A theory of causal explanation*. New York: Oxford University Press.