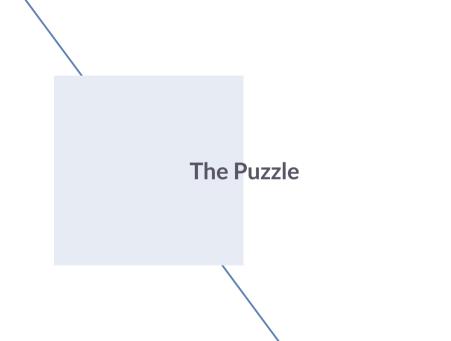
Modeling Uncertainty, Unawareness, and Underspecification among Structural Causal Models Todd Snider and Michael Franke

Sinn und Bedeutung 2024



Structural Causal Models (SCMs) have been used

- to provide truth conditions for conditionals (Pearl, 2000; Hiddleston, 2005; Briggs, 2012, a.o.)
- to explain causal selection (Woodward, 2003, a.o.)
- to give semantics for **causal verbs** like *make* (Nadathur and Lauer, 2020, a.o.)

In all of these contexts, regardless of the packaging, we are interested in the (causal) relations between the events/propositions mentioned

- (1) If Alex buys a turtle, Charlie will be upset.
- (2) Alex bought a turtle. That's why Charlie is upset.
- (3) Alex's buying a turtle would **make** Charlie upset.

All of (1-3) involve a dependency between Alex's buying a turtle (A) and Charlie's being upset (C), so we can turn to SCMs to model relations like that dependency

Even more reasons to love 'em

Prominent work in Cognitive Science suggests that causal intuitions arise from people performing mental simulations, modelled as sampling possibilities based on SCMs (Gerstenberg, 2022; Quillien and Lucas, 2023, o.a.).

Recent contributions have also used SCMs as ingredients of pragmatic reasoning about the meaning of both causal and non-causal language (Grusdt, Lassiter, and Franke, 2022; Beller and Gerstenberg, 2023).

Even more reasons to love 'em

Prominent work in Cognitive Science suggests that causal intuitions arise from people performing mental simulations, modelled as sampling possibilities based on SCMs (Gerstenberg, 2022; Quillien and Lucas, 2023, o.a.).

Recent contributions have also used SCMs as ingredients of pragmatic reasoning about the meaning of both causal and non-causal language (Grusdt, Lassiter, and Franke, 2022; Beller and Gerstenberg, 2023).

SCMs are useful tools for modeling a variety of phenomena

Even more reasons to love 'em

Prominent work in Cognitive Science suggests that causal intuitions arise from people performing mental simulations, modelled as sampling possibilities based on SCMs (Gerstenberg, 2022; Quillien and Lucas, 2023, o.a.).

Recent contributions have also used SCMs as ingredients of pragmatic reasoning about the meaning of both causal and non-causal language (Grusdt, Lassiter, and Franke, 2022; Beller and Gerstenberg, 2023).

SCMs are useful tools for modeling a variety of phenomena but they have limits (as we'll see in a moment)

Even more reasons to love 'em

Prominent work in Cognitive Science suggests that causal intuitions arise from people performing mental simulations, modelled as sampling possibilities based on SCMs (Gerstenberg, 2022; Quillien and Lucas, 2023, o.a.).

Recent contributions have also used SCMs as ingredients of pragmatic reasoning about the meaning of both causal and non-causal language (Grusdt, Lassiter, and Franke, 2022; Beller and Gerstenberg, 2023).

SCMs are useful tools for modeling a variety of phenomena but they have limits (as we'll see in a moment) so we aim to supplement them (our goal for today)

Uncertainty

Many early treatments using SCMs for the interpretation of conditionals presume that agents have a single mental model of the world: one SCM that they use to evaluate utterances containing conditionals

- Identify the causal statement conveyed by the utterance
- (Perform interventions on the model, if necessary)
- ▶ If the SCM (in)validates the statement, then the utterance is true(/false)

There is no representation of an agent's potentially being uncertain about which SCM is the right model of the facts and dependencies of the world

Representing Uncertainty

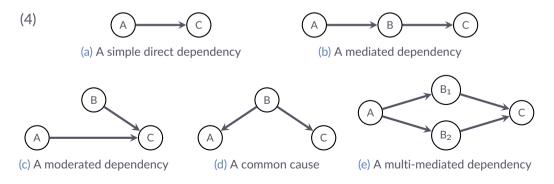
Bjorndahl and Snider 2016

4

Inspired by the observation that a single counterfactual like (1) is compatible with a number of possible explanations (each with a different SCM), Bjorndahl and Snider 2016 incorporated SCMs into a Stalnakerian framework

[A > C]

(1) If Alex buys a turtle, Charlie will be upset.



Representing Uncertainty

Bjorndahl and Snider 2016

Adopting Starr 2013, each SCM can be treated as a possible world, fixing the facts and dependencies at that world

In a Stalnakerian turn, then, Bjorndahl and Snider 2016 presents agents as maintaining uncertainty over a set of SCMs: the live candidate worlds in the Context Set

When an agent interprets an utterance containing a conditional, they:

- Identify the set of SCMs compatible with the utterance
 - On the basis of both node-values and dependencies
- If they accept it as true,) Intersect that set with the prior Context Set
- Take the new Context Set as the backdrop for a subsequent conversational move

Unawareness and Inattention

Even with uncertainty over models, there is more we might want to represent

(4a)



(4a) does not include a node *B* as part of the explanation of A > C.

- ▶ Is this because the agent has an **explicit belief** that *B* is unrelated?
- Or only an implicit belief that B is unrelated?
 In other words, are they not aware that B even exists as a node worth modeling?
 Put a third way, are they not attending to the possibility that B exists?

We know that awareness and attention are key factors of an agent's reasoning and decision-making (Ciardelli, Groenendijk, and Roelofsen, 2009; Franke and Jager, 2011; Westera, 2022), but we can't read this off of the SCM; this framework doesn't allow us to represent the state of an agent's awareness or attention

Unawareness and Inattention

Awareness and attention can also affect how agents respond in conversations

- (5) If Alex buys a turtle, Charlie will be upset.
 - a. Skeptical Sam: What do you mean, how so?
 - b. Naïve Nat: Oh, okay. Good to know.
 - c. Guessing Gal: Oh, because Charlie would be jealous?

An agent who has broad reasons to disbelieve or disprefer a direct A-to-C dependency (like Sam) might question the asserted [A > C] covariance, and ask for more detail

An agent who isn't even aware that there might (need to) be intervening/moderating factors (like Nat) wouldn't even think to ask

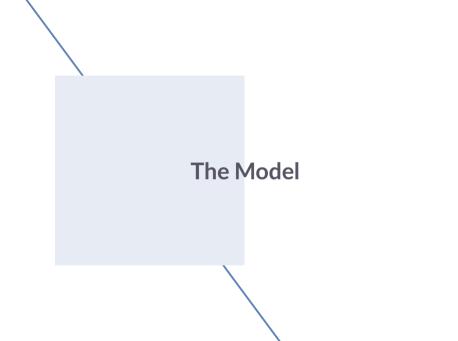
An agent who is aware of the presence of some additional factor but has uncertainty among them (like Gal) might ask a directed question to try to decide among candidates

Unawareness and Inattention Desideratum

At least broadly speaking, the behavior of [Sam vs. Gal vs. Nat] is driven by (or at least, bounded by) their states of awareness, what information they are attending to

- They maintain uncertainty over SCMs, for those issues they don't have enough information to resolve
 - Either for the truth values of nodes
 - Or for the (non-)existence of edges between nodes
- But their (un)awareness determines which SCMs they are 'even considering' or at least, which SCMs the agent can distinguish from one another
 - If an agent isn't aware that they have uncertainty, they won't make conversational moves to resolve that uncertainty

What we want is a way to model both the **information** and the **attention** state of an agent, as they consider the possibility space of SCMs which represent the actual world



The Model Setting up

We start with the setup as in Bjorndahl and Snider 2016, using Starr's (2013) Structured Possible Worlds to incorporate SCMs into a Stalnakerian picture

- a possible world $w \in W$ is a fully-specified SCM
- Equivalently, a world is a function from exogenous variables to {0,1} and from endogenous variables to a dependence function, which in turn maps situations (a node's parents' values) to {0,1}

As is standard, we'll take propositions to be sets of worlds

We'll say that an agent believes some proposition $\varphi \subseteq W$ ($B\varphi$), iff all of the worlds in the agent's belief-set $\beta \subseteq W$ make that proposition true: $\beta \subseteq \varphi$.

We'll model awareness by using an equivalence relation \sim on W, which induces the partition \aleph : our **awareness partition**

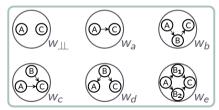
Intuitively, \sim and \aleph together represent the indistinguishability of worlds: if $w_1 \sim w_2$, then the agent cannot tell w_1 and w_2 apart (= is unaware of any difference)

$$\begin{array}{c} \textcircled{\textcircled{}} \end{array}{\end{array}}$$

A space of possible worlds

We'll model awareness by using an equivalence relation \sim on W, which induces the partition \aleph : our **awareness partition**

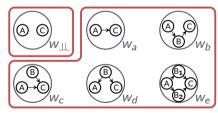
Intuitively, \sim and \aleph together represent the indistinguishability of worlds: if $w_1 \sim w_2$, then the agent cannot tell w_1 and w_2 apart (= is unaware of any difference)



A space of possible worlds An undifferentiating awareness partition x

We'll model awareness by using an equivalence relation \sim on W, which induces the partition \aleph : our **awareness partition**

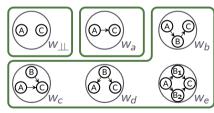
Intuitively, \sim and \aleph together represent the indistinguishability of worlds: if $w_1 \sim w_2$, then the agent cannot tell w_1 and w_2 apart (= is unaware of any difference)



A space of possible worlds An undifferentiating awareness partition X Differentiating independence

We'll model awareness by using an equivalence relation \sim on W, which induces the partition \aleph : our **awareness partition**

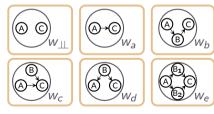
Intuitively, \sim and \aleph together represent the indistinguishability of worlds: if $w_1 \sim w_2$, then the agent cannot tell w_1 and w_2 apart (= is unaware of any difference)



A space of possible worlds An undifferentiating awareness partition X Differentiating independence Further differentiating a direct cause

We'll model awareness by using an equivalence relation \sim on W, which induces the partition \aleph : our **awareness partition**

Intuitively, \sim and \aleph together represent the indistinguishability of worlds: if $w_1 \sim w_2$, then the agent cannot tell w_1 and w_2 apart (= is unaware of any difference)



A space of possible worlds An undifferentiating awareness partition & Differentiating independence Further differentiating a direct cause A maximally differentiated partition &

Explicit and Implicit Beliefs

This new tool allows us better model an agent's beliefs relative to their awareness

First, call two partitions Q_1 and Q_2 orthogonal (in the sense of Lewis, 1988) iff for all $X \in Q_1$ and $Y \in Q_2$, if $X \neq \emptyset$ and $Y \neq \emptyset$, then $X \cap Y \neq \emptyset$

We'll say that an agent is **unaware of a proposition** φ iff the partition of W induced by the issue of φ (i.e., the set $\{\varphi, \overline{\varphi}\}$) is orthogonal to \aleph (their prior awareness partition)

- In other words: if differentiating on the basis of φ subdivides every cell of the prior partition \aleph , then the agent wasn't aware of φ as a factor along which to distinguish
- On the other hand, if any two worlds were already differentiated only by $\varphi,$ then the agent must already have been aware of φ

With this, we can say:

An agent has an **explicit belief** in φ iff they believe φ and are aware of φ An agent has an **implicit belief** in φ iff they believe φ and are **not** aware of φ

(Un)Awareness of Different Sorts And underspecification

This awareness partition captures an agent's (in)attention to the distinctions among worlds, including in the identities of nodes known to them

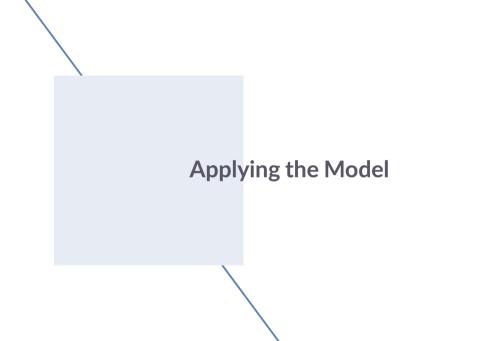
 e.g., an agent might attend to the possibility of Alex buying a turtle, but might not bother to differentiate among species or individuals being bought (though those are theoretically differentiable possibilities)

The same is true for edges, dependencies that may or may obtain between nodes

The partition also captures what Franke and Jager 2011 calls the "filtering" of "unmentionables": collapsing over propositional variables an agent is not even aware enough of to include in their model

- this includes variables that an agent simply doesn't know about to even model
- as well as variables collapsed to a single node, i.e., complexes treated as simplex

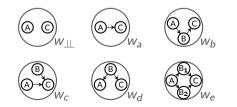
In this sense, each SCM is **underspecified** for the values of variables not modeled



The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - a. Skeptical Sam: What do you mean, how so?

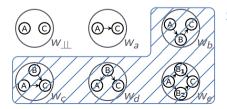
Skeptical Sam is willing to believe that there is *some* dependency between *A* and *C*, but maintains uncertainty among the models that make that true (other than simply)



The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - a. Skeptical Sam: What do you mean, how so?

Skeptical Sam is willing to believe that there is *some* dependency between *A* and *C*, but maintains uncertainty among the models that make that true (other than simply)



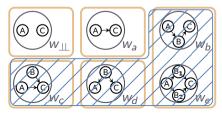
Sam's belief-set

The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - a. Skeptical Sam: What do you mean, how so?

Skeptical Sam is willing to believe that there is *some* dependency between *A* and *C*, but maintains uncertainty among the models that make that true (other than simply)

Crucially, Sam is *aware* of their uncertainty: their belief-set crosses multiple cells of \aleph



Sam's belief-set

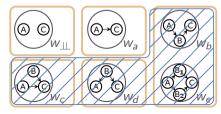
Sam's awareness partition ℵ

The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - a. Skeptical Sam: What do you mean, how so?

Skeptical Sam is willing to believe that there is *some* dependency between *A* and *C*, but maintains uncertainty among the models that make that true (other than simply)

Crucially, Sam is *aware* of their uncertainty: their belief-set crosses multiple cells of ℵ



Sam's belief-set

Sam's awareness partition ℵ

Skeptical doesn't mean a narrow belief-set; it means not ruling things out lightly

The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - b. Naïve Nat: Oh, okay. Good to know.

In coming to believe that [A > C], Naïve Nat must distinguish w_{\perp} from the other worlds in order to then be able to rule it out

But not making any other distinctions, Nat is unaware of any uncertainty to resolve

14

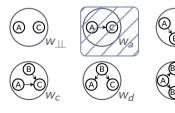
The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - b. Naïve Nat: Oh, okay. Good to know.

In coming to believe that [A > C], Naïve Nat must distinguish w_{\perp} from the other worlds in order to then be able to rule it out

But not making any other distinctions, Nat is unaware of any uncertainty to resolve

Nat's belief-set



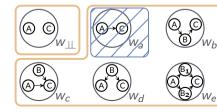
14

The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - b. Naïve Nat: Oh, okay. Good to know.

In coming to believe that [A > C], Naïve Nat must distinguish w_{\perp} from the other worlds in order to then be able to rule it out

But not making any other distinctions, Nat is unaware of any uncertainty to resolve



14

Nat's belief-set

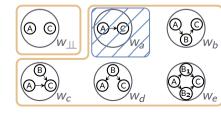
Nat's awareness partition 🕅

The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - b. Naïve Nat: Oh, okay. Good to know.

In coming to believe that [A > C], Naïve Nat must distinguish w_{\perp} from the other worlds in order to then be able to rule it out

But not making any other distinctions, Nat is unaware of any uncertainty to resolve



14

Nat's belief-set

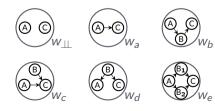
Nat's awareness partition X

This Nat has an **implicit belief** in simplicity, as they would reject any B (yet unconsidered)

The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - c. Guessing Gal: Oh, because Charlie would be jealous?

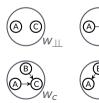
Gal is aware of some uncertainty yet to be resolved, but is less uncertain than Sam, enough so as to wager a guess (among those differentiable remaining worlds)

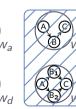


The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - c. Guessing Gal: Oh, because Charlie would be jealous?

Gal is aware of some uncertainty yet to be resolved, but is less uncertain than Sam, enough so as to wager a guess (among those differentiable remaining worlds)



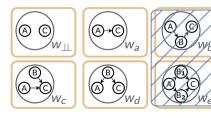


Gal's belief-set

The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - c. Guessing Gal: Oh, because Charlie would be jealous?

Gal is aware of some uncertainty yet to be resolved, but is less uncertain than Sam, enough so as to wager a guess (among those differentiable remaining worlds)



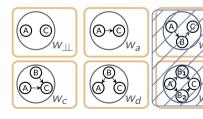
Gal's belief-set

Gal's awareness partition ℵ

The differing behaviors of our prototypical agents correspond to different states of awareness and belief—even while they all accept [A > C] as true

- (5) If Alex buys a turtle, Charlie will be upset.
 - c. Guessing Gal: Oh, because Charlie would be jealous?

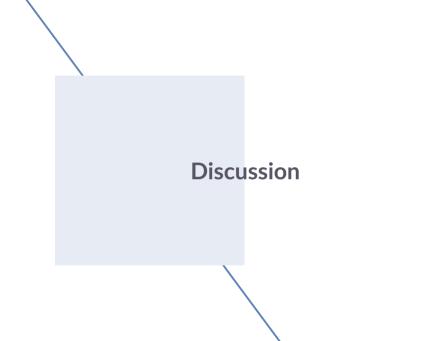
Gal is aware of some uncertainty yet to be resolved, but is less uncertain than Sam, enough so as to wager a guess (among those differentiable remaining worlds)



Gal's belief-set

Gal's awareness partition ℵ

Here Gal has an **explicit belief** that there is some intervening factor, and guesses among them



Minimal Representations

Introducing such awareness sets helps keep representations **tractable**: agents aren't necessarily modeling at the level of granularity that we are as theorists (though our representations, too, are bounded, and could always be more detailed)

We can model an agent's stance regarding a world (e.g., implicit belief) without their having to have a mental representation of that world

Per Swanson 2010, the interpretation of causal talk is guided by conversational norms regarding what count as "good representatives" for causal paths

We aim for our model to offer a formal scaffolding for expressing these sorts of ideas, as we take economical representations to be a crucial component in the pragmatics of causal talk

Awareness and Reasoning

We know that awareness plays a role in human behavior:

- awareness of alternatives affects judgments about causal consequences (Fernbach and Darlow, 2010)
- determining the set of relevant alternatives influences pragmatic reasoning about conditionals (Grusdt, Lassiter, and Franke, 2022)

Having the ability to model an agent's awareness, and to differentiate between their explicit and implicit beliefs, allows us to better describe (and make predictions about) behavior in such paradigms

Differentiating between an agent's knowledge and attention in this way also suggests a cline in terms of the **ease of acceptability** of conversational moves



Differentiating between an agent's knowledge and attention in this way also suggests a cline in terms of the **ease of acceptability** of conversational moves

easier updates with simple information gain (about node values or dependencies) that don't affect an agent's awareness or contradict prior beliefs

harder

Differentiating between an agent's knowledge and attention in this way also suggests a cline in terms of the **ease of acceptability** of conversational moves

easier updates with simple information gain (about node values or dependencies) that don't affect an agent's awareness or contradict prior beliefs
 harder updates that contradict an agent's explicit beliefs, require belief revision

Differentiating between an agent's knowledge and attention in this way also suggests a cline in terms of the **ease of acceptability** of conversational moves

easier updates with simple information gain (about node values or dependencies) that don't affect an agent's awareness or contradict prior beliefs
 updates that expand an agent's awareness, without challenging any of their implicit beliefs
 harder updates that contradict an agent's explicit beliefs, require belief revision

Differentiating between an agent's knowledge and attention in this way also suggests a cline in terms of the **ease of acceptability** of conversational moves

easier updates with simple information gain (about node values or dependencies) that don't affect an agent's awareness or contradict prior beliefs
 updates that expand an agent's awareness, without challenging any of their implicit beliefs
 updates that expand an agent's awareness in ways that challenge their implicit beliefs, and so may also require belief revision updates that contradict an agent's explicit beliefs, require belief revision

Differentiating between an agent's knowledge and attention in this way also suggests a cline in terms of the **ease of acceptability** of conversational moves

easier updates with simple information gain (about node values or dependencies) that don't affect an agent's awareness or contradict prior beliefs
 updates that expand an agent's awareness, without challenging any of their implicit beliefs
 updates that expand an agent's awareness in ways that challenge their implicit beliefs, and so may also require belief revision updates that contradict an agent's explicit beliefs, require belief revision

An open question whether this cline maps onto distinct patterns of responses

To Do List

- As it stands now, this model is static; we hope to develop a version which is dynamic, tracking changes in awareness across discourse moves
- The model at present tracks a single agent; we hope to develop a multi-speaker extension of the model, to model both mutual attention and how discrepancies affect how discourse unfolds
- Testing the empirical predictions associated with the spectrum of the ease of acceptability
- Making precise the linking hypotheses between Swanson 2010-style norms and our model—and then developing methods to test them



References I

Beller, Ari and Tobias Gerstenberg (2023). A counterfactual simulation model of causal language. PsyArXiv Preprint. DOI: 10.31234/osf.io/xv8hf.

Bjorndahl, Adam and Todd Snider (2016). "Informative Counterfactuals". In: Semantics and Linguistic Theory (SALT). Ed. by Sarah D'Antonio, Mary Moroney, and Carol Rose Little. Vol. 25. LSA and CLC Publications, pp. 1–17. DOI: 10.3765/salt.v25i0.3077.

Briggs, Rachael (2012). "Interventionist counterfactuals". In: *Philosophical Studies* 160.1, pp. 139–166.
Ciardelli, Ivano, Jeroen Groenendijk, and Floris Roelofsen (2009). "Attention! 'Might' in Inquisitive Semantics". In: *Semantics and Linguistic Theory (SALT)*. Ed. by Ed Cormany, Satoshi Ito, and David Lutz. Vol. 19. CLC Publications, pp. 91–108. DOI: 10.3765/salt.v19i0.2520.

References II

Fernbach, Philip and Adam Darlow (2010), "Causal Conditional Reasoning and Conditional Likelihood". In: *Cognitive Science Society*. Ed. by Stellan Ohlsson and Richard Catrambone. Vol. 32, pp. 1088–1093. Franke, Michael and Tikitu de Jager (2011), "Now That You Mention It: Awareness Dynamics in Discourse and Decisions". In: Language, Games, and Evolution: Trends in Current Research on Language and Game Theory, Ed. by Anton Benz et al. Berlin, Heidelberg: Springer, pp. 60–91. Gerstenberg, Tobias (2022). "What would have happened? Counterfactuals, hypotheticals and causal judgements". In: Philosophical Transactions of the Royal Society B: Biological Sciences 377.1866. Grusdt, Britta, Daniel Lassiter, and Michael Franke (Oct. 2022). "Probabilistic modeling of rational communication with conditionals". In: Semantics and Pragmatics 15.13. DOI: 10.3765/sp.15.13. Hiddleston, Eric (2005), "A causal theory of counterfactuals". In: Noûs 39.4, pp. 632–657.

References III

Lewis, David (1988). "Relevant Implication". In: Theoria 54.3, pp. 161–174.

Nadathur, Prerna and Sven Lauer (June 2020). "Causal necessity, causal sufficiency, and the implications

of causative verbs". In: Glossa: A journal of general linguistics 5.1. ISSN: 2397-1835. DOI:

10.5334/gjgl.497.

Pearl, Judea (2000). Causality: Models, Reasoning and Inference. Cambridge University Press.

Quillien, Tadeg and Christopher G. Lucas (2023). "Counterfactuals and the logic of causal selection". In:

Psychological Review. DOI: 10.1037/rev0000428.

Stalnaker, Robert (1978). "Assertion". In: *Syntax and Semantics* 9. Ed. by Peter Cole, pp. 315–332. Starr, W. (2013). "Structured Possible Worlds". Ms. Cornell University.

References IV

Swanson, Eric (2010). "Lessons from the context sensitivity of causal talk". In: *The Journal of Philosophy* 107.5, pp. 221–242.

Westera, Matthijs (Sept. 2022). "Attentional Pragmatics: A pragmatic approach to exhaustivity". In: Semantics and Pragmatics 15.10, pp. 1–51. ISSN: 1937-8912. DOI: 10.3765/sp.15.10.
Woodward, James (2003). Making Things Happen: A Theory of Causal Explanation. Oxford University Press.