Connecting probabilistic models and inference datasets to semantic theory

Julian Grove

EMG (Northwestern U.) / LEAP (U. of Chicago), October 15, 2024

FACTS.lab, Department of Linguistics, University of Rochester

Joint work (Grove and White 2024c)

Probabilistic dynamic semantics

Julian Grove and Aaron Steven White
University of Rochester

Abstract We introduce the framework of Probabilistic Dynamic Semantics (PDS), which we use to seamlessly integrate dynamic semantic analyses of lexical and discourse phenomena in the Montagovian tradition with probabilistic models of linguistic inference datasets. We show how PDS provides a general standpoint from which to understand uncertainty and context sensitivity in discourse and briefly illustrate applications to anaphora and vagueness.

Aaron Steven White
University of Rochester

Motivation

Gradience

(1) a. Whenever anyone laughed, the magician scowled and their assistant smirked.

b. They were secretely pleased.

(2) Was the magician pleased?

- *they* (sg) $\rightsquigarrow$ the magician ✓ ("Yes!")
- *they* (sg) $\rightsquigarrow$ the assistant ✓ ("No!")
- *they* (pl) $\rightsquigarrow$ the magician + the assistant ✓ ("Yes!")
- *they* $\rightsquigarrow$ a person who laughed X

Average response:

"No!" [—————————|-|—————] "Yes!"

Big questions

- How should we model gradience, as it pertains to inference (e.g., semantic) judgments?
  - Can we use the usual semantic toolkit (λ)?
  - Should we construct statistical (e.g., mixed effects) models?
  - Should we use something new, such as probabilistic programming languages (e.g., Church (Goodman, Mansinghka, et al. 2008) or WebPPL (Goodman and Stuhlmüller 2014))?

$\downarrow\downarrow\downarrow$

- Regardless, how should we integrate semantic theories into probabilistic models of human inference judgment data?

Zooming in

How should we integrate semantic theories into probabilistic models of human inference judgment data?

- What are the parameters?
- What generative model relates them to one another?

- What distributions of values do semantic (and pragmatic) parameters take on?
- Which semantic theories best fit a given dataset?

Common practice

- What are the parameters?
- What generative model relates them to one another?

- Not always clear what role the parts of the semantic theory play in the generative model.

What I'm advocating

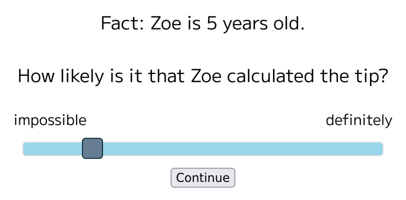

- Idea: once the semantic theory is fixed, so are the parameters and the generative model.
- Semantic theories may then be quantitatively compared, directly.
- Case studies:
    - inferences about world knowledge
    - the veridicality inferences generated by clause-embedding predicates

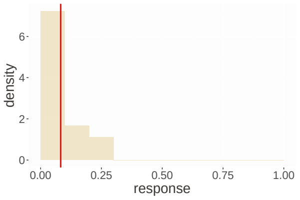
Gradience in inference judgments

World knowledge (Degen and Tonhauser 2021)

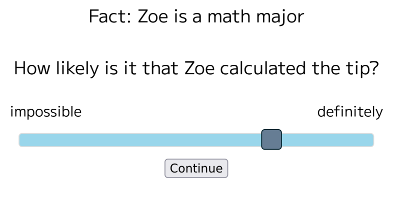
Task: judge some inference, given a background fact…

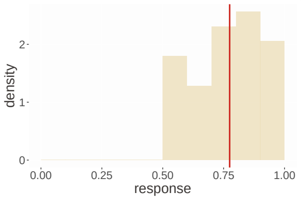
- “low” condition





- “high” condition





Factivity

(1) Jo loves that Mo Left.
$\rightsquigarrow$ Mo left.

This inference patterns like a presupposition, using family-of-sentence tests (Chierchia and McConnell-Ginet 1990):

(2) a. Jo doesn't love that Mo Left.

b. Does Jo love that Mo left?

c. If Jo loves that Mo Left, she'll also love that Bo left.
$\rightsquigarrow$ Mo left.

Factive inference judgments

What sorts of inference patterns arise from uses of clause-embedding predicates in an experimental setting?

- What happens if you ask someone to rate the likelihood that Mo left, given that *Jo loves that Mo left* is true.

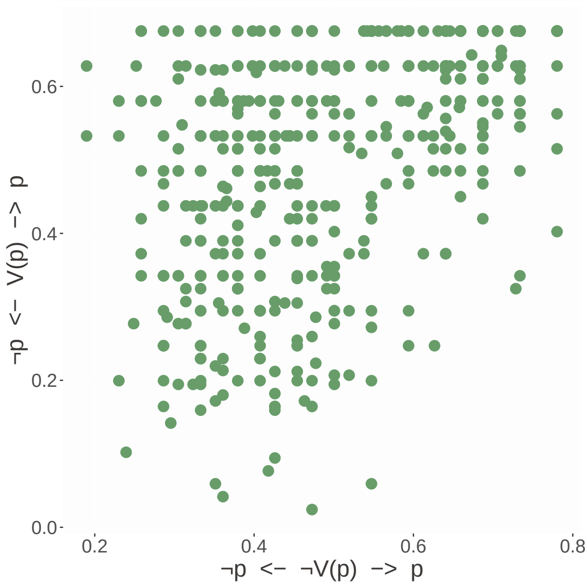
White and Rawlins (2018)

‘Someone {discovered, didn’t discover} that a particular thing happened.’

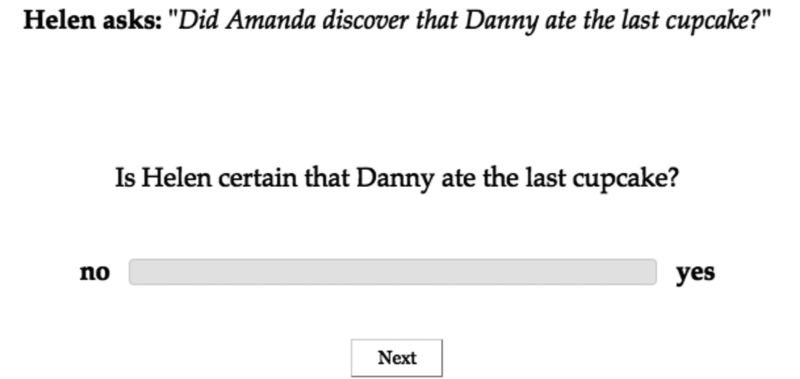
‘Did that thing happen?’

(“yes” / “maybe or maybe not” / “no”)

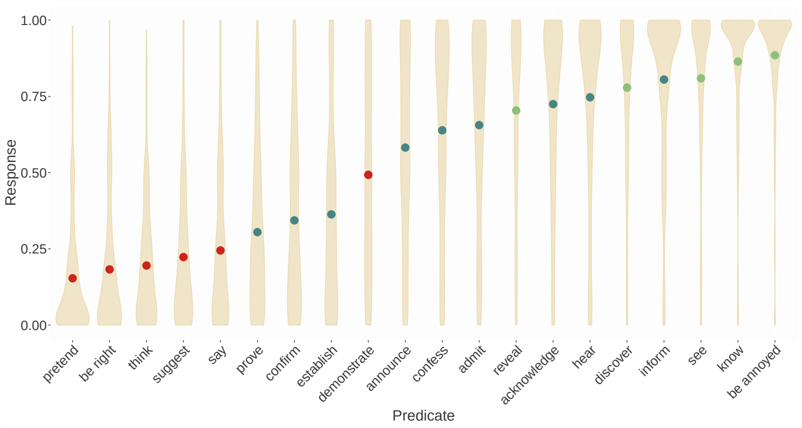
White and Rawlins (2018)



Degen and Tonhauser (2022)



Degen and Tonhauser (2022)



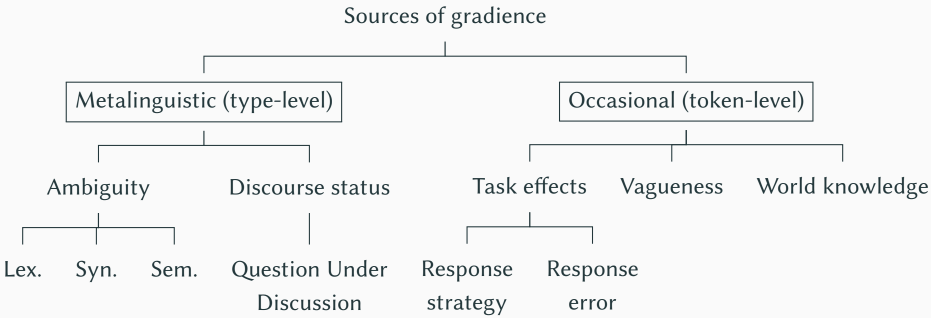
Why gradience?

Why do inference judgments display gradience?

And can semantic theory be made to countenance it?

??

Bird's-eye view



(1) Whenever anyone laughed, the magician scowled and their assistant smirked. They were secretely pleased.

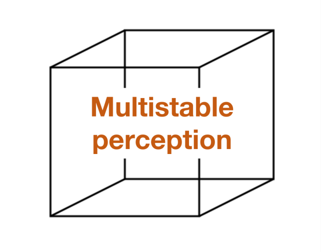
(2) My uncle ran the race.

(3) Al is tall.

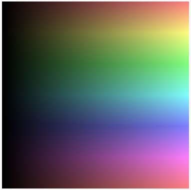
(4) Zoe is a math major $\rightsquigarrow$ Zoe is good at math.

The intuition

- Metalinguistic (type-level) sources:



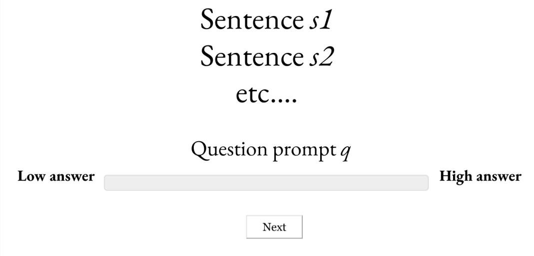
- Occasional (token-level) sources:



Probabilistic dynamic semantics

Basic idea

We should think of an experimental trial as a little discourse.



- Sentences: Start with a prior distribution over *discourse states*. Update this prior with $[\![s1]\!]$, then $[\![s2]\!]$, etc.
- Question: Push $[\![q]\!]$ onto the QUD stack (Farkas and Bruce 2010).
- Answer: Grab $[\![q]\!]$ from the QUD stack; respond.

Upshot: probabilistic models of data and semantic analyses are one and the same.

Components of the system

- Discourses
- Denotations of basic expressions
- Assertions (a type of discourse)
- Questions (a type of discourse)
- Responding to questions

A discourse

A discourse is a function of type

$$\sigma \longrightarrow \mathsf{P}\sigma'$$

- reads in a discourse state
- gives back a probability distribution over new discourse states

We can grab things from the discourse state:

- Given the state s, $\mathrm{CG}(s)$ is the common ground of s.

$$\mathrm{CG}(s) : \mathsf{P}\iota$$

Basic meanings

A basic meaning is a function of type

$$\sigma \longrightarrow \mathsf{P}(\alpha \times \sigma')$$

- reads in a discourse state
- gives back a probability distribution over ordinary meanings, together with new discourse states

Example: *tall*

(1) Jo is tall.
$\rightsquigarrow$ Jo's height exceeds some contextually salient threshold.

$$[\![\textit{tall}]\!] = \lambda s.\boxed{\langle \lambda x, i.\mathrm{height}(i)(x) \geq \mathrm{d}_{\textit{tall}}(i), s\rangle}$$
$$: \sigma \longrightarrow \mathsf{P}((e \longrightarrow \iota \longrightarrow t) \times \sigma)$$

$$[\![\textit{Jo is tall}]\!] = \lambda s.\boxed{\langle \lambda i.\mathrm{height}(i)(\mathrm{j}) \geq \mathrm{d}_{\textit{tall}}(i), s\rangle}$$
$$: \sigma \longrightarrow \mathsf{P}((\iota \longrightarrow t) \times \sigma)$$

Making an assertion

An assertion is a discourse derived from a sentence meaning.

$$\texttt{assert} : (\sigma \rightarrow \mathsf{P}((\iota \rightarrow t) \times \sigma')) \rightarrow \sigma \rightarrow \mathsf{P}\sigma'$$

- reads in a sentence meaning
- reads in a discourse state
- gives back a probability distribution over new discourse states just like the old one, but whose common ground is updated with the sentence meaning

Asking a question

Asking a question involves deriving a discourse from a *question meaning*.
Questions denote sets of true short answer meanings (Hausser and Zaefferer 1978; Hausser 1983; Xiang 2021).

$$\begin{aligned} \mathsf{ask} \; : \; & (\sigma \to \mathsf{P}((\alpha \to \iota \to t) \times (\sigma_1' \times \delta \times \sigma_2'))) \to \\ & \sigma \to \mathsf{P}(\sigma_1' \times ((\alpha \to \iota \to t) \times \delta) \times \sigma_2') \end{aligned}$$

- reads in a question meaning
- adds it to the top of the QUD stack

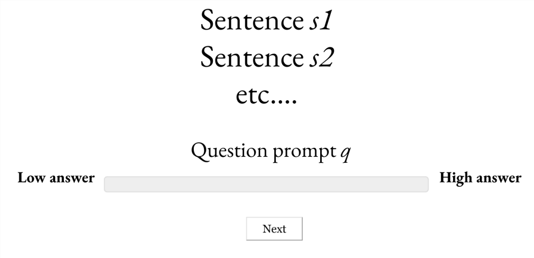
Responding to a question

You can respond to a question at the top of the QUD stack.

$$\texttt{respond}^{f_\Phi : \alpha \to \mathsf{P}\rho} : \mathsf{P}\sigma \to (\sigma \to \mathsf{P}\sigma') \to \mathsf{P}\rho$$

- reads in a probability distribution over starting states
- reads in some discourse
- gives back a probability distribution over responses to the QUD at the top of the stack, given some testing instrument (e.g., a slider scale, a Likert scale, etc.)

Modeling a trial

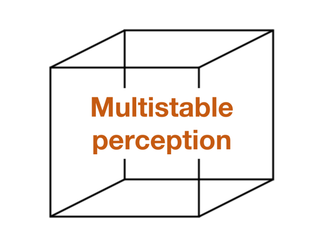


- Start with some *prior* over discourse states.
- $\texttt{assert}([\![s1]\!]); \texttt{assert}([\![s2]\!])$
- $\texttt{ask}([\![q]\!])$
- $\texttt{respond}^{f_\Phi}(\mathit{prior})(\texttt{assert}([\![s1]\!]); \texttt{assert}([\![s2]\!]); \texttt{ask}([\![q]\!]))$

This is a probability distribution over responses
(an actual model of the data)!

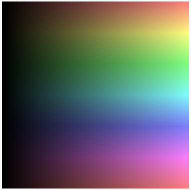
Sources of probabilistic uncertainty

- Metalinguistic (type-level) sources:



tied to the discourse state

- Occasional (token-level) sources:



tied to the common ground

Case studies

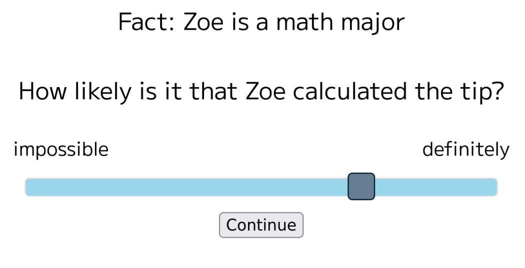
Get your phone (Grove and White 2024a)

Factivity, presupposition projection, and the role of discrete knowledge in gradient inference judgments*

Julian Grove and Aaron Steven White
University of Rochester

Abstract We investigate whether the factive presuppositions associated with some clause-embedding predicates are fundamentally discrete in nature—as classically assumed—or fundamentally gradient—as recently proposed (the Gradient Projection Principle of Tonhauser, Beaver, and Degen 2018). To carry out this investigation, we develop statistical models of presupposition projection that implement these two hypotheses, fit these models to existing inference judgment data aimed at measuring factive presuppositions (Degen and Tonhauser

World knowledge (Degen and Tonhauser 2021)

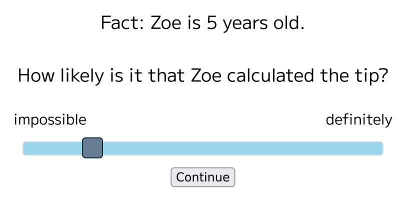


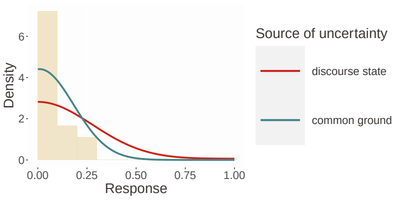
- `assert`(⟦*Zoe is a math major*⟧)
- `ask`(⟦*how likely is it that Zoe calculated the tip*⟧)

We can vary whether uncertainty about world knowledge is tied to the common ground vs. the discourse state.

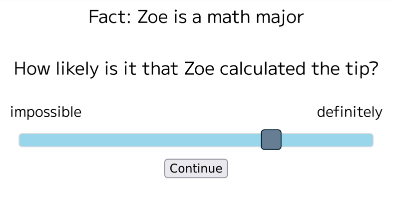
World knowledge (Degen and Tonhauser 2021)

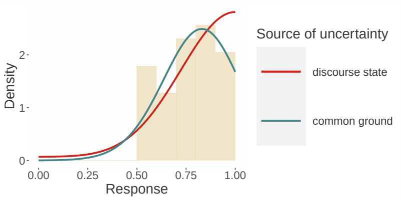
- “low” condition



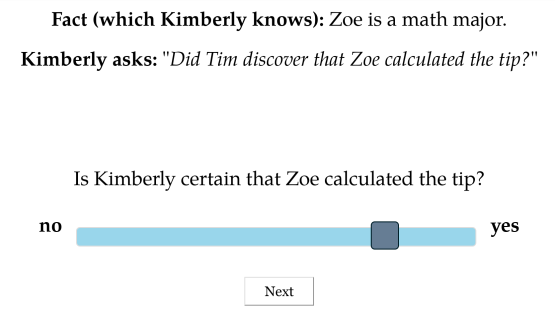


- “high” condition





Factivity (Degen and Tonhauser 2021)

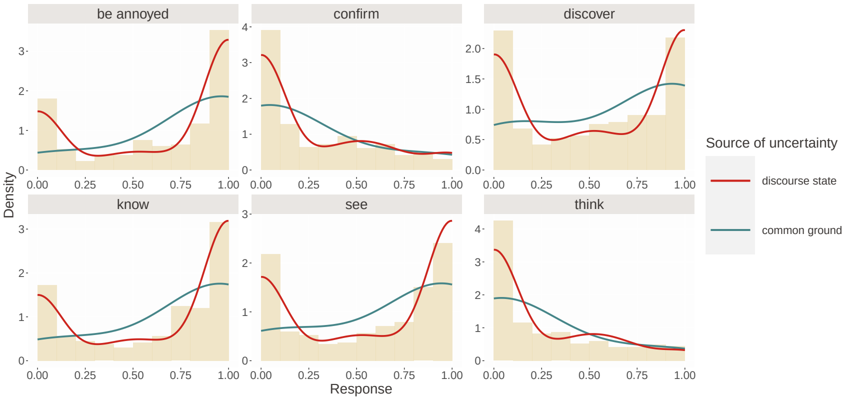


- $\texttt{assert}(\llbracket \textit{Zoe is a math major} \rrbracket)$
- $\texttt{assert}(\llbracket \textit{Tim discovered that that Zoe calculated the tip} \rrbracket)$
- $\texttt{ask}(\llbracket \textit{how likely is it that Zoe calculated the tip} \rrbracket)$

We can vary whether uncertainty about *veridicality inferences* are tied to the common ground vs. the discourse state.

Factivity (Degen and Tonhauser 2021)

(For each predicate, averaged across complement clauses and background facts.)



We don't quite model the question prompt correctly.

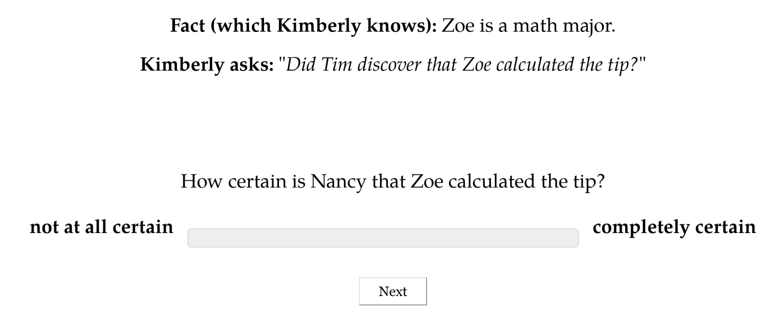
We can do better!

Modeling the prompt in inference judgment tasks

Julian Grove & Aaron Steven White*

Abstract. We show that when analyzing data from inference judgment tasks, it can be important to incorporate into one's data analysis regime an explicit representation of the semantics of the natural language prompt used to guide participants on the task. To demonstrate this, we conduct two experiments within an existing experimental paradigm focused on measuring factive inferences, while manipulating the prompt participants receive in small but semantically potent ways. In statistical model comparisons couched within the framework of probabilistic dynamic semantics, we find that probabilistic models structured, in part, by the semantics of the prompt fit better to

Modeling the question prompt (Grove and White 2024b)

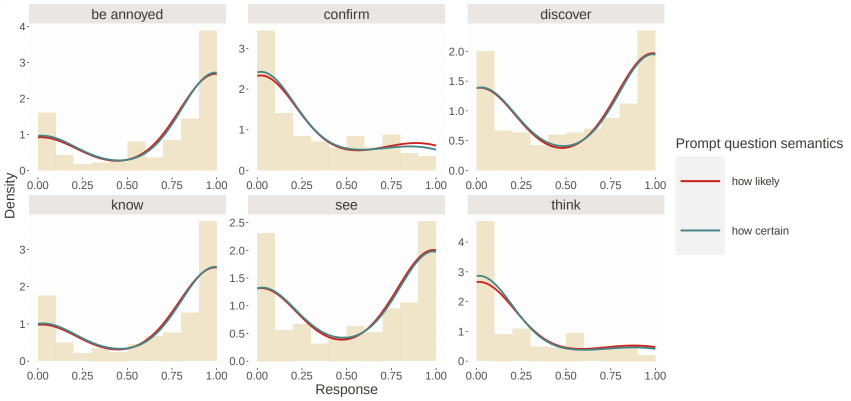


- assert($[\![$*Zoe is a math major*$]\!]$)
- assert($[\![$*Tim discovered that that Zoe calculated the tip*$]\!]$)
- ask($[\![$*how certain is Kimberly that Zoe calculated the tip*$]\!]$)

We can vary whether uncertainty about *veridicality inferences* are tied to the common ground vs. the discourse state.

Modeling the question prompt (Grove and White 2024b)

(Again, averaged across complement clauses and background facts.)



A reliable (although visually subtle) difference.

Conclusion

Summary

- We can connect semantic theory to the gradience observed experimentally using only the traditional semantic toolkit (more or less).
- Further, we can take advantage of dynamic theories of discourse in order to model the dynamic semantics of experimental trials.
- This strategy—of using an already-available formal apparatus—allows our linking assumptions between semantic theories and probabilistic models to be made formally precise.
- Crucially, using semantic theory to characterize inference judgment data directly allows for immediate theory comparison.

Thanks!

References i

- Chierchia, Gennaro and Sally McConnell-Ginet (1990). ***Meaning and Grammar: An Introduction to Semantics.*** Cambridge: MIT Press.
- Degen, Judith and Judith Tonhauser (2021). **"Prior Beliefs Modulate Projection".** In: *Open Mind* 5, pp. 59–70. DOI: 10.1162/opmi_a_00042.
- Degen, Judith and Judith Tonhauser (2022). **"Are there factive predicates? An empirical investigation".** en. In: *Language* 98.3. Publisher: Linguistic Society of America, pp. 552–591. DOI: 10.1353/lan.0.0271.

References ii

Farkas, Donka F. and Kim B. Bruce (2010). **"On Reacting to Assertions and Polar Questions".** In: *Journal of Semantics* 27.1, pp. 81–118. DOI: 10.1093/jos/ffp010.

Goodman, Noah D., Vikash K. Mansinghka, et al. (2008). **"Church: a language for generative models".** In: *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence.* UAI'08. Arlington, Virginia, USA: AUAI Press, pp. 220–229.

Goodman, Noah D. and Andreas Stuhlmüller (2014). ***The Design and Implementation of Probabilistic Programming Languages.*** http://dippl.org.

- Grove, Julian and Aaron Steven White (2024a). ***Factivity, presupposition projection, and the role of discrete knowlege in gradient inference judgments.*** LingBuzz Published In: submitted.
- Grove, Julian and Aaron Steven White (2024b). ***Modeling the prompt in inference judgment tasks.*** LingBuzz Published In: Proceedings of ELM 3.
- Grove, Julian and Aaron Steven White (2024c). ***Probabilistic dynamic semantics.*** LingBuzz Published In:

Hausser, Roland and Dietmar Zaefferer (1978). **"Questions and Answers in a Context-Dependent Montague Grammar".** en. In: *Formal Semantics and Pragmatics for Natural Languages.* Ed. by F. Guenthner and S. J. Schmidt. Dordrecht: Springer Netherlands, pp. 339–358. DOI: 10.1007/978-94-009-9775-2_12.

Hausser, Roland R. (1983). **"The Syntax and Semantics of English Mood".** en. In: *Questions and Answers.* Ed. by Ferenc Kiefer. Dordrecht: Springer Netherlands, pp. 97–158. DOI: 10.1007/978-94-009-7016-8_6.

References v

White, Aaron Steven and Kyle Rawlins (2018). **"The role of veridicality and factivity in clause selection".** In: *NELS 48: Proceedings of the Forty-Eighth Annual Meeting of the North East Linguistic Society*. Ed. by Sherry Hucklebridge and Max Nelson. Vol. 48. University of Iceland: GLSA (Graduate Linguistics Student Association), Department of Linguistics, University of Massachusetts, pp. 221–234.

Xiang, Yimei (2021). **"A hybrid categorial approach to question composition".** en. In: *Linguistics and Philosophy* 44.3, pp. 587–647. DOI: 10.1007/s10988-020-09294-8.