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

Julian Grove

LiMe Lab, September 8, 2023

FACTS.lab, University of Rochester

Joint work

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 (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 2021), and compare the models' fit to the data using standard statistical model comparison metrics. We

Aaron Steven White
UofR

Motivation

Big question

How should we use inference datasets to test semantic theories?

Big question

How should we use inference datasets to test semantic theories?

A couple of sub-questions:

Big question

How should we use inference datasets to test semantic theories?

A couple of sub-questions:

- How do we get at "true" inference judgments?

Big question

How should we use inference datasets to test semantic theories?

A couple of sub-questions:

- How do we get at "true" inference judgments?
  - Judgment data tends to be influenced by non-semantic factors, e.g., participant response strategies, participant accuracy.

Big question

How should we use inference datasets to test semantic theories?

A couple of sub-questions:

- How do we get at "true" inference judgments?
  - Judgment data tends to be influenced by non-semantic factors, e.g., participant response strategies, participant accuracy.
- Different question: how do we *interpret* inference data?

Big question

How should we use inference datasets to test semantic theories?

A couple of sub-questions:

- How do we get at "true" inference judgments?
  - Judgment data tends to be influenced by non-semantic factors, e.g., participant response strategies, participant accuracy.
- Different question: how do we *interpret* inference data?
  - Do we compare mean responses among conditions? If so, what do such means represent?

Big question

How should we use inference datasets to test semantic theories?

A couple of sub-questions:

- How do we get at "true" inference judgments?
  - Judgment data tends to be influenced by non-semantic factors, e.g., participant response strategies, participant accuracy.
- Different question: how do we *interpret* inference data?
  - Do we compare mean responses among conditions? If so, what do such means represent?
  - Should we somehow take into account the whole response distribution?

Big question

How should we use inference datasets to test semantic theories?

A couple of sub-questions:

- How do we get at "true" inference judgments?
  - Judgment data tends to be influenced by non-semantic factors, e.g., participant response strategies, participant accuracy.
- Different question: how do we *interpret* inference data?
  - Do we compare mean responses among conditions? If so, what do such means represent?
  - Should we somehow take into account the whole response distribution?

Big question

How should we use inference datasets to test semantic theories?

A couple of sub-questions:

- How do we get at "true" inference judgments?
  - Judgment data tends to be influenced by non-semantic factors, e.g., participant response strategies, participant accuracy.
- Different question: how do we *interpret* inference data?
  - Do we compare mean responses among conditions? If so, what do such means represent?
  - Should we somehow take into account the whole response distribution?

We need linking assumptions…

This talk

We will do a few things:

This talk

We will do a few things:

- We will look at particular empirical domain: factive inferences.

This talk

We will do a few things:

- We will look at particular empirical domain: factive inferences.
- We will develop a compositional probabilistic semantics that allows us to formulate Bayesian models of inference data (following Grove and Bernardy (2023)).

This talk

We will do a few things:

- We will look at particular empirical domain: factive inferences.
- We will develop a compositional probabilistic semantics that allows us to formulate Bayesian models of inference data (following Grove and Bernardy (2023)).
- We will use this semantics to *combine* theories of factivity with linking assumptions seamlessly.

Factivity and gradience

Factivity

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

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.

Gradience

What sorts of inference patterns arise from uses of factive predicates in an experimental setting?

- E.g., 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.'

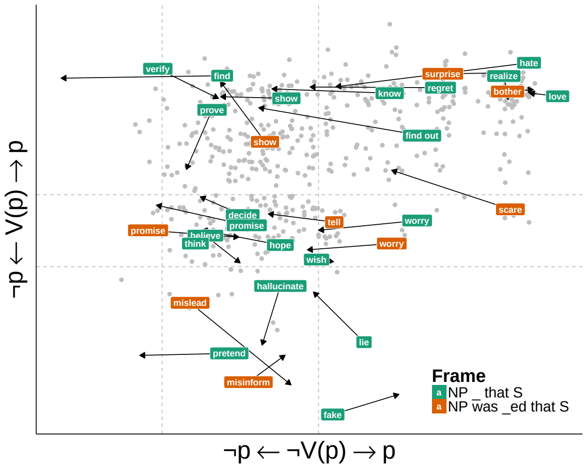
White and Rawlins (2018)

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

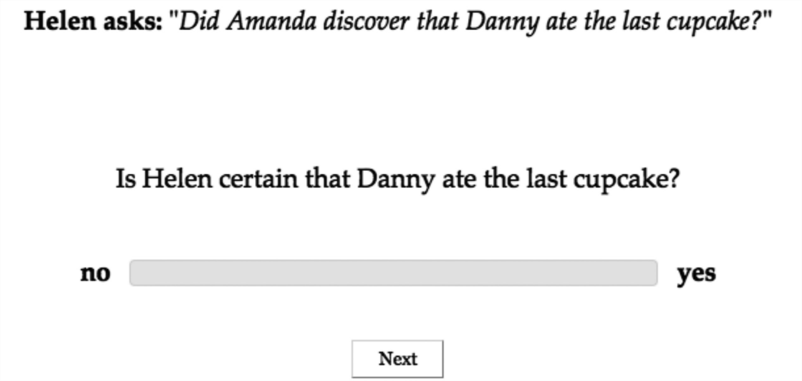
'Did that thing happen?'

(*yes*, *maybe or maybe not*, *no*)

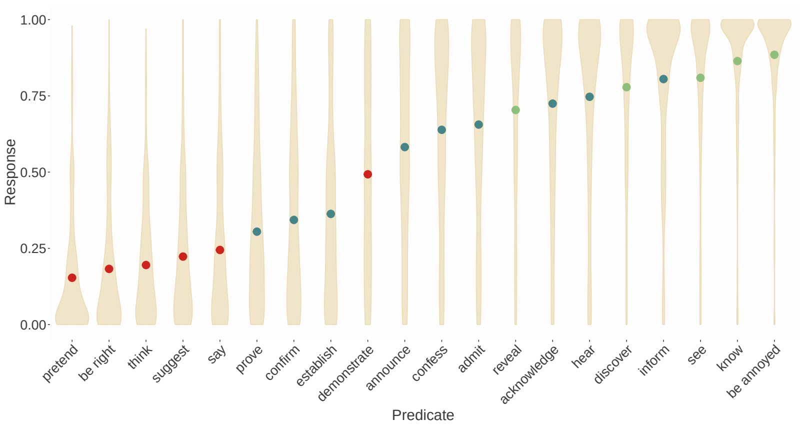
White and Rawlins (2018)



x-axis: negative polarity; *y*-axis: positive polarity


Helen asks: "*Did Amanda discover that Danny ate the last cupcake?*"
Is Helen certain that Danny ate the last cupcake?
no
yes
Next

Degen and Tonhauser (2022)



Why is there gradience?

One possibility:

Why is there gradience?

One possibility:

- Different predicates boost the certainty associated with an inference to different degrees (Tonhauser, Beaver, and Degen 2018).

Why is there gradience?

One possibility:

- Different predicates boost the certainty associated with an inference to different degrees (Tonhauser, Beaver, and Degen 2018).
- For example, *discover* makes its complement clause more likely to be true. (But not as much more likely as, say, *know*.)

Why is there gradience?

One possibility:

- Different predicates boost the certainty associated with an inference to different degrees (Tonhauser, Beaver, and Degen 2018).
- For example, *discover* makes its complement clause more likely to be true. (But not as much more likely as, say, *know*.)

Why is there gradience?

One possibility:

- Different predicates boost the certainty associated with an inference to different degrees (Tonhauser, Beaver, and Degen 2018).
- For example, *discover* makes its complement clause more likely to be true. (But not as much more likely as, say, *know*.)

Another possibility:

Why is there gradience?

One possibility:

- Different predicates boost the certainty associated with an inference to different degrees (Tonhauser, Beaver, and Degen 2018).
- For example, *discover* makes its complement clause more likely to be true. (But not as much more likely as, say, *know*.)

Another possibility:

- Predicates are generally ambiguous between being factive or non-factive. But different predicates are factive with different frequencies, and it is these frequencies which differ among one another in a gradient fashion.

Why is there gradience?

One possibility:

- Different predicates boost the certainty associated with an inference to different degrees (Tonhauser, Beaver, and Degen 2018).
- For example, *discover* makes its complement clause more likely to be true. (But not as much more likely as, say, *know*.)

Another possibility:

- Predicates are generally ambiguous between being factive or non-factive. But different predicates are factive with different frequencies, and it is these frequencies which differ among one another in a gradient fashion.
- It is just that *discover*'s average factivity is less than *know*'s.

Hypotheses

Two hypotheses about the source of gradience among by-predicate means:

Hypotheses

Two hypotheses about the source of gradience among by-predicate means:

- *The Fundamental Discreteness Hypothesis*
  Factivity is a discrete semantic property of at least some token occurrences clause-embedding predicates. (A given occurrence of a particular predicate either triggers a projective inference, or it does not trigger a projective inference.)

Hypotheses

Two hypotheses about the source of gradience among by-predicate means:

- *The Fundamental Discreteness Hypothesis*
  Factivity is a discrete semantic property of at least some token occurrences clause-embedding predicates. (A given occurrence of a particular predicate either triggers a projective inference, or it does not trigger a projective inference.)
- *The Fundamental Gradience Hypothesis*
  The gradient distinctions among predicates reflect the different gradient contributions specific predicates make to inferences about the truth of their complement clauses.

Hypotheses

The fundamental discreteness hypothesis represents the classical view (Kiparsky and Kiparsky 1970; Karttunen 1971, i.a.).

Hypotheses

The fundamental discreteness hypothesis represents the classical view (Kiparsky and Kiparsky 1970; Karttunen 1971, i.a.).

The fundamental gradience hypothesis represents a more recent view, i.e., that presupposition triggers can trigger inferences gradiently (Tonhauser, Beaver, and Degen 2018).

How to proceed

To investigate and compare these two hypotheses, we will do two things:

How to proceed

To investigate and compare these two hypotheses, we will do two things:

- Provide a modular probabilistic semantics that allows the two hypotheses to be stated precisely.

How to proceed

To investigate and compare these two hypotheses, we will do two things:

- Provide a modular probabilistic semantics that allows the two hypotheses to be stated precisely.
  - I.e., give a characterization of Bayesian models that encode the two hypotheses, and formulate explicit linking assumptions using the very same semantic repertoire.

How to proceed

To investigate and compare these two hypotheses, we will do two things:

- Provide a modular probabilistic semantics that allows the two hypotheses to be stated precisely.
    - I.e., give a characterization of Bayesian models that encode the two hypotheses, and formulate explicit linking assumptions using the very same semantic repertoire.
- Fit these models to inference data and compare the fits.

A modular probabilistic semantics

Grove and Bernardy (2023)

Following Grove and Bernardy (2023), we provide an interface for stating a probabilistic semantics using *monads*.

Grove and Bernardy (2023)

Following Grove and Bernardy (2023), we provide an interface for stating a probabilistic semantics using *monads*.

This just means we have the following ingredients:

Following Grove and Bernardy (2023), we provide an interface for stating a probabilistic semantics using *monads*.

This just means we have the following ingredients:

- A type constructor P that takes any type α onto the type $\mathsf{P}\alpha$ of *probabilistic programs* that return α's.

Grove and Bernardy (2023)

Following Grove and Bernardy (2023), we provide an interface for stating a probabilistic semantics using *monads*.

This just means we have the following ingredients:

- A type constructor P that takes any type α onto the type $\mathsf{P}\alpha$ of *probabilistic programs* that return α's.
- An operator 'bind'

$$(\sim) : \mathsf{P}\alpha \to (\alpha \to \mathsf{P}\beta) \to \mathsf{P}\beta$$

  allowing us to bind a probabilistic program of type $\mathsf{P}\alpha$ to a value of type α used inside another probabilistic program.

Grove and Bernardy (2023)

Following Grove and Bernardy (2023), we provide an interface for stating a probabilistic semantics using *monads*.

This just means we have the following ingredients:

- A type constructor P that takes any type α onto the type $\mathsf{P}\alpha$ of *probabilistic programs* that return α's.
- An operator 'bind'

$$(\leadsto) : \mathsf{P}\alpha \to (\alpha \to \mathsf{P}\beta) \to \mathsf{P}\beta$$

  allowing us to bind a probabilistic program of type $\mathsf{P}\alpha$ to a value of type α used inside another probabilistic program.

Grove and Bernardy (2023)

Following Grove and Bernardy (2023), we provide an interface for stating a probabilistic semantics using *monads*.

This just means we have the following ingredients:

- A type constructor P that takes any type α onto the type $\mathsf{P}\alpha$ of *probabilistic programs* that return α's.
- An operator 'bind'

$$(\sim) : \mathsf{P}\alpha \to (\alpha \to \mathsf{P}\beta) \to \mathsf{P}\beta$$

  allowing us to bind a probabilistic program of type $\mathsf{P}\alpha$ to a value of type α used inside another probabilistic program.

  (Basically allows us to write sampling statements: $x \sim m$.)

Grove and Bernardy (2023)

Following Grove and Bernardy (2023), we provide an interface for stating a probabilistic semantics using *monads*.

This just means we have the following ingredients:

- A type constructor P that takes any type α onto the type $\mathsf{P}\alpha$ of *probabilistic programs* that return α's.
- An operator 'bind'

  $$(\sim) : \mathsf{P}\alpha \to (\alpha \to \mathsf{P}\beta) \to \mathsf{P}\beta$$

  allowing us to bind a probabilistic program of type $\mathsf{P}\alpha$ to a value of type α used inside another probabilistic program.

  (Basically allows us to write sampling statements: $x \sim m$.)
- An operator 'return'

  $$\boxed{(\cdot)} : \alpha \to \mathsf{P}\alpha$$

tall as an example

In a non-probabilistic semantics, since *tall* is an adjective, you might give it a meaning of type $e \rightarrow t$:

$$\lambda x.\text{height}(x) \geq d$$

tall as an example

In a non-probabilistic semantics, since *tall* is an adjective, you might give it a meaning of type $e \to t$:

$$\lambda x.\mathrm{height}(x) \geq d$$

But where does the threshold d come from?

tall as an example

Solution: the meaning of *tall* is represented by a probability distribution.

tall as an example

Solution: the meaning of *tall* is represented by a probability distribution.

- A probabilistic program of type $\mathsf{P}(e \longrightarrow t)$:

$$d \sim \text{thresholdPrior}$$
$$\boxed{\lambda x.\text{height}(x) \geq d}$$

tall as an example

Solution: the meaning of *tall* is represented by a probability distribution.

- A probabilistic program of type $\mathsf{P}(e \rightarrow t)$:

$$d \sim \text{thresholdPrior}$$
$$\boxed{\lambda x.\text{height}(x) \geq d}$$

- Returns the property true of an individual if their height exceeds the threshold d.

Solution: the meaning of *tall* is represented by a probability distribution.

- A probabilistic program of type $\mathsf{P}(e \to t)$:

$$d \sim \text{thresholdPrior}$$
$$\boxed{\lambda x.\text{height}(x) \geq d}$$

- Returns the property true of an individual if their height exceeds the threshold d.
  - But now d takes on the probability distribution represented by thresholdPrior.

tall as an example

Solution: the meaning of *tall* is represented by a probability distribution.

- A probabilistic program of type $\mathsf{P}(e \to t)$:

$$d \sim \text{thresholdPrior}$$
$$\boxed{\lambda x.\text{height}(x) \geq d}$$

- Returns the property true of an individual if their height exceeds the threshold d.
  - But now d takes on the probability distribution represented by thresholdPrior.
  - So the meaning of *tall* represents a probability distribution over *properties*, each one fixed by some threshold d.

Two levels of probabilistic involvement

The plan is now to make the following move:

Two levels of probabilistic involvement

The plan is now to make the following move:

- Model linguistic meanings as programs, not of type $\mathsf{P}\alpha$, but of type $\mathsf{P}(\mathsf{P}\alpha)$.

Two levels of probabilistic involvement

The plan is now to make the following move:

- Model linguistic meanings as programs, not of type $\mathsf{P}\alpha$, but of type $\mathsf{P}(\mathsf{P}\alpha)$.
  - The "inner" P — $\mathsf{P}(\mathsf{P}\alpha)$ — represents uncertainty that arises on individual occasions of use and interpretation, even after the meanings of expressions have been fixed.

Two levels of probabilistic involvement

The plan is now to make the following move:

- Model linguistic meanings as programs, not of type $\mathsf{P}\alpha$, but of type $\mathsf{P}(\mathsf{P}\alpha)$.
    - The "inner" P — $\mathsf{P}(\mathsf{P}\alpha)$ — represents uncertainty that arises on individual occasions of use and interpretation, even after the meanings of expressions have been fixed.
    - The "outer" P — $\mathsf{P}(\mathsf{P}\alpha)$ — represents *metalinguistic* uncertainty; that is, uncertainty about what interpretation to apply in the first place.

Two levels of probabilistic involvement: example

Say you're in a noisy bar. Someone says *Jo is -all*…

Two levels of probabilistic involvement: example

Say you're in a noisy bar. Someone says *Jo is -all…*

$$[\![\textit{Jo is -all}]\!] : \mathtt{P}(\mathtt{P}t)$$

$$[\![\textit{Jo is -all}]\!] = \tau \sim \mathtt{Bernoulli}(0.7)$$

$$\begin{cases} \boxed{\begin{array}{l} d \sim \text{heightThreshold} \\ \boxed{\text{height(j)} \geq d} \end{array}} & \tau \\ \boxed{\begin{array}{l} d \sim \text{sizeThreshold} \\ \boxed{\text{size(j)} \leq d} \end{array}} & \neg\tau \end{cases}$$

Two levels of probabilistic involvement: example

Say you're in a noisy bar. Someone says *Jo is -all*…

$$[\![\textit{Jo is -all}]\!] : \mathtt{P}(\mathtt{P}t)$$

$$[\![\textit{Jo is -all}]\!] = \quad \tau \sim \mathtt{Bernoulli}(0.7)$$

$$\begin{cases} \boxed{\begin{array}{l} d \sim \text{heightThreshold} \\ \boxed{\text{height}(\text{j}) \geq d} \end{array}} & \tau \\ \boxed{\begin{array}{l} d \sim \text{sizeThreshold} \\ \boxed{\text{size}(\text{j}) \leq d} \end{array}} & \neg\tau \end{cases}$$

- 70% chance the utterances was *Jo is tall*, 30% chance it was *Jo is small.*

Two levels of probabilistic involvement: example

Say you're in a noisy bar. Someone says *Jo is -all*...

$$[\![\textit{Jo is -all}]\!] : \mathtt{P}(\mathtt{P}t)$$

$$[\![\textit{Jo is -all}]\!] = \tau \sim \mathtt{Bernoulli}(0.7)$$

$$\begin{cases} \boxed{\begin{array}{l} d \sim \text{heightThreshold} \\ \boxed{\text{height(j)} \geq d} \end{array}} & \tau \\ \boxed{\begin{array}{l} d \sim \text{sizeThreshold} \\ \boxed{\text{size(j)} \leq d} \end{array}} & \neg\tau \end{cases}$$

- 70% chance the utterances was *Jo is tall*, 30% chance it was *Jo is small*.
- Having fixed one adjective or the other, there is uncertainty that arises from the relevant degree threshold.

Models of factivity

Inference data: Degen and Tonhauser (2021)

Degen and Tonhauser (2021) investigate the projection behavior of twenty clause-embedding predicates, systematically varying the contexts in which the predicates are used.

Inference data: Degen and Tonhauser (2021)

Degen and Tonhauser (2021) investigate the projection behavior of twenty clause-embedding predicates, systematically varying the contexts in which the predicates are used.

- For any given complement clause, a background fact is provided which is intended to make the clause either likely or unlikely to be true…

Inference data: Degen and Tonhauser (2021)

Fact (which Elizabeth knows): Zoe is a math major.

Elizabeth asks: "*Did Tim pretend that Zoe calculated the tip?*"

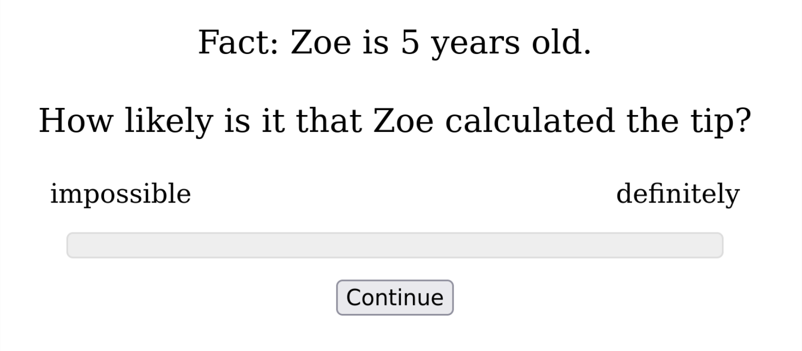
Is Elizabeth certain that Zoe calculated the tip?

no **yes**

Next

Inference data: Degen and Tonhauser (2021)

They additionally conduct a separate norming study intended to assess the prior certainties associated with such complement clauses, paired with their background facts.


Fact: Zoe is 5 years old.
How likely is it that Zoe calculated the tip?
impossible
definitely
Continue

Factivity and world knowledge

Our models of Degen et al.'s data vary whether two different sources of uncertainty are *metalinguistic* in nature versus uncertainty that is tied to particular interpretations:

Factivity and world knowledge

Our models of Degen et al.'s data vary whether two different sources of uncertainty are *metalinguistic* in nature versus uncertainty that is tied to particular interpretations:

- Uncertainty about factivity

Factivity and world knowledge

Our models of Degen et al.'s data vary whether two different sources of uncertainty are *metalinguistic* in nature versus uncertainty that is tied to particular interpretations:

- Uncertainty about factivity
- Uncertainty about world knowledge (i.e., the prior probability of the complement clause being true)

Factivity and world knowledge

Our models of Degen et al.'s data vary whether two different sources of uncertainty are *metalinguistic* in nature versus uncertainty that is tied to particular interpretations:

- Uncertainty about factivity
- Uncertainty about world knowledge (i.e., the prior probability of the complement clause being true)

Factivity and world knowledge

Our models of Degen et al.'s data vary whether two different sources of uncertainty are *metalinguistic* in nature versus uncertainty that is tied to particular interpretations:

- Uncertainty about factivity
- Uncertainty about world knowledge (i.e., the prior probability of the complement clause being true)

The fundamental discreteness hypothesis says that uncertainty about factivity is metalinguistic.

Factivity and world knowledge

Our models of Degen et al.'s data vary whether two different sources of uncertainty are *metalinguistic* in nature versus uncertainty that is tied to particular interpretations:

- Uncertainty about factivity
- Uncertainty about world knowledge (i.e., the prior probability of the complement clause being true)

The fundamental discreteness hypothesis says that uncertainty about factivity is metalinguistic.

The fundamental gradience hypothesis says that it is tied to particular interpretations.

Factivity and world knowledge

Our models of Degen et al.'s data vary whether two different sources of uncertainty are *metalinguistic* in nature versus uncertainty that is tied to particular interpretations:

- Uncertainty about factivity
- Uncertainty about world knowledge (i.e., the prior probability of the complement clause being true)

The fundamental discreteness hypothesis says that uncertainty about factivity is metalinguistic.

The fundamental gradience hypothesis says that it is tied to particular interpretations.

This gives us four models...

The discrete-factivity model

$$\texttt{discrete-factivity} : \mathtt{P}(\mathtt{P}\kappa)$$

$$\begin{aligned}\texttt{discrete-factivity} \;=\; & \langle \mathbf{v}, \mathbf{w} \rangle \sim \texttt{priors} \\ & \tau_{\mathbf{v}} \sim \texttt{Bernoulli}(\mathbf{v}) \\ & \boxed{\begin{aligned} & \tau_{\mathbf{w}} \sim \texttt{Bernoulli}(\mathbf{w}) \\ & \boxed{\tau_{\mathbf{v}} \vee \tau_{\mathbf{w}}} \end{aligned}}\end{aligned}$$

Important: The Bernoulli variable $\tau_{\mathbf{v}}$ associated determining whether or not a predicate is factive is sampled at the "outer" P.

The wholly-gradient model

$$\texttt{wholly-gradient} : \mathsf{P}(\mathsf{P}\kappa)$$

$$\texttt{wholly-gradient} = \langle \mathbf{v}, \mathbf{w} \rangle \sim \texttt{priors}$$

$$\boxed{\begin{array}{l} \tau_{\mathbf{v}} \sim \texttt{Bernoulli}(\mathbf{v}) \\ \tau_{\mathbf{w}} \sim \texttt{Bernoulli}(\mathbf{w}) \\ \boxed{\langle \tau_{\mathbf{v}}, \tau_{\mathbf{w}} \rangle} \end{array}}$$

Important: The Bernoulli variable $\tau_{\mathbf{v}}$ associated determining whether or not a predicate is factive is sampled at the "inner" P.

The discrete-world model

$$\texttt{discrete-world} : \mathsf{P}(\mathsf{P}\kappa)$$

$$\begin{aligned}\texttt{discrete-world} = \; & \langle \mathbf{v}, \mathbf{w} \rangle \sim \texttt{priors} \\ & \tau_{\mathbf{w}} \sim \texttt{Bernoulli}(\mathbf{w}) \\ & \boxed{\begin{aligned} & \tau_{\mathbf{v}} \sim \texttt{Bernoulli}(\mathbf{v}) \\ & \boxed{\langle \tau_{\mathbf{v}}, \tau_{\mathbf{w}} \rangle} \end{aligned}}\end{aligned}$$

The wholly-discrete model

$$\texttt{wholly-discrete} : \mathsf{P}(\mathsf{P}\kappa)$$

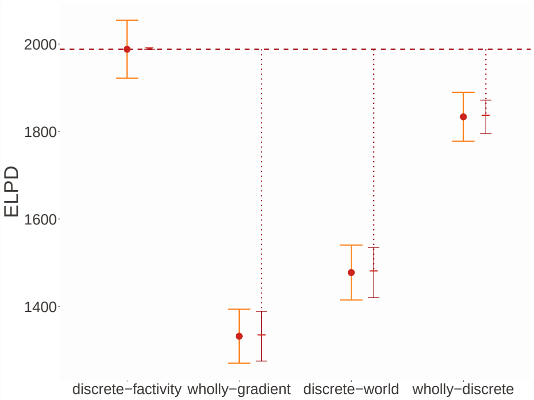
$$\begin{aligned}\texttt{wholly-discrete} = \; & \langle \mathbf{v}, \mathbf{w} \rangle \sim \texttt{priors} \\ & \tau_{\mathbf{v}} \sim \texttt{Bernoulli}(\mathbf{v}) \\ & \tau_{\mathbf{w}} \sim \texttt{Bernoulli}(\mathbf{w}) \\ & \boxed{\boxed{\langle \tau_{\mathbf{v}}, \tau_{\mathbf{w}} \rangle}}\end{aligned}$$

Comparisons

We compare the four models in terms of their expected log pointwise predictive densities (ELPDs) computed under the widely applicable information criterion (Gelman, Hwang, and Vehtari 2014; Watanabe 2013).

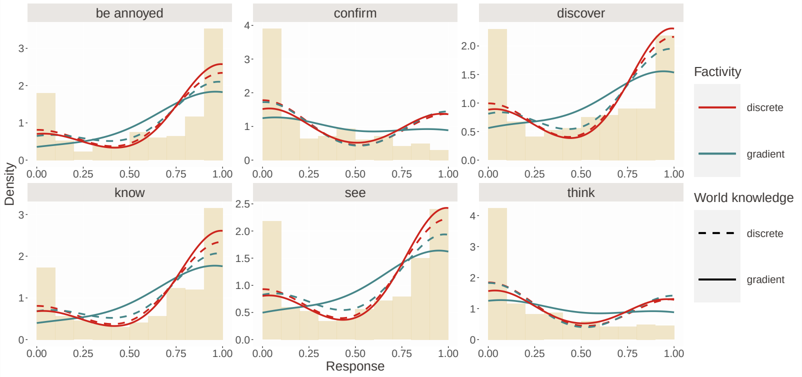
Comparisons

We compare the four models in terms of their expected log pointwise predictive densities (ELPDs) computed under the widely applicable information criterion (Gelman, Hwang, and Vehtari 2014; Watanabe 2013).



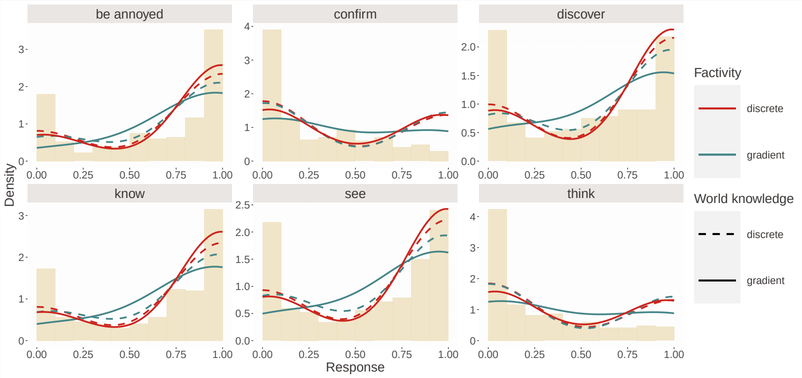
Why these differences?

Posterior predictive distributions, collapsing across all complement clauses and background facts:



Why these differences?

Posterior predictive distributions, collapsing across all complement clauses and background facts:



A few follow-ups

Granted, this investigation is *post hoc*. To draw firmer conclusions, we should test these models against newly collected data.

A few follow-ups

Granted, this investigation is *post hoc*. To draw firmer conclusions, we should test these models against newly collected data.

We do this in two ways:

A few follow-ups

Granted, this investigation is *post hoc*. To draw firmer conclusions, we should test these models against newly collected data.

We do this in two ways:

- Replicate Degen et al.'s experiment and use the posterior parameter distributions from our first set of models as priors in a new set of models.

A few follow-ups

Granted, this investigation is *post hoc*. To draw firmer conclusions, we should test these models against newly collected data.

We do this in two ways:

- Replicate Degen et al.'s experiment and use the posterior parameter distributions from our first set of models as priors in a new set of models.
- Modify the methodology slightly, so that complement clauses have minimal lexical content.

A few follow-ups

Granted, this investigation is *post hoc*. To draw firmer conclusions, we should test these models against newly collected data.

We do this in two ways:

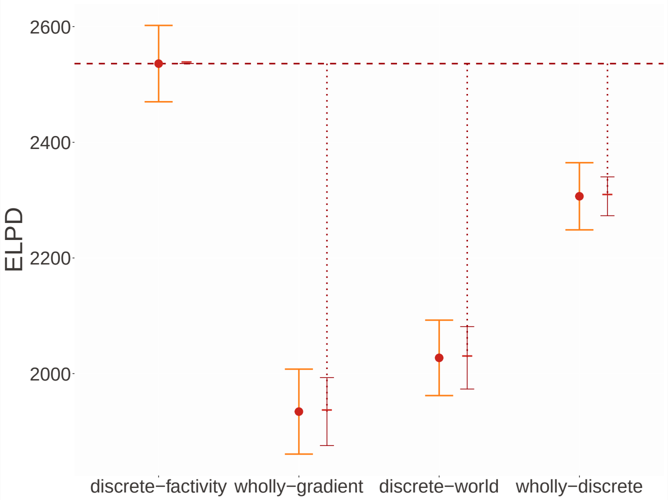
- Replicate Degen et al.'s experiment and use the posterior parameter distributions from our first set of models as priors in a new set of models.
- Modify the methodology slightly, so that complement clauses have minimal lexical content.
    - Helps get rid of the effect of world knowledge.

Our replication

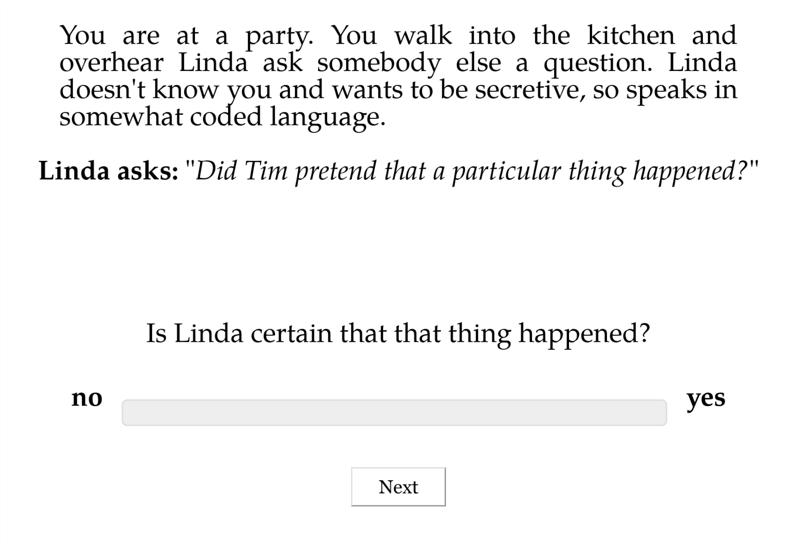
Given data from 288 new participants:

Our replication

Given data from 288 new participants:

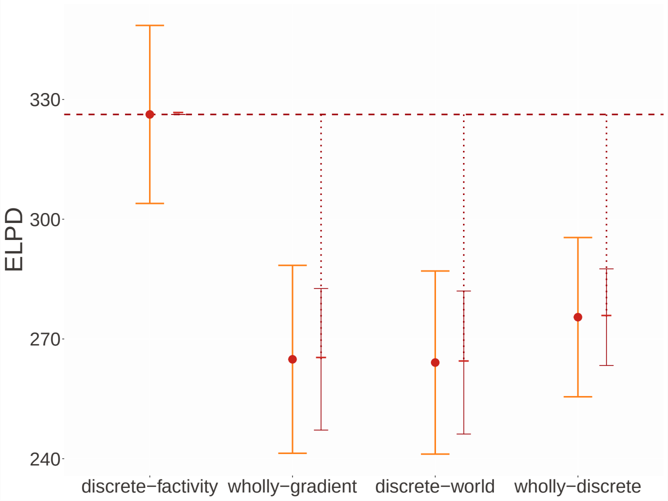


First modification: the bleached task



Bleached comparisons

Given data from 47 participants:



Second modification: the templatic task

You are at a party. You walk into the kitchen and overhear William ask somebody else a question. The party is very noisy, and you only hear part of what is said. The part you don't hear is represented by the 'X'.

William asks: "*Did Ray pretend that X happened?*"

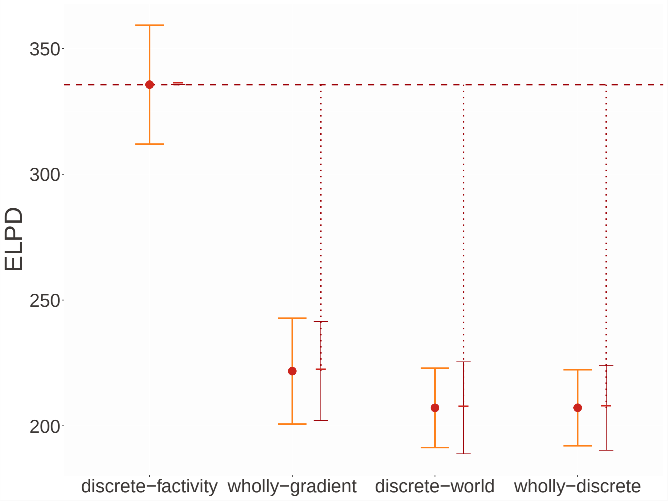
Is William certain that X happened?

no **yes**

Next

Templatic comparisons

Given data 49 participants:



Summing up

Conclusions

We find pretty firm evidence in support of the fundamental discreteness hypothesis.

Conclusions

We find pretty firm evidence in support of the fundamental discreteness hypothesis.

- Classical semantic accounts of the behavior of factive predicates (Kiparsky and Kiparsky 1970; Karttunen 1971, i.a.) can remain intact.

Conclusions

We find pretty firm evidence in support of the fundamental discreteness hypothesis.

- Classical semantic accounts of the behavior of factive predicates (Kiparsky and Kiparsky 1970; Karttunen 1971, i.a.) can remain intact.

Conclusions

We find pretty firm evidence in support of the fundamental discreteness hypothesis.

- Classical semantic accounts of the behavior of factive predicates (Kiparsky and Kiparsky 1970; Karttunen 1971, i.a.) can remain intact.

Broader point:

Conclusions

We find pretty firm evidence in support of the fundamental discreteness hypothesis.

- Classical semantic accounts of the behavior of factive predicates (Kiparsky and Kiparsky 1970; Karttunen 1971, i.a.) can remain intact.

Broader point:

- We can connect semantic theory to experimental data using the traditional semantic toolkit (i.e., typed λ-calculus).

Conclusions

We find pretty firm evidence in support of the fundamental discreteness hypothesis.

- Classical semantic accounts of the behavior of factive predicates (Kiparsky and Kiparsky 1970; Karttunen 1971, i.a.) can remain intact.

Broader point:

- We can connect semantic theory to experimental data using the traditional semantic toolkit (i.e., typed λ-calculus).
- But we have to integrate semantic analyses into theories of inference carefully… here, we choose to integrate them in a modular fashion, using monads.

Conclusions

We find pretty firm evidence in support of the fundamental discreteness hypothesis.

- Classical semantic accounts of the behavior of factive predicates (Kiparsky and Kiparsky 1970; Karttunen 1971, i.a.) can remain intact.

Broader point:

- We can connect semantic theory to experimental data using the traditional semantic toolkit (i.e., typed λ-calculus).
- But we have to integrate semantic analyses into theories of inference carefully… here, we choose to integrate them in a modular fashion, using monads.
- The strategy of using an already-available formal apparatus allows linking assumptions to be made explicit and testable.

References

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

Gelman, Andrew, Jessica Hwang, and Aki Vehtari (2014). **"Understanding predictive information criteria for Bayesian models".** en. In: *Statistics and Computing* 24.6, pp. 997–1016. DOI: 10.1007/s11222-013-9416-2.

Grove, Julian and Jean-Philippe Bernardy (2023). **"Probabilistic Compositional Semantics, Purely".** en. In: *New Frontiers in Artificial Intelligence*. Ed. by Katsutoshi Yada et al. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, pp. 242–256. DOI: 10.1007/978-3-031-36190-6_17.

Karttunen, Lauri (1971). **"Some observations on factivity".** In: *Paper in Linguistics* 4.1. Publisher: Routledge _eprint: https://doi.org/10.1080/08351817109370248, pp. 55–69. DOI: 10.1080/08351817109370248.

Kiparsky, Paul and Carol Kiparsky (1970). **"FACT".** en. In: *Progress in Linguistics*. De Gruyter Mouton, pp. 143–173.

Tonhauser, Judith, David I. Beaver, and Judith Degen (2018). **"How Projective is Projective Content? Gradience in Projectivity and At-issueness".** en. In: *Journal of Semantics* 35.3, pp. 495–542. DOI: 10.1093/jos/ffy007.

Watanabe, Sumio (2013). **"A Widely Applicable Bayesian Information Criterion".** In: *Journal of Machine Learning Research* 14.27, pp. 867–897.

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.