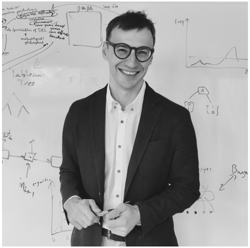
Investigation of semantic representations of quantifiers with the Diffusion Decision Model

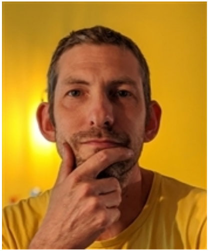
Fabian Schlotterbeck[1]
Sonia Ramotowska[2]

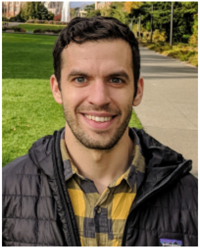
University of Tübingen[1], Heinrich Heine University Düsseldorf[2]

Procedural and computational models of semantic and pragmatic processes
August 1, 2023

Collaborators



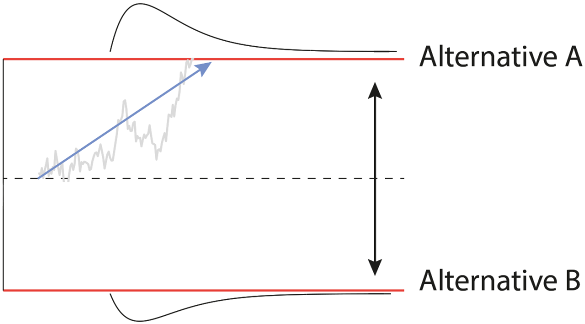




Diffusion decision model

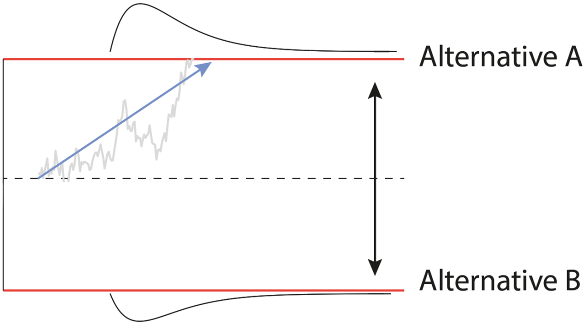
Diffusion decision model (DDM)

- The DDM is a **sequential sampling model** of decision making. (Ratcliff, 1978)



Diffusion decision model (DDM)

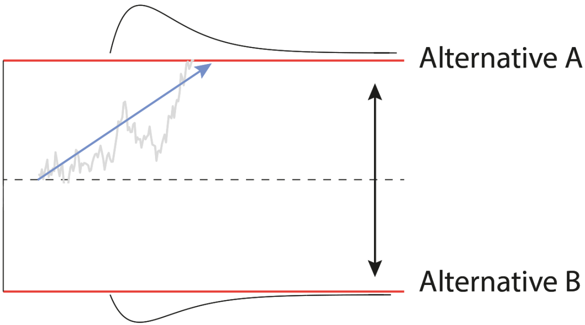
- The DDM is a **sequential sampling model** of decision making. (Ratcliff, 1978)



- Decision processes, such as truth-value judgments, are described as the **accumulation of a noisy signal** over time **until a decision boundary** is reached and a response is initiated.

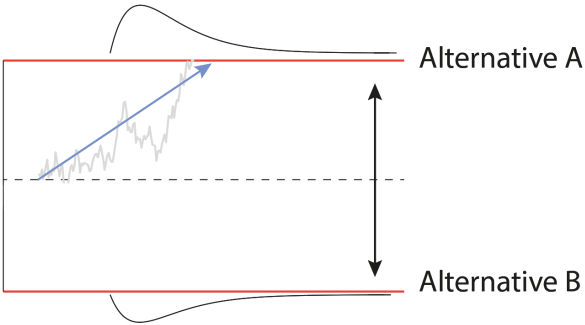
DDM explains common observations

- Applied successfully to a large variety of decision tasks



DDM explains common observations

- Applied successfully to a large variety of decision tasks



- Explains a range of observations:
  - Accuracy and RT distributions of both response alternatives
  - **Speed-accuracy trade-off**
  - Effects of instructions (speed vs accuracy)
  - ...

Sequential sampling models

Theoretical motivation

- **Accuracy and RT can be modeled jointly** if we generalize ideas from Signal Detection Theory (see e.g. Bogacz, 2006).

Optimal response strategy

$$Z_2 < \frac{P(R_1|x_n \qquad\qquad)}{P(R_2|x_n \qquad\qquad)} = \frac{P(x_n|R_1) \qquad\qquad\qquad\qquad P(R_1)}{P(x_n|R_2) \qquad\qquad\qquad\qquad P(R_2)} < Z_1$$

Sequential sampling models

Theoretical motivation

- **Accuracy and RT can be modeled jointly** if we generalize ideas from Signal Detection Theory (see e.g. Bogacz, 2006).

Optimal response strategy

1. Continuously draw samples x_i from an internal representation, as long as:

$$Z_2 < \frac{P(R_1|x_n \ldots x_1)}{P(R_2|x_n \ldots x_1)} = \frac{P(x_n|R_1) \ldots P(x_1|R_1)P(R_1)}{P(x_n|R_2) \ldots P(x_1|R_2)P(R_2)} < Z_1$$

Sequential sampling models

Theoretical motivation

- **Accuracy and RT can be modeled jointly** if we generalize ideas from Signal Detection Theory (see e.g. Bogacz, 2006).

Optimal response strategy

1. Continuously draw samples x_i from an internal representation, as long as:

$$Z_2 < \frac{P(R_1|x_n \dots x_1)}{P(R_2|x_n \dots x_1)} = \frac{P(x_n|R_1)\dots P(x_1|R_1)P(R_1)}{P(x_n|R_2)\dots P(x_1|R_2)P(R_2)} < Z_1$$

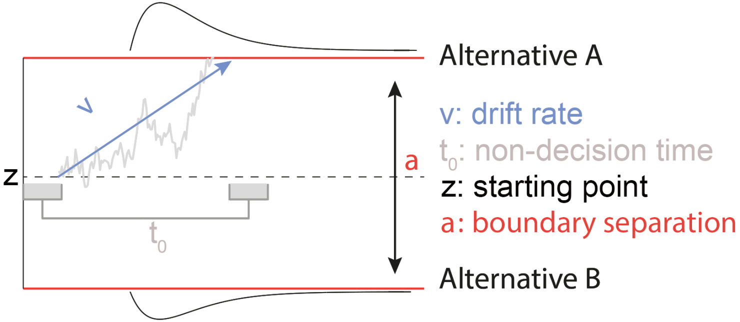
$$\log Z_2 < \underbrace{\log \frac{P(R_1|x_n \dots x_1)}{P(R_2|x_n \dots x_1)}}_{\text{posterior log odds}} = \sum \log \frac{P(x_i|R_1)}{P(x_i|R_2)} + \log \frac{P(R_1)}{P(R_2)} < \log Z_1$$

2. Choose R_1 or R_2 if posterior log odds reach Z_1 or Z_2, resp.

- This stochastic process defines the DDM.

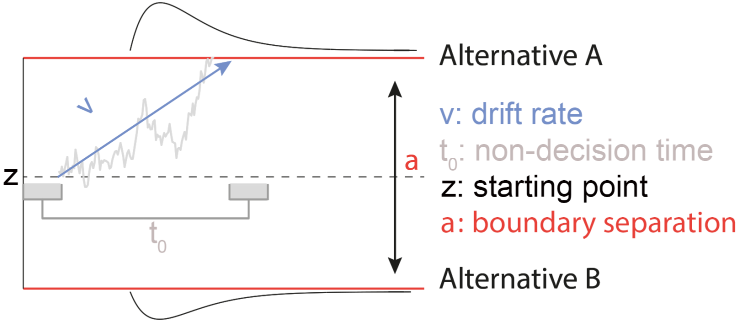
Components of the DDM

- **Model parameters represent processing components**:



Components of the DDM

- **Model parameters represent processing components**:

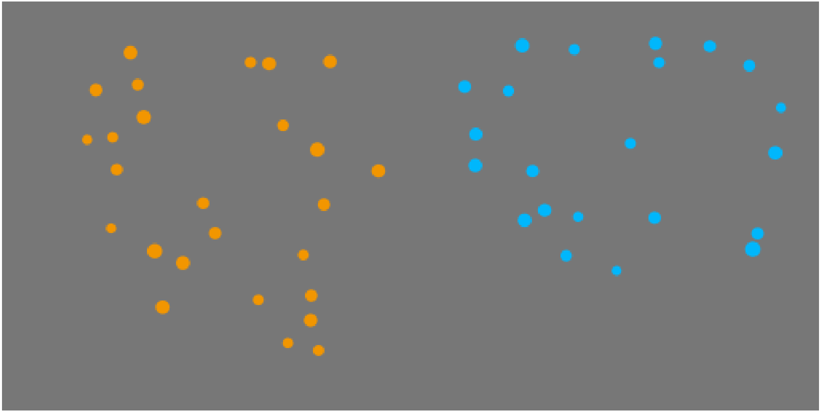


- They **can be estimated from data**.

  In this way, the diffusion model can provide a meeting point between a model for stimulus encoding and representation and decision processes. (Ratcliff & McKoon, 2008)

Application to number comparison and quantifier verification

Number comparison



Which set has more dots?

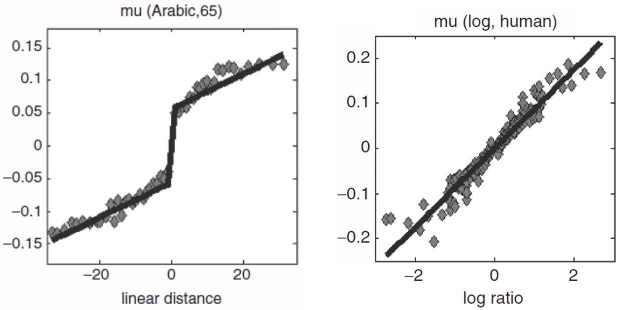
orange *blue*

Application of DDM to number comparison

- **Good model fit** to number comparison tasks

Application of DDM to number comparison

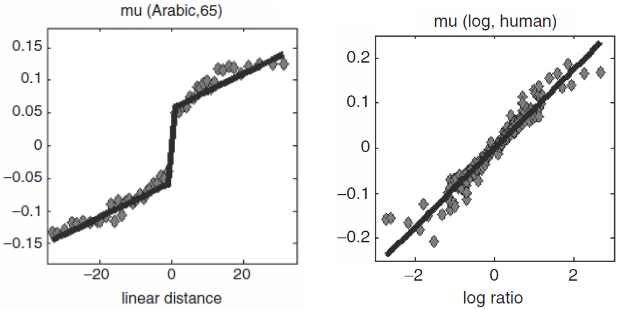
- **Good model fit** to number comparison tasks
- **Monotonic relation** between numerical distance and drift rate: **step-like for exact**, and **linear for approximate** comparison.



Application of DDM to number comparison

- **Good model fit** to number comparison tasks
- **Monotonic relation** between numerical distance and drift rate: **step-like for exact**, and **linear for approximate** comparison.

⇒ Supports common assumptions about representations of exact and approximate number (Dehaene, 2007)



Application of DDM to quantifier verification

- **Verification of quantifiers often boils down to number comparison**
- Famous example:

  $[\![\text{Most of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \mid A \cap \neg B \mid)$

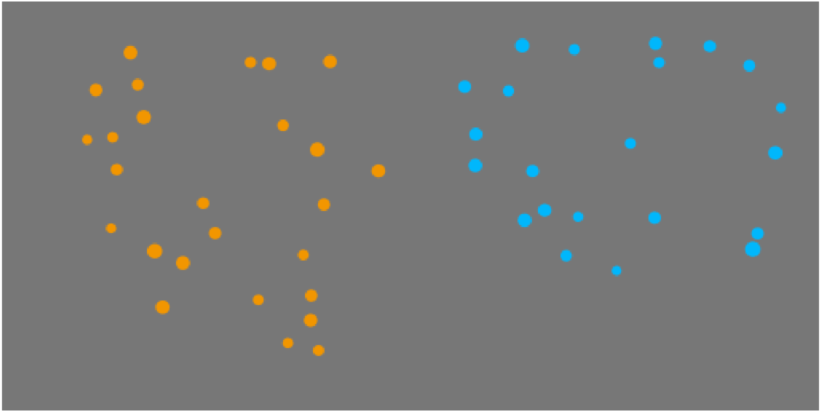
Application of DDM to quantifier verification

- **Verification of quantifiers often boils down to number comparison**
- Famous example:

  $[\![\text{Most of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \mid A \cap \neg B \mid)$

  $\Rightarrow$ Comparison of two cardinalities

Quantifier verification



Most of the dots are blue.

true?

no *yes*

Advantages of DDM analysis

- Previous studies combined SDT with characteristics of number comparison to test semantic representations of quantifiers like *most* (see e.g. Pietroski et al., 2009; Lidz et al. 2011).

Advantages of DDM analysis

- Previous studies combined SDT with characteristics of number comparison to test semantic representations of quantifiers like *most* (see e.g. Pietroski et al., 2009; Lidz et al. 2011).
- In these studies, conclusions were based on the amount of errors people make in verification tasks.

Advantages of DDM analysis

- Previous studies combined SDT with characteristics of number comparison to test semantic representations of quantifiers like *most* (see e.g. Pietroski et al., 2009; Lidz et al. 2011).
- In these studies, conclusions were based on the amount of errors people make in verification tasks.
- DDM extracts more information because errors and RT are modeled jointly.

Advantages of DDM analysis

- Previous studies combined SDT with characteristics of number comparison to test semantic representations of quantifiers like *most* (see e.g. Pietroski et al., 2009; Lidz et al. 2011).
- In these studies, conclusions were based on the amount of errors people make in verification tasks.
- DDM extracts more information because errors and RT are modeled jointly.

$\Rightarrow$ **Can we use DDM to study representations underlying quantifier verification?**

General methodology

How can we apply DDM to study quantifier representations?

1. Start with an hypothesis regarding DDM parameters

How can we apply DDM to study quantifier representations?

1. Start with an hypothesis regarding DDM parameters
2. Collect quantifier verification data (RT and accuracy)

How can we apply DDM to study quantifier representations?

1. Start with an hypothesis regarding DDM parameters
2. Collect quantifier verification data (RT and accuracy)
3. Fit model to data

How can we apply DDM to study quantifier representations?

1. Start with an hypothesis regarding DDM parameters
2. Collect quantifier verification data (RT and accuracy)
3. Fit model to data
4. Constrain parameters

How can we apply DDM to study quantifier representations?

1. Start with an hypothesis regarding DDM parameters
2. Collect quantifier verification data (RT and accuracy)
3. Fit model to data
4. Constrain parameters
5. Check model fit

How can we apply DDM to study quantifier representations?

1. Start with an hypothesis regarding DDM parameters
2. Collect quantifier verification data (RT and accuracy)
3. Fit model to data
4. Constrain parameters
5. Check model fit
6. If model fit is good, perform hypothesis tests

How can we apply DDM to study quantifier representations?

1. Start with an hypothesis regarding DDM parameters
2. Collect quantifier verification data (RT and accuracy)
3. **Fit model to data**
4. **Constrain parameters**
5. **Check model fit**
6. **If model fit is good, perform hypothesis tests**

Step 3: Fit model to data

- Model can be fitted using. . .

Step 3: Fit model to data

- Model can be fitted using...
  - **maximum likelihood**...
  - **Bayesian inference**...
- on the basis of...

Step 3: Fit model to data

- Model can be fitted using...
  - **maximum likelihood**...
  - **Bayesian inference**...
- on the basis of...
  - **individual participants**.
  - **the entire set of participants in a hierarchical manner** (can be useful with quasi complete separation).

Step 3: Fit model to data

- Model can be fitted using...
  - **maximum likelihood**...
  - **Bayesian inference**...
- on the basis of...
  - **individual participants**.
  - **the entire set of participants in a hierarchical manner** (can be useful with quasi complete separation).
- Depending on which one we choose, the following steps differ slightly, although not substantially.

Step 3: Fit model to data

- Model can be fitted using...
  - **maximum likelihood**...
  - **Bayesian inference**...
- on the basis of...
  - **individual participants**.
  - **the entire set of participants in a hierarchical manner** (can be useful with quasi complete separation).
- Depending on which one we choose, the following steps differ slightly, although not substantially.

Step 4: Constrain parameters

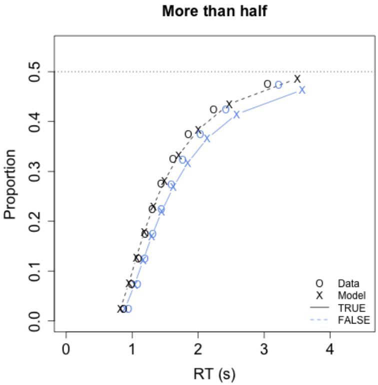
- **Parameters that are not directly relevant to the hypothesis** under investigation **can be constrained** to be equal across conditions.

Step 4: Constrain parameters

- **Parameters that are not directly relevant to the hypothesis** under investigation **can be constrained** to be equal across conditions.
- We do this **only if model fit is not affected negatively**, as determined by model comparisons.

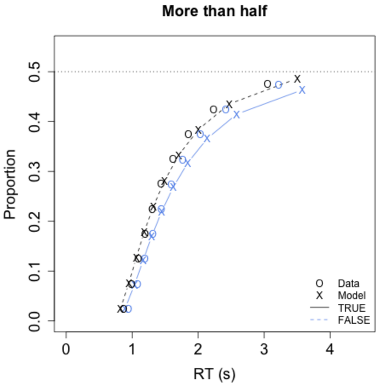
Step 5: Check posterior predictive distribution

- Model fit can be visualized using *Defective Cumulative Density* plots of experimental data and the fitted model
- For example:



Step 5: Check posterior predictive distribution

- Model fit can be visualized using *Defective Cumulative Density* plots of experimental data and the fitted model
- For example:



- CDFs show the joint distribution of RT and responses in the data and the fitted model.

Assumption about drift rates in quantifier verification

Drift rates determined by numerical distance

- As in number comparison, judgments in quantifier verification (e.g. *more than half*) depend on the distance between two numbers (e.g. $\mid A \cap B \mid$ and $\frac{|A|}{2}$)

$$[\![\text{More than half of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \frac{|A|}{2})$$

Drift rates determined by numerical distance

- As in number comparison, judgments in quantifier verification (e.g. *more than half*) depend on the distance between two numbers (e.g. $\mid A \cap B \mid$ and $\frac{|A|}{2}$)

$$[\![\text{More than half of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \frac{|A|}{2})$$

- In the DDM, this is reflected in drift rates
  - Positive drift for *true* vs. negative drift for *false* (or vice versa)
  - Larger drift rates for larger absolute distance

$\Rightarrow$ We assume a **monotonic relationship** between drift rates and numerical distance.

Drift rates determined by numerical distance

- As in number comparison, judgments in quantifier verification (e.g. *more than half*) depend on the distance between two numbers (e.g. $|\ A \cap B\ |$ and $\frac{|A|}{2}$)

$$[\![\text{More than half of the As are B}]\!] = 1 \leftrightarrow (|\ A \cap B\ | > \frac{|A|}{2})$$

- In the DDM, this is reflected in drift rates
  - Positive drift for *true* vs. negative drift for *false* (or vice versa)
  - Larger drift rates for larger absolute distance

$\Rightarrow$ We assume a **monotonic relationship** between drift rates and numerical distance.

- We specified this relation using a **generalized logistic regression function**.

Generalized logistic function

$$v(p) = V_{\mathrm{l}} + \frac{V_{\mathrm{u}} - V_{\mathrm{l}}}{1 + e^{-s(p - p_0)}}$$

$\Rightarrow$ Introduces additional parameters (V_u, V_l, s and p_0) but can cover multiple conditions and thus actually decrease the number of parameters

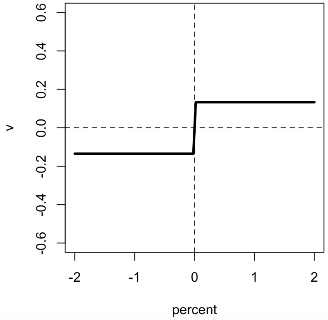


Figure: Example of logistic regression function fitted to one participant for *more than half* (0 is a normalized percentage and stands for 50%).

Three case studies

Case study 1

Monotonicity and processing load

Representational complexity and pragmatics cause the monotonicity effect

Fabian Schlotterbeck* (fabian.schlotterbeck@uni-tuebingen.de)
University of Tübingen

Sonia Ramotowska* (s.ramotowska@uva.nl)
University of Amsterdam

Leendert van Maanen (lvmaanen@gmail.com)
Utrecht University

Jakub Szymanik (jakub.szymanik@gmail.com)
University of Amsterdam

Monotonicity: Entailment relations licensed by quantifiers

(1) More than half of the audience attended all talks.
$\Longrightarrow$ More than half of the audience attended an invited talk.
$\not\Longrightarrow$ More than half of the audience attended all talks and dinner.

Monotonicity: Entailment relations licensed by quantifiers

(1) More than half of the audience attended all talks.
$\Longrightarrow$ More than half of the audience attended an invited talk.
$\not\Longrightarrow$ More than half of the audience attended all talks and dinner.

- **Upward entailing** (UE) quantifiers (e.g. *more than half*) license inferences from subsets to supersets.

Monotonicity: Entailment relations licensed by quantifiers

(1) More than half of the audience attended all talks.
$\Longrightarrow$ More than half of the audience attended an invited talk.
$\not\Longrightarrow$ More than half of the audience attended all talks and dinner.

(2) Fewer than half of the audience attended all talks.
$\not\Longrightarrow$ Fewer than half of the audience attended an invited talk.
$\Longrightarrow$ Fewer than half of the audience attended all talks and dinner.

- **Upward entailing** (UE) quantifiers (e.g. *more than half*) license inferences from subsets to supersets.

Monotonicity: Entailment relations licensed by quantifiers

(1) More than half of the audience attended all talks.
$\Longrightarrow$ More than half of the audience attended an invited talk.
$\not\Longrightarrow$ More than half of the audience attended all talks and dinner.

(2) Fewer than half of the audience attended all talks.
$\not\Longrightarrow$ Fewer than half of the audience attended an invited talk.
$\Longrightarrow$ Fewer than half of the audience attended all talks and dinner.

- **Upward entailing** (UE) quantifiers (e.g. *more than half*) license inferences from subsets to supersets.
- **Downward entailing** (DE) quantifiers (e.g. *fewer than half*) license inferences from supersets to subsets.

Monotonicity effect

- **Psycholinguistic studies** showed that **DE quantifiers are more difficult to process** than UE ones (e.g. Just & Carpenter, 1971; Geurts & Van der Slik, 2005; Urbach & Kutas 2010).

Monotonicity effect

- **Psycholinguistic studies** showed that **DE quantifiers are more difficult to process** than UE ones (e.g. Just & Carpenter, 1971; Geurts & Van der Slik, 2005; Urbach & Kutas 2010).
- **This monotonicity effect** is found in various tasks, including **verification tasks.**

Monotonicity effect

- **Psycholinguistic studies** showed that **DE quantifiers are more difficult to process** than UE ones (e.g. Just & Carpenter, 1971; Geurts & Van der Slik, 2005; Urbach & Kutas 2010).
- **This monotonicity effect** is found in various tasks, including **verification tasks.**
- Cognitive processes behind the monotonicity effect are still debated.

Two classes of models explaining the monotonicity effect

- Main question: **Which cognitive processes cause the monotonicity effect?**

Two classes of models explaining the monotonicity effect

- Main question: **Which cognitive processes cause the monotonicity effect?**
- We **apply the DDM to data from verification experiments** to answer the following questions:

Two classes of models explaining the monotonicity effect

- Main question: **Which cognitive processes cause the monotonicity effect?**
- We **apply the DDM to data from verification experiments** to answer the following questions:
  1. Is DDM **applicable to quantifier verification**?

Two classes of models explaining the monotonicity effect

- Main question: **Which cognitive processes cause the monotonicity effect?**
- We **apply the DDM to data from verification experiments** to answer the following questions:
  1. Is DDM **applicable to quantifier verification**?
  2. **Which parameters are affected** by monotonicity?

Two classes of models explaining the monotonicity effect

- Main question: **Which cognitive processes cause the monotonicity effect?**
- We **apply the DDM to data from verification experiments** to answer the following questions:
  1. Is DDM **applicable to quantifier verification**?
  2. **Which parameters are affected** by monotonicity?
- **Predictions** regarding 2 were **derived from two broad classes of accounts:**.

Two classes of models explaining the monotonicity effect

- Main question: **Which cognitive processes cause the monotonicity effect?**
- We **apply the DDM to data from verification experiments** to answer the following questions:
  1. Is DDM **applicable to quantifier verification**?
  2. **Which parameters are affected** by monotonicity?

- **Predictions** regarding 2 were **derived from two broad classes of accounts:**.

Two step models

An **extra step** in verification of DE quantifiers that is due to their representational complexity leads to delay.

Two classes of models explaining the monotonicity effect

- Main question: **Which cognitive processes cause the monotonicity effect?**
- We **apply the DDM to data from verification experiments** to answer the following questions:
  1. Is DDM **applicable to quantifier verification**?
  2. **Which parameters are affected** by monotonicity?

- **Predictions** regarding 2 were **derived from two broad classes of accounts:**.

Two step models

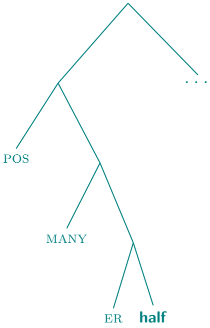
An **extra step** in verification of DE quantifiers that is due to their representational complexity leads to delay.

Pragmatic processing models

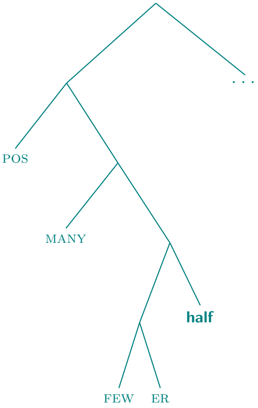
Contextual fit or pragmatic felicity determines processing difficulty.

Representational “complexity” of comparative quantifiers

more than half



fewer than half



- **Additional operation** in DE case (antonym operator FEW) **may correspond to an extra processing step**.

Two step models

An extra step in the verification of DE quantifiers

- **Additional processing step** in the verification of DE vs. UE quantifiers.

Two step models

An extra step in the verification of DE quantifiers

- **Additional processing step** in the verification of DE vs. UE quantifiers.
- This additional step is due to **more complex semantic representations** (e.g. covert negation).

Two step models

An extra step in the verification of DE quantifiers

- **Additional processing step** in the verification of DE vs. UE quantifiers.
- This additional step is due to **more complex semantic representations** (e.g. covert negation).

Prediction

Monotonicity affects **non-decision time**

Two step models

An extra step in the verification of DE quantifiers

- **Additional processing step** in the verification of DE vs. UE quantifiers.
- This additional step is due to **more complex semantic representations** (e.g. covert negation).

Prediction

Monotonicity affects **non-decision time**

Reason

Truth evaluation of DE quantifiers involves an extra step in addition to the actual verification step.

Pragmatic models

- **DE** quantifiers cause processing difficulties because they are **systematically dispreferred** to suitable UE alternatives.

Pragmatic models

- **DE** quantifiers cause processing difficulties because they are **systematically dispreferred** to suitable UE alternatives.
- This is **due to violation of pragmatic principles**, such as **avoidance of infrequent or uninformative utterances**, in various contexts (cf. Nieuwland, 2017; Nieuwland & Kuperberg 2008; Nordmeyer & Frank, 2014).

Pragmatic models

- **DE** quantifiers cause processing difficulties because they are **systematically dispreferred** to suitable UE alternatives.
- This is **due to violation of pragmatic principles**, such as **avoidance of infrequent or uninformative utterances**, in various contexts (cf. Nieuwland, 2017; Nieuwland & Kuperberg 2008; Nordmeyer & Frank, 2014).
- **Verification treated as production**: Participants judge whether they would utter the sentence to describe the context (e.g. Degen & Goodman, 2014; Waldon & Degen, 2020).

Pragmatic models

- **DE** quantifiers cause processing difficulties because they are **systematically dispreferred** to suitable UE alternatives.
- This is **due to violation of pragmatic principles**, such as **avoidance of infrequent or uninformative utterances**, in various contexts (cf. Nieuwland, 2017; Nieuwland & Kuperberg 2008; Nordmeyer & Frank, 2014).
- **Verification treated as production**: Participants judge whether they would utter the sentence to describe the context (e.g. Degen & Goodman, 2014; Waldon & Degen, 2020).

Prediction

Monotonicity affects **drift rate**

Pragmatic models

- **DE** quantifiers cause processing difficulties because they are **systematically dispreferred** to suitable UE alternatives.
- This is **due to violation of pragmatic principles**, such as **avoidance of infrequent or uninformative utterances**, in various contexts (cf. Nieuwland, 2017; Nieuwland & Kuperberg 2008; Nordmeyer & Frank, 2014).
- **Verification treated as production**: Participants judge whether they would utter the sentence to describe the context (e.g. Degen & Goodman, 2014; Waldon & Degen, 2020).

Prediction

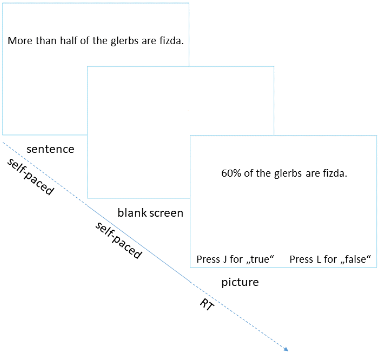
Monotonicity affects **drift rate**

Reason

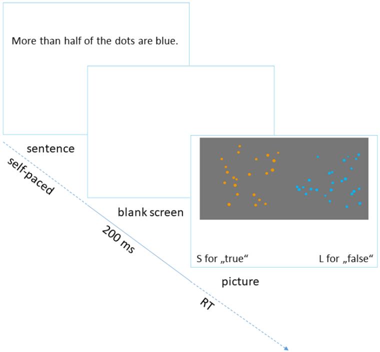
Empirical results in related domains (e.g. lexical selection in picture naming tasks; Anders et al., 2015, 2019) and theoretical considerations (Bogacz et al. 2006, Bitzer et al., 2014)

Why two experiments?

Linguistic task

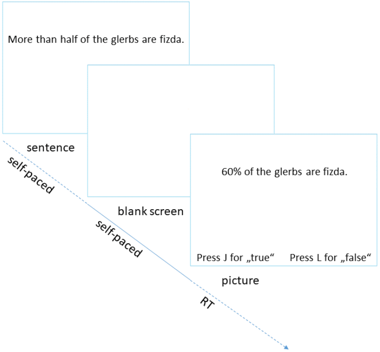


Visual task

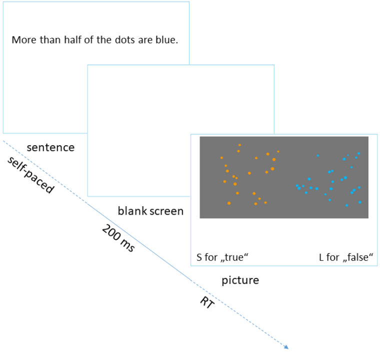


Why two experiments?

Linguistic task



Visual task



Do results generalize across the processing of precise and approximate numbers (e.g. Dehaene, 2007)?

Experimental design

- **Two web-based experiments**

Experimental design

- **Two web-based experiments**
- **Visual task** (N=56, 240 trials per quantifier):
  - Two-factorial within-participants design
  - MONOTONICITY (*more* vs. *fewer than half*)
  - RATIO of the colored dots (28:20, 26:22, 22:26 and 20:28)
- **Linguistic task** (N=72, 50 trials per quantifier):
  - MONOTONICITY (*more* vs. *fewer than half*)
  - PROPORTION (randomly drawn from 1-99%, excluding 50%)

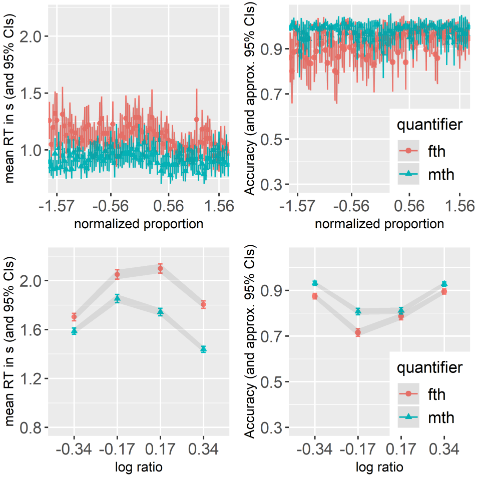
Same procedure

1. Start with an hypothesis regarding DDM parameters $\surd$
2. Collect quantifier verification data (RT and accuracy) $\surd$
3. Fit model to data $\surd$
4. Constrain parameters $\surd$
5. Check model fit $\surd$
6. If model fit is good, perform hypothesis tests...

Results

Replication of monotonicity effect

MONOTONICITY effect on RT and accuracy in both tasks



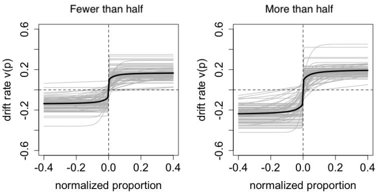
Linguistic vs. visual task

- **Expected differences** between linguistic and visual task

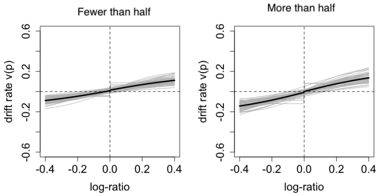
Linguistic vs. visual task

- **Expected differences** between linguistic and visual task
- Reflected in the drift rate:

Linguistic task



Visual task



Pragmatic processing models hypothesis test

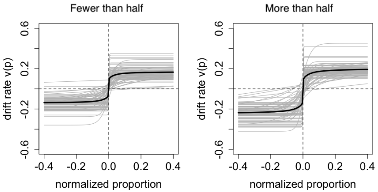
Pragmatic processing models predict **difference in drift rate**.

Pragmatic processing models hypothesis test

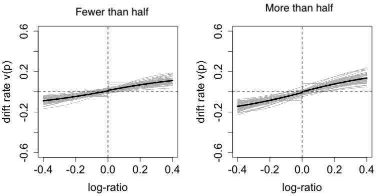
Pragmatic processing models predict **difference in drift rate**.

- **Drift rates higher** for UE than DE
  Linguistic: $t(71) = 9.10; p < .001;$
  Visual: $t(55) = 8.46; p < .001.$

Linguistic task



Visual task



Two-step models hypothesis test

Two-step models predict **difference in non-decision time**.

Two-step models hypothesis test

Two-step models predict **difference in non-decision time**.

- **Non-decision time longer** for DE than UE
  Linguistic: 34 ms, $t(71) = 5.53$; $p < .001$;
  Visual: 43 ms, $t(55) = 5.74$; $p < .001$

Interim Discussion

Summary and implications

- The **same pattern of results across tasks**: Both **non-decision times and drift rates** were affected by the monotonicity manipulation.

Interim Discussion

Summary and implications

- The **same pattern of results across tasks**: Both **non-decision times and drift rates** were affected by the monotonicity manipulation.
- Our **modeling results support both two-step and pragmatic processing models**!

Interim Discussion

Summary and implications

- The **same pattern of results across tasks**: Both **non-decision times and drift rates** were affected by the monotonicity manipulation.
- Our **modeling results support both two-step and pragmatic processing models**!
- More generally, they indicate **two separate sources of the monotonicity effect** that map onto different DDM parameters.

Interim Discussion

Monotonicity vs. polarity

- To pass the exam, you should make **few** mistakes.
- #To pass the exam, you should make **a small number** of mistakes.

Interim Discussion

Monotonicity vs. polarity

- To pass the exam, you should make **few** mistakes.
- #To pass the exam, you should make **a small number** of mistakes.
- **A small number** is not consistent with zero (Agmon et al. (2019) example)

Interim Discussion

How to disentangle the two sources of difficulty

- Our findings are consistent with recent findings of Agmon et al. (2019), who compared "quantifiers" to "adjectives" and found larger RT differences in cases like (3) vs. (4).

(3) a. More than half...
b. Fewer than half...

(4) a. A large proportion of the dots are blue.
b. A small proportion of the dots are blue.

Interim Discussion

How to disentangle the two sources of difficulty

- Our findings are consistent with recent findings of Agmon et al. (2019), who compared "quantifiers" to "adjectives" and found larger RT differences in cases like (3) vs. (4).

(3) a. More than half...
b. Fewer than half...

(4) a. A large proportion of the dots are blue.
b. A small proportion of the dots are blue.

- Both types of expressions differ in polarity, but only the quantifiers in (3) also differ in monotonicity.

$\Rightarrow$ Negative polarity and downward monotonicity may be two separate sources of the observed effects

Interim Discussion

How to disentangle the two sources of difficulty

- Our findings are consistent with recent findings of Agmon et al. (2019), who compared "quantifiers" to "adjectives" and found larger RT differences in cases like (3) vs. (4).

(3) a. More than half...
b. Fewer than half...

(4) a. A large proportion of the dots are blue.
b. A small proportion of the dots are blue.

- Both types of expressions differ in polarity, but only the quantifiers in (3) also differ in monotonicity.

$\Rightarrow$ Negative polarity and downward monotonicity may be two separate sources of the observed effects

$\Rightarrow$ Compare "quantifiers" to "adjectives"

Visual task with “adjectives” (work in progress)

Experimental design

- Identical to first visual experiment but with **different linguistic stimuli**

Experimental design

- Identical to first visual experiment but with **different linguistic stimuli**
- **Visual task with adjectives** (N=68, 240 trials per quantifier):
  - Two-factorial within-participants design
  - **Polarity** (**a large** vs. **a small proportion**)
  - RATIO of the colored dots (28:20, 26:22, 22:26 and 20:28)

Prediction

If observed effects in drift rates and non-decision time for *fewer* vs. *more than half* **reflect two separate sources of difficulty**

Prediction

If observed effects in drift rates and non-decision time for *fewer* vs. *more than half* **reflect two separate sources of difficulty** and each of them is either due to MONOTONICITY or to POLARITY,

Prediction

If observed effects in drift rates and non-decision time for *fewer* vs. *more than half* **reflect two separate sources of difficulty** and each of them is either due to MONOTONICITY or to POLARITY, an **effect in only one parameter is expected for** *a large* vs. *a small proportion* because these expressions differ only in POLARITY.

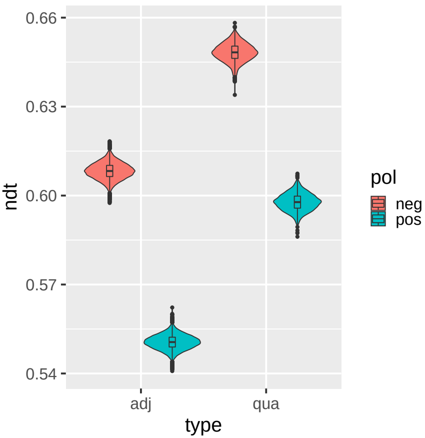
Same procedure, almost

1. Start with an hypothesis regarding DDM parameters $\surd$
2. Collect quantifier verification data (RT and accuracy) $\surd$
3. Fit model to data $\surd$
4. Constrain parameters $\surd$
5. Check posterior predictive distribution $\surd$
6. If model fit is good, perform hypothesis tests. . .

Parameter estimates: “quantifiers” vs. “adjectives”

Main effects in non-decision time (NDT)

- **Longer NDT** for...

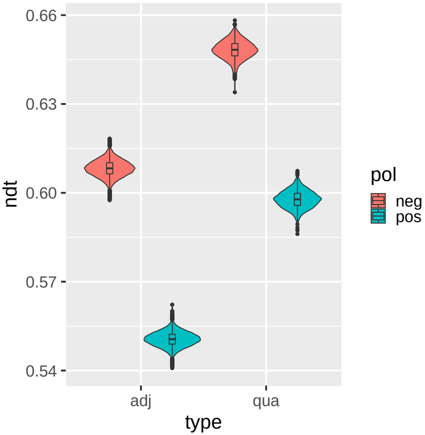


- negative vs. positive expressions ($\sim$ 50 ms)

Parameter estimates: “quantifiers” vs. “adjectives”

Main effects in non-decision time (NDT)

- **Longer NDT** for...

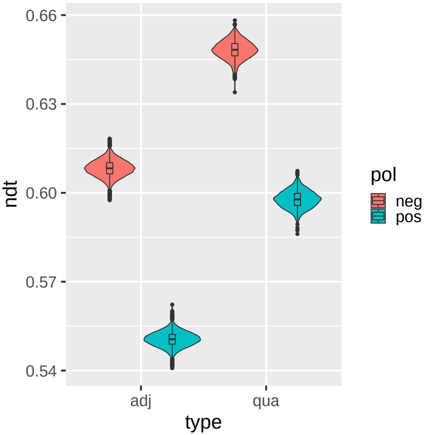


- negative vs. positive expressions ($\sim$ 50 ms)
- “quantifiers” vs. “adjectives” ($\sim$ 50 ms)

Parameter estimates: “quantifiers” vs. “adjectives”

Main effects in non-decision time (NDT)

- **Longer NDT** for...



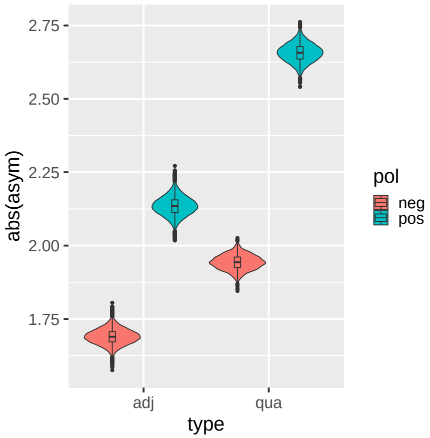
- negative vs. positive expressions ($\sim$ 50 ms)
- “quantifiers” vs. “adjectives” ($\sim$ 50 ms)
- matches representational complexity

$$
\begin{aligned}
\mathit{lprop}(X, Y) &:= \textsc{pos}(\textsc{size}(\textsc{prop}(\mathrm{X},\mathrm{Y})))\\
\mathit{sprop}(X, Y) &:= \textsc{pos}(\textsc{antonym}(\textsc{size}(\textsc{prop}(\mathrm{X},\mathrm{Y}))))\\
\mathit{mth}(X, Y) &:= \textsc{pos}(\textsc{er}(\textsc{many}(\mathrm{X},\mathrm{Y}),\ \textsc{half}(\mathrm{X})))\\
\mathit{fth}(X, Y) &:= \textsc{pos}(\textsc{antonym}(\textsc{er}(\textsc{many}(\mathrm{X},\mathrm{Y}),\ \textsc{half}(\mathrm{X}))))
\end{aligned}
$$

Parameter estimates: "quantifiers" vs. "adjectives"

Interaction in drift rate

- Larger **drift rates** for...

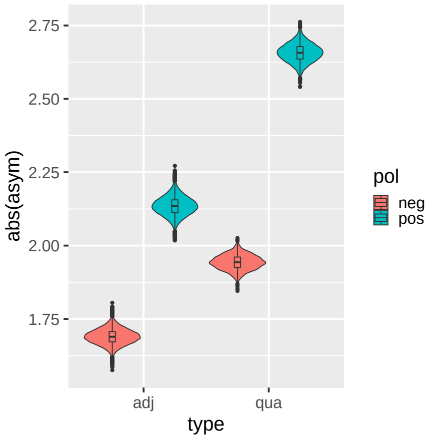


- positive vs. negative expressions

Parameter estimates: “quantifiers” vs. “adjectives”

Interaction in drift rate

- Larger **drift rates** for...

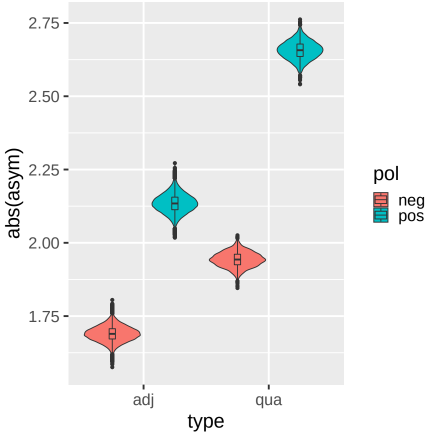


- positive vs. negative expressions
- “quantifiers” vs. “adjectives”

Parameter estimates: “quantifiers” vs. “adjectives”

Interaction in drift rate

- Larger **drift rates** for...



- positive vs. negative expressions
- “quantifiers” vs. “adjectives”
- Additional effect of MONOTONICITY for “quantifiers” vs. “adjectives”

Discussion

Discussion

Summary and implications

- Results **consistent effects across tasks**

$\Rightarrow$ Indicates linguistic character of effects

Discussion

Summary and implications

- Results **consistent effects across tasks**

⇒ Indicates linguistic character of effects

- Across expression types, both **drift rates and non-decision times showed effects** of POLARITY

⇒ Challenges simple mapping between semantic properties and effects in DDM parameters.

Discussion

Summary and implications

- Results **consistent effects across tasks**

$\Rightarrow$ Indicates linguistic character of effects

- Across expression types, both **drift rates and non-decision times showed effects** of POLARITY

$\Rightarrow$ Challenges simple mapping between semantic properties and effects in DDM parameters.

- **Interaction in drift rate** parameters indicates extra penalty of downward MONOTONICITY.

$\Rightarrow$ Compatible finding with Agmon et al. (2019).

Case study 2

Meaning representations of quantifiers

Uncovering the Structure of Semantic Representations Using a Computational Model of Decision-Making

Sonia Ramotowska,[a] Shane Steinert-Threlkeld,[b] Leendert van Maanen,[c] Jakub Szymanik[d]

[a] *Institute für Linguistik, Heinrich Heine Universität Düsseldorf*
[b] *Department of Linguistics, University of Washington*
[c] *Department of Experimental Psychology, Utrecht University*
[d] *Center for Mind/Brain Sciences and Department of Information Engineering and Computer Science, University of Trento*

Received 9 July 2022; received in revised form 29 October 2022; accepted 13 December 2022

Motivation: The case of *most* and *more than half*

- According to standard semantic theories *most* and *more than half* are truth conditionally equivalent

Motivation: The case of *most* and *more than half*

- According to standard semantic theories *most* and *more than half* are truth conditionally equivalent
  - $[\![\text{Most of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \mid A \cap \neg B \mid)$
  - $[\![\text{More than half of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \frac{|A|}{2})$

Motivation: The case of *most* and *more than half*

- According to standard semantic theories *most* and *more than half* are truth conditionally equivalent
  - $[\![\text{Most of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \mid A \cap \neg B \mid)$
  - $[\![\text{More than half of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \frac{|A|}{2})$
- This means that both should be true for proportions above 50% and false for proportions below 50%.

Motivation: The case of *most* and *more than half*

- According to standard semantic theories *most* and *more than half* are truth conditionally equivalent
  - $[\![\text{Most of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \mid A \cap \neg B \mid)$
  - $[\![\text{More than half of the As are B}]\!] = 1 \leftrightarrow (\mid A \cap B \mid > \frac{|A|}{2})$
- This means that both should be true for proportions above 50% and false for proportions below 50%.
- The proportion for which the truth value changes is the quantifier **threshold**

Motivation: The case of *most* and *more than half*

- According to standard semantic theories *most* and *more than half* are truth conditionally equivalent
  - $[\![\text{Most of the As are B}]\!] = 1 \leftrightarrow (|\ A \cap B\ | > |\ A \cap \neg B\ |)$
  - $[\![\text{More than half of the As are B}]\!] = 1 \leftrightarrow (|\ A \cap B\ | > \frac{|A|}{2})$
- This means that both should be true for proportions above 50% and false for proportions below 50%.
- The proportion for which the truth value changes is the quantifier **threshold**
- This standard view on truth conditions of *most* and *more than half* has not been experimentally tested

Goal: The case of *most* and *more than half*

- Previous studies have shown that. . .
  - *most* and *more than half* are verified using different verification strategies (cf. Hackl, 2009)

Goal: The case of *most* and *more than half*

- Previous studies have shown that...
  - *most* and *more than half* are verified using different verification strategies (cf. Hackl, 2009)
  - they are preferred to be used with different proportions (cf. Solt, 2016)

Goal: The case of *most* and *more than half*

- Previous studies have shown that...
  - *most* and *more than half* are verified using different verification strategies (cf. Hackl, 2009)
  - they are preferred to be used with different proportions (cf. Solt, 2016)
  - *most*, but not *more than half*, is vague (Denic & Szymanik, 2022)

Goal: The case of *most* and *more than half*

- Previous studies have shown that...
  - *most* and *more than half* are verified using different verification strategies (cf. Hackl, 2009)
  - they are preferred to be used with different proportions (cf. Solt, 2016)
  - *most*, but not *more than half*, is vague (Denic & Szymanik, 2022)
- Current goals
  - Experimentally test the truth conditional equivalence of *most* and *more than half*

Goal: The case of *most* and *more than half*

- Previous studies have shown that...
  - *most* and *more than half* are verified using different verification strategies (cf. Hackl, 2009)
  - they are preferred to be used with different proportions (cf. Solt, 2016)
  - *most*, but not *more than half*, is vague (Denic & Szymanik, 2022)
- Current goals
  - Experimentally test the truth conditional equivalence of *most* and *more than half*
  - Capture quantifier threshold and vagueness by different model parameters

Goal: The case of *most* and *more than half*

- Previous studies have shown that...
  - *most* and *more than half* are verified using different verification strategies (cf. Hackl, 2009)
  - they are preferred to be used with different proportions (cf. Solt, 2016)
  - *most*, but not *more than half*, is vague (Denic & Szymanik, 2022)
- Current goals
  - Experimentally test the truth conditional equivalence of *most* and *more than half*
  - Capture quantifier threshold and vagueness by different model parameters
  - Test whether all participants have the same truth conditional representations of quantifiers

Quantifier verification task

- Five English quantifiers *more than half*, *most*, *fewer than half*, *many*, and *few*

Quantifier verification task

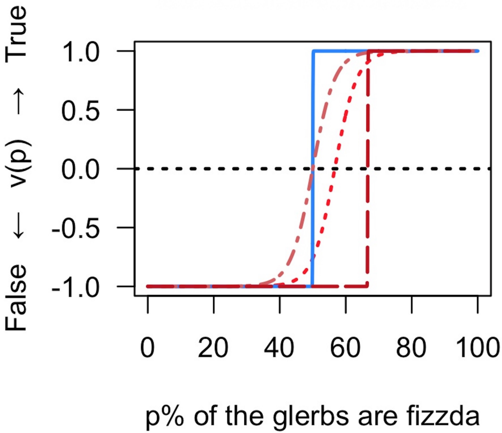
- Five English quantifiers *more than half*, *most*, *fewer than half*, *many*, and *few*
- The same verification task as linguistic task in Schlotterbeck et al. (2020)

Quantifier verification task

- Five English quantifiers *more than half*, *most*, *fewer than half*, *many*, and *few*
- The same verification task as linguistic task in Schlotterbeck et al. (2020)
- Experiment: test individual differences in meaning representations of *most* and *more than half*

Meaning representations - threshold and vagueness

$$v(p) = V_L + \frac{V_U - V_L}{1 + e^{(-s*(p-p_0))}}$$



Same procedure

1. Start with an hypothesis regarding DDM parameters $\surd$
2. Collect quantifier verification data (RT and accuracy) $\surd$
3. Fit model to data $\surd$
4. Constrain parameters $\surd$
5. Check model fit $\surd$
6. If model fit is good, perform hypothesis tests...

Difference in thresholds between *most* and *more than half*

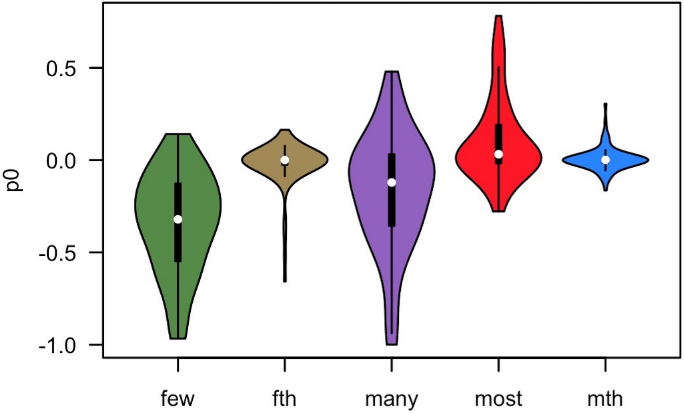


Figure: Individual differences in threshold of *most*.

Difference in thresholds between *most* and *more than half*

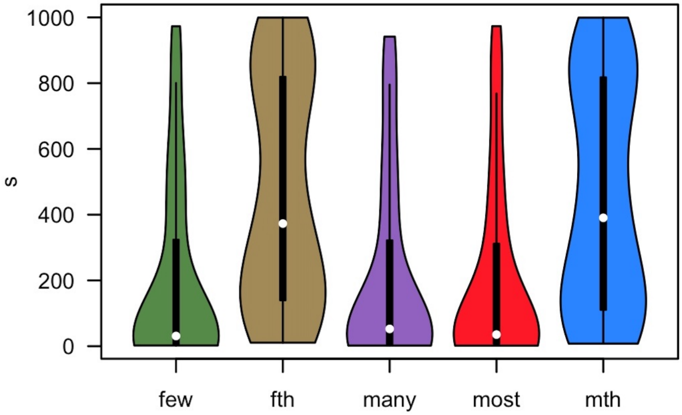


Figure: *Most* is more vague than *more than half*.

Most is not truth conditionally equivalent to *more than half*

- For some participants, *most* had a higher threshold than *more than half*

Most is not truth conditionally equivalent to *more than half*

- For some participants, *most* had a higher threshold than *more than half*
- For those participants, constraining threshold to 50% did not improve model fit for *most*

Most is not truth conditionally equivalent to *more than half*

- For some participants, *most* had a higher threshold than *more than half*
- For those participants, constraining threshold to 50% did not improve model fit for *most*
- In addition, our modeling approach captured differences between quantifiers in vagueness

Most is not truth conditionally equivalent to *more than half*

- For some participants, *most* had a higher threshold than *more than half*
- For those participants, constraining threshold to 50% did not improve model fit for *most*
- In addition, our modeling approach captured differences between quantifiers in vagueness
- This finding is challenging for semantic theories that do not allow for individual differences

Most is not truth conditionally equivalent to *more than half*

- For some participants, *most* had a higher threshold than *more than half*
- For those participants, constraining threshold to 50% did not improve model fit for *most*
- In addition, our modeling approach captured differences between quantifiers in vagueness
- This finding is challenging for semantic theories that do not allow for individual differences
- How stable are semantic representations?

Most is not truth conditionally equivalent to *more than half*

- For some participants, *most* had a higher threshold than *more than half*
- For those participants, constraining threshold to 50% did not improve model fit for *most*
- In addition, our modeling approach captured differences between quantifiers in vagueness
- This finding is challenging for semantic theories that do not allow for individual differences
- How stable are semantic representations?
  - Over time

Most is not truth conditionally equivalent to *more than half*

- For some participants, *most* had a higher threshold than *more than half*
- For those participants, constraining threshold to 50% did not improve model fit for *most*
- In addition, our modeling approach captured differences between quantifiers in vagueness
- This finding is challenging for semantic theories that do not allow for individual differences
- How stable are semantic representations?
  - Over time
  - Across different experimental manipulations

Case study 3

Representations of quantifiers under speed stress

Time-pressure Does Not Alter the Bias Towards the Canonical Interpretation of Quantifiers

Ruben D. Potthoff (r.potthoff@nki.nl)
Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Plesmanlaan 121
Amsterdam, 1066 CX The Netherlands &
Department of Experimental Psychology, Utrecht University, Heidelberglaan 8
Utrecht, 3584 CS The Netherlands

Sonia Ramotowska (sonia.ramotowska@hhu.de)
Department of Linguistics, Heinrich Heine University, Universitätsstraße 1
Düsseldorf, 40225 Germany

Jakub Szymanik (jakub.szymanik@gmail.com)
Center for Brain/Mind Sciences and the Computer Science Department, University of Trento, Corso Bettini 31
Rovereto TN, 38068 Italy

Leendert van Maanen (l.vanmaanen@uu.nl)
Department of Experimental Psychology, Utrecht University, Heidelberglaan 8
Utrecht, 3584 CS The Netherlands

Motivation

Interface Transparency Thesis (ITT)

"Speakers exhibit a bias towards the verification procedures provided by canonical specifications of truth conditions."
Lidz, Pietroski, Halberda, and Hunter (2011, p.229)

Motivation

Interface Transparency Thesis (ITT)

"Speakers exhibit a bias towards the verification procedures provided by canonical specifications of truth conditions."
Lidz, Pietroski, Halberda, and Hunter (2011, p.229)

How stable is this bias?

Speed-accuracy manipulation in DDM

- DDM gives a clear prediction about the effect of speed-accuracy manipulation on model parameters.
- In speed condition:
  - The decisions are faster but more prone to errors.
  - The distance between decision boundaries is shorter.
- In accuracy condition:
  - The decisions are slower but more accurate.
  - The distance between decision boundaries is larger.

$\Rightarrow$ Speed-accuracy trade off manipulation should affect the a parameter.

Speed-accuracy manipulation and linguistic representations

- Speed-accuracy manipulation can also have an effect on linguistic representations.

Speed-accuracy manipulation and linguistic representations

- Speed-accuracy manipulation can also have an effect on linguistic representations.
- In speed condition, participants derive scalar implicatures less often than in accuracy condition (Bott & Noveck, 2004).

Speed-accuracy manipulation and linguistic representations

- Speed-accuracy manipulation can also have an effect on linguistic representations.
- In speed condition, participants derive scalar implicatures less often than in accuracy condition (Bott & Noveck, 2004).
- Some verification strategies may be affected by speed stress (e.g., because they require more precision).

Speed-accuracy manipulation and linguistic representations

- Speed-accuracy manipulation can also have an effect on linguistic representations.
- In speed condition, participants derive scalar implicatures less often than in accuracy condition (Bott & Noveck, 2004).
- Some verification strategies may be affected by speed stress (e.g., because they require more precision).
- Will speed-stress affect DDM parameters related to meaning representations: threshold or vagueness?

Quantifier verification task

- Four Dutch quantifiers *more than half* (meer dan de helft), *most* (de meeste), *less than half* (minder dan de helft), and *least* (de minste)

Quantifier verification task

- Four Dutch quantifiers *more than half* (meer dan de helft), *most* (de meeste), *less than half* (minder dan de helft), and *least* (de minste)
- The same verification task as linguistic task in Schlotterbeck et al. (2020) and Ramotowska et al. (2023)

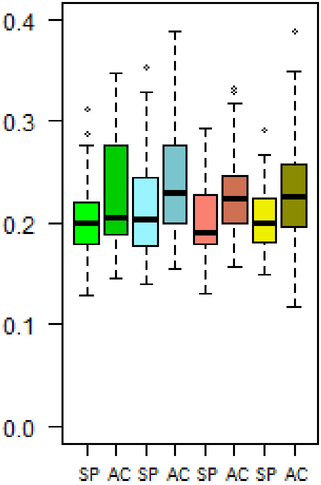
Quantifier verification task

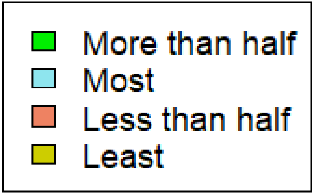
- Four Dutch quantifiers *more than half* (meer dan de helft), *most* (de meeste), *less than half* (minder dan de helft), and *least* (de minste)
- The same verification task as linguistic task in Schlotterbeck et al. (2020) and Ramotowska et al. (2023)
- Two with-subject conditions: speed-stress and accuracy-stress

Same procedure

1. Start with an hypothesis regarding DDM parameters $\surd$
2. Collect quantifier verification data (RT and accuracy) $\surd$
3. Fit model to data $\surd$
4. Constrain parameters $\surd$
5. Check model fit $\surd$
6. If model fit is good, perform hypothesis tests. . .

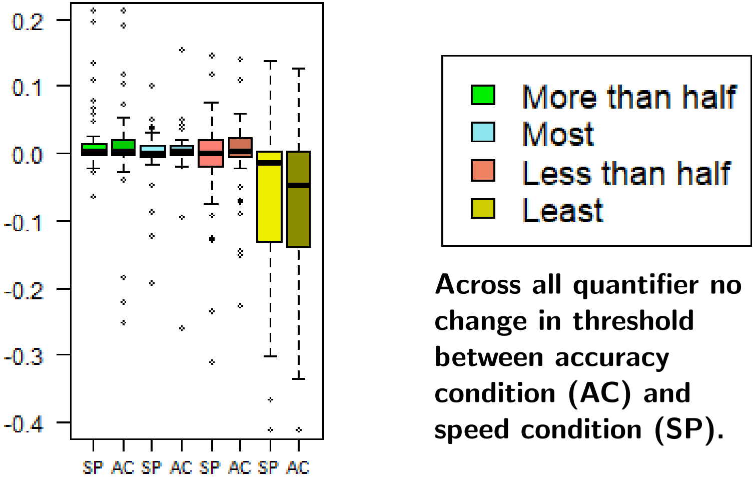
Results - boundaries separation





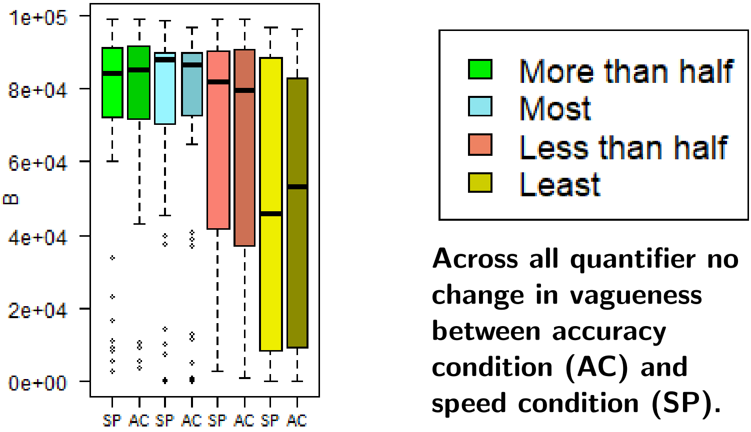
Across all quantifier larger boundaries separation in accuracy condition (AC) than in speed condition (SP).

Results - thresholds



Across all quantifier no change in threshold between accuracy condition (AC) and speed condition (SP).

Results - vagueness



Across all quantifier no change in vagueness between accuracy condition (AC) and speed condition (SP).

Representations of quantifiers under speed stress

- The speed-accuracy trade-off manipulation was successful as indicated by the change in *a* parameter

Representations of quantifiers under speed stress

- The speed-accuracy trade-off manipulation was successful as indicated by the change in *a* parameter
- Representations of quantifiers under speed-stress stay the same

Representations of quantifiers under speed stress

- The speed-accuracy trade-off manipulation was successful as indicated by the change in *a* parameter
- Representations of quantifiers under speed-stress stay the same
- The result in line with ITT

Conclusions

Towards a hybrid model

- Computational
  - Optimal response strategy in Bayesian terms
- Procedural
  - Stochastic process of evidence accumulation
  - Non-decision time as a separate processing step

DDM and beyond

- How can we model multiple processing stages in this framework?
  - E.g., the dual-stage two-phase (DSTP) model
  - Linguistic phenomena involving two-stage processing e.g., negation, scalar implicatures
- What are the implications for time insensitive models?

Some open questions

- Under which set of assumptions can we systematically relate drift rate to production probabilities (specifically in computational pragmatic models; cf. van Tiel et al. 2021)?
- What determines shifts in decision threshold of quantifiers? (cf. Ramotowska et al. 2023, Schoeller & Franke, 2015)
- How robust are verification algorithms related to drift rates (cf. Potthoff et al., 2023; Lidz et al., 2011)?

Thank you!