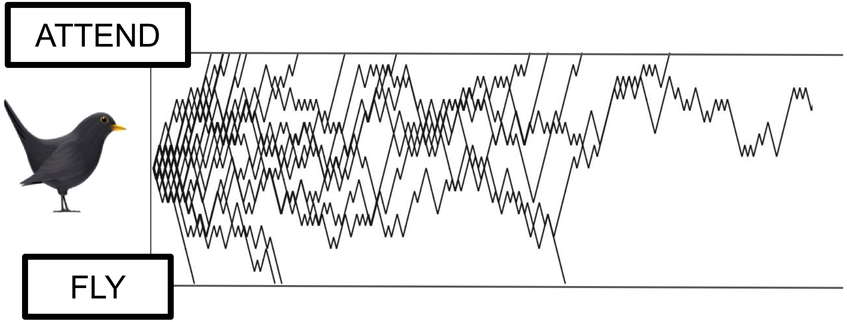
European Research Council
Established by the European Commission

Jointly modeling maze RT and accuracy using diffusion models



A first case study

Jack Duff (UCLA) & Laura Pissani (Saarland)
AMLaP 2025, Prague

The maze task: Measuring word-by-word comprehension difficulty through RTs on decisions between target continuations and foils.

In this talk:
- Data demonstrating **decision difficulty** effects as a function of the foil (!)
- How we can use cognitive models of decision-making to understand these effects

Upshot: A better understanding of the multiple mechanisms underlying maze data, and how we can tease them apart.

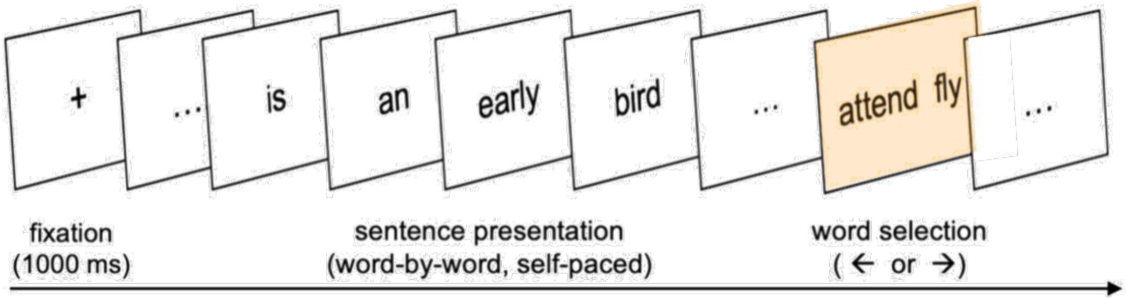
The metaphor awakening effect

FLY

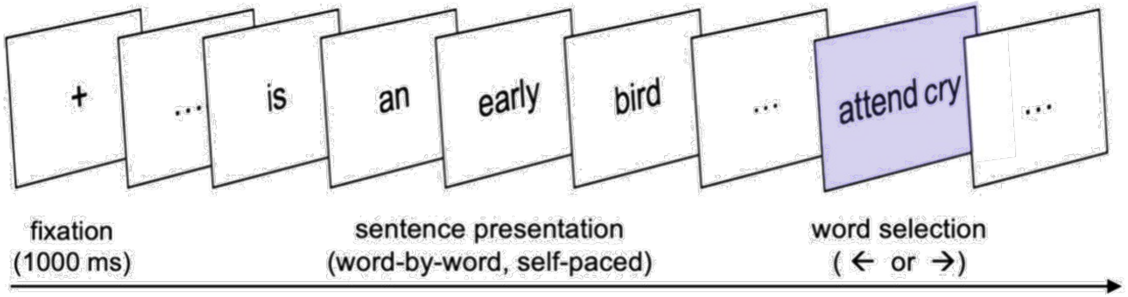
John is an *early bird* so he can attend morning classes.

The maze task: conditions

TARGET: *attend*
RELATED-foil: *fly*



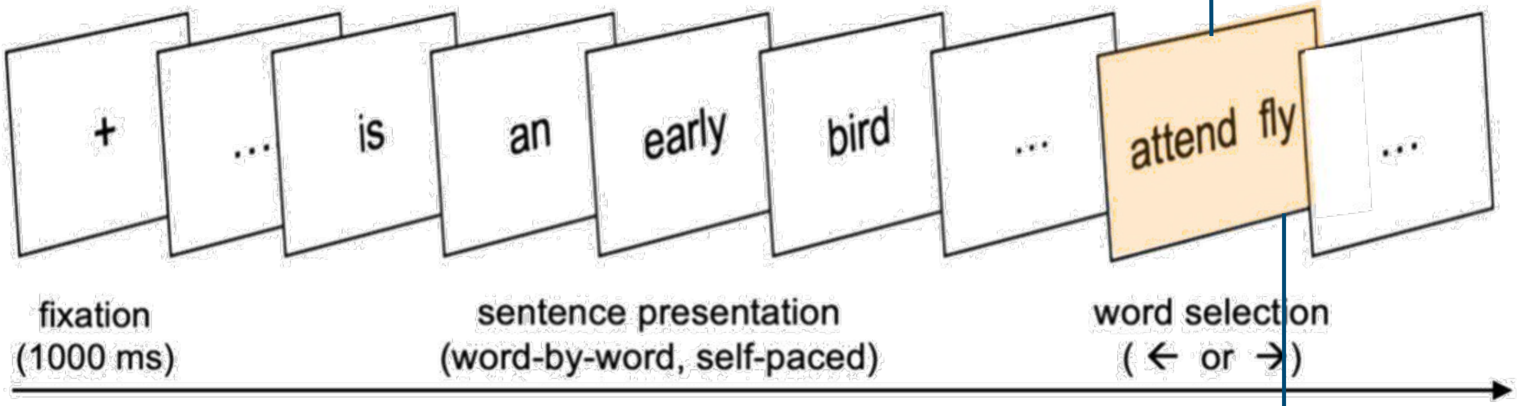
TARGET: *attend*
UNRELATED-foil: *cry*



(Pissani & de Almeida 2022)

The maze task: measures

Response Time: Time taken to select the target *attend*
Accuracy: Proportion of selecting the target *attend*

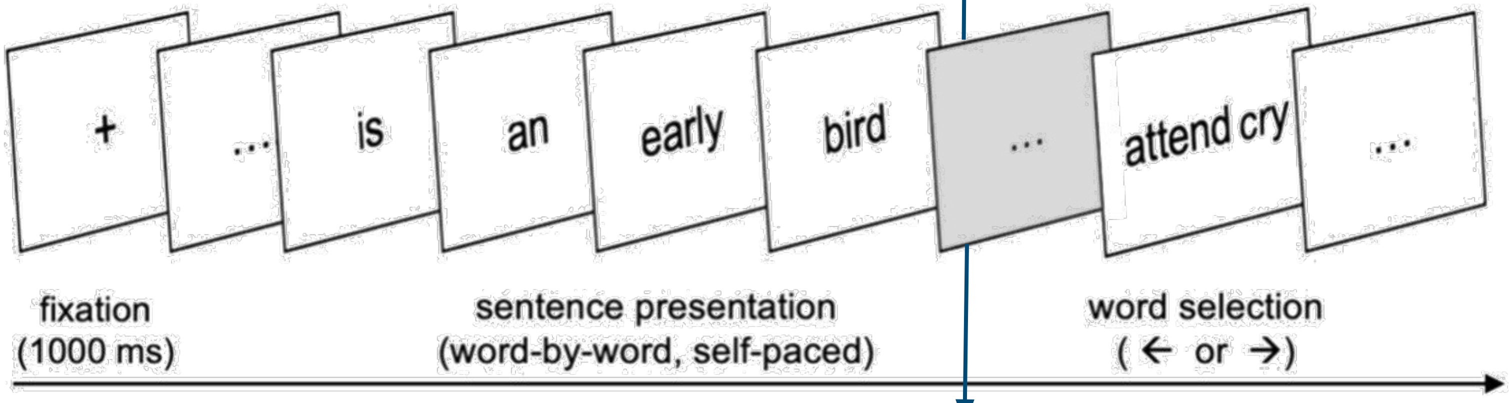


When the target is paired with the related foil *fly*, we expect participants to be **slower** and **less accurate**.

(Pissani & de Almeida 2022)

The maze task: delays

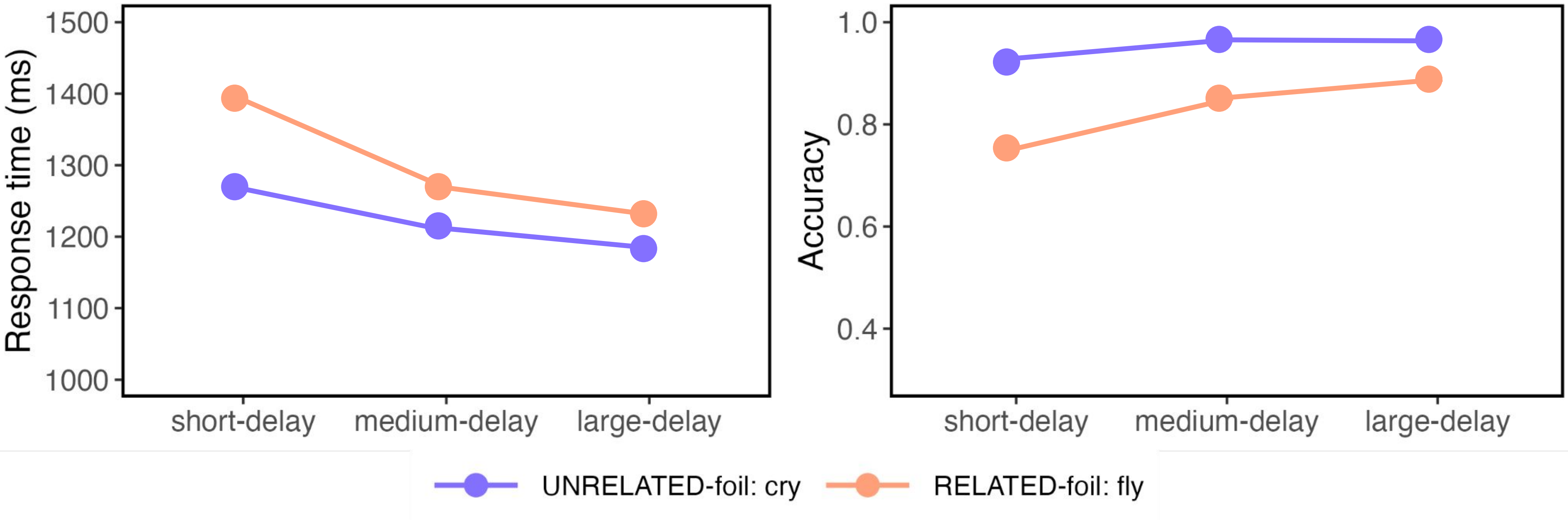
Short delay: 1 - 3 words
Medium delay: 6 - 8 words
Long delay: 10 - 13 words



As the distance between the metaphor early bird and the maze decision decreases, we expect **the awakening effect to diminish.**

(Pissani & de Almeida 2022, 2023)

Results



(Pissani & de Almeida 2022, 2023)

Key findings

The metaphor awakening effect (MAE) reflects the relationship between the literal meaning of the metaphor and the subsequent literal cue.

The literal meaning of metaphors may be initially available and remain available further downstream, however it fades rapidly as the conventional content takes priority.

1 - 3 words

MAE: 116 ms↑ 16%↓

6 - 8 words

MAE: 51 ms↑ 13%↓

11 - 13 words

MAE: 55 ms↑ 8%↓

Diffusion decision modeling

Modeling core decision “difficulty”

Dependent measures

Rate of choosing target

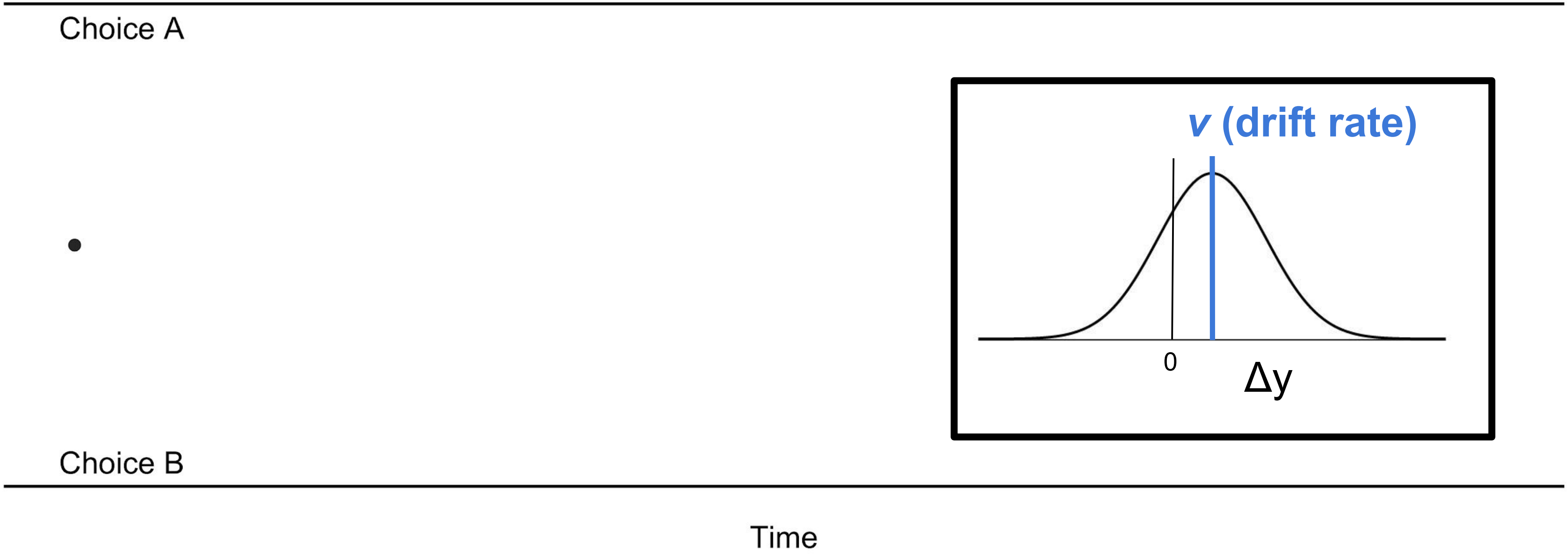
Target response time

Hypothesized source

Difficulty of deciding between target and foil

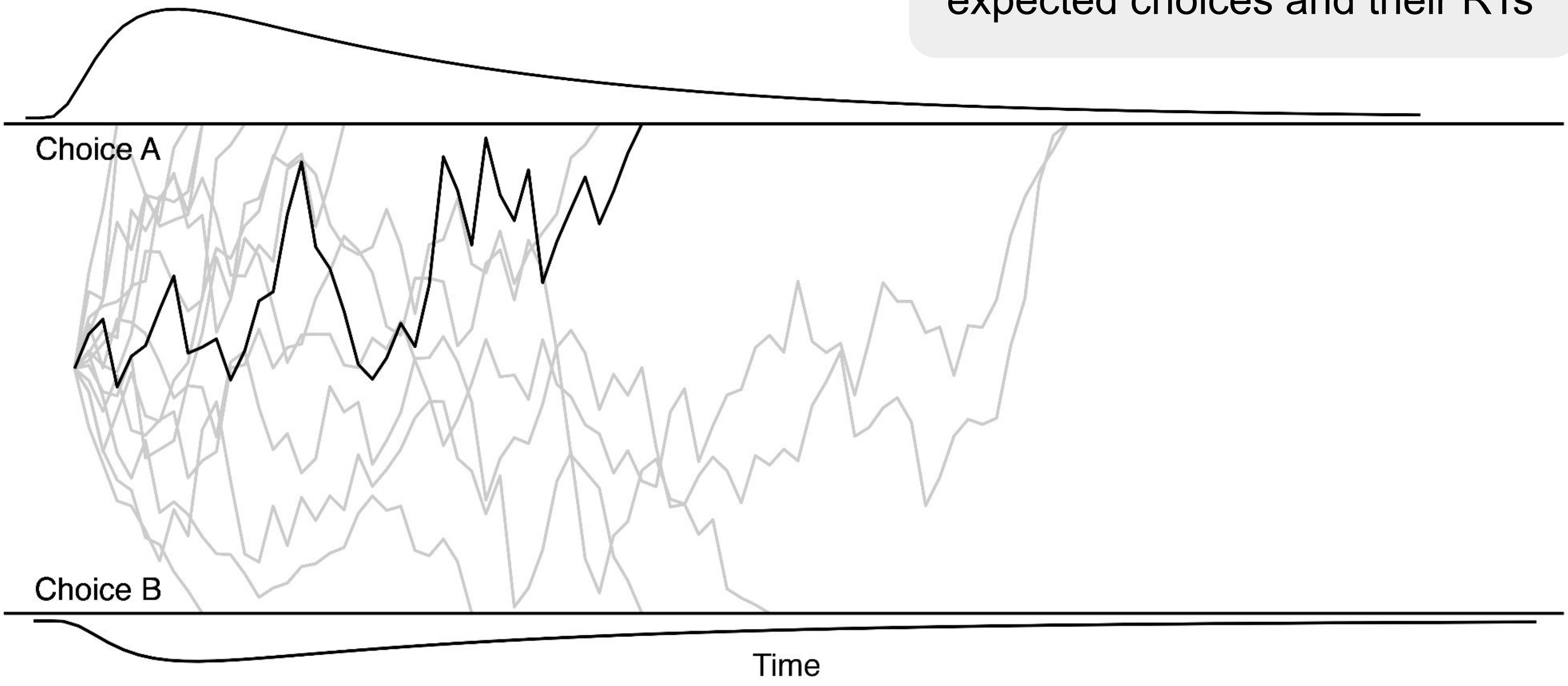
The diffusion model of decision-making

Continuous preference accumulation between two choices as noisy particle motion



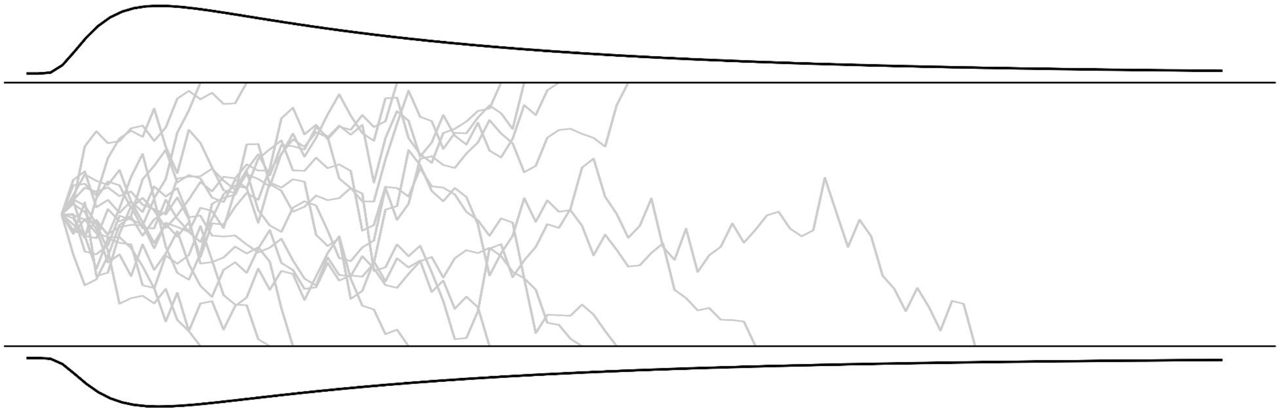
(Ratcliff 1978; Ratcliff & McKoon 2008; Ratcliff et al. 2016; Myers et al. 2022)

Result: Joint distribution of expected choices and their RTs



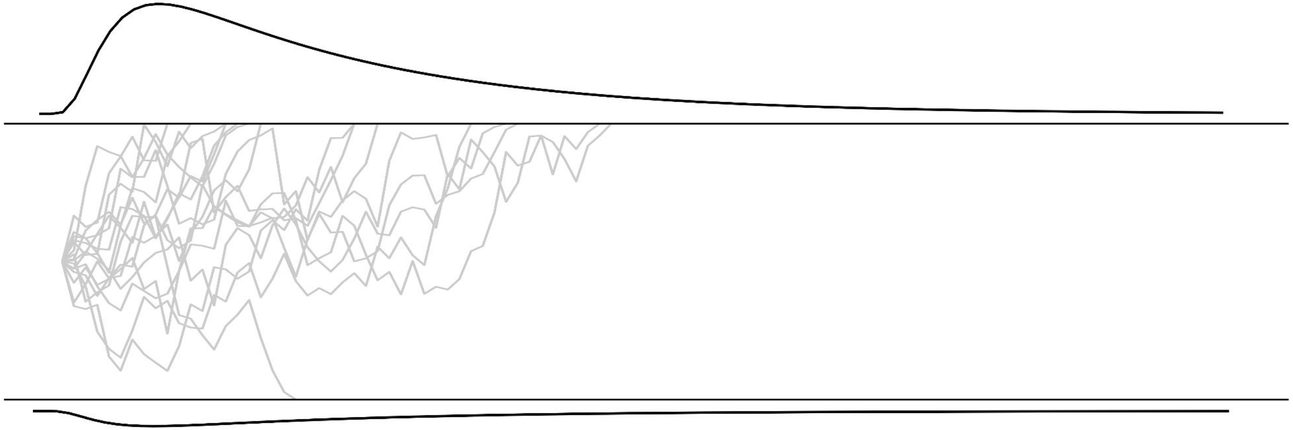
Harder decisions:
Lower *v*

- more errors
- slower RTs

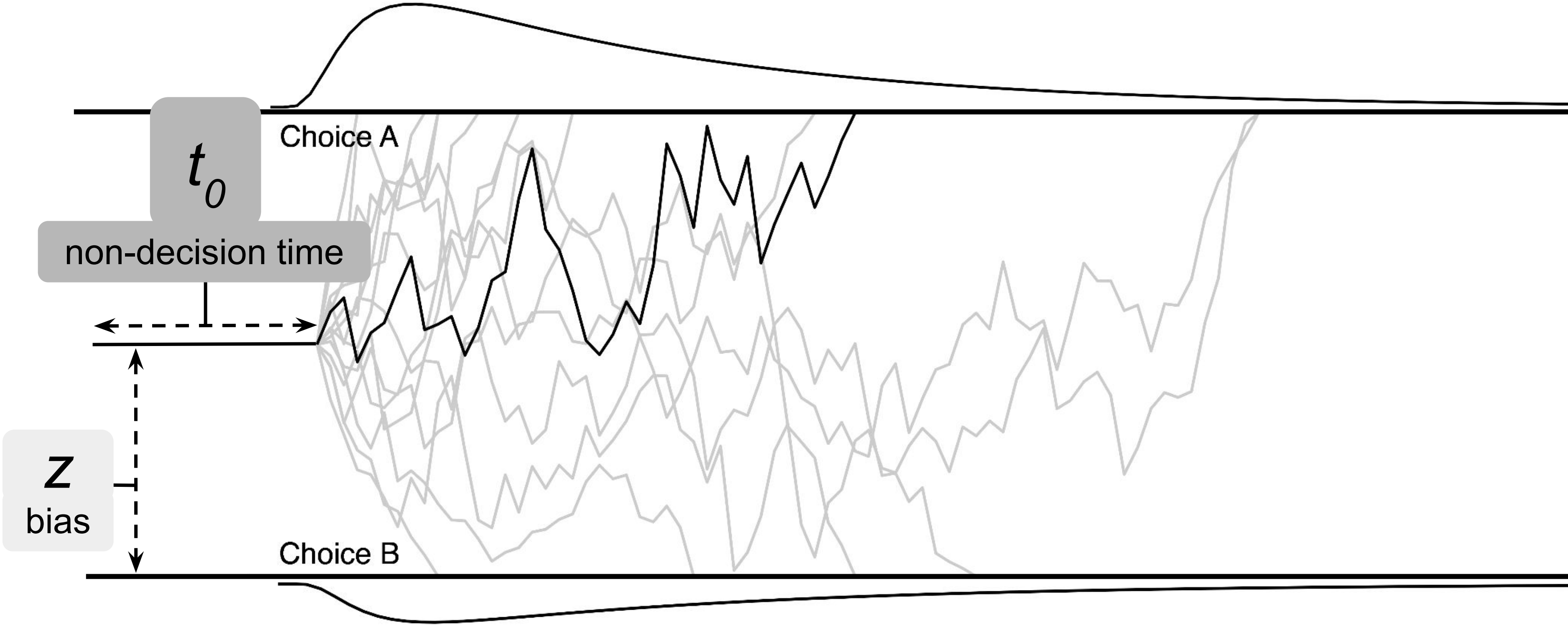


Easier decisions:
Higher *v*

- fewer errors
- faster RTs



Other parameters of a diffusion model

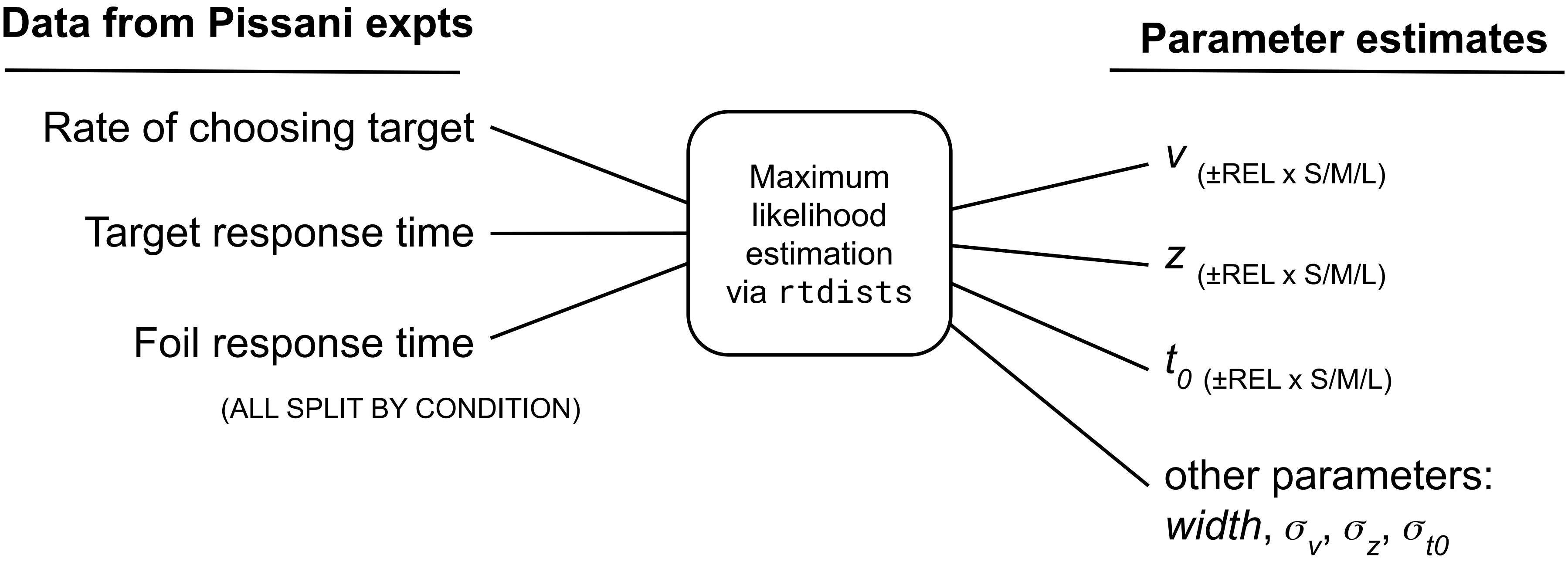


Applying a diffusion model

Our goals

- Can diffusion models capture the observed joint accuracy/RT patterns well?
- What is causing the effect: drift rate (v), non-decision time (t_0), initial bias (z)?
- Underlying everything: Exactly how are maze decisions sensitive to context?

Model-fitting procedure



Model comparison

- AIC compared across models with different numbers of condition-varying parameters
- All diffusion models outperformed GLM regression baseline
- Top models by AIC all allowed 6-way *v* contrasts
- Best model (#5) needed only 2-way contrast for *z* and t_0

Table 1. Model comparisons, best AIC in bold.

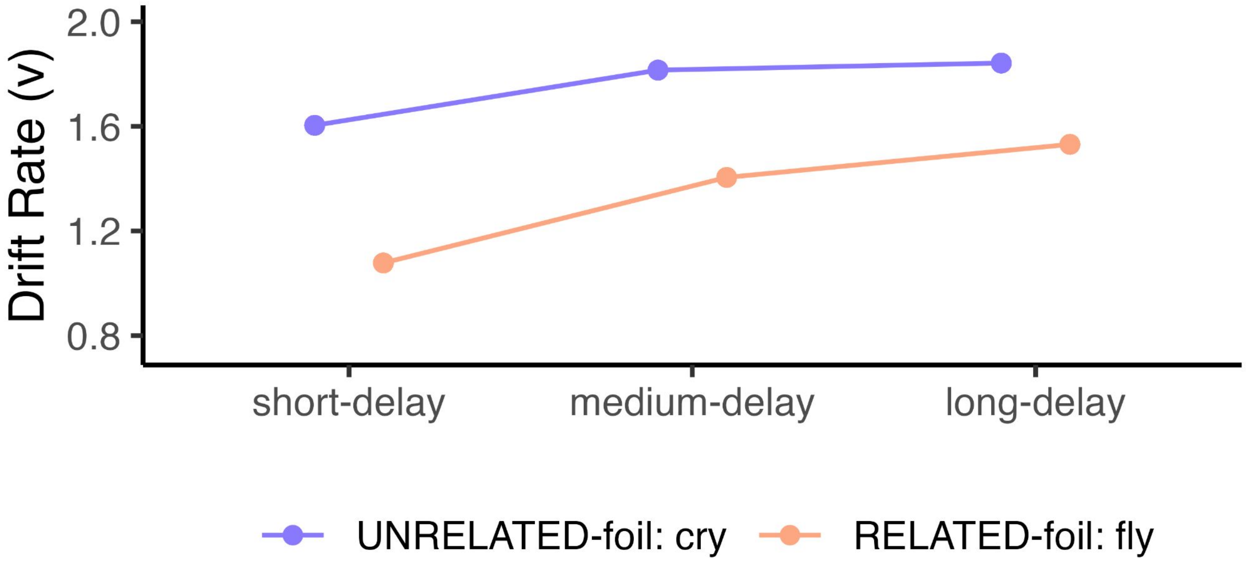
$_{\pm REL}$: parameter fit with two values, for related vs. unrelated foils.

$_{\pm REL\ x\ D}$: parameter fit with six values, crossing ±related and {S/M/L} delay.

	Model	∑ LogLik	AIC
1	$t_{0\ \pm REL\ x\ D}$, $z_{\pm REL\ x\ D}$, $v_{\pm REL\ x\ D}$	–2016.04	4076.08
2	$t_{0\ \pm REL}$, $z_{\pm REL\ x\ D}$, $v_{\pm REL\ x\ D}$	–2019.54	4075.07
3	$t_{0\ \pm REL\ x\ D}$, $z_{\pm REL}$, $v_{\pm REL\ x\ D}$	–2017.09	4070.18
4	$t_{0\ \pm REL\ x\ D}$, $z_{\pm REL\ x\ D}$, $v_{\pm REL}$	–2027.52	4091.05
5	$\boldsymbol{t_{0\ \pm REL}}$, $\boldsymbol{z_{\pm REL}}$, $\boldsymbol{v_{\pm REL\ x\ D}}$	**–2020.46**	**4068.92**
6	$t_{0\ \pm REL}$, $z_{\pm REL\ x\ D}$, $v_{\pm REL}$	–2032.00	4092.00
7	$t_{0\ \pm REL\ x\ D}$, $z_{\pm REL}$, $v_{\pm REL}$	–2036.45	4100.90
8	t_0, $z_{\pm REL}$, $v_{\pm REL\ x\ D}$	–2029.35	4084.69
9	$t_{0\ \pm REL}$, z, $v_{\pm REL\ x\ D}$	–2028.01	4082.02
10	t_0, z, $v_{\pm REL\ x\ D}$	–2031.16	4086.32
B	combination of GLMs	–2138.54	4317.09

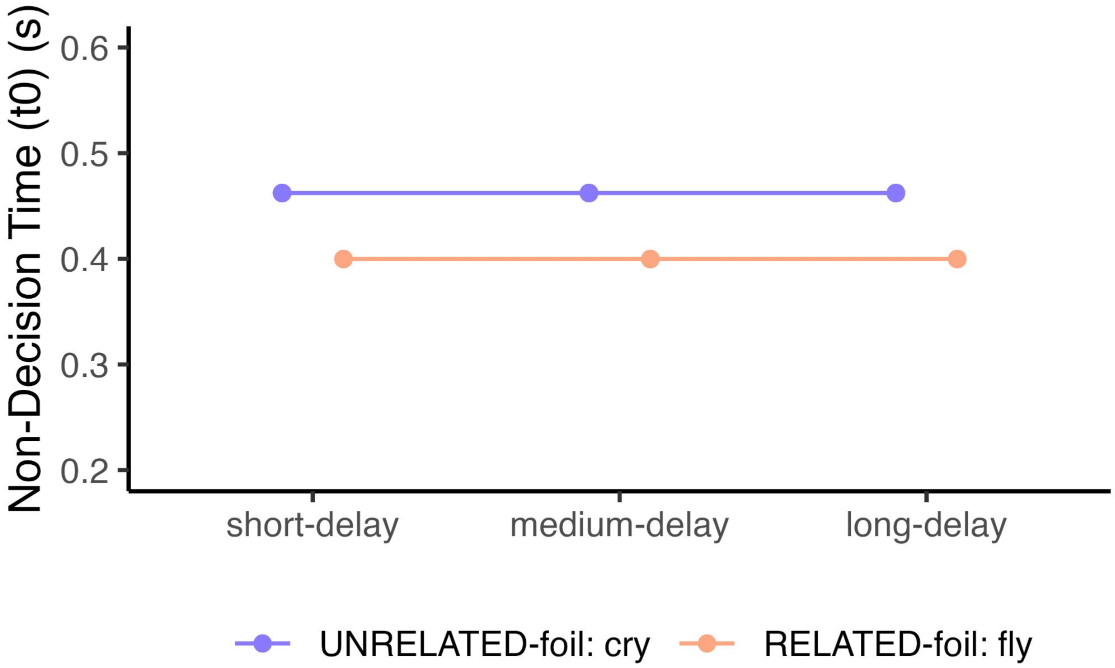
Properties of the best model: Drift rates

- **Lower drift rate** with related foils
  - Longer RTs
  - More errors
- Related/unrelated contrast is **smaller at later delays**



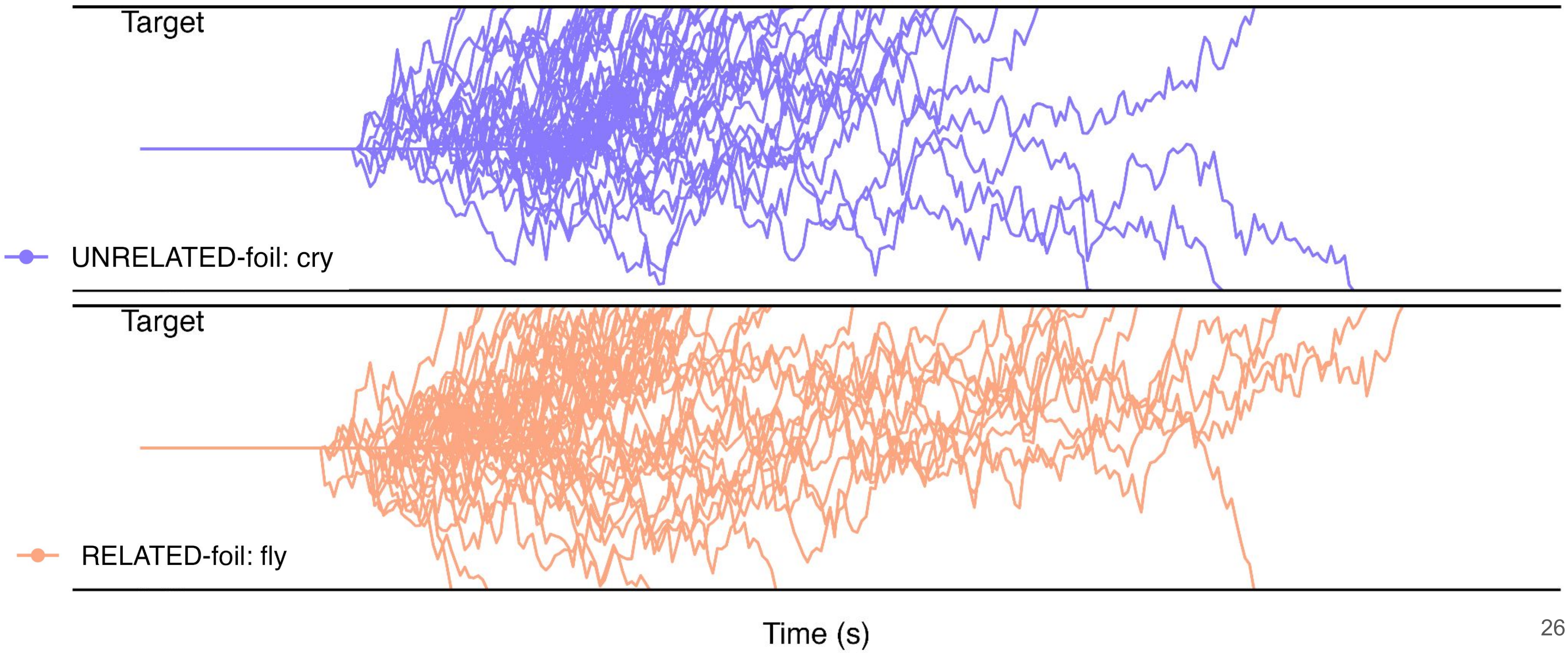
Properties of the best model: Non-decision times

- Counter-intuitively: Non-decision processing of target and foil is **faster** with related foils
- Possible explanation: Initial processing of related foils is primed



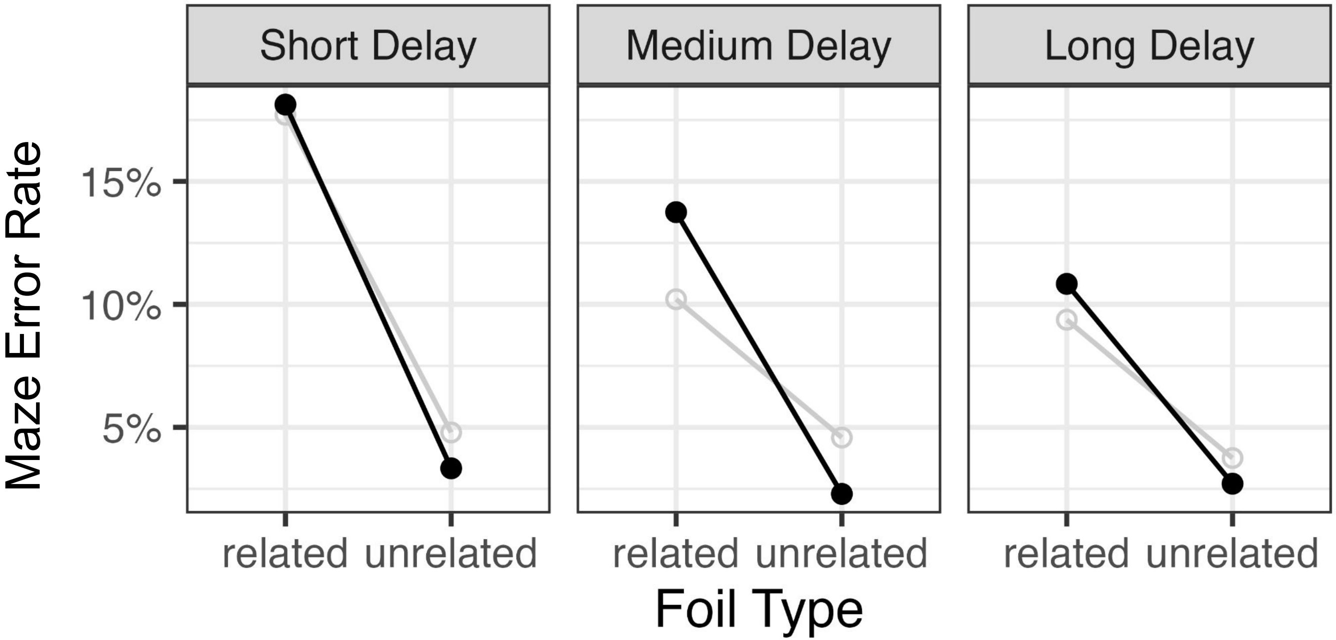
Simulated decision behavior with best model

Faster t_0 + slower v: many more errors + noise, but just a subtle increase in time



Comparing best model predictions to real data: Error Rate

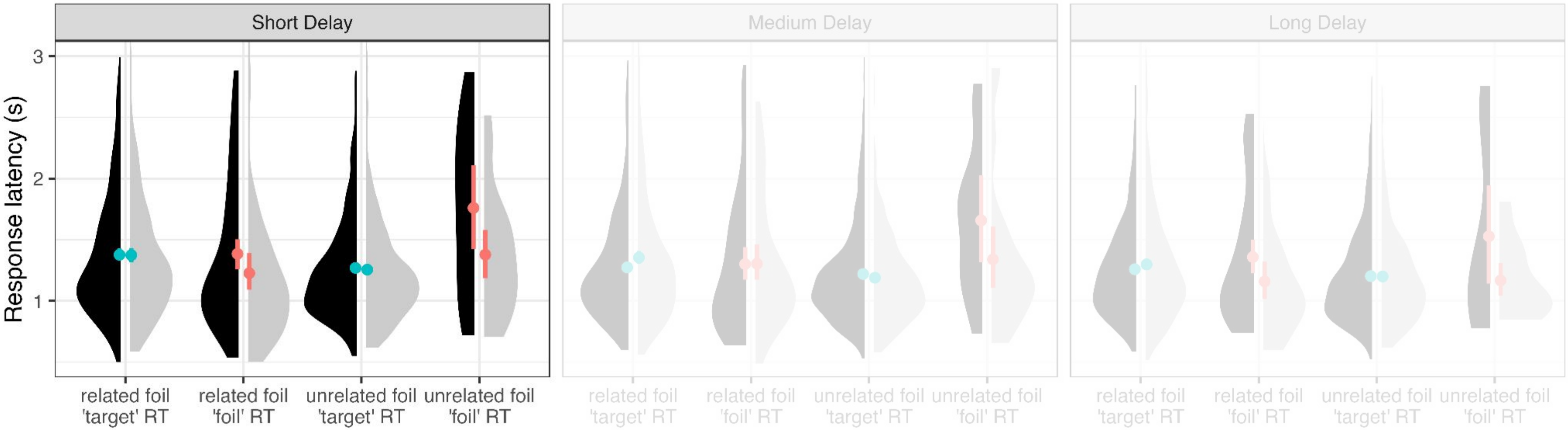
- Captures main accuracy effect, some aspects of delay curve
- Not quite capturing gradient shrinking from medium to long



(**Black** = Human data; **Grey** = Model predictions)

(**Black** = Human data; **Grey** = Model predictions)

Comparing best model predictions to real data: RT dists

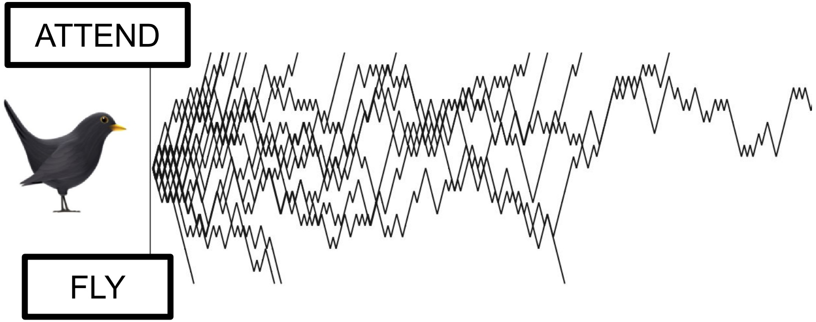


- Generally: Impressive fit of both center and shape of distributions
- Closest for “target” RTs
- Consistently understimates “foil” RTs, though note these were rarer

Our conclusions

- Diffusion models can indeed account for RT and error rate patterns as a joint consequence of contextually-activated foils.
  - Good evidence for a decision-difficulty explanation of the critical effect.
  - It’s good to use models which reflect the full complexity of your hypotheses!
- **Specifically:** Particular foils increased the difficulty of picking the target, even as they sped up other trial components (perhaps primed foil reading)
- **Overall**: Evidence for real complexity in maze latencies! Processing and decision-making are sensitive in different ways to target, foil, and context.
  - Consider your linking hypotheses with care when interpreting maze data!

Thanks!



Thanks to Amanda Rysling for first suggesting diffusion models for maze data, and thanks to Vera Demberg and many others at Saarland University for helpful feedback.

Funding

European Research Council
Established by the European Commission

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (Grant Agreement No. 948878 to V. Demberg).

Contact us: duff@ucla.edu, laura.pissani@uni-saarland.de

Back-up slides: Another case study

Another case study: Mazing homonyms in context

Homonyms after contexts which biased towards dominant/subordinate meanings.

```
                               ┌ M1: made his toast soggy        ┐
Reportedly, after it           ┤                                 ├ the jam displeased Tom.
                               └ M2: doubled his morning commute ┘
```

Classic eye-tracking finding (Duffy et al. 1988): slower reading of **jam** in M2 context.

Maze replication (Duff 2023): slower Maze decisions at **jam** in M2 context.

- Why? Slower target reading (within t_0) or harder decisions (v)?

Another case study: Diffusion model results

Best fit: modulating t_0 (or z), but not v.

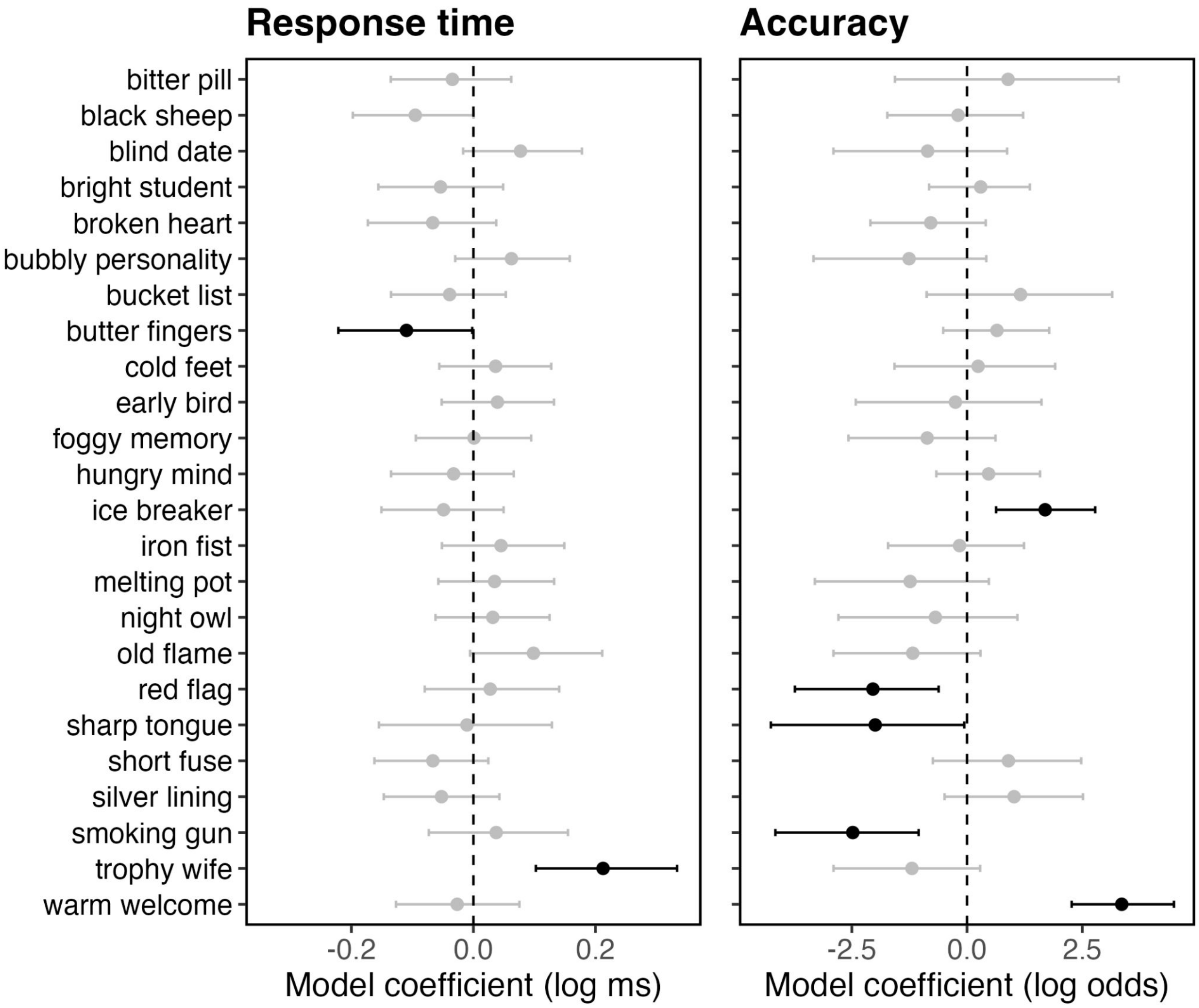
Tentative conclusion: Some Maze effects are not about decision difficulty.

A very rough cheat sheet, under diffusion model assumptions:

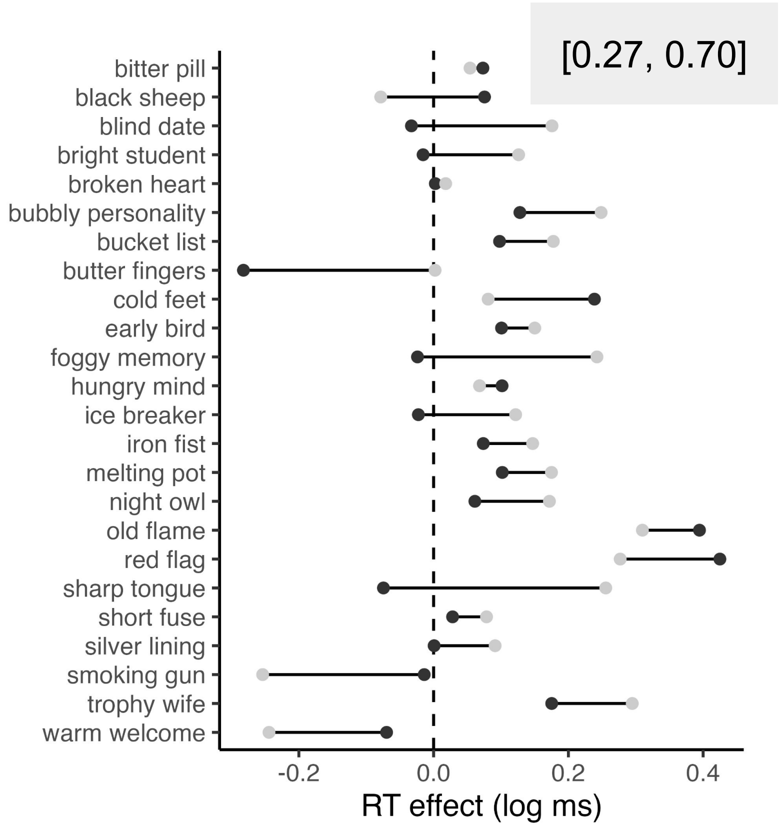
- RT slowdowns associated with increases in error rate: decision difficulty.
- RT slowdowns without increases in error rate: initial processing difficulty.

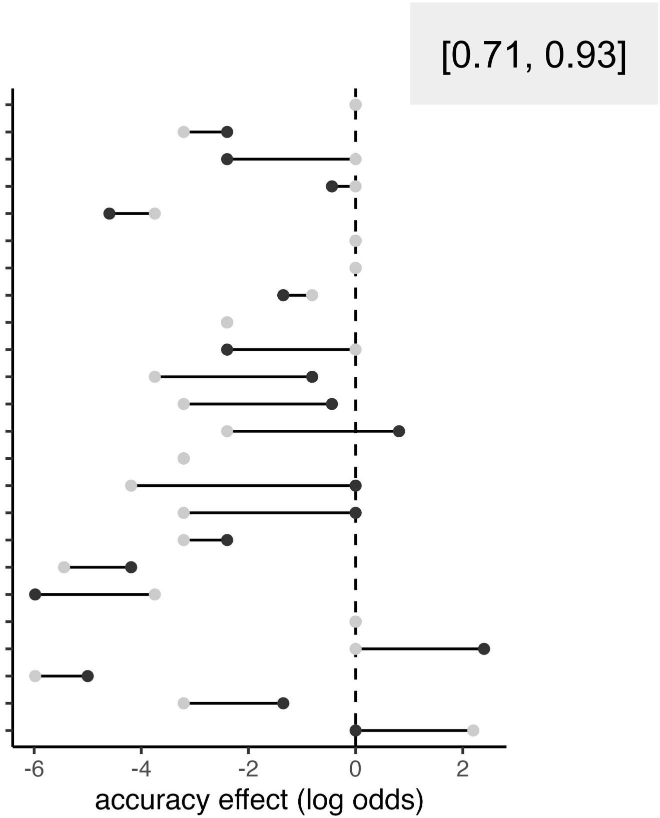
Back-up slides: Item variance + reliability

By-item random slopes for distractor



But: Split-half item reliability on RTs is very low, patterns unclear





Other extra material

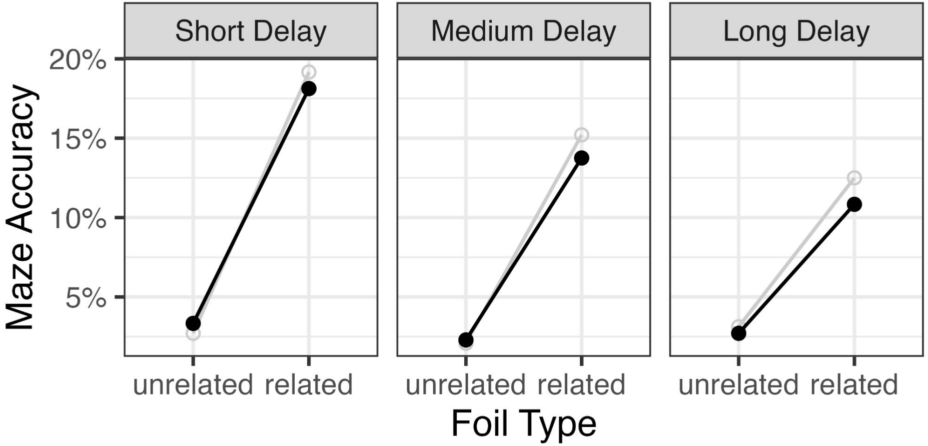
Details on our GLM baseline

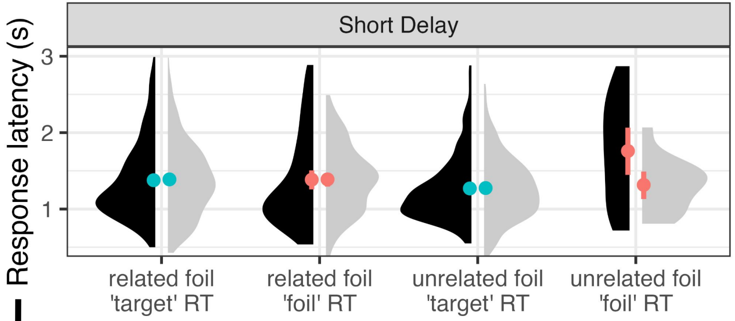
Joint likelihood

P(Response & RT | Condition) = P(Response | Condition) x P(RT | Response & Condition)

Logistic regression:
Response ~ FoilType * Delay

Gamma regression:
RT ~ FoilType * Delay * Response

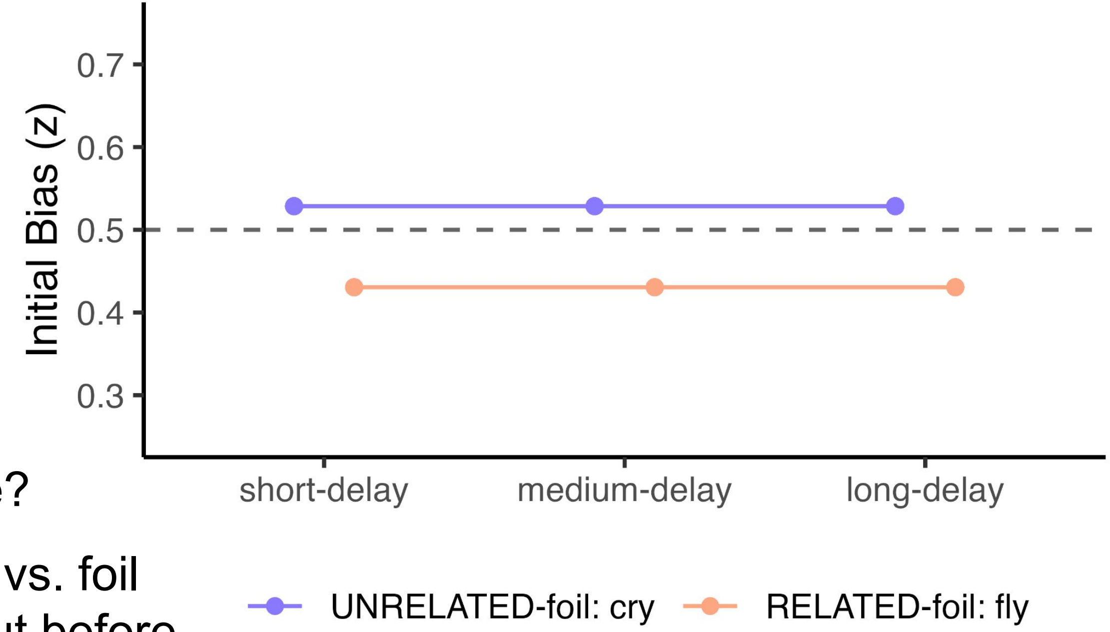




Captures center but not shape.

Properties of the best model: Initial bias

- **Bias away from target** with related foils
  - Longer RTs
  - More errors
  - Faster errors



- What could “bias” mean here?
  - Maybe: Appeal of target vs. foil after initial processing but before decision-making (driven by low-level activation?)

Awakening the literal meaning of familiar metaphors: Evidence from eye movements

Metaphor-related: Some people say that a *sharp tongue* can for sure CUT rather abruptly a friendship.
Metaphor-unrelated: Some people say that a *sharp tongue* can for sure RUIN rather abruptly a friendship.
Literal-related: Some people say that a *sharp tool* can for sure CUT rather abruptly a fabric.
Literal-unrelated: Some people say that a *sharp tool* can for sure RUIN rather abruptly a fabric.