Models Connect Verbal Theories to Empirical Data:

Examples and Principles

Pablo Gomez

January 2018

Agenda

- A big picture view
- Specific examples of my work

[0] I will use footnotes to invite you to ask questions about topics that are not central to this talk.

The forest

1. **Guiding principle: Models connect theory to data.**
   - Models are most useful when they can be compared to other models.
2. Big Question: How do external stimuli interact with internal representations?

Theories

- They are usually stated verbally in psychology.

Theories

- They are usually stated verbally in psychology.
- Example: “Frequency of exposure facilitates processing.”

Theories

- They are usually stated verbally in psychology.
- Example: “Frequency of exposure facilitates processing.”
- Example: Cognitive Dissonance Theory.

Models

- Instantiations of the theories.

Models

- Instantiations of the theories.
- Include probabilistic statements.

Models

- Instantiations of the theories.
- Include probabilistic statements.
- At their best, they give us insights and clarity.

Models

- Instantiations of the theories.
- Include probabilistic statements.
- At their best, they give us insights and clarity.
- They allow us to make principled inferences.

Models

- Instantiations of the theories.
- Include probabilistic statements.
- At their best, they give us insights and clarity.
- They allow us to make principled inferences.
- Examples: SEMs, GLM, **process models**, connectionist models.

Data

- It comes from our experimental work or from datified behaviors.

Data

- It comes from our experimental work or from datified behaviors.
- One can compute $p(data|Model)$.

Data

- It comes from our experimental work or from datified behaviors.
- One can compute $p(data|Model)$.
- $p(data - or - more - extreme)$ assuming H_0 if we are frequentists (p-value).

Data

- It comes from our experimental work or from datified behaviors.
- One can compute $p(data|Model)$.
- $p(data - or - more - extreme)$ assuming H_0 if we are frequentists (p-value).
- $\frac{p(data|H_1)}{p(data|H_0)}$ if we are bayesians (BF_{10}).

Data

- It comes from our experimental work or from datified behaviors.
- One can compute $p(data|Model)$.
- $p(data - or - more - extreme)$ assuming H_0 if we are frequentists (p-value).
- $\frac{p(data|H_1)}{p(data|H_0)}$ if we are bayesians (BF_{10}).
- Some form of goodness of fit

Data

- It comes from our experimental work or from datified behaviors.
- One can compute $p(data|Model)$.
- $p(data - or - more - extreme)$ assuming H_0 if we are frequentists (p-value).
- $\frac{p(data|H_1)}{p(data|H_0)}$ if we are bayesians (BF_{10}).
- Some form of goodness of fit
- BIC, AIC, etc..

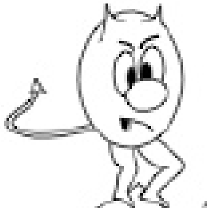
The forest

1. Guiding principle: Models connect theory to data.
    - **Models are most useful when they can be compared to other models.**
2. Big Question: How do external stimuli interact with internal representations?

Demonstration with a counterexample:

A demon comes to me and threatens to destroy earth unless I play a riddle, and I answer his question correctly.

- Demon: *I have selected a human, and you have to guess who that person is*
- Me: *That's too hard*.
- Demon: *Ok, fine, just answer on question about them: Is/was this person an American citizen?*
- Me: *I need some data!*



Data and inference

- With some data I can apply Principle 1, and use a model to evaluate such data.

Data and inference

- With some data I can apply Principle 1, and use a model to evaluate such data.
- $Model_A$: She is an American citizen.

Popperian falsification

If the demon gives me data that is unlikely under $Model_A$, I can reject that model.

The demon is a good guy

- He agrees to tell me three facts.

The demon is a good guy

- He agrees to tell me three facts.
- The person is a woman.

The demon is a good guy

- He agrees to tell me three facts.
- The person is a woman.
- Her initials are/were J. R.

Popperian falsification

- If data is unlikley under $Model_A$, reject $Model_A$.

Popperian falsification

- If data is unlikley under $Model_A$, reject $Model_A$.
- The demon gives me the third fact: J.R. is/was a congresswoman.

Popperian falsification

- If data is unlikley under $Model_A$, reject $Model_A$.
- The demon gives me the third fact: J.R. is/was a congresswoman.
- It is very unlikely to be a congresswoman if you are/were an American. $p(congresswoman|American) = \frac{300}{400{,}000{,}000}$

Popperian falsification

- If data is unlikley under $Model_A$, reject $Model_A$.
- The demon gives me the third fact: J.R. is/was a congresswoman.
- It is very unlikely to be a congresswoman if you are/were an American. $p(congresswoman|American) = \frac{300}{400{,}000{,}000}$
- (Conclusion): J.R. is probably not an American citizen: Reject $Model_A$.

The demon destroys the earth

Of course, the previous slide is absurd

1. We must consider $Model_N$: J.R. is not an America citizen.
2. We could calculate $p(congresswoman|American) = 300/400000000$ and $p(congresswmn|NON - American) = 0$
3. The data: *congresswoman* is only possible under one of the models. [1]

[1] Ask me about how this example can be used to teach about the frequentist vs. Bayesian debate

Jeannette Rankin

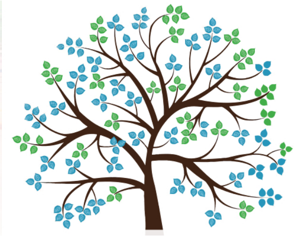
The forest

1. Guiding principle: Models connect theory to data.
   - Models are most useful when they can be compared to other models.
2. **Big Question: How do external stimuli interact with internal representations?**

The trees

1. **Lexical decision task**
2. Priming
3. Letter position coding

- Dyslexia
- Braille
- Cross-linguistic research



1. Lexical decision

- A string of letters like **MARCH** is presented to an observer.
- The observer decides if the string of letters is a word or a nonword.

Repetition

The strings could be presented once or repeated.

Word Frequency

INVECTIVE is a word that we see infrequently.

Type of nonword

- **JUGDE** really looks like “judge”.
- **TRTTH** only vaguely looks like “truth”.

Instructions

- *Respond as fast vs. as accurately as possible.*
- *Only respond to words (go/no-go).*
- *75% of items will be words.*

Understanding these tasks is challenging and fascinating

- Performance is a function of many factors (e.g., perceptual, semantic, strategic, individual).

Understanding these tasks is challenging and fascinating

- Performance is a function of many factors (e.g., perceptual, semantic, strategic, individual).
- Performance can be measured in many ways: latency, accuracy, ERPs, etc. . .

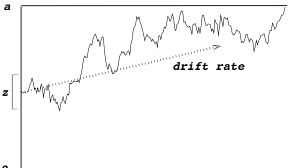
How do we make sense of these data?

Models connect theory to data

Diffusion Model. A process model that assumes that evidence is accumulated noisily until a decision threshold is reached. One fits the model to the data and obtains diffusion model parameters (Ratcliff, 1978).

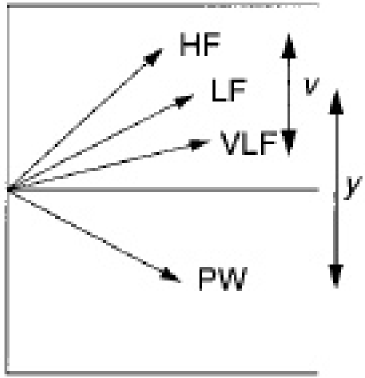
Diffusion Model

- There is an encoding of sensory information (T_{ER}).
- There is a process of accumulation of evidence as a function of time.
- Tokens of information provide evidence for one or the other response alternatives.
- Such accumulation of evidence is noisy, meaning that it does not grow monotonically.
- When the evidence reaches one of the two decision thresholds, a response is initiated.
- The position of these boundaries relates to the amount of evidence needed to make a decision.



Ratcliff, Gomez & McKoon (2004, Psych Review)

*"In summary, application of the diffusion model to lexical decision data shows that the major effects... are **all** captured by a single component of processing: drift rate."*



Looking at a tree from the forest.

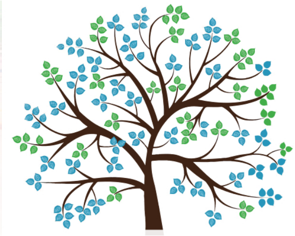
- **Theory.** The lexical decision task involves a resonance[2] between the stimulus and mental representations in the lexicon.
- **Model.** Data is accounted for by drift rate parameter.
- **Data.** LDT data from many experiments (8 in Psych Rev paper, 4 in Psych Ageing, 4 in JEPG...)

[2] Ask me about speed dating

The trees

1. Lexical decision task
2. **Priming**
3. Letter position coding

- Dyslexia
- Braille
- Cross-linguistic research



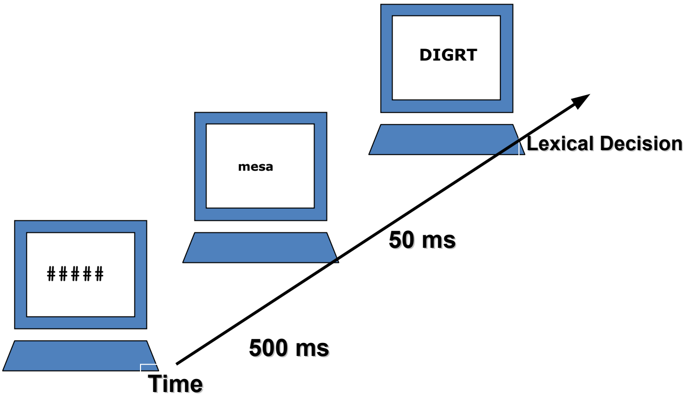
Now, using model comparison

- In the last 10-15 years, the masked priming procedure has been widely used in visual word recognition research.
- What is the consequence of priming on the lexical decision?

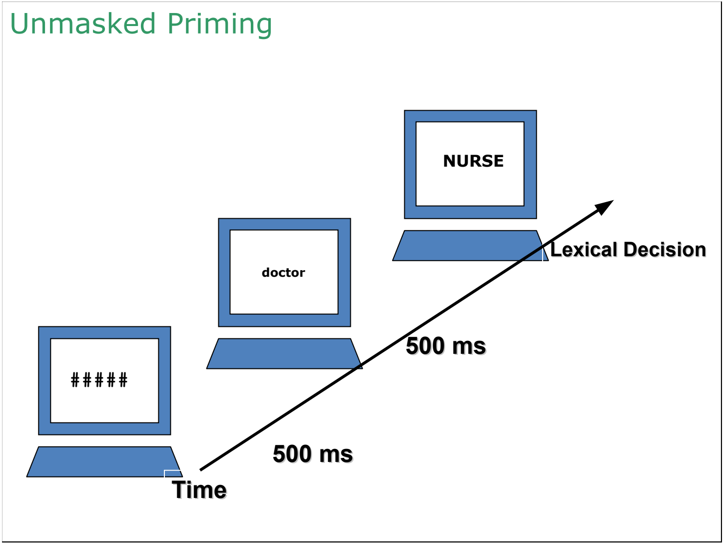
Two methods of priming

Masked Priming

Masked Priming (Forster y Davis, 1984)



Unmasked Priming



. . . and two types of relationship between primes and targets

We hoped to identify what aspect or component of processing is affected by priming. [3]

- Associate Priming: **doctor** primes **NURSE**

[3] Ask me about the effects of ageing.

. . . and two types of relationship between primes and targets

We hoped to identify what aspect or component of processing is affected by priming. [3]

- Associate Priming: **doctor** primes **NURSE**
- Masked

[3] Ask me about the effects of ageing.

… and two types of relationship between primes and targets

We hoped to identify what aspect or component of processing is affected by priming. [3]

- Associate Priming: **doctor** primes **NURSE**
- Masked
- Unmasked

[3] Ask me about the effects of ageing.

. . . and two types of relationship between primes and targets

We hoped to identify what aspect or component of processing is affected by priming. [3]

- Associate Priming: **doctor** primes **NURSE**
- Masked
- Unmasked
- Identity priming: **nurse** primes **NURSE**

[3] Ask me about the effects of ageing.

. . . and two types of relationship between primes and targets

We hoped to identify what aspect or component of processing is affected by priming. [3]

- Associate Priming: **doctor** primes **NURSE**
- Masked
- Unmasked
- Identity priming: **nurse** primes **NURSE**
- Masked

[3] Ask me about the effects of ageing.

. . . and two types of relationship between primes and targets

We hoped to identify what aspect or component of processing is affected by priming. [3]

- Associate Priming: **doctor** primes **NURSE**
- Masked
- Unmasked
- Identity priming: **nurse** primes **NURSE**
- Masked
- Unmasked

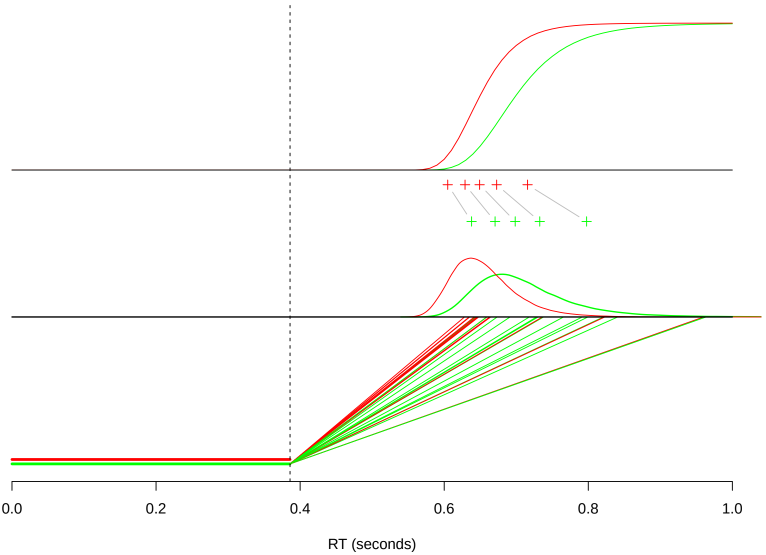
[3] Ask me about the effects of ageing.

RT distributions

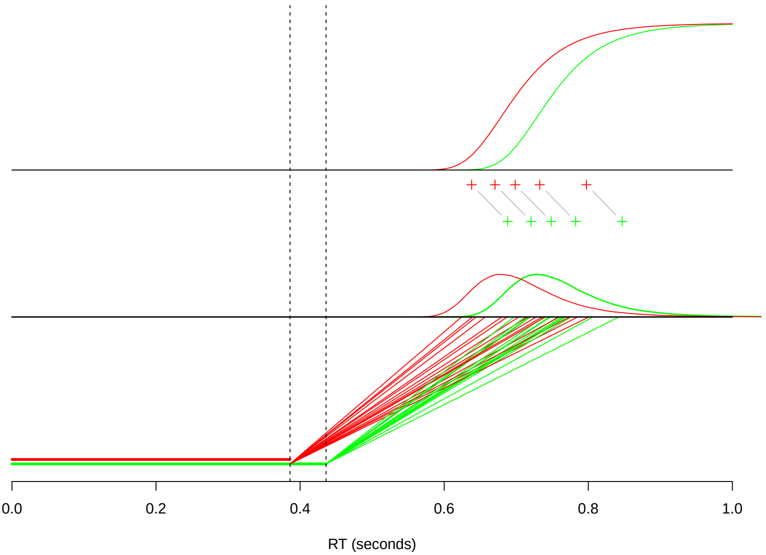
Distributional properties are critical for the model.

[4] Ask me about data visualization and individual variability issues. Exciting techniques for future research.

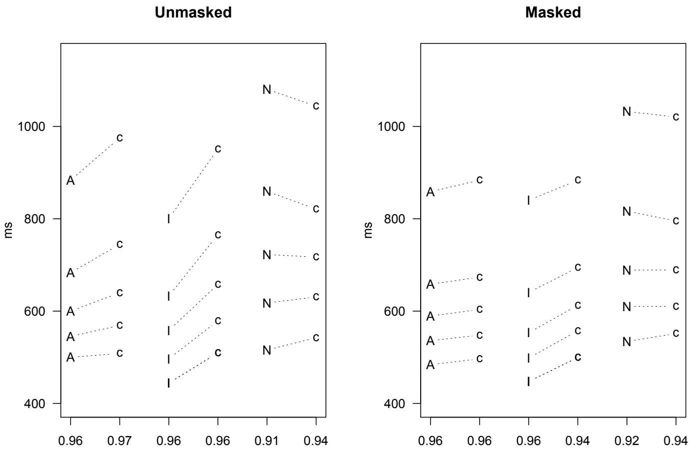
Distributions primer. Drift Rate effect



Distributions primer. Encoding time effect



Data



Unmasked
Masked
ms
1000
800
600
400
A
c
I
N
0.96
0.97
0.96
0.96
0.91
0.94
0.96
0.96
0.96
0.94
0.92
0.94

Models

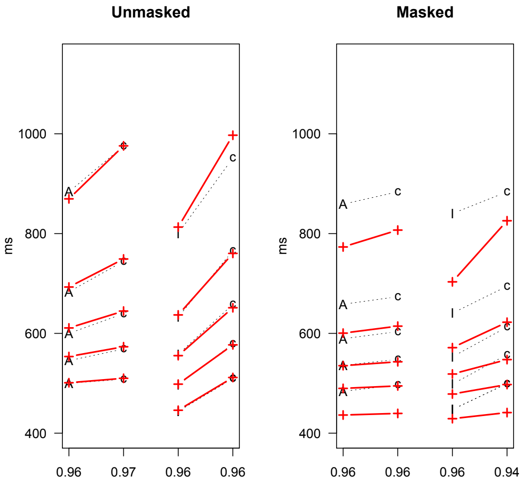
Two models were fit to data:

1. Priming effects occur in the drift rate.
2. Priming effects occur only in the encoding process.

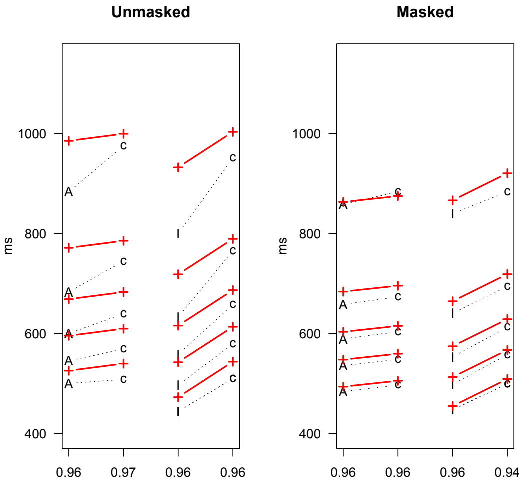
RT distributions and diffusion model

- T_{er} changes relate to shifts in RT distributions.
- Drift changes relate to changes in the spread in RT distributions.
- Drift changes relate to differences in accuracy.

The fits of the drift rate model



The fits of the encoding time model



Interpretation of the fits of the model

- Masked Priming produces changes in the encoding time which are more substantial for identity priming than for semantic priming.
- Unmasked priming creates changes in the rate of accumulation of evidence which is larger for identity priming than for semantic priming.

Looking at a tree from the forest.

Models connect theory to data.

- **Theory.**
  - Single process: All priming is a consequence of "savings" (Bodner & Masson).
  - Differential processes: Masked priming is perceptual, while unmasked priming relates to spreading activation or accessing compounds in memory.
- **Model.**
- **Data.**

Looking at a tree from the forest.

Models connect theory to data.

- **Theory.**
- **Model.** The priming procedure (masked vs. unmasked) changed the locus of the effect. A qualitative change.
- **Data.**

Looking at a tree from the forest.

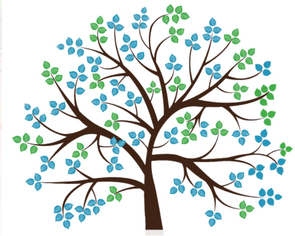
Models connect theory to data.

- **Theory.**
- **Model.**
- **Data.** There is a ubiquitous feature of RTs across hundreds of experiments: -The variance of the RT distribution is highly correlated with the mean RT in almost all two-choice tasks.
  - The tail of the RT distribution is wider for slower conditions than for faster conditions.
  - There are no quantitative models or theories that can explain a shift in the RT distribution AND a change in the scale with the same mechanism.

The trees

1. Lexical decision task
2. Priming
3. **Letter position coding**

- **Dyslexia**
- Braille
- Cross-linguistic research



One trial LDT

Tap your right leg if the next string is an English word.

JUGDE

INRO

Letter position

- Classic word recognition theories were not concerned with letter position coding.
- The Interactive activation model, for example, assumed that *INRO* is more similar to INFO than to *JUGDE* is to JUDGE.

Letter position coding

- We know that letter position encoding is not perfect. (Andrews, 1996; Chambers, 1979; O'Connor & Forster, 1981; Holmes & Ng, 1993; O'Connor & Forster, 1981; Perea, Rosa, & Gomez, 2005; Forster et al., 1987; Whitney, 2001; Davis, 1999)

Theory

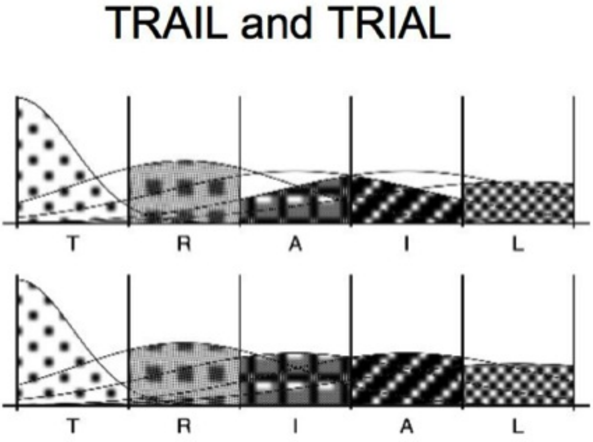
There are two ways to think about letter position coding and why it is a process prone to error.

1. It is a consequence of how the orthography of words is represented in our cognitive system (open bigrams)
2. It is a consequence of perceptual noise not specific to reading.

Model (PsychRev 2008)

The basic assumption of the overlap model is that locations of objects (in our case, letters) are best understood as distributions along a dimension, rather than as precise points.

Model



Beyond the Lab

What happens if we increase the space between letters?

- Bigrams are the same regardless of space.
- Is the position uncertainty reduced?

Dyslexia

Contents lists available at SciVerse ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

The effects of inter-letter spacing in visual-word recognition: Evidence with young normal readers and developmental dyslexics

Manuel Perea [a,*], Victoria Panadero [a], Carmen Moret-Tatay [b], Pablo Gómez [c]

[a] ERI-Lectura, Universitat de València, Valencia, Spain
[b] Universidad Católica de Valencia, Valencia, Spain
[c] DePaul University, Chicago, USA

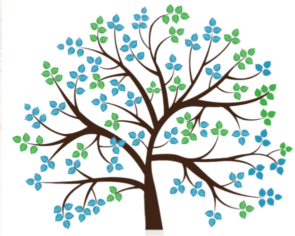
Results

Slight increases in inter-letter spacing improved the readability of texts for dyslexic children (but not for non-dyslexic).

The trees

1. Lexical decision task
2. Priming
3. Letter position coding

- Dyslexia
- **Braille**
- Cross-linguistic research



Two competing theories

1. The JUGDE effect is due to the orthographic representation of words.
2. The JUDGE effect is due to the limitations of the visual system.

Haptic Reading

If some of these effects are visual in origin should they be present in Braille reading?

Braille

OPEN ACCESS Freely available online

PLOS | ONE

Letter Position Coding Across Modalities: The Case of Braille Readers

Manuel Perea[1]*, Cristina García-Chamorro[1], Miguel Martín-Suesta[2], Pablo Gómez[3]

1 ERI-Lectura and Departamento de Metodología, Universitat de València, Valencia, Spain, 2 Organización Nacional de Ciegos (ONCE), Valencia, Spain, 3 DePaul University, Chicago, Illinois, United States of America

No JUGDE effect

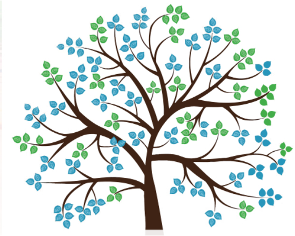
- Unlike experiments in the visual modality, we failed to find any clear signs of transposed-letter confusability effects [4]. In two experiments: single items lexical decision, and sentence reading.
- The effect is quite robust in the visual modality.

[5] Ask me about development for ERP recordings during Braille reading

The trees

1. Lexical decision task
2. Priming
3. Letter position coding

- Dyslexia
- Braille
- **Crosslinguistic research**



Thai vs English (PBR, 2017) [5]

In Thai, there are no spaces between words and no case.

นักศึกษาไปซื้อตะไคร้มาจากตล

[6] Ask me about the effects of literacy

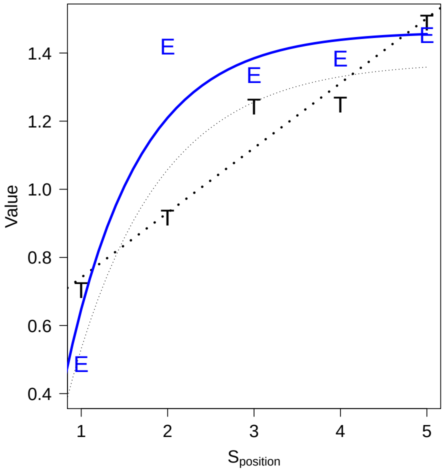
The overlap model

- The position uncertainty for each letter position is a parameter of the model: $s_1 \ldots s_L$.
- With English readers, we found that we can reduce the number of parameters.
  - $s = f(position)$
  - Nicely described as an exponential approach to a limit

With Thai readers

- $s = f(position)$
- Nicely described as an linear function

Comparing the linear vs the exponential approach to a limit



Final word

Good things happen when...

If we think of scientific inference as model comparison, and not just as a set of procedures, we become more mindful, and the next research idea presents itself.

Thank you!

Questions

Frequentist vs. Bayesian debate

Motivation

One of your friends runs an experiment on a priming task and finds a $t(299) = 2.06, p = .04$. Your friend happily goes to your advisor, but your advisor does not seem too impressed. Later, the friend asks you, "Could a p-value in the neighborhood of .04 be more consistent with the H_0 than with the H_1?"

A strange question

It is almost like asking:

Are five popsicles a lot?

We can specify a hypothesis H_δ

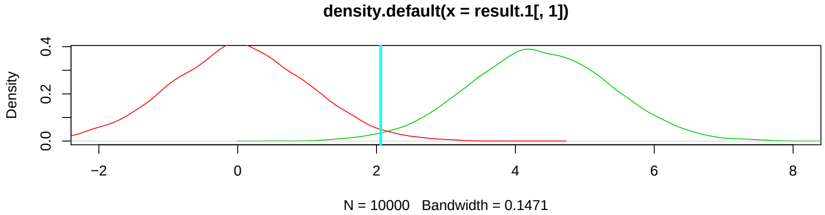
Example

```r
nsims <- 10000
result.1 <- matrix(ncol = 2, nrow = nsims)
result.0 <- matrix(ncol = 2, nrow = nsims)

# Number of subjects
N = 300
# Effect size
delta = 0.25
for (i in 1:nsims) {
    x.1 = rnorm(N, delta, 1)
    x.0 = rnorm(N, 0, 1)
    result.1[i, 2] = t.test(x.1, mu = 0)$p.value
    result.1[i, 1] = t.test(x.1, mu = 0)$statistic

    result.0[i, 2] = t.test(x.0, mu = 0)$p.value
    result.0[i, 1] = t.test(x.0, mu = 0)$statistic
}
```

Given and $N = 300$, a t of 2.06 is almost equally as consistent with the H_0 than with $H_{\delta=.25}$



A gateway to Bayes factors

- This ratio is not that useful in reality as it does not seem judicious to pick a point H_1.
- I picked 0.25 just for pedagogical reasons as it worked well with $N = 300$
- But my prior belief about the effect size is likely not that precise.

Prior beliefs about H_1

- Instead, my prior belief might be that there is a real non-zero effect, but I am unsure of its size or even direction.
- Different researchers might have different *priors*

Speed Dating

Selective vs. Unselective Romantic Desire: Not All Reciprocity is Created Equal. Eastwick et al.

Lets consider two pools of daters

- x, y
- a, b

Speed dates

- a meets x
- a meets y
- b meets x
- b meets y

And each member of the pair rates their new friend.

Two forms of correlation

- Generalized
- Dyadic

Dyadic

The correlation of the ratings across each speed-date:

->	<-
a to x	x to a
a to y	y to a
b to x	x to b
b to y	y to b

Generalized

The correlation between the average rating given and received across all daters:

Given	Received
a to x,y	x,y to a
b to x,y	x,y to b
x to a,b	a,b to x
x to a,b	a,b to x

Their finding

- Dyadic ratings are positively correlated.
- But people that give high ratings tend to get low ratings.

Their claim

"... selectivity is an aphrodisiac."

Note, however, that dater x, didn't see the interaction between a and y.

I don't buy it...

- Multi-agent model with
  $rating_{a,x} = \Psi(selectivitiy_x, selectivity_y, attr_x)$

I don't buy it...

- Multi-agent model with $rating_{a,x} = \Psi(selectivitiy_x, selectivity_y, attr_x)$
- We could not get it to produce the dyadic/generalized dissociation.

Instead (Gomez & Erber)

- Theory: We like people that match our own attractivenes, and people more attractive than ourselves.

Instead (Gomez & Erber)

- Theory: We like people that match our own attractivenes, and people more attractive than ourselves.
- Model: Resonance model of attraction.

Instead (Gomez & Erber)

- "match is an aphrodisiac"

Instead (Gomez & Erber)

- "match is an aphrodisiac"
- If you are a cynic: "... narcissism is more important than selectivity."

Ageing

Ageing

- Older adults slow down in most tasks.
- LDT, memory, etc..
- A diffusion model analysis revealed that in the case of LDT, a lot of the slow down could be attributed to conservative criteria setting.

Data visualization and individual variability issues

Consider an experiment with many observers and many items per condition in which RT is the DV.

- IAT (stereotype consistent vs inconsistent).
- Stroop (color consistent vs inconsistent).
- Priming (primed vs control).

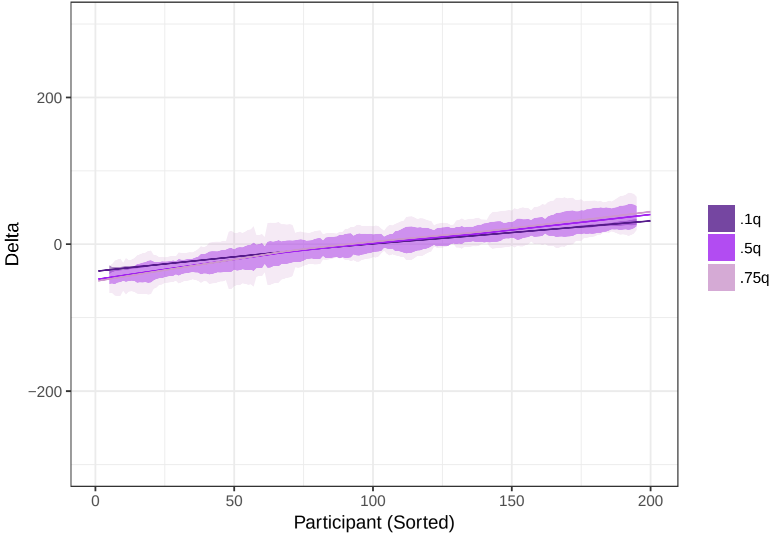
Two forms of variability

- Across participants.
- Across trials.

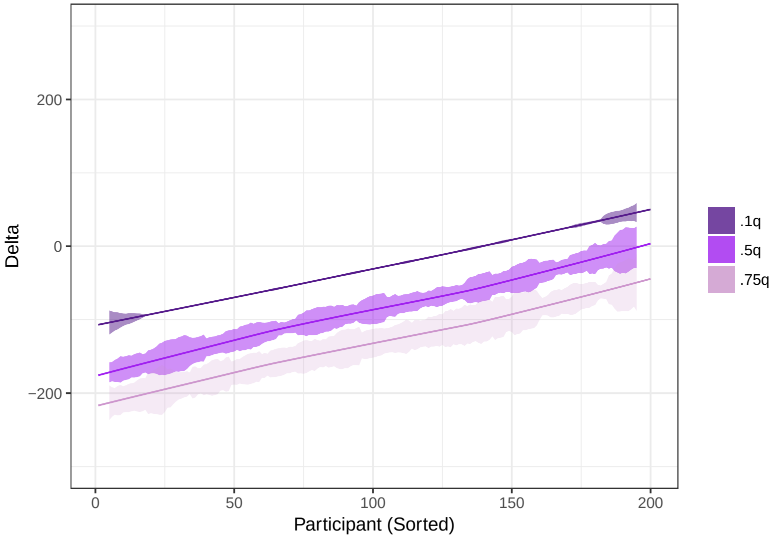
EDA

I am a big believer in Exploratory Data Analysis as the first step in mindful statistical inference.

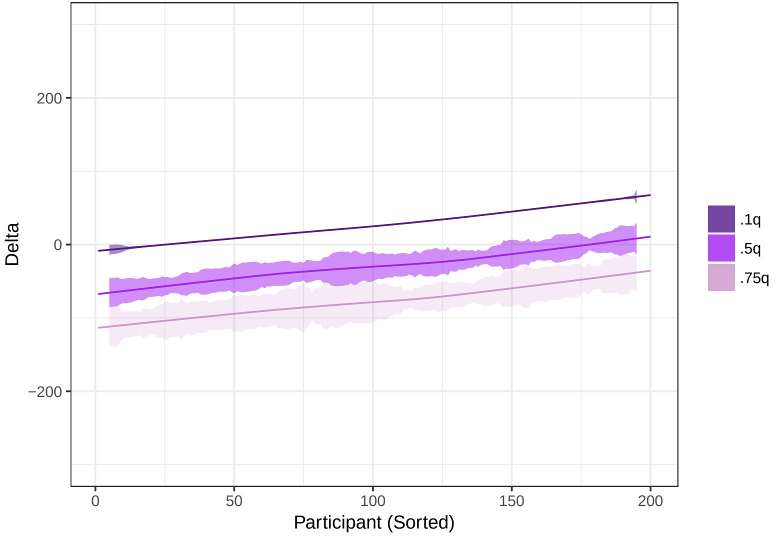
Examples



Examples



Examples



An ignored set of questions

- Does everybody IAT?
- Does everybody prime?
- Does everybody flanker?
- Does everybody show word frequency effects?

Research on Braille Reading

The need for research into Braille reading is well documented:

- Large rates of illiteracy for blind people with large quality of life consequences: $< 10\%$ of the blind population in the U.S. can read Braille.
- Important consequences for our understanding of reading.

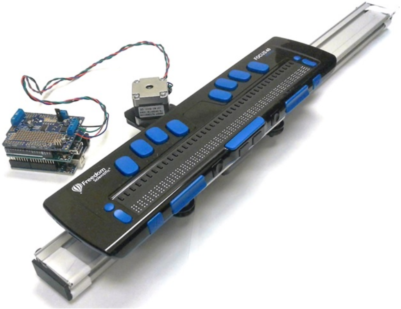
Visual modality vs haptic modality

Researchers of visual word recognition have an important tool at their disposal: they can control the timing of presentation of visual stimuli.

Not so in the haptic modality…

- Researchers need to have control over when the participants are exposed to the critical stimuli.
- EEG recording = EEG signal + Artifacts.
    - Signals unrelated to the cognitive activity that obscure the signal of interest.
- Electromyography (EMG).
    - Muscle movement-related signals.
    - Larger size than effects of cognitive tasks.

Hardware



Fitting can be done very procedurally.

When the model is fit to data, one needs to have these four questions in mind:

1. Does it make sense?
2. What is being diffused? In other words, what is the evidence that is being accumulated?
3. Within the diffusion model framework, what are the possible mechanisms in play?
4. The model is not a model of XXXXX, how does it map into my research question?

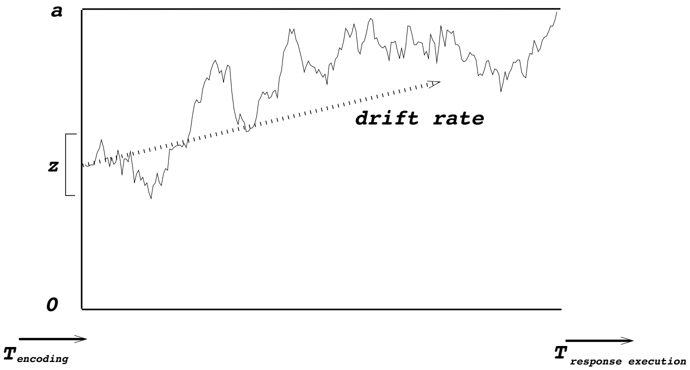
Without answers to these questions, diffusion modeling becomes useless even if there is output from the fitting algorithm.

Diffusion Modeling

Advantages of the diffusion model

- It has been successfully applied to many 2-choice tasks.
- It can produce predictions concerning RT distributions and accuracy levels.
- It provides us with a coherent account of the relationship between latency and accuracy.
- The model can provide us with estimates of the impact of manipulation on the underlying processes.

The diffusion model



The diffusion model

Symbol	Parameter
a	Boundary separation
z	Starting point
s_z	Variability in starting point across trials
T_{er}	Non-decision component of response time
s_t	Variability in non-decision component of response time
ν	Drift rate
η	Variability in drift rate across trials
s	Variability in drift within each trial (SD)