Explaining item-wise variability in Moses illusions

Hanna Muller, Philip Resnik,
Colin Phillips

Answer the question (or, if you can’t, why not?)

Answer the question (or, if you can’t, why not?)

What is the name of the holiday during which children dress up and walk from door to door to give candy?

Answer the question (or, if you can't, why not?)

What is the name of the holiday during which children dress up and walk from door to door to **give** candy?

Shallow processing?

What is the name of the holiday during which children dress up and walk from door to door to **give** candy?

Barton & Sanford (1993), Sanford & Sturt (2002), Hannon & Daneman (2004), Sturt, Sanford, Stewart, & Dawydiak, (2004), Sanford & Graesser (2006)

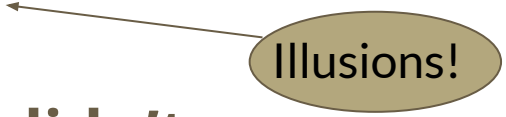
Aside: you're not *always* shallow

*The authors that **no** critics recommended have ***ever*** written a best-selling novel.

*The authors that the critics **didn't** recommend have ***ever*** written a best-selling novel.

Aside: you're not *always* shallow

*The authors that **no** critics recommended have ***ever*** written a best-selling novel.



*The authors that the critics **didn't** recommend have ***ever*** written a best-selling novel.

Aside: you're not *always* shallow

*The authors that **no** critics recommended have ***ever*** written a best-selling novel.

Illusions!

*The authors that the critics **didn't** recommend have ***ever*** written a best-selling novel.

No illusions!

Shallow processing?

What is the name of the holiday during which children dress up and walk from door to door to **give** candy?

Shallow processing?

What is the name of the holiday during which children dress up and walk from door to door to give candy?

Why "Moses" illusions?

What is the name of the holiday during which children dress up and walk from door to door to **give** candy?

How many animals of each kind did **Moses** bring on the ark?

Erickson & Mattson (1981)

Does knowledge matter?

How many animals of each kind did **Moses** bring on the ark?

Does knowledge matter?

How many animals of each kind did **Moses** bring on the ark?

🤷‍♀️

Does knowledge matter?

What is the name of the holiday during which children dress up and walk from door to door to **give** candy?

How many animals of each kind did **Moses** bring on the ark?

Does knowledge matter?

What is the name of the holiday during which children dress up and walk from door to door to **give** candy?

How many animals of each kind did **Moses** bring on the ark?

Outline

1. Some errors are harder to detect than others (Experiment 1)
2. What factors influence the probability of an illusion?
    a. Lexical similarity
    b. Attention to bottom-up lexical processing (Experiments 2 & 3)

Shallow processing

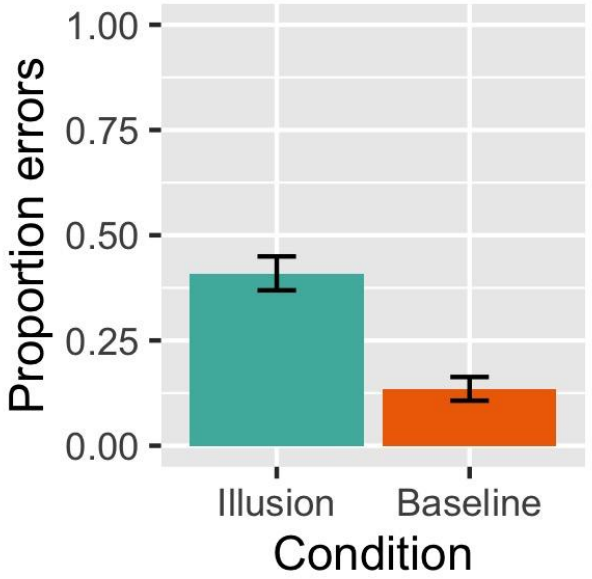
- Sometimes we mean things like not selecting one of multiple possible parses / meanings (quantifier scope, polysemy)
- Sometimes we mean selecting prematurely (garden paths)
- Sometimes we mean using more non-linguistic information than we might have expected (head injury)
- Sometimes we mean using less non-linguistic information than we might have expected (role reversals)
- Sometimes we mean using “heuristics” rather than “algorithms” to get from a form to a meaning (“good enough” parsing, agent-first bias in passives, local coherence)
- Sometimes we mean only representing a piece of linguistic input at a certain level of abstraction/representation, when other representations are available (proof reading vs comprehension)

Variability in the illusion

Some illusions can be “turned on and off”

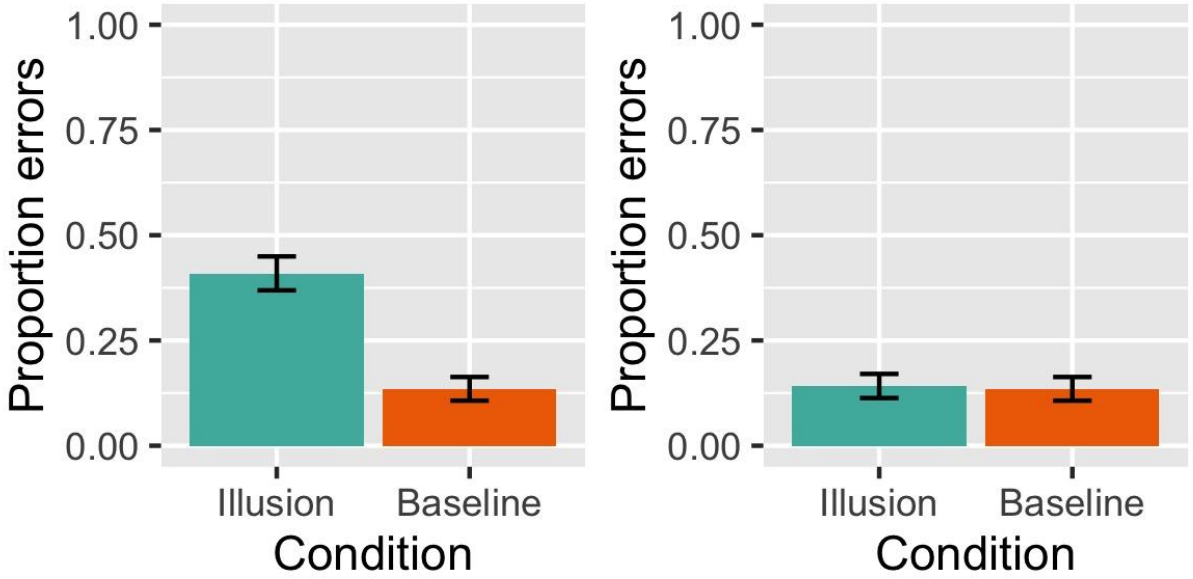
Variability in the illusion

Some illusions can be “turned on and off” - negative polarity illusions



Variability in the illusion

Some illusions can be “turned on and off” - negative polarity illusions are sensitive to subtle changes



Variability in the illusion

Some illusions can be “turned on and off” - negative polarity illusions are sensitive to subtle changes

What makes Moses illusions stronger or weaker?

Variability in the illusion

50 stimuli drawn from prior literature

- What is the name of the holiday during which children dress up and walk from door to door to **give** candy?
- What is the name of the holiday during which children dress up and walk from door to door to **collect** candy?

Variability in the illusion

50 stimuli drawn from prior literature

- What is the name of the holiday during which children dress up and walk from door to door to **give** candy?
- What is the name of the holiday during which children dress up and walk from door to door to **collect** candy?

100 Mechanical Turk workers

Variability in the illusion

50 stimuli drawn from prior literature

- What is the name of the holiday during which children dress up and walk from door to door to **give** candy?
- What is the name of the holiday during which children dress up and walk from door to door to **collect** candy?

100 Mechanical Turk workers

2 part task:

Variability in the illusion

50 stimuli drawn from prior literature

- What is the name of the holiday during which children dress up and walk from door to door to **give** candy?
- What is the name of the holiday during which children dress up and walk from door to door to **collect** candy?

100 Mechanical Turk workers

2 part task: answer questions

Variability in the illusion

50 stimuli drawn from prior literature

- What is the name of the holiday during which children dress up and walk from door to door to **give** candy?
- What is the name of the holiday during which children dress up and walk from door to door to **collect** candy?

100 Mechanical Turk workers

2 part task: answer questions, then knowledge check

Part One

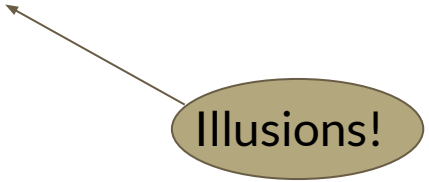
What is the name of the holiday during which children dress up and walk from door to door to give candy?

- ❏
- ❏ I cannot answer
- ❏ I don’t know

Part One

What is the name of the holiday during which children dress up and walk from door to door to give candy?

- ❑ `Halloween`
- ❑ I cannot answer
- ❑ I don’t know



Part Two

The name of the holiday during which children dress up and walk from door to door to XXX candy is Halloween.

- ❏ steal
- ❏ offer
- ❏ collect
- ❏ give
- ❏ none of these
- ❏ I don’t know

Example stimuli

What is the name of the raised bumps on paper which enable deaf people to read?

Example stimuli

What is the name of the raised bumps on paper which enables deaf people to read?

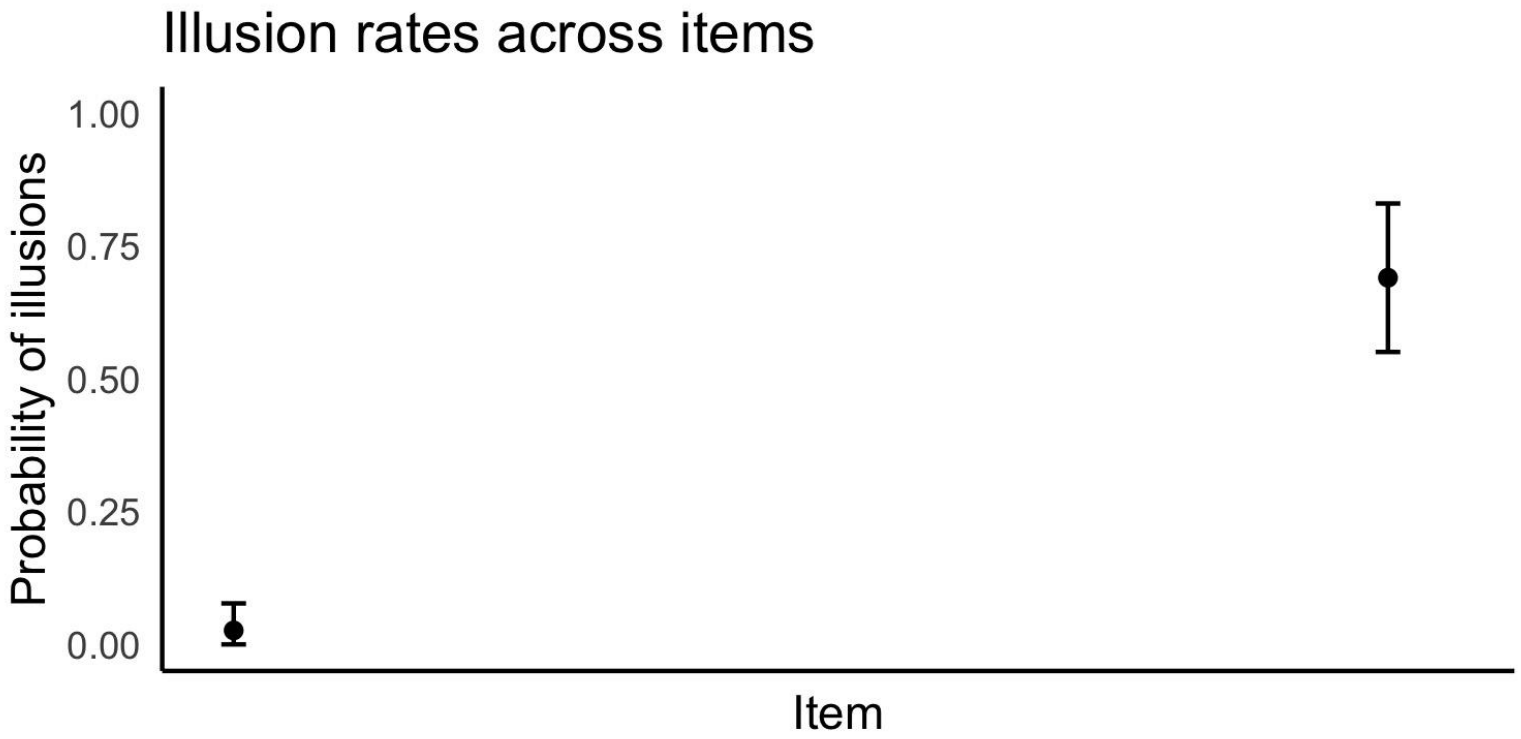
How did painter Vincent van Gogh lose his eye during his life?

Example stimuli

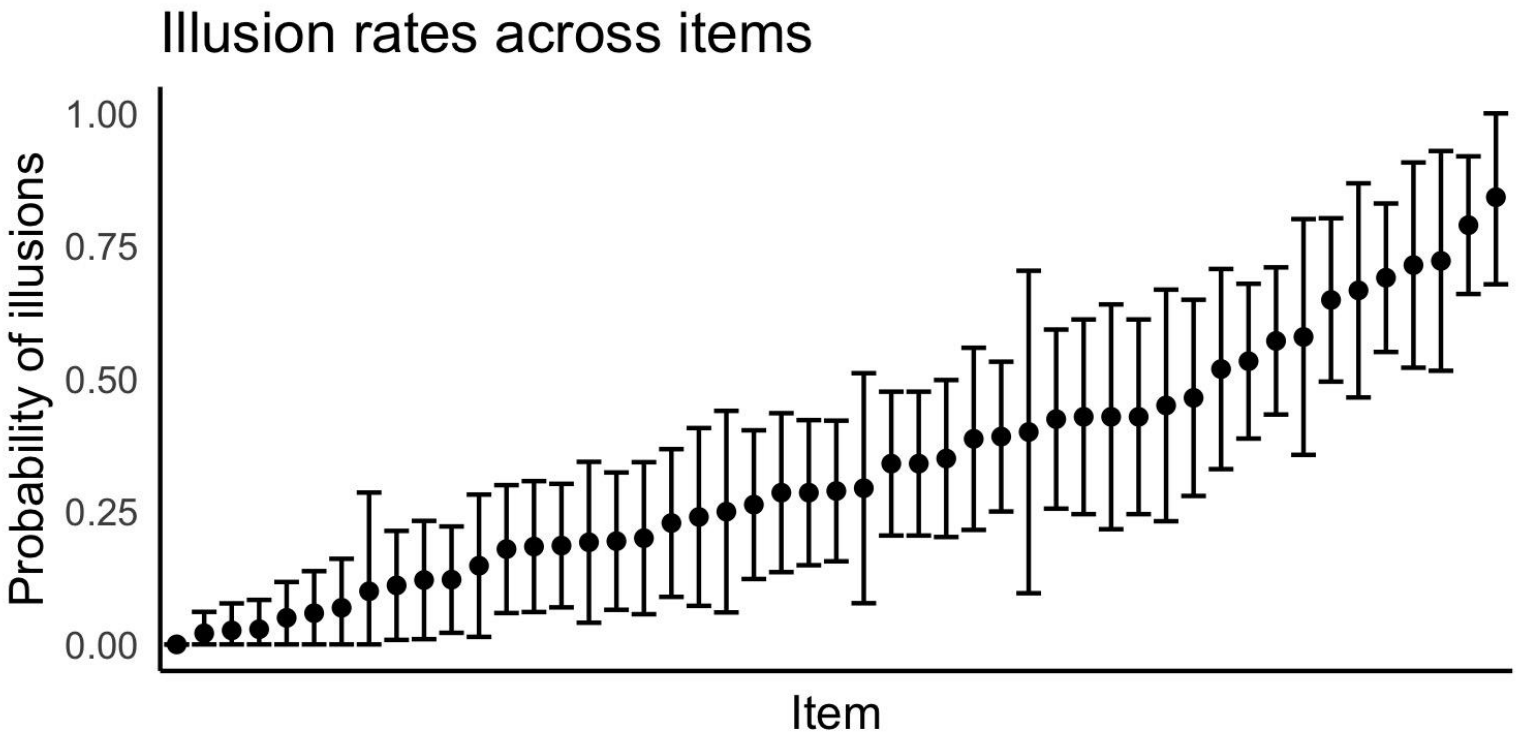
What is the name of the raised bumps on paper which enables **deaf** people to read?

How did painter Vincent van Gogh lose his **eye** during his life?

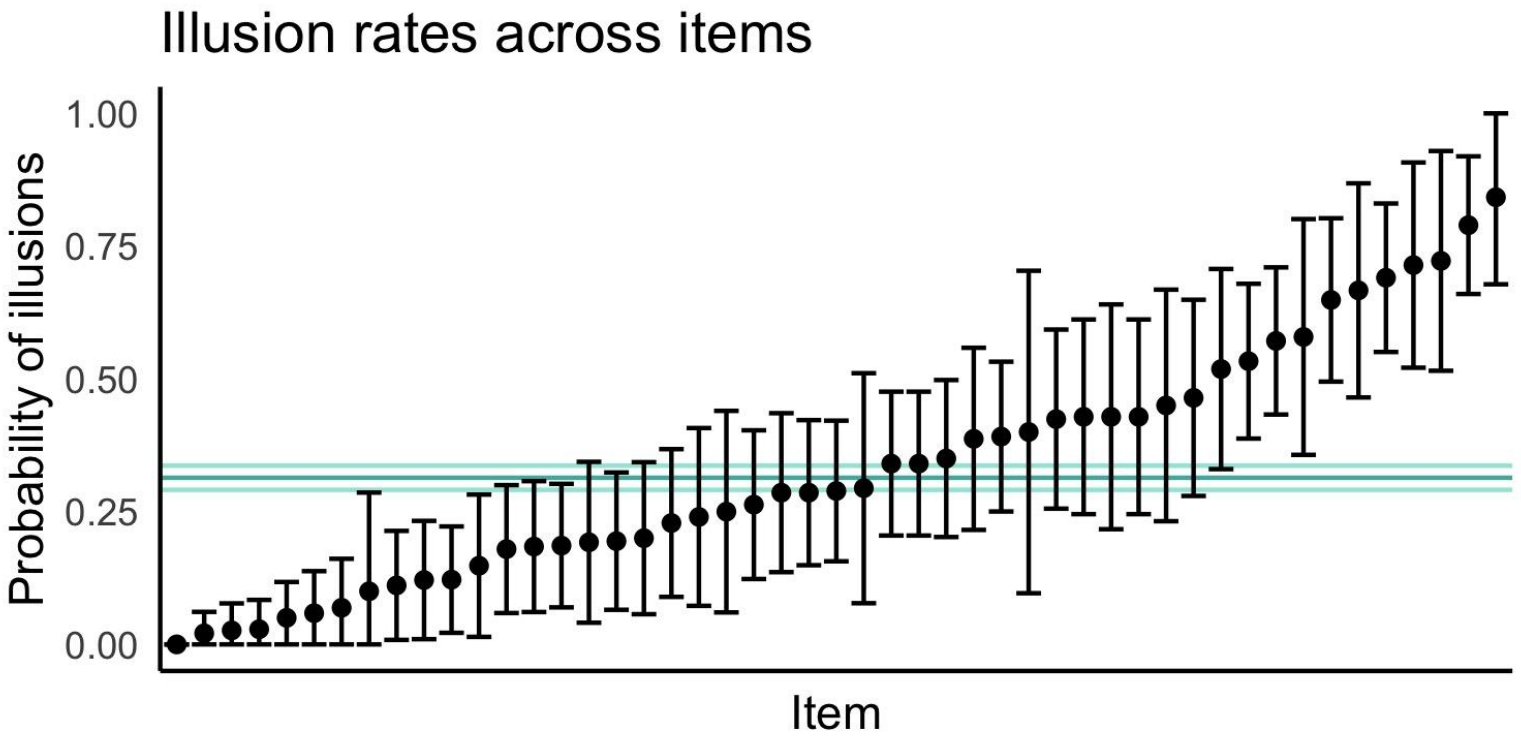
Results


Illusion rates across items
Probability of illusions
1.00
0.75
0.50
0.25
0.00
Item

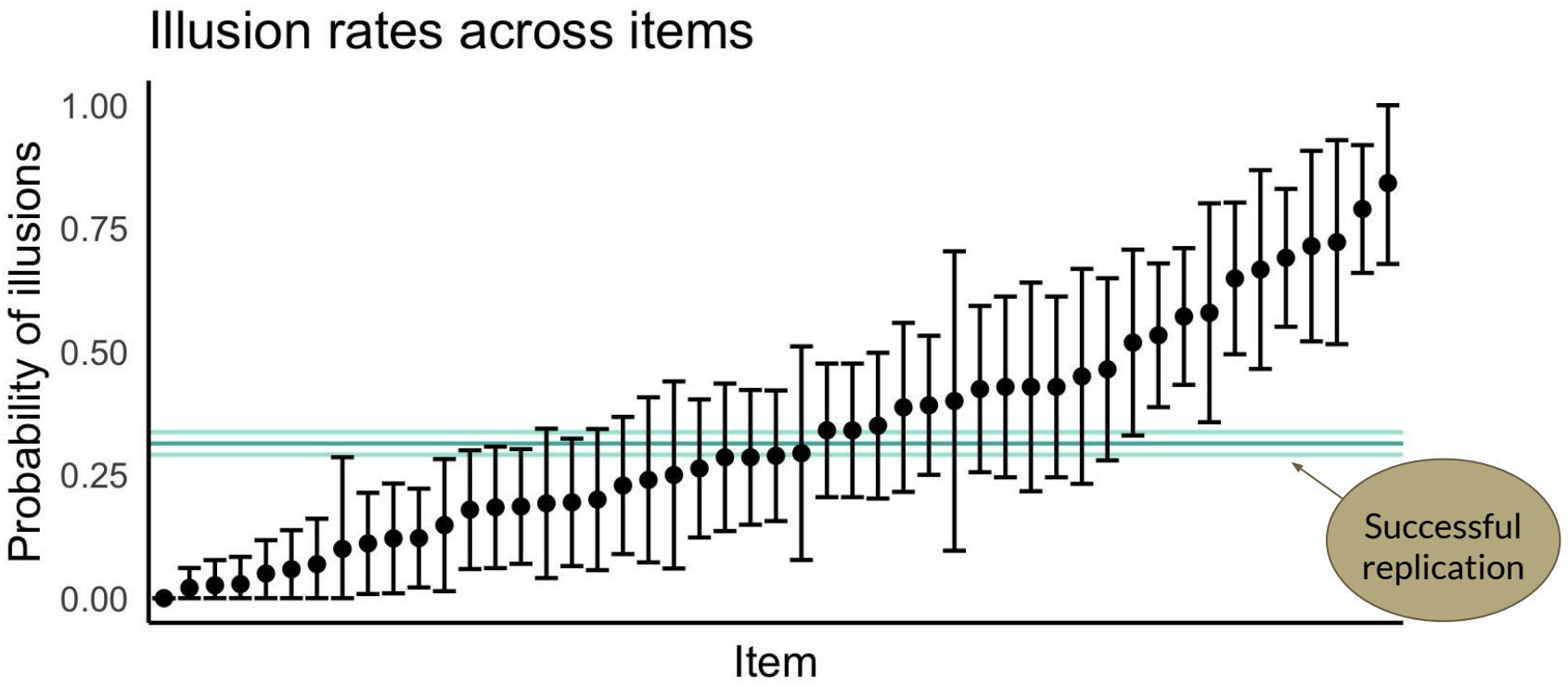
Results


Illusion rates across items
Probability of illusions
1.00
0.75
0.50
0.25
0.00
Item

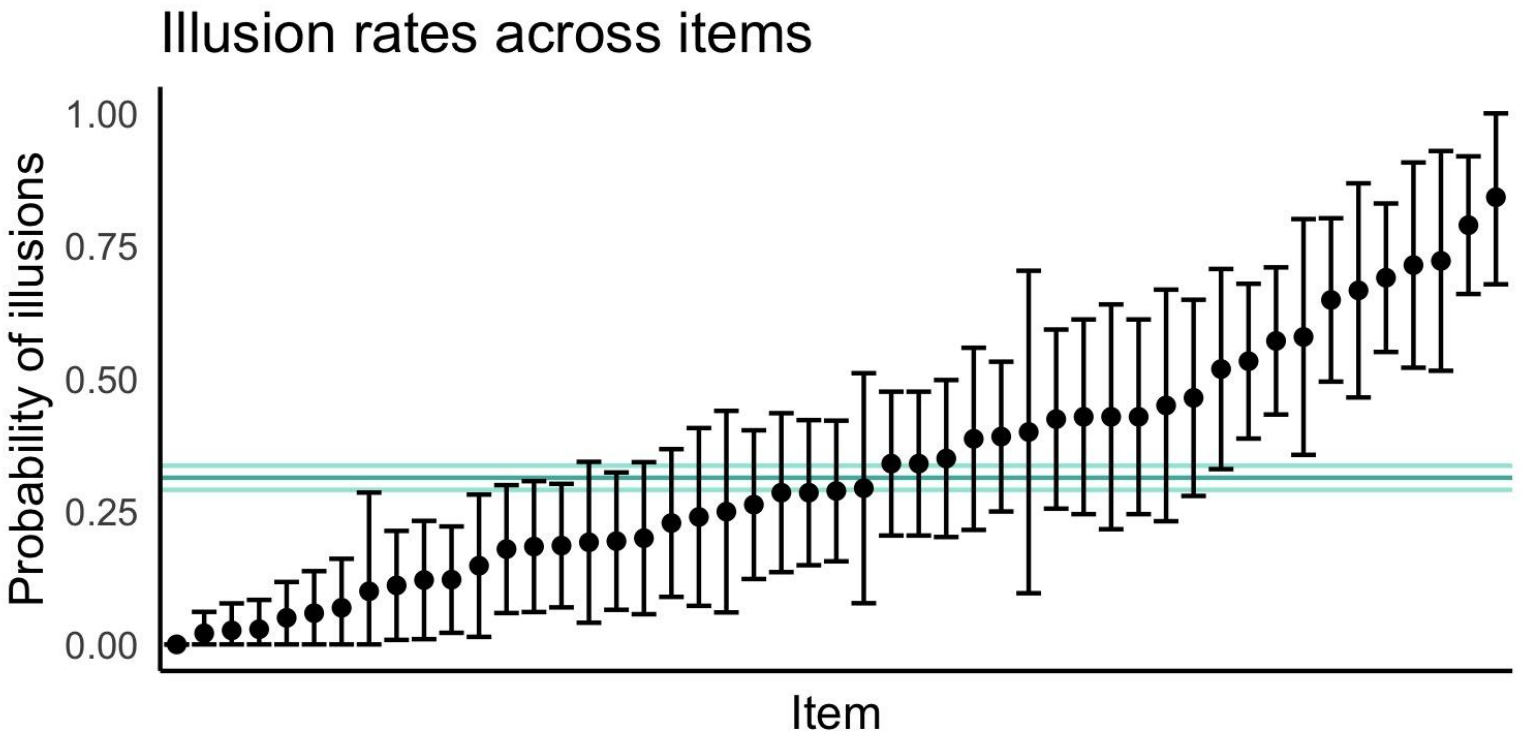
Results


Illusion rates across items
Probability of illusions
1.00
0.75
0.50
0.25
0.00
Item

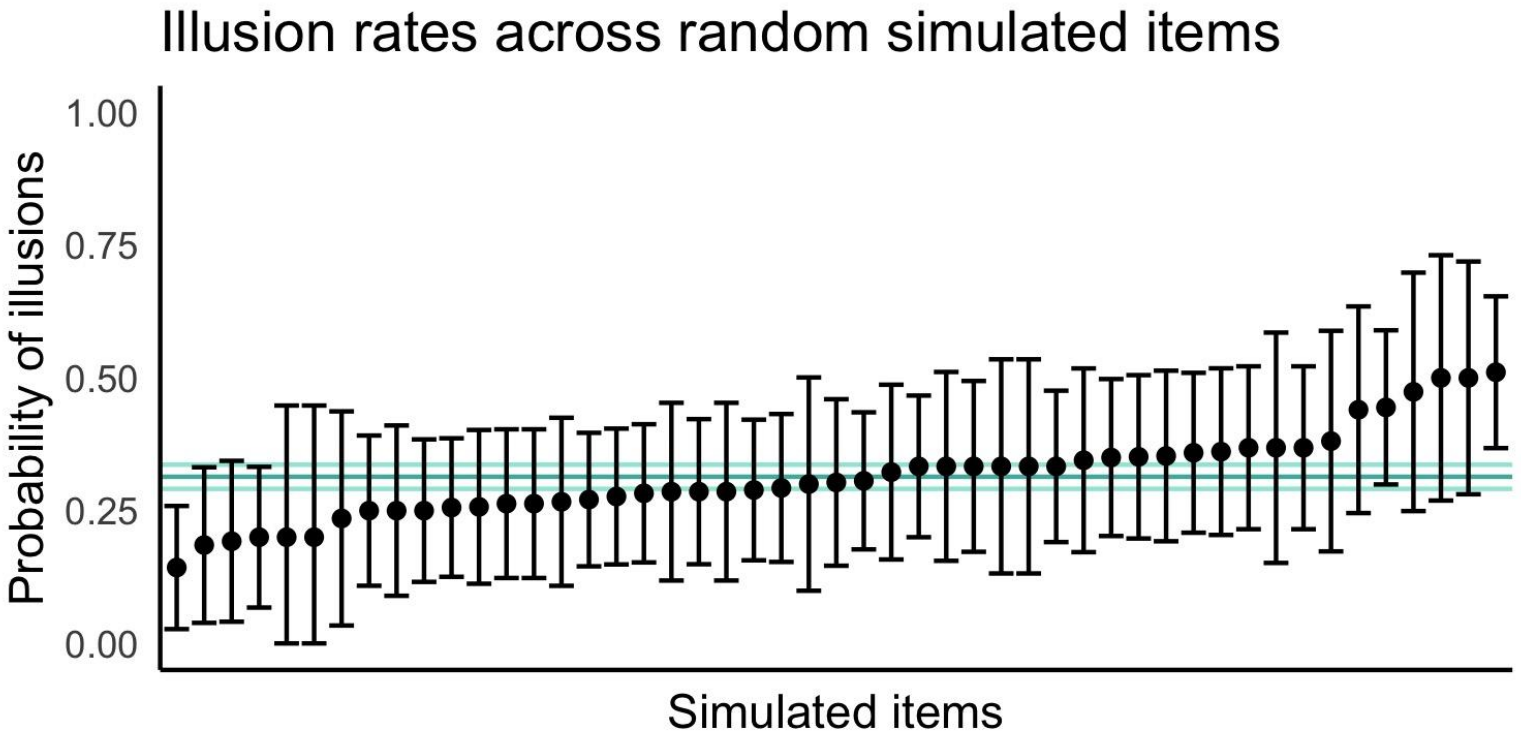
Results


Illusion rates across items
Probability of illusions
1.00
0.75
0.50
0.25
0.00
Item
Successful replication

Results


Illusion rates across items
Probability of illusions
1.00
0.75
0.50
0.25
0.00
Item

Results


Illusion rates across random simulated items
Probability of illusions
1.00
0.75
0.50
0.25
0.00
Simulated items

Why do Moses illusions occur?

- Attention shifted away from bottom-up lexical processing

Why do Moses illusions occur?

- Attention shifted away from bottom-up lexical processing
  - Do items that give you more information early on make you more likely to experience illusions?

Why do Moses illusions occur?

- Attention shifted away from bottom-up lexical processing
  - Do items that give you more information early on make you more likely to experience illusions?

- Anomalous lexical item has some semantic features in common with what you expected (or should have seen instead)

Why do Moses illusions occur?

- Attention shifted away from bottom-up lexical processing
  - Do items that give you more information early on make you more likely to experience illusions?

- Anomalous lexical item has some semantic features in common with what you expected (or should have seen instead)
  - Do items that have more semantic overlap between the anomalous word and the intended word make you more likely to experience illusions?

Why do Moses illusions occur?

- Attention shifted away from bottom-up lexical processing
  - Do items that give you more information early on make you more likely to experience illusions?

- Anomalous lexical item has some semantic features in common with what you expected (or should have seen instead)
  - Do items that have more semantic overlap between the anomalous word and the intended word make you more likely to experience illusions?

Explaining the variability

- How many animals of each kind did **Moses** take on the ark?
- How many animals of each kind did **Nixon** take on the ark?

Explaining the variability

Words that are highly dissimilar to the word they're swapped in for are much less likely to yield illusions

- How many animals of each kind did **Moses** take on the ark?
- How many animals of each kind did **Nixon** take on the ark?

Explaining the variability

Words that are highly dissimilar to the word they're swapped in for are much less likely to yield illusions

- How many animals of each kind did **Moses** take on the ark?
- How many animals of each kind did **Nixon** take on the ark?

Erickson & Mattson (1981), Van Oostendorp & De Mul (1990), Van Jaarsveld, Dijkstra, & Hermans (1997), Cook, Walsh, Bills, Kircher, & OBrien (2016)

Explaining the variability

Words that are highly dissimilar to the word they're swapped in for are much less likely to yield illusions

- How many animals of each kind did **Moses** take on the ark?
- How many animals of each kind did **Nixon** take on the ark?

Does variability in word similarity accross our items explain the differing illusion rates?

Erickson & Mattson (1981), Van Oostendorp & De Mul (1990), Van Jaarsveld, Dijkstra, & Hermans (1997), Cook, Walsh, Bills, Kircher, & OBrien (2016)

Measuring semantic similarity

Word2vec

- Assign a vector to each word based on the distribution of contexts in which that word appears
- Words that appear in similar contexts are likely to be semantically similar

Measuring semantic similarity

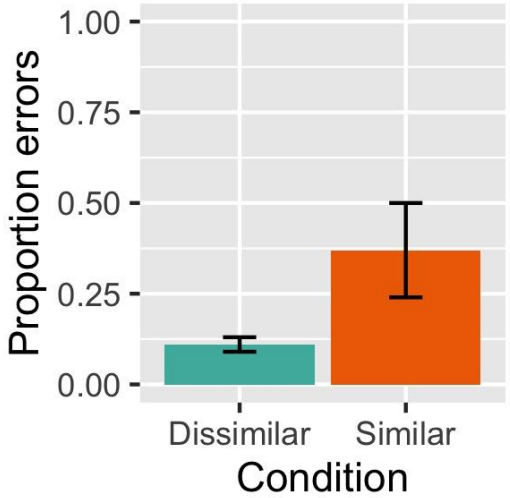
Cook, Walsh, Bills, Kircher, & OBrien (2016)

- Halloween is the holiday when children go door to door **giving out** candy and treats.
- Halloween is the holiday when children go door to door **stealing** candy and treats.

Measuring semantic similarity

Cook, Walsh, Bills, Kircher, & OBrien (2016)

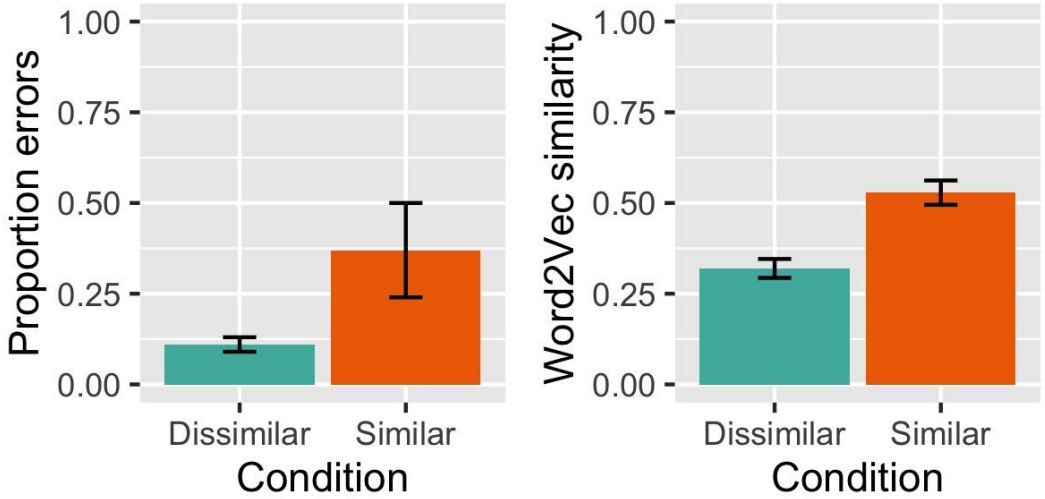
- Halloween is the holiday when children go door to door **giving out** candy and treats.
- Halloween is the holiday when children go door to door **stealing** candy and treats.



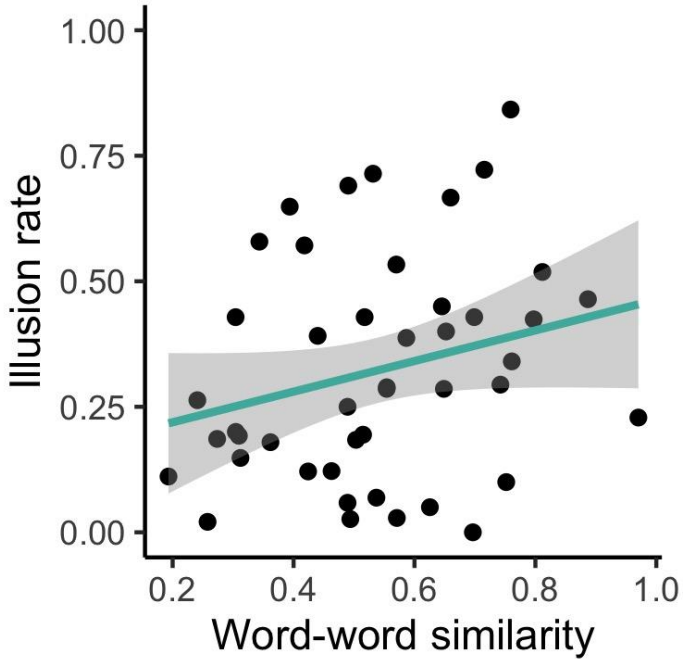
Measuring semantic similarity

Cook, Walsh, Bills, Kircher, & OBrien (2016)

- Halloween is the holiday when children go door to door **giving out** candy and treats.
- Halloween is the holiday when children go door to door **stealing** candy and treats.



Semantic similarity (word2vec)



Semantic similarity

Weakly predicts illusion item-to-item illusion rates

Semantic similarity

Weakly predicts illusion item-to-item illusion rates

(Though more sophisticated measures of similarity may correlate better)

Why do Moses illusions occur?

- Attention shifted away from bottom-up lexical processing
  - Do items that give you more information early on make you more likely to experience illusions?

- Anomalous lexical item has some semantic features in common with what you expected (or should have seen instead)
  - Do items that have more semantic overlap between the anomalous word and the intended word make you more likely to experience illusions?

Why do Moses illusions occur?

- Attention shifted away from bottom-up lexical processing
  - Do items that give you more information early on make you more likely to experience illusions?

- Anomalous lexical item has some semantic features in common with what you expected (or should have seen instead)
  - Do items that have more semantic overlap between the anomalous word and the intended word make you more likely to experience illusions?

Attention

Which

Attention

Which British monarch

Attention

Which British monarch formally opened

Attention

Which British monarch formally opened the Olympic

Attention

Which British monarch formally opened the Olympic

Attention

Which British monarch formally opened the Olympic

Attention

Which British monarch formally opened the Olympic winter games

Attention

Which British monarch formally opened the Olympic winter games in London

Attention

Which British monarch formally opened the Olympic winter games in London in 2012?

Attention

Which British monarch formally opened the Olympic **winter** games in London in 2012?

Attention

Shifting to answer generation leads to less attention for lexical representations

- Later substitutions should be harder to detect

van Jaarsveld Dijkstra, & Hermans (1997)

Attention

Shifting to answer generation leads to less attention for lexical representations

- Later substitutions should be harder to detect

Other ways of directing attention to various parts of the sentence

Attention

Shifting to answer generation leads to less attention for lexical representations

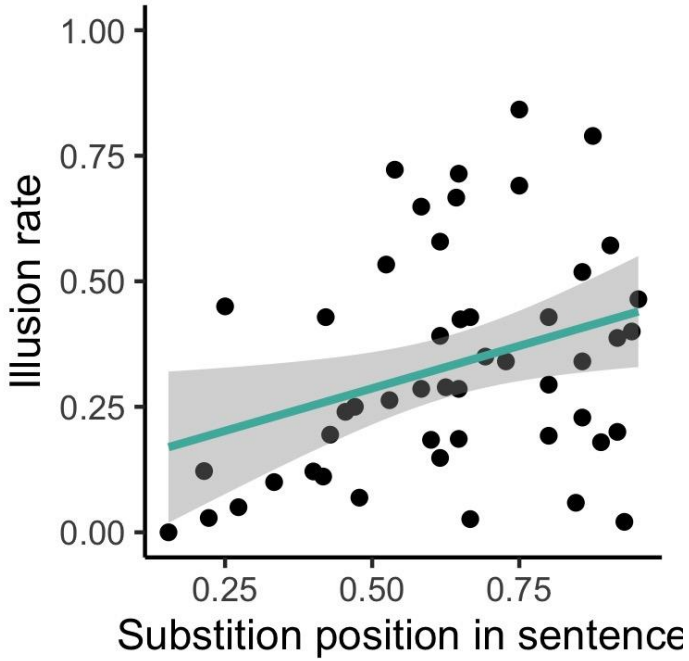
- Later substitutions should be harder to detect

Other ways of directing attention to various parts of the sentence

- Linguistic focus, cleft structures

Bredart & Modolo (1988), Sturt, Sanford, Stewart, & Dawydiak (2004), Raposo & Marques (2013)

Position effects



Attention

Four broad categories of evidence we're looking for:

1. Evidence that for some stimuli but not others, comprehenders have enough information that they *could* stop prioritizing the bottom-up input prior to the substitution
2. Evidence that when there's enough information, comprehenders *do* shift attention
3. Evidence that shifts in attention lead to shallower representations of incoming information (i.e. more illusions)
4. Evidence for what specific level of representation is reached when shallow processing strategies are used

Attention

Four broad categories of evidence we're looking for:

1. Evidence that for some stimuli but not others, comprehenders have enough information that they *could* stop prioritizing the bottom-up input prior to the substitution
2. Evidence that when there's enough information, comprehenders *do* shift attention
3. Evidence that shifts in attention lead to shallower representations of incoming information (i.e. more illusions)
4. Evidence for what specific level of representation is reached when shallow processing strategies are used

How much do you know before the anomaly?

Contemporary language models can generate believable sentences (using a lead-in sentence fragment)

Unclear whether the mechanisms they use are a reasonable approximation of human mechanisms, but their performance is a reasonable approximation of human-generated corpora

Language models have the key advantage that they don't just give you one predicted next word, they assign a probability to every word in the lexicon

Language models

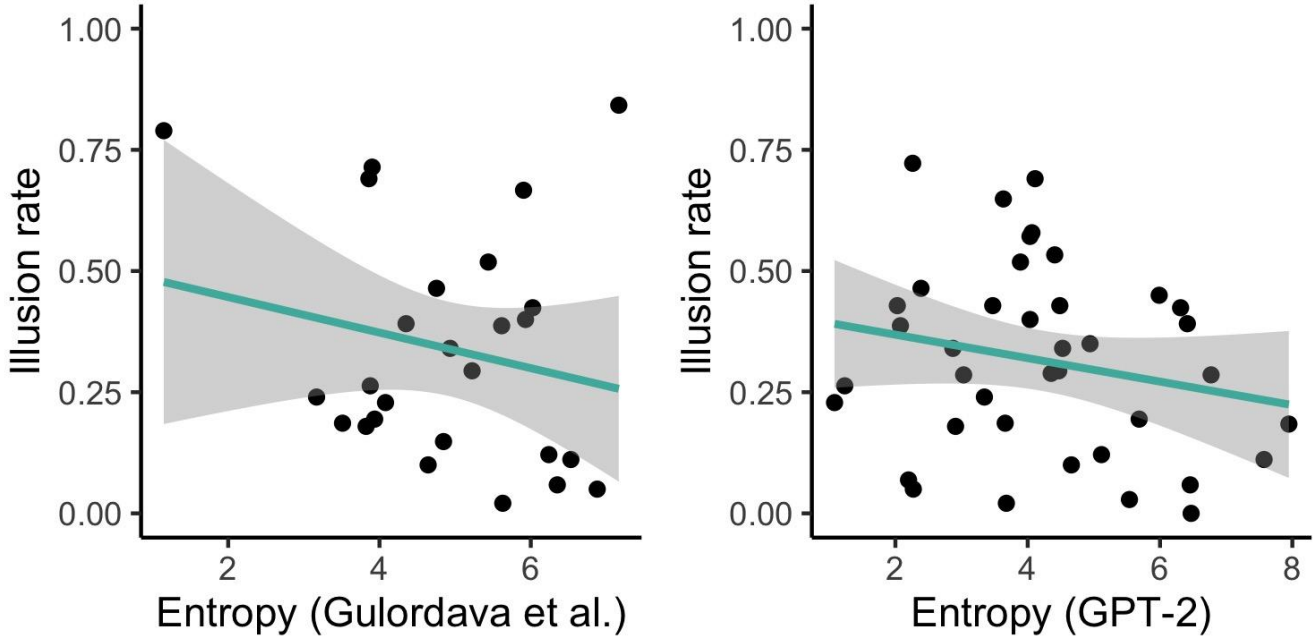
Entropy - higher means less certainty about what's coming next

So we expect that the stimuli with lower entropy (you have a good idea about what's coming next) will have higher illusion rates (because you know what's coming, you don't have to attend to it as much)

Two models: LSTM that has been used to approximate human language behavior, and GPT-2

Gulordava et al, Van Schijndel & Linzen, GPT-2

Language model predictions



Knowing the next word vs knowing the answer

What is the name of the holiday during which children dress up and walk from door to door to

- ask for candy
- beg for candy
- door
- get candy
- receive candy
- trick or treat
- collect candy from strangers

Cloze task (Experiment 2)

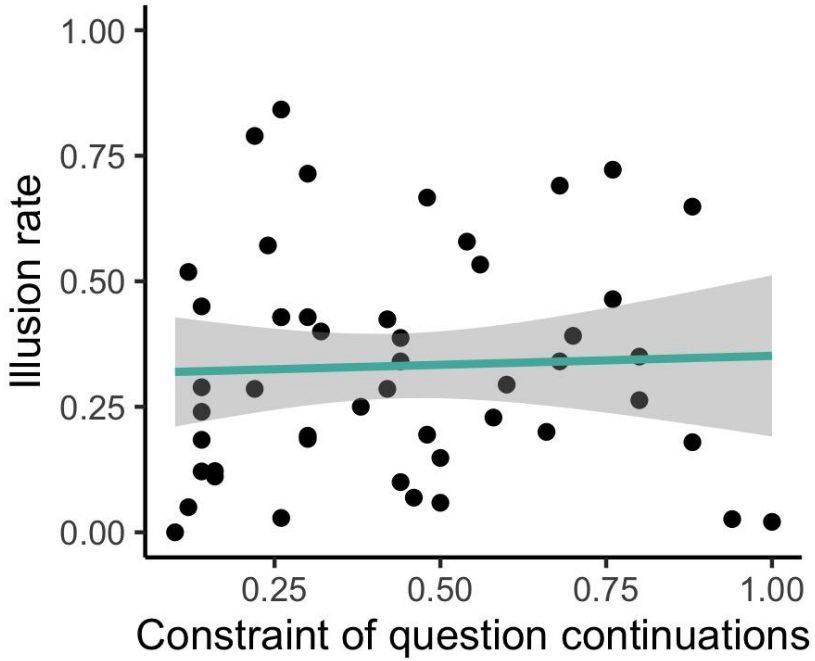
50 Mechanical Turk workers

The same 50 stimuli as experiment 1, but cut off just before the substitution

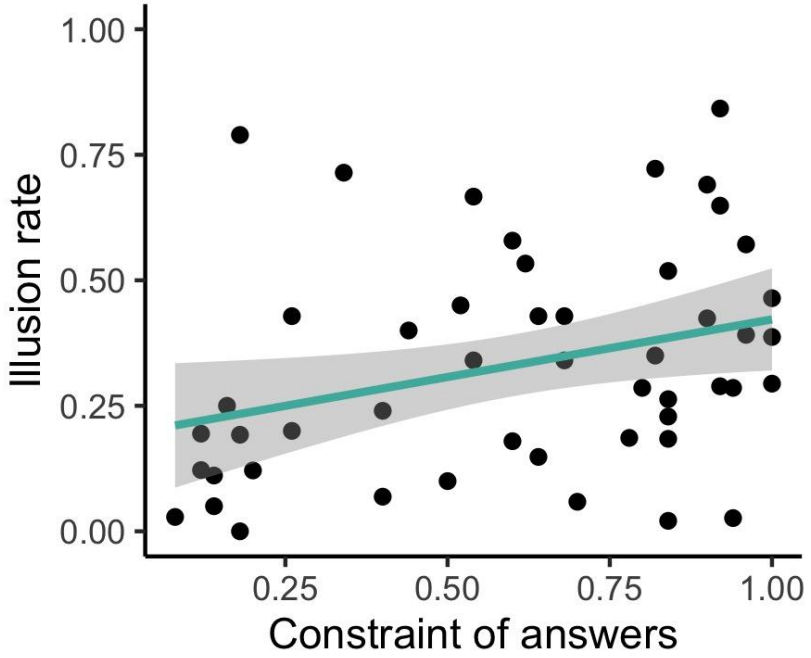
What is the name of the holiday during which children dress up and walk from door to door to _____

Task: complete the question AND answer it

Cloze task


1.00
0.75
0.50
0.25
0.00
Illusion rate
0.25
0.50
0.75
1.00
Constraint of question continuations

Cloze task


1.00
0.75
0.50
0.25
0.00
Illusion rate
0.25
0.50
0.75
1.00
Constraint of answers

Attention

Four broad categories of evidence we're looking for:

1. Evidence that for some stimuli but not others, comprehenders have enough information that they *could* stop prioritizing the bottom-up input prior to the substitution
2. Evidence that when there's enough information, comprehenders *do* shift attention
3. Evidence that shifts in attention lead to shallower representations of incoming information (i.e. more illusions)
4. Evidence for what specific level of representation is reached when shallow processing strategies are used

Attention

Four broad categories of evidence we're looking for:

1. Evidence that for some stimuli but not others, comprehenders have enough information that they *could* stop prioritizing the bottom-up input prior to the substitution
2. Evidence that when there's enough information, comprehenders *do* shift attention
3. Evidence that shifts in attention lead to shallower representations of incoming information (i.e. more illusions)
4. Evidence for what specific level of representation is reached when shallow processing strategies are used

When you only have one task (Experiment 3)

50 Mechanical Turk workers

The same 50 stimuli as experiment 1

Task: Participants don’t need to answer the question, they just report whether there’s an error

When you only have one task (Experiment 3)

50 Mechanical Turk workers

The same 50 stimuli as experiment 1

Task: Participants don’t need to answer the question, they just report whether there’s an error

Minor improvement in accuracy: overall illusion rate goes down from 33% to 30%

Key take aways

1. Illusion rates vary substantially across items
2. Some of that variability correlates with how similar the anomalous word is to the word you should have said
3. Some variability correlates with your certainty about how the question will unfold
4. These findings suggest that the comprehender actively modulates how deeply upcoming information will be processed as a function of how much is known and other task demands

Thank you

This work was supported by NSF DGE-1449815, a University of Maryland Flagship Fellowship awarded to HM, and ONR MURI Award N00014-18-1-2670.

Thanks to Adam Liter, Masato Nakamura, and Nika Jurov for contributions to the modeling of Moses illusion sentences with GPT-2 and LSTMs

Audiences at UMD's Language Science Lunch Talks and Cognitive Neuroscience of Language lab for helpful feedback on earlier versions of this talk

References

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. Memory & cognition, 21(4), 477-487.

Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. Acta Psychologica, 67(2), 135-144.

Cook, A. E., Walsh, E. K., Bills, M. A., Kircher, J. C., & O'Brien, E. J. (2016). Validation of semantic illusions independent of anomaly detection: Evidence from eye movements. The Quarterly Journal of Experimental Psychology, 1-11.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. Journal of Verbal Learning and Verbal Behavior, 20(5), 540-551.

Gulordava, Kristina, Piotr Bojanowski, Edouard Grave, Tal Linzen & Marco Baroni. 2018. Colorless green recurrent networks dream hierarchically. In Marilyn Walker, Heng Ji & Amanda Stent (eds.), Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: Human language technologies, 1195–1205. New Orleans, Louisiana: Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1108.

Hannon, B., & Daneman, M. (2004). Shallow semantic processing of text: An individual-differences account. Discourse Processes, 37(3), 187-204.

Kamas, E. N., Reder, I. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. Memory & Cognition, 24(6), 687-699.

Radford, Alec, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Raposo, A., & Marques, J. F. (2013). The contribution of fronto-parietal regions to sentence comprehension: Insights from the Moses illusion. NeuroImage, 83, 431-437.

Reder, L. M., & Cleeremans, A. (1990). The role of partial matches in comprehension: The Moses illusion revisited. In Psychology of Learning and Motivation (Vol. 25, pp. 233-258). Academic Press.

Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match?. Journal of Memory and Language, 30(4), 385-406.

Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and underspecification. Discourse Processes, 42(2), 99-108.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. Trends in cognitive sciences, 6(9), 382-386.

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. Psychonomic bulletin & review, 11(5), 882-888.

van Jaarsveld, H. J., Dijkstra, T., & Hermans, D. (1997). The detection of semantic illusions: Task-specific effects for similarity and position of distorted terms. Psychological research, 59(4), 219-230.

Van Oostendorp, H., & De Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. Acta Psychologica, 74(1), 35-46.

van Schijndel, Marten & Tal Linzen. 2018. A Neural Model of Adaptation in Reading. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 4704–4710. Brussels, Belgium: Association for Computational Linguistics. https: //doi.org/10.18653/v1/D18-1499.