Modeling Task Effects in Human Reading with Neural Attention

Michael Hahn	Frank Keller
Stanford University	University of Edinburgh
mhahn2@stanford.edu	keller@inf.ed.ac.uk

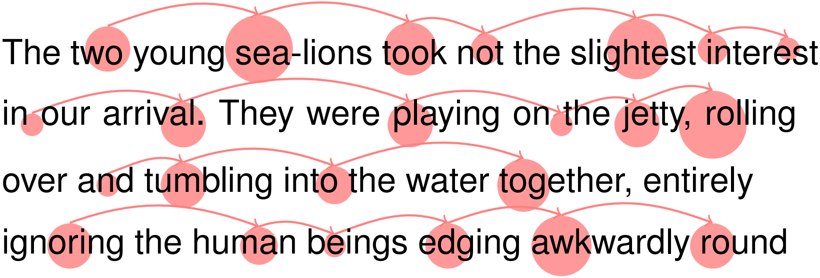
CUNY Conference 2017

Eye Movements in Reading

The two young sea-lions took not the slightest interest in our arrival. They were playing on the jetty, rolling over and tumbling into the water together, entirely ignoring the human beings edging awkwardly round

adapted from the Dundee corpus [Kennedy and Pynte, 2005]

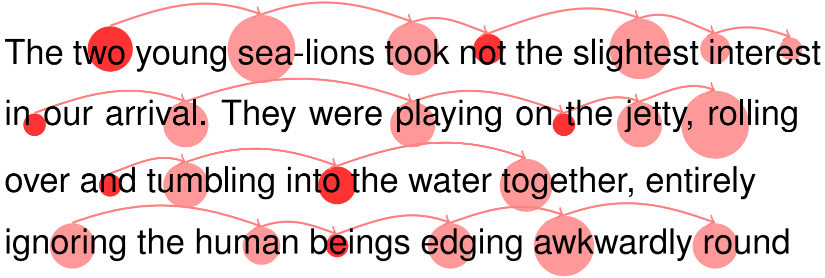
Eye Movements in Reading



adapted from the Dundee corpus [Kennedy and Pynte, 2005]

- **Fixations** static
- **Saccades** take 20–40 ms, no information obtained from text

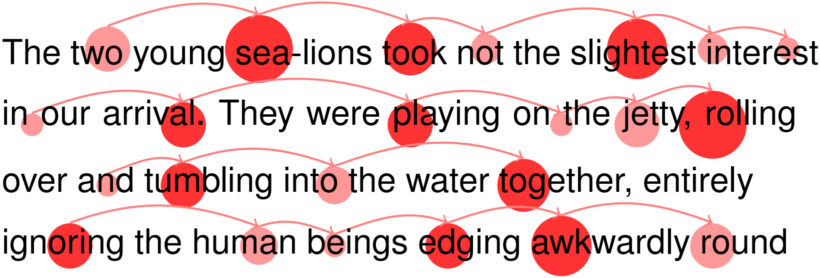
Eye Movements in Reading



adapted from the Dundee corpus [Kennedy and Pynte, 2005]

- **Fixations** static
- **Saccades** take 20–40 ms, no information obtained from text
- **Fixation times** vary from $\approx$ **100 ms** to $\approx$ 300ms

Eye Movements in Reading



adapted from the Dundee corpus [Kennedy and Pynte, 2005]

- **Fixations** static
- **Saccades** take 20–40 ms, no information obtained from text
- **Fixation times** vary from $\approx$ 100 ms to $\approx$ **300ms**

Eye Movements in Reading

The two young sea-lions took not the slightest interest in our arrival. They were playing on the jetty, rolling over and tumbling into the water together, entirely ignoring the human beings edging awkwardly round

adapted from the Dundee corpus [Kennedy and Pynte, 2005]

- **Fixations** static
- **Saccades** take 20–40 ms, no information obtained from text
- **Fixation times** vary from $\approx$ 100 ms to $\approx$ 300ms
- $\approx 40\%$ of words are **skipped**

Computational Models

1. Models of saccade generation in cognitive psychology:
   - EZ-Reader [Reichle et al., 1998, 2003, 2009]
   - SWIFT [Engbert et al., 2002, 2005]
2. Machine learning models trained on eye-tracking data [Nilsson and Nivre, 2009, 2010, Hara et al., 2012, Matthies and Søgaard, 2013]

Computational Models

1. Models of saccade generation in cognitive psychology:
   - EZ-Reader [Reichle et al., 1998, 2003, 2009]
   - SWIFT [Engbert et al., 2002, 2005]
2. Machine learning models trained on eye-tracking data [Nilsson and Nivre, 2009, 2010, Hara et al., 2012, Matthies and Søgaard, 2013]

These models

- require selection of relevant eye-movement features, and
- estimate parameters from eye-tracking corpora

Computational Models

1. Models of saccade generation in cognitive psychology:
    - EZ-Reader [Reichle et al., 1998, 2003, 2009]
    - SWIFT [Engbert et al., 2002, 2005]
2. Machine learning models trained on eye-tracking data [Nilsson and Nivre, 2009, 2010, Hara et al., 2012, Matthies and Søgaard, 2013]

These models

- require selection of relevant eye-movement features, and
- estimate parameters from eye-tracking corpora

3. Bayesian inference (Bicknell and Levy, 2010)
    - maximize speed of reading while reliably identifying the text
    - replicates predictability, frequency effects

Computational Models

1. Models of saccade generation in cognitive psychology:
   - EZ-Reader [Reichle et al., 1998, 2003, 2009]
   - SWIFT [Engbert et al., 2002, 2005]
2. Machine learning models trained on eye-tracking data [Nilsson and Nivre, 2009, 2010, Hara et al., 2012, Matthies and Søgaard, 2013]

These models

- require selection of relevant eye-movement features, and
- estimate parameters from eye-tracking corpora

3. Bayesian inference (Bicknell and Levy, 2010)
   - maximize speed of reading while reliably identifying the text
   - replicates predictability, frequency effects
   - not evaluated on wide-coverage reading data
   - assumes fixed task of word identification

Computational Models: Surprisal

Surprisal measures predictability of word w_i in context $w_1 w_2 \ldots w_{i-1}$:

$$\text{Surprisal}(w_i) = -\log P(w_i | \boldsymbol{w}_{1 \ldots i-1}) \tag{1}$$

Computational Models: Surprisal

Surprisal measures predictability of word w_i in context $w_1 w_2 \ldots w_{i-1}$:

$$\text{Surprisal}(w_i) = -\log P(w_i | \mathbf{w}_{1\ldots i-1}) \tag{1}$$

- predicts word-by-word reading times [Hale, 2001, McDonald and Shillcock, 2003a,b, Levy, 2008]

Computational Models: Surprisal

Surprisal measures predictability of word w_i in context $w_1 w_2 \ldots w_{i-1}$:

$$\text{Surprisal}(w_i) = -\log P(w_i | \boldsymbol{w}_{1\ldots i-1}) \tag{1}$$

- predicts word-by-word reading times [Hale, 2001, McDonald and Shillcock, 2003a,b, Levy, 2008]
- designed as a model of processing effort, hence can't explain:
  - regressions
  - re-fixations
  - spillover
  - skipping
    $\approx$ 40 % of words are skipped

Tradeoff Hypothesis

Goal

Build unsupervised model that accounts for **reading times** and **skipping.**

Tradeoff Hypothesis

Goal

Build unsupervised model that accounts for **reading times** and **skipping.**

Hypothesis

Human reading optimizes a tradeoff between:

- **Precision** of language understanding:
  Perform a language-related task as well as possible
- **Economy** of attention:
  Fixate as few words as possible

Tradeoff Hypothesis

We assume that the default task in reading is to **memorize the text,** i.e., to reconstruct the input as accurately as possible.

Approach: NEAT (NEural Attention Tradeoff)

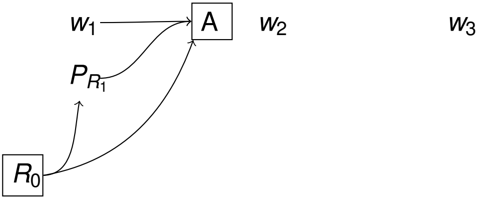
1. Develop generic reading architecture integrating
   - neural language modeling
   - attention mechanism
2. Train end-to-end to optimize tradeoff between precision and economy
3. Evaluate on human eye-tracking corpus

Architecture

w_1 w_2 w_3

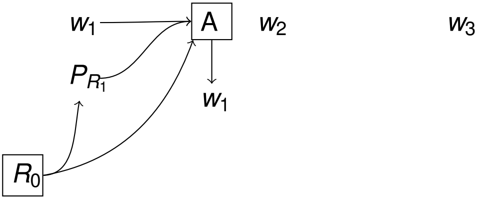
$\boxed{R_0}$

Architecture



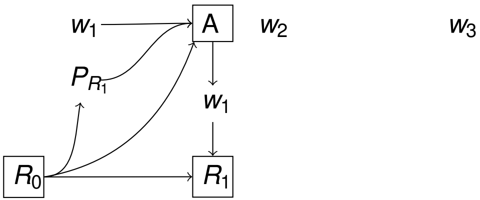
- Attention module A **shows word** to R or **skips** it

Architecture



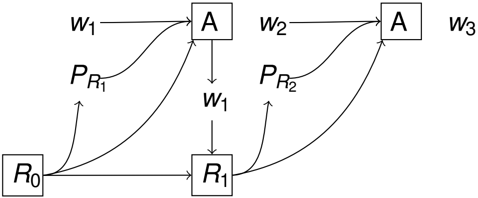
- Attention module *A* **shows word** to *R* or **skips** it

Architecture



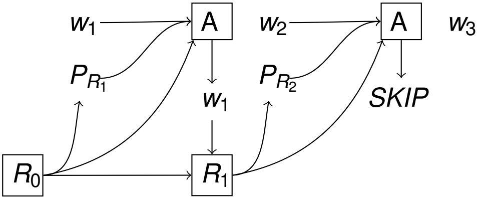
- Attention module *A* **shows word** to *R* or **skips** it

Architecture



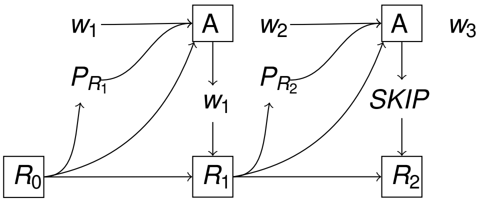
- Attention module A **shows word** to R or **skips** it

Architecture



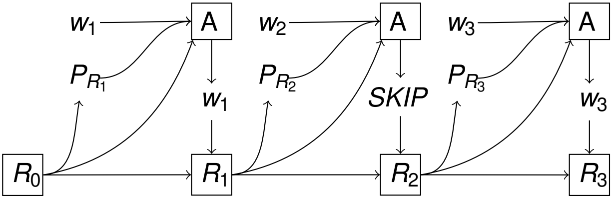
- Attention module *A* **shows word** to *R* or **skips** it
- *R* receives special SKIPPED representation when skipping

Architecture



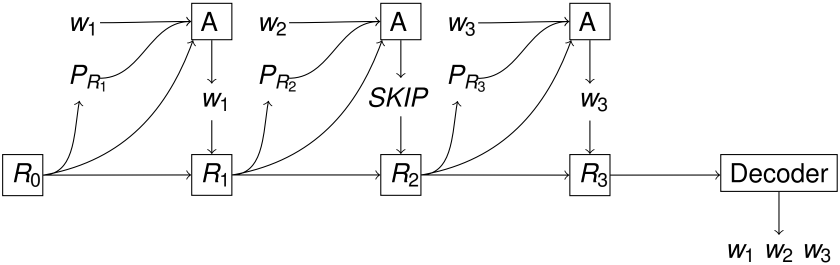
- Attention module A **shows word** to R or **skips** it
- R receives special SKIPPED representation when skipping

Architecture



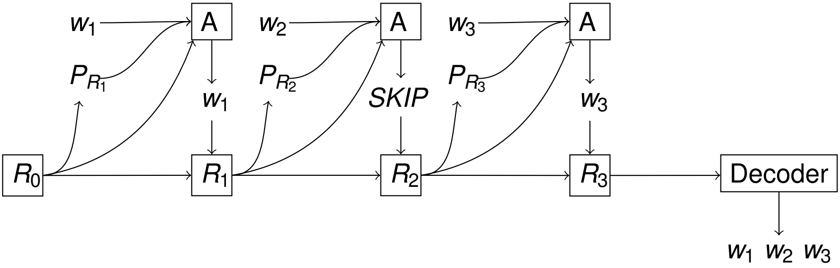
- Attention module A **shows word** to R or **skips** it
- R receives special SKIPPED representation when skipping

Architecture



- Decoder tries to **reconstruct** full text

Architecture



- Decoder tries to **reconstruct** full text
- Reader, Attention, Decoder implemented as **neural networks** (LSTM)

Implementing the Tradeoff Hypothesis

Training Objective

Solve prediction and reconstruction with minimal attention:

$$\arg_\theta \min \left\{ \mathrm{E}_{\boldsymbol{w},\boldsymbol{\omega}} \left[\underbrace{\boxed{L(\boldsymbol{\omega} | \boldsymbol{w}, \theta)}}_{\text{loss for prediction + reconstruction}} + \alpha \cdot \underbrace{\boxed{\|\boldsymbol{\omega}\|_{\ell_1}}}_{\text{number of fixated words}} \right] \right\}$$

Implementing the Tradeoff Hypothesis

Training Objective

Solve prediction and reconstruction with minimal attention:

$$\arg_\theta \min \{ \mathrm{E}_{\boldsymbol{w},\boldsymbol{\omega}} [\boxed{L(\boldsymbol{\omega}|\boldsymbol{w},\theta)} + \alpha \cdot \boxed{||\boldsymbol{\omega}||_{\ell_1}}] \}$$

loss for prediction + reconstruction

number of fixated words

- neural network components trained on newstext ($\approx$ 200 million words)
- training is unsupervised: no lexicon, grammar, eye-tracking data, etc. required

Evaluation

Setup

- English section of the Dundee corpus [Kennedy and Pynte, 2005]
  - 20 texts from *The Independent*
  - eye-movement data from ten readers
- 360,000 words
- Fixation rate: 61.3 %

Evaluation

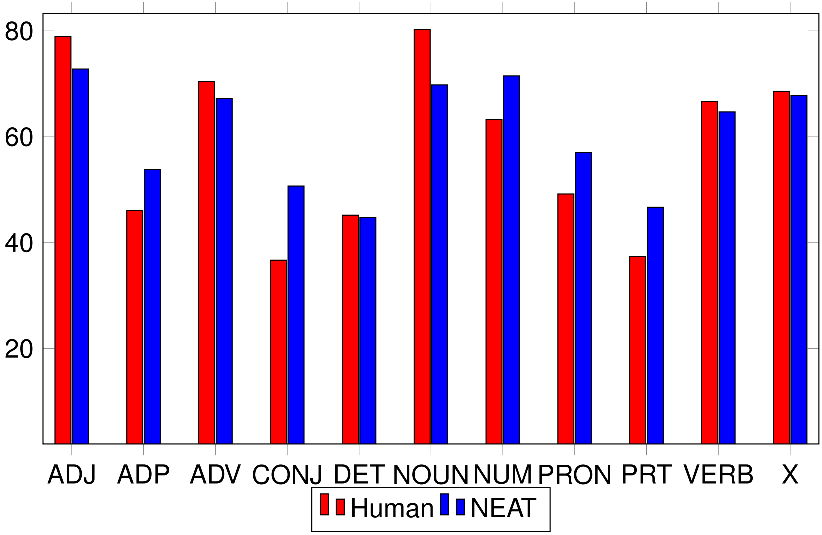
Setup

- English section of the Dundee corpus [Kennedy and Pynte, 2005]
  - 20 texts from *The Independent*
  - eye-movement data from ten readers
- 360,000 words
- Fixation rate: 61.3 %

Results

- NEAT predicts human fixations with accuracy of 63.7 % (random baseline 52.6 %, supervised models 69.9 %)
- surprisal derived from NEAT predicts reading times
- NEAT predicts
  - effects of frequency, length, and predictability
  - correlations between successive fixations
  - differential skipping rates across part-of-speech categories

Fixation Rates by POS Categories



Evaluating Skipping: Heatmaps

HUMAN

The decision of the Human Fertility and Embryology Authority (HFEA) to allow a couple to select genetically their next baby was bound to raise concerns that advances in biotechnology are racing ahead of our ability to control the consequences. The couple at the centre of this case have a son who suffers from a potentially fatal disorder and whose best hope is a marrow transplant from a sibling, so the stakes of this decision are particularly high. The HFEA's critics believe that it sanctions 'designer babies' and does not show respect for the sanctity of individual life.

Evaluating Skipping: Heatmaps

HUMAN

The decision of the Human Fertility and Embryology Authority (HFEA) to allow a couple to select genetically their next baby was bound to raise concerns that advances in biotechnology are racing ahead of our ability to control the consequences. The couple at the centre of this case have a son who suffers from a potentially fatal disorder and whose best hope is a marrow transplant from a sibling, so the stakes of this decision are particularly high. The HFEA's critics believe that it sanctions ’designer babies’ and does not show respect for the sanctity of individual life.

MODEL

The decision of the Human Fertility and Embryology Authority (HFEA) to allow a couple to select genetically their next baby was bound to raise concerns that advances in biotechnology are racing ahead of our ability to control the consequences. The couple at the centre of this case have a son who suffers from a potentially fatal disorder and whose best hope is a marrow transplant from a sibling, so the stakes of this decision are particularly high. The HFEA's critics believe that it sanctions ’designer babies’ and does not show respect for the sanctity of individual life.

Task Effects in Reading

Hypothesis

Human reading optimizes a tradeoff between

- **Precision** of language understanding:
  Perform a language-related task as well as possible
- **Economy** of attention:
  Fixate as few words as possible

We assumed that the task is to **memorize the text,** i.e., to reconstruct the input as accurately as possible

Task Effects in Reading

Hypothesis

Human reading optimizes a tradeoff between

- **Precision** of language understanding:
  Perform a language-related task as well as possible
- **Economy** of attention:
  Fixate as few words as possible

We assumed that the task is to **memorize the text,** i.e., to reconstruct the input as accurately as possible

Prediction: if we change the task, the precision/economy tradeoff changes, and we observe different reading behavior

Task Effects in Reading

Hypothesis

Human reading optimizes a tradeoff between

- **Precision** of language understanding:
  Perform a language-related task as well as possible
- **Economy** of attention:
  Fixate as few words as possible

We assumed that the task is to **memorize the text,** i.e., to reconstruct the input as accurately as possible

Prediction: if we change the task, the precision/economy tradeoff changes, and we observe different reading behavior

There is independent evidence for **task effects**, e.g., reading vs proofreading [Schotter et al., 2014].

Eye-tracking Experiment: Question Answering

Experimentally test NEAT's prediction about task effects using a **question answering task** in two conditions:

- **No Preview:** participants read text, then answer question
- **Preview:** participants see question, read text, answer question

Eye-tracking Experiment: Question Answering

Experimentally test NEAT's prediction about task effects using a **question answering task** in two conditions:

- **No Preview:** participants read text, then answer question
- **Preview:** participants see question, read text, answer question

Experimental Setup

1. Participants read 20 newspaper texts and answer one multiple-choice question per text
2. 10 participants in Preview condition, 10 in No Preview condition
3. Texts and questions taken from DeepMind question answering corpus [Hermann et al., 2015]
4. Eye-movements recorded using an Eyelink 2000 tracker

Eye-tracking Experiment: Example (Preview Condition)

Question: A random sample from a __________ store tested positive for Listeria monocytogenes.

Eye-tracking Experiment: Example (Preview Condition)

Sabra is recalling 30,000 cases of hummus due to possible contamination with Listeria, the U.S. said Wednesday. The nationwide recall is voluntary. So far, no illnesses caused by the hummus have been reported. The potential for contamination was discovered when a routine, random sample collected at a Michigan store on March 30 tested positive for Listeria monocytogenes. The FDA issued a list of the products in the recall. Anyone who has purchased any of the items is urged to dispose of or return it to the store for a full refund. Listeria monocytogenes can cause serious and sometimes fatal infections in young children, frail or elderly people, and others with weakened immune systems, the FDA says. Although some people may suffer only short-term symptoms such as high fever, severe headache, nausea, abdominal pain and diarrhea, Listeria can also cause miscarriages and stillbirths among pregnant women.

Eye-tracking Experiment: Example (Preview Condition)

Question: A random sample from a __________ store tested positive for Listeria monocytogenes.

Answers: (1) Michigan (2) Washington (3) Ohio (4) Georgia

Eye-tracking Experiment: Example (Preview Condition)

Question: A random sample from a __________ store tested positive for Listeria monocytogenes.

Answers: (1) Michigan (2) Washington (3) Ohio (4) Georgia

Results: Descriptive Statistics

	No Preview	Preview
Fixation rate	0.50	0.34
First fixation	221.3	194.9
First pass	260.7	210.8
Total time	338.0	263.0
Regression path	463.9	362.5
Quest. accuracy	0.70	0.89

Results: Descriptive Statistics

	No Preview	Preview
Fixation rate	0.50	0.34
First fixation	221.3	194.9
First pass	260.7	210.8
Total time	338.0	263.0
Regression path	463.9	362.5
Quest. accuracy	0.70	0.89

- More skipping, faster reading, shorter regression paths in Preview condition

Results: Descriptive Statistics

	No Preview	Preview
Fixation rate	0.50	0.34
First fixation	221.3	194.9
First pass	260.7	210.8
Total time	338.0	263.0
Regression path	463.9	362.5
Quest. accuracy	0.70	0.89

- More skipping, faster reading, shorter regression paths in Preview condition

Results: Descriptive Statistics

	No Preview	Preview
Fixation rate	0.50	0.34
First fixation	221.3	194.9
First pass	260.7	210.8
Total time	338.0	263.0
Regression path	463.9	362.5
Quest. accuracy	0.70	0.89

- More skipping, faster reading, shorter regression paths in Preview condition
- nevertheless improved accuracy

Results: Mixed Effects Model for First Pass and Total Time

Predictor	First Pass	Total Time
(Intercept)	225.72***	284.73***
NoPreview	23.61***	39.23***
PositionText	−0.85	−14.97**
IsNamedEntity	14.72**	53.09***
IsCorrectAnswer	−19.93	−12.40
OccursInQuestion	−4.42	−4.25
Surprisal	8.98***	20.39***
IsFunctionWord	−1.89	−2.80
NoPreview:PositionText	−5.19*	−4.06
NoPreview:IsNamedEntity	4.56	15.62**
NoPreview:IsCorrectAnswer	−12.74	−70.15***
NoPreview:OccursInQuestion	1.69	5.24
NoPreview:Surprisal	1.49	4.18**
NoPreview:IsFunctionWord	−0.62	−1.84

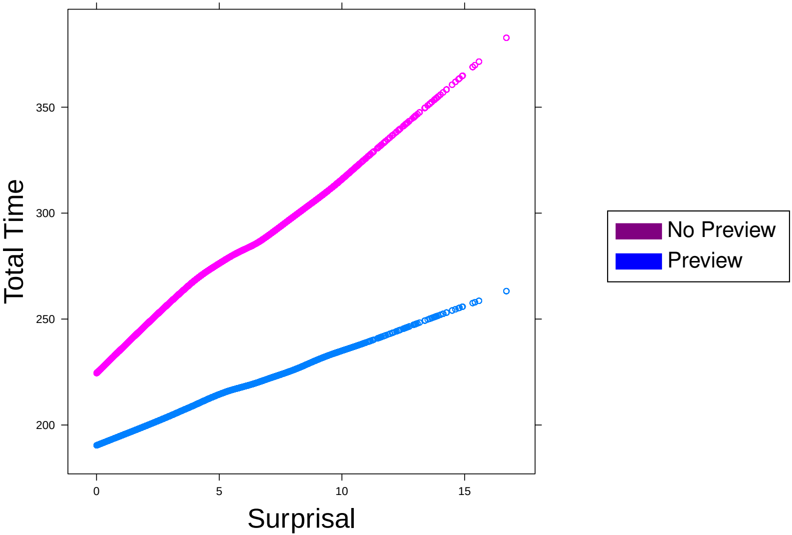
Results: Mixed Effects Model for First Pass and Total Time

Predictor	First Pass	Total Time
(Intercept)	225.72***	284.73***
NoPreview	23.61***	39.23***
PositionText	−0.85	−14.97**
IsNamedEntity	14.72**	53.09***
IsCorrectAnswer	−19.93	−12.40
OccursInQuestion	−4.42	−4.25
Surprisal	8.98***	20.39***
IsFunctionWord	−1.89	−2.80
NoPreview:PositionText	−5.19*	−4.06
NoPreview:IsNamedEntity	4.56	15.62**
NoPreview:IsCorrectAnswer	−12.74	−70.15***
NoPreview:OccursInQuestion	1.69	5.24
NoPreview:Surprisal	1.49	4.18**
NoPreview:IsFunctionWord	−0.62	−1.84

Results: Mixed Effects Model for First Pass and Total Time

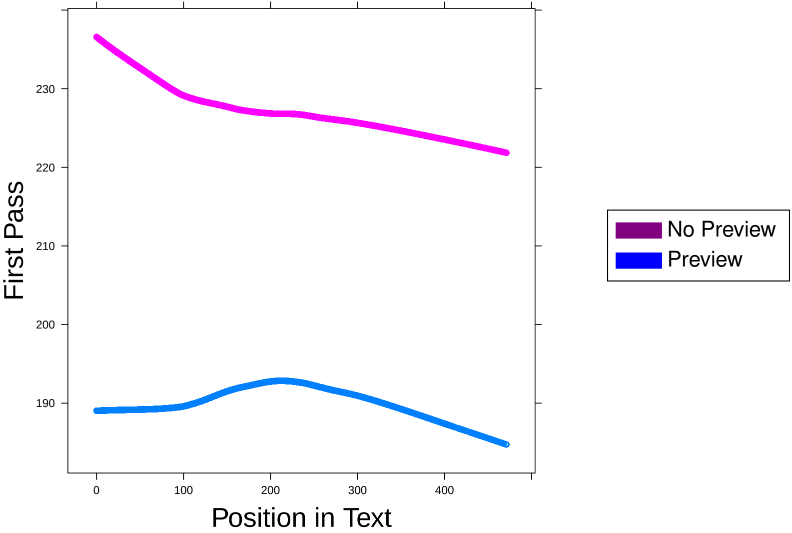
Predictor	First Pass	Total Time
(Intercept)	225.72***	284.73***
NoPreview	23.61***	39.23***
PositionText	−0.85	−14.97**
IsNamedEntity	14.72**	53.09***
IsCorrectAnswer	−19.93	−12.40
OccursInQuestion	−4.42	−4.25
Surprisal	8.98***	20.39***
IsFunctionWord	−1.89	−2.80
NoPreview:PositionText	−5.19*	−4.06
NoPreview:IsNamedEntity	4.56	15.62**
NoPreview:IsCorrectAnswer	−12.74	−70.15***
NoPreview:OccursInQuestion	1.69	5.24
NoPreview:Surprisal	1.49	4.18**
NoPreview:IsFunctionWord	−0.62	−1.84

Interaction Condition:Surprisal (Total Time)



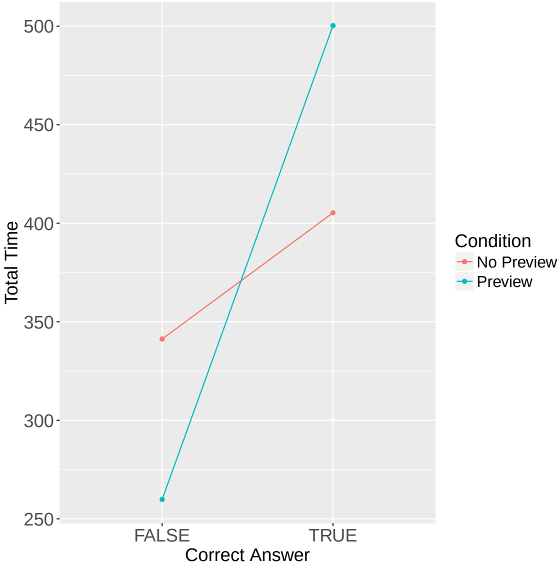
In **No Preview** condition, larger effect of **surprisal**.

Interaction Condition:PositionText (First Pass Time)



In **Preview** condition, readers slow down in the **middle** (highest likelihood of finding the answer)

Interaction Condition:IsCorrect Answer (Total Time)



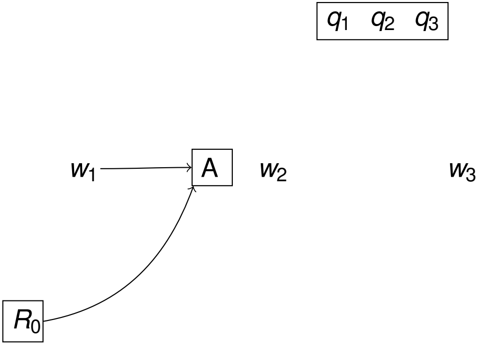
In **Preview** condition, readers slow down more for words that occur in the **correct answer**.

NEAT Architecture: Question Answering (Preview)

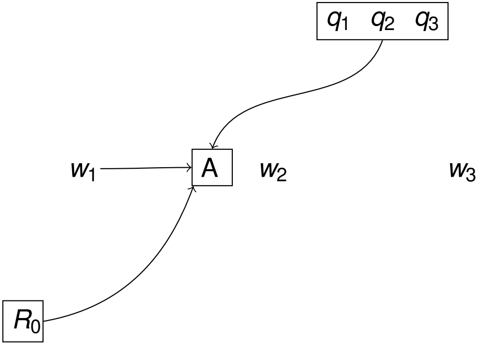
$\boxed{q_1 \quad q_2 \quad q_3}$

w_1 w_2 w_3

NEAT Architecture: Question Answering (Preview)

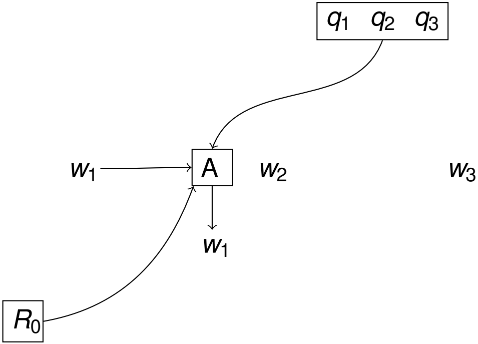


NEAT Architecture: Question Answering (Preview)



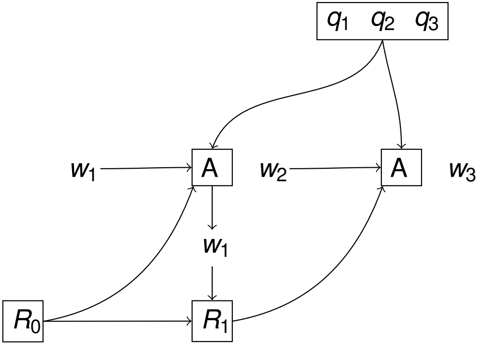
- **Preview**: Fixating and skipping conditioned on question

NEAT Architecture: Question Answering (Preview)



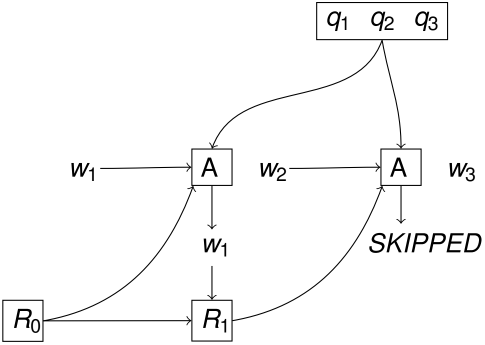
- **Preview**: Fixating and skipping conditioned on question

NEAT Architecture: Question Answering (Preview)



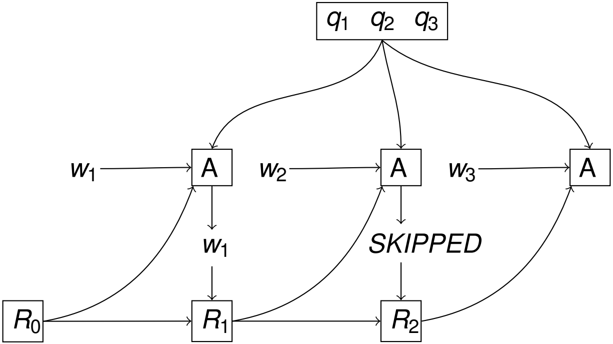
- **Preview**: Fixating and skipping conditioned on question

NEAT Architecture: Question Answering (Preview)



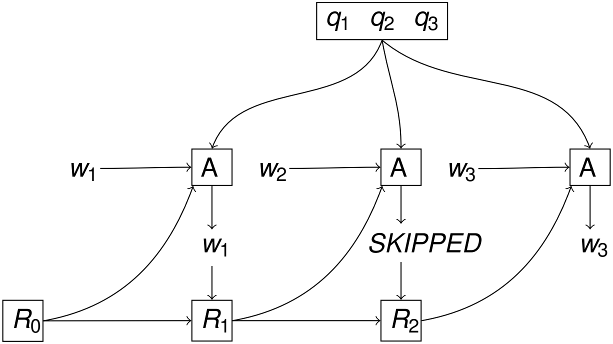
- **Preview**: Fixating and skipping conditioned on question

NEAT Architecture: Question Answering (Preview)



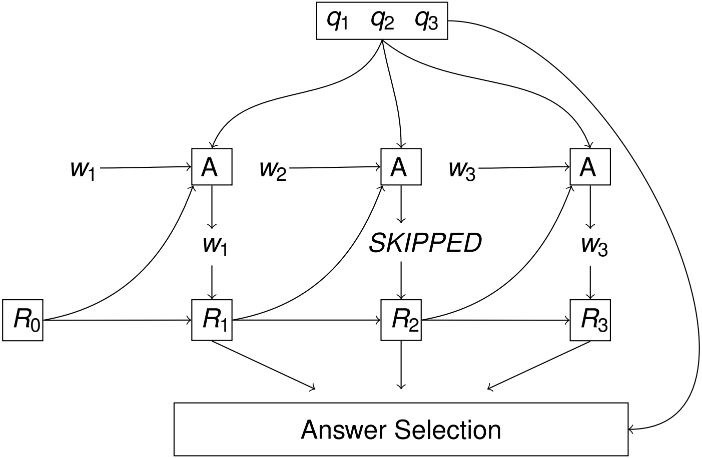
- **Preview**: Fixating and skipping conditioned on question

NEAT Architecture: Question Answering (Preview)



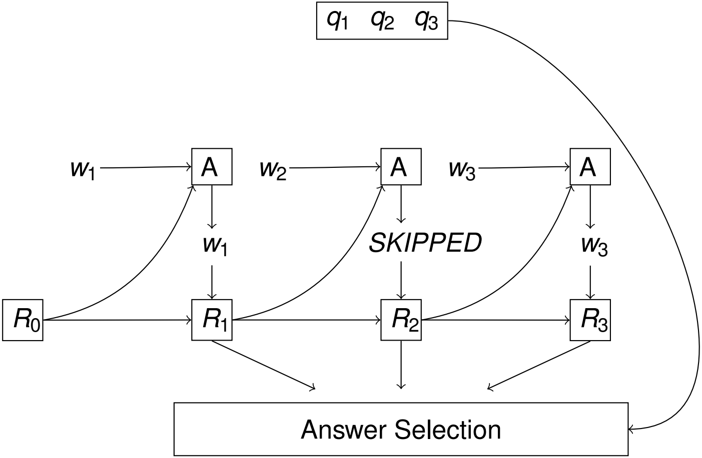
- **Preview**: Fixating and skipping conditioned on question

NEAT Architecture: Question Answering (Preview)



We replace the task module with a **question answering module,** based on the Attentive Reader model [Hermann et al., 2015]

NEAT Architecture: Question Answering (No Preview)



In the **No Preview** condition, we remove the connections from the question to *A*.

Implementing Task Differences in NEAT

Training Objective

Question answering with minimal attention:

$$\arg_\theta \min \left\{ \mathrm{E}_{(\boldsymbol{t},\boldsymbol{q},a),\boldsymbol{\omega}} \left[-\log P(a|\boldsymbol{\omega},\boldsymbol{t},\boldsymbol{q},\theta) + \alpha \cdot \frac{\|\boldsymbol{\omega}\|_{\ell_1}}{N} \right] \right\}$$

Implementing Task Differences in NEAT

Training Objective

Question answering with minimal attention:

$$\arg_\theta \min \left\{ \mathrm{E}_{(\boldsymbol{t},\boldsymbol{q},a),\boldsymbol{\omega}} \boxed{\left[-\log P(a|\boldsymbol{\omega}, \boldsymbol{t}, \boldsymbol{q}, \theta) \right.} + \alpha \cdot \boxed{\frac{\|\boldsymbol{\omega}\|_{\ell_1}}{N}} \left. \right] \right\}$$

loss for question answering

fixation rate

Implementing Task Differences in NEAT

Training Objective

Question answering with minimal attention:

$$\arg_\theta \min \left\{ \mathrm{E}_{(\boldsymbol{t},\boldsymbol{q},a),\boldsymbol{\omega}} \left[\boxed{-\log P(a|\boldsymbol{\omega},\boldsymbol{t},\boldsymbol{q},\theta)} + \alpha \cdot \boxed{\frac{\|\boldsymbol{\omega}\|_{\ell_1}}{N}} \right] \right\}$$

loss for question answering

fixation rate

- trained separately for Preview and No Preview conditions

Implementing Task Differences in NEAT

Training Objective

Question answering with minimal attention:

$$\arg_\theta \min \left\{ \mathrm{E}_{(\boldsymbol{t},\boldsymbol{q},a),\boldsymbol{\omega}} \left[\underbrace{-\log P(a|\boldsymbol{\omega},\boldsymbol{t},\boldsymbol{q},\theta)}_{\text{loss for question answering}} + \alpha \cdot \underbrace{\frac{\|\boldsymbol{\omega}\|_{\ell_1}}{N}}_{\text{fixation rate}} \right] \right\}$$

- trained separately for Preview and No Preview conditions
- on the DeepMind newstext corpus [Hermann et al., 2015]:
  - 380,298 article-question pairs from CNN
  - $\approx$ 290 million words

Evaluation on Question-answering Data

Test if the attention variable (which is task dependent) improves prediction over surprisal (computed by LSTM, as before).

Mixed effects model for **first-pass times:**

Predictor	Mean	SD	
(Intercept)	231.59	7.50	*
PositionText	1.54	3.74	
IsNamedEntity	19.70	4.82	*
IsCorrectAnswer	-23.23	12.06	
OccursInQuestion	-4.78	3.48	
Surprisal	8.85	1.32	*
IsFunctionWord	-1.76	1.39	
WordLength	8.76	0.55	*
PositionText:IsFunctionWord	4.71	3.79	
Residualized NEAT Attention	78.33	4.87	*

Evaluation on Question-answering Data

Test if the attention variable (which is task dependent) improves prediction over surprisal (computed by LSTM, as before).

Mixed effects model for **first-pass times:**

Predictor	Mean	SD	
(Intercept)	231.59	7.50	*
PositionText	1.54	3.74	
IsNamedEntity	19.70	4.82	*
IsCorrectAnswer	-23.23	12.06	
OccursInQuestion	-4.78	3.48	
Surprisal	8.85	1.32	*
IsFunctionWord	-1.76	1.39	
WordLength	8.76	0.55	*
PositionText:IsFunctionWord	4.71	3.79	
Residualized NEAT Attention	78.33	4.87	*

Evaluation on Question-answering Data

Test if the attention variable (which is task dependent) improves prediction over surprisal (computed by LSTM, as before).

Mixed effects model for **skipping:**

Predictor	Mean	SD	
(Intercept)	−0.45	0.16	**
PositionText	0.19	0.05	***
IsNamedEntity	0.01	0.07	
IsCorrectAnswer	0.35	0.18	
OccursInQuestion	0.02	0.05	
Surprisal	0.09	0.02	***
IsFunctionWord	−0.17	0.02	***
WordLength	0.24	0.01	***
PositionText:IsFunctionWord	0.06	0.05	
Residualized NEAT Attention	1.45	0.06	***

Evaluation on Question-answering Data

Test if the attention variable (which is task dependent) improves prediction over surprisal (computed by LSTM, as before).

Mixed effects model for **skipping:**

Predictor	Mean	SD	
(Intercept)	−0.45	0.16	**
PositionText	0.19	0.05	***
IsNamedEntity	0.01	0.07	
IsCorrectAnswer	0.35	0.18	
OccursInQuestion	0.02	0.05	
Surprisal	0.09	0.02	***
IsFunctionWord	−0.17	0.02	***
WordLength	0.24	0.01	***
PositionText:IsFunctionWord	0.06	0.05	
Residualized NEAT Attention	1.45	0.06	***

Qualitative Analysis

HUMAN: NO PREVIEW

Sabra is recalling 30,000 cases of hummus due to possible contamination with Listeria, the U.S. said Wednesday. The nationwide recall is voluntary. So far, no illnesses caused by the hummus have been reported. The potential for contamination was discovered when a routine, random sample collected at a Michigan store on March 30 tested positive for Listeria monocytogenes. The FDA issued a list of the products in the recall. Anyone who has purchased any of the items is urged to dispose of or return it to the store for a full refund. Listeria monocytogenes can cause serious and sometimes fatal infections in young children, frail or elderly people, and others with weakened immune systems, the FDA says. Although some people may suffer only short-term symptoms such as high fever, severe headache, nausea, abdominal pain and diarrhea, Listeria can also cause miscarriages and stillbirths among pregnant women.

Qualitative Analysis

HUMAN: PREVIEW

Sabra is recalling 30,000 cases of hummus due to possible contamination with Listeria, the U.S. said Wednesday. The nationwide recall is voluntary. So far, no illnesses caused by the hummus have been reported. The potential for contamination was discovered when a routine, random sample collected at a Michigan store on March 30 tested positive for Listeria monocytogenes. The FDA issued a list of the products in the recall. Anyone who has purchased any of the items is urged to dispose of or return it to the store for a full refund. Listeria monocytogenes can cause serious and sometimes fatal infections in young children, frail or elderly people, and others with weakened immune systems, the FDA says. Although some people may suffer only short-term symptoms such as high fever, severe headache, nausea, abdominal pain and diarrhea, Listeria can also cause miscarriages and stillbirths among pregnant women.

Qualitative Analysis

MODEL: NO PREVIEW

Sabra is recalling 30,000 cases of hummus due to possible contamination with Listeria, the U.S. said Wednesday. The nationwide recall is voluntary. So far, no illnesses caused by the hummus have been reported. The potential for contamination was discovered when a routine, random sample collected at a Michigan store on March 30 tested positive for Listeria monocytogenes. The FDA issued a list of the products in the recall. Anyone who has purchased any of the items is urged to dispose of or return it to the store for a full refund. Listeria monocytogenes can cause serious and sometimes fatal infections in young children, frail or elderly people, and others with weakened immune systems, the FDA says. Although some people may suffer only short-term symptoms such as high fever, severe headache, nausea, abdominal pain and diarrhea, Listeria can also cause miscarriages and stillbirths among pregnant women.

Qualitative Analysis

MODEL: PREVIEW

Sabra is recalling 30,000 cases of hummus due to possible contamination with Listeria, the U.S. said Wednesday. The nationwide recall is voluntary. So far, no illnesses caused by the hummus have been reported. The potential for contamination was discovered when a routine, random sample collected at a Michigan store on March 30 tested positive for Listeria monocytogenes. The FDA issued a list of the products in the recall. Anyone who has purchased any of the items is urged to dispose of or return it to the store for a full refund. Listeria monocytogenes can cause serious and sometimes fatal infections in young children, frail or elderly people, and others with weakened immune systems, the FDA says. Although some people may suffer only short-term symptoms such as high fever, severe headache, nausea, abdominal pain and diarrhea, Listeria can also cause miscarriages and stillbirths among pregnant women.

Conclusions

- NEAT: unsupervised neural net model of reading
- based on tradeoff between **precision** of understanding and **economy** of attention
- evaluation on Dundee corpus (normal reading):
  - accurately predicts human skipping behavior
  - known qualitative properties of skipping emerge
- NEAT predicts task-effects in reading:
  - tested in question-answering eye-tracking experiment
  - preview interacts with text position, named entities, surprisal
- NEAT can capture these results using task model that performs question answering

References I

K. Bicknell and R. Levy. Rational eye movements in reading combining uncertainty about previous words with contextual probability. In *Proceedings of the 32nd annual conference of the cognitive science society*, pages 1142–1147, 2010.

R. Engbert, A. Longtin, and R. Kliegl. A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, 42(5):621–636, 2002. URL http://www.sciencedirect.com/science/article/pii/S0042698901003017.

R. Engbert, A. Nuthmann, E. M. Richter, and R. Kliegl. SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, 112(4):777–813, 2005. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.112.4.777.

J. Hale. A Probabilistic Earley Parser as a Psycholinguistic Model. In *Proceedings of NAACL*, volume 2, pages 159–166, 2001.

T. Hara, D. M. Y. Kano, and A. Aizawa. Predicting word fixations in text with a CRF model for capturing general reading strategies among readers. In *Proceedings of the First Workshop on Eye-tracking and Natural Language Processing*, pages 55–70, 2012. URL http://anthology.aclweb.org/W/W12/W12-49.pdf#page=65.

K. M. Hermann, T. Kočiskỳ, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom. Teaching machines to read and comprehend. *arXiv preprint arXiv:1506.03340*, 2015. URL http://arxiv.org/abs/1506.03340.

A. Kennedy and J. Pynte. Parafoveal-on-foveal effects in normal reading. *Vision Research*, 45(2):153–168, January 2005. URL http://linkinghub.elsevier.com/retrieve/pii/S0042698904003979.

R. Levy. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177, March 2008. URL http://linkinghub.elsevier.com/retrieve/pii/S0010027707001436.

F. Matthies and A. Søgaard. With Blinkers on: Robust Prediction of Eye Movements across Readers. In *EMNLP*, pages 803–807, 2013. URL http://www.aclweb.org/website/old_anthology/D/D13/D13-1075.pdf.

S. A. McDonald and R. C. Shillcock. Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14(6):648–652, November 2003a.

S. A. McDonald and R. C. Shillcock. Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research*, 43(16):1735–1751, July 2003b. URL http://www.sciencedirect.com/science/article/pii/S0042698903002372.

References II

M. Nilsson and J. Nivre. Learning where to look: Modeling eye movements in reading. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 93–101. Association for Computational Linguistics, 2009. URL http://dl.acm.org/citation.cfm?id=1596392.

M. Nilsson and J. Nivre. Towards a data-driven model of eye movement control in reading. In *Proceedings of the 2010 workshop on cognitive modeling and computational linguistics*, pages 63–71. Association for Computational Linguistics, 2010. URL http://dl.acm.org/citation.cfm?id=1870073.

E. D. Reichle, A. Pollatsek, D. L. Fisher, and K. Rayner. Toward a model of eye movement control in reading. *Psychological Review*, 105(1):125–157, January 1998.

E. D. Reichle, K. Rayner, and A. Pollatsek. The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences*, 26(04):445–476, 2003. URL http://journals.cambridge.org/abstract_S0140525X03000104.

E. D. Reichle, T. Warren, and K. McConnell. Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16(1):1–21, February 2009. URL http://www.springerlink.com/index/10.3758/PBR.16.1.1.

E. R. Schotter, K. Bicknell, I. Howard, R. Levy, and K. Rayner. Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition*, 131(1):1–27, 2014.