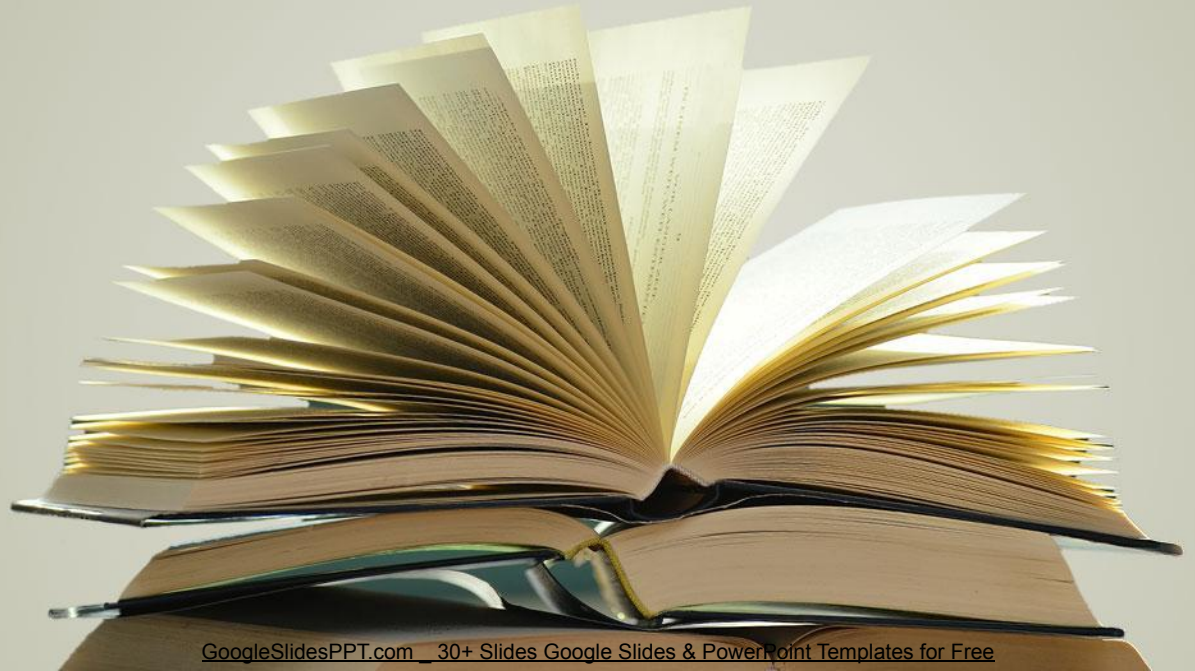
Real Relevant Reviews (R^3)

Only The Most Helpful Product Reviews

W266 Section 003
Jen Darrouzet & Bradley Nott



GoogleSlidesPPT.com _ 30+ Slides Google Slides & PowerPoint Templates for Free

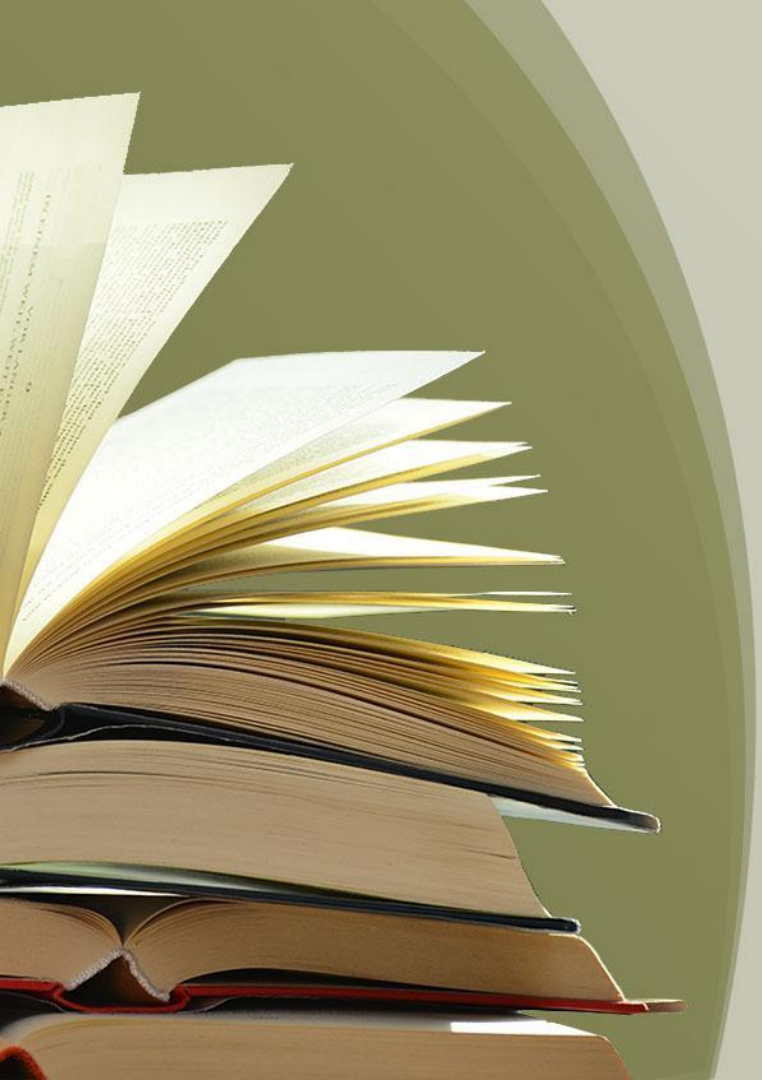
The R^3 Project

1. **The Problem**
   Too many online reviews
2. **The Data**
   8 Million book reviews from Amazon.com
3. **The Approach**
   Helpful Vote Arrival Rates + Neural Nets
4. **The Results**
   To be revealed...
5. **The Next Steps**
   Real, relevant, crowd-sourced advice for buyers

R^3 System

The Problem

Need an efficient method to make product reviews useful

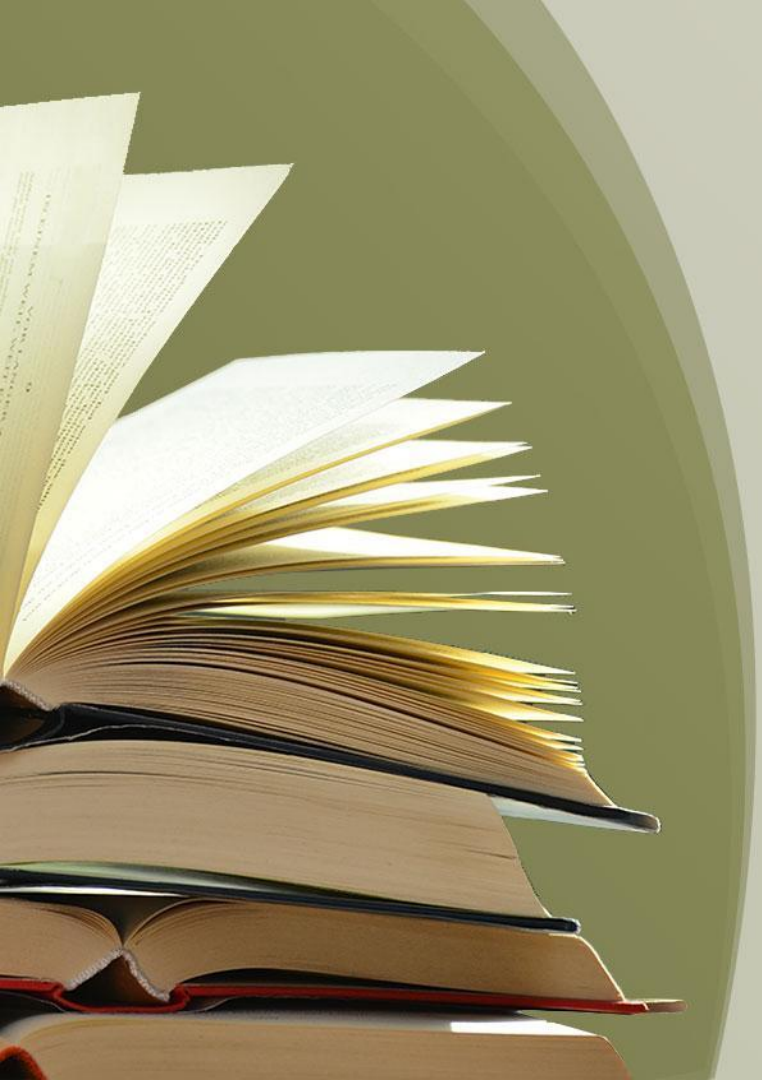


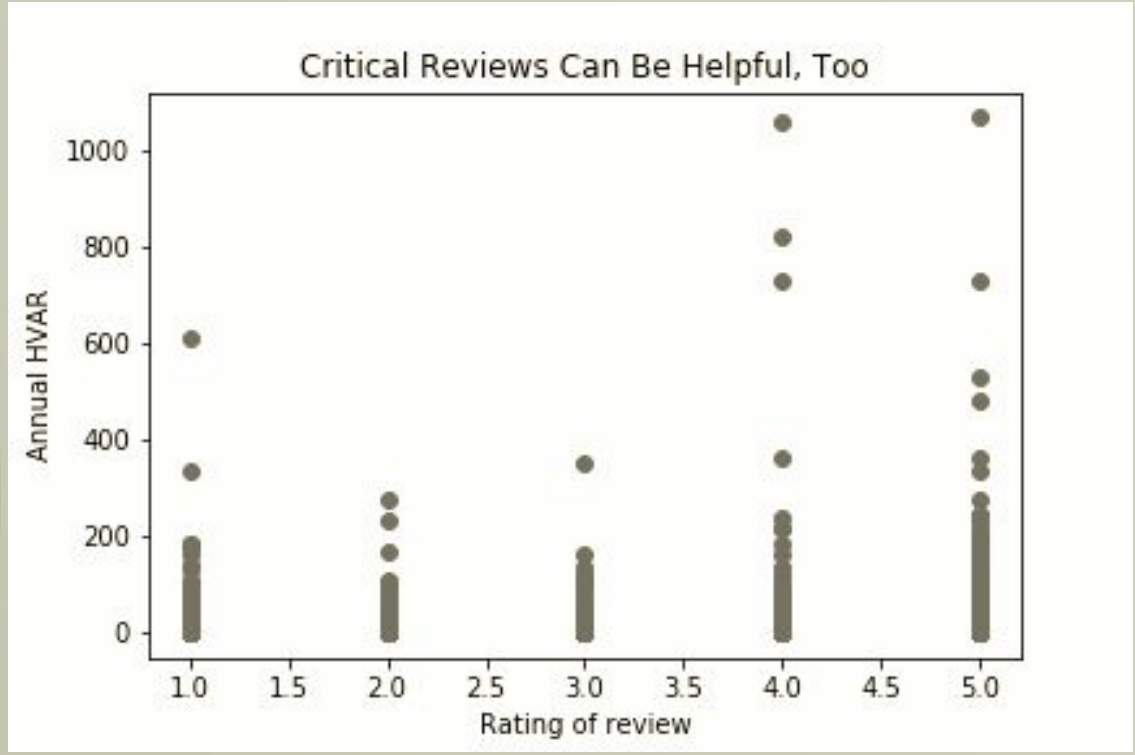
The Data

8 Million book reviews from Amazon.com (2006-2014) that contain a helpful vote tally

The Approach

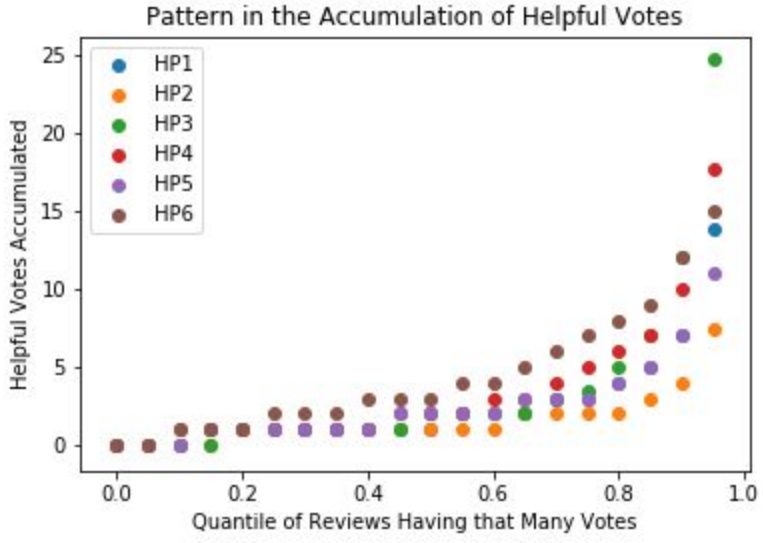
Transform the helpful vote tally into a metric that enables comparisons

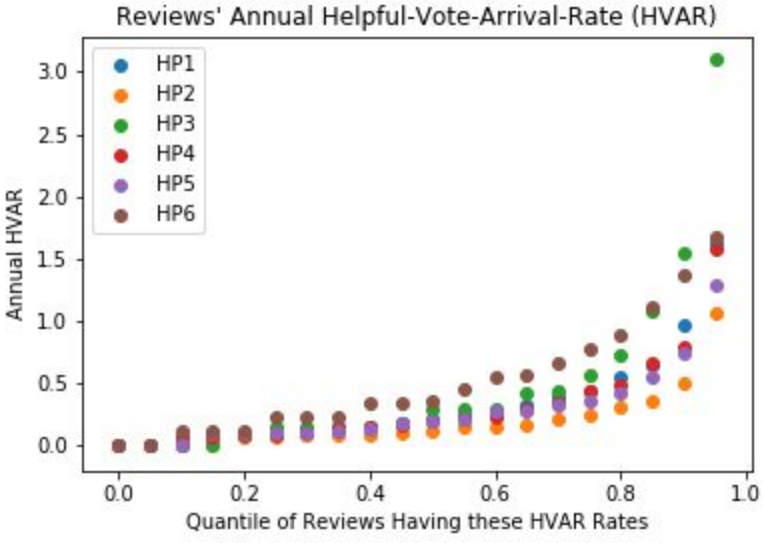



Critical Reviews Can Be Helpful, Too
Annual HVAR
0
200
400
600
800
1000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Rating of review

How to Train for Helpfulness

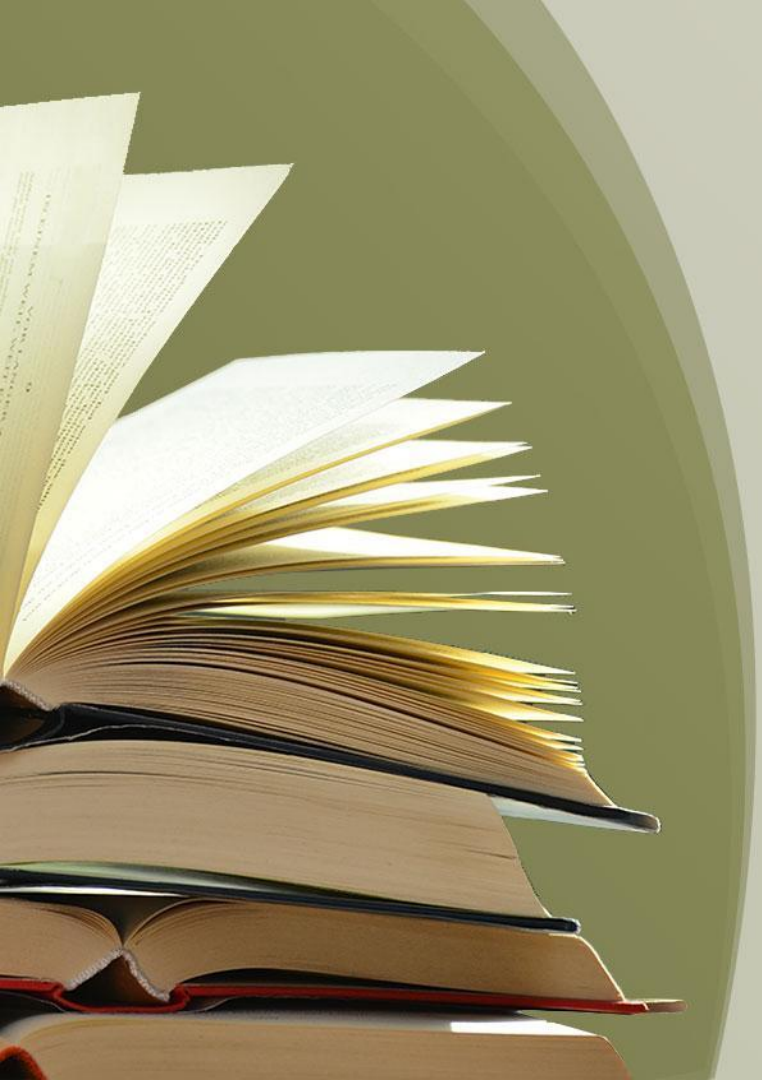
Counts? Proportions? Apples-to-apples?





Experiments

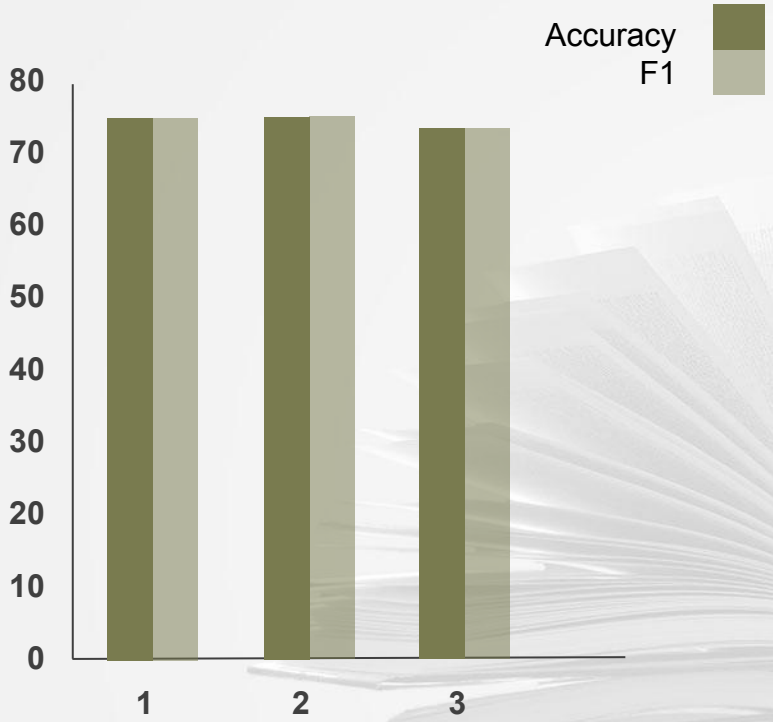
- Data is noisy and the signal is rare
- What is “means” to be helpful varies
- Difficult to be confident about signal
- Smaller models have a hard time with variations in review length



The Results

Classifier Performance

Predicting helpfulness of a review, given review text and 1-5★ rating



Modeling Notes

None of the models fit the noisy data well.
Sorting by probability of helpfulness yields good top N.
Asking shoppers if we got it right could help, too.

1 tf.idf & LogReg	Super fast; unlimited vocab was 1M+
2 Bi-LSTM	Insanely slow (used alt CuDNNLSTM); 200 tokens
3 BERT	Medium speed; 200 tokens; 30K vocab; extra benefit of NSP

BERT's added benefit

If BERT is substantially same, **get relevance for "free"**

Query Relevance

with BERT's Next Sentence Prediction (NSP)

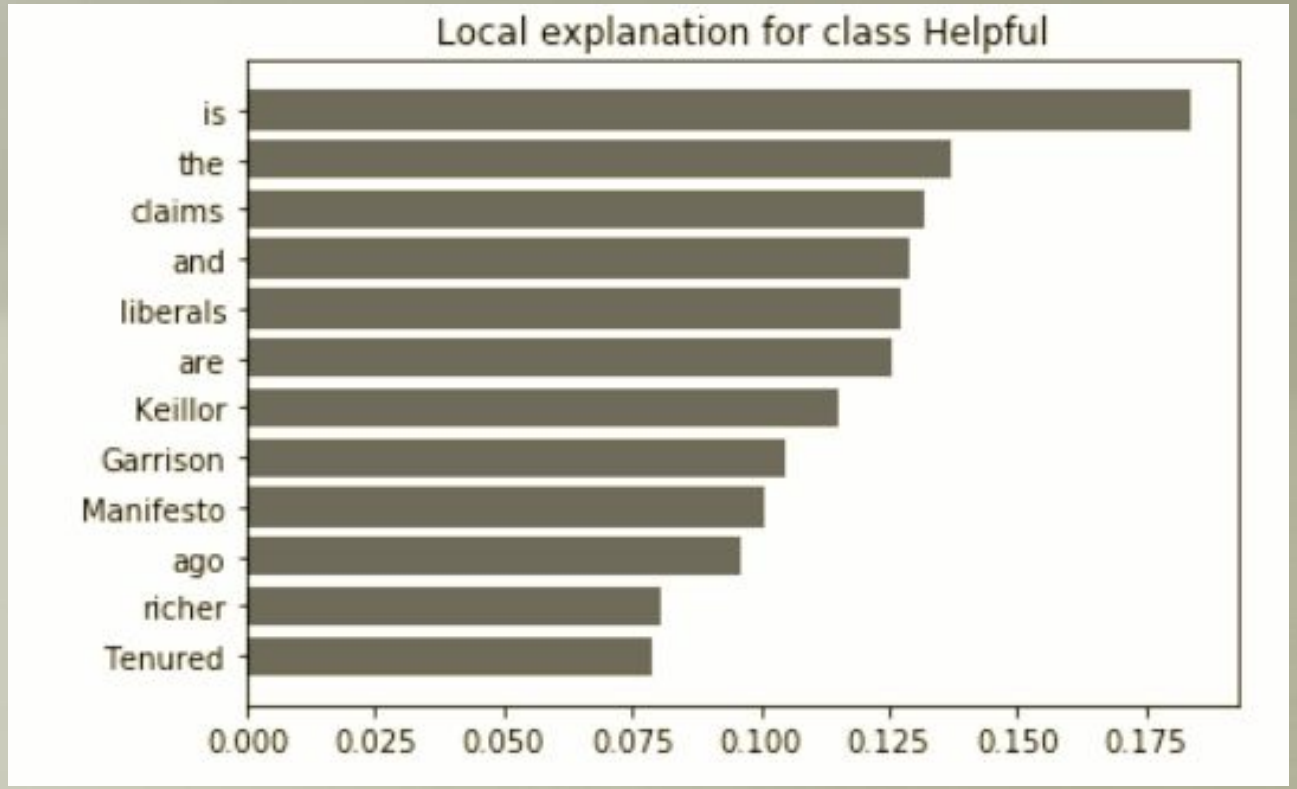
Query: *Is Water for Elephants appropriate for a teenager headed off to veterinary school?*

4★: ...Jacob's parents died just as he was about to write his final exams in **veterinary school** at Cornell University....**because of his vet training**, he ends up hired on to care for the horses and other animals...I read it quickly and found it very sweet and entertaining; **recommended for teens and adults**... Caveat: **there are descriptions of mistreatment and abuse of animals** in this book. If that is unpalatable to you, you should skip this book...

2★: Jacob, now an old man, begins the book by going back into his youth to the time **he was attending vet school**. ... He winds up with a circus and becomes the circus vet, although, **he rarely spent any time with or working on the animals**. **He just stood and watched animals be slaughtered and then felt bad** that he did not do anything...

Helpfulness indicator

LIME helped identify that reviews with claims were >80% helpful



Claims

Not just for nonfiction reviews, either

BEST Someone with an impressive array of murderous skills is systematically assassinating right-wing political figures one after another. The FBI and the CIA, who suspect that the killer comes from somewhere in their ranks, join forces to try to track down the assassin. An imprisoned former CIA traitor, who claims to know the identity of the assassin, wants to trade that information

Next steps

Tune R^3 for identified model weaknesses (e.g. length)

- ❏ Customer tolerance for review length may be product-category dependent
  - ❏ 📚📖 Book lovers embrace long-form text
  - ❏ 💄💋 Beauty product shoppers could prefer short text/user-submitted photos

Evaluate R^3's ranked reviews using IR metrics like Mean Average Precision (MAP)

- ❏ Could not automatically measure precision/recall as dataset did not have relevance labels

Optimize R^3 for shopper happiness (both **useful and fast**)

- ❏ Once we can reliably deliver a TOP N HELPFUL + RELEVANT review list, SPEED is paramount
  - ❏ Can we process new reviews immediately upon submission? (Important for new releases.)
  - ❏ What is the fastest BERT model to compare most helpful reviews to a shopper query?
  - ❏ Confident we can pre-calculate a BERT representation of each incoming review such that shoppers' query-time comparisons are fast...but how fast is the question.