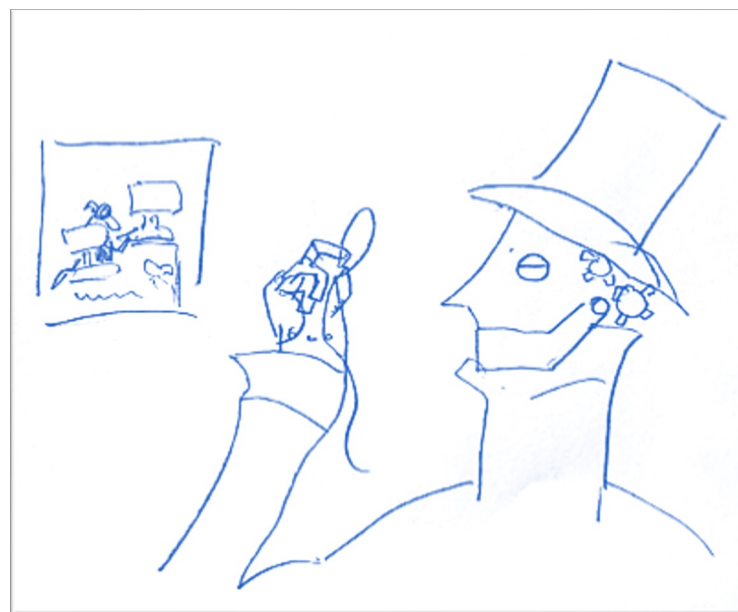# Do Androids Laugh at Electric Sheep?
Humor "Understanding" Benchmarks from The New Yorker Caption Contest

Lillian Lee, Cornell University
Eugene Charniak Memorial Symposium

Joint work with Jack Hessel, Ana Marasović, Jena D. Hwang, Jeff Da, Rowan Zellers, Robert Mankoff, Yejin Choi

My attempt to draw a robot Eustace Tilley examining the "On the Internet, nobody knows you're a dog" cartoon.

My DALL-E trials were not quite on point and/or horrifying.

Do Androids Laugh at Electric Sheep?

Humor "Understanding" Benchmarks from The New Yorker Caption Contest

Does this topic/talk honor the legacy of
Eugene Charniak?

Do Androids Laugh at Electric Sheep?

Humor "Understanding" Benchmarks from The New Yorker Caption Contest

Does this topic/talk honor the legacy of
Eugene Charniak?

I mean c'mon, it's **right after** a talk entitled
"The title may be dull, but it is a very good description', or,
some of Eugene's invaluable advice on interdisciplinary NLP research".

And where's any mention of simplicity, sparsity, or other symposium themes?

Do Androids Laugh at Electric Sheep?

Humor "Understanding" Benchmarks from The New Yorker Caption Contest

Does this topic/talk honor the legacy of
Eugene Charniak?

I mean c'mon, it's **right after** a talk entitled
*"The title may be dull, but it is a very good description', or,
some of Eugene's invaluable advice on interdisciplinary NLP research".*

And where's any mention of simplicity, sparsity, or other symposium themes?
My angle today: the interplay between *baselines* and *benchmarks.*
(Plus, TIL that Eugene told a lot of jokes)
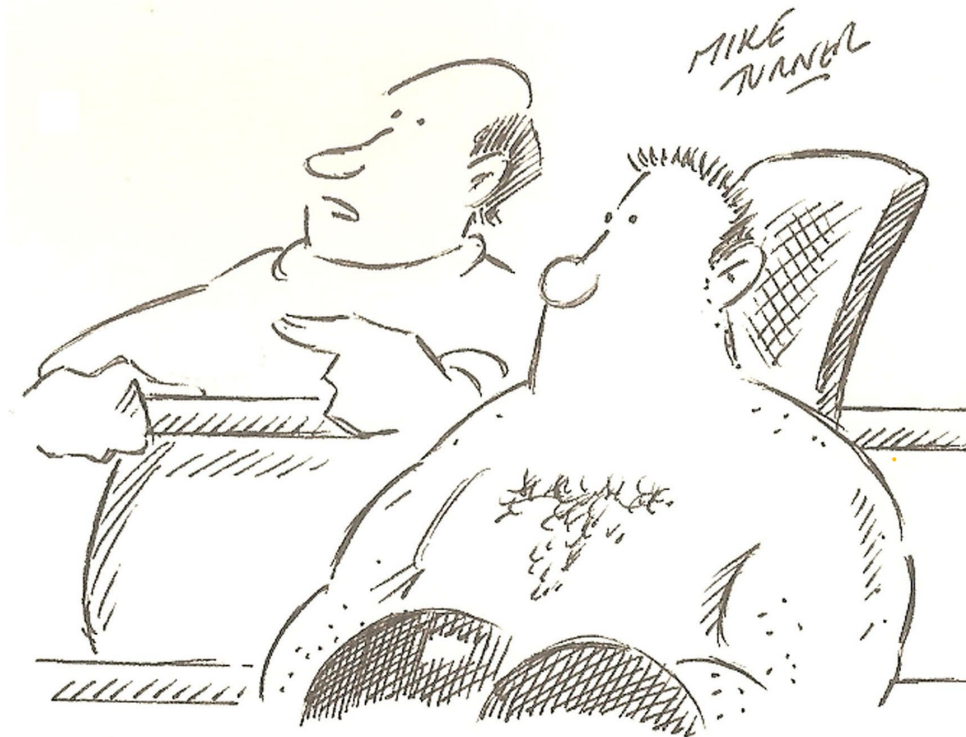
Humor "understanding" as model benchmark

(These days, we/I desperately need a word for the concept "quote-understand-endquote".)

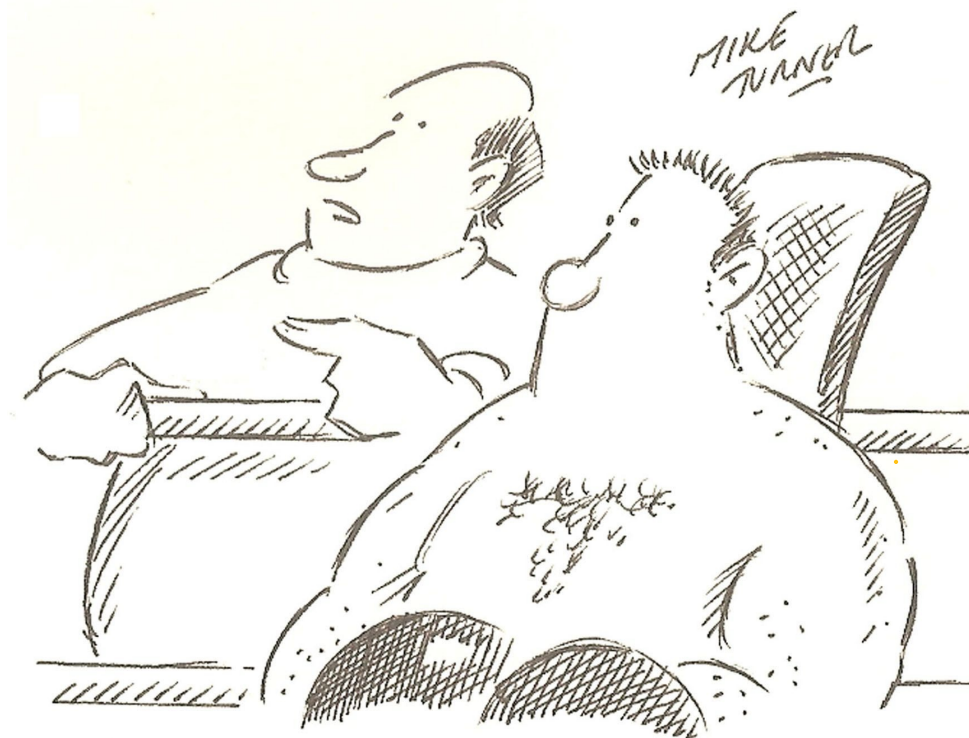To recognize or explain humor seems to require:
- recognizing a situation
- "knowing" what is expected and unexpected in the situation

Humor often draws on knowledge and the ability to reason about physical, social, and cultural worlds.

→ An AI-complete problem. *Cartoons* add multimodality.

"And your job is to hit the one who isn't wearing a bow tie."

MIKE
TURNER

"And your job is to hit the one
who isn't wearing a bow tie."

To get the joke, you need some boxing+social knowledge.

- The one without the bow tie = the boxing opponent.
- The one <u>with</u> the bow tie = the referee.

- It is odd that the boxer pictured doesn't already know how to recognize their opponent

(and perhaps some might find it funny to picture them going for the ref instead)

Incidentally, despite with all our advances in NLP, searches for "bowtie" vs. "bow tie" images/cartoons yield different results ☹

'I think the cat wants to go out.'

ALBOROZO

'I think the cat wants to go out.'

Note the "non-linear" impact of text+image combination. Alone …

- The cat's outfit is unusual, but it's not *especially* funny to dress a cat this way.
- The owner's utterance is quite typical: cats need to go outside.

But the *juxtaposition* changes our interpretation:

- The cat dressed itself!
- "go out" = "go out on the town" (have a swell time)
- [And I think the cat's expression becomes more meaningful.]

You both know Jane.

You both know Jane.

What about when you don't "get it"?

Here, we need the knowledge that "Jane" is probably Jane Goodall, famed chimp researcher.

Or, perhaps this is a Tarzan/Jane reference. (Thanks, Graeme Hirst)

An arguably common cause: not getting the **relationship** between image and text.

Data source: The New Yorker Caption Contest

"Each week, we provide a cartoon in need of a caption. [Thousands of captions are submitted], we choose three finalists [=human labels]", and you vote for your favorite [=human labels]"

Three formulations of "understanding": Different **baselines**

[Cartoon by Mick Stevens]

Matching

A) Can I interest you in an offshore account?

B) So how much of the story is autobiographical?

C) Don't give me that holier-than-thou attitude!

D) They give me free drinks if I keep my tray table down.

E) *Publicly, we are still saying there are no side effects*

Quality Ranking

I'll admit he may look ugly, but his resume is beautiful.

-vs-

🏆

Publicly, we are still saying there are no side effects

Explanation Generation

Publicly, we are still saying there are no side effects

This is a board meeting of a shady pharmaceutical company. The drug the company makes has the side effect of turning people into cartoon monsters, and most everyone at the company has taken it. Nonetheless, they are choosing not to warn the public. This plays upon a common belief that pharmaceutical companies care more about profits than they do the well-being of their patients.

Quality ranking with cartoon data isn't new

NEXT: A System for Real-World Development, Evaluation, and Application of Active Learning

Kevin Jamieson
UC Berkeley
kjamieson@berkeley.edu

Lalit Jain, Chris Fernandez, Nick Glattard, Robert Nowak
University of Wisconsin - Madison
{ljain,crfernandez,glattard,rdnowak}@wisc.edu

[Jamieson et al. 2015; NeurIPS]

Humor in Collective Discourse: Unsupervised Funniness Detection in the New Yorker Cartoon Caption Contest

Dragomir Radev[1], Amanda Stent[2], Joel Tetreault[2], Aasish Pappu[2]
Aikaterini Iliakopoulou[3], Agustin Chanfreau[3], Paloma de Juan[2], Jordi Vallmitjana[2]
Alejandro Jaimes[2], Rahul Jha[1], Robert Mankoff[4]

[Radev et al. 2016; LREC]

Inside Jokes: Identifying Humorous Cartoon Captions

Dafna Shahaf
Microsoft Research
dshahaf@microsoft.com

Eric Horvitz
Microsoft Research
horvitz@microsoft.com

Robert Mankoff
The New Yorker Magazine
bob_mankoff@newyorker.com

[Shahaf et al. 2015; KDD]

Random Walk Factoid Annotation for Collective Discourse

Ben King   Rahul Jha
Department of EECS
University of Michigan
Ann Arbor, MI
benking@umich.edu
rahuljha@umich.edu

Dragomir R. Radev
Department of EECS
School of Information
University of Michigan
Ann Arbor, MI
radev@umich.edu

Robert Mankoff *
The New Yorker Magazine
New York, NY
bob_mankoff
@newyorker.com

[King et al. 2013; ACL]

The New Yorker Cartoon Caption Contest Dataset

Thanks for your interest in this dataset. The data may not be used for commercial purposes. If you do use it for academic research, we would appreciate you referencing the dataset as follows:

Jain, L., Jamieson, K., Mankoff, R., Nowak, R., Sievert, S., (2020). The New Yorker Cartoon Caption Contest Dataset. https://nextml.github.io/caption-contest-data/

These data were gathered as part of the The New Yorker (TNY) cartoon caption contest. The crowdsourced ratings inform the final judging of the captions, but the official winner is decided by the TNY editorial staff.

[Jain et al. 2020; hosting contest data from 500-800]

A) Can I interest you in an offshore account?

B) So how much of the story is autobiographical?

C) Don't give me that holier-than-thou attitude!

D) They give me free drinks if I keep my tray table down.

E) Publicly, we are still saying there are no side effects

+

[Cartoon by Mick Stevens]

A new baseline task: Matching

- Recall the "Jane" example: "getting" a joke requires realizing the relation between image and text

- Correct choice: the winning caption.

- The others: winning captions for **different** images (contests).

- Thus, the alternatives are "natural" --- all good captions --- but (likely) wrong for this instance.

We provide highly-annotated data

"Publicly, we are still saying there are no side effects."

This is a board meeting of a shady pharmaceutical company. The drug the company makes has the side effect of turning people into cartoon monsters, and most everyone at the company has taken it. Nonetheless, they are choosing not to warn the public. This plays upon a common belief that pharmaceutical companies care more about profits than they do the well-being of their patients.

Explanation of the joke
"Explain the joke in 2-4 sentences as if to a friend who doesn't 'get it' yet"

All 650 written by Jack (crowdsourced quality too low; he deserves (another) medal)

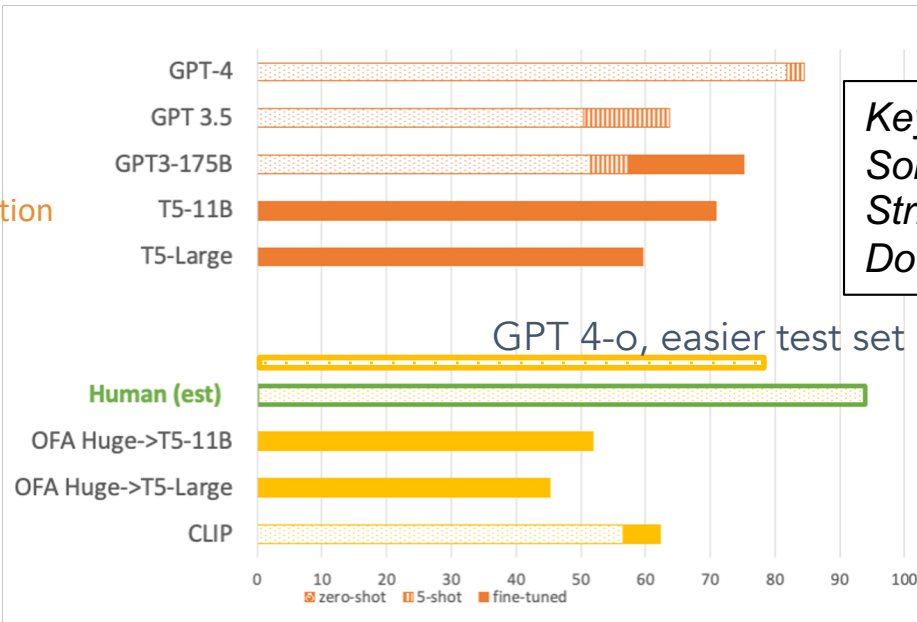| a board room	A bunch of monsters are sitting around a table. There is a man standing to the side.	Everyone is a monster except for the human.	wiki/Meeting wiki/Animal
Locations *"This scene takes place in/at/on a(n)..."*	**Descriptions** *"Describe the literal contents of the image in 2-3 sentences"*	**Uncanny** *"Highlight/explain any unusual/out-of-place elements in 1-2*	**Entity Links** *"These wikipedia links would be helpful for a robot to understand the*

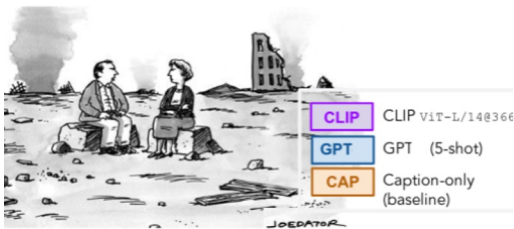## Simulations of human preprocessing and knowledge (crowdsourced)

[Cartoon by Mick Stevens]
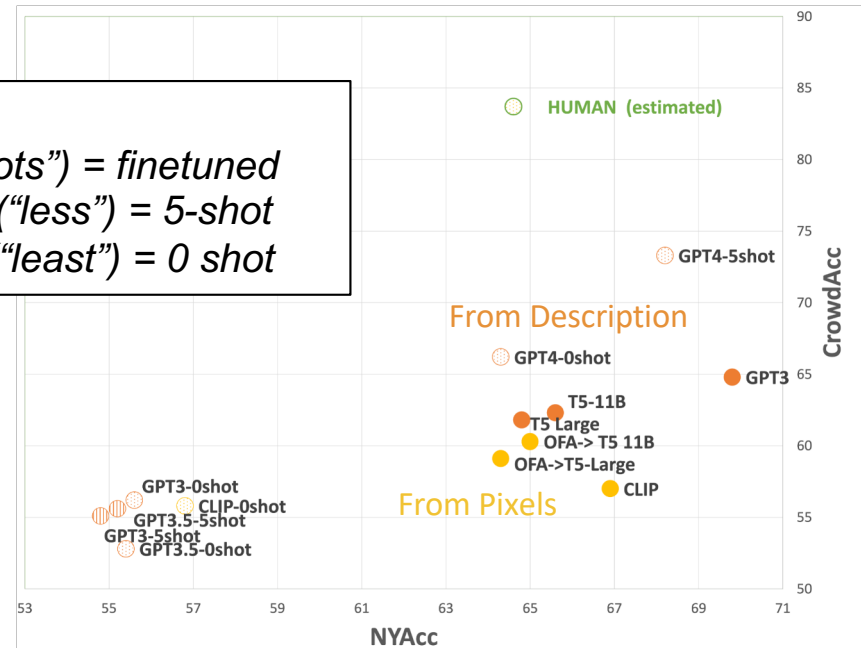
Matching Results (left)

Ranking Results (right)

From Description

From Pixels

Key:
Solid ("lots") = finetuned
Striped ("less") = 5-shot
Dotted ("least") = 0 shot

GPT 4-o, easier test set

zero-shot   5-shot   fine-tuned

From Description

From Pixels

CLIP   CLIP ViT-L/14@366
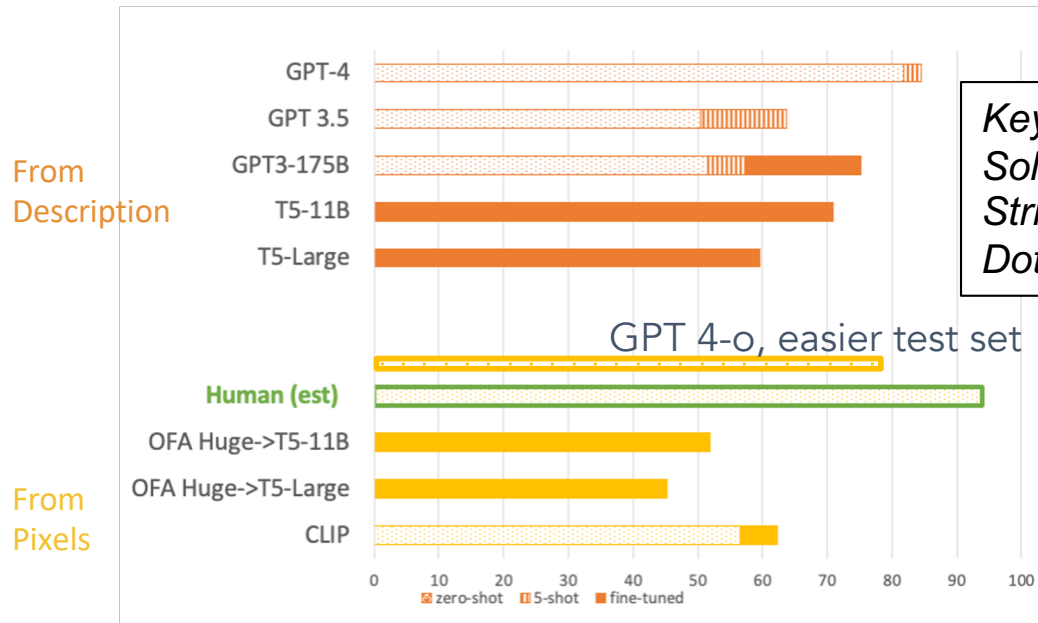GPT   GPT (5-shot)
CAP   Caption-only (baseline)

Matching

A) You should be happy. How many husbands even notice window treatments?   CAP ✗

B) I've led a good life, but now it's time to meet my raker.

C) *I'd like to see other people.*   GPT ✓

D) I think it's called an air B&B.

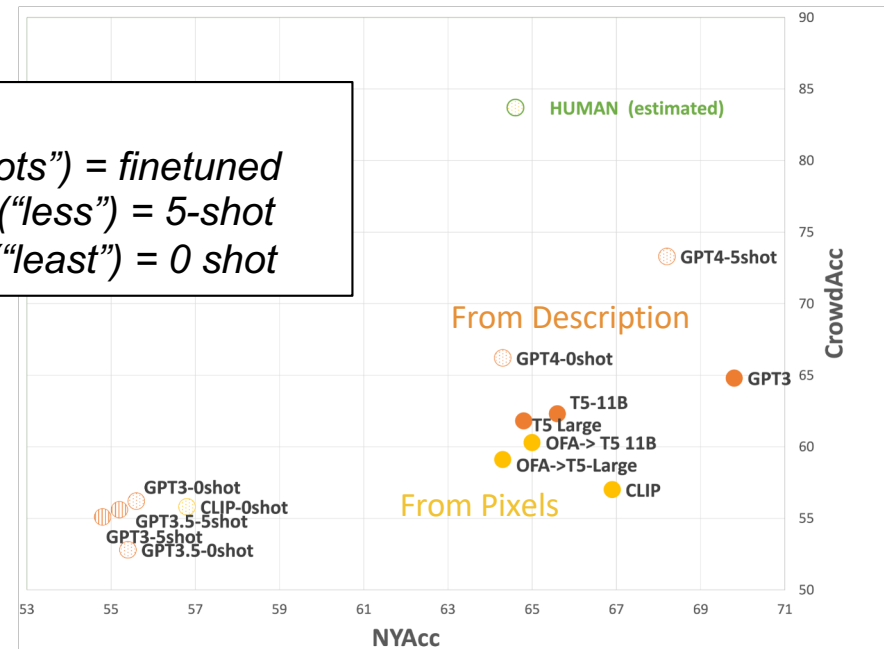E) We have to turn back. I forgot my scarf.   CLIP ✗

Quality Ranking

I'd like to see other people.   vs.   Oh well, we've survived worse

CLIP GPT ✓   CAP ✗

Matching Results (left)

Ranking Results (right)

Key:
Solid ("lots") = finetuned
Striped ("less") = 5-shot
Dotted ("least") = 0 shot

From Description

From Pixels

GPT 4-o, easier test set

zero-shot   5-shot   fine-tuned

HUMAN (estimated)

GPT4-5shot

From Description

GPT4-0shot

GPT3

T5-11B
T5 Large
OFA-> T5 11B
OFA->T5-Large
CLIP

GPT3-0shot
CLIP-0shot
GPT3.5-5shot
GPT3-5shot
GPT3.5-0shot

From Pixels

Explanations results:
2/3 of the time, human judges preferred human-authored to GPT4.

Kanin

"I want to be feared as a tyrant, loved as a father, and revered as
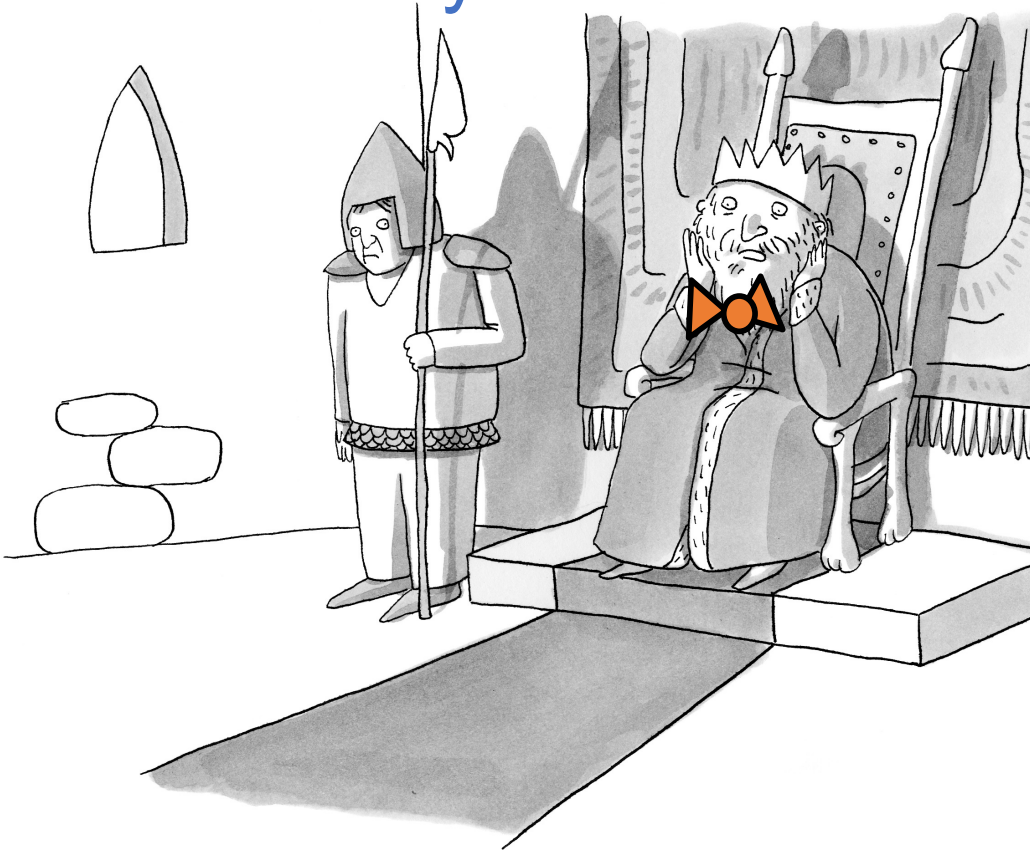a god, but I also want them to think I'm funny."

One of my all-time favorites … (and back to) Eugene

- Many people did love Eugene as a father: the testimonials I've heard from Brown alumni!

- We all revered him as a god of our field.

- And, I've learned today that Eugene liked jokes

"I want to be feared as a tyrant, loved as a father, and revered as a god, but I also want them to think I'm funny."

One of my all-time favorites … (and back to) Eugene

"I want to be feared as a tyrant, loved as a father, and revered as a god, but I also want them to think I'm funny."

- Many people did love Eugene as a father: the testimonials I've heard from Brown alumni!

- We all revered him as a god of our field.

- And, I've learned today that Eugene liked jokes