CONCEPTUAL ORGANIZATION IN THE SUPERMARKET

Adam Hornsby (@adamnhornsby), Thomas Evans, Peter Riefer, Rosie Prior & Brad Love

https://arxiv.org/abs/1810.08577

INTRODUCTION

WHAT IS A TOMATO?

WHAT IS A TOMATO?

It's a **fruit**, it's **red**, it's **fleshy** and **juicy**...

WHAT IS A TOMATO?

It's a **fruit**, it's **red**, it's **fleshy** and **juicy**...
...but is this **how consumers think?**

WHAT IS A TOMATO?

It's a **fruit**, it's **red**, it's **fleshy** and **juicy**...
...but is this **how consumers think?**

The answer helps us to **optimize in-store and online search** for customers

TOMATOES ELUDE US BECAUSE...

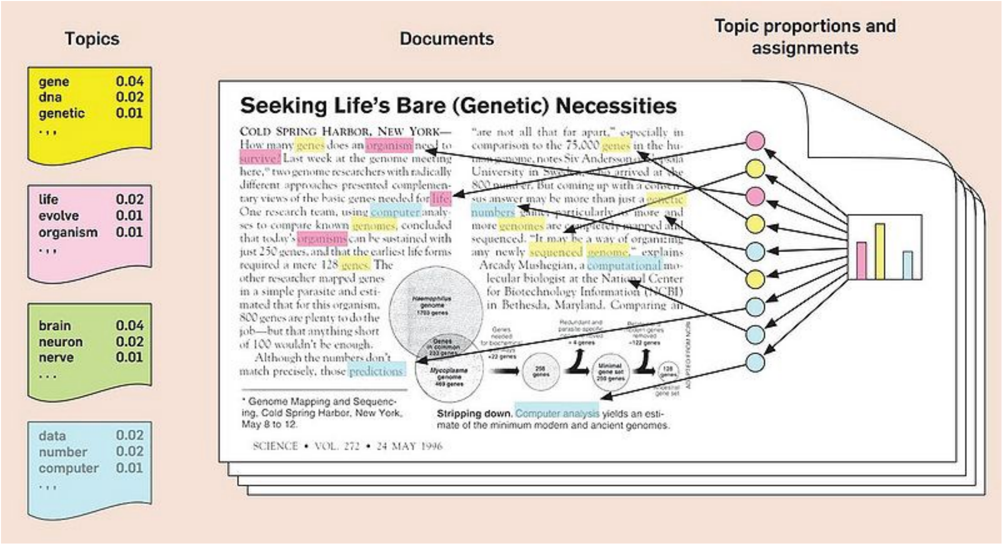
TOMATOES ELUDE US BECAUSE...

Objects gain meaning by their **interactions** with other objects (Wittgenstein, 1967, Jones & Love, 2007)

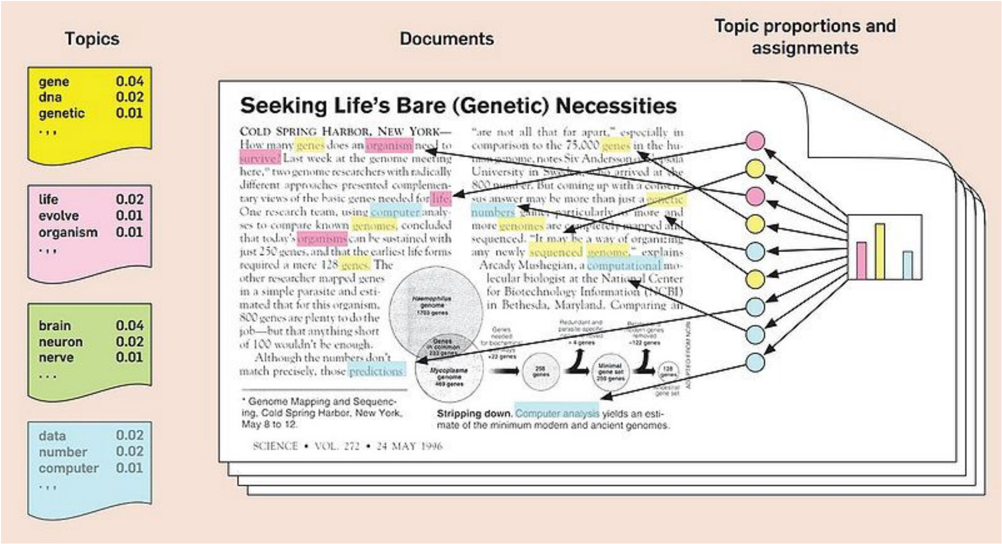
TOMATOES ELUDE US BECAUSE...

Objects gain meaning by their **interactions** with other objects (Wittgenstein, 1967, Jones & Love, 2007)
How do people **mentally categorise** products in the supermarket?

NLP RESEARCHERS KNOW ABOUT INTERACTIONS

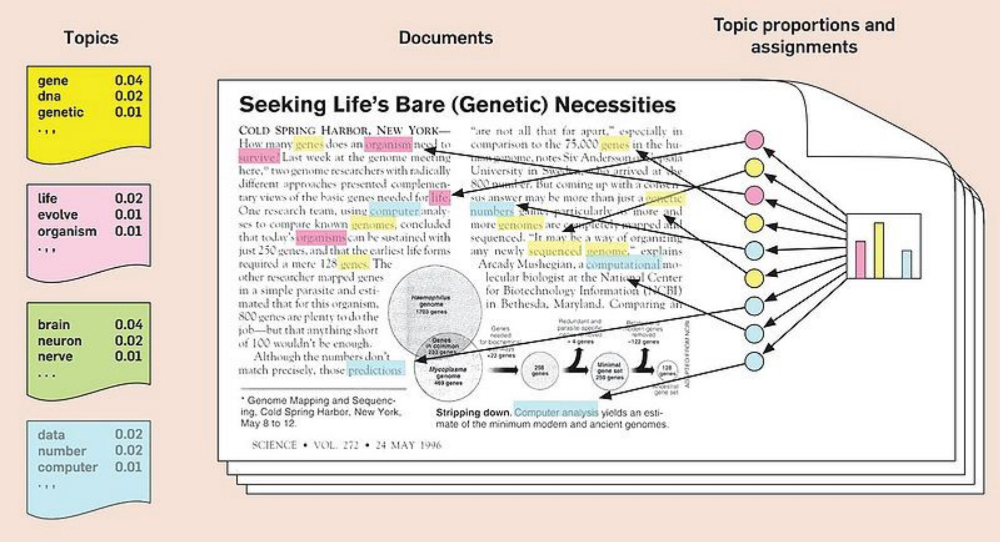


NLP RESEARCHERS KNOW ABOUT INTERACTIONS



*"You shall know a word **by the company it keeps**"* (Firth, 1957)
(i.e. Distributional Hypothesis)

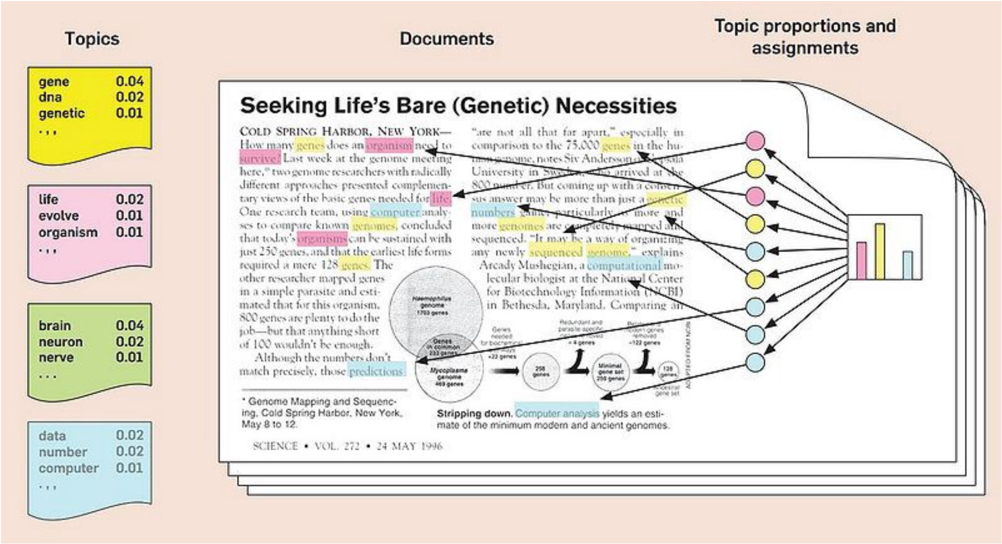
NLP RESEARCHERS KNOW ABOUT INTERACTIONS



"You shall know a word ***by the company it keeps****"* (Firth, 1957)
(i.e. Distributional Hypothesis)

Topic models (e.g. LDA) use this premise to learn high-level categories from language data

NLP RESEARCHERS KNOW ABOUT INTERACTIONS



*"You shall know a word **by the company it keeps**"* (Firth, 1957)
(i.e. Distributional Hypothesis)

Topic models (e.g. LDA) use this premise to learn high-level categories from language data
So maybe these algorithms can help us to **understand how consumers think about products?**

IMAGINE A BASKET INSTEAD OF A SENTENCE

Item	Count
Dog	1
Cat	0
Man	1
Bites	1
...	...

Item	Count
Chili	2
Lime	1
Milk	0
Banana	1
...	...

IMAGINE A BASKET INSTEAD OF A SENTENCE

Item	Count
Dog	1
Cat	0
Man	1
Bites	1
...	...

Item	Count
Chili	2
Lime	1
Milk	0
Banana	1
...	...

It is straightforward to **use existing NLP algorithms** on basket data

IMAGINE A BASKET INSTEAD OF A SENTENCE

Item	Count
Dog	1
Cat	0
Man	1
Bites	1
...	...

Item	Count
Chili	2
Lime	1
Milk	0
Banana	1
...	...

It is straightforward to **use existing NLP algorithms** on basket data
Basket data requires **less preprocessing** and is **unordered**, which suits many NLP algorithms better

IMAGINE A BASKET INSTEAD OF A SENTENCE

Item	Count
Dog	1
Cat	0
Man	1
Bites	1
...	...

Item	Count
Chili	2
Lime	1
Milk	0
Banana	1
...	...

It is straightforward to **use existing NLP algorithms** on basket data
Basket data requires **less preprocessing** and is **unordered**, which suits many NLP algorithms better

Will a **topic model** recover meaningful **categories** from basket data directly?

RESULTS

OUR TOPIC MODEL DISCOVERED SEMANTIC GROUPS

OUR TOPIC MODEL DISCOVERED SEMANTIC GROUPS

We tuned an LDA model (through Spark 1.6) on **1.2m real supermarket transactions**

OUR TOPIC MODEL DISCOVERED SEMANTIC GROUPS

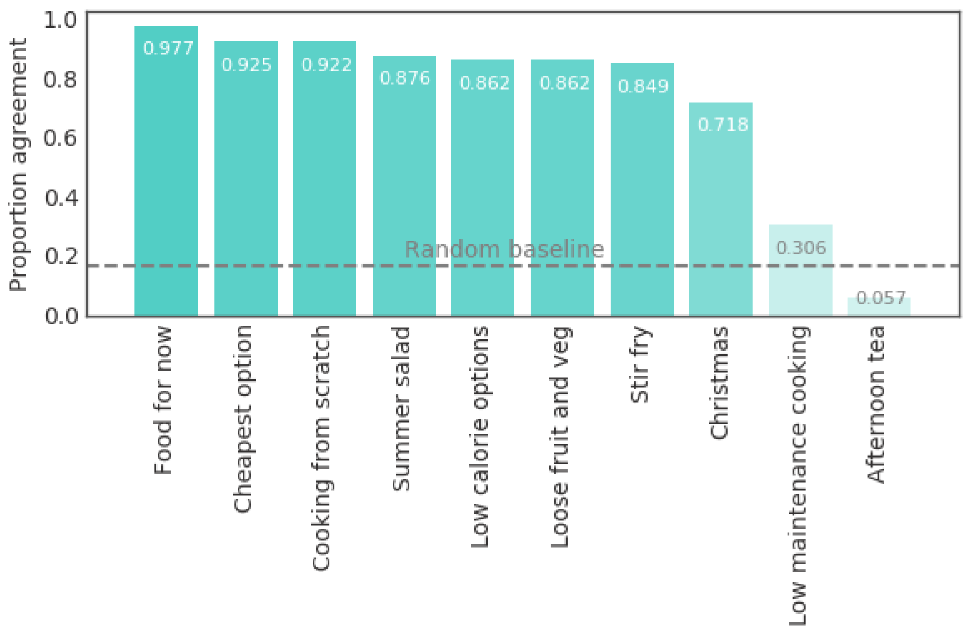
We tuned an LDA model (through Spark 1.6) on **1.2m real supermarket transactions**
The final **25 topics** appeared coherent and grouped around **specifc** (e.g. *Stir fry*) **and general** (e.g. *Cooking from scratch*), **goal-directed themes**

OUR TOPIC MODEL DISCOVERED SEMANTIC GROUPS

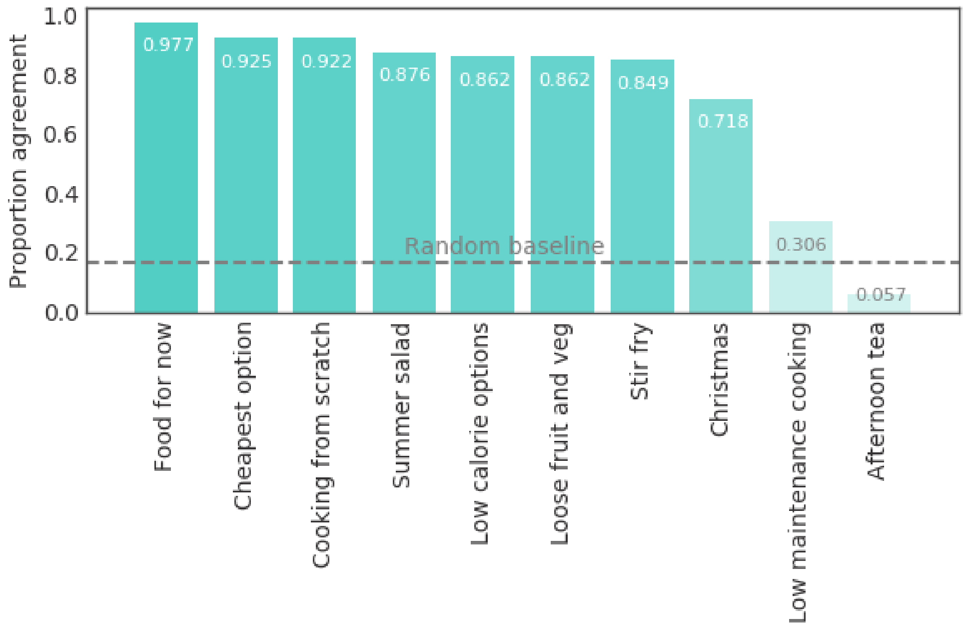
We tuned an LDA model (through Spark 1.6) on **1.2m real supermarket transactions**
The final **25 topics** appeared coherent and grouped around **specifc** (e.g. *Stir fry*) **and general** (e.g. *Cooking from scratch*), **goal-directed themes**

So **did they make sense to consumers**?

CONSUMERS AGREED WITH LDA'S TOPICS

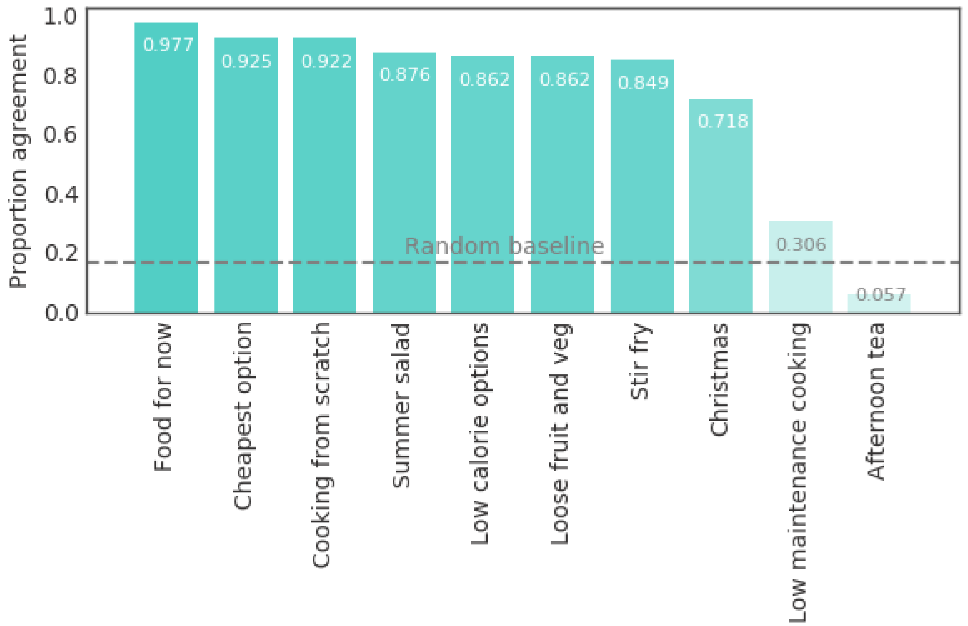


CONSUMERS AGREED WITH LDA'S TOPICS



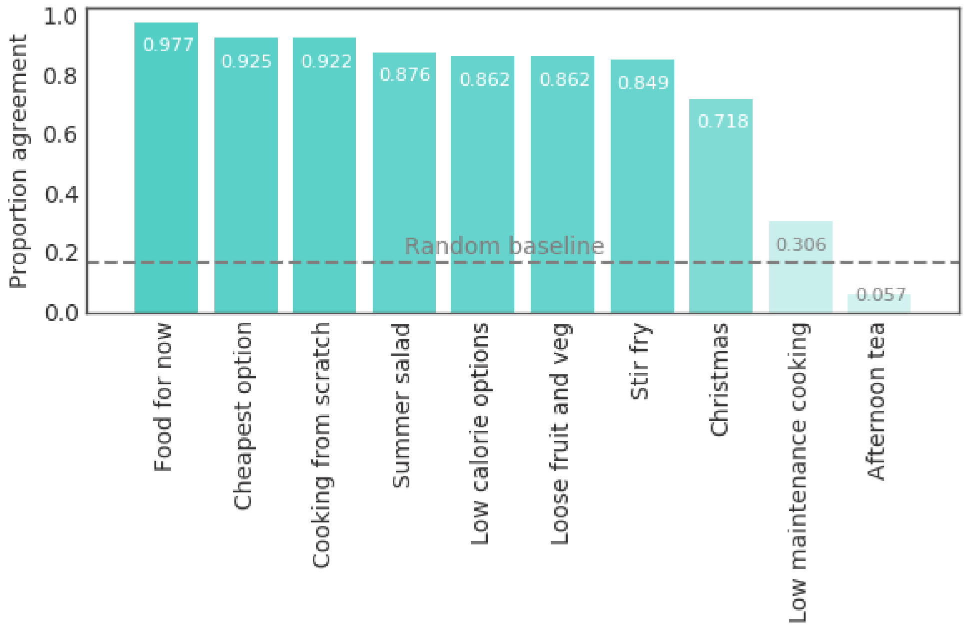
An experiment with **1000 real consumers** showed that they **could identify "intruder" products** accurately

CONSUMERS AGREED WITH LDA'S TOPICS



An experiment with **1000 real consumers** showed that they **could identify "intruder" products** accurately
Suggests that most topics were **similar to mental categories** held by consumers

CONSUMERS AGREED WITH LDA'S TOPICS



An experiment with **1000 real consumers** showed that they **could identify "intruder" products** accurately
Suggests that most topics were **similar to mental categories** held by consumers
Some were difficult (e.g. *Afternoon tea*), perhaps
due to **individual differences**

TOPICS ALSO PREDICTED INDIVIDUAL DIFFERENCES

TOPICS ALSO PREDICTED INDIVIDUAL DIFFERENCES

The topics customers purchased **predicted their self-reported age, location and gender**

TOPICS ALSO PREDICTED INDIVIDUAL DIFFERENCES

The topics customers purchased **predicted their self-reported age, location and gender**
Suggests that people's **mental representations of products may differ** between individuals and groups

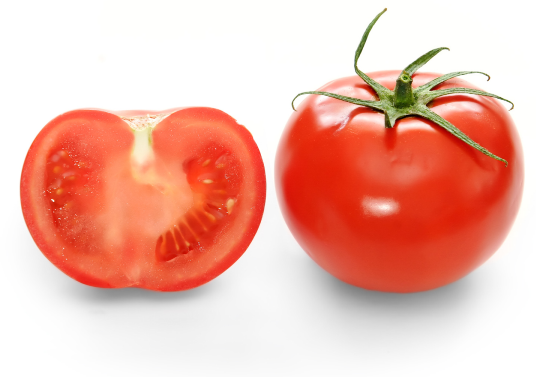
TOPICS ALSO PREDICTED INDIVIDUAL DIFFERENCES

The topics customers purchased **predicted their self-reported age, location and gender**
Suggests that people's **mental representations of products may differ** between individuals and groups

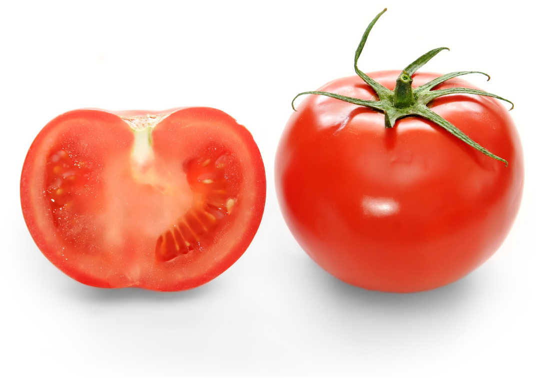
This can help us to **personalise search** algorithms

CONCLUSIONS

CONCLUSIONS

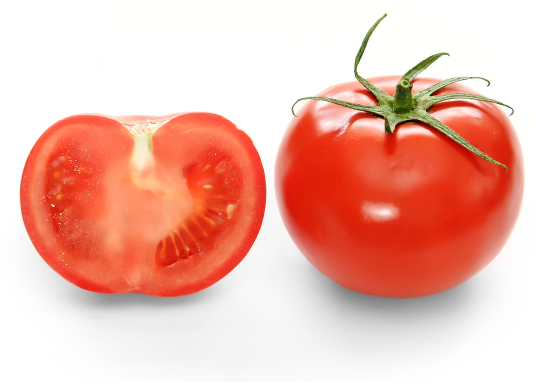


CONCLUSIONS



Consumers think about products in terms of their **roles with other products**
(e.g. "goes well in a salad")

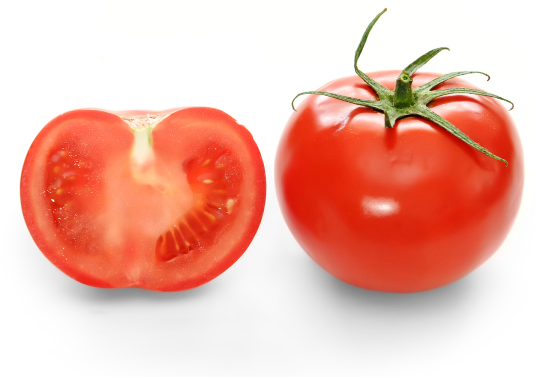
CONCLUSIONS



Consumers think about products in terms of their **roles with other products** (e.g. "goes well in a salad")

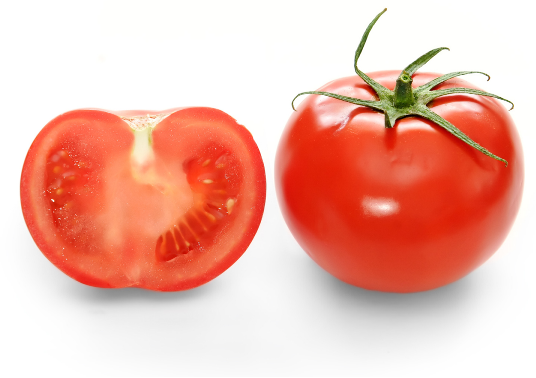
Differences in **experience may lead to different mental categories**

CONCLUSIONS



Consumers think about products in terms of their **roles with other products**
(e.g. "goes well in a salad")
Differences in **experience may lead to different mental categories**
Topic models apply to **other data sources** (e.g. baskets)

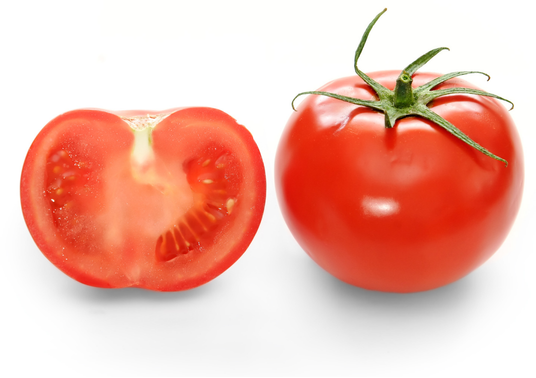
CONCLUSIONS



Consumers think about products in terms of their **roles with other products**
(e.g. "goes well in a salad")
Differences in **experience may lead to different mental categories**
Topic models apply to **other data sources** (e.g. baskets)

We are using this insight to **optimize customer's routes through the store** (both online and offline) (e.g. dual siting)

CONCLUSIONS



Consumers think about products in terms of their **roles with other products**
(e.g. "goes well in a salad")
Differences in **experience may lead to different mental categories**
Topic models apply to **other data sources** (e.g. baskets)

We are using this insight to **optimize customer's routes through the store** (both online and offline) (e.g. dual siting)
An **alternative source of data** for evaluating NLP models
(data & code available soon)

Conceptual Organization is Revealed by Consumer Activity Patterns

Adam N. Hornsby[a,b], Thomas Evans[a], Peter S. Riefer[a], Rosie Prior[a], and Bradley C. Love[b,c,1]

[a]dunnhumby, 184 Shepherds Bush Road, London W6 7NL, United Kingdom; [b]Department of Experimental Psychology, University College London, London WC1H 0AP, United Kingdom; [c]The Alan Turing Institute, United Kingdom

This manuscript was compiled on October 22, 2018

https://arxiv.org/abs/1810.08577

Supported by...

ROYAL COMMISSION EXHIBITION OF 1851

NIH National Institutes of Health

The Leverhulme Trust

wellcome trust