Message Distortion in Information Cascades

❖ Manoel Horta Ribeiro.....UFMG
Work done while interning at EFFL
❖ Kristina Gligorić.....EPFL
❖ Robert West.....EPFL

Motivation

Information often gets distorted as it propagates, this happens due to:

word of mouth

wərd-ə(v)-ˈmau̇th

Information commonly spreads in a cascading fashion, from person to person, or from platform to platform, rather than directly from the original source to every person or platform

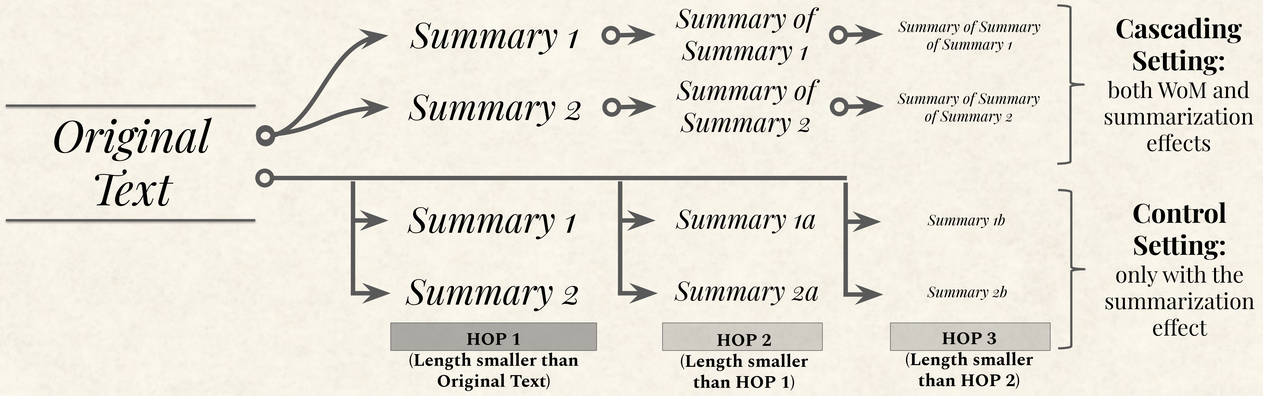
summarization

sə-mə-rə-ˈzā-shən

When an original message is passed on, it is frequently compressed, focusing on the essence while omitting unnecessary details.

Q: ***But how much of the distortion is due to each of the factors?***
A: *Don't you worry! We developed an experimental framework to study just that!*

1st idea | a crowdsourced experiment



2nd idea | ways to track information

Given the excerpt below, we illustrate two ways of tracking info., **keyphrases** and **facts**:

> We examined the association of **coffee** drinking with subsequent **total/cause-especific mortality** among **229,119** men and **173,141 women** in the NIH AARP Diet & Health Study who were **50 to 71** years of age at baseline.

Keyphrases *are short* pieces of the original text *that we track along the experiment. They are* highlighted *in the example.*

Facts *are short* statements *about the original text, for example:*

There were 229119 men & 173141 women in the participants.

For a summary, each fact is evaluated by crowdworkers as:

(A) entirely captured; *(C) insufficiently captured;* *(B) partially captured;* *(D) contradictory*

Experiment: NEJM abstracts

We experiment with our framework using abstracts from the *New England Journal of Medicine*.
As medical abstracts are structured we group keyphrases and facts into categories:

Participants	***Intervention***	***Outcomes***	***Conclusion***
Sex|Age|Condition|Location|Sample	*Duration|Intensity|Control*	*Strength|Adverse-Ef.*	

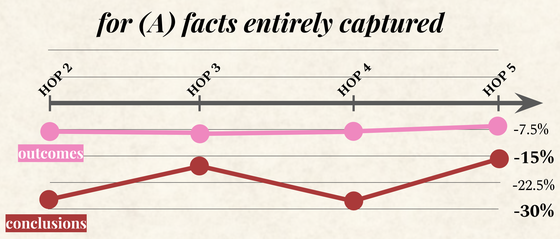
We run the experiment over 5 hops of text lengths described as follows:

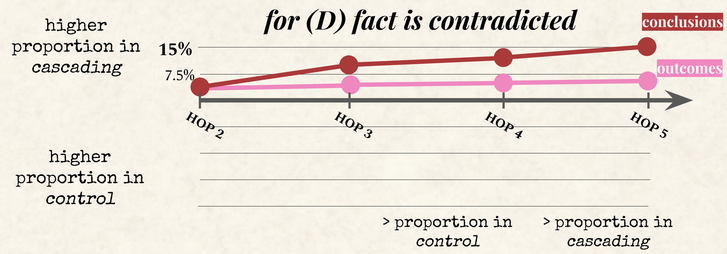
Abstract ~2048 chars → **Hop 1** ~1024 chars → **Hop 2** ~512 chars → **Hop 3** ~256 chars → **Hop 4** ~128 chars → **Hop 5** ~64 chars

Results

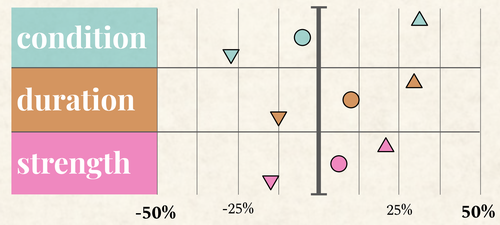
The telephone effect impacts crucial info (**conclusions** & **outcomes**) the most:

Difference in percentage points for facts in the summarization experiment:





Yet, there are nuances. Once you have a good summary, the follow-up summary (▲) is often better than control (●). Once you have a bad one, the follow-up (▼) is often way worse. We show this for **(A)** facts across all hops:

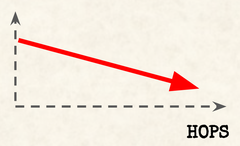


observational takeaway #1

Hop-wise, the conditional probability of

a keyphrase *to persist in a summary increases:*

a fact *to persist in a summary decreases:*



observational takeaway #2

Summaries that retain a lot of **facts** often retain a lot of **keyphrases**. Moreover, our analyses indicate that extractive summarization strategies perform better

check out the paper for more...