Message Distortion in Information Cascades

Manoel Horta Ribeiro*
UFMG
manoelribeiro@dcc.ufmg.br

Kristina Gligorić
EPFL
kristina.gligoric@epfl.ch

Robert West
EPFL
robert.west@epfl.ch

ABSTRACT

Information diffusion is usually modeled as a process in which immutable pieces of information propagate over a network. In reality, however, messages are not immutable, but may be morphed with every step, potentially entailing large cumulative distortions. This process may lead to misinformation even in the absence of malevolent actors, and understanding it is crucial for modeling and improving online information systems. Here, we perform a controlled, crowdsourced experiment in which we simulate the propagation of information from medical research papers. Starting from the original abstracts, crowd workers iteratively shorten previously produced summaries to increasingly smaller lengths. We also collect control summaries where the original abstract is compressed directly to the final target length. Comparing cascades to controls allows us to separate the effect of the length constraint from that of accumulated distortion. Via careful manual coding, we annotate lexical and semantic units in the medical abstracts and track them along cascades. We find that iterative summarization has a negative impact due to the accumulation of error, but that high-quality intermediate summaries result in less distorted messages than in the control case. Different types of information behave differently; in particular, the conclusion of a medical abstract (i.e., its key message) is distorted most. Finally, we compare extractive with abstractive summaries, finding that the latter are less prone to semantic distortion. Overall, this work is a first step in studying information cascades without the assumption that disseminated content is immutable, with implications on our understanding of the role of word-of-mouth effects on the misreporting of science.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**.

KEYWORDS

Information cascades; message distortion; information distortion

ACM Reference Format:
Manoel Horta Ribeiro, Kristina Gligorić, and Robert West. 2019. Message Distortion in Information Cascades. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3308558.3313531

* Work done during an internship at EPFL. Code/data: `github.com/epfl-dlab/mdic`

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313531

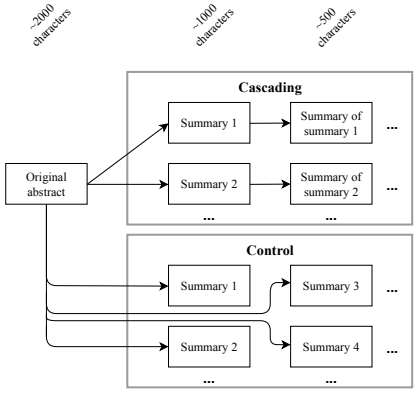


Figure 1: Schema of the crowdsourced experiment for simulating information cascades. In the cascading setting, workers summarize texts iteratively, reducing the number of characters hop by hop. In the control setting, workers always summarize the original text for all target lengths.

1 INTRODUCTION

The spread of information, online and offline, is a noisy process. As a message is passed on from person to person, or from platform to platform, errors creep in, and facts are distorted, oftentimes to an extent such that, after a few hops of propagation, downstream messages may be entirely different from—or even contradict—the original message. This way, valuable information may turn into harmful misinformation, even without purposeful interference.

Misconceptions and sensationalism add their part to compound the problem, as frequently observed in the context of health-related topics such as dieting and vaccination. For example, in 2006, the first results of the *Women's Health Initiative Dietary Intervention* trial were published. The trial found little impact of diets lower in fat and higher in fruits, vegetables, and grains in the incidence of certain diseases in women between 50 and 79 years old [25]. Shortly after its publication, a sequence of press releases and news stories increasingly distorted the nuanced and cautious findings of the study, overlooking methodological caveats and benefits found [35]. Throughout the diffusion process, news overwhelmingly reported that food and nutrition had little to do with health and disease [9].

This anecdote portrays how information may be distorted as it propagates over the news and social media. These distortion processes are overlooked in the existing literature on information diffusion [21], which treats information disseminated through a network as consisting of immutable pieces of content (e.g., *memes* [34] or topics [10]). Previous research in communication studies [14, 22] indicates, however, that the way information is altered, interpreted, or framed along its diffusion may have a significant impact.

Two orthogonal factors are at play during message propagation:

(1) **Word of mouth:** 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.
(2) **Summarization:** When an original message is passed on, it is frequently compressed, focusing on the essence while omitting unnecessary details.

Both factors can introduce errors. First, word-of-mouth propagation usually takes place on noisy channels (unless messages are forwarded unmodified, e.g., via retweets), and when an error occurs, it is passed on via what we term the **telephone effect,** named after the *telephone game,* in which *"players form a line, and the first player comes up with a message and whispers it to the ear of the second person in the line. The second player repeats the message to the third player, and so on. When the last player is reached, they announce the message they heard to the entire group. The first person then compares the original message with the final version. Although the objective is to pass around the message without it becoming garbled along the way, this usually ends up happening"* [1]. Second, summarization can be seen as lossy compression and thus induces an additional loss of information, which we term the **summary effect.**

Consider, e.g., this three-hop cascade: a ten-page medical research paper is promoted in a one-page university press release, which is picked up by a half-page newspaper article, which is finally mentioned in a tweet with 280 characters. It is clear that the tweet at the end of the cascade will be different from the original research paper. What is less clear is whether the difference stems from the fact that the message was passed on three times (the telephone effect), or from the fact that ten pages were compressed to 280 characters (the summary effect). Disentangling the telephone and summary effects is difficult when working with observationally collected information cascades as studied in prior work, e.g., URLs spreading via tweets [50] or quotations spreading via news articles and blog posts [34]. Moreover, assembling an appropriate dataset in the first place is challenging, too. For instance, in the case of the aforementioned *Women's Health* trial, it is hard even to identify the structure of cascades, i.e., the graph that specifies from which other node each node received the message. If an article misreported the findings, it is unclear if the author read the original scientific report and misunderstood it, or if they read other articles that had already introduced the error. Also, the level of coverage may vary widely: a special feature on women's health in a tabloid may briefly mention the trial, whereas an editorial in *Science* may be dedicated to it entirely. Meaningfully comparing such different formats is hard.

Present work: an experimental framework for studying message distortion in information cascades. These challenges associated with observational settings motivated us to adopt an experimental approach. In our experimental design, inspired by the telephone game, we aim to track the distortion of messages as they propagate hop by hop. Starting from an original message, we simulate an information cascade by asking a crowd worker to shorten the original message to a prescribed target length while preserving the essential information. The resulting summary is then passed on to another worker, who is asked to condense it to an even shorter target length, and so forth. This way, we obtain chains of **cascading summaries.** Along the chains, the original message is distorted by the telephone and summary effects. To tease the two effects apart, we also collect **control summaries,** of the same lengths as the cascading summaries, but produced by directly summarizing the original message to the respective target length, without any intermediate summaries. This setup is depicted in Fig. 1. Cascading summaries are subject to both the telephone and the summary effects, whereas control summaries are only subject to the summary effect, so the difference in error rates in cascading vs. control summaries can be ascribed to the telephone effect.

This experimental design allows us to address the following **research questions:**

RQ1: Measuring the telephone effect. What part of information distortion is due to the cascading of messages (telephone effect), rather than to length restrictions (summary effect)?
RQ2: Information persistence. Given that a piece of information has already survived k summarization steps, how likely is it to survive one more? What factors impact its survival?
RQ3: Extractive vs. abstractive summarization. Broadly, there are two ways of summarizing text, (1) by subselecting keyphrases (*extractive*), (2) by paraphrasing essential information (*abstractive*). How effective are these strategies in mitigating distortion introduced by the telephone effect?

Application to medical information. Motivated by the importance of medical information, we apply the above framework to a scenario where original messages are abstracts of papers published in the *New England Journal of Medicine* (NEJM). Automatically identifying the key facts contained in a medical abstract is an open challenge [45], and so is determining the presence vs. absence of those facts in subsequent summaries. Hence, in order to reliably address the above research questions, we manually annotate lexical and semantic units in the abstracts and track them along cascades.

We find that, overall, cascading summarization has a detrimental effect due to the accumulation of error. High-quality intermediate summaries, however, can have a positive effect, by isolating the essential information and discarding noise, entailing less distorted subsequent messages than in the control case. Different types of information behave differently; in particular, the conclusion of a medical abstract—the most critical information —is distorted most: the conclusion is correctly represented in cascading summaries by about 25 percentage points less, compared to control summaries, for a fixed target length. Moreover, we find that the prior knowledge of the crowd workers impact the persistence of information in the control setting, but not in the cascading setting. Finally, comparing extractive with abstractive summaries, we find that extractive summaries (where more keyphrases are preserved) are less prone to semantic distortion.

Implications. Beyond the special case of medical information, our work has general implications for the study of information cascades in a setting where information is not immutable and atomic but subject to distortion and omission. In this context, modeling the distortion of content through its diffusion in a network can help us understand the nature of viral content and the processes through which biased or erroneous information arises from a correct source. Furthermore, these insights could be used to purposefully create content that is less prone to distortion as it is diffused.

Table 1: Papers used and associated topics: vaccination (VA), breast cancer (BC), cardiovascular diseases (CD), nutrition (NU).

Paper	Topic	Paper	Topic
A population-based study of measles, mumps, and rubella vaccination and autism [36]	VA	Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes [44]	CD
Response to a monovalent 2009 influenza A (H1N1) vaccine [20]	VA	Azithromycin and the risk of cardiovascular death [48]	CD
First results of phase 3 trial of RTS, S/AS01 malaria vaccine in African children [51]	VA	Global sodium consumption and death from cardiovascular causes [41]	CD
Waning protection after fifth dose of acellular pertussis vaccine in children [31]	VA	Effect of sibutramine on cardiovascular outcomes in overweight and obese subjects [29]	CD
Adjuvant exemestane with ovarian suppression in premenopausal breast cancer [46]	BC	Changes in diet and lifestyle and long-term weight gain in women and men [42]	NU
Effect of screening mammography on breast-cancer mortality in Norway [30]	BC	Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates [52]	NU
Exemestane for breast-cancer prevention in postmenopausal women [19]	BC	Primary prevention of cardiovascular disease with a Mediterranean diet [15]	NU
Effect of three decades of screening mammography on breast-cancer incidence [7]	BC	Association of coffee drinking with total and cause-specific mortality [17]	NU

2 RESEARCH DESIGN

We describe our experimental design for collecting cascading and control summaries (Sec. 2.1), the dataset of cascades of medical information (Sec. 2.2), and our method for extracting keyphrases and facts from abstracts and tracking them along cascades (Sec. 2.3).

2.1 Collecting information cascades

We perform a controlled experiment to simulate an information diffusion process through text summarization tasks. Starting from an original text of length l_0, we leverage crowd workers to shorten it to a sequence of prescribed target lengths $l_1 > l_2 > \cdots > l_n$. More specifically, workers were given the following instructions:

> You will be given a short text (l_k characters) with medicine-related information. Your tasks are these:
>
> (1) Read the text carefully.
> (2) Write a summary of the text. Your summary should
> (a) convey the most important information in the text, as if you are trying to inform another person about what you just read;
> (b) contain between $l_{k+1} - \Delta_{k+1}$ and $l_{k+1} + \Delta_{k+1}$ characters.
>
> We expect high-quality summaries and will manually inspect some of them. Copy-pasting is disabled.

Cascading and control summaries differ with respect to the input given to workers. In cascading summaries, a crowd worker whose task is to produce a summary of length l_k is given as input a summary of length l_{k-1} that was previously produced by another crowd worker. Only for target length l_1 is the original text used as input. In control summaries, on the contrary, the input is always the original text, regardless of the target length.

For each original text, target length, and condition, we collect multiple summaries. Cascading summaries form a set of chains rooted in the original abstracts, whereas control summaries are flat broadcast trees. This is depicted in Fig. 1.

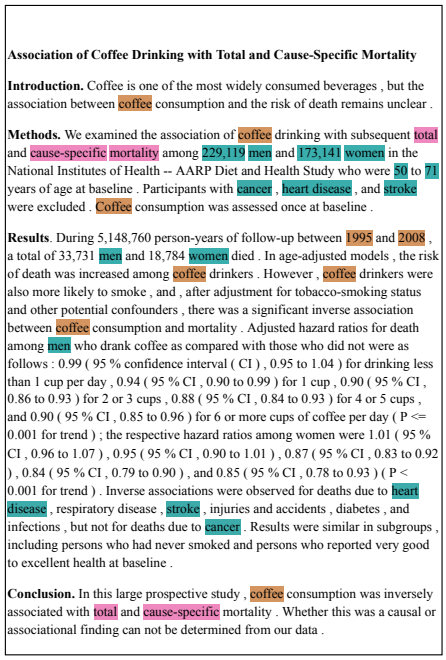
Cascading summaries are affected by both the telephone and the summary effect (as defined in Sec. 1), whereas control summaries are affected only by the summary effect, such that comparing the two cases allows us to quantify the telephone effect. One might argue that the telephone effect could be isolated in a more straightforward fashion by using a constant target length throughout, asking workers to simply rephrase the input, and thus eliminating the summary effect altogether. This, however, would allow for trivial solutions on behalf of crowd workers: nothing would keep them from simply copying the input, such that by eliminating the summary effect, the telephone effect would also vanish. This shortcoming could be addressed using a different research design, e.g., by showing workers the input text for a certain amount of time, then hiding it and asking them to reproduce it from memory. This setup, however, is more complex to implement, requires workers to spend idle time before finishing the task, and runs the risk of cheating (e.g., even when disabling copy–paste, users might take a screenshot or a picture). Also, for longer original texts (such as research paper abstracts), summarization is a step users would naturally perform in real cascades. For all these reasons, we adopt the summarization-based paradigm. Due to budget constraints, we create cascades such that each node has one descendant, since otherwise, the cost of the experiment would increase exponentially.

2.2 Dataset: cascades of medical information

In our concrete application of the above experimental design, we collect cascades whose root nodes consist of medical abstracts from the *New England Journal of Medicine* (NEJM). We select four research fields of public interest (vaccination, breast cancer, cardiovascular diseases, and nutrition), and choose 4 impactful papers per field, for a total of 16 abstracts, listed in Table 1. Analyzing the number of links to these papers in a large corpus of blog posts and news pieces, we select papers that were widely discussed online.

For each abstract, we generate 8 chains of cascading summaries, so that for each target length, each abstract had 8 summaries associated with it. In parallel, we collect another 8 independent control summaries per target length, totaling $16 \times 8 \times 2 = 256$ summaries per target length. Original abstracts were $l_0 \approx 2000$ characters long, and we consider five target lengths (which we also refer to as *hops*): $l_1 = 1000$, $l_2 = 500$, $l_3 = 250$, $l_4 = 125$, and $l_5 = 64$ characters. We allow slacks $\Delta_1 = 100$, $\Delta_2 = 50$, $\Delta_3 = 25$, $\Delta_4 = 13$ and $\Delta_5 = 9$, for each budget respectively.

For several reasons, we enforce that workers only summarize one text per hop per abstract. Firstly, we do not want a single worker to be involved in several summarization chains, as this would imply that a single unskilled or malicious worker could jeopardize the quality of all the chains. Secondly, we do not want information to leak from one summary to another. As different chains are related to the same paper, it could be that workers understood the text better if they summarized multiple summaries originating from the same paper. Due to the latter reason, we also limit workers to do summaries related to a paper in different hops 36 hours apart.



Association of Coffee Drinking with Total and Cause-Specific Mortality

Introduction. Coffee is one of the most widely consumed beverages , but the association between coffee consumption and the risk of death remains unclear .

Methods. We examined the association of coffee drinking with subsequent total and cause-specific mortality among 229,119 men and 173,141 women in the National Institutes of Health -- AARP Diet and Health Study who were 50 to 71 years of age at baseline . Participants with cancer , heart disease , and stroke were excluded . Coffee consumption was assessed once at baseline .

Results. During 5,148,760 person-years of follow-up between 1995 and 2008 , a total of 33,731 men and 18,784 women died . In age-adjusted models , the risk of death was increased among coffee drinkers . However , coffee drinkers were also more likely to smoke , and , after adjustment for tobacco-smoking status and other potential confounders , there was a significant inverse association between coffee consumption and mortality . Adjusted hazard ratios for death among men who drank coffee as compared with those who did not were as follows : 0.99 (95 % confidence interval (CI) , 0.95 to 1.04) for drinking less than 1 cup per day , 0.94 (95 % CI , 0.90 to 0.99) for 1 cup , 0.90 (95 % CI , 0.86 to 0.93) for 2 or 3 cups , 0.88 (95 % CI , 0.84 to 0.93) for 4 or 5 cups , and 0.90 (95 % CI , 0.85 to 0.96) for 6 or more cups of coffee per day (P <= 0.001 for trend) ; the respective hazard ratios among women were 1.01 (95 % CI , 0.96 to 1.07) , 0.95 (95 % CI , 0.90 to 1.01) , 0.87 (95 % CI , 0.83 to 0.92) , 0.84 (95 % CI , 0.79 to 0.90) , and 0.85 (95 % CI , 0.78 to 0.93) (P < 0.001 for trend) . Inverse associations were observed for deaths due to heart disease , respiratory disease , stroke , injuries and accidents , diabetes , and infections , but not for deaths due to cancer . Results were similar in subgroups , including persons who had never smoked and persons who reported very good to excellent health at baseline .

Conclusion. In this large prospective study , coffee consumption was inversely associated with total and cause-specific mortality . Whether this was a causal or associational finding can not be determined from our data .

Coarse category	Fine category	Facts	Keyphrases
Participants	Sex	The study was performed in women and men.	men; women
	Age	The participants were between 50 to 71 years of age at baseline.	50; 71
	Condition	Participants with cancer, heart disease and stroke were excluded.	cancer; heart disease; stroke
	Location	N/A	
	Sample size	There were 229,119 men and 173,141 women among the participants.	229,119; 173,141
Intervention	General	The study assessed the impact of coffee consumption.	coffee
	Duration	The study followed up individuals between 1995 and 2008.	1995; 2008
	Intensity	N/A	
	Control	N/A	
Outcomes	General	The study measured long-term total and cause-specific mortality.	total; cause-specific; mortality
	Effect strength	N/A	
	Adverse effects	N/A	
Conclusion	General	Coffee consumption was inversely associated with total and cause-specific mortality in non-smokers.	We annotate no keyphrases for this category.

Figure 2: *Left:* one of the 16 abstracts used in this study. *Right:* the categories employed, as well as facts and keyphrases associated with the abstract. Keyphrases are highlighted in the abstract.

To assess participants' level of domain knowledge, we use a questionnaire for each topic. For the topics *Breast cancer* and *Vaccination,* we use online quizzes approved by University of Rochester medical reviewers [2, 5]. For the topic *Cardiovascular diseases*, we use the sections *Epidemiology* and *Risk factors* of Bergman et al. [6]. For the topic *Nutrition,* we use the first section on expert nutrition advice of the UCL Nutrition Knowledge Questionnaire [32].

Table 2: Values used for annotating facts.

Fact value	Explanation
A	The fact is entirely captured in the text, omitting only insignificant details.
B	The essence of the fact is captured in the text, but a significant amount of detail was omitted.
C	The fact is not, or only insufficiently, captured in the text.
D	The fact contradicts the original text.

2.3 Annotating and tracking information

Studying information distortion requires quantifying the completeness and truthfulness of a summary: which facts from the original text are present in a given summary, and which facts have been omitted or even contradicted?

To address this challenge, we exploit the fact that medical abstracts are highly structured: key information can be attributed to a few main categories (e.g., *Participants* or *Intervention*), which may be further decomposed into multiple lower-level ones [45]. For example, *Participants* may be decomposed into subcategories like *Age, Sex,* and *Condition.* We develop two methods for tracking information related to these categories, which we explain in detail.

We annotate **keyphrases** (e.g., "mortality") in the abstracts with the categories they belong to (e.g., *Outcomes*). The phrases are then matched in subsequent summaries. Tracking keyphrases is computationally simple, but runs the risk of low recall: a fact may be expressed in a summary, but in words different from those in the input text. In other words, semantic completeness does not require lexical completeness. Moreover, it may be the case that the keyphrase is present, but the actual information is lost, e.g., if the summary contradicts the source text but shares keyphrases with it.

We also extract **facts,** short sentences stating the essential information a text conveys about a specific category. For example, for *Participants/Sample size,* a fact may be *"There were 229,119 men and 173,141 women among the participants"*. Facts of a given category may partially overlap with others; e.g., in the above example, it is mentioned that the study was performed with men and women, information which may also be present in the fact for the *Participants/Sex* category. We present a full abstract with annotated keyphrases and facts in Fig. 2. On the right-hand side, we show the hierarchical categories for both keyphrases and facts, inspired by the categories proposed in Nye et al. [45]. Notice that *Conclusion/General* is a category for which we have only facts but not keyphrases. This is the case as the conclusion often involves keyphrases from various categories; e.g., the statement *"Coffee consumption was inversely associated with total and cause-specific mortality in nonsmokers"* contains keyphrases about the intervention (*coffee*) and the outcome (*cause-specific; mortality*).

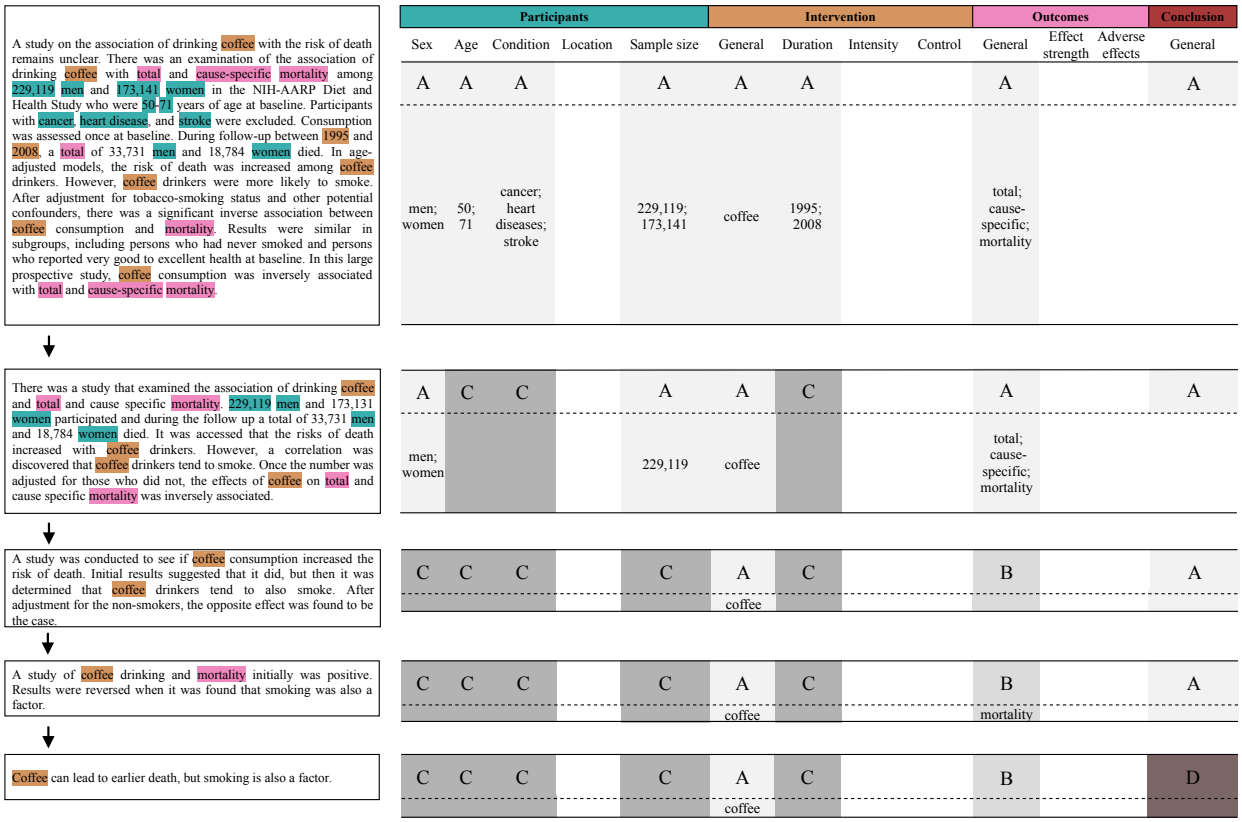


Figure 3: *Left:* a real summarization cascade obtained from crowd workers. *Right:* the annotated facts and the matched keyphrases for each of the summaries. Fact values (*A, B, C, D*) are explained in Table 2.

Tracking facts along the cascade is more complex and ambiguous than tracking keyphrases. For each summary, we assign each fact a **fact value** (*A*, *B*, *C*, or *D*). The meaning of these values is defined in Table 2. We attributed values to facts via crowdsourcing, instructing crowd workers as follows:

> You will be given several statements (around 9) and a series of short texts (of different sizes) related to medicine. Your tasks are these:
>
> (1) Read the statement and the texts carefully.
> (2) Judge how well the statement is captured by the different texts, choosing between *A*, *B*, *C*, and *D* (as explained in Table 2).
>
> The above will be repeated for several statements (the texts will be the same). Quality checks will be performed on the answers.

Attributing values to the facts is subjective: it is often not trivial to decide between *A* and *B*, or between *B* and *D*; e.g., the information that the study involved 1,067 participants can be distorted along the cascade by stating that the study involved a thousand participants. Is this fact partially preserved, as the order of magnitude is still conveyed, or does this contradict the initial fact, since $1{,}067 \neq 1{,}000$?

We take several steps to ensure annotation quality: *(i)* We add a qualification test where we explain the task thoroughly, gave examples, and assessed workers' ability to understand the subtleties of the task. *(ii)* We show texts of different sizes at once to ensure that workers could compare what a piece of information would look like in several lengths, making the annotation of a given fact consistent across several summaries. *(iii)* We introduce quality checks (facts that are wrong for all statements), allowing us to filter workers who failed often. *(iv)* Lastly, we assign three workers per fact and manually review all facts, giving emphasis to those where not all three workers had annotated a given fact with the exact same value (33% of all cases).

Example. We present a summary cascade on the left-hand side of Fig. 3. The summaries refer to the abstract from Fig. 2. The findings are nuanced: coffee is associated with an increased risk of death; yet, when accounting for confounding variables, such as smoking, the association is reversed. In the cascade, as the summaries become shorter, this subtlety becomes increasingly difficult to grasp. In the last summary, we eventually even arrive at a contradiction.

Annotated keyphrases and extracted facts are shown on the right-hand side of Fig. 3. Firstly, notice that the sudden contradiction we just described is captured, as the label *D* was assigned to the *Conclusion/General* fact in the last summary. Another interesting transition here is how the *Outcomes/General* fact shifts from *A* to *B* in the third summary: both previous summaries indicated that the summary measured total and cause-specific mortality, whereas the third one simply indicates that it measured the risk of death, which is less complete.

To see the limitations of tracking keyphrases and how they are overcome by tracking facts, consider the second summary, which

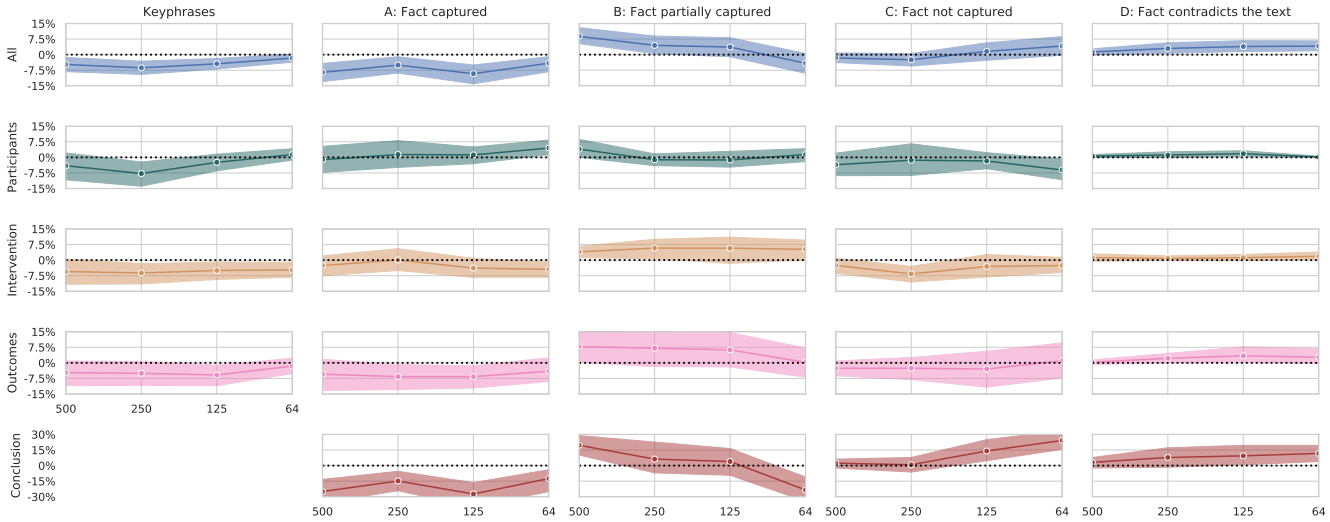


Figure 4: Differences in percentage of preserved keyphrases (column 1) and fact values (columns 2–5) between cascading and control summaries for fact categories (rows) and hops (x-axes), with bootstrapped 95% confidence intervals.

contains a typo: it says that the study was performed in 173,131 women instead of 173,141. Also, the worker used the expression "risk of death", instead of "mortality". These examples show that keyphrases can be difficult to track when there are typos or when expressions are rephrased.

3 RESULTS

3.1 RQ1: Strength of the telephone effect

In our first analysis, we compare the distortion of information in the cascading and control conditions.

Differences across hops. We calculate the percentage of keyphrases in the summaries of a given hop and the percentage of facts with a given value for all papers in the cascading and control settings. We depict the difference between the two settings in percentage points in Fig. 4. Values above 0% mean that the percentage of keyphrases, or of facts of a given value, is higher in the cascading summaries. We observe several interesting patterns. Cascading summaries preserve fewer keyphrases in all categories at all hops. This difference decreases in the last hops, perhaps because they require users to adopt more abstractive summarization strategies in both settings. The percentage of facts labeled as *A* is lower for cascading than for control summaries. This effect is particularly strong for the *Outcomes* and *Conclusion* categories. In the latter, the difference is greater than 15%. The difference in the percentage of facts labeled as *D* has an increasing trend. Here too, the effect is strongest for the *Outcomes* and *Conclusion* categories. Lastly, facts labeled as either *B* or *C* do not present a clear trend for the *Participants, Intervention,* and *Outcomes* categories. For the *Conclusion* category, however, the difference in *B*-valued facts decreases along the hops, while the percentage of *C*-valued facts increases. Overall, these differences suggest that the "telephone effect" impacts the quality of summaries significantly, particularly harming the essential facts—the outcomes and conclusion of a study.

Category distribution. We also analyze how the distribution of categories differs for the distinct settings (Table 3). Consider the percentage of facts labeled as *A* that belong to the *Participants* and *Conclusion* categories. For cascading summaries, the percentage of *Participants* facts consistently stays above 29%, while this percentage drops significantly to 14% for control summaries. For the *Conclusion* category, we have the inverse: the percentage of *A* statements in the *Conclusion* category decreases in the cascading group and remains stable (above 10%) in the control. This suggests that the type of message that ends up getting spread differs between the cascading and control summaries: cascading summaries seem to "remember" less pertinent facts about participants, whereas control summaries preserve more crucial facts about study conclusions.

Table 3: Distribution over categories (*Participants, Intervention, Outcomes, Conclusion*) for fact values (*A, B, C, D*). Results for control setting in bold.

Tgt. len.		500	250	125	64
A	**P**	38%/**35%**	36%/**32%**	32%/**25%**	29%/**14%**
	I	34%/**33%**	36%/**34%**	41%/**37%**	50%/**52%**
	O	17%/**18%**	16%/**18%**	18%/**20%**	17%/**21%**
	C	10%/**14%**	12%/**16%**	8%/**18%**	4%/**13%**
B	**P**	23%/**23%**	17%/**23%**	14%/**19%**	13%/**11%**
	I	24%/**26%**	27%/**21%**	25%/**20%**	24%/**15%**
	O	35%/**39%**	37%/**37%**	38%/**37%**	39%/**36%**
	C	18%/**12%**	19%/**19%**	23%/**25%**	24%/**38%**
C	**P**	46%/**47%**	48%/**46%**	47%/**47%**	45%/**48%**
	I	24%/**24%**	27%/**29%**	28%/**29%**	28%/**29%**
	O	27%/**28%**	23%/**23%**	20%/**21%**	21%/**21%**
	C	2%/**1%**	3%/**2%**	5%/**2%**	6%/**2%**
D	**P**	22%/**20%**	21%/**20%**	18%/**0%**	2%/**0%**
	I	26%/**20%**	15%/**15%**	13%/**17%**	15%/**0%**
	O	19%/**27%**	24%/**20%**	24%/**0%**	24%/**18%**
	C	33%/**33%**	41%/**40%**	45%/**83%**	59%/**82%**

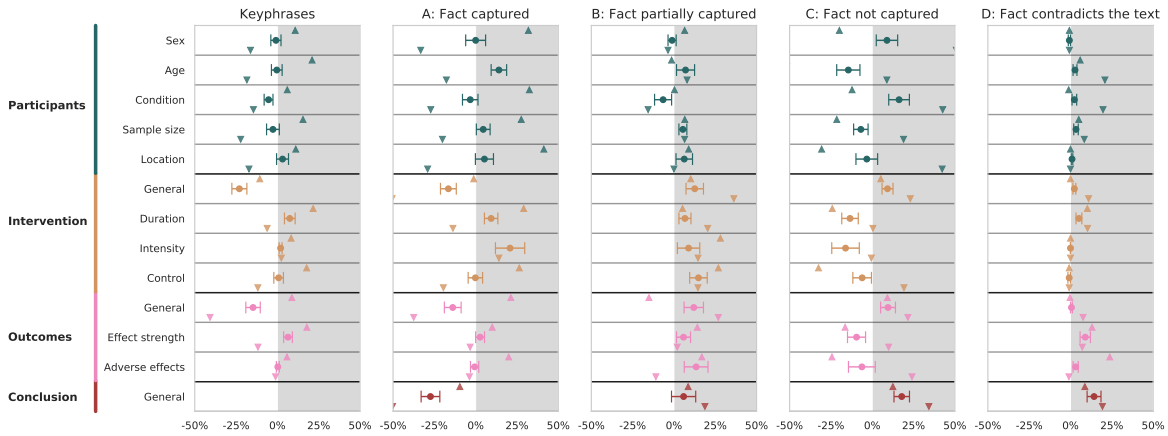


Figure 5: Differences in the percentage of facts and keyphrases between cascading and control summaries for fine-grained categories (averaged over all target lengths). On the positive side of the x-axis (in gray), the percentage of facts/keyphrases is higher in the cascading scenario. Each category is associated with three values (with bootstrapped 95% CIs): The circle ○ represents the actual difference between the cascading and control settings. The upward △ (downward ▽) triangle represents the difference between the successor of the best (worst) cascading summary from the previous hop and control.

A closer look. We now zoom into the fine-grained categories to drill deeper with regard to the differences between the cascading and control settings. We study the difference between the percentage of keyphrases and facts, averaged across hops, as depicted by the circles (○) in Fig. 5. We find that the differences are largest in what can be considered the most important facts. For example, the categories *Conclusion/General*, *Intervention/General,* and *Outcomes/General* have the biggest difference for facts labeled as *A* and for keyphrases (recall that there are no conclusion-related keyphrases). The percentage of facts labeled as *C* and *D* for these categories is also higher in the cascading scenario. This is interesting as one might expect that the telephone effect should impact peripheral categories such as *Duration* the most, whereas the opposite is the case. Analyzing the telephone effect from this perspective, we find that iterative summarization creates a "tunnel vision" effect, where less important information often moves to the fore when multiple summarization steps are involved.

The bright side of the telephone effect. Cascading summaries clearly have the disadvantage of propagating errors. However, they also have a potential advantage: given that one person did an excellent job at summarizing a text, it may be that the text they created is more straightforward for the next person to understand and thus to summarize. To determine the existence and strength of such a hypothetical positive effect, we proceed as follows. For each paper and each hop in the cascading setting, we consider only the summary whose source was the "best" of all summaries in the previous hop, where we consider the best summary to be the one with the largest number of *A*-labeled facts. We then compare the percentage of facts and keyphrases per fine-grained category between these summaries and the ones in the control setting as previously done. We show the values for this comparison for each fine-grained category as an upward triangle (△) in Fig. 5. Similarly, we also show the values for an analogous comparison where the source summary was the "worst" summary, depicted as a downward triangle (▽).

By inspecting facts labeled as *A*, we find that the "best-ancestor" scenario preserves facts better or similarly to the control setting (except for the *Conclusion/General* and *Intervention/General*category). This shows that a summary can, in fact, be a better reference text than the original abstract itself. A potential cause for this improvement might be that, although the telephone setting causes distortions, it also lowers the cognitive load, as summarizing jargon-filled 2000-character abstract is significantly harder than summarizing a 500-character text that was written by someone whose knowledge on the subject may be closer to one's own. On the other hand, the "worst-ancestor" scenario highlights the dangers of the telephone effect, where one sloppy summarization may distort the facts present in the text and harm subsequent summaries.

3.2 RQ2: Hop-to-hop information persistence

Next, we calculate the *conditional chance of survival* for keyphrases and facts across hops. That is, given that a fact or a keyphrase is present in a given hop k, what is the chance that it will also be present in hop $k+1$? In the case of keyphrases, for a given summarization process and category, we define the survival probability as the percentage of keyphrases that continue to exist in the summary, out of those that already exist in the reference text. For instance, if there are three keyphrases in the *Outcomes* category in hop 1, and two remain at hop 2, the probability of survival for this category from hop 1 to hop 2 is 2/3. In the case of facts, we consider the number of statements of a given category that are fully (fact value *A*) or partially (*B*) preserved in the source text and continue to be so in the resulting summary. The mean values along with a linear regression line for different categories can be seen in the first column

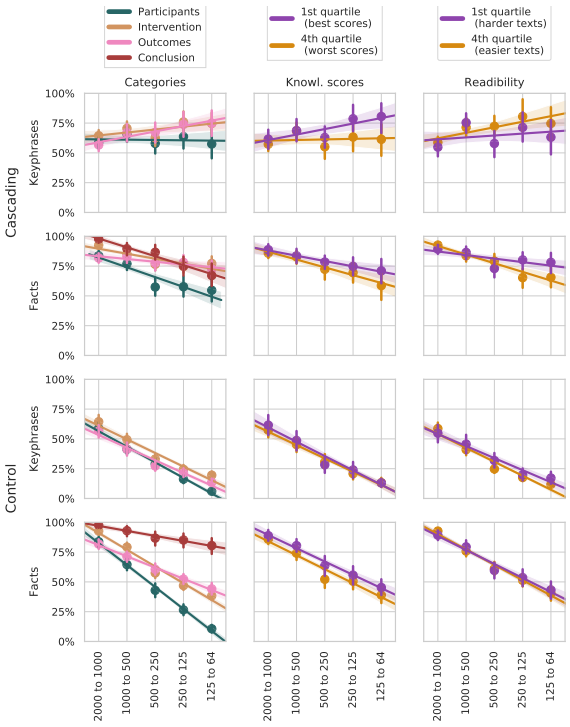


Figure 6: Probabilities of surviving one more hop for keyphrases/facts, alongside regression lines. Each row represents a combination of keyphrases/facts and a setting (cascading/control). In the first column, we compare the survival chance across different categories. In the other two, we compare it for different values of workers' questionnaire scores and of the readability index of the original abstract.

of Fig. 6. Interestingly, we observe that, while by construction the target length of texts decreases exponentially hop by hop, survival probabilities decrease or increase only linearly.

Cascading vs. control. We continue to inspect the first column of Fig. 6. The probability of survival of keyphrases increases for cascading summaries across hops for all coarse categories. This makes sense intuitively: if a keyphrase was selected for a summary, it is likely that it is relevant and thus will be selected again. In the control scenario, keyphrases always exist in the reference text (because it is the original abstract), so the probability of survival decreases as target lengths decrease. Analyzing facts, however, we find a significant difference between our experimental conditions: the probability of survival for facts in the control *and* cascading settings decreases across hops, implying that different dynamics govern distortions associated with lexical vs. semantic elements of a text. Whereas survival probabilities increase for keyphrases in cascading summaries, they decrease for facts. We conjecture that some salient keyphrases have an inherent "fitness" for survival, but that this is less true for abstract facts, which may be more or less salient, and thus "fit", in their concrete surface forms.

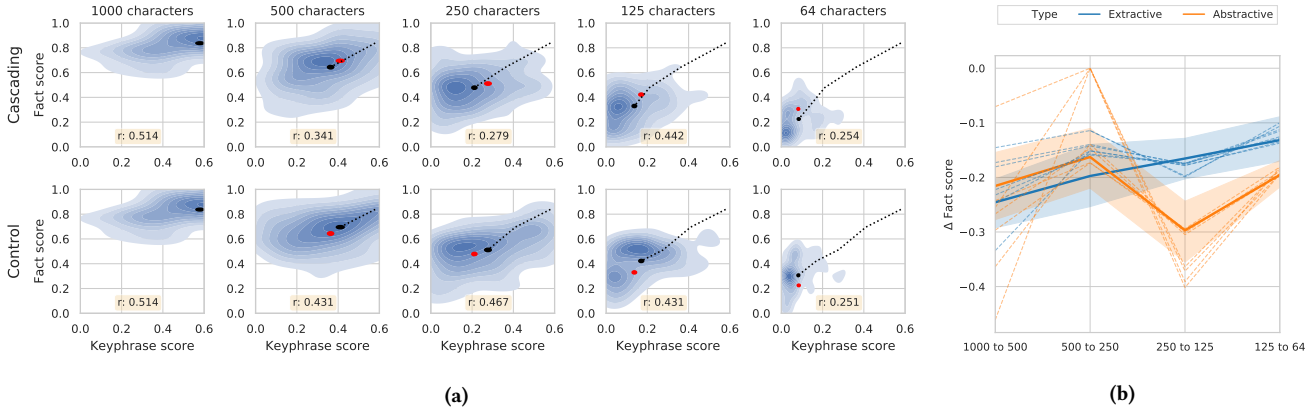
Categories. Moreover, we can also find differences in the dynamics of distortion across different categories. We observe that categories have different survival probabilities, especially for facts. For control summaries, the survival chance of *Conclusion* facts is higher than for the facts of other categories. Also, for *Conclusion*, the survival chance in cascading summaries is lower than for control summaries. This suggests that using the original abstract as a reference makes it more likely that the reader will understand the conclusion. Notice that this is different from what we observed in Sec. 3.1, as there, it could be the case that the conclusion prevailed in the control setting simply because it got lost in the cascade. Here we see that, even when the conclusion fact is present in a given summary in the cascading setting, it is less likely that it will survive.

Knowledge questionnaire. We consider the influence of the worker's knowledge questionnaire score (Sec. 2.2) on the level of distortion. At each hop, we rank the texts according to the score of the workers who summarized them on the questionnaire of the topic associated with the original abstract, and then compare the survival chance of facts and keyphrases of two extremes: the texts summarized by workers whose score is in the first quartile (who performed the best in the test) and the texts summarized by workers whose score is in the fourth quartile (who performed the worst). This is shown in the second column of Fig. 6. We use Chow's test [13] to assess whether the coefficients of the regression are significantly different, finding that, in the control setting, workers who scored better in the test lost significantly fewer facts than those who did poorly ($p < 0.01$). For cascading summaries, keyphrases survive more for workers that scored well in the text ($p < 0.05$), whereas the difference in the survival of facts is not significant. A hypothesis for this disparity is that in-depth knowledge on the subject is essential to read the original abstract (as in the control case), but that, once the text has been summarized by someone else (who may not possess that knowledge), the noise is already introduced, and the effect of the level of knowledge fades away.

Readability. We also consider the influence of the number of difficult words *in the original abstract* on the survival of facts and keyphrases. We order texts according to the percentage of words that *(i)* have more than one syllable and *(ii)* are not in a list of the most frequent English words (as done for readability metrics such as the Dale-Chall Readability Score [12]). We compare the cascades of the four most readable abstracts (first quartile) with the four least readable (fourth quartile). This is depicted in the third column of Fig. 6. We find a different effect to what we observed when inspecting the knowledge questionnaires: facts in more readable abstracts do not a have a significantly different of survival in any setting according to Chow's test. Keyphrases survive more for more readable abstracts in the cascading setting only ($p < 0.05$). Altogether, these results suggest that the telephone effect alters not only the facts and keyphrases of the summaries, but also how other factors influence the distortion processes over these facts and keyphrases.

3.3 RQ3: Extractive vs. abstractive strategies

Another aspect that may influence distortion effects in information cascades is the summarization strategy employed. Here we compare *abstractive* and *extractive* summarization strategies. To be able to do so, we have to distinguish abstractive and extractive summaries, as well as successful and unsuccessful ones. To this end, we define



(a) (b)

Figure 7: (a) Kernel density estimates of summaries according to their fact scores and keyphrase scores for different hops in cascading and control settings. Ellipses mark centroids (height and width: 95% CIs; black: experimental setting in question; red: other experimental setting). (b) Loss of fact score per summarization step for extractive vs. abstractive texts in the cascading setting. Lines correspond to multiple combinations of $\alpha \in [0.4, 0.7]$ and $\beta \in [0.0, 0.15]$ (95% CIs plotted for $\alpha = 0.3, \beta = 0.05$).

keyphrase scores and *fact scores* for each summary. The keyphrase score is defined as the percentage of keyphrases that the summary retained from the original abstract, whereas for fact scores we take a weighted average of the percentages of *A*- and *B*-valued facts retained from the original abstract (assigning weight 1 to *A*-valued facts and weight 0.5 to *B*-valued facts). With these definitions, we can plot summaries as points in a two-dimensional plane. We depict the kernel density estimate of these points in Fig. 7a. In this figure, we consider points with higher fact scores (y-axis) to be better summaries, and points with higher keyphrase scores (x-axis) to be more extractive. This way, the figure allows us to visually track the quality and extractiveness of summaries.

In each plot of Fig. 7a, we display a black ellipsis marking the centroid of the summary cloud (the height and width of the ellipsis represent 95% confidence intervals), and a red ellipsis marking the centroid for the other experimental setting. We also plot the hop-by-hop trajectory of the centroid as a dotted line. Notice the diagonal trajectory: summaries go from better and more extractive to worse and less extractive. Comparing the centroids for the cascading and the control settings, we also see that the centroid of the cascading setting always lies southwest of that of the control setting: cascading summaries are worse as well as less extractive.

In all scenarios, the fact scores are correlated with keyphrase scores (Pearson's r between 0.25 and 0.51), as can be seen in Fig. 7a. This makes it particularly hard for us to fairly assess summarization strategies, as it may be that the worst summaries are considered "abstractive" simply because they do not contain keyphrases (while also not containing the facts). Thus, our analysis has to condition the comparison in a way that we compare summaries that are similar in "goodness", but that differ in the level of extractiveness (which we capture using the defined scores).

We proceed as follows: for each hop and abstract in the cascading setting, we match summaries in pairs. Summaries are only matched if *(i)* the relative difference in their keyphrase score is *greater* than α and *(ii)* the relative difference in their fact score is *less* than β. We greedily match each summary once. Within each pair, we consider the summary with the higher keyphrase score to be more extractive (while of similar quality). We then compare how the fact score of each summary in the pair decreases with the subsequent summarization. The idea is that if extractive texts lose more of their fact score than abstractive texts, this is evidence that extractive summarization is less effective (and vice versa).

We show the results for each of summarization step in Fig. 7b for several values of α and β. The plot suggests that even when we account for the correlation between keyphrases and facts, we observe that extractive summaries are beneficial to latter hops, but similar in the first ones. In our data, thus, although abstracts were often full of jargon and technicalities, trying to change important lexical parts of the text (which we tagged as keyphrases) seems to be detrimental Our analysis here is limited as we are considering the survival of keyphrases to be a proxy for extractiveness, which may not necessarily hold in all cases.

4 DISCUSSION

4.1 Summary of findings and implications

In this paper, we propose an experimental framework for studying message distortion in information cascades and assess the diffusion of information from selected medical abstracts. Our analyses suggest that information cascades, as captured by iterative summarizations, distort the examined medical texts lexically and semantically. Facts and keyphrases frequently are not captured or even contradict the original text. The "telephone effect" impacts the most essential information the most, in particular the conclusion of abstracts. Overall, the content of the message after cascading summarization differs considerably from the content after direct summarization.

The telephone effect is, however, nuanced. Firstly, it is not necessarily bad: good summaries may serve as stepping stones toward better downstream summaries. Moreover, cascading summaries attenuate the impact of the complexity of the original text as well as the impact of users' topic-specific knowledge. Our findings suggest

that influential platforms or users may have a disparate impact on the quality of information being spread, resonating with the narrative that the rise of online social networks, where messages written by anyone can have far-reaching impact, has diminished the quality of the information [26].

We also investigated the success of abstractive vs. extractive summarization in information cascades. Even when accounting for the correlation between keyphrases and facts, we still find that extractive summarization performs better. This insight has implications on how scientists and the press should cooperate to convey research to the general public. Our findings suggest that scientific media coverage should make an effort not to distort the key terms with which authors use in their research.

Lastly, although keyphrases and facts are correlated, the conditional probability of survival of keyphrases increases hop by hop in information cascades, whereas it decreases for facts—a finding that should be taken into consideration when modeling the diffusion of information using keyphrases as proxies for associated facts.

4.2 Related work

Similar methodologies. Previous work has considered similar experimental settings to study distinct phenomena. Mesoudi and Whiten [38] propose a setting similar to our iterative cascades (among others) to study cultural transmission and evolution. Moussaïd et al. [40] use the equivalent of what we call a one-hop iterative cascade to study the propagation of risk information. A similar analysis of how textual content mutates was done in an observational setting by Adamic et al [3]. The researchers studied how memes evolve through Facebook analyzing the adaptation of textual features, as well as their impact on the propagation of the meme.

Word of mouth and customer behavior. Word of mouth is an important phenomenon driving customer behavior. Prior work has studied how the level of customer satisfaction impacts engagement in word-of-mouth information diffusion [4] and how effects of word of mouth together with other factors impact customer persuasion [24]. Word-of-mouth effects have also been analyzed in a networked framework to demonstrate different roles played by weak and strong social ties and the relational properties of homophily on referral behavior [23].

***Bona fide* vs. intentional information distortion.** Agents involved in word of mouth distort the information even given best intents. In contrast, a rich body of work seeks to detect, model, and prevent intentional message distortion, e.g., the dissemination of misinformation, and fake news [11, 33, 54]. Our experimental setting abstracts away from the complex social media landscape with heterogeneous agents including bots and trolls [16, 55], in order to understand the fundamental patterns that govern information distortion. On the middle ground between best intent and intentional distortion, media outlets distort information in more a more nuanced way [39]. In *agenda setting*, more attention is allocated to stories that fit a biased narrative, and in *framing*, the facts are more subtly distorted due to how they are presented.

Telephone and summarization effects. The telephone effect and the summarization effect have been studied from a linguistic standpoint. Breck and Cardie [8] note that facts, events, and opinions appearing in news articles are often known only second- or third-hand. Agents reporting them resort to using two kinds of expression that can filter information: perspective and speech expressions. They propose a learning approach that correctly determines the hierarchical structure of such information filtering expressions emerging due to the telephone effect. More recently, Gligorić et al. [18] found that length constraints cause Twitter users to summarize their content. This process significantly alters linguistic aspects of the text, such as the use of abbreviations, contracted forms, and articles.

Science communication. Media coverage of science has been investigated by communication research. The focus of the field is on scientists' attitude towards the media and on patterns of interactions with journalists and other key players such as news organizations and science information professionals [47, 56]. Evidence suggests that most scientists consider visibility in the media important and responding to journalists a professional attitude that is reinforced by universities and other science organizations. Exaggeration in the news is strongly associated with exaggeration in press releases, and improving the accuracy of academic press releases could represent a key opportunity for reducing misleading health-related news [53].

Summarization and paraphrasing. Finally, the summarization task at the heart of our experimental design is tightly related to the classic natural language processing tasks of paraphrasing [28, 49, 57] and summarization [37, 43]. Prior research introduced the distinction between abstractive and extractive summarization approaches upon which we rely in some of our analyses [27].

4.3 Limitations and future work

Our experimental setup allows us to finely measure and characterize message distortion effects, which would be difficult in observational setups. Still, it remains unclear to what extent our findings generalize to a wider spectrum of real-world information cascades, where three important distinctions are likely to have an impact: *(i)* the sequence-of-laypeople mechanism used here may differ from the way information diffuses in the real world; *(ii)* different kinds of information (e.g., news) may be subject to different dynamics; *(iii)* the assumed *bona fide* scenario may not hold in all cases.

Future work will face the challenge of tracking message distortion in real-world cascades (such as the one described in Sec. 1). By taking advantage of methods and insights developed in this paper, we plan to establish the extent to which our findings reflect the complex universe of social media, thus removing the *bona fide* assumption. Moreover, we will explore whether the observed effects can be modeled mathematically to predict which lexical and semantic units of a text that are likely to be distorted.

ACKNOWLEDGEMENTS

We gratefully acknowledge support from CNPq, Atmosphere, a Google Research Award for Latin America (Manoel Horta Ribeiro), and a Google Faculty Research Award (Robert West).

REFERENCES

[1] 2018. Chinese whispers. https://en.wikipedia.org/w/index.php?title=Chinese_whispers&oldid=865329255 Page Version ID: 865329255.

[2] Levy Adam S and Kim Stump-Sutliff. [n. d.]. Breast Cancer Quiz - Health Encyclopedia - University of Rochester Medical Center. https://web.archive.org/

web/20181116150802/https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=40&contentid=BreastCancerCancerBrQuiz

[3] Lada A. Adamic, Thomas M. Lento, Eytan Adar, and Pauline C. Ng. 2016. Information Evolution in Social Networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM '16)*. ACM, New York, NY, USA, 473–482.

[4] Eugene W. Anderson. 1998. Customer Satisfaction and Word of Mouth. *Journal of Service Research* 1, 1 (Aug. 1998), 5–17.

[5] Lentnek Arnold and Rita Sather. [n. d.]. Immunization Quiz - Health Encyclopedia - University of Rochester Medical Center. https://web.archive.org/web/20190222050418/https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=40&contentid=ImmunQuiz

[6] Hannah E. Bergman, Bryce B. Reeve, Richard P. Moser, Sarah Scholl, and William M. P. Klein. 2011. Development of a Comprehensive Heart Disease Knowledge Questionnaire. *American journal of health education / American Alliance for Health, Physical Education, Recreation, and Dance* 42, 2 (March 2011), 74–87.

[7] Archie Bleyer and H. Gilbert Welch. 2012. Effect of three decades of screening mammography on breast-cancer incidence. *New England Journal of Medicine* 367, 21 (2012), 1998–2005.

[8] Eric Breck and Claire Cardie. 2004. Playing the Telephone Game: Determining the Hierarchical Structure of Perspective and Speech Expressions. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04)*. Association for Computational Linguistics, Stroudsburg, PA, USA.

[9] Jane E. Brody. 2015. The More We Learn on Nutrition, the More We Ignore. https://well.blogs.nytimes.com/2015/10/12/the-more-we-learn-on-nutrition-the-more-we-ignore/

[10] Mario Cataldi, Luigi Di Caro, and Claudio Schifanella. 2010. Emerging Topic Detection on Twitter Based on Temporal and Social Terms Evaluation. In *Proceedings of the Tenth International Workshop on Multimedia Data Mining (MDMKDD '10)*. ACM, New York, NY, USA, 4:1–4:10.

[11] Stefano Cresci, Fabrizio Lillo, Daniele Regoli, Serena Tardelli, and Maurizio Tesconi. 2018. $FAKE: Evidence of Spam and Bot Activity in Stock Microblogs on Twitter. In *Twelfth International AAAI Conference on Web and Social Media*.

[12] Edgar Dale and Jeanne S. Chall. 1948. A Formula for Predicting Readability: Instructions. *Educational Research Bulletin* 27, 2 (1948), 37–54.

[13] Christopher Dougherty. 2016. *Introduction to Econometrics* (fifth edition, new to this edition: ed.). Oxford University Press, Oxford, New York.

[14] Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication* 43, 4 (Dec. 1993), 51–58.

[15] Ramón Estruch, Emilio Ros, Jordi Salas-Salvadó, Maria-Isabel Covas, Dolores Corella, Fernando Arós, Enrique Gómez-Gracia, Valentina Ruiz-Gutiérrez, Miquel Fiol, and José Lapetra. 2013. Primary prevention of cardiovascular disease with a Mediterranean diet. *New England Journal of Medicine* 368, 14 (2013), 1279–1290.

[16] Claudia I. Flores-Saviaga, Brian C. Keegan, and Saiph Savage. 2018. Mobilizing the Trump Train: Understanding Collective Action in a Political Trolling Community. In *Twelfth International AAAI Conference on Web and Social Media*.

[17] Neal D. Freedman, Yikyung Park, Christian C. Abnet, Albert R. Hollenbeck, and Rashmi Sinha. 2012. Association of coffee drinking with total and cause-specific mortality. *New England Journal of Medicine* 366, 20 (2012), 1891–1904.

[18] Kristina Gligorić, Ashton Anderson, and Robert West. 2018. How Constraints Affect Content: The Case of Twitter's Switch from 140 to 280 Characters. In *Twelfth International AAAI Conference on Web and Social Media*.

[19] Paul E. Goss, James N. Ingle, José E. Alés-Martínez, Angela M. Cheung, Rowan T. Chlebowski, Jean Wactawski-Wende, Anne McTiernan, John Robbins, Karen C. Johnson, and Lisa W. Martin. 2011. Exemestane for breast-cancer prevention in postmenopausal women. *New England Journal of Medicine* 364, 25 (2011), 2381–2391.

[20] Michael E. Greenberg, Michael H. Lai, Gunter F. Hartel, Christine H. Wichems, Charmaine Gittleson, Jillian Bennet, Gail Dawson, Wilson Hu, Connie Leggio, and Diane Washington. 2009. Response to a monovalent 2009 influenza A (H1N1) vaccine. *New England Journal of Medicine* 361, 25 (2009), 2405–2413.

[21] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A. Zighed. 2013. Information Diffusion in Online Social Networks: A Survey. *SIGMOD Rec.* 42, 2 (July 2013), 17–28.

[22] Stuart Hall. 2009. Encoding/Decoding. In *Media and Cultural Studies: Keyworks*. John Wiley & Sons. Google-Books-ID: I8dPhB88Sx4C.

[23] Thorsten Hennig-Thurau, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler. 2004. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing* 18, 1 (Jan. 2004), 38–52.

[24] Paul M. Herr, Frank R. Kardes, and John Kim. 1991. Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research* 17, 4 (March 1991), 454–462.

[25] Barbara V. Howard, Linda Van Horn, Judith Hsia, JoAnn E. Manson, Marcia L. Stefanick, Sylvia Wassertheil-Smoller, Lewis H. Kuller, Andrea Z. LaCroix, Robert D. Langer, Norman L. Lasser, Cora E. Lewis, Marian C. Limacher, Karen L. Margolis, W. Jerry Mysiw, Judith K. Ockene, Linda M. Parker, Michael G. Perri, Lawrence Phillips, Ross L. Prentice, John Robbins, Jacques E. Rossouw, Gloria E. Sarto, Irwin J. Schatz, Linda G. Snetselaar, Victor J. Stevens, Lesley F. Tinker, Maurizio Trevisan, Mara Z. Vitolins, Garnet L. Anderson, Annlouise R. Assaf, Tamsen Bassford, Shirley A. A. Beresford, Henry R. Black, Robert L. Brunner, Robert G. Brzyski, Bette Caan, Rowan T. Chlebowski, Margery Gass, Iris Granek, Philip Greenland, Jennifer Hays, David Heber, Gerardo Heiss, Susan L. Hendrix, F. Allan Hubbell, Karen C. Johnson, and Jane Morley Kotchen. 2006. Low-Fat Dietary Pattern and Risk of Cardiovascular Disease: The Women's Health Initiative Randomized Controlled Dietary Modification Trial. *JAMA: Journal of the American Medical Association* 295, 6 (2006), 655–666.

[26] Lee Howell. 2018. Opinion | Only You Can Prevent Digital Wildfires. *The New York Times* (Oct. 2018). https://www.nytimes.com/2013/01/09/opinion/only-you-can-prevent-digital-wildfires.html

[27] Wan-Ting Hsu, Chieh-Kai Lin, Ming-Ying Lee, Kerui Min, Jing Tang, and Min Sun. 2018. A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, 132–141.

[28] Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional Neural Network Architectures for Matching Natural Language Sentences. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 2042–2050.

[29] W. Philip T. James, Ian D. Caterson, Walmir Coutinho, Nick Finer, Luc F. Van Gaal, Aldo P. Maggioni, Christian Torp-Pedersen, Arya M. Sharma, Gillian M. Shepherd, and Richard A. Rode. 2010. Effect of sibutramine on cardiovascular outcomes in overweight and obese subjects. *New England Journal of Medicine* 363, 10 (2010), 905–917.

[30] Mette Kalager, Marvin Zelen, Frøydis Langmark, and Hans-Olov Adami. 2010. Effect of screening mammography on breast-cancer mortality in Norway. *New England Journal of Medicine* 363, 13 (2010), 1203–1210.

[31] Nicola P. Klein, Joan Bartlett, Ali Rowhani-Rahbar, Bruce Fireman, and Roger Baxter. 2012. Waning protection after fifth dose of acellular pertussis vaccine in children. *New England Journal of Medicine* 367, 11 (2012), 1012–1019.

[32] N. Kliemann, J. Wardle, F. Johnson, and H. Croker. 2016. Reliability and validity of a revised version of the General Nutrition Knowledge Questionnaire. *European Journal of Clinical Nutrition* 70, 10 (Oct. 2016), 1174–1180.

[33] Adam Kucharski. 2016. Post-truth: Study epidemiology of fake news. *Nature* 540 (Dec. 2016), 525.

[34] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. 2009. Meme-tracking and the Dynamics of the News Cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*. ACM, New York, NY, USA, 497–506.

[35] Lauren Lissner, Lluis Serra Majem, Maria Daniel Vaz de Almeida, Christina Berg, Roger Hughes, Geoffrey Cannon, Inga Thorsdottir, John Kearney, Jan-Å Gustafsson, Joseph Rafter, Lbrahim Elmadfa, and Nick Kennedy. 2006. The Women's Health Initiative. What is on trial: nutrition and chronic disease? Or misinterpreted science, media havoc and the sound of silence from peers? *Public Health Nutrition* 9, 2 (April 2006), 269–272.

[36] Kreesten Meldgaard Madsen, Anders Hviid, Mogens Vestergaard, Diana Schendel, Jan Wohlfahrt, Poul Thorsen, Jørn Olsen, and Mads Melbye. 2002. A population-based study of measles, mumps, and rubella vaccination and autism. *New England Journal of Medicine* 347, 19 (2002), 1477–1482.

[37] Inderjeet Mani. 2009. Summarization evaluation: an overview. In *Proceedings of the NTCIR Workshop*, Vol. 2.

[38] Mesoudi Alex and Whiten Andrew. 2008. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 1509 (Nov. 2008), 3489–3501.

[39] Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen R. Corman, and Huan Liu. 2018. Identifying Framing Bias in Online News. *Trans. Soc. Comput.* 1, 2 (June 2018), 5:1–5:18.

[40] Mehdi Moussaïd, Henry Brighton, and Wolfgang Gaissmaier. 2015. The amplification of risk in experimental diffusion chains. *Proceedings of the National Academy of Sciences* (April 2015), 201421883.

[41] Dariush Mozaffarian, Saman Fahimi, Gitanjali M. Singh, Renata Micha, Shahab Khatibzadeh, Rebecca E. Engell, Stephen Lim, Goodarz Danaei, Majid Ezzati, and John Powles. 2014. Global sodium consumption and death from cardiovascular causes. *New England Journal of Medicine* 371, 7 (2014), 624–634.

[42] Dariush Mozaffarian, Tao Hao, Eric B. Rimm, Walter C. Willett, and Frank B. Hu. 2011. Changes in diet and lifestyle and long-term weight gain in women and men. *New England Journal of Medicine* 364, 25 (2011), 2392–2404.

[43] Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Caglar Gulcehre, and Bing Xiang. 2016. Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. Association for Computational Linguistics, Berlin, Germany, 280–290.

[44] Steven E. Nissen and Kathy Wolski. 2007. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England*

Journal of Medicine 356, 24 (2007), 2457–2471.

[45] Benjamin Nye, Junyi Jessy Li, Roma Patel, Yinfei Yang, Iain Marshall, Ani Nenkova, and Byron Wallace. 2018. A Corpus with Multi-Level Annotations of Patients, Interventions and Outcomes to Support Language Processing for Medical Literature. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Melbourne, Australia, 197–207.

[46] Olivia Pagani, Meredith M. Regan, Barbara A. Walley, Gini F. Fleming, Marco Colleoni, István Láng, Henry L. Gomez, Carlo Tondini, Harold J. Burstein, and Edith A. Perez. 2014. Adjuvant exemestane with ovarian suppression in premenopausal breast cancer. *New England Journal of Medicine* 371, 2 (2014), 107–118.

[47] Hans Peter Peters. 2013. Gap between science and media revisited: Scientists as public communicators. *Proceedings of the National Academy of Sciences of the United States of America* 110, Suppl 3 (Aug. 2013), 14102–14109.

[48] Wayne A. Ray, Katherine T. Murray, Kathi Hall, Patrick G. Arbogast, and C. Michael Stein. 2012. Azithromycin and the risk of cardiovascular death. *New England Journal of Medicine* 366, 20 (2012), 1881–1890.

[49] Philip Resnik, Olivia Buzek, Chang Hu, Yakov Kronrod, Alex Quinn, and Benjamin B. Bederson. 2010. Improving Translation via Targeted Paraphrasing. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP '10)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 127–137.

[50] Tiago Rodrigues, Fabrício Benevenuto, Meeyoung Cha, Krishna Gummadi, and Virgílio Almeida. 2011. On Word-of-mouth Based Discovery of the Web. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference (IMC '11)*. ACM, New York, NY, USA, 381–396.

[51] S. Clinical Trials Partnership RTS. 2011. First results of phase 3 trial of RTS, S/AS01 malaria vaccine in African children. *New England Journal of Medicine* 365, 20 (2011), 1863–1875.

[52] Frank M. Sacks, George A. Bray, Vincent J. Carey, Steven R. Smith, Donna H. Ryan, Stephen D. Anton, Katherine McManus, Catherine M. Champagne, Louise M. Bishop, and Nancy Laranjo. 2009. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *New England Journal of Medicine* 360, 9 (2009), 859–873.

[53] Petroc Sumner, Solveiga Vivian-Griffiths, Jacky Boivin, Andy Williams, Christos A. Venetis, Aimée Davies, Jack Ogden, Leanne Whelan, Bethan Hughes, Bethan Dalton, Fred Boy, and Christopher D. Chambers. 2014. The association between exaggeration in health related science news and academic press releases: retrospective observational study. *BMJ* 349 (Dec. 2014), g7015.

[54] Sebastian Tschiatschek, Adish Singla, Manuel Gomez Rodriguez, Arpit Merchant, and Andreas Krause. 2018. Fake News Detection in Social Networks via Crowd Signals. In *Companion Proceedings of the The Web Conference 2018 (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 517–524.

[55] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online Human-Bot Interactions: Detection, Estimation, and Characterization. In *Eleventh International AAAI Conference on Web and Social Media*.

[56] Michael F. Weigold. 2001. Communicating Science: A Review of the Literature. *Science Communication* 23, 2 (Dec. 2001), 164–193.

[57] Wenpeng Yin and Hinrich Schütze. 2015. Convolutional Neural Network for Paraphrase Identification. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Denver, Colorado, 901–911.