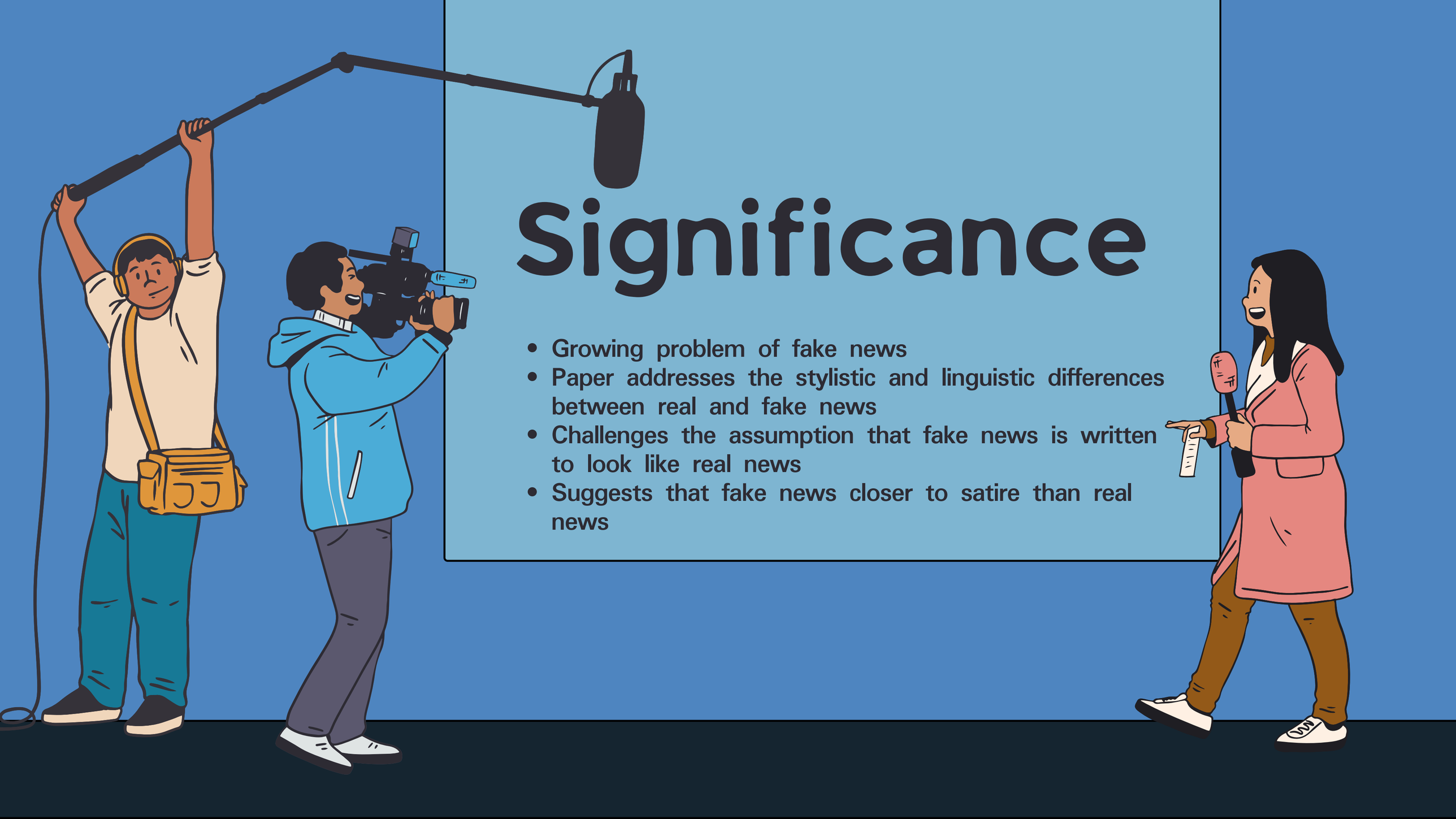

ON AIR
This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News
Paper: Benjamin D. Horne, Sibel Adalı
Prsented by Ella Mahnke



Significance

- Growing problem of fake news
- Paper addresses the stylistic and linguistic differences between real and fake news
- Challenges the assumption that fake news is written to look like real news
- Suggests that fake news closer to satire than real news

The Problem

Information overload:

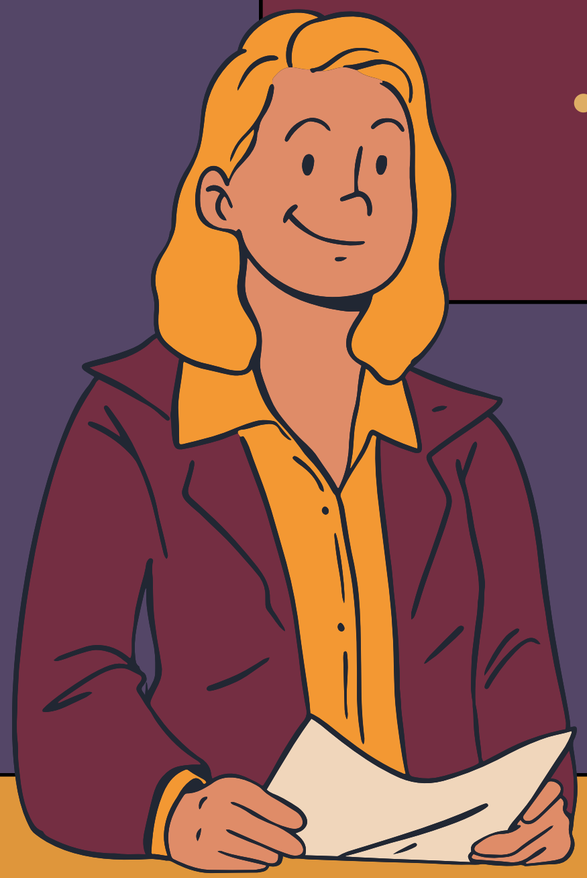
- Abundance of quick news can overwhelm observers
- Forces the use of quick heuristics to gain information
- Trust decisions often combine these heuristics
- This supports the notion of echo chambers

Prior Research

- Focused on the spread of misinformtion in social networks
- For example, echo chambers, homophily
- Relativley little work understanding content and differences in real news and fake news
- Studies that did research this, focused on classifiying fake vs real news

Core definitions

- Real News – Known to be true, from well-trusted news sources & adheres to a journalistic style
- Fake news – Known to be false, from websites intentionally trying to spread misinformation with the goal to deceive.
- Satire news – Explicitly states it is satirical, produced for entertainment, and relies on absurdity rather than sound arguments



Research Question

- Is there any systematic stylistic and other content differences between fake and real news?
- Aims to understand whether fake news differs systematically from real news in style and language use
- Seeks to understand the similarities between fake news and satire to uncover the different persuasive heuristics they employ

Data Set 1 - Buzzfeed Election Data Set

- Source: Buzzfeed's 2016 analysis of high-engagement Facebook election stories
- Used the analysis tool BuzzSumo
- Top-engagement real and fake stories over 9 months before the 2016 U.S. election
- Any opinion based stories or articles were filtered as satire
- 36 Real stories 35 fake stories
- Possible bias from collection procedure & Engagement ≠ actual traffic

Data Set 2 - Political News Data Set

- Created to strengthen the analysis
- Contains an equal number of stories for each category (75 each, 225 total)
- Randomly selected "hard" news articles related to US articles
- Fake → sites from Zimdars' list; each with at least one previously debunked story.
- Real → outlets from Business Insider's "Most Trusted" list.
- Satire → sites explicitly labeled as satirical.
- Avoids data collection bias

Real sources	Fake sources	Satire sources
Wall Street Journal	Ending the Fed	The Onion
The Economist	True Pundit	Huff Post Satire
BBC	abcnews.com.co	Borowitz Report
NPR	DC Gazette	The Beaverton
ABC	libertywritersnews	SatireWire
CBS	Before its News	Faking News
USA Today	Infowars	
The Guardian	Real News Right Now	
NBC		
Washington Post		

Table 1: Data set 2 sources

Data Set 3 - Burfoot and Baldwin

- Source: Dataset from Burfoot & Baldwin (2009) used for satire vs. real news classification
- 233 satire stories and 4,000 real news stories.
- Real → sampled from the English Gigaword Corpus (newswire documents)
- Satire → hand-selected stories topically matched to real articles
- Filtered to exclude "non-newsy" satire
- Does include non political content, some sources are not fully traceable

Features

Stylistic

- Capture syntax, text style, and grammatical structure
- POS tag counts using NLTK
- Counts of: Stopwords, punctuation, quotes, Negations (no, never, not), Informal/swear words, interrogatives (who/what/why/how), and ALL-CAPS words
- Dictionary-based features via LIWC (2015)

Complexity

- Measure sentence-level and word-level intricacy
- Words per sentence
- Syntax tree depth (overall, noun phrase, verb phrase) using Stanford Parser
- Readability scores: Gunning Fog, SMOG, Flesch–Kincaid
- Type–Token Ratio (TTR) → lexical diversity
- Fluency (common vs. specialized vocabulary) using COCA frequency data

Psychological

- Capture cognitive, emotional, and sentiment related elements.
- LIWC categories: cognitive processes, drives, personal concerns
- LIWC sentiment (bag-of-words).
- SentiStrength for sentiment intensity:
  Positive: +1 to +5
  Negative: –1 to –5

Analysis and Classification

Statistical Analysis

- One-way ANOVA:

Compares means across groups using variance ratios.

Applied to features that pass normality tests.

- Wilcoxon Rank Sum Test:

Non parametric test for non normal distributions.

Statistical Analysis

- Selected top 4 features from hypothesis testing for body and title text
- Applied Linear SVM with:

Linear kernel and 5-fold cross validation

- Simple model and small feature set to reduce overfitting
- Aimed to assess predictive power of linguistic features.

Findings

- Systemic differences in style and language between real and fake news

Body differences:

- Real news articles are significantly longer than fake ones. Fake news uses fewer technical words, punctuation/quotes, and analytic words
- Fake news articles show more lexical redundancy (low Type–Token Ratio/TTR). They also require a slightly lower education level to read.
- Fake news articles use significantly more personal pronouns

Title Differences:

- Fake news titles are longer than real news titles
- Fake titles use simpler words. Use significantly more proper nouns, more all capitalized words, but fewer stop–words and fewer nouns overall.
- Fake news packs the main claim and substance of the article into its title, often involving specific people/entities and actions, allowing readers to skip the body text.

Complexity and style of fake news content is more closely related to satire than to real news

- Both satire and fake news use smaller, fewer technical/analytic words, fewer quotes/punctuation, and significantly more lexical redundancy than real articles.

- Challenges the notion that fake news is written to look like real news to fool the reader
- This paper provides practical applications for debunking fake news
- However, the data sets are small
- Since this paper did a content analysis, I owuld like to see user studies in the future

Commentary

Question 1

What specific stylistic strategy did the authors find fake news uses in its titles to facilitate quick persuasion?

A. Titles are significantly shorter, using fewer adverbs and exclamation marks.

B. Titles use more stop-words and function words, similar to traditional journalistic style.

C. Titles are longer, pack the main claims, and use significantly more proper nouns and words in all capital letters.

D. Titles rely heavily on complex syntax structures and low-fluency technical terms.

Question 2

What did the authors find regarding the readability of fake news articles compared to real news articles in the body text?

A. Fake news articles require a significantly higher education level to read, indicating higher complexity.

B. Fake articles need a slightly lower education level to read, aligning with the goal of peripheral persuasion.

C. The readability scores were identical across all three news categories (Real, Fake, Satire).

D. Real news articles consistently used simpler words, resulting in lower (easier) readability scores

Question 3

Which of these was NOT a feature category observed in the fake news paper?

A. Stylistic

B. Psychological

C. Veracity

D. Complexity

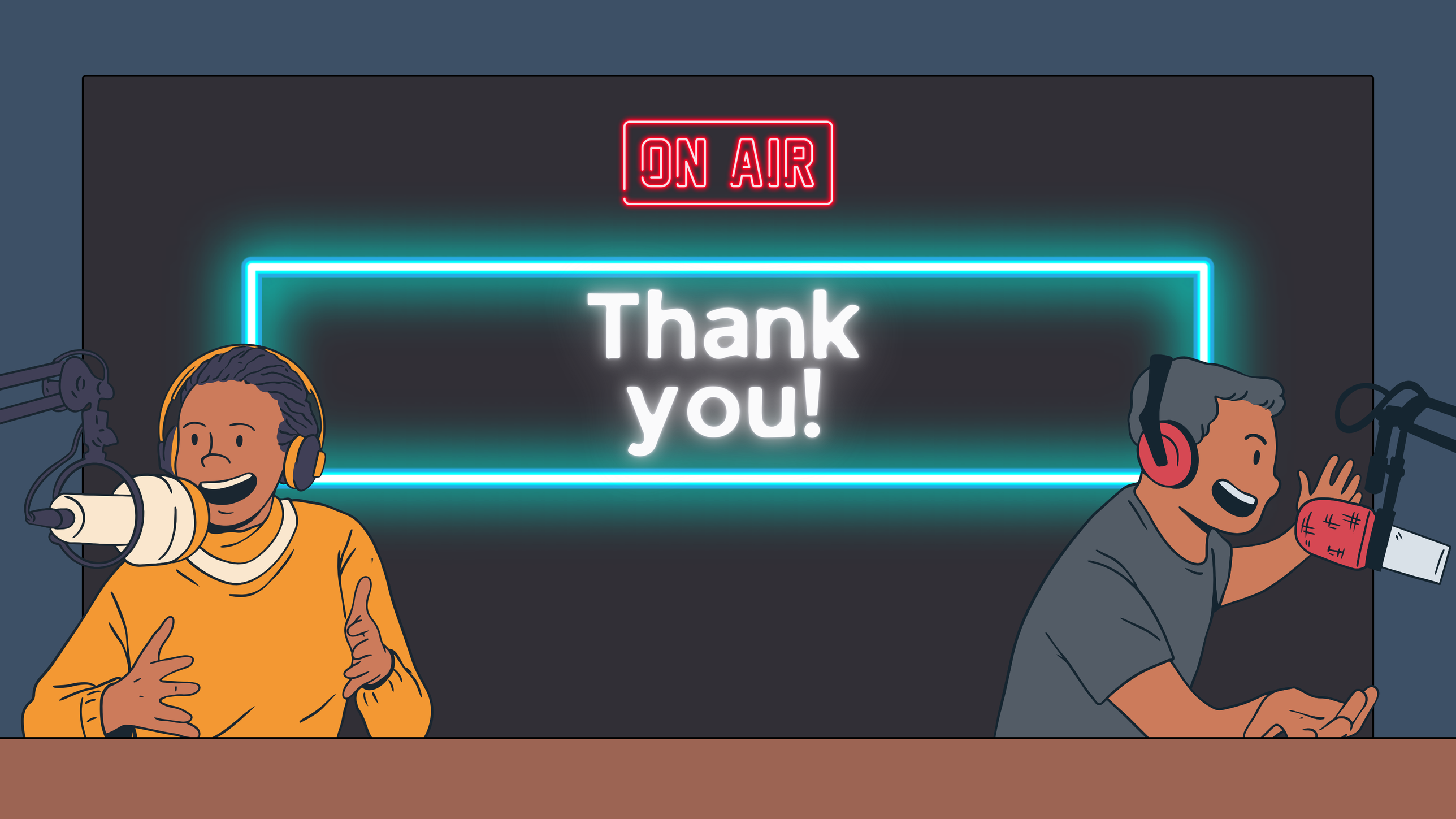
Question 4
How does fake news differ from satire news?

A. Fake news is intended for entertainment, satire is meant to inform

B. Fake news tries to decieve people, satire is made clear to be false

C. They are synonomous

B. Fake news is imaginary, satire is not


ON AIR
Thank you!