



Misleading Information + Twitter

A designers toolkit to managing misleading information



A little bit about the Twitterverse

Twitter is where people come to see and tweet what's happening... right now



But it's very hard to know who and what to trust...

[Redacted] · Sep 4

If we not supposed to drink bleach why they make it in these to go bottles 🤔



[Redacted] · Sep 19

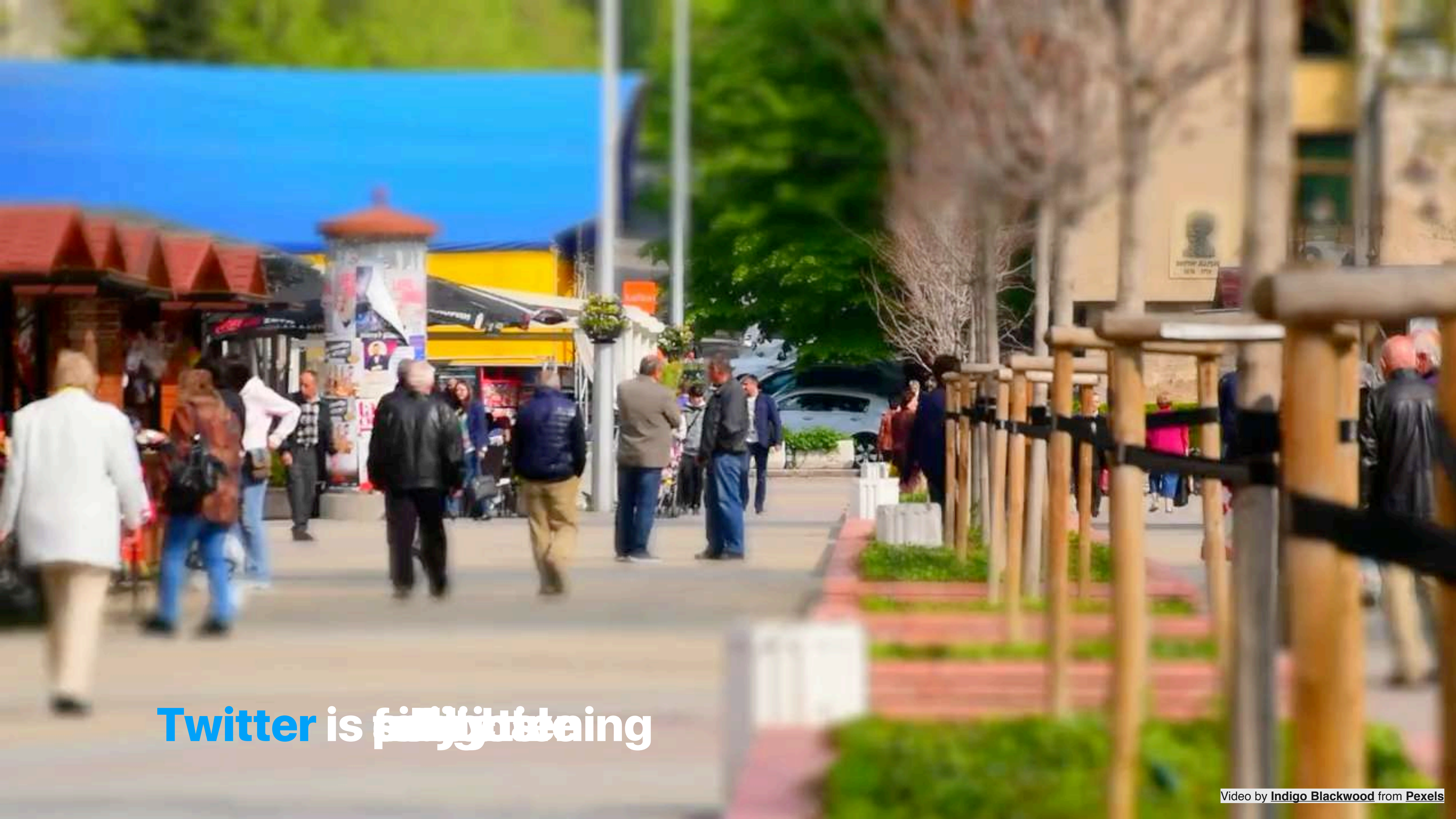
I'm not going to forget that the British media completely ignored the fact that prince william cheated on kate and instead nitpicked at the fact that **megan markle** breathed

21 11.1K 31K [Share]



[Share] 1 1 [Share]

More severe misleading info



Twitter is family ~~friendly~~ **entertaining**

Twitter's mission

Keeping the world's conversations safe and reliable information

What people think

"If it is determined that a person is a
bait and switch, then it's
time to take it out of
circulation so it doesn't hurt anyone."

— *Facebook* 4,438

What people think and expect from Twitter



People define misleading information as untrue content that can be harmful



They believe Twitter has a responsibility to help stop the spread of harmful misleading info



And that different types of misleading information require different types of action or intervention

What people think and expect from Twitter



To identify misleading information on the platform



To indicate when information is disputed or unverified



To set context so people can make up their own minds

Our role as designers

Our role is to create a tool kit for managing misleading information



Design interventions that help reduce the spread of misleading information



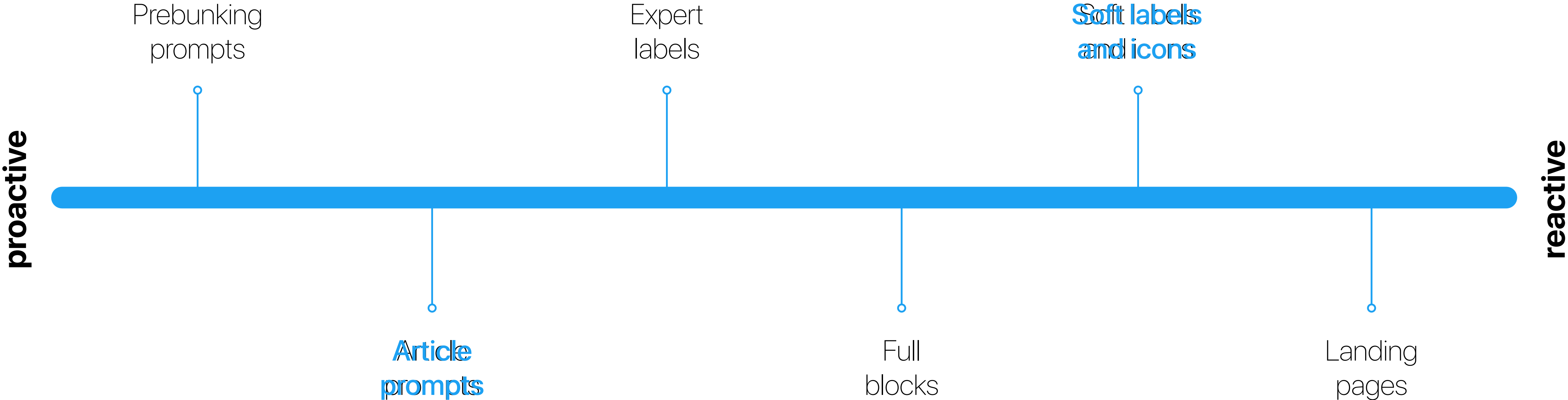
Design annotations (aka labels) that signal content that is credible *as well as* misleading

Our tool kit

Prompts and labels

What is in our design toolkit?

We want to *elevate* reliable content and *de-amplify* misleading information



Prompts

Prompts help us elevate credible information and sources. They provide context if we think something might be misleading

TOOLS

Prebunking

[Article prompts](#)

What people do

"I stumble upon a news headline about the number of coronavirus cases in China and I immediately share it."

— *Male, 29*



Prompts

Article prompts

Many people will share articles and only read the headline.

OUR HYPOTHESIS

Customers are less likely to retweet unreliable sources if they read the article and notice facts or evidence missing.

DESIGNERS

Andrew Courter
@covrter

Ashleigh Kaneski
@ashleighkaneski

CS + RESEARCH

Lisa Young
@RealOaktownGirl

Cody Elam
@codyelam



Our design principles

- Encourage good behavior
- Be explicit about consequences and rewards
- Provide clarity
- Be primary
- Balance flow and friction

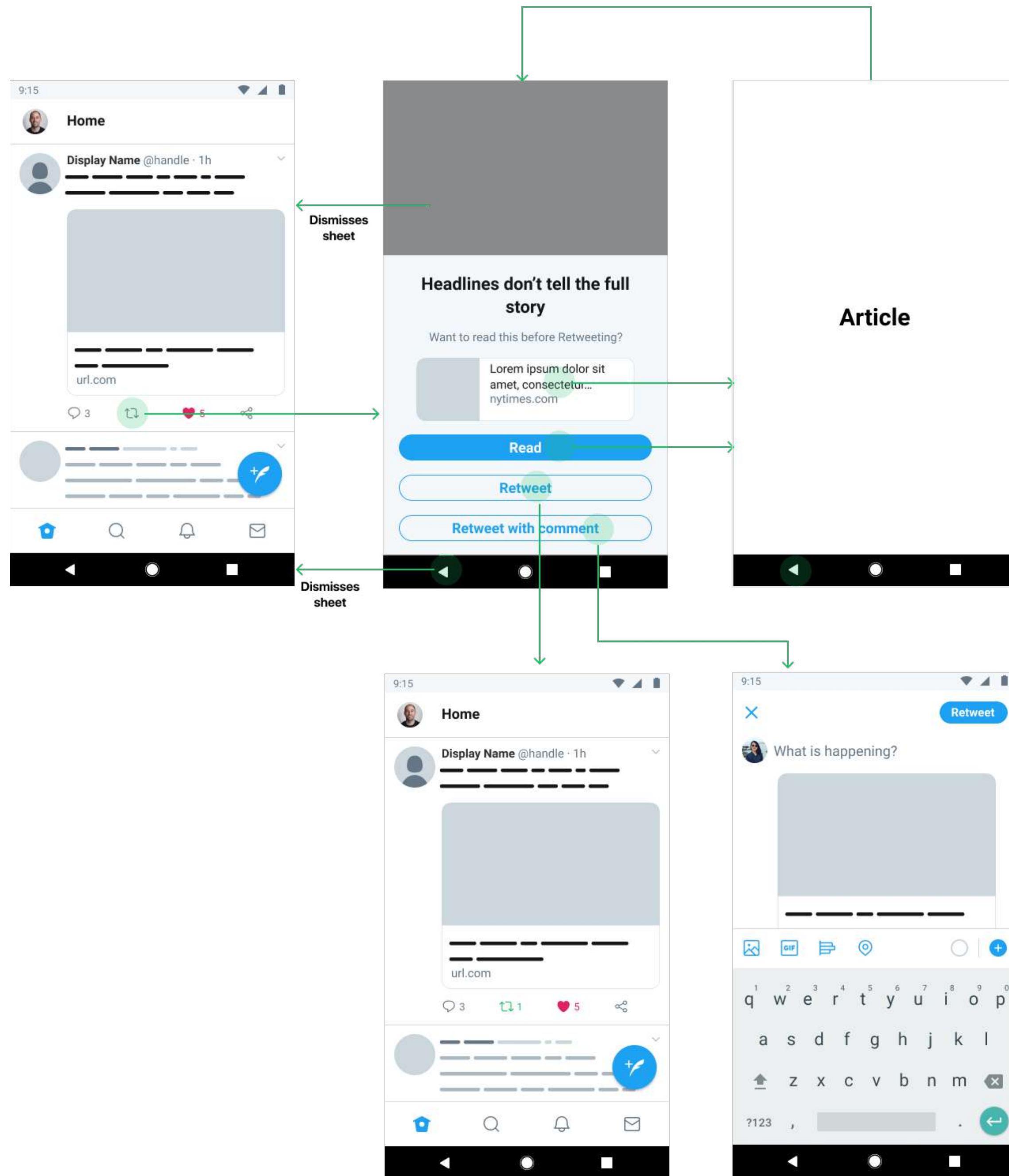




Prompts

The idea

When Twitter detects a URL in someone's tweet, the reader is prompted to consider reading the article before re-tweeting it.

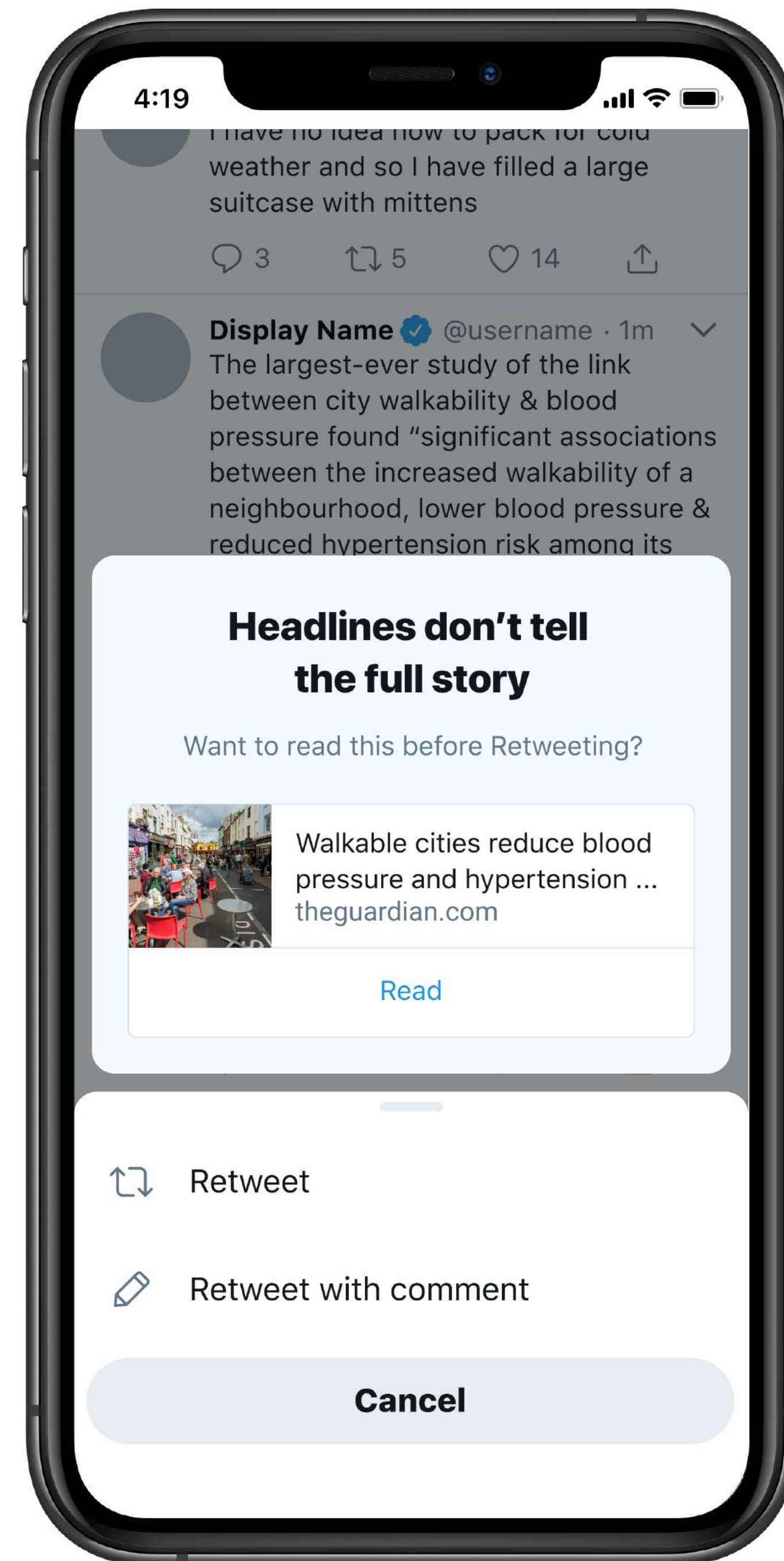




Prompts

Iterations

- **UI**
Distinct or merged
- **Primary CTA**
Read the article
- **Color + copy matter**
Test options

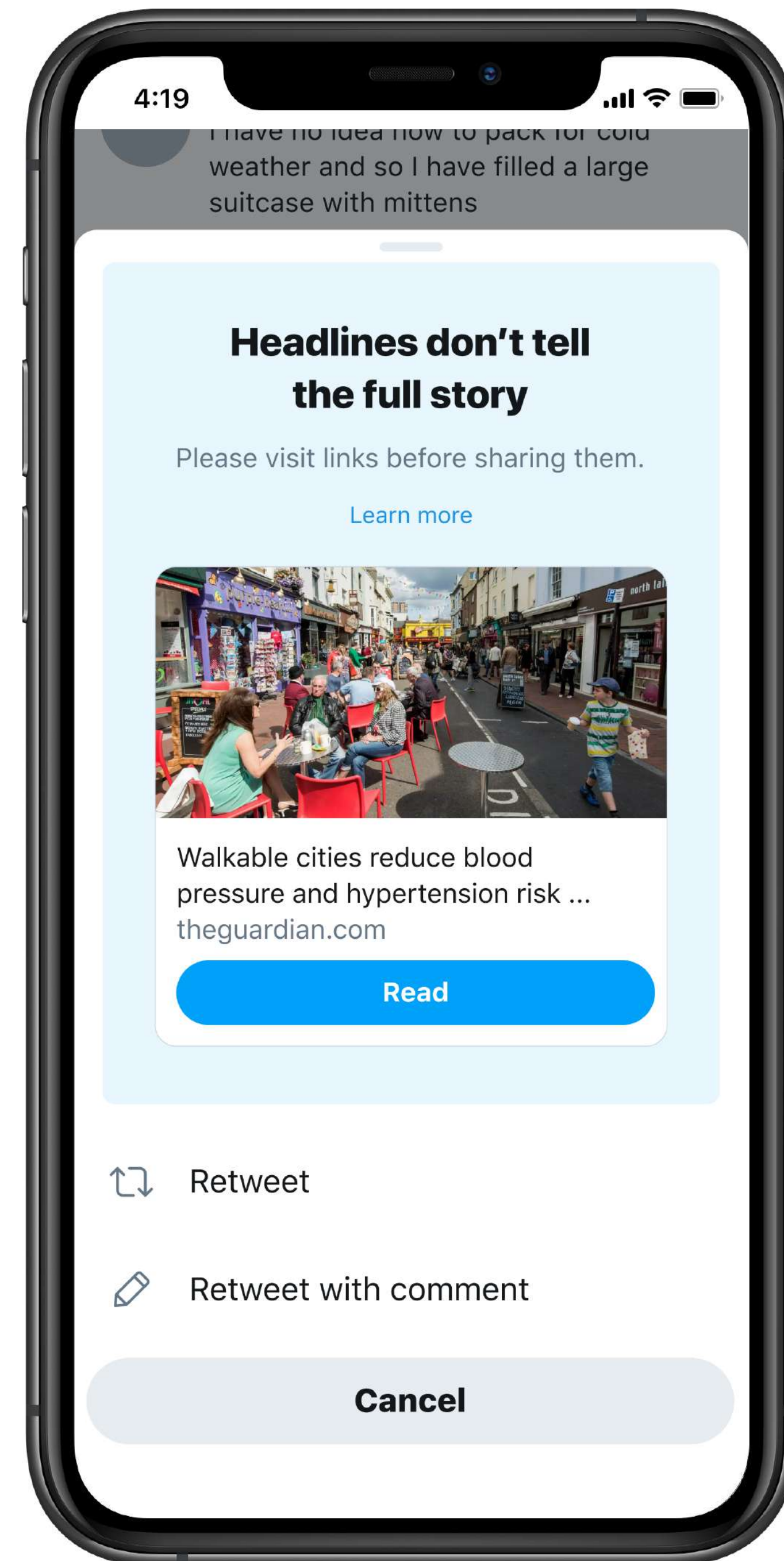




Prompts

Iterations

- **UI**
Distinct or merged
- **Primary CTA**
Read the article
- **Color + copy matter** Test options

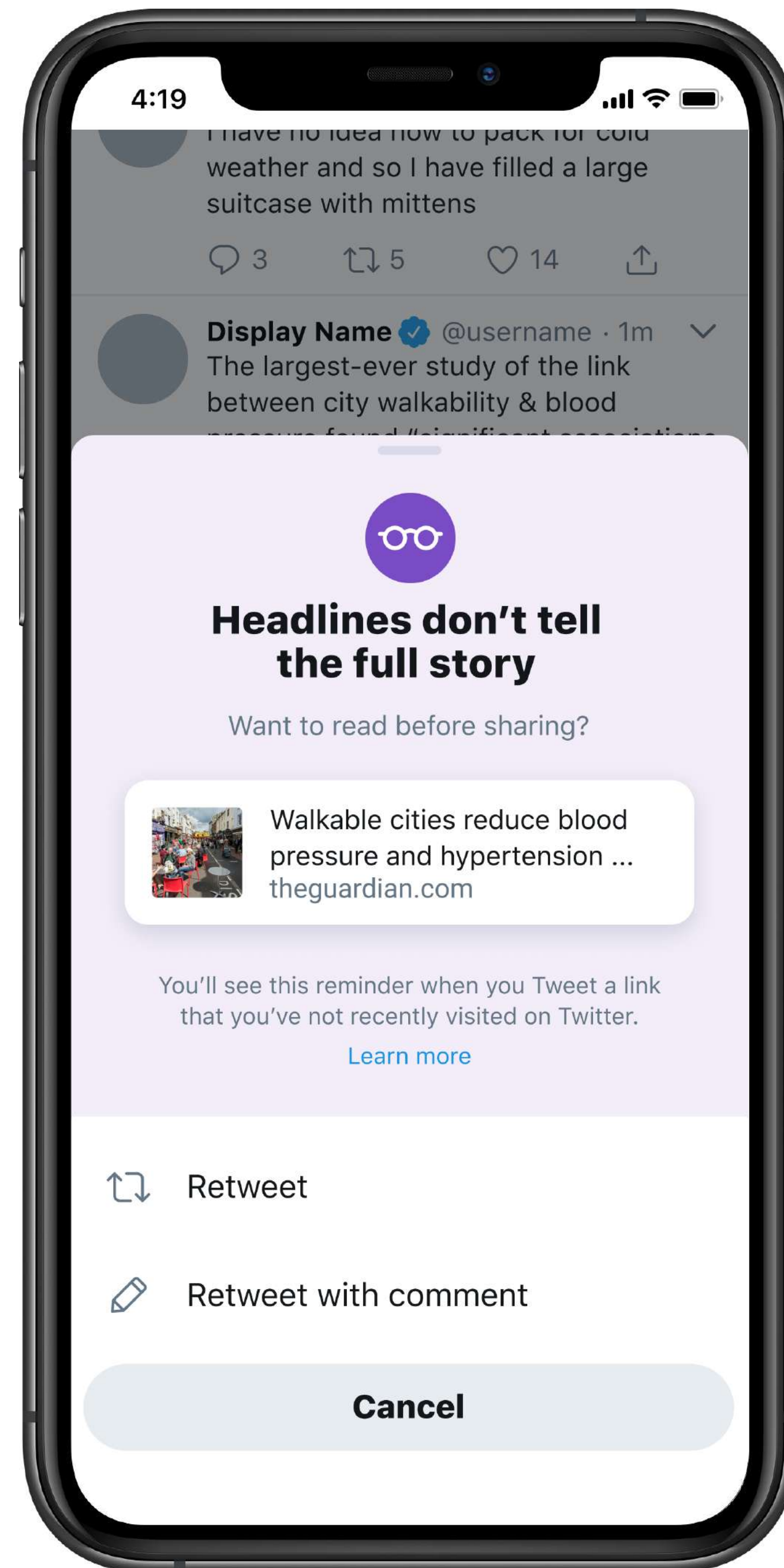




Prompts

Iterations

- **UI**
Distinct or merged
- **Primary CTA**
Read the article
- **Color + copy matter**
Test options





Prompts

The one that launched

on Android

Coming soon to iOS and Web

DESIGNERS

Andrew Courter
@covrter

Ashleigh Kaneski
@ashleighkaneski

CS + RESEARCH

Lisa Young
@RealOaktownGirl

Cody Elam
@codyelam



Article prompt results



People who read the article first



People who still retweeted



People who quote tweeted



People who didn't share

What people are saying...❤️❤️❤️

The image shows a vertical scroll of three tweets. The top tweet is from a user with a grey profile picture, dated August 6, with a downward arrow in the top right corner. The text of the tweet is: "This new twitter feature that encourages you to read the article before you RT it is dope." Below the text is a blacked-out image. The middle tweet is from a user with a grey profile picture, dated July 11, with a downward arrow in the top right corner. The text of the tweet is: "Not a bad start. How about if you add an indicator so that others can see if the retweeted article has been read by the person retweeting?" Below the text are icons for reply, retweet, like, and share. The bottom tweet is partially visible, showing a video thumbnail of two people and engagement metrics: 12 replies, 158 retweets, and 1K likes.

[Profile Picture] · Aug 6

This new twitter feature that encourages you to read the article before you RT it is dope.

[Profile Picture] · Jul 11

Not a bad start. How about if you add an indicator so that others can see if the retweeted article has been read by the person retweeting?

12 158 1K

Labels

Labels are a tool that help us elevate credible information as well as identify misleading content

TOOLS

Soft annotations

Hard annotations

What people are feeling

"I'm angry....I wish there was a way we could know what is real and what is not."

— *Female, 43*



Labels

Soft annotations

We label a tweet when its content may be misleading and provide more information from verified sources.

OUR HYPOTHESIS

People will be better informed if they know which tweets are misleading and why.

DESIGNERS

Brittany Forks
@brittanyforks

Ashleigh Kaneski
@ashleighkaneski

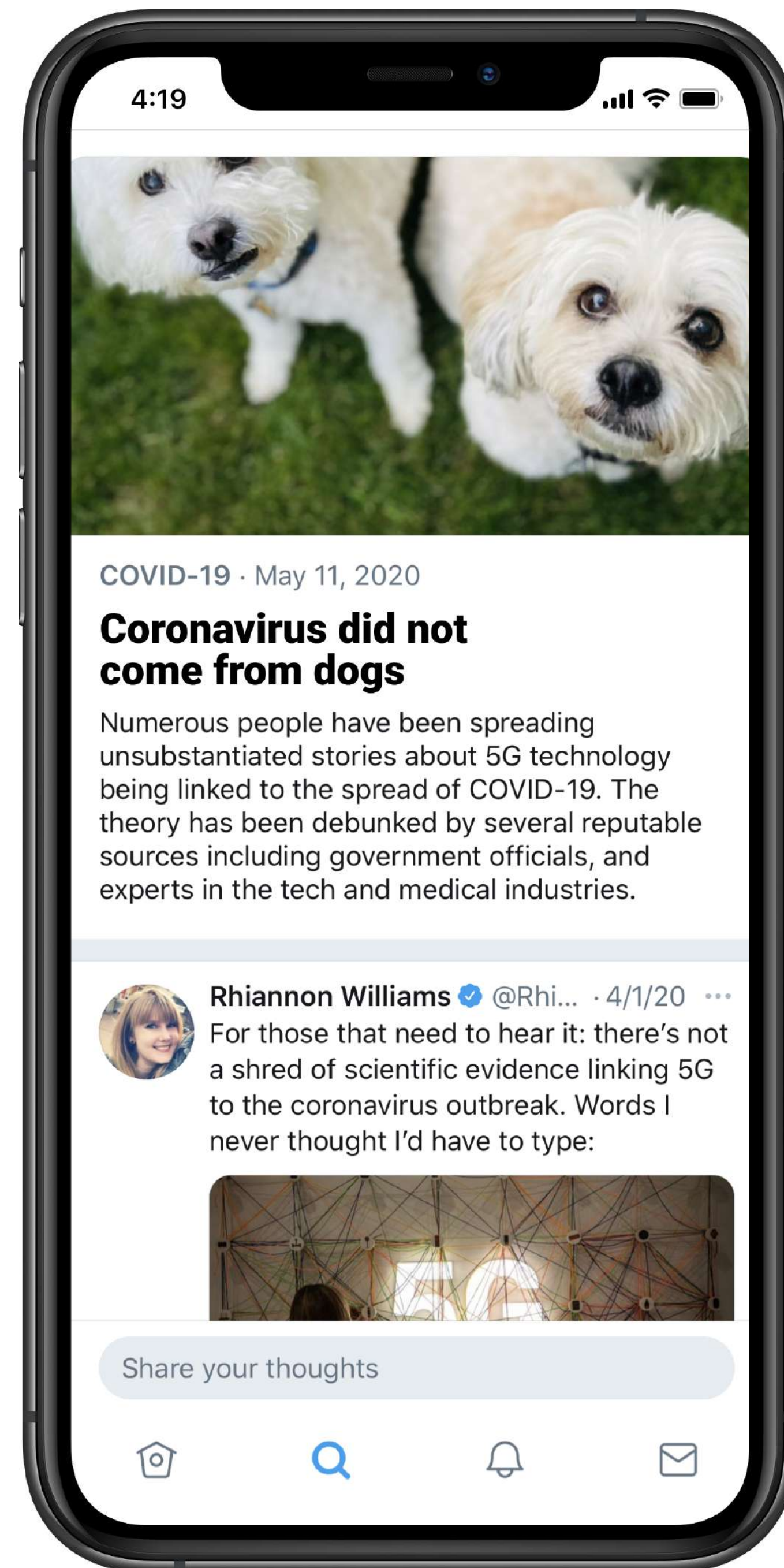
Adebola Falade
@bolafalade

CS + RESEARCH

Lisa Young
@RealOaktownGirl

Sarah Vieweg
@velofemme

Ruben Gomez
@rubenzo1



Our design principles

- Keep it simple
- Provide context and be transparent
- Balance prominence with amplification





Labels

Our existing system

We have a system to label tweets for a variety of reasons.

We wanted a softer label option for misleading information.

This Tweet is unavailable

This Tweet is no longer available.
[Learn more](#)

This Tweet from @username has been withheld in <country> in response to legal demand.
[Learn more](#)

This Tweet may include sensitive content
In order to view it, you'll need to change your sensitive media selection. You can find this under the Safety section of your Privacy and safety account settings.
[Change setting](#)

You reported this Tweet **View**

This Tweet is not available because it's from someone you blocked. **View**

This Tweet violated the Twitter Rules about {_specific rule_}. Twitter doesn't condone this Tweet. However, because of its public interest, you can still view it if you want. [Learn more](#) **View**



Labels

Iterations

- Spectrum

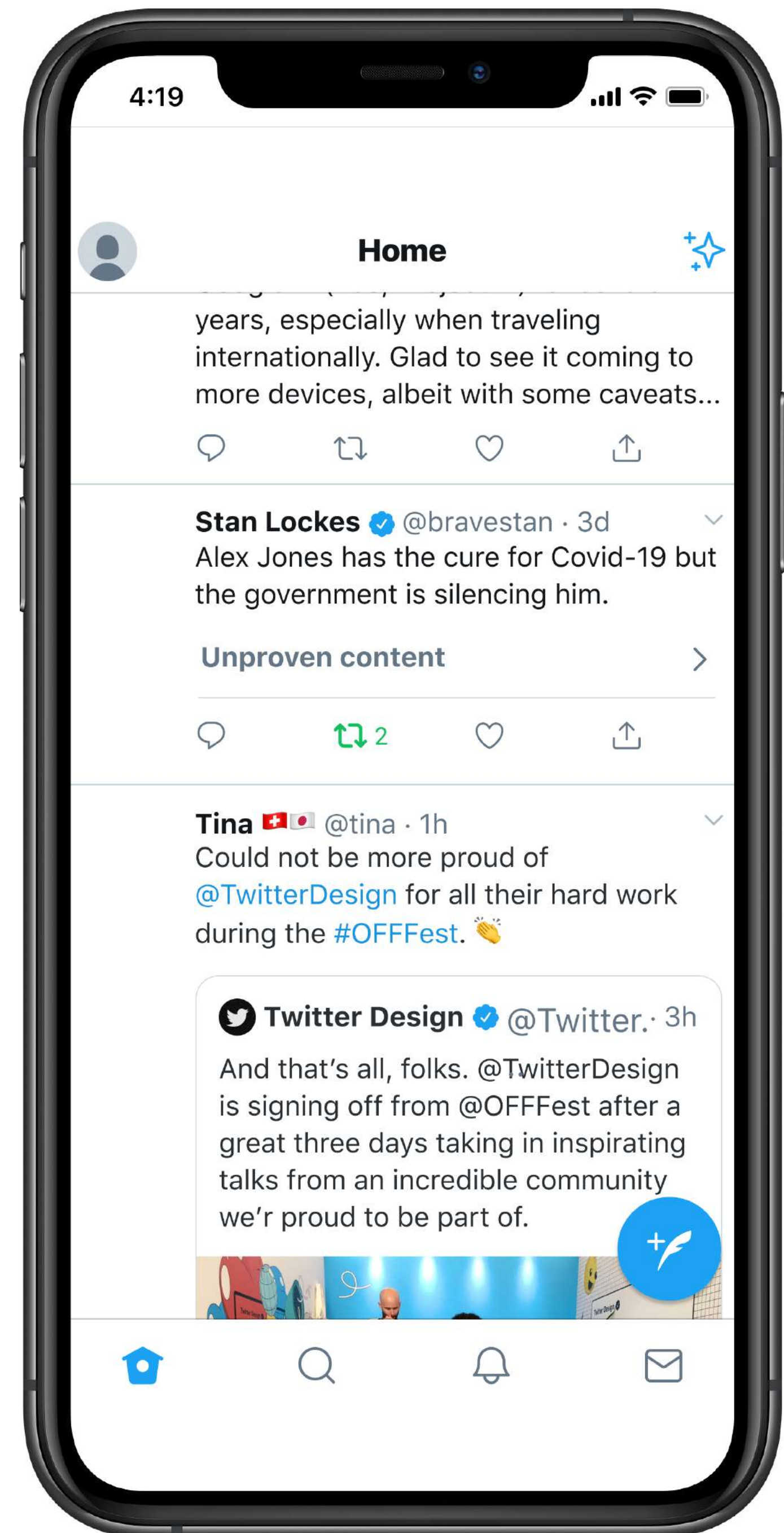
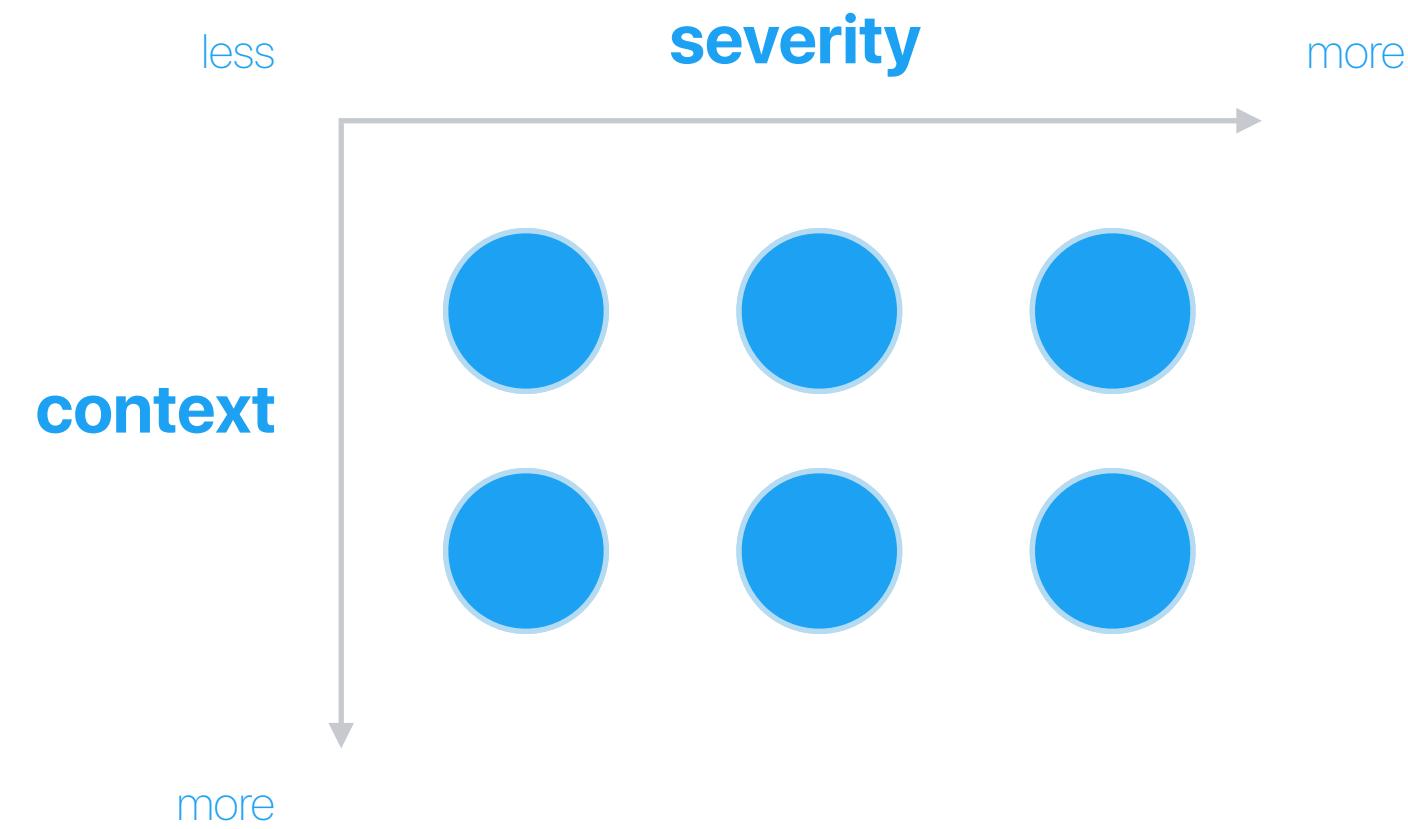
Less to more

- Clarity

Position and content variations

- Visual system

Start with color + icon





Labels

Iterations

- Spectrum

Less severe to more severe

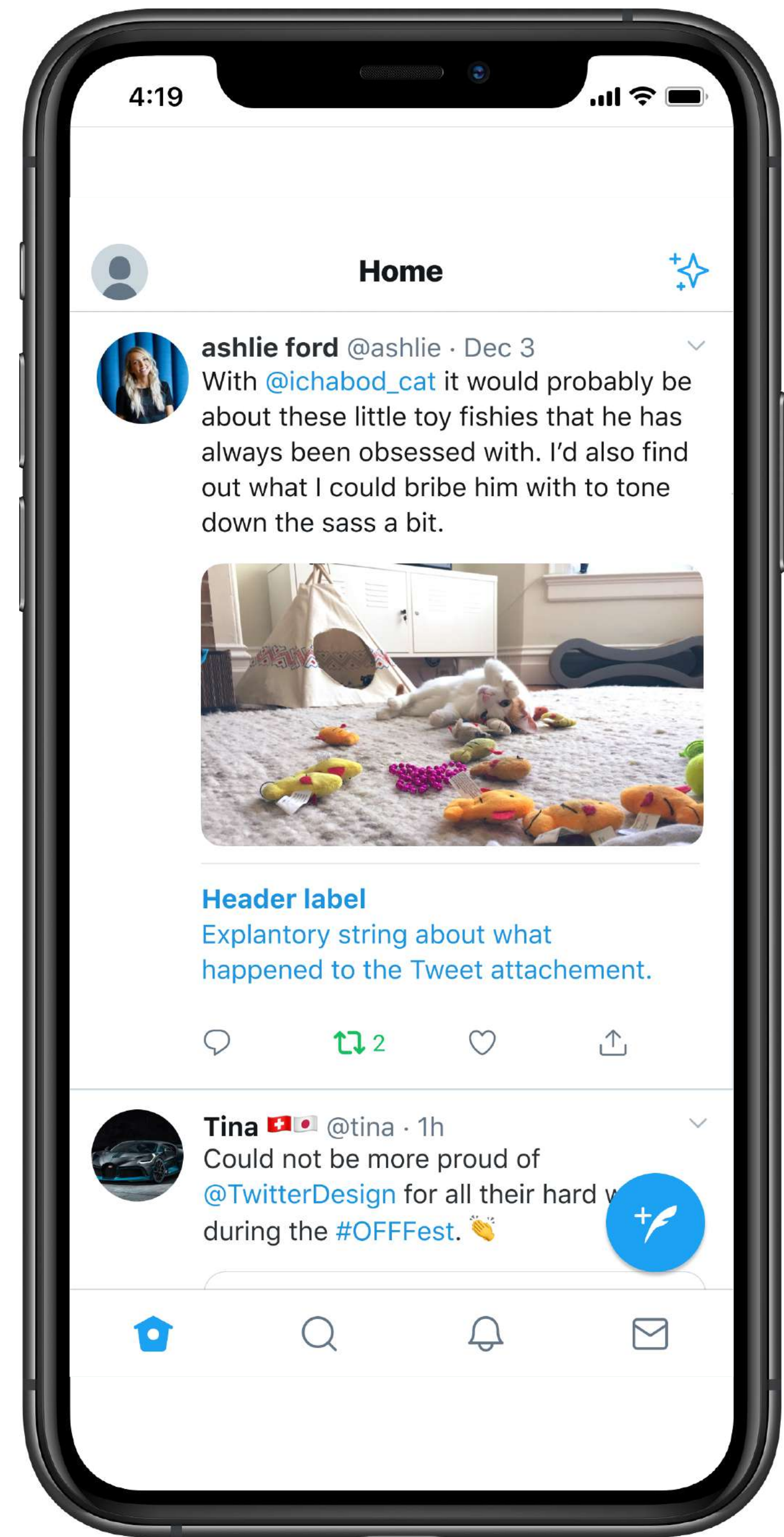
- Clarity

Position and content variations

- Visual system

Start with color + icon

Body message only	Explanatory string about what happened to the Tweet attachment. It can wrap to multiple lines of text.
Body message Action link	Explanatory string about what happened to the Tweet attachment. Action label
Body message More pivot	Explanatory string about what happened to the Tweet attachment. >
Header label Body message	Header label Explanatory string about what happened to the Tweet attachment.
Header label Body message Action link	Header label Explanatory string about Tweet. Action link
Header label Body message More pivot	Header label > Explanatory string about what happened to the Tweet.
Header label Body message Action link	i Explanatory string about what happened to the Tweet attachment.
Icon Header label Body message Action link	i Header label Explanatory string about what happened to the Tweet. Action link
Header label Body message More pivot	i Explanatory string about what happened to the Tweet attachment. >
Icon Header label Body message	i Header label Explanatory string about what happened to the Tweet.
Icon Header label Body message Action link	i Header label Explanatory string about Tweet. Action link
Icon Header label Body message More pivot	i Header label > Explanatory string about what happened to the Tweet.





Labels

Iterations

- **Spectrum**
Less severe to more severe
- **Clarity**
Position and content variations
- **Visual system**
Start with color + icon





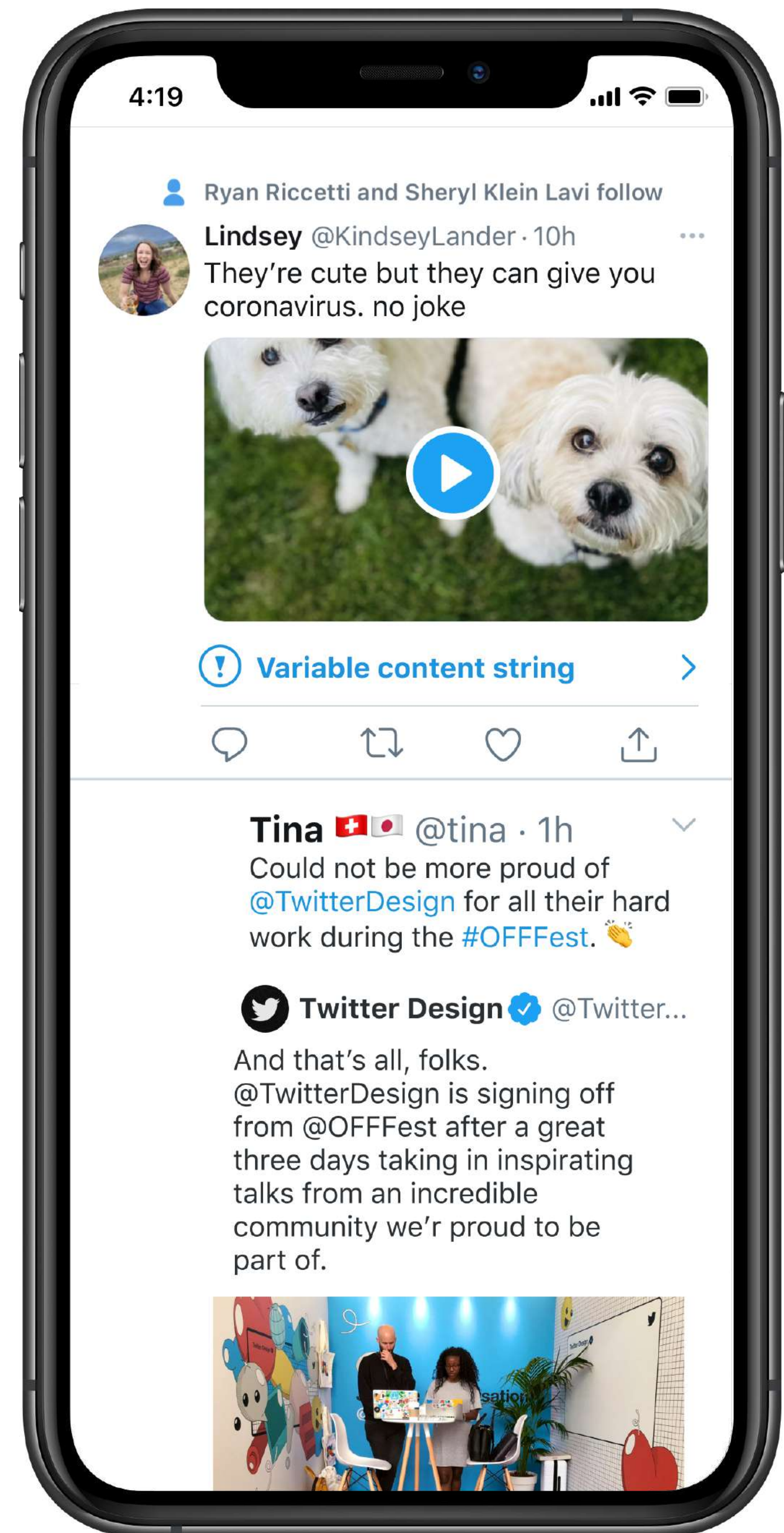
Labels

The color test

We tested the blue label and people couldn't see it. We have been running a color test with a stronger red.

VARIATIONS

- ⓘ Blue icon with blue text
- ⓘ Red icon with blue text
- ⓘ Red icon with red text



Soft annotation results 🙄



Red icon with red text



Red icon with blue text






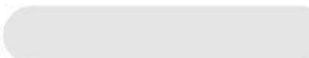
People who still retweeted







People who replied

What people are saying...❤️❤️❤️

  · Aug 31 

While  is worse when it comes to actually adding labels to manipulated and false stories, at least its labels are big and can't be easily ignored. Twitter just adds a teeeeeeeny tiiiiiiiny note at the bottom with zero additional explanation.

 4  33  428 

A low-angle, close-up shot of a person's feet on a skateboard. The person is wearing dark pants, white socks, and dark sneakers with white soles. The skateboard has a light-colored deck and bright green wheels. The skateboard is on a paved road, and the background is blurred, suggesting motion. The text "It's a journey" is overlaid on the left side of the image.

It's a journey

There is no silver bullet

Thank you 🙏