OLD DOMINION UNIVERSITY

How Archivists Could Stop Deep Fakes from Rewriting History

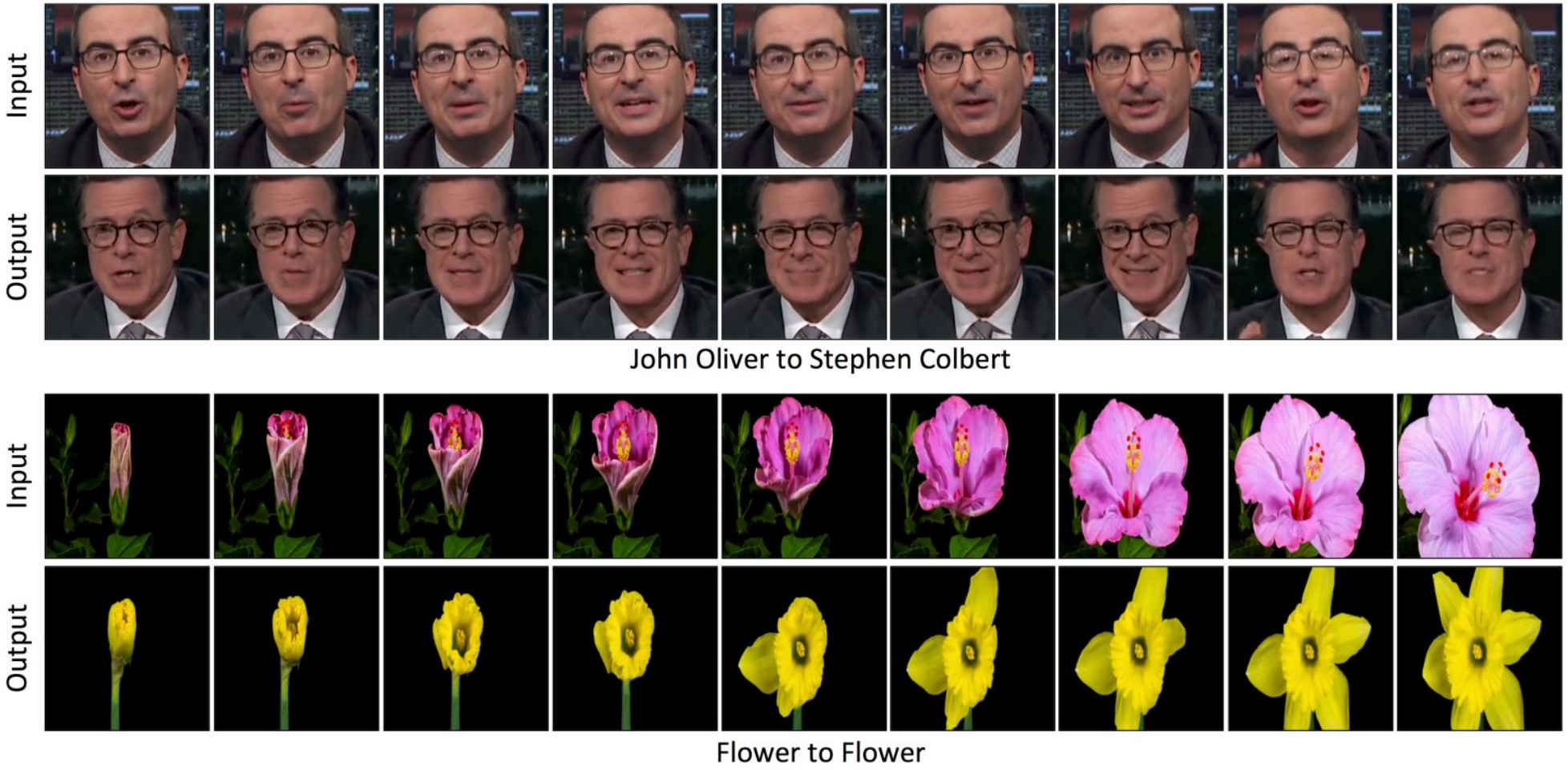
By Melanie Ehrenkranz

https://gizmodo.com/how-archivists-could-stop-deepfakes-from-rewriting-hist-1829666009

Presentation: Justin Whitlock

14 April 2019

Deep Fakes are sometimes easy to spot



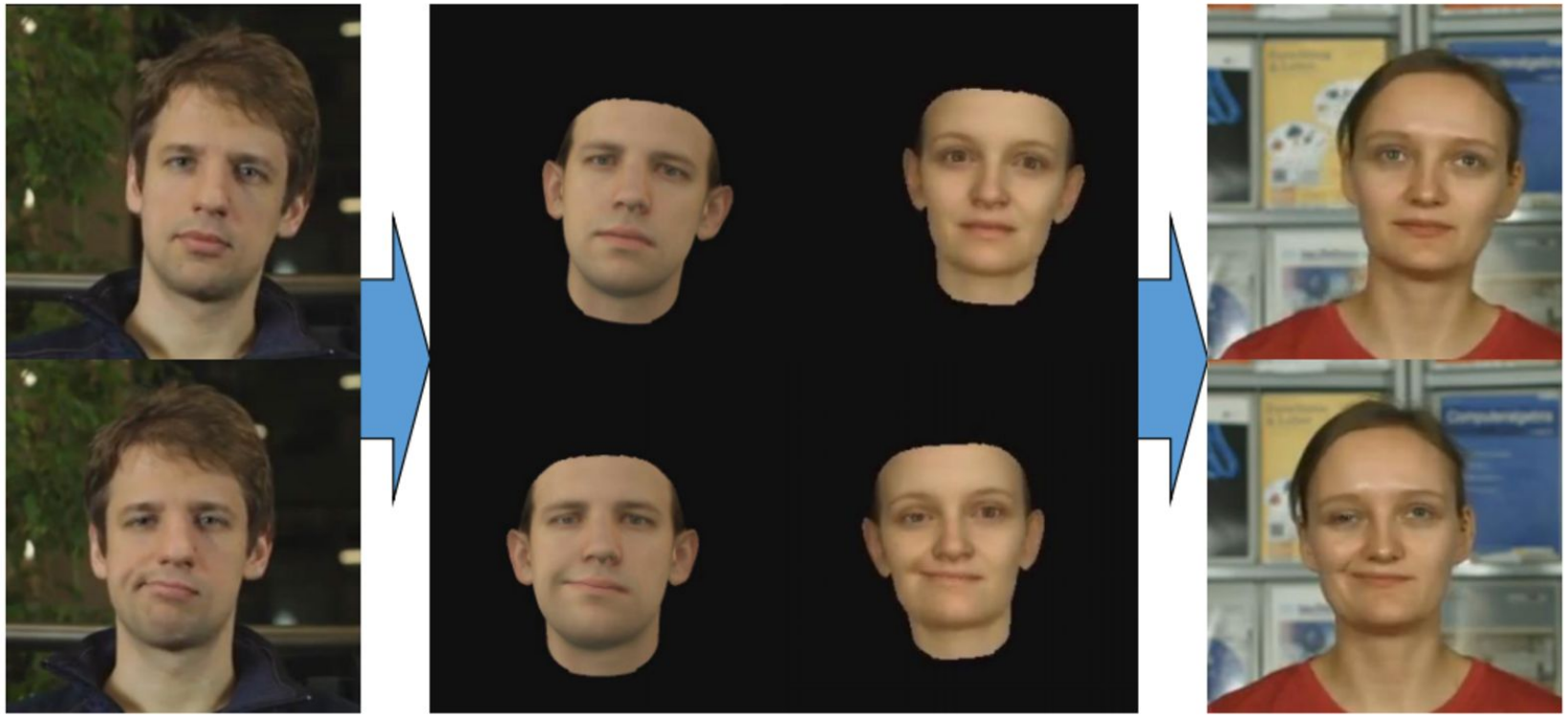
https://youtu.be/ehD3C60i6lw

Deep Fakes are becoming more convincing

This research maps facial expressions and mouth movements to video of Barack Obama, using recorded audio to create fake videos that look almost legitimate.

https://youtu.be/9Yq67CjDqvw

An improvement which includes adjusting the head motion to match dialogue.



https://youtu.be/qc5P2bvfl44

Concern over Deep Fakes is two-fold

- Prolific generation of misinformation in an attempt to influence current events
  - The spread of fake news on social media.
  - Election meddling by foreign powers.
- Insertion of false documents into the historical record as factual accounts.
  - How do archivists ensure that Deep Fakes are labeled as such?
  - How to prevent bad actors from tampering with archival records by inserting Deep Fakes into the record or changing the provenance of known Deep Fakes?

History is full of fakes; The Hitler Diaries, 1983

- "An archive of great historical significance" - Hugh Trevor, UK historian, retracted almost immediately.
- Forgeries, penned by Konrad Kujau, a man with a history of falsifying Nazi Memorabilia
- 60 volumes forged between 1981 and 1983.

https://en.wikipedia.org/wiki/Hitler_Diaries

What if the Hitler Diaries entered the historical archive as accepted fact? Why?

Historical Negationism:

- The purposeful distortion of the historical record to support a specific ideology or political cause.

  "...presenting known forged documents as genuine,

  inventing ingenious but implausible reasons for distrusting genuine documents,

  attributing conclusions to books and sources that report the opposite,

  manipulating statistical series to support the given point of view,

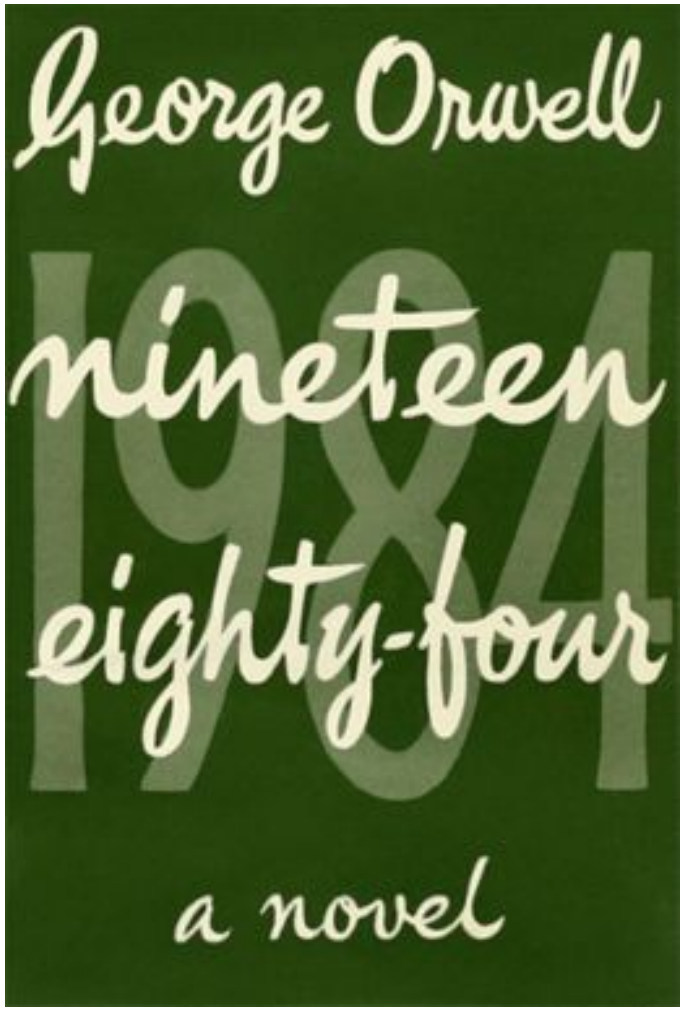
  deliberately mis-translating texts (in languages other than the revisionists)."

Wikipedia contributors. (2019, April 13). Historical negationism. In *Wikipedia, The Free Encyclopedia*. Retrieved 17:55, April 14, 2019, from https://en.wikipedia.org/w/index.php?title=Historical_negationism&oldid=892348289

One of the most famous examples of Negationism

1984

- Published in 1949
- World history is erased.
- History is then fabricated by the Ministry of Truth to support the narrative of a totalitarian government.
- Inspired vernacular:
  - Orwellian
  - Big Brother
  - Double Think



Lost Cause: Mis-remembering the American Civil War

- Reframing of the Civil War as a conflict over states rights, downplaying the role of slavery as a factor in the formation of the Confederacy.
- Recasting of Confederate leadership as being against slavery.
  - The idea that Robert E. Lee was against the institution of slavery.
- Describing the Reconstruction government as ineffectual because of Federal mismanagement.
  - In Reality: Undermined by white militias who terrorized and assassinated lawmakers.
  - Federal government did very little to back up the Reconstruction government.
- Paint Jim Crow laws as well meaning attempts to protect black Americans by keeping them seperate.
- Downplaying the atrocities of chattel slavery.
- Generally just romanticizing the Confederacy as "Southern Heritage".

Mlldred Lewis Rutherford (1851-1928)

Every atrocity has its negationists

- **Holocaust Denial:** Dismiss historical accounts of the extent of the Holocaust as either exaggeration, or that it is entirely fiction; Particularly the use of gas chambers.
- **Armenian Genocide**: Turkey denies there was a genocide, while Pakistan denies the existence of Armenia. United States refuses to take an official position on the matter.
- **Japanese War Crimes**: Ultranationalists wrote a number of school textbooks denying or omitting many events involving war crimes committed by the Japanese government prior to and during WWII.

http://bit.ly/2KDU40n	Turkey
https://bbc.in/2UFVQCR	U.S.
http://bit.ly/2GmZGru	Pakistan
https://stanford.io/2Uyvz9k	Japan

How to circumvent negationism?

- Days of Remembrance:
  - Many EU nations hold a day of remembrance for the Holocaust and Armenian Genocide.
- Legal Action:
  - Germany, France, Belgium, Spain, and Austria have made negationism a crime.
- Education:
  - Ensuring that textbooks do not exclude the ugly details behind many historic events.
    - If you leave a vacuum it will get filled with something.
- Thorough documentation that is
  - Available to the public.
  - Its provenance is transparent.
  - Past errors are carefully recorded

Fronza, Emanuela. "The punishment of negationism: the difficult dialogue between law and memory." *Vt. L. Rev.* 30 (2005): 609.
Bloch, Pascale. "Response to Professor Fronza's The Punishment of Negationism." *Vt. L. Rev.* 30 (2005): 627.
Teachout, Peter R. "Making Holocaust Denial a Crime: Reflections on European Anti-Negationist Laws from the Perspective of US Constitutional Experience." *Vt. L. Rev.* 30 (2005): 655.

How can archivists circumvent tampering? Lots and lots of copies.

- **LOCKSS:** Lots Of Copies Keep Stuff Safe
  - Stanford Web Archiving Program's guiding principle.
  - Changes are evident by comparing the copies to each other.
  - An attacker would need to change ALL of the copies at ALL of the archives at the same time for changes to go unnoticed.

"We want to build systems where tampering is evident and it prompts alerts…We want to build a system where the integrity of the information is determined by the consensus of peers." ~Nicolas Taylor, Program Manager, LOCKSS

https://www.lockss.org

Using fixity to ensure data has not been changed

- Checksums:
  - When transferring files, checksums are used to ensure the file has been received correctly
  - They can also be used to determine whether a file has changed since it was created.
- Manifest.org
  - Stores the fixity of archival data at a certain date.
- Make multiple copies of the manifest!
  - Push the manifest fixity records to multiple archives.

https://www.slideshare.net/phonedude/weaponized-web-archives-provenance-laundering-of-short-order-evidence-93083713

Forensic evidence can help determine what is real

- Yvonne Ng at WITNESS compares the Deep Fake issue with traditional methods of determining authenticity of paintings.
  - Materials used.
  - Objects in the picture.
  - Setting of the picture, rooms, buildings, architecture.
- All of these can help determine when a painting was made, or when it definitely was not made.

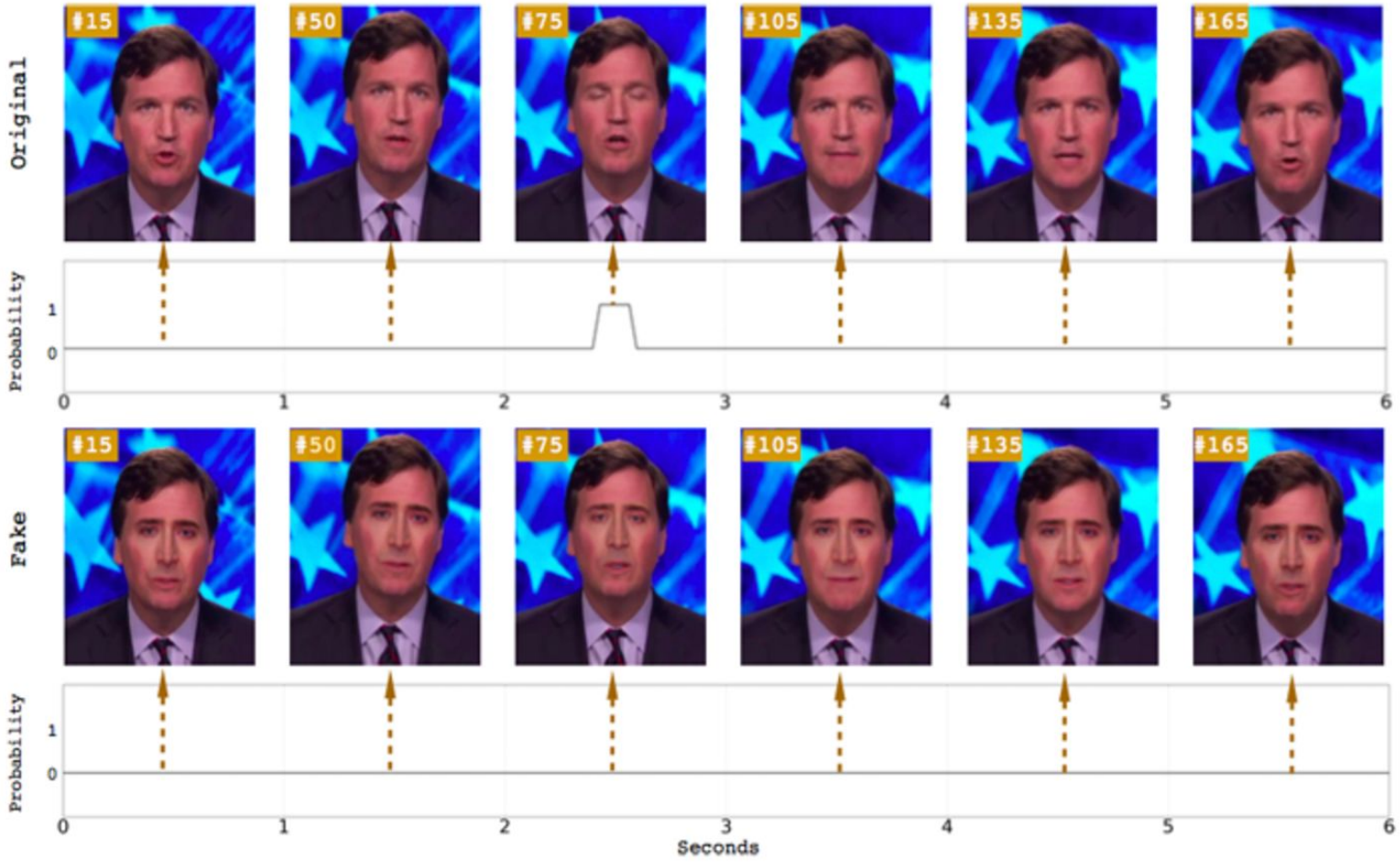
Traditional forensics also apply to digital resources

Similar techniques can be used to authenticate the provenance of a particular digital record.

- Method of creation.
- Objects in the frame
  - Say, Barack Obama is wearing a pin given to him by the Queen of England, but the video is portrayed as happening BEFORE she gave it to him.
- Background, Decor, etc.
  - White House regularly undergoes redecoration.
  - Buildings get repainted, torn down, etc.
- Language used. Particularly phrases that were not in use at the time of the video's supposed creation.

Machine learning can be used to spot the fakes created by machine learning

Researchers were able to spot Deep Fakes by detecting the number of eyeblinks.



Li, Yuezun, et al. "In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking." *arXiv preprint arXiv:1806.02877* (2018).

Most institutions are more concerned with preservation

- Data preservation over long periods of time requires a certain degree of foresight in the here and now.
  - Natural disasters
  - Bit rot
  - Environmental threats
  - Obsolescence of hardware or software

Redundancy could serve both purposes

- More copies means more of a chance of survival for data and physical objects.
  - In the post apocalypse, Dean Koontz, Danielle Steele, and Louis L'Amour will be well remembered.
- Carefully recording and labeling the Deep Fakes in multiple archives can ensure that future generations understand the context in which a Deep Fake was generated.
- More copies can also help in cases of bit rot, or other forms of data loss.
- Fixity manifests can help maintain provenance as well as alert to changes due to bit rot or data loss.

Archives can watch each other

- Push documents to multiple archives.
- Record fixity at time of push.
- Push fixity of documents to multiple archives.
- If a document is changed or damaged
  - Fixity will show which documents have been changed.
  - If fixity doesn't agree, multiple copies can help determine which fixity manifest is most likely original.

https://www.slideshare.net/phonedude/weaponized-web-archives-provenance-laundering-of-short-order-evidence-93083713

Take Away:

- Archivists already have some defense against Deep Fakes.
  - Forensic tools.
  - Careful documentation of previous fakes.
- The tools used to create Deep Fakes can also be used to combat them.
  - Machine learning can be leveraged to spot Deep Fakes.
- Redundancy and fixity can serve both the need for clear documentation of provenance, and data preservation.
- Archives can essentially police each other through consensus.