Secrets, Lies, and Account Recovery: Lessons from the Use of Personal Knowledge Questions at Google

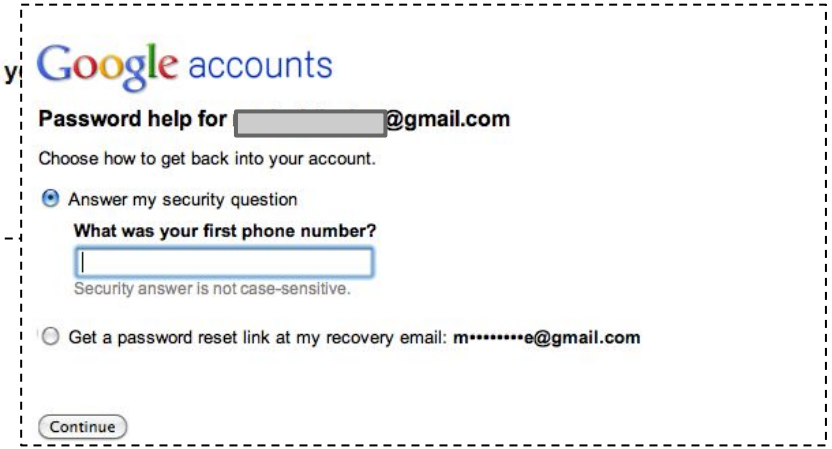
Joseph Bonneau, Elie Bursztein (elieb@google.com), Ilan Caron, Rob Jackson, Michael Williamson

Secret question goal: users use “secret” knowledge to recover their accounts

only in specific cases

Facebook

Yahoo



Google

Targeted attack

Sarah Palin's Yahoo account hacked in 2008 via secret question

her 1st question: "date of birth" - 2nd: "where did you meet your spouse"

Sarah Palin

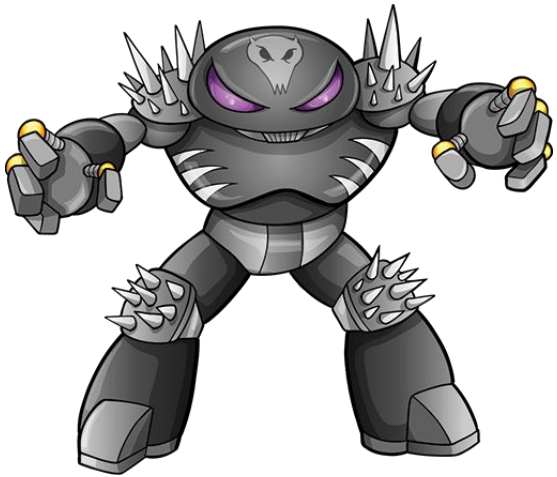
From Wikipedia, the free encyclopedia

Sarah Louise Palin (/ˈpeɪlɪn/; *née* **Heath**; born February 11, 1964) is an American politician, commentator, and author who served as the ninth Governor of Alaska, from 2006 to her resignation in 2009. As the Republican Party nominee for Vice President in the 2008 Presidential election, alongside Arizona Senator John McCain, she was the first Alaskan on the national ticket of a major political party, and the first Republican woman nominated for the Vice Presidency. Her book *Going Rogue* has sold more than two million copies. Since January 2010, she has provided political commentary for Fox News, and starred in a reality television show, *Sarah Palin's Alaska*.

She was elected to the Wasilla City Council in 1992, and became Mayor of Wasilla in 1996. In 2003, after an unsuccessful run for Lieutenant Governor, she was appointed Chairman of the Alaska Oil and Gas Conservation

Sarah Palin

Large scale attack



Attempt to hijack accounts at scale by guessing answers to secret questions

Not that simple in practice!

Most companies enforce some rate limiting

Attackers have only a few attempts per-account/IP etc...

Secret questions are combined with other factors

At Google and possibly other places, the secret question answer is not enough to recover an account

Still important to understand the security - usability at scale

Tailor risk analysis systems - compare to other recovery methods

Dataset used

Security analysis: Hundreds of millions of secret questions answers

Each data buckets has above 100.000 answers

Usability analysis: ~11 million of account recovery claims

Data from 2013, used to measure success rate

Crowdsourcing attack: 1000 respondents from crowdflower

Used to evaluate the effectiveness of crowdsourced distributions

Outline

How secure are secret questions?

For real

How successful are people at answering their questions?

By reviewing account recovery claims

Is there any hope? and what is the future?

Can we fix secret questions? What can replace it

For more analysis please read the paper
http://goo.gl/EDqkVC

1 Secret question security

How attackers can build answer dataset?

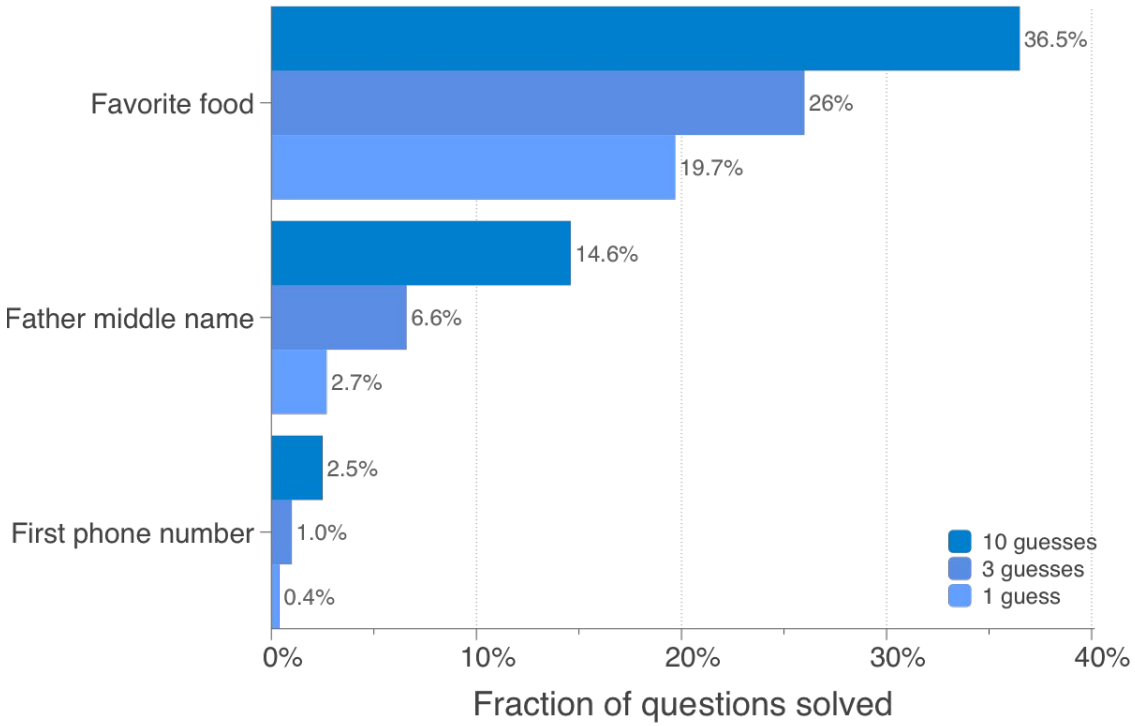
Scrape public sources

Birth registry, social profiles, yellow pages, school yearbooks ....

Use crowd-sourcing

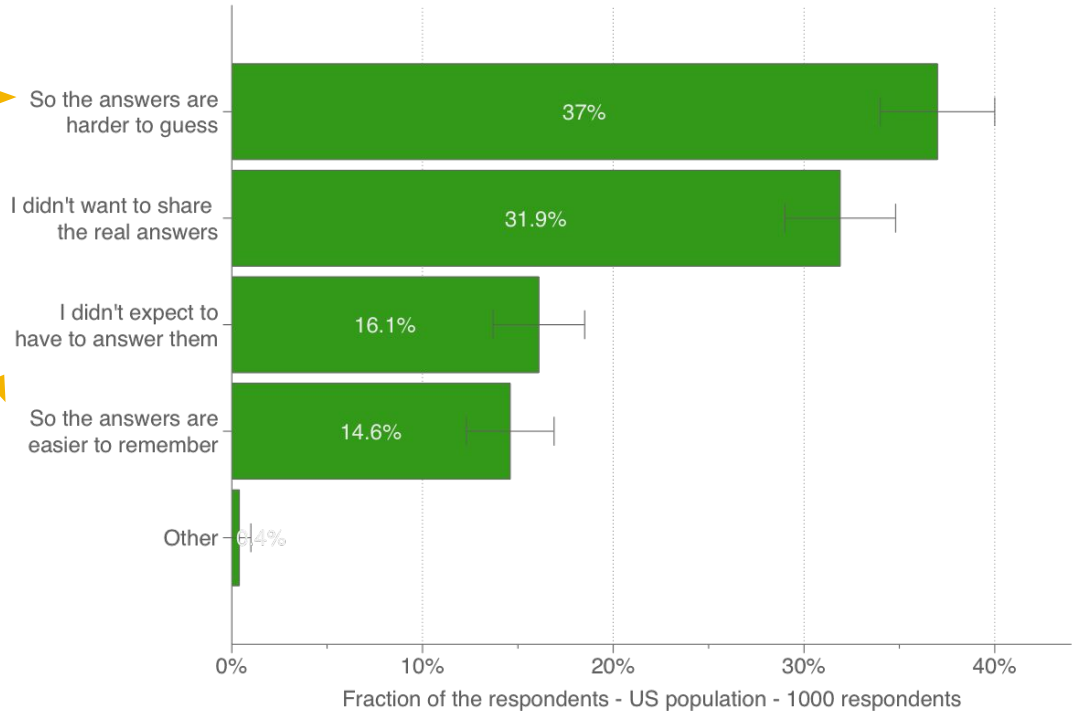
Ask internet users the same questions to be targeted

Security inequality

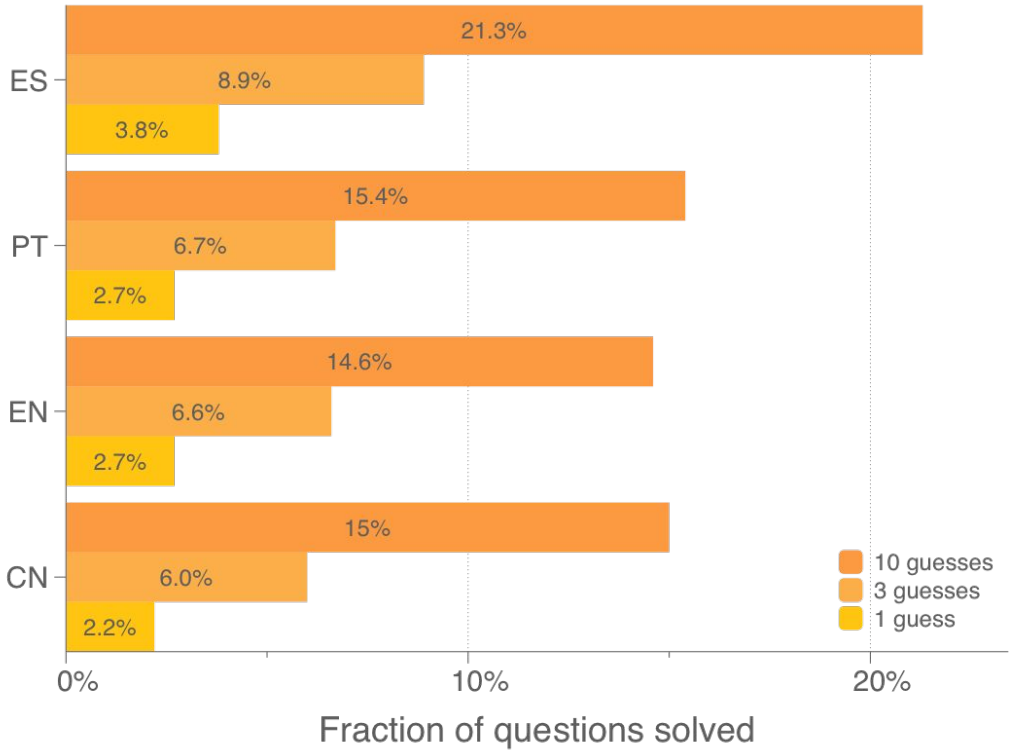


Why people provide inaccurate answers - survey

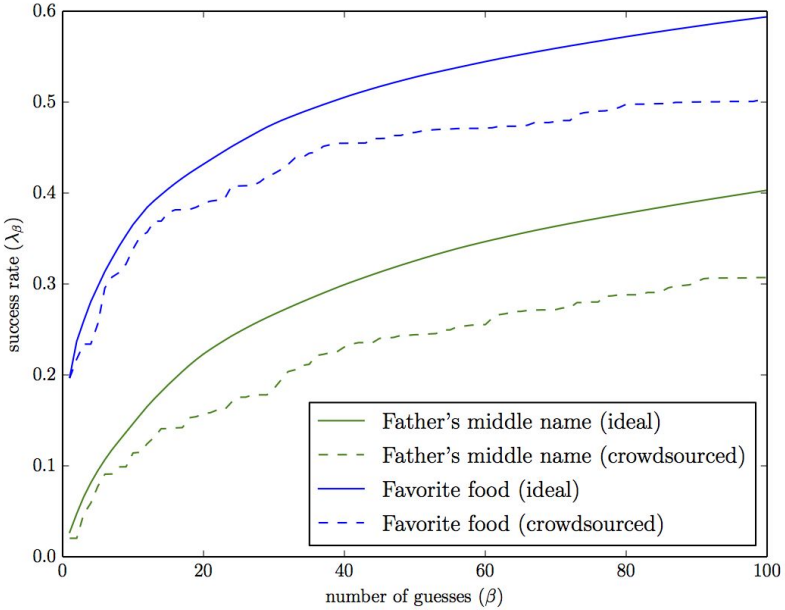
achieve the opposite



Father middle name? - country specificity



True distribution vs crowd source



Crowdsourcing can be used to approximate the true distribution for the easy questions

Takeaway

Most questions have weak resistance to guess-based attacks

This is inherent from the underlying distribution

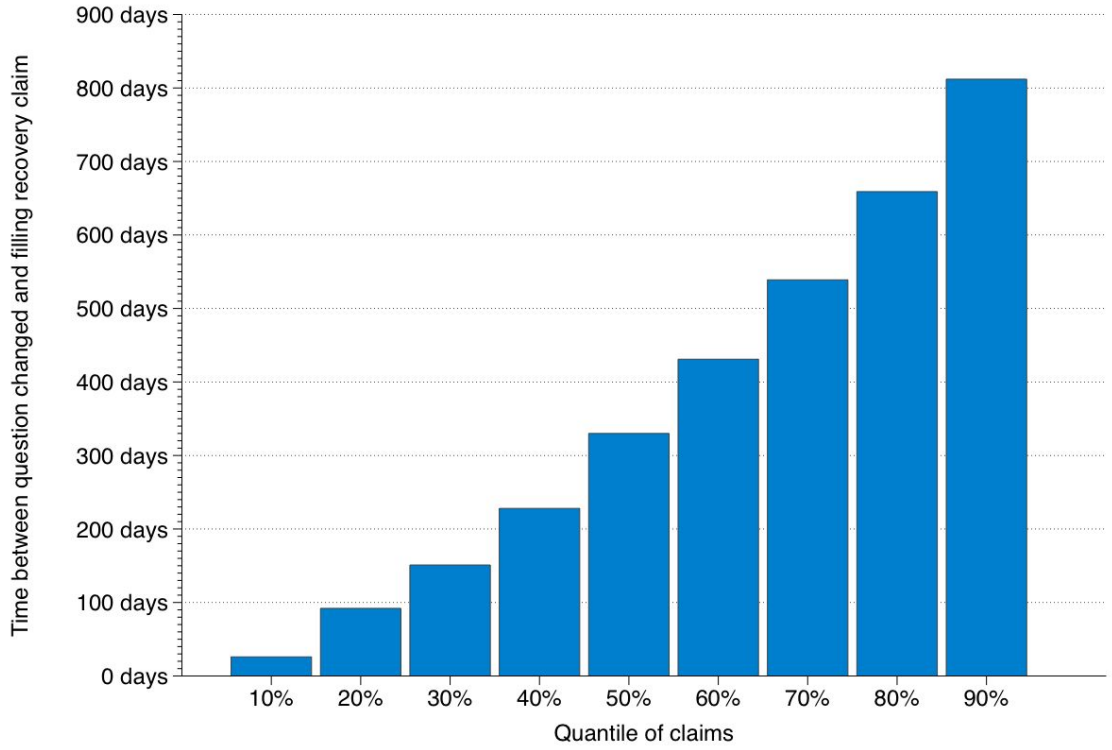
Strongest questions security is degraded by unexpected user answers

This is due to people's behavior, not the underlying distribution

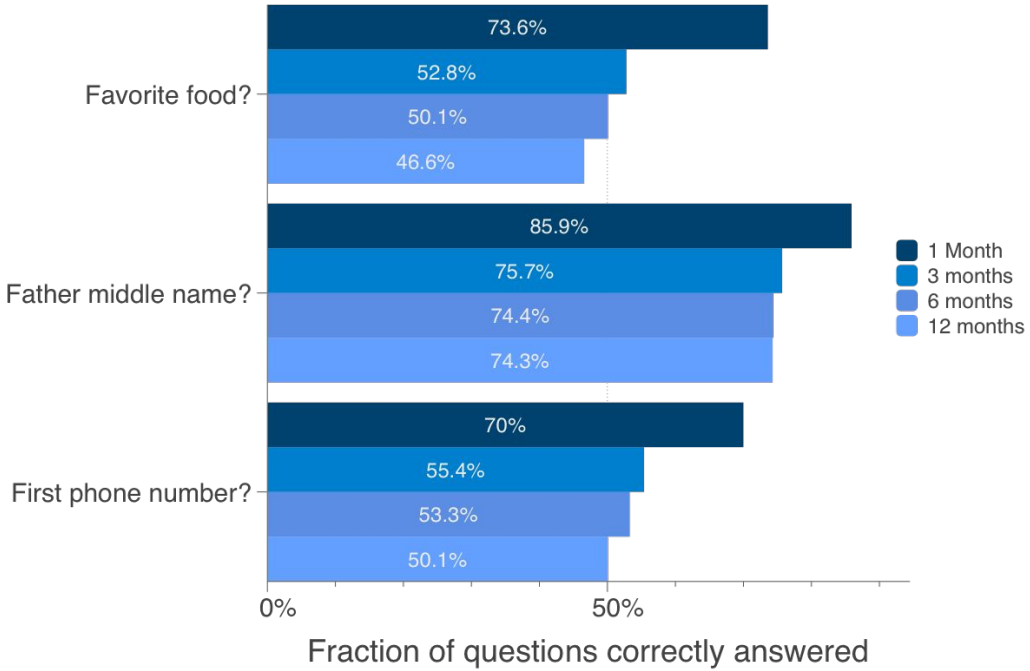
Crowd source and public data is an efficient proxy to approximate true distribution

2 Secret question usability

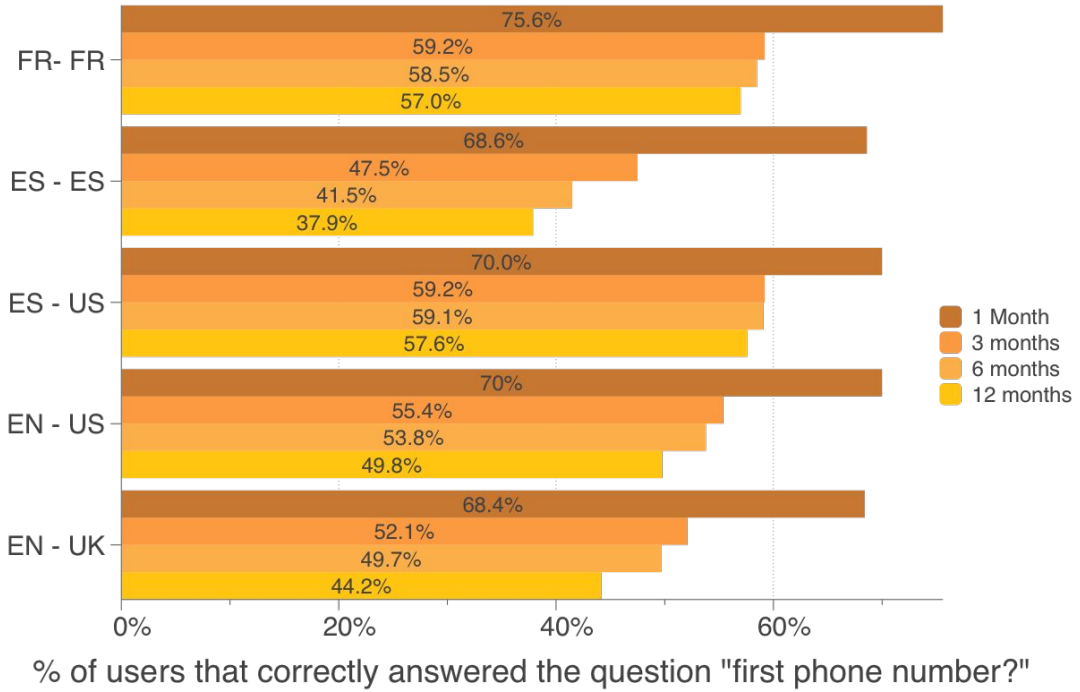
When do people recover their account?



Recall rate for some US questions



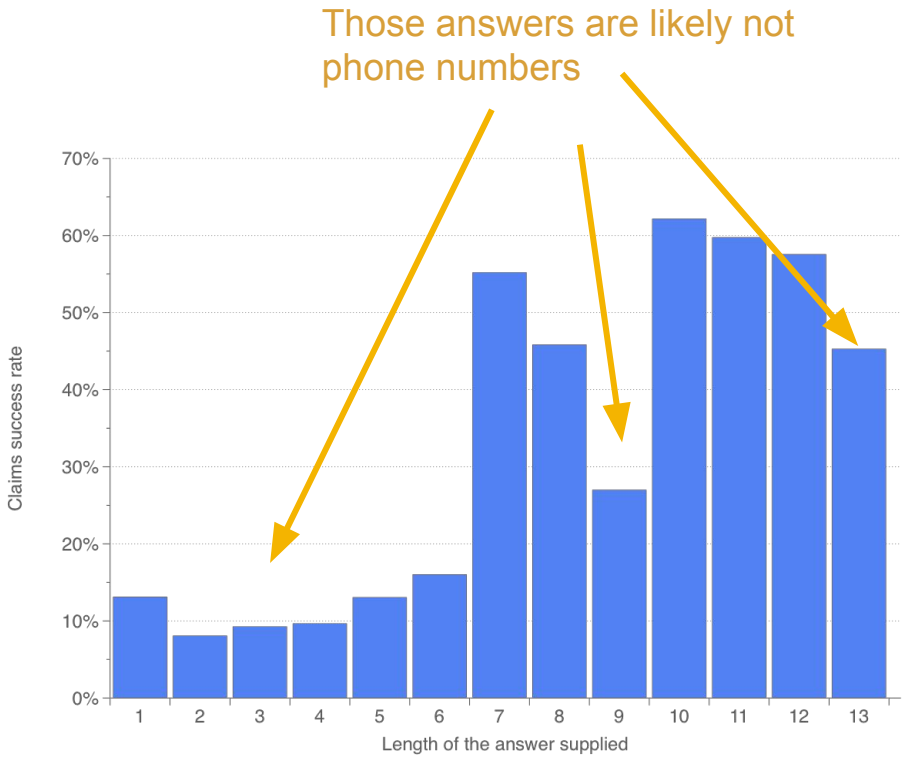
Language & country effect on answer recall



Inaccurate answers yield to poor recall

US phone number format:
(123) 456 7890

valid formating (len):
4567890 (7)
456-7890 (8)
1234567890 (10)
123-4567890 (11) < odd
123-456-7890 (12)



Takeaway

Secret questions' recall decreases over time - some of them faster

Human and place are better remembered

Answer recall is country dependent

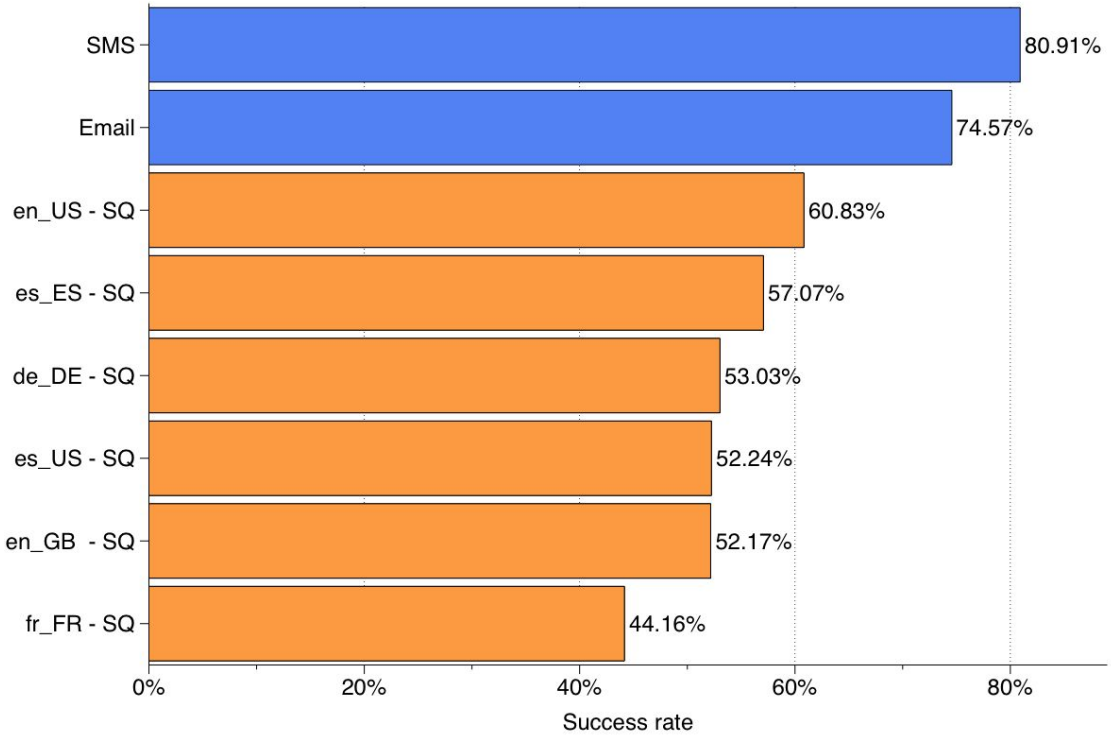
Might be due to regional specificity e.g language structure

Providing inaccurate answers yields worse recall

Inaccurate answers are a key issue

3 Moving forward

Alternatives offer better usability (and security)



Conclusion

Secret questions are not secure

Either because of the underlying distribution or inaccurate answers

Secret questions have poor recall - strong ones having the worst recall

Inaccurate answers also significantly decrease answer recall

Alternative options provide better recall and are more secure

Use secret questions only if you can combine with other signals

Thank you - questions?