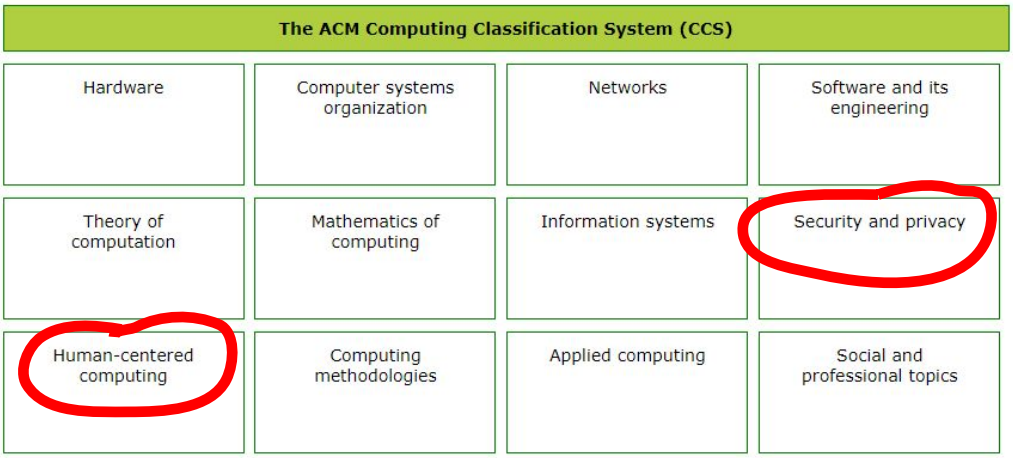
Snooping on Mobile Phones: Prevalence and Trends

Diogo Marques	University of Lisbon	dmarques@di.fc.ul.pt
Ildar Muslukhov	University of British Columbia	ildarm@ece.ubc.ca
Tiago Guerreiro	University of Lisbon	tjvg@di.fc.ul.pt
Konstantin Beznosov	University of British Columbia	beznosov@ece.ubc.ca
Luís Carriço	University of Lisbon	lmc@di.fc.ul.pt

[Slides & transcript of presentation for students in the Data Science Masters program at FC/ULisboa, November 2017. Based on earlier presentation at the Twelfth Symposium On Usable Privacy and Security, Denver, CO, June 2016. Report at https://www.usenix.org/conference/soups2016/technical-sessions/presentation/marques]

I'm going to be speaking about this report, Snooping on Mobile Phones: Prevalence and Trends, which was all about quantifying how common it is for people to snoop on one another's phones.



"Human and Societal Aspects of Security and Privacy"
a.k.a. "Usable Privacy and Security"

2

Before we start let me just give you some context for the kind of work we do.
This is ACM's classification system for the discipline of Computer Science. ACM is the scholarly society that aims to advance the field of Computer Science. For instance, ACM gives out the Turing Award, the Nobel prize of Computer Science. As you can see there are many domains within Computer Science. Our research lies in the intersection of these two domains, "Security and Privacy", and "Human-centered computing". ACM calls this intersection "Human and Societal Aspects of Security and Privacy", but the field is usually just called "Usable Privacy and Security". This is mainly for historical reasons. In the early days, what people do was to look at whether people could use security technologies, like encrypted email or passwords. Nowadays, of course, the field is much broader.

Still, when I say I work on security and privacy people conjure up images like this.
A lone hacker doing something shady.
People expect me to again tell them they should fear hackers, and they should fear ransomware, and they should definitely fear government surveillance, and online tracking by private companies and all that.
Indeed, there are issues of concern.
But notice this. In these threats we are told to fear, we are victims, and there some big bad wolf working against us.

Consider this other picture.
This is clearly a very different kind of security problem, and one that is a bit more uncomfortable to talk about. We can look at this picture and think:
“Well, she is the bad guy here, she is snooping on his phone! She should not be doing that under any circumstances.”
Or we may think: “What has that guy been up to that justifies her curiosity. Maybe he has done something bad before, so her behavior is justifiable”.
The point is, this is a much messier situation than the threat of hackers or ransomware. There is no clear big bad wolf. It's just people trying to live their lives.
This messiness is also a problem for us when we want to measure these types of phenomenons. How do we measure a phenomenon like people snooping on one another’s phones?

I have looked through someone else's phone without their permission. Y/N?

5

So let's make an experiment here. Look at this statement: "I have looked through someone else's phone without their permission. Yes or no?.
Raise your hand if your answer is yes.
~
I'm pretty sure that some of you have done this but not raised your hand.
Which is a problem, because this is exactly what we wanted to ask people, and if they do not raise their hands, then how can we measure it?
Well, we of course found a way to do it, otherwise I wouldn't be here.
First, I'm going to show you our topline result, the prevalence rate we found, and then I'll go back and tell you how we got there.

An estimated

30%

identified with "In the past 12 months, I have looked through someone else's phone without their permission"

(n = 1,381, MTurk)

6

Here it is:
In a sizeable group of people, we estimated that 30% identified with the following statement:
In the past 12 months, I have looked through someone else's phone, without their permission.
Notice that I said this is an estimate, not the proportion of participants who explicitly indicated that they identified with the statement.

Perfectly legitimate questions

Q1: How can that question be asked?

Q2: Can it be asked to MTurk workers?

Q3: How was the prevalence estimate obtained?

Q4: Who is more likely to conduct snooping attacks?

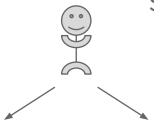
I imagine you have some questions about this number, so the rest of this presentation is structured like a Q&A.
To answer question 1 (“How can that question be asked”), I’m going to talk a bit about a survey technique we used, called the list experiment, which is designed to address sensitive questions.
To answer question 2, I’ll speak a little bit about this MTurk thing, which is where we recruited participants. We did some validation to see if we could get good quality data from that source.
To answer questions 3 and 4 (“How was the prevalence estimate obtained”, “Who is more likely to conduct snooping attacks”), I’m going to go into the empirical data we gathered.

Q1: How was the question asked?

How was the question asked? How can we ask people if they snooped on others, and expect them to answer honestly.

List experiment

D. Raghavarao and W. T. Federer (1979). "Block Total Response as an Alternative to the Randomized Response Method in Surveys". Journal of the Royal Statistical Society, Series B 41 (1): 40–45.



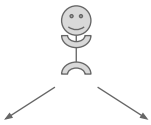
Control	Treatment
Bananas	Bananas
Standing in lines	Standing in lines
Rainbows	Rainbows
	Marijuana

How many of these items do you love?

9

One way to do it with a list experiment. (Also sometimes called the unmatched count technique, or item count technique). It is a well-known survey technique for self-incriminating questions, and has been around since at the least the 1979, when it was formally described by Raghavarao and Federer. Here's how it works.
Each participant is randomly assigned to be in either the control group, or the treatment group.
Participants in the control group, receive a list of items, and they are asked to say how many, not which ones, they identify with.
Participants in the treatment group, receive an extra item in the list. This will be the item that we're trying to measure.
Now we can estimate the estimated proportion of people who identified with the extra item just by calculating the difference in means.

List experiment



Control	Treatment
Bananas Standing in lines Rainbows	Bananas Standing in lines Rainbows Marijuana
$\bar{x} = 2.0$	$\bar{x} = 2.5$

How many of these items do you love?

Estimated proportion of respondents who identify with loving marijuana: (2.5 - 2.0) = 0.5

To see it, imagine a very uniform population, where everyone loves bananas and rainbows and everybody hates to stand in line. One aspect in which this population is not uniform, is that only half loves Marijuana, and the other half does not care for it.
In the control group everyone would answer two, referring to bananas and rainbows, so the mean would be 2.
In the treatment group, half the people would still answer 2 (bananas and rainbows), and half the people would answer 3 (bananas, rainbows and marijuana). So the mean would be 2.5.
The difference, as expected, it then 0.5, or half the population.

Perfectly legitimate questions

Q1: How can that question be asked?

A1: With a list experiment.

Q2: Can it be asked to MTurk workers?

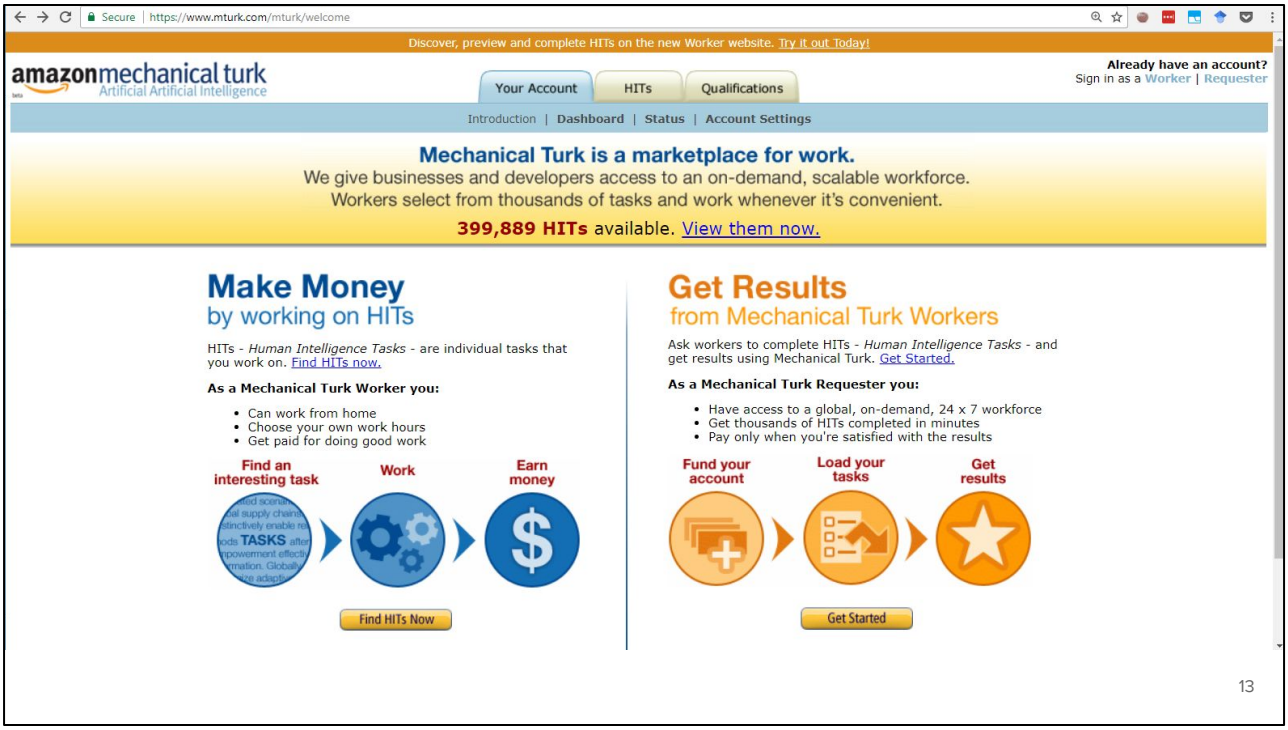
Q3: How was the prevalence estimate obtained?

Q4: Who is more likely to conduct snooping attacks?

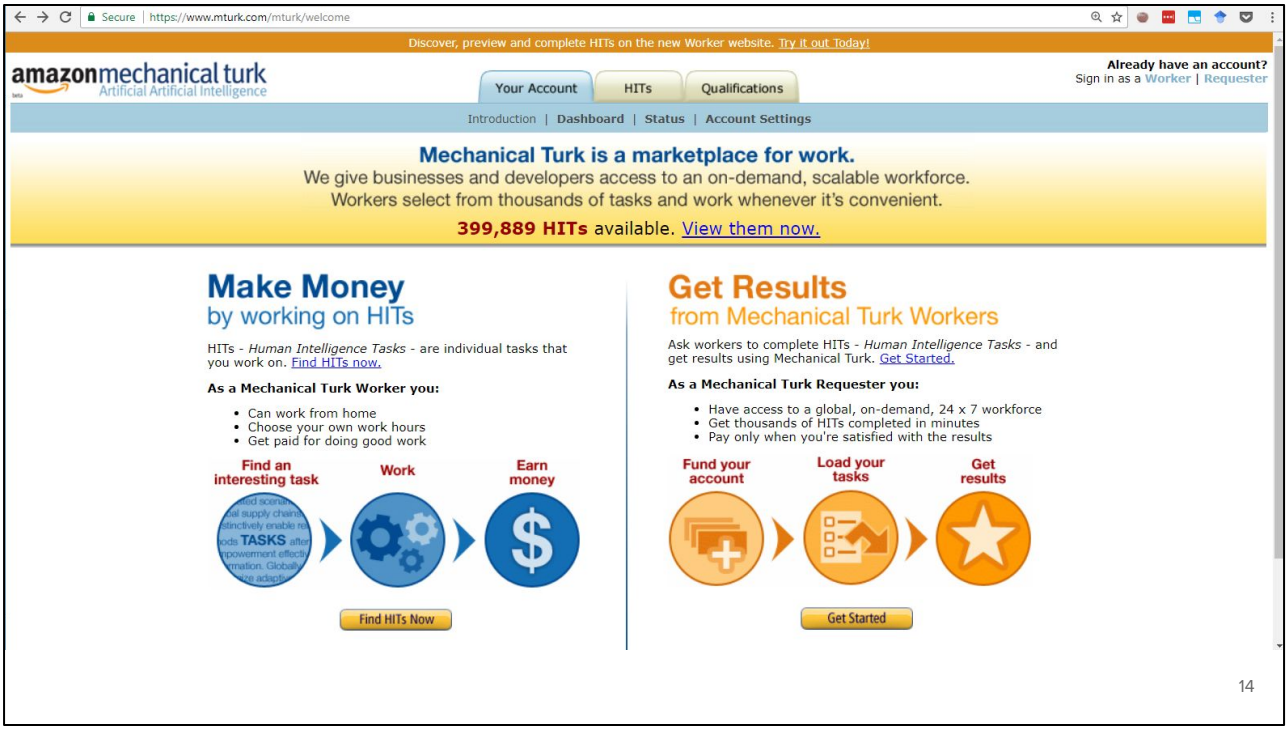
To answer question 1, “How can the question be asked?”: one way is with a list experiment.

Q2: Can it be asked to MTurk workers? (i.e., are list experiments on MTurk trustworthy?)

Moving on to question two. “Can it be asked to Mechanical Turk workers”, or, “are list experiments on MTurk trustworthy”.



What is MTurk. I'm sure some of you know it, but please bear with me. Mechanical Turk is a service ran by Amazon. They have a pool of hundreds of thousands of people who do small tasks in exchange for a fee. The idea behind this service was that if you had some kind of problem to solve in which humans are better than computers, you could actually offload it as a microtask to a human through the MTurk service. There are actually lots of things you think are automated, that in fact have humans behind it to work well. Google uses it to improve search results, to feed their indexes, to detect fraud. Other big companies use it to moderate content. Lots of companies that do translation. Accounting companies use it to add up receipts, because even with OCR it can be difficult for a computer to find the total for receipts when they are digitized and in all shapes and sizes. This was what MTurk was created for.



MTurk was launched in 2005, and it still looks like a website from 12 years ago. But meanwhile it started to be used for a host of other things. One of them was, of course, to train machine learning algorithms. If you're creating an algorithm to detect pictures of cats, it sure would be nice to have a huge dataset of pictures labeled "cat" or "not cat". And if that database has thousands or millions of images, you can't label them yourself, but maybe you can pay 1 cent per picture to have a distributed workforce label it for you.

The other thing that happened was that Mechanical Turk started being used on every field where surveys are used. This includes marketing, political science, cognitive science, psychology, public opinion, and so on. Of course, a microtask can be classifying an image in 3 seconds, but it can also be to respond to a one-hour survey, as long as you pay enough.

Satisficing

But here's the rub. In survey research, there is always this problem of satisficing. That is, participants answering more or less at random, instead of pondering the question and answering it. On paper, they answered the question, but the data you get can be meaningless. In Mechanical Turk there are actually financial incentives to act this way. The less time they spend on task, the quicker they can move to another one, and make more money.

Design: control and treatment groups

Control

"Below is a list of experiences you might have had in the past 12 months.
- I've shared pictures from my cell phone on social media.
- I've purposefully made phone calls or sent text messages to 1-900 numbers.
- I've received at least one text message with unsolicited advertising (spam) on my cell phone.
- I've lost or had my cell phone stolen from me.

Please count how many you have had and indicate below."

Treatment-0 Expected prevalence: ~0%

Control list + "I've been to space, aboard and interplanetary vessel that I built myself."

Treatment-1 Expected prevalence: ~100%

Control list + "I've opened my eyes in the morning at least once (for instance, after waking up)."

16

Therefore, before we went any further, we decided to test if a list experiment like the one we wanted to deploy would do well on Mechanical Turk. We designed an experiment as follows.
We had a control group, that received the same list of items that we were going to use in our study of snooping attacks. We then had two treatment groups. Each of these groups had an extra item for which we had an expected prevalence.
The group treatment-0 had an extra item about travelling to space, for which the expected prevalence would be 0, or near 0. The group treatment-1 had an extra item about opening eyes in the morning, for which the expected prevalence would be 1, or close to 1.
If respondents answered carefully, we would expect that the prevalences estimates obtained with the list experiment matched these known prevalences of 0 and 100%.

Results: estimate vs. ground truth

	Flown in space (Treatment0 - Control)	Opened eyes once (Treatment1 - Control)	Opened eyes once (Treatment1 - Treatment0)
Proportion estimated by list experiment	0.10	1.07	0.97
Expected proportion	0.00	1.00	1.00
Bias (difference)	+0.10	+0.07	-0.03

(MTurk sample n = 434)

Here is what we found. In the first row you can see the proportion that was estimated by the list experiment, and in the second row we can see what we know that proportion to be, because we know that almost no-one has travelled in space, and that almost everyone has opened their eyes in the morning at least once in a year. As you can see in the third row, which shows the difference, the estimates are very close to the expected proportions. That is good news.

That is especially so when we take the difference in means between the two treatment groups, as you can see in the third column. We know the difference should be 1. The estimate we get comparing these two groups is 0.97.

What's happening here, we think, is that the two extra items we created for the treatment groups, are being interpreted by participants as attention checks. So when we compare the two groups where those extra items were shown, we get estimates that are closer to what we expect. For that reason, in our study of snooping attacks, we added these two extra items both to the control and the treatment group.

Perfectly legitimate questions

Q1: How can that question be asked?

A1: With a list experiment.

Q2: Can it be asked to MTurk workers?

A2: Yes, with caution.

Q3: How was the prevalence estimate obtained?

Q4: Who is more likely to conduct snooping attacks?

To answer the second question, in the Q&A “Can it be asked to MTurk workers. The answer is yes, list experiments do very well. They provide especially good estimates if there are attention checks in the list of items itself.

Q3: How was the prevalence estimate obtained?

Moving on to question 3: “How was the prevalence estimate obtained”. I’ll just give an outline of our final design, before we go into more detail about the results.

Design

Control group question

Below is a list of experiences you might have had in the past 12 months.

- I've shared pictures from my cell phone on social media.
- I've opened my eyes in the morning at least once (for instance, after waking up).
- I've purposefully made phone calls or sent text messages to 1-900 numbers.
- I've received at least one text message with unsolicited advertising (spam) on my cell phone.
- I've been to space, aboard and interplanetary vessel that I built myself.
- I've lost or had my cell phone stolen from me.

Please count how many you have had and indicate below.

Here is the list that we used.
The control list had 4 items relating to mobile security and privacy, which are the ones in grey. These items were selected based on a study that I'm not discussing today, but you can find the paper. The basic idea is that there is a mix items of very low prevalence and medium prevalence, so as to avoid ceiling and floor effects, which is to say, so as to increase plausible deniability.
The control list also had two attention checks, which you can see in blue. Those attention checks were added based on the study on satisficing.

Design

Treatment group question

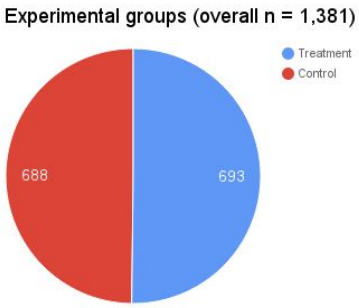
Below is a list of experiences you might have had in the past 12 months.

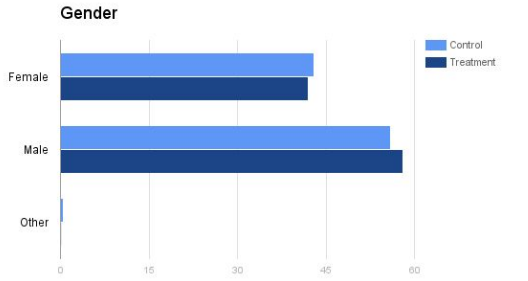
- I've shared pictures from my cell phone on social media.
- I've opened my eyes in the morning at least once (for instance, after waking up).
- I've purposefully made phone calls or sent text messages to 1-900 numbers.
- I've received at least one text message with unsolicited advertising (spam) on my cell phone.
- **I've looked through someone else's cell phone without their permission.**
- I've been to space, aboard and interplanetary vessel that I built myself.
- I've lost or had my cell phone stolen from me.

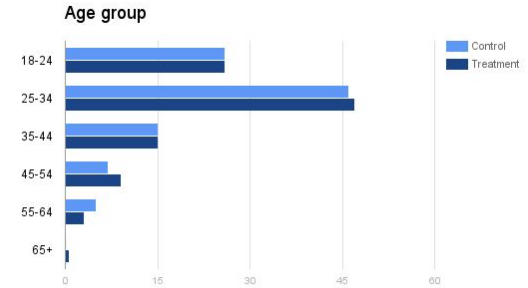
Please count how many you have had and indicate below.

The treatment group had the same items as the control, plus the statement about snooping attacks, which you can see underlined. The particular wording that was used was also determined by a study that's in the paper.
The survey was fielded to MTurk, only to workers in the US.

Participants







12

After the list question, we asked just a few demographic questions. We did not have many questions, or questions that were too specific, as we thought it would decrease the sense of anonymity. So we asked only about gender, age, level of education, and state of residency, and whether participants were smartphone owners or not.
We had 1381 valid responses, about evenly divided between the experimental groups. Demographically, we had some diversity; and the experimental groups were very similar demographically, as would be expected from random assignment.

Results: prevalence

- Control group mean = 2.517
- Treatment group mean = 2.825
- Difference in means (estimate) = 2.825 - 2.517 = 0.30

So here's how that topline result was calculated:
- The mean number items selected by participants in the control group was 2.517
- The mean number of items selected by participants in treatment group was 2.825
So there is a difference, and it is 0.308, which is to say that the estimated prevalence is 30%.

1 in 5

U.S. adults estimated to have conducted snooping attacks in the year before the survey was conducted.

(MTurk sample n = 1,381; Sample adjusted by cell-based post-stratification weighting to the 2010 Census by age and gender.)

Just before we move on to the analysis of trends, I also wanted to show you this. We adjusted the Mechanical Turk data to the US 2010 census data, and got an adjusted estimate of 20% of US adults having had conducted snooping attacks in the last year before the survey. This was done by cell-based post-stratification weighting, a technique that is often used in election polls. The difference in the prevalence rates happens because the US population is older and more gender-balanced that our sample. Although our sample, as is common in MTurk samples, reflects very closely what is usually called the US Internet Population, which is just the US population that regularly used the internet.

Perfectly legitimate questions

Q1: How can that question be asked?

A1: With a list experiment.

Q2: Can it be asked to MTurk workers?

A2: Yes, with caution.

Q3: How was the prevalence estimate obtained?

A2: Sizeable MTurk list experiment.

Q4: Who is more likely to conduct snooping attacks?

We've now answered question 3, "How was the prevalence estimate obtained", and can move on to analyzing trends in the data.

Q4: Who is more likely to conduct snooping attacks?

The question is: “Who is more likely to conduct snooping attacks”. To answer this question, we needed to build a concise model of the likelihood of snooping on others, based on the demographic variables that we had collected.

Variable selection

Available variables:

- Gender
- Age
- Education
- Geographical region
- Smartphone ownership

Strong predictors:

- Gender
- Age
- Education
- Geographical region
- Smartphone ownership

So we first had to look at which variables had more predictive power. We did that by running, with each demographic variable, linear regressions of how many items participants selected, controlling for group assignment.
In this exercise, we found that gender, education and region were not a very strong predictors. Age and smartphone ownership, on the contrary, seemed to be the best predictors.

Variable selection

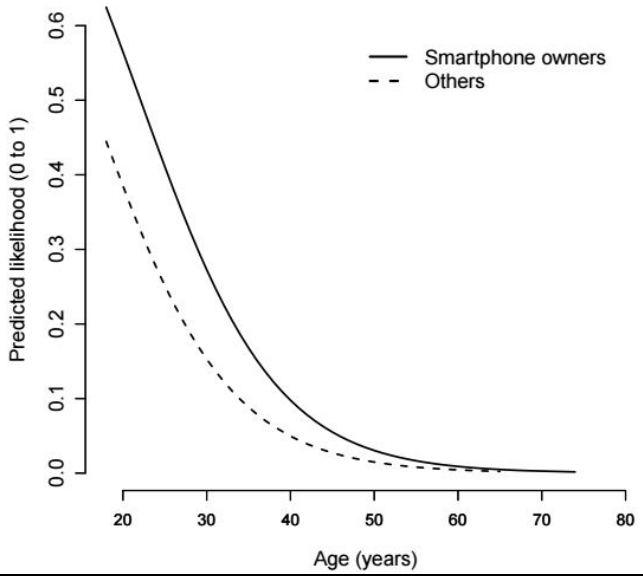
Linear regression models of number of items selected. The first row indicates the proportion of variance explained by being in the treatment or control group. In the remaining rows, a variable is added to that model. F statistic from an ANOVA of the smaller and larger models.

Predictor variables	R^2	ΔR^2	*F*	D.f.	*P*-value
GROUP	0.022				
GROUP + GENDER	0.025	0.003	1.87	2	0.1542
GROUP + AGE	0.053	0.031	44.78	1	**<0.0001**
GROUP + EDUCATION	0.031	0.009	2.47	5	**0.0306**
GROUP + REGION	0.025	0.003	1.32	3	0.2671
GROUP + OWNER	0.100	0.077	118.38	1	**<0.0001**

Here you can that in more detail. We started with the model in the first row, which is just a model of items selected per group - control or treatment. We then built a model for each of the other variables, plus group assignment, and compared to that model in the first row.

If you notice, in this table it looks like education might also be a good predictor, since the p-value is below .05. But the reality is that if we break down that regression, there is no ordinal pattern, like if you're more educated you're more likely to select more items. It's basically random, so education was also not considered a good predictor for a model.

Trends: age and ownership



(MTurk sample n = 1,381)

So we those two variables, age and smartphone ownership, we built this other model. This is a specialized regression for list experiments developed by Kosuke Imai. This is a model of the estimated likelihood of having conducted snooping attacks, based on age and smartphone ownership. This graph shows what that model looks like.
We can see some trends in this figure:
- Younger people are more likely to snoop on others. In fact, for the youngest participants, this behavior seems to be prevalent.
- The likelihood is sharply less as people get older.
- People that have smartphones, are themselves more likely to snoop on others.

Follow-up study: depth of adoption

- Usage of young smartphone owners is often “deep”
  - Deep: usage that makes smartphone keep lots of private data
- Is there a relationship between deep usage and likelihood of snooping?
  - Rationale: users learn by their own experience what information they could gather in a snooping attack
- New study:
  - Same list experiment question
  - Only smartphone users
  - “Depth of adoption” measured with 10-item scale

Looking at this model, it appeared to us that if it was younger smartphone owners that were snooping more on others, then there must be something different about this population in terms of usage.
So we decided to do a follow-up study, to see if there was a relationship between snooping, and having a kind of usage which we called “deep”, meaning that the person uses the phone in a such a way that it keeps lots of private data.
The logic behind this is that when people have a “deep” kind of usage, they will learn by their own experiences what kinds of sensitive information they might get access to if they were to snoop.
To see if this was true, we conducted another list experiment. This time, we asked specifically for smartphone owners. We removed demographic questions aside from age. And finally we added a scale of depth of adoption that we created based on existing survey data on privacy-sensitive usage.

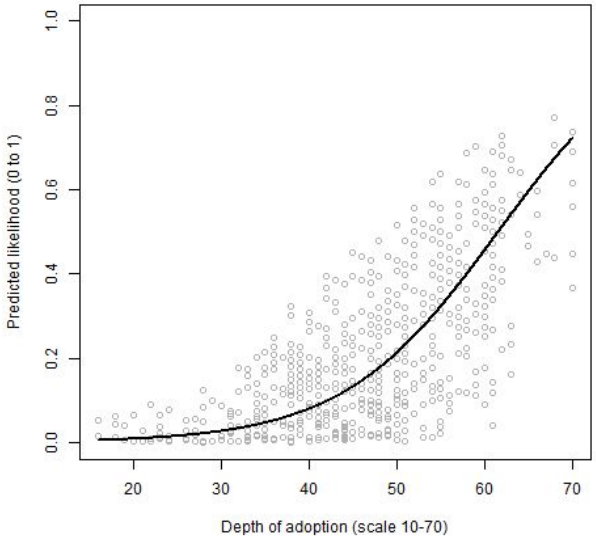
“Depth of adoption” scale

Please answer on a scale from 1 to 7, where a 1 means that the statement indicates something you feel like you never do, and a 7 means that the statement indicates something you feel like you do all the time.

- I use my smartphone to check my personal email account
- … to take pictures of myself or of people close to me.
- … to go on social networks (like Facebook, Twitter, Snapchat) with my personal account.
- … to exchange instant messages with people that are close to me.
- … to to look up information about health conditions.
- … to do online banking on my personal accounts.
- … to look up jobs or submit job applications.
- … to look up government services or information.
- … to look up directions to places, or to get turn-by-turn navigation.
- … to organize personal affairs (for instance, access personal notes, calendar or shopping list).

Just to give you an idea, this is what the scale looked like. There were 10 items that participants rated from 1 to 7, based on how frequently they felt they did each of these things. For instance, one of the items asked if participants went on social networks on their phones, another if they looked up health conditions, and so forth. For each participant, we added the responses to individual items, to got a “depth of adoption variable” with range from 10 to 70.

Trends: depth of adoption



(MTurk sample n = 653)

We deployed this new list experiment to MTurk, and this time we gathered 653 valid responses. With this new data we built a regression model of likelihood of having engaged in a snooping attacks in the year prior, based on response to the scale, and controlling for age.

We can clearly see that there is indeed a positive relationship between adopting smartphones more deeply, and engaging in snooping attacks. That is illustrated by the trend line going up as the depth of adoption score goes up.

What we can conclude is that among smartphone users, those that adopted the technology more deeply are also more likely to snoop on others.

This is an interesting result because it's a more stable relationship. So the rates of 20% or 30% may increase or decrease as times goes, but this relationship between having a “deep” kind of usage, and attempting to snoop on others, and more structural and possibly lasting.

Perfectly legitimate questions

Q1: How can that question be asked?

A1: With a list experiment.

Q2: Can it be asked to MTurk workers?

A2: Yes, with caution.

Q3: How was the prevalence estimate obtained?

A2: Sizeable MTurk list experiment.

Q4: Who is more likely to conduct snooping attacks?

A3: Young smartphone owners, who use them in ways that generate privacy-sensitive data.

We can now answer our last question: “Who is more likely to conduct snooping attacks?”. Our models indicate that younger people, smartphone owners, and especially those who adopted smartphones more deeply, are more likely to snoop on others.

Summary of empirical work

Study 1: Selection of items for list experiment
(Google Consumer Surveys sample, n = 1,140)

Study 2: Trustworthiness of list experiments on MTurk
(MTurk sample, n = 434)

Study 3: Prevalence of snooping attacks, model of age and ownership
(MTurk sample, n = 1,381)

Study 4: Snooping and depth of adoption model
(MTurk sample, n = 653)

Before I conclude, just to give you a sense of what was involved in generating these findings, here's what we did.
We did a first study to select appropriate items for list question.
We then did an experiment I told you about to see if list experiments on MTurk actually are good at estimating behaviors for which we already know the prevalence.
We then did the two studies about snooping that I talked about.
In study 3 we obtained the estimate I showed you initially and also the first model.
And finally, with study 4 we created the last model I showed you.
So before the days of online tools like MTurk and Google Consumer Surveys having samples like this could cost you 10's or 100's of thousands of dollars, we were able to do all this for less than a 1000.

Main findings:

- Snooping attacks are common
- “Digital natives” more likely to snoop

To conclude, in a series of survey experiments, we have found that snooping attacks are very common, especially among some population segments.
For snooping attacks to be so common, it means that they must be relatively easy to do. Which for me means that we’re not providing users with what they need to secure their data.

Snooping on Mobile Phones: Prevalence and Trends

Diogo Marques	University of Lisbon	dmarques@di.fc.ul.pt
Ildar Muslukhov	University of British Columbia	ildarm@ece.ubc.ca
Tiago Guerreiro	University of Lisbon	tjvg@di.fc.ul.pt
Konstantin Beznosov	University of British Columbia	beznosov@ece.ubc.ca
Luís Carriço	University of Lisbon	lmc@di.fc.ul.pt