Answering Mobile Surveys With Images: *An Exploration Using a Computer Vision API*

Oriol J. Bosch | RECSM-UPF/ ISER- U. Essex

Melanie Revilla | RECSM-UPF

Ezequiel Paura | Netquest

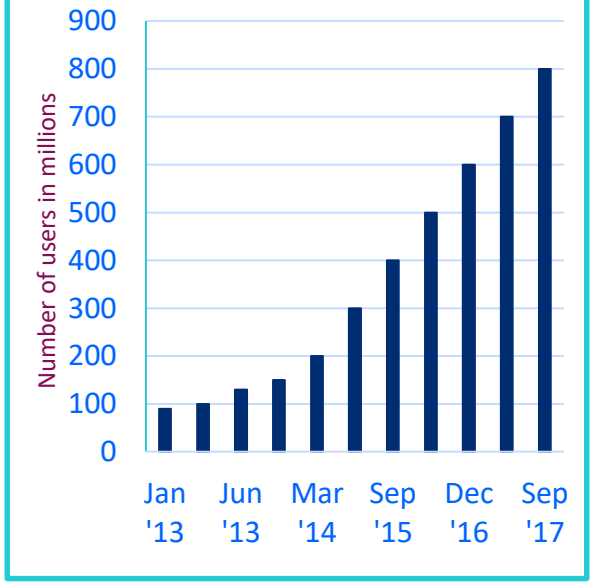
BACKGROUND

Posting and Sharing Images

Two opportunities

1. People familiar with the use of smartphone cameras and uploading interfaces (e.g. Instagram, Facebook).
    - ➢USA: 62% of the adult internet users post or share pictures on the Internet (Duggan, 2013)
    - ➢UK: 70% of the adult internet users engage on these activities (Dutton, Grant, & Groselj, 2013)

2. New tools that allow respondents to upload images.



Number of monthly active Instagram users 2013-2017

Source: Instagram

BACKGROUND

Using Images in Web Surveys

Several potential advantages of images in web surveys

- Richer information than what people can describe by text.

- Substitute factual questions distorted by **memory constraints**
  - i.e. “select all the items from this list that you have in your fridge.”

- Involve young participants
  - Used to new communicational approaches
  - Lower participation rates in web surveys (Bosch, Revilla and Paura, 2018)

BACKGROUND

Using Images in Web Surveys: Potential Limits

Stated willingness to participate

- Revilla et al. (2017): 48.2% of Netquest panelists in Spain reported willingness to share stored images and 54.9% to share photos of products
- Wenz et al. (2017): 65.0% respondents from Britain reported willingness to take and share photos

Processing and analyzing visual data

- How to process the amount of information of having +1000 images?

BACKGROUND

Computer Vision APIs

Use of machine to collect/analyze images and videos to extract information from visual data (Mulfari et al., 2016).

Examples of information we can extract from images:

- Face and landmark recognition
- Optical character recognition
- Tagging of explicit content

OUR STUDY

OUR STUDY

Goals

General goal

Investigate the feasibility of using images in the frame of mobile web surveys targeting Millennials

3 aspects considered

1) **Estimate** the proportion of respondents who:
   - Upload an image in line with the question
   - Upload an image not in line with the question
   - Did not understand how to upload an image
   - Skipped the question
2) **Investigate** which variables affect in which of the 4 behaviors a respondent is engaging
3) **Assess** the adequacy of using a computer vision API (e.g. Google Vision)

OUR STUDY

Data

Population of interest

Millennial panelists who have smartphone and use Internet more than once a week

Quotas

Cross quotas for country, age (16-34) and gender based on the general Internet population

Data collection

2^{th} – 19^{th} June, 2017

Netquest opt-in online panels

Final sample for analyses

1,614 respondents who completed the survey until the end: Spain = 808; Mexico = 806

OUR STUDY

Data

Population of interest

Millennial panelists who have smartphone and use Internet more than once a week

Quotas

Cross quotas for country, age (16-34) and gender based on the general Internet population

Data collection

2th – 19th June, 2017

Netquest opt-in online panels

Final sample for analyses

1,614 respondents who completed the survey until the end: Spain = 808; Mexico = 806

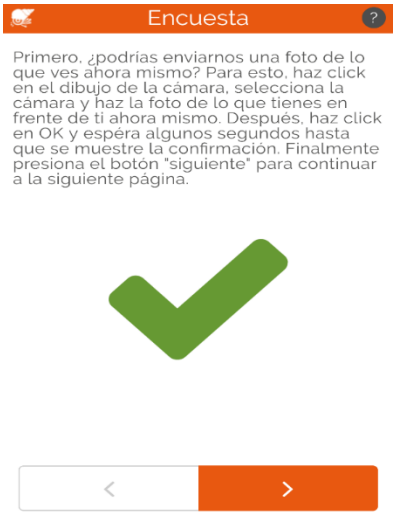
Survey

Total of 62 questions (skipping allowed)

Task	Request for an answer
T1. Taking a picture	**FOTO1.** Firstly, could you send us a photo of what you see right now?
T2. Uploading a stored image	**FOTO2.** Secondly, could you send us a photo of something that made you laugh and that you have already saved on your mobile phone?

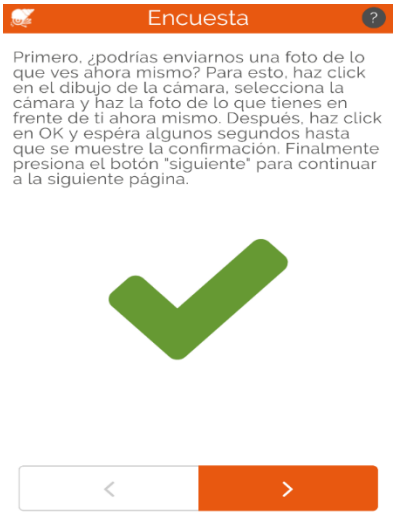
OUR STUDY

Design



OUR STUDY

Design



OUR STUDY

Codification approach

The approach

2 steps to compare the similitude of the codification of human coder (HC) and Google Vision (GV).

2-step procedure

1. The HC and GV generated a maximum of 5 tags (simple concept for each tag). The HC was introduced to the style of tags that GV produced but codified the images in a separate file.
2. A second HC not involved in step 1 analyzed the similitude of both codifications. Criterion: if both targeted the main concept represented in the image= similar.

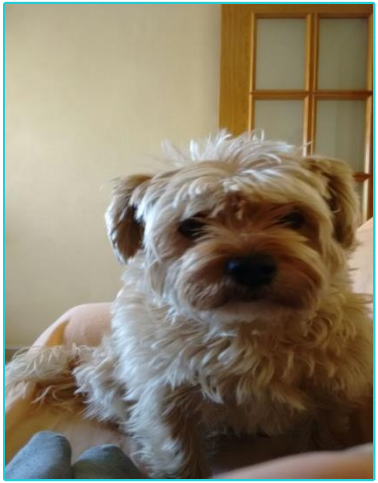
OUR STUDY

Codification approach

		Camera's Photo	Stored Image
Similar	HC	City; Street; Buildings	
	GV	Sky; City; Urban area; Landmark; Building	
Different	HC		
	GV		

OUR STUDY

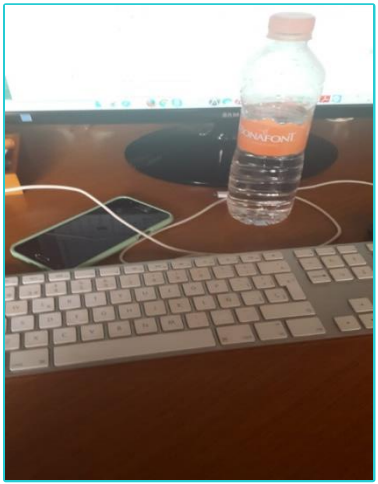
Codification approach



		Camera's Photo	Stored Image
Similar	HC	City; Street; Buildings	Dog
	GV	Sky; City; Urban area; Landmark; Building	Dog; Dog like mammal; Dog breed group; Morkie
Different	HC		
	GV		

OUR STUDY

Codification approach



		Camera's Photo	Stored Image
Similar	HC	City; Street; Buildings	Dog
	GV	Sky; City; Urban area; Landmark; Building	Dog; Dog like mammal; Dog breed group; Morkie
Different	HC	Keyboard; Screen; Smartphone; Water	
	GV	Bottle; Product design	

OUR STUDY

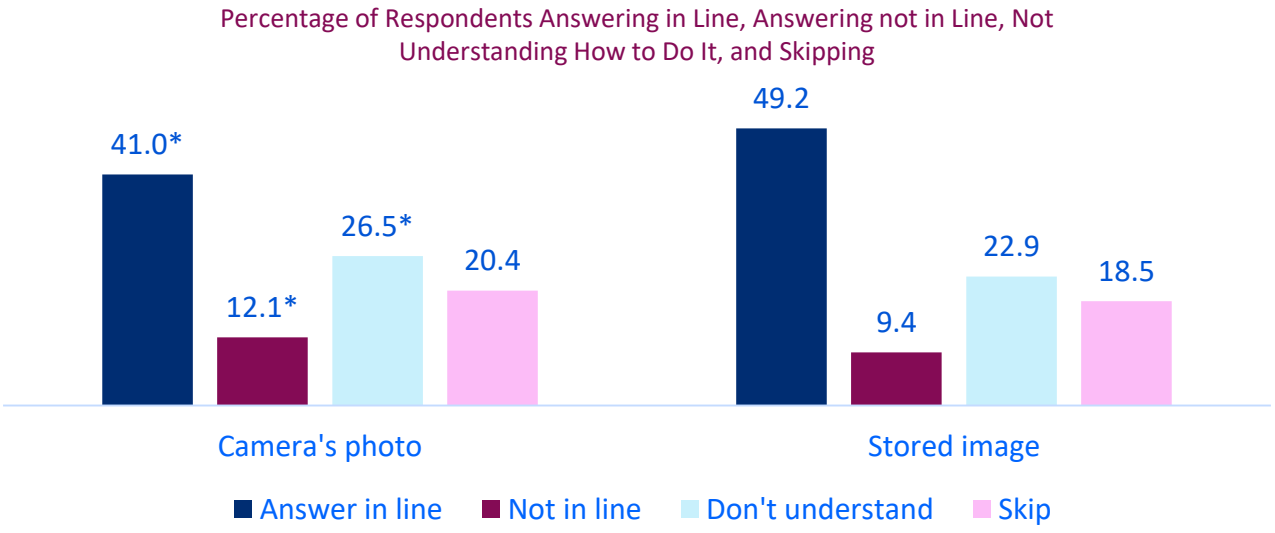
Codification approach

		Camera's Photo	Stored Image
Similar	HC	City; Street; Buildings	Dog
	GV	Sky; City; Urban area; Landmark; Building	Dog; Dog like mammal; Dog breed group; Morkie
Different	HC	Keyboard; Screen; Smartphone; Water	R2D2; Drugs; Joke; Text
	GV	Bottle; Product design	Product; Product; Technology; Product design; Plastic

MAIN RESULTS

MAIN RESULTS

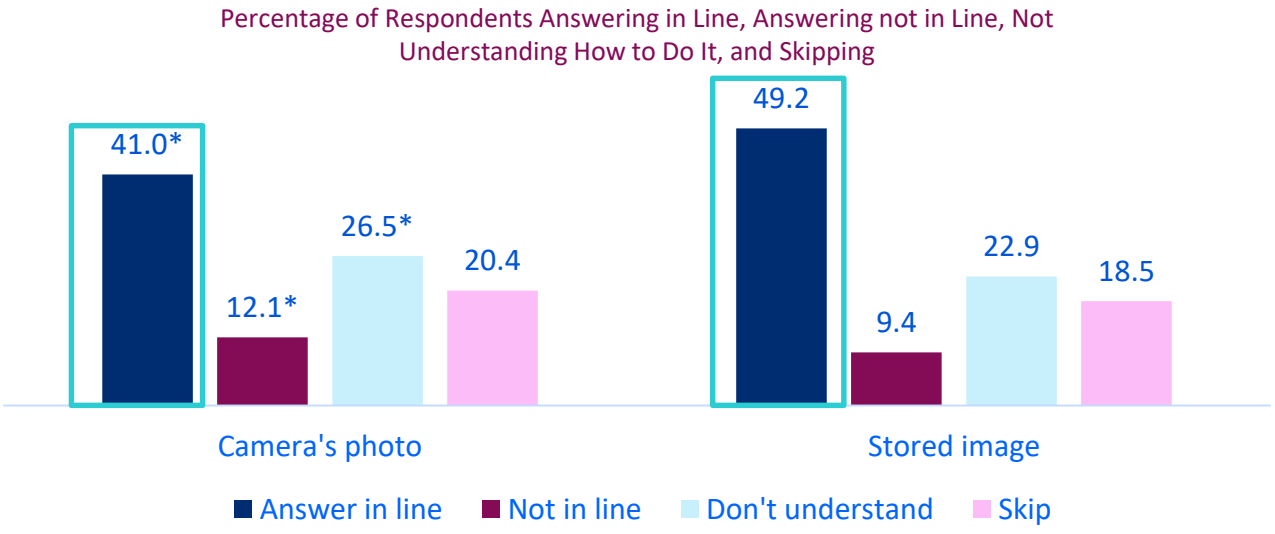
Evaluation of the performance of uploading pictures within the survey context



Note: indicates significant differences between tasks at p<.05*

MAIN RESULTS

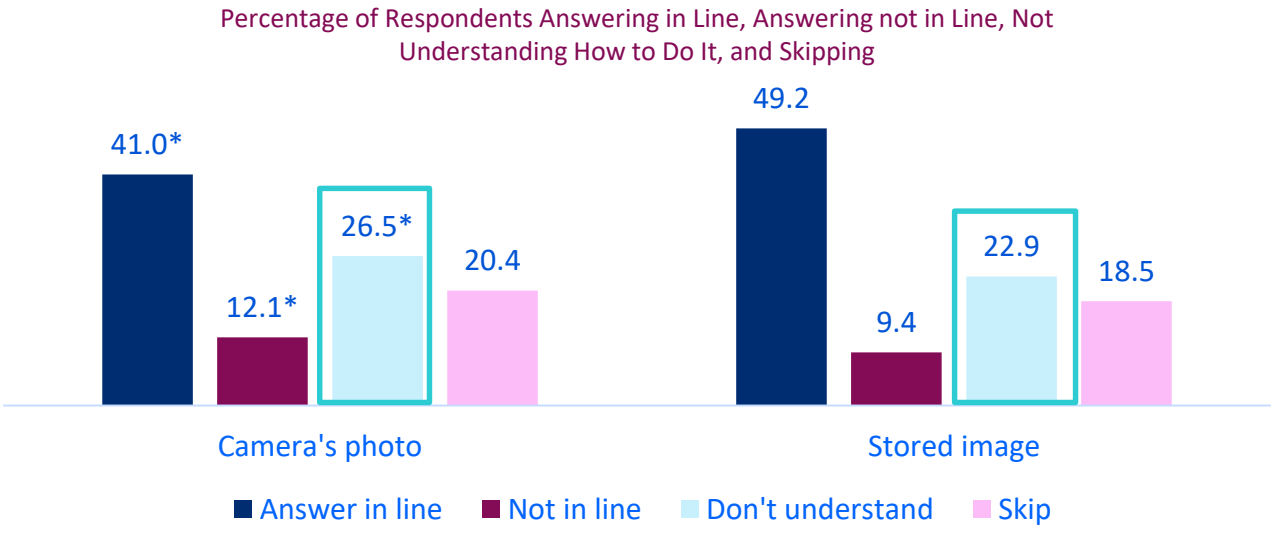
Evaluation of the performance of uploading pictures within the survey context



Note: indicates significant differences between tasks at p<.05*

MAIN RESULTS

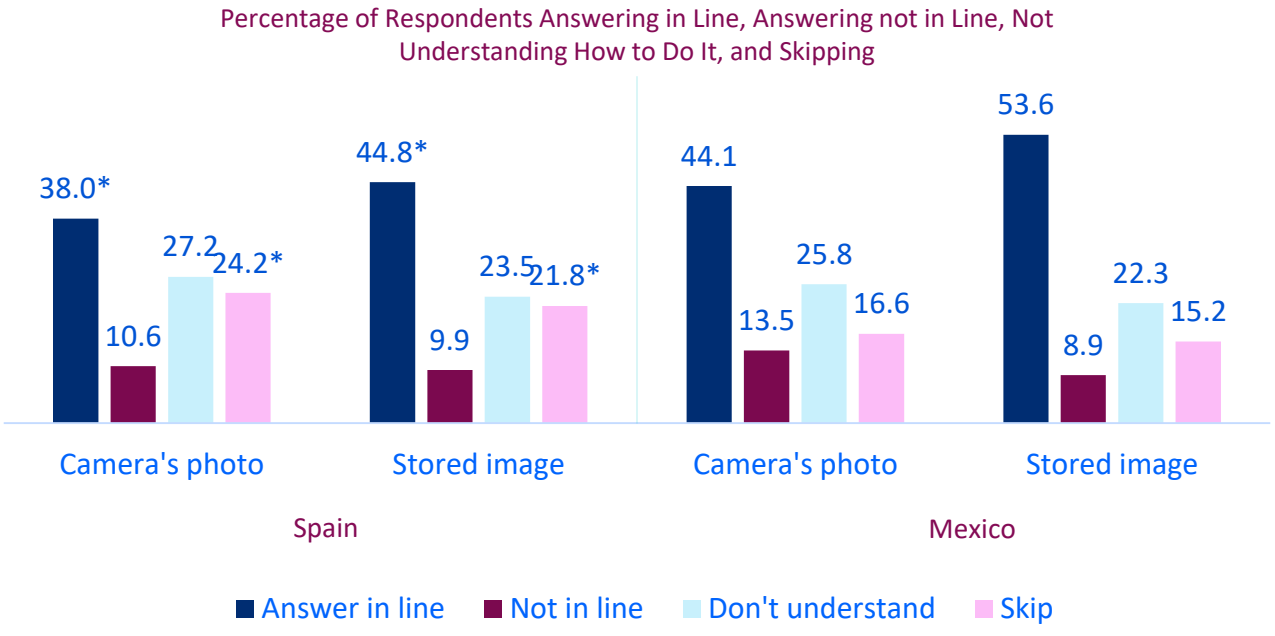
Evaluation of the performance of uploading pictures within the survey context



Note: indicates significant differences between tasks at p<.05*

MAIN RESULTS

Evaluation of the performance of uploading pictures within the survey context



Note: indicates significant differences between tasks at p<.05*

MAIN RESULTS

Determinants of Uploading an Image not in Line, not Understanding How to Do It, and Skipping, Relative to Answering in Line.

	T1: Taking and Sharing a Picture			T2: Uploading a Stored Image		
	Not in Line	Do Not Know	Skipping	Not in Line	Do Not Know	Skipping
Alone	1.36	1.15	1.00	0.86	1.14	1.02
Public	0.78	1.03	0.79	1.21	1.35*	0.90
Women	1.24	0.98	1.39*	0.81	0.94	1.17
Age	1.00	0.94	1.00	1.01	0.97*	1.03*
Mexico	1.21	0.92	0.81	0.78	0.92	0.75
Internet usage	0.92	0.97	0.94	0.96	0.94	0.95
Social media usage	0.86	0.85**	0.82**	0.96	0.90	0.83**
Extroversion	1.02	0.99	0.99	1.04	0.98	1.00
Creativity	0.96	0.99	0.99	0.90**	0.96	0.97
Laziness	0.97	0.99	0.98	1.00	1.03	1.00
Constant	0.25*	5.46**	0.88	0.56	2.08	0.44
Negelkerke R^2			0.05			
N			1,511			

*Note: Multinomial logistic regression. Coefficients expressed in relative risk ratios. * p<.05, **p<0.1*

MAIN RESULTS

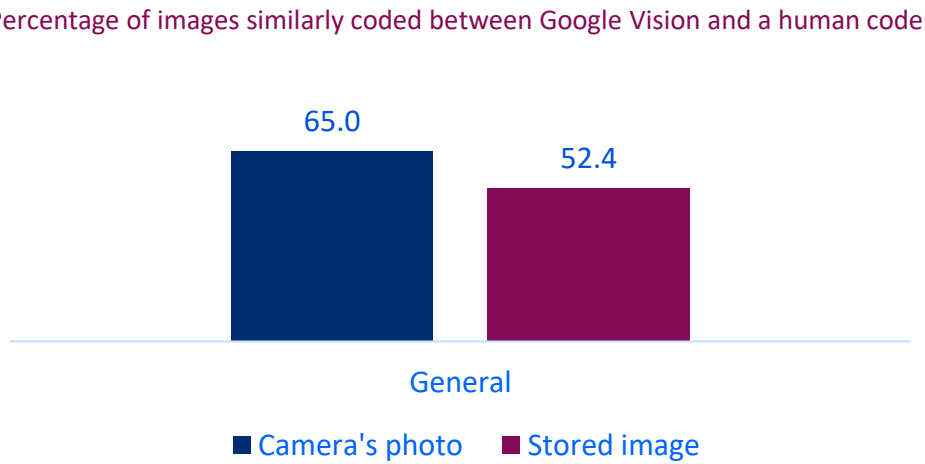
Determinants of Uploading an Image not in Line, not Understanding How to Do It, and Skipping, Relative to Answering in Line.

	T1: Taking and Sharing a Picture			T2: Uploading a Stored Image		
	Not in Line	Do Not Know	Skipping	Not in Line	Do Not Know	Skipping
Alone	1.36	1.15	1.00	0.86	1.14	1.02
Public	0.78	1.03	0.79	1.21	1.35*	0.90
Women	1.24	0.98	1.39*	0.81	0.94	1.17
Age	1.00	0.94	1.00	1.01	0.97*	1.03*
Mexico	1.21	0.92	0.81	0.78	0.92	0.75
Internet usage	0.92	0.97	0.94	0.96	0.94	0.95
Social media usage	0.86	0.85**	0.82**	0.96	0.90	0.83**
Extroversion	1.02	0.99	0.99	1.04	0.98	1.00
Creativity	0.96	0.99	0.99	0.90**	0.96	0.97
Laziness	0.97	0.99	0.98	1.00	1.03	1.00
Constant	0.25*	5.46**	0.88	0.56	2.08	0.44
Negelkerke R^2			0.05			
N			1,511			

*Note: Multinomial logistic regression. Coefficients expressed in relative risk ratios. * p<.05, **p<0.1*

MAIN RESULTS

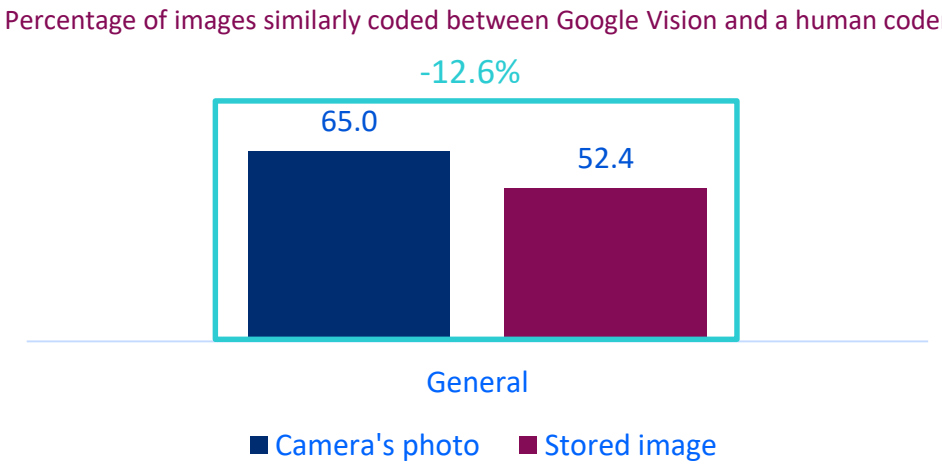
Evaluation of the similarity between the human and Google Vision codes



Note: The difference between tasks is significant ($p < .05$).

MAIN RESULTS

Evaluation of the similarity between the human and Google Vision codes



Note: The difference between tasks is significant ($p < .05$).

MAIN RESULTS

Evaluation of the effectiveness of each method

	Google Vision	Human	Difference
Time			
Total (In minutes)	< 5	2,100	-2,095 (41,900% faster)
Per image (In minutes)	<0.005	1.2	-1.195 (52,900% faster)
Cost			
Total (In euros)	2.50	553	-551.73 (21,671% cheaper)
Per image properly coded (In euros)	0.002	0.30	-0.2998 (14,880% cheaper)

Cost:

- GV: pack of 1,000 images= €1.27
  Analyzing 1,818= €2,50
- HC: Avg. hourly (gross) wage in Spain= €15.8.
  Multiplied by the time spent= €553

Images properly coded:

- GV: 1,059
- HC: 1,818

MAIN RESULTS

Substantive evaluation: combining survey answers with image labeling

Top 5 tags generated for each context by Google Vision and Human Coder

	Google Vision	Human
Answering from…		
Home (69.6%)	Technology; Electronic device; Display device; Floor; Room	Furniture; Laptop; TV Program; Room; House
Work (20.6%)	Technology; Electronic device; Floor; Display device; Product	Keyboard; PC; Office; Laptop; Furniture
Study centre (3.0%)	Technology; Text; Electronic device; Font; Display device	Laptop; Documents; Library; Classroom; PC
Bar/Restaurant (1.7%)	Technology; Electronic device; City; Food; Tree	TV Program; City; Bar; Food; Furniture
Public transport (2.23%)	Car; Sky; Vehicle; Light; Mode of transport	Car; Street; Road; Train; Window
Other (2.89%)	Cup; Architecture; Sky; Wall; Tree	Sky; Building; Street; Legs; Tree

CONCLUSIONS

CONCLUSIONS

Main Results

Around half of the panelists uploaded an image when asked

- 55% of the non-respondents were incapable of understanding how to upload an image. Satisficing?
- 40-50% sending an image in line with the question
- Higher number of social networks used per week increases the probability of uploading an image.

Between 52 and 65% of the images similarly coded by the human coder and Google Vision

- Google Vision performed better when coding photos of objects, places or people than when faced with complex images as cartoonish vignettes and memes
- Even if Google Vision is cheaper and faster, codes have to be processed and interpreted, increasing final costs.

CONCLUSIONS

Limits

- Millennials
- Opt-in panel
- Focused only on the labeling option, a combination of options (e.g. labeling+text recognition) could gave better results
- Subjective codification approach. Single coder.

CONCLUSIONS

Next steps

Improve response rates

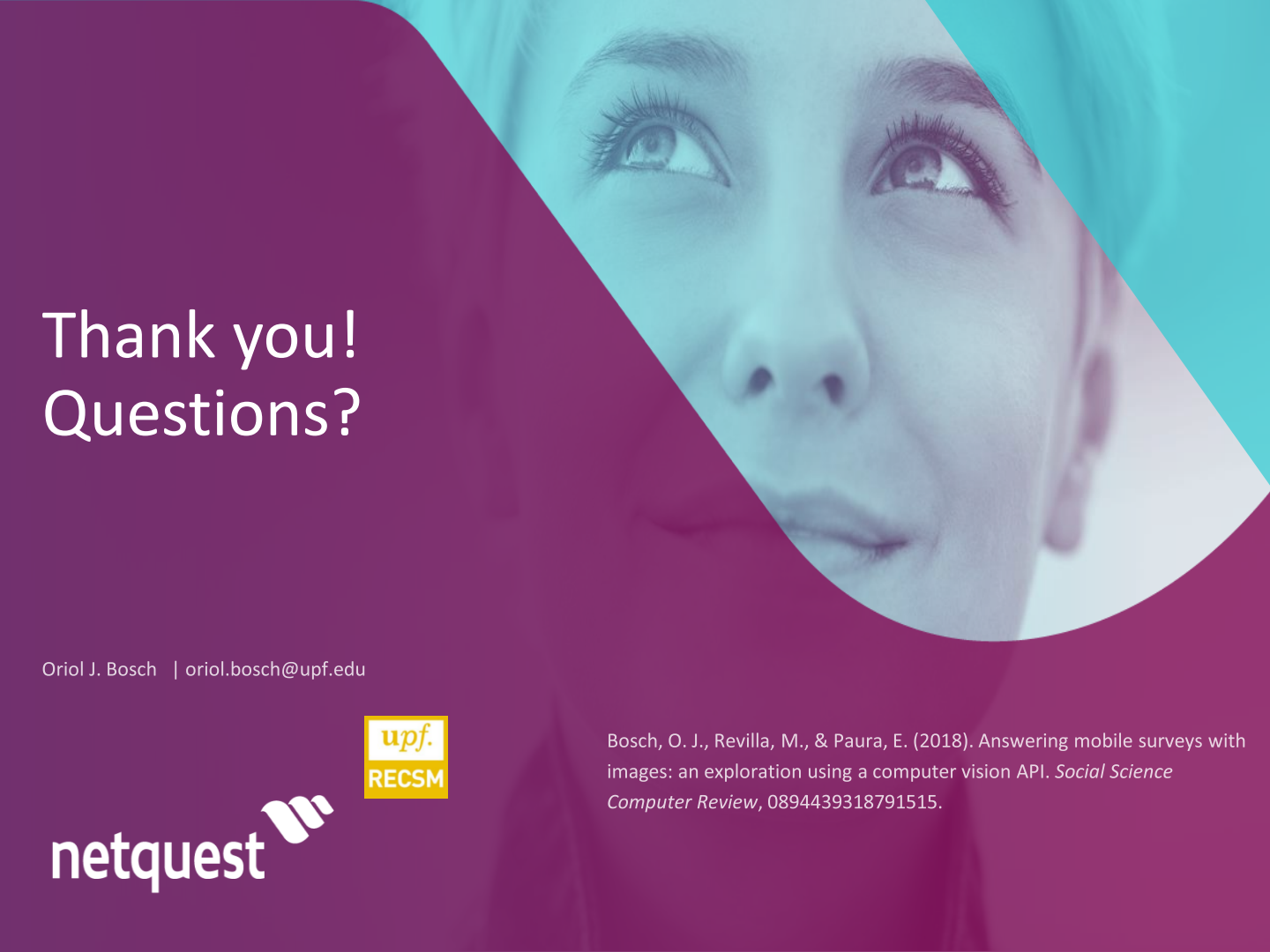
- Compare response rates for this new approach with traditional text answer
- Test ways to improve response rates: better interfaces and instructions, motivational messages, etc.

Data quality, respondent's behaviour and survey evaluation

- Compare these aspects with traditional text answer
- Test with which approach gives richer and more insightful information

API

- Assess computer vision API's performance when facing other types of images and analyses, such as sentiment or text analysis



Thank you!
Questions?

Oriol J. Bosch | oriol.bosch@upf.edu

upf.
RECSM

netquest

Bosch, O. J., Revilla, M., & Paura, E. (2018). Answering mobile surveys with images: an exploration using a computer vision API. *Social Science Computer Review*, 0894439318791515.

OUR STUDY

Data

Population of interest

Millennial panelists who have smartphone and use Internet more than once a week

Quotas

Cross quotas for country, age (16-34) and gender based on the general Internet population

Data collection

2th – 19th June, 2017
Netquest opt-in online panels

Final sample for analyses

1,614 respondents who completed the survey until the end: Spain = 808; Mexico = 806

% of those who started: Spain= 66.4; Mexico= 59.7

% of those who answered the first main survey question: Spain= 97.3 ; Mexico= 97.7

Survey

- Total of 62 questions (skipping allowed)
- Focused on a set of two questions that asked respondents to upload two types of images
- Both questions were placed at the end of the questionnaire to avoid break-offs. Limit= respondents more tired at the end
-We added the response option “I do not understand how to do it”

Task	Request for an answer
T1. Taking a picture	**FOTO1.** Firstly, could you send us a photo of what you see right now?
T2. Uploading a stored image	**FOTO2.** Secondly, could you send us a photo of something that made you laugh and that you have already saved on your mobile phone?

NEW STUDY

Collaboration with GIP

- Survey experiment within a survey conducted using Respondi.
- Both PC and Mobile devices allowed
- Experimental design to test: 1) the impact of answering with images on data quality, nonresponse, completion time and survey evaluation; 2) the impact of showing a motivational message on these aspects when answering with images
- Differences between PC and smartphone
- Differences between uploading already saved images or taking pictures
- Reasons behind nonresponse
- Etc

APPENDIX

Differences between respondents and non-respondents

	Camera's photo		Stored image	
Sociodemographic variables	Respondent	Non-respondent	Respondent	Non-respondent
Male	1.50	1.50	1.50	1.49
	(.02)	(.02)	(.02)	(.02)
Age	25.94**	25.24	25.53	25.78
(Continuous scale)	(.19)	(.19)	(.17)	(.20)
Internet use	1.03	1.06	1.04*	1.06
(6 categories: higher-lower)	(0.01)	(0.01)	(.01)	(.01)
Urbanity	2.20	2.26	2.21	2.24
(5 categories: Urban-Rural)	(0.04)	(0.04)	(.04)	(.04)

Note: the stars in column 'Respondent' indicates when differences between means are statistically significant between groups; * p > 0.05 ** p > 0.01

APPENDIX

Practical advice

Five-steps procedure

1. Code the whole set of images using Google Vision.
2. Generate a small random sample of the whole set and code it manually.
3. Compare the similarity of the codif and try to understand the logic behind Google Vision's codification.
4. Interpret the whole codification generated by Google Vision.
5. Acknowledge that the results are based on an interpretation of a Google Vision's codification and report the similarity rate of the random sample.