Going beyond conventional web surveys:
Opportunities and challenges of using new types of data within the frame of web surveys

HERSS Summer School

Hannover, 11 September 2024

Melanie Revilla | IBEI

Acknowledgments:
This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 849165). I want to thank Oriol Bosch, Patricia Iglesias, and Carlos Ochoa for all their helpful feedback and inputs.

Why is there an interest in using new types of data?

Importance of (web) surveys

- Most frequently used method for collecting data in many disciplines
  - Sociology: 69.7% of the published articles use survey data
  - Political sciences: 41.9% (Saris & Gallhofer, 2007)
- Web surveys: more and more common nowadays
  - 35% spent on research using (mobile) web, *vs* 11% for telephone and 8% for face-to-face (ESOMAR, 2019)
  - With pandemic, switch from other modes to web mode even quicker
- Results potentially used by key actors to take decisions

Problem: surveys suffer from errors

- Both on representation and measurement sides (TSE framework)

Final sample ≠ target population →

Lot of surveys (especially web) use nonprobability-based samples

Even when probability-based sampling is used, response rates have been going down drastically in most countries → possible selection bias in who participate

Weighting can sometimes be used but is often not sufficient

Problem: surveys suffer from errors

- Both on representation and measurement sides (TSE framework)

To err is human

People do not know everything surveys ask about

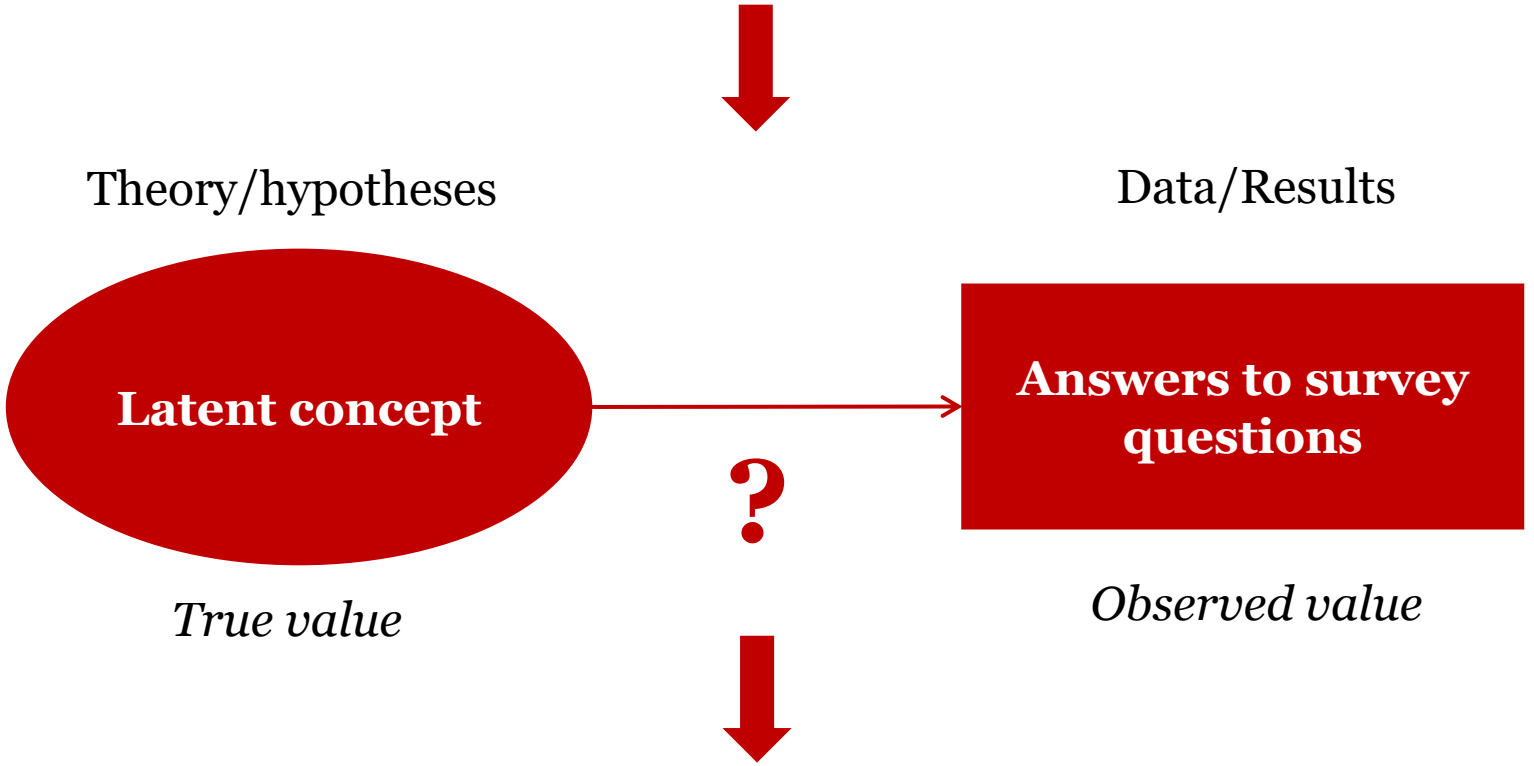
Remembering-self ≠ experiencing-self (Kahneman & Riis, 2005)

Lack of effort / satisficing

Social desirability

Problem: surveys suffer from errors

- Both on representation and measurement sides (TSE framework)



Average **measurement quality** for 67 ESS questions across up to 41 country-language groups (Poses et al. 2021) **= 0.65**

Measurement errors in surveys

- This gap between observed and true values can affect the results substantially

- Crucial to consider measurement errors

Table 6: Estimates of the parameters with and without correction

	Without correction On Allow immigration	With correction On Allow immigration
By		
Better life	-.265*	-.609*
Economic threat	-.133*	.001
Cultural threat	-.154*	-.140*
Total explained (R^2)	.254	.547

→ Wrong conclusions

Source: Saris & Revilla, 2016

Overall, need to improve quality of (web) survey data

web data opp

- But… How?
  - Need for improvement has been clear for decades
  - Lot of knowledge already on survey errors
    - How to reduce + correct for them (see e.g., the work of Willem Saris or Duane Alwin)
    - Lot also known about web surveys (e.g., Couper 2008; Tourangeau et al. 2013)
  - But still large errors, especially on the measurement side
  - **What else can we do?**

How could we enhance or extend web survey data?

Main idea that we will discuss

Taking advantage of **new measurement opportunities linked mainly to the growing use of smartphones** to reduce measurement errors in web surveys

Main idea that we will discuss

web data opp

Taking advantage of **new measurement opportunities linked mainly to the growing use of smartphones** to reduce measurement errors in web surveys

Smartphones are **everywhere**

More people have smartphones than toilets worldwide[1] 🤭

Including in **web surveys**

On average, Millennials answer 79% of the surveys using smartphones and Boomers 36% (US Netquest panel 2017/2018; Bosch et al., 2019)

→ **Create both new challenges and new opportunities**

[1] https://www.globalcitizen.org/en/content/access-denied-toilets-Harpic-Waterorg-RB/

New opportunities

- Opportunities at different levels
  - Phone number: can be used to **contact** respondents
    - E.g., through SMS, sending a link to the survey, allows random digit dialing (RDD)
  - People take/use their phones everywhere
    - Possible to contact them **anytime/anyplace**
    - Invitations can be seen very quickly after being sent
  - Possibility to collect **new data types**
    - Sensors + apps → different types of data

Which new types of data?

- Data collected **passively**
  - Participants only need to accept to share such data and/or set up a tracking app
  - Mainly **digital traces data** = records of activity undertaken through an online information system, including digital footprints left behind by users as they interact with technology (Howison et al., 2011)
    - Ex: browsing history, search queries, social media interactions, GPS data, app usage…
  - Allows studying online behaviors, traveling patterns, etc.

Most of those data can also be collected for PC!

- Data collected **actively**
  - Respondents need to actively provide these data

HOW COULD WE ENHANCE?

New data types considered in the WEB DATA OPP project

VISUAL DATA

Screenshots
Photos/videos taken during the survey
Visual files saved on (or accessible from) the device

VOICE DATA

Dictation
Voice recording

4 new types of data

METERED DATA

Obtained through a tracking application ("meter") installed by the participants on their devices to register at least the URLs of the webpages visited. Usually collected in metered panels.

GEOLOCATION DATA

Obtained through a tracking application installed on participants' mobile devices to register at least the GPS coordinates

IN-THE-MOMENT SURVEYS triggered by such data

web data opp

How could these new data types help?

Benefits expected only for some concepts, not all!

HOW COULD THE NEW DATA TYPES HELP?

Expected **benefits** (Revilla, 2022)

Researchers

- Reduce some of the issues related to measurement errors

Information people do not know

[https://www.mosquitoalert.com/en/]

Social desirability bias

	EUR
Banana	1,44 A
1,148 kg x 1,25 EUR/kg	
Freshona/Espinacas	1,15 A
Vemondo/Bebida soja 0%	1,60 B
2 x 0,80	
Vemondo/Tofu ecológico	0,95 B
Mandarina Ebre	2,79 A
Dentalux/Crema dental	0,95 C
Chef Select/Trio de humm	2,19 B
Edulis/Ensalada dúo	1,15 A
Alesto/Mezcla frutos sec	1,89 B
Floralys/Servill 2capas	0,95 C
Favorina/Huevos chocolat	1,49 B
Champiñón	0,65 A
Huevos L suelo	1,79 A
Floralys/Papel higiénico	2,55 C
Total	21,54

Mistakes and satisficing

HOW COULD THE NEW DATA TYPES HELP?

Expected **benefits** (Revilla, 2022)

Researchers

- Reduce some of the issues related to measurement errors
- Provide data for new concepts (not measured so far)
- Massive amount of data / granular data
- Real time / continuous (passive data)
- Answer new research questions

Nutrition Facts

Chicken with Mushroom Gravy

Serving Size: 1 Serving (328g)

Amount Per Serving	
Calories 398	Calories from Fat 155
	% Daily Value*
Total Fat 17g	**26%**
Saturated Fat 7.8g	**39%**
Trans Fat 0.4g	
Polyunsaturated Fat 6g	
Monounsaturated Fat 1.8g	
Cholesterol 152mg	**51%**
Sodium 730mg	**30%**
Potassium 569mg	**16%**
Total Carbohydrates 8.5g	**3%**
Dietary Fiber 0.9g	**4%**
Sugars 0.7g	
Protein 50g	
Vitamin A	6.3%
Vitamin C	2.3%
Calcium	1%
Iron	15%

* Percent Daily Values are based on a 2000 calorie diet.

HOW COULD THE NEW DATA TYPES HELP?

Expected **benefits** (Revilla, 2022)

Researchers

- Reduce some of the issues related to measurement errors
- Provide data for new concepts (not measured so far)
- Massive amount of data / granular data
- Real time / continuous (passive data)
- Answer new research questions

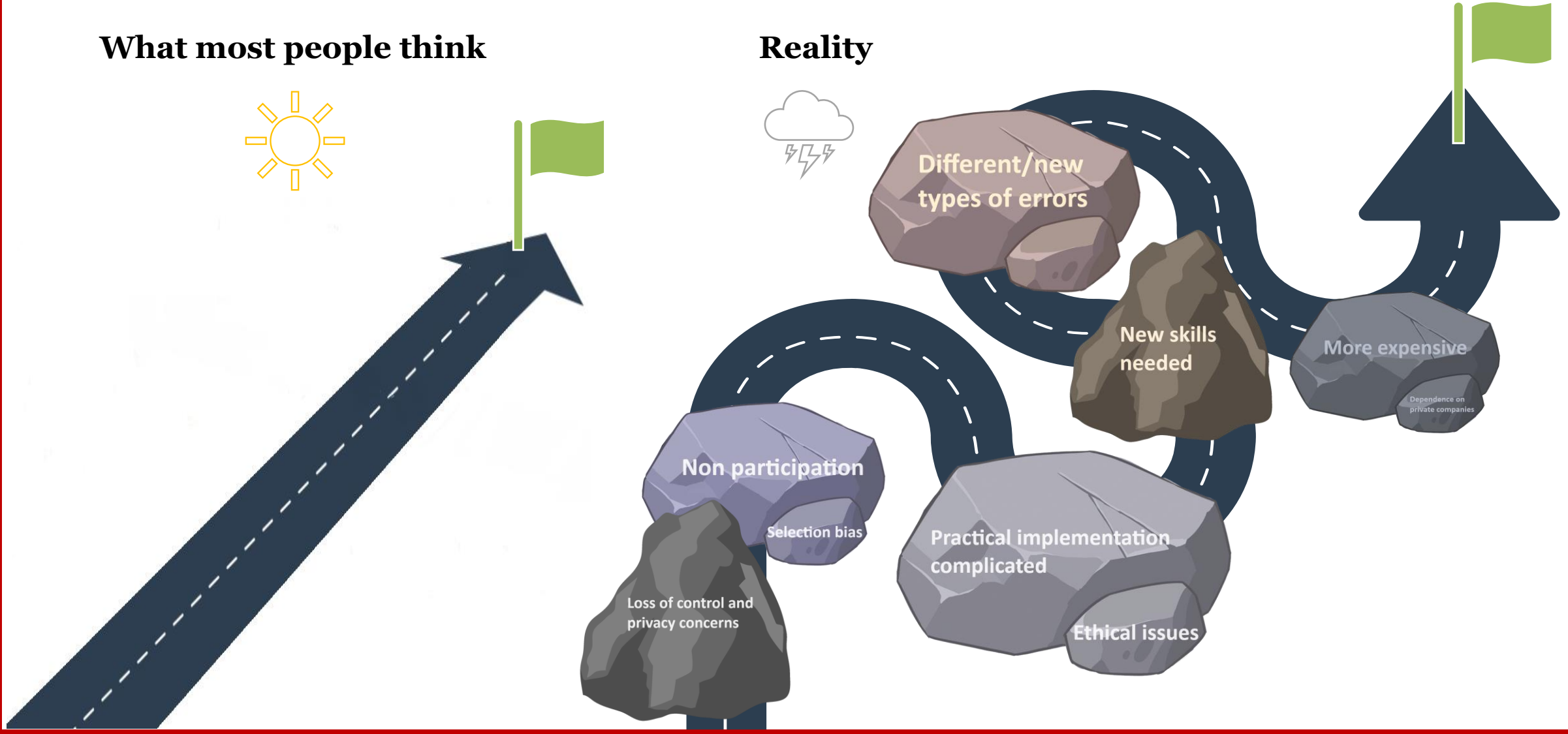
Participants

- Reduce time dedicated to provide information
- Reduce efforts
- More enjoyable

But this is not that easy…

What most people think

Reality



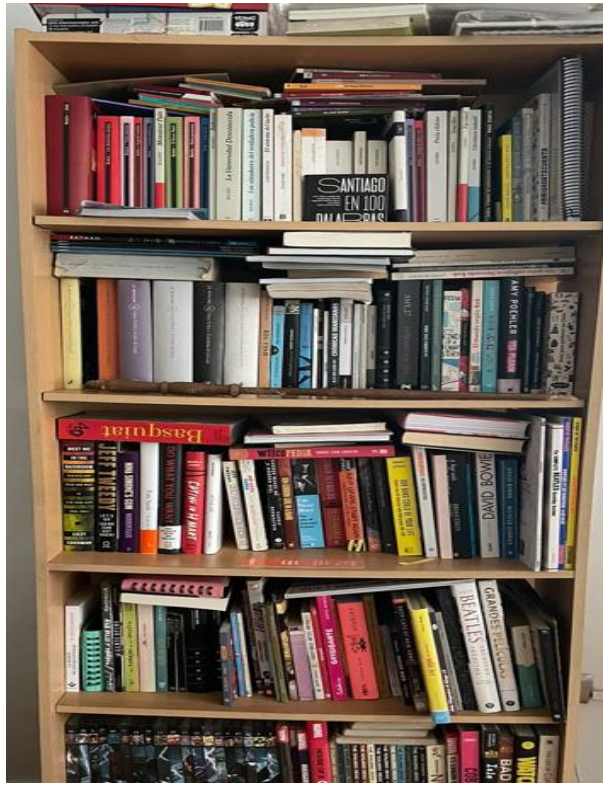
Examples of research with the new types of data

EXAMPLE 1: COLLECTING PHOTOS OF THE BOOKS PEOPLE HAVE AT HOME (IGLESIAS ET AL., 2023)

Potential benefits of asking for photos of the books at home

- Number of books often used as indicator of cultural or economic capital
- But people do not know how many books they have
- Social desirability bias expected → over-reporting
- Kind of books also matter (cooking vs history books)
- Asking for photos of the books has the potential to:
  - Provide more accurate information about the number of books
  - While also providing extra information (kind of books, language, storage, etc.)

A *picture* is **worth** a **thousand** words



EXAMPLE 1: COLLECTING PHOTOS OF THE BOOKS PEOPLE HAVE AT HOME (IGLESIAS ET AL., 2023)

Lot of challenges

How can the information be extracted from the images?

Will participants share the images?

How can the images be stored safely?

Which tool can be used to collect the photos?

EXAMPLE 1: COLLECTING PHOTOS OF THE BOOKS PEOPLE HAVE AT HOME (IGLESIAS ET AL., 2023)

Lot of challenges

How can the information be extracted from the images?

We developed a practical guide (Iglesias et al., 2024)

Will participants share the images?

We studied different factors that can affect participation (Iglesias & Revilla, 2024)

How can the images be stored safely?

Which tool can be used to collect the photos?

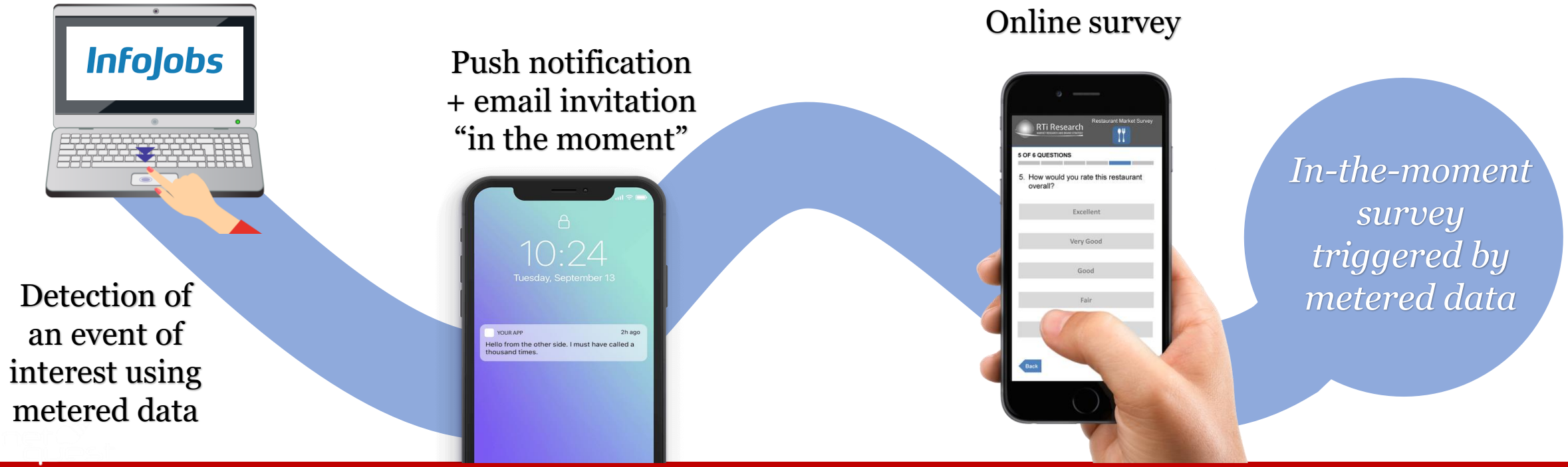
We created the WebdataVisual tool (Revilla et al., 2022a)

Main learnings from the books-at-home experiment

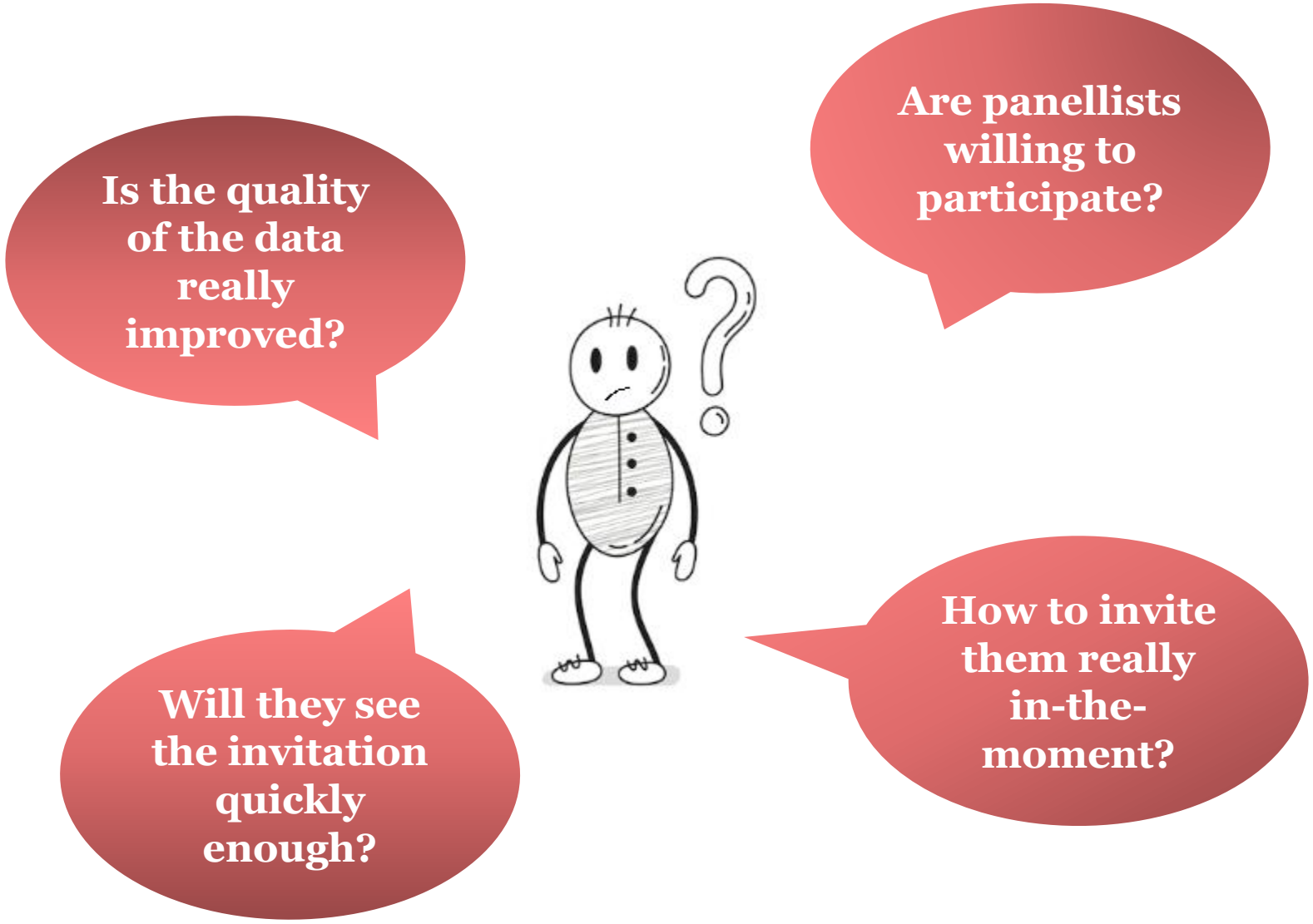
- Participation was lower than expected
  - 66% out of 703 participants asked for photos did not send a photo or left the survey
- Classification was very challenging
  - Problem of overlap
  - Problem to differentiate books from CDs, DVDs…
  - Problem to differentiate between illiterate and literate children's books
- Extra-information was available but not systematically
  - Languages of the books, titles, authors, etc, were visible for some books but not all
  - So difficult to use that in quantitative analyses
- At this day, photos can complement more than replace

Potential benefits of using an in-the-moment survey

- Reduce the **time** between an event of interest and the questions about this event
- Potential for decreasing **recall errors**, thereby enhancing data quality
- Potential new insights as well



Lot of challenges



EXAMPLE 2: STUDYING JOB APPLICATION THROUGH AN IN-THE-MOMENT SURVEY (OCHOA, 2023)

Lot of challenges

We implemented an experiment (Ochoa, 2023)

Is the quality of the data really improved?

Are panellists willing to participate?

We studied willingness to participate (Ochoa & Revilla, 2022a; Ochoa 2022)

Will they see the invitation quickly enough?

We studied invitation methods (Ochoa & Revilla, 2022b)

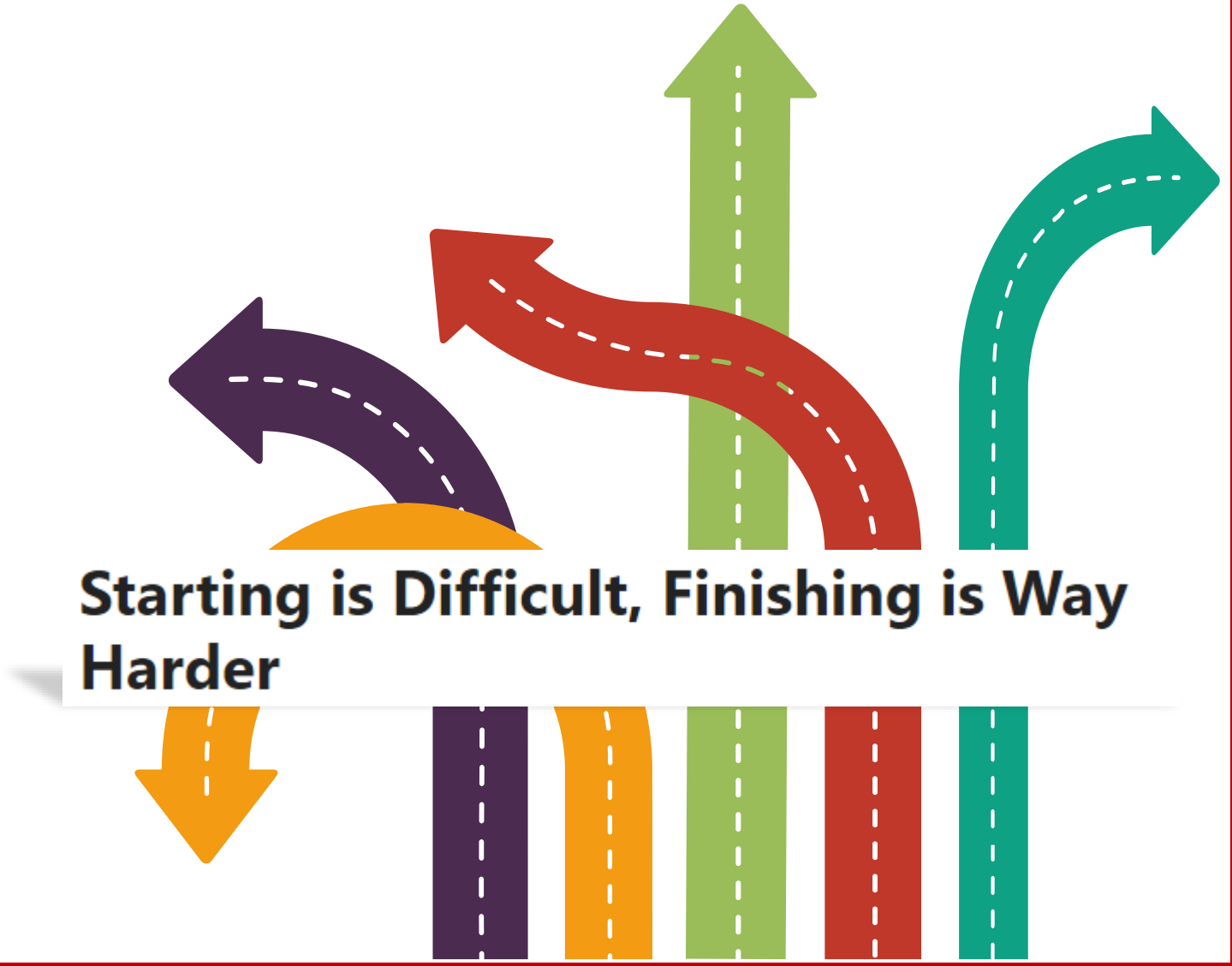
How to invite them really in-the-moment?

We created the WebdataNow tool (Revilla et al., 2022b)

Main learnings from the job search experiment

- In-the-moment surveys can be implemented: we now have the necessary panels and tools to do so
- However, data collection is more complex and take much more time
- Participants are satisfied with their experience (no issues with intrusiveness)
- Fewer improvements in data quality indicators than expected
  - Slight improvement in % DK
  - Some improvement in length of answers to open-ended questions
- Differences in substantive results that suggest people might not chose to answer "DK" even if they do not recall (properly)

Conclusions

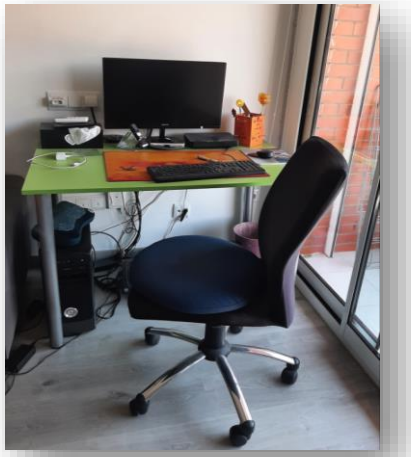


CONCLUSIONS

Increasing interest in new data types

- Issues in conventional surveys
  - Decreasing participation + poor data quality in measures of many concepts
- Push researchers to consider **new data types**
  - Could reduce some types of errors + provide new/more detailed data
- Potentially **broad applications** and new insights

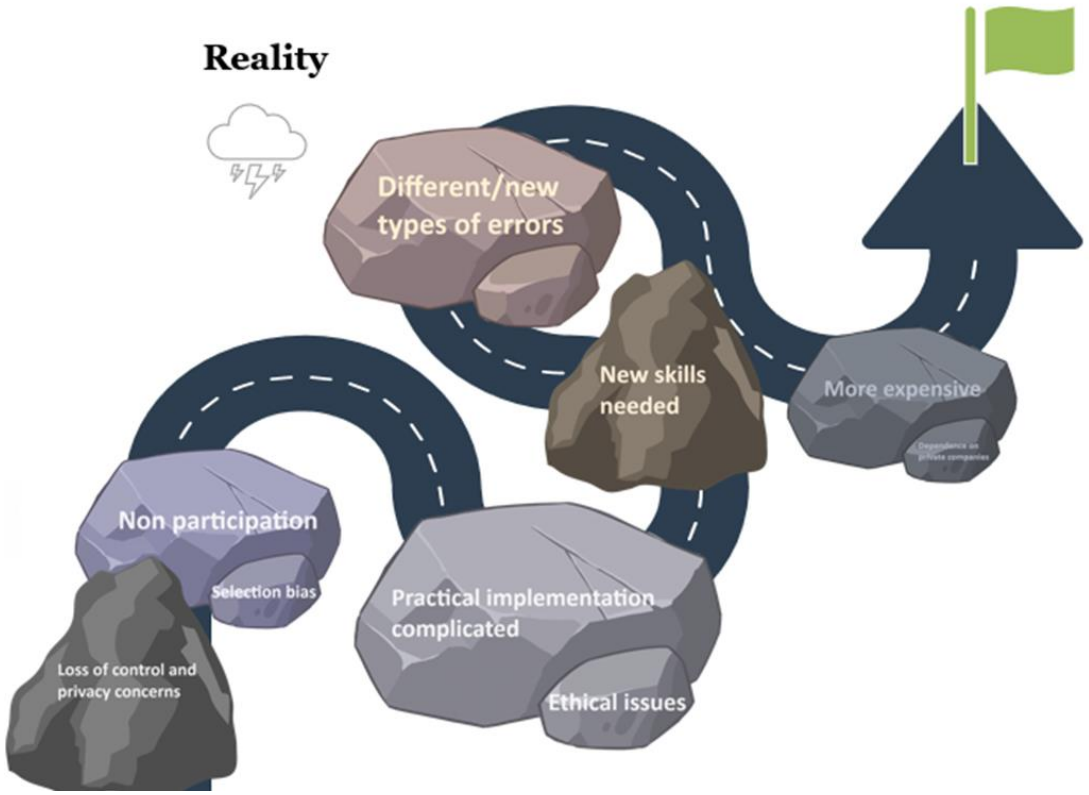




CONCLUSIONS

But this is a complicated road

Remember…

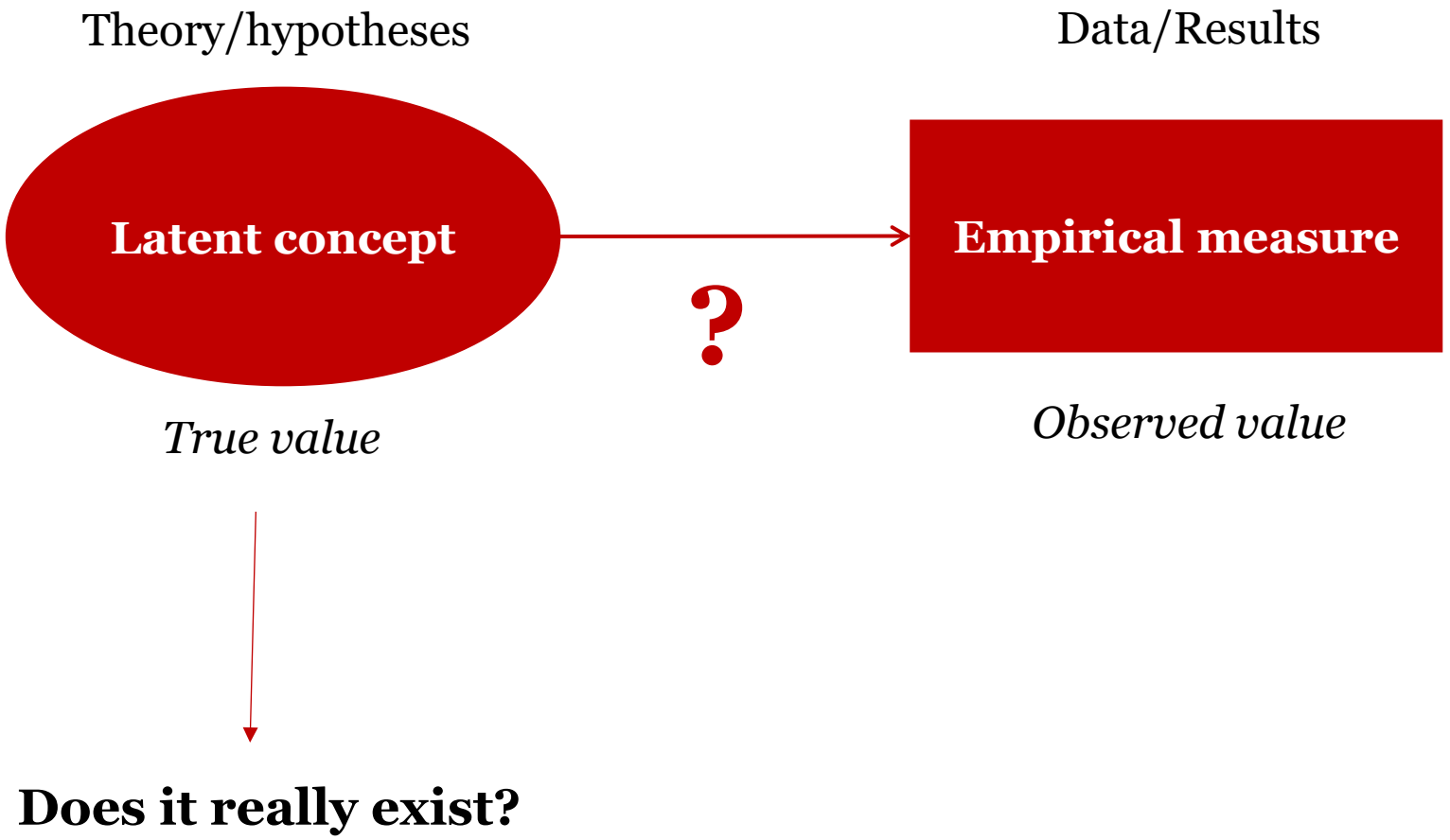


Lot of further research needed

We need to get **similar level of knowledge for the new data types** as we have for conventional questions

Necessary to make informed decisions about their use + to improve data quality

Even if we improve data quality, errors will remain…



CONCLUSIONS

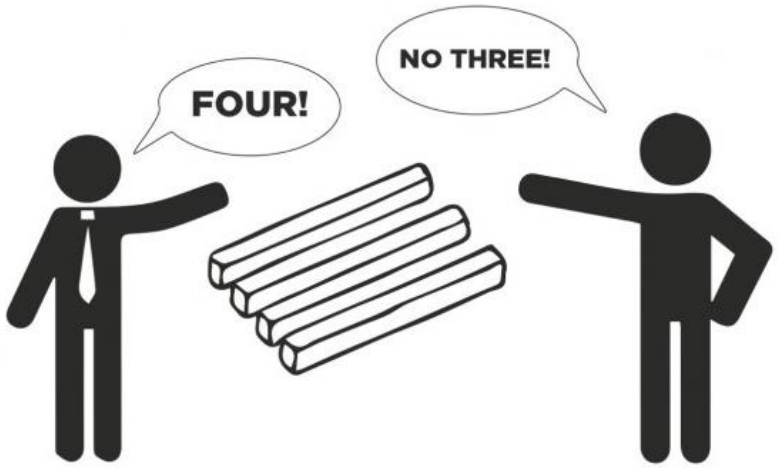
Does it really exist?

https://www.colorado.edu/asmagazine/2023/11/27/loch-ness-monster-myth-or-reality

CONCLUSIONS

Look from different perspectives

Different types of data provide **different but complementary information**



Combine *several* types of data!

Thanks!

Questions?

Melanie Revilla | IBEI

mrevilla@ibei.org

https://www.upf.edu/web/webdataopp

INSTITUT BARCELONA ESTUDIS INTERNACIONALS

web data opp

References

- Bosch, O.J., Revilla, M., & Paura , E. (2019). Do Millennials differ in terms of survey participation? *International Journal of Market Research,* 61(4): 359-365. https://doi.org/10.1177/1470785318815567
- Couper, M. (2008). Designing Effective Web Surveys, Cambridge: Cambridge University Press.
- ESOMAR (2019). Global Market Research Report. Amsterdam: ESOMAR. ISBN: 978-90-903-2259-9 . Retrieved from (June 2021): https://www.esomar.org/knowledge-center/library?publication=2926
- Howison, J., Wiggins, A., & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12(12), 767–797. https://doi.org/10.17705/1jais.00282
- Iglesias, P.A., Revilla, M., Heppt, B., Volodina, A., & Lechner, C. (2023). Protocol for a web survey experiment studying the feasibility of asking respondents to capture and submit photos of the books they have at home and the resulting data quality [version 1; peer review: 2 approved, 1 approved with reservations]. Open Res Europe 2023, **3**:202 (https://doi.org/10.12688/openreseurope.16507.1)
- Iglesias, P.A., Ochoa, C, & Revilla, M. (2024). A practical guide to (successfully) collect and process images through online surveys. *Social Sciences & Humanities Open*, 9, 100792. https://doi.org/10.1016/j.ssaho.2023.100792
- Iglesias, P., & Revilla., M. (2024). Skills, availability, willingness, expected participation and burden of sharing visual data within the frame of web surveys. *Quality and Quantity,* 58, 1071–1092. https://doi.org/10.1007/s11135-023-01670-3
- Kahneman, D., & Riis, J. (2005). "Living, and thinking about it: Two perspectives on life". In F. Huppert, B. Keverne, & N. Baylis (Eds.), *The science of well-being*. Oxford, England: Oxford University Press.
- Ochoa, C. (2023). Researching the moment of truth. An experiment comparing in-the-moment and conventional surveys to investigate online job applications. ESRA conference, 20th July 2023. Presentation (preliminary results).

References

- Ochoa, C. (2022). Willingness to participate in geolocation-based research. PLoS ONE 17(12): e0278416. https://doi.org/10.1371/journal.pone.0278416
- Ochoa, C., Revilla, M. (2022a). Willingness to participate in in-the-moment surveys triggered by online behaviors. Behavior Research Methods (2022). https://doi.org/10.3758/s13428-022-01872-x
- Ochoa, C., Revilla, M. (2022b). Acceptance and coverage of fast invitation methods to in-the-moment surveys. International Journal of Market Research. https://doi.org/10.1177/14707853221085204 .
- Poses, C., Revilla, M., Asensio, M., Schwarz, H., and Weber, W. (2021). Measurement quality of 67 common social sciences questions across countries/languages: Results of 28 Multitrait-Multimethod experiments implemented in the European Social Survey. *Survey Research Methods*, 15(3): 235-256. https://doi.org/10.18148/srm/2021.v15i3.7816
- Revilla, M. (2022). How to enhance web survey data using metered, geolocation, visual and voice data?. *Survey Research Methods*, 16(1): 1-12. https://doi.org/10.18148/srm/2022.v16i1.8013.
- Revilla, M., Iglesias, P., Ochoa, C., & Antón, D. (2022a). WebdataVisual: a tool to gathervisual data within the frame of web surveys. OSF. http://doi.org/10.17605/OSF.IO/R7CAX
- Revilla, M., Ochoa, C., Iglesias, P. Antón, D. (2022b). WebdataNow: a tool to send in-the-moment surveys triggered by passive data. OSF. http://doi.org/10.17605/OSF.IO/G3MSC
- Saris, W. E., & Gallhofer, I. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. New York: Wiley
- Saris, W.E., & Revilla, M. (2016). Correction for measurement errors in survey research: necessary and possible. *Social Indicators Research*, 127(3): 1005-1020.
- Tourangeau, R., Conrad, F. & Couper, M.P. (2013). *The Science of Web Surveys*, Oxford: Oxford University Press.