Google

Using paradata to improve the quality of web surveys

Some examples and applications

Mario Callegaro,
Senior Survey Research Scientist
Google Ltd, London

Royal Statistical Society,
June 26, 2014

Today's agenda

Google

- Taxonomy of web survey paradata
- Examples of usage of paradata types
- The concept of adaptive scripting
- Longitudinal paradata
- Software to collect paradata
- Privacy issues with paradata
- Implication for research

Taxonomy of paradata types

Google

Paradata for web surveys can be classified into the following groups:

1. **Direct paradata**
    A. Contact-info
    B. Device-type paradata
    C. Questionnaire navigation paradata
2. **Indirect paradata**
    A. E.g. eye tracking, video recording, behavioral coding

1 A. Contact-info paradata

Similar to call-records for CATI or CAPI

For list samples with email invitations you can have these possible outcomes:

- Nothing back (probably delivered)
- Bounce back
- Temporary away (Out of office reply)
- Other undeliverable status

Useful to study nonresponse, quality of the list, eligibility…

Studies with time of the day and day of the week for sending email invitations and increase response rates

1 B. Device type: Browser used, OS type and OS language

Google

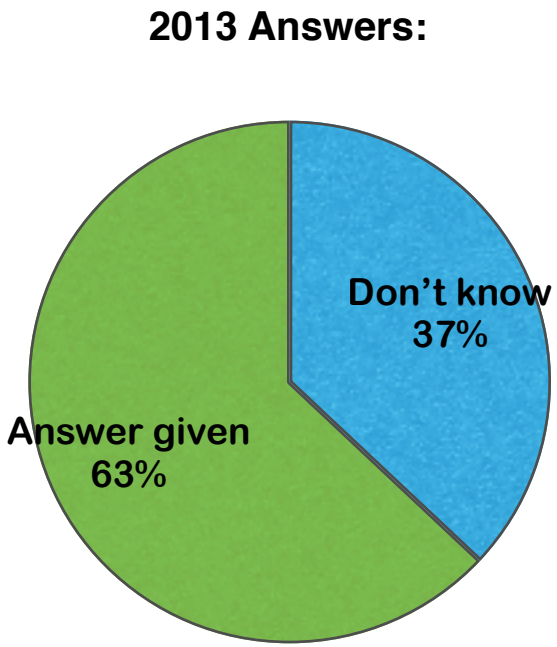
Every time a browser connects to a website, it sends a string of text called user agent string.
From that we can infer the device type associated with it, whether it is a:

- Desktop/laptop/netbook
- Smartphone
- Tablet
- Other devices (book reader, game console...)

```
Mozilla/5.0 (iPhone; CPU iPhone OS 5_0 like Mac OS X)
AppleWebKit/534.46 (KHTML, like Gecko) Version/5.1
Mobile/9A334 Safari/7534.48.3
```

2011 - 2013 proportion of participants taking a web survey from a smartphone

Meaning Ltd Survey of market research companies



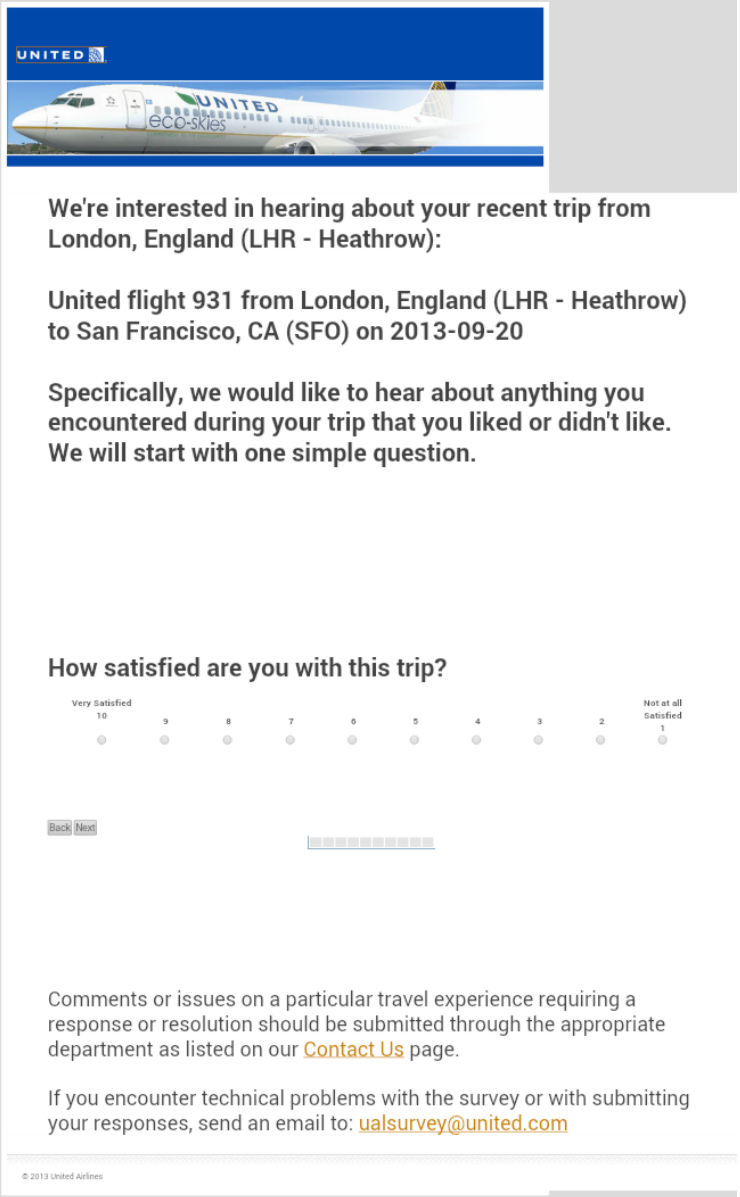
Data from Macer 2011-2013 Survey of market research companies global (N=240)

What happens when a survey is not optimized to be taken on a mobile device

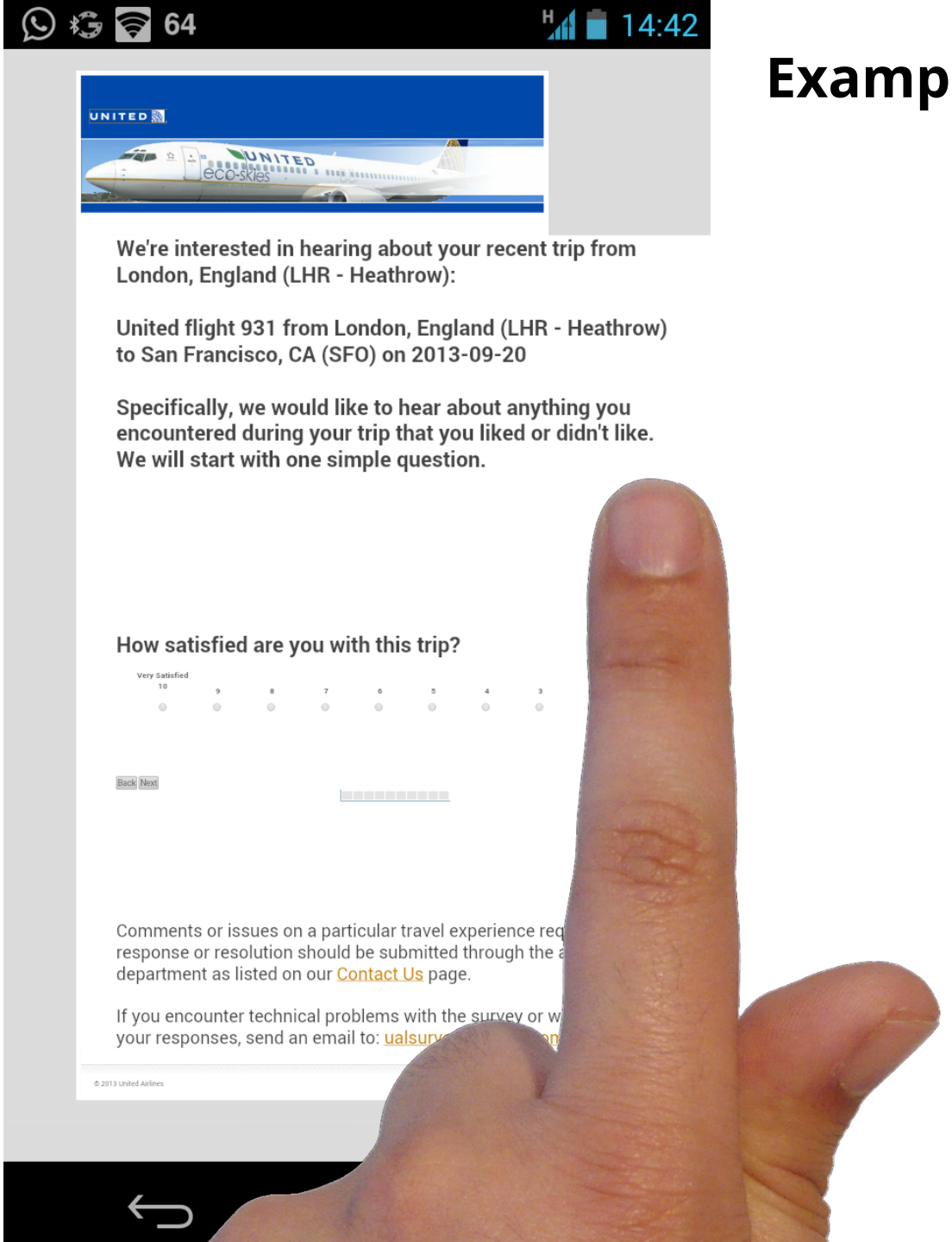
Google

- User agent strings can be used to measure the behavior of respondents taking the survey from different devices
- Completion, breakoff and partial interview rates can be computed by device
- Other survey quality criteria can be computed by device such as number of missing items
- When a survey is not optimized to be taken from a mobile devices it generates higher breakoffs

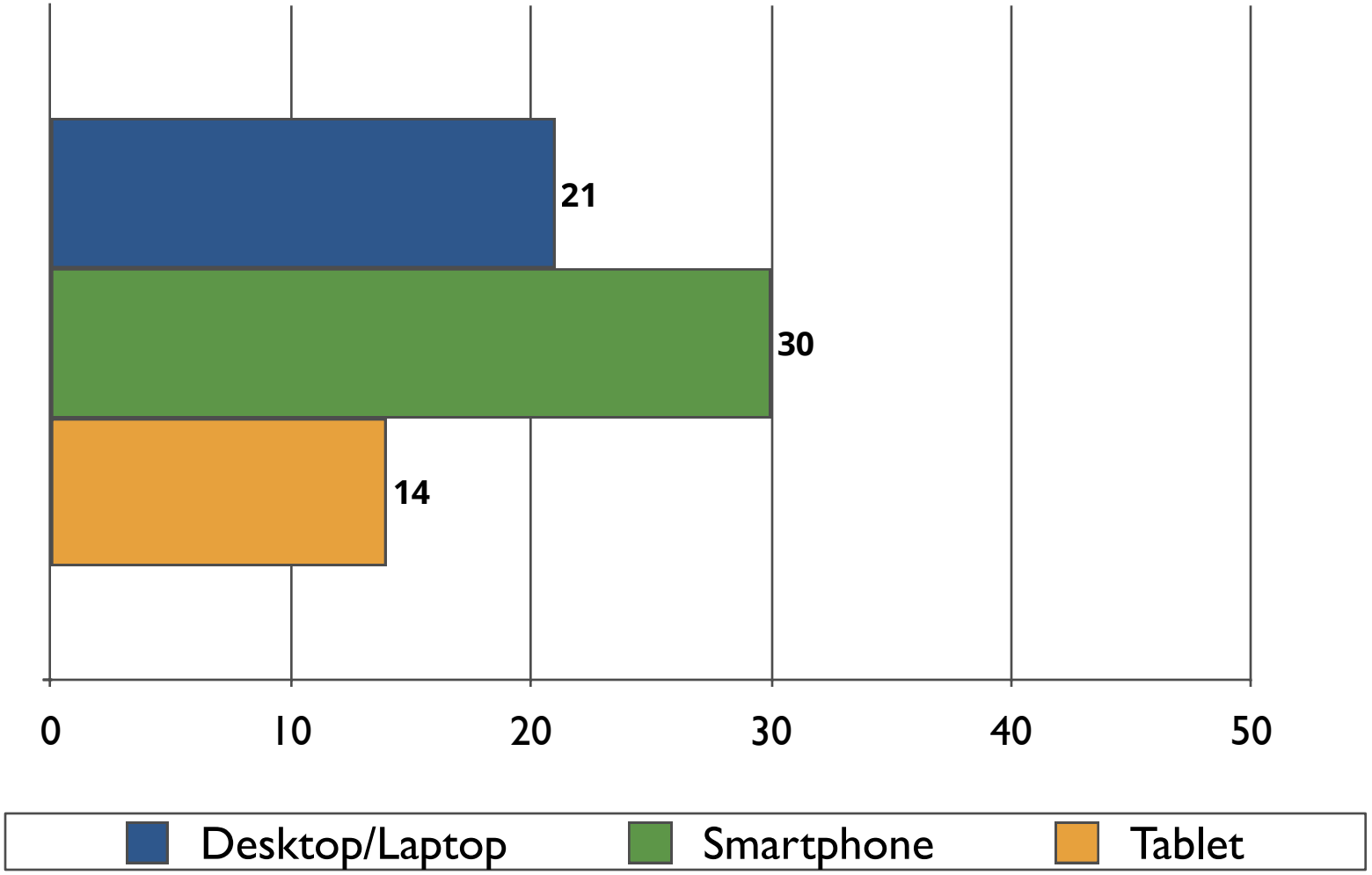
Example


64
14:42
UNITED
UNITED
eco-skies
We're interested in hearing about your recent trip from London, England (LHR - Heathrow):
United flight 931 from London, England (LHR - Heathrow) to San Francisco, CA (SFO) on 2013-09-20
Specifically, we would like to hear about anything you encountered during your trip that you liked or didn't like. We will start with one simple question.
How satisfied are you with this trip?
Very Satisfied
10
9
8
7
6
5
4
3
2
Not at all Satisfied
1
Back
Next
Comments or issues on a particular travel experience requiring a response or resolution should be submitted through the appropriate department as listed on our Contact Us page.
If you encounter technical problems with the survey or with submitting your responses, send an email to: ualsurvey@united.com
© 2013 United Airlines

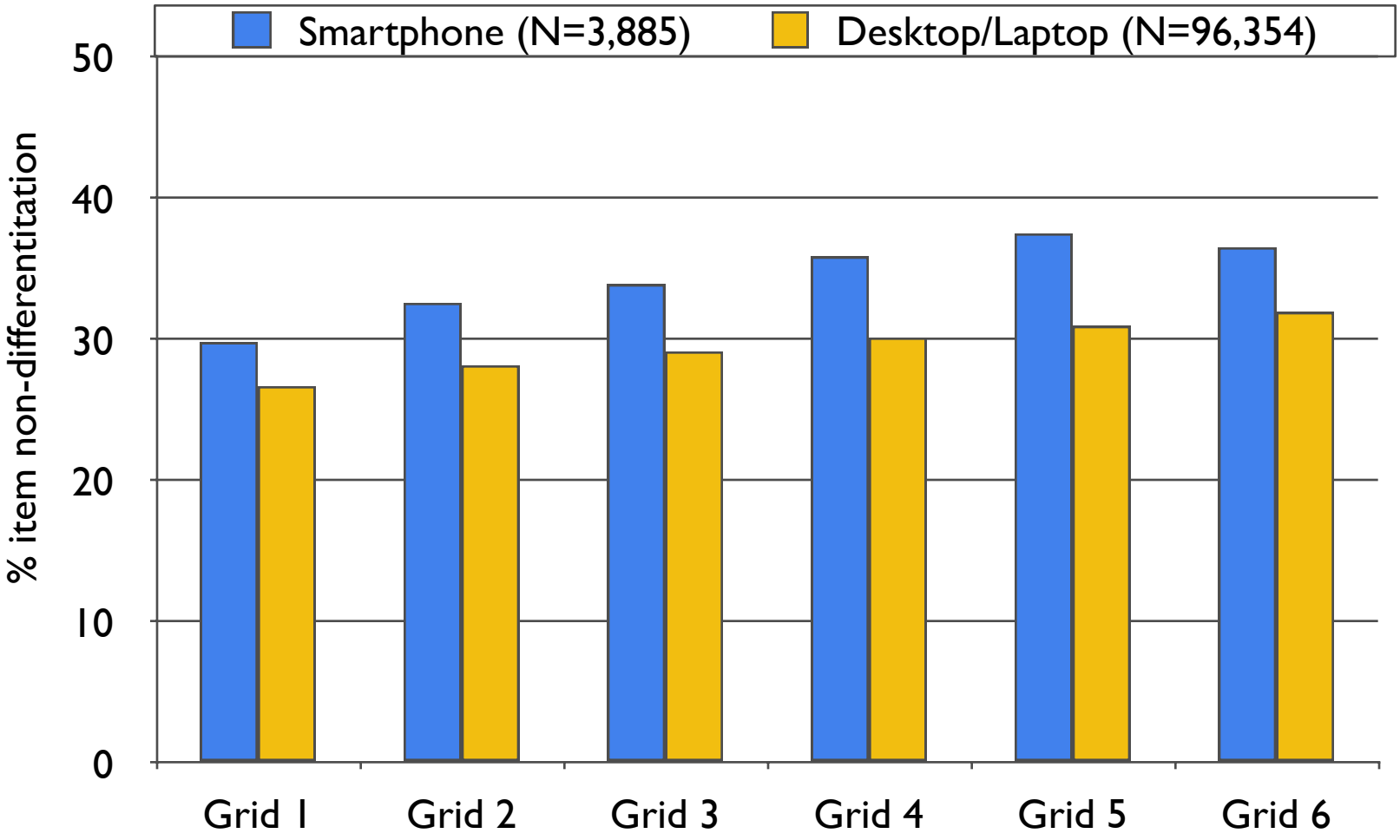
Example


64
14:42
UNITED
UNITED
eco-skies
We're interested in hearing about your recent trip from London, England (LHR - Heathrow):
United flight 931 from London, England (LHR - Heathrow) to San Francisco, CA (SFO) on 2013-09-20
Specifically, we would like to hear about anything you encountered during your trip that you liked or didn't like. We will start with one simple question.
How satisfied are you with this trip?
Very Satisfied
10
9
8
7
6
5
4
3
Back
Next
Comments or issues on a particular travel experience req
response or resolution should be submitted through the a
department as listed on our Contact Us page.
If you encounter technical problems with the survey or w
your responses, send an email to: ualsurv
© 2013 United Airlines

Survey suspend by device
Foundation of Data Quality 2, 2014



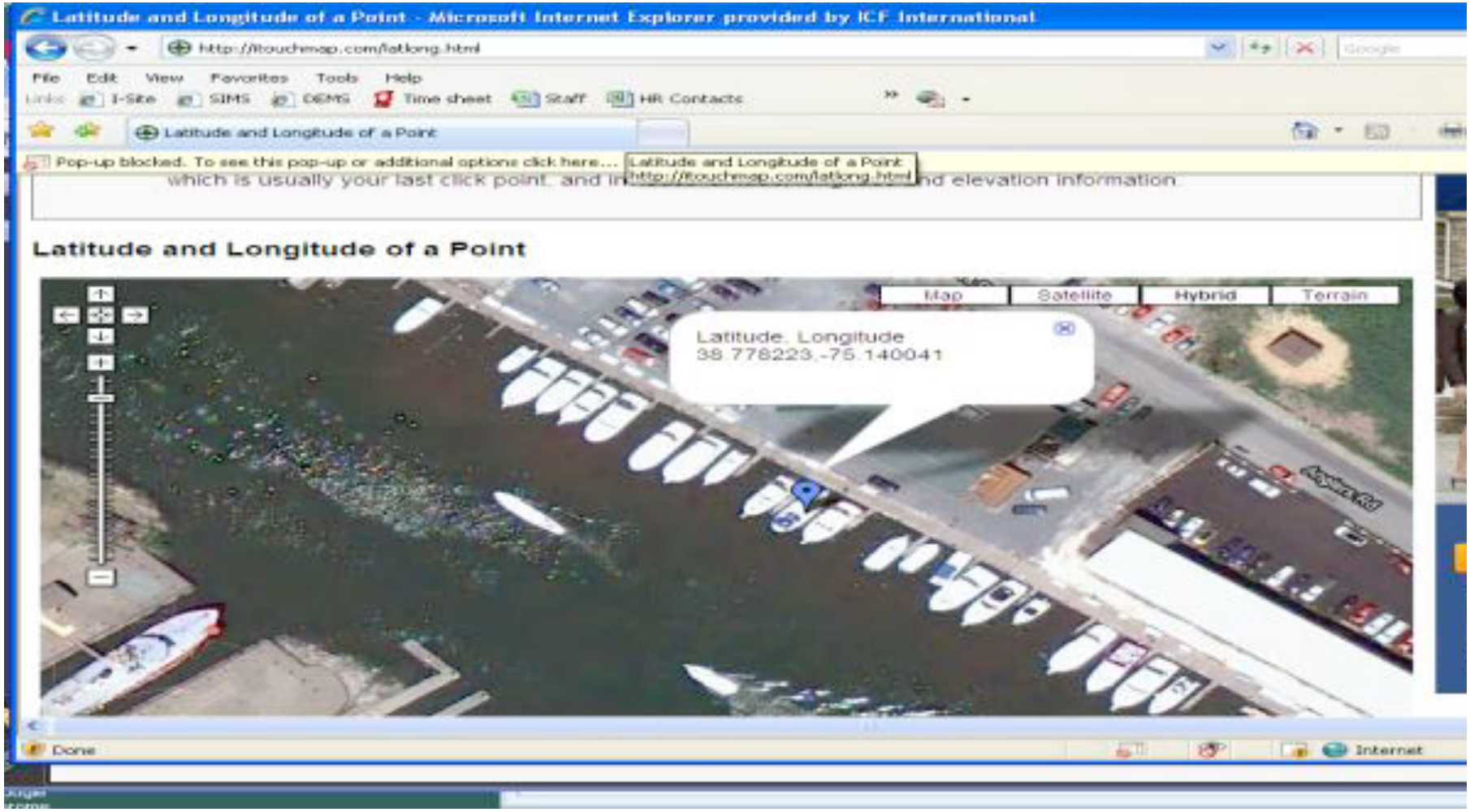
Item non-differentiation by device



All differences are statistically significant
Item non-differentiation = answered the same point for the whole grid (39 items in total)

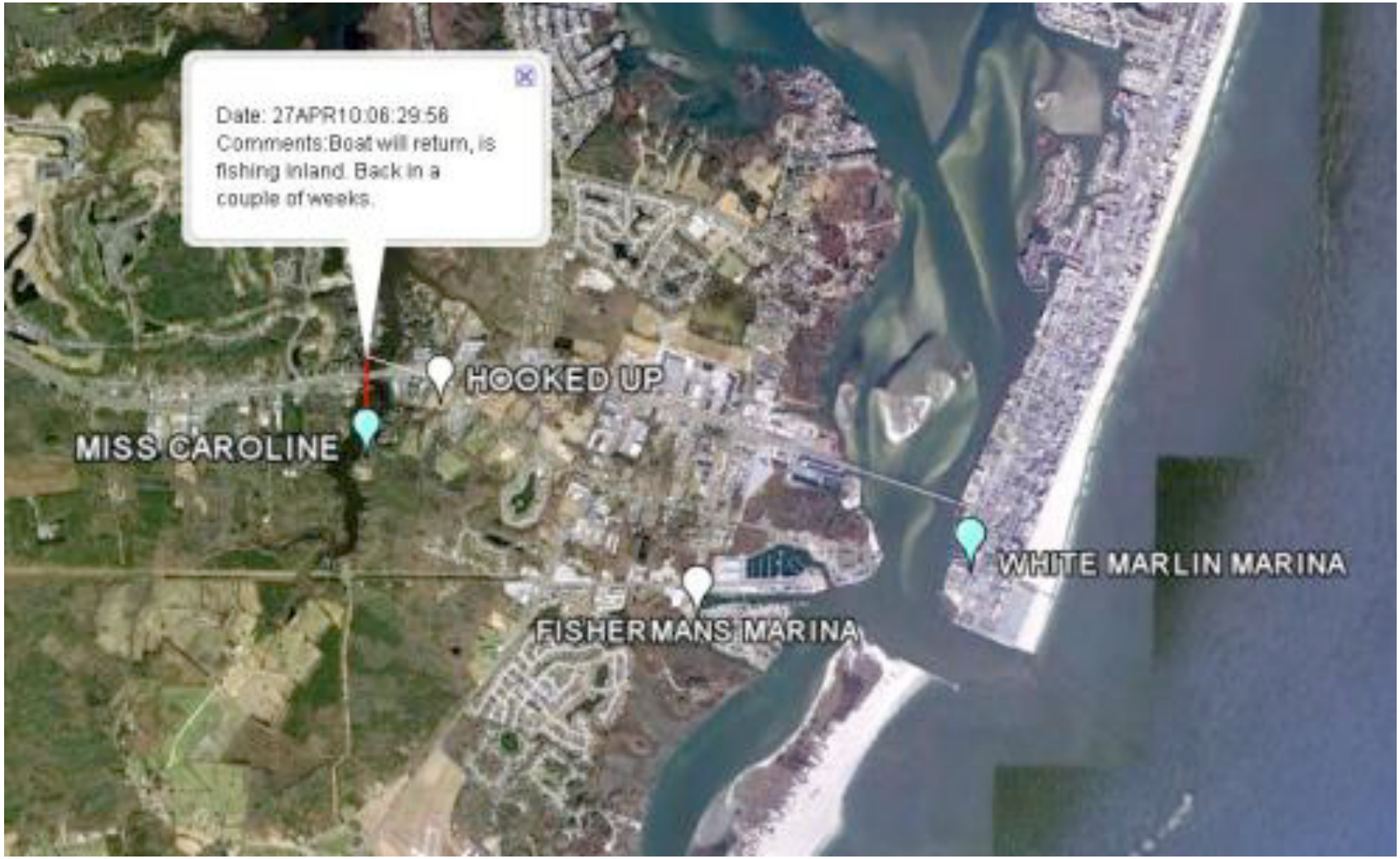
McClaine, Crawford and Dugan study of US college students (2012)

Device type: GPS coordinates example



Dayton, J & H. Driscoll: *The Next CAPI Evolution - Completing Web Surveys on Cell-Enabled iPads*. AAPOR 2011

Device type: GPS coordinates example (cont.)



Dayton, J & H. Driscoll: *The Next CAPI Evolution - Completing Web Surveys on Cell-Enabled iPads*. AAPOR 2011

Device type: Cookies

- Cookies are used on survey sites to help the researcher recognize the respondent as a prior user as well as for other survey control or quality functions
- Some companies use them to ensure that only one survey is done from the same device, for example
- The use of cookies in web surveys is, however, not very well documented
- ESOMAR Practical Guide on Cookies (July 2012) is a very good starting point in order to comply to the European e-privacy initiative (Cookie law) in terms of web surveys

ESOMAR PRACTICAL GUIDE ON COOKIES

Google

JULY 2012

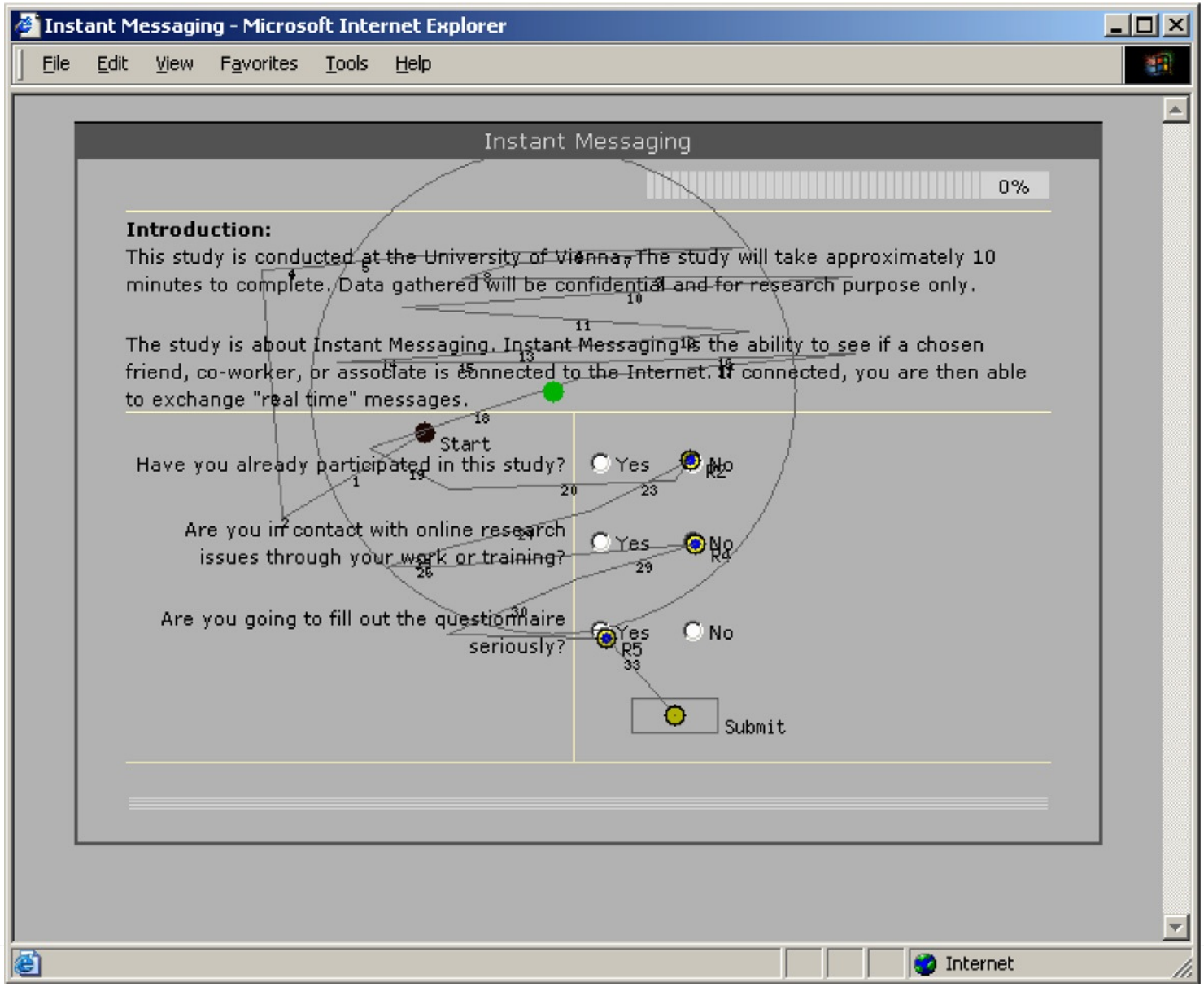
- Make an inventory of cookies used by your website/survey platform
- Identify the information stored in Cookies
- Check if some cookies are exempt:
  "for the sole purpose of carrying out the transmission of a communication over an electronic communication network, or as strictly necessary in order for the provider of an information society service explicitly requested by"
- Prepare a privacy policy web page explaining the users what cookies are used for
- Create a consent mechanism and opt-out mechanism

Questionnaire navigation paradata example

```
lXNtoilre7_2|1|M677|13|1320#
M548|174|830#
M160|101|1750#
M366|192|550#
M728|4|7690#
M489|247|610#
C493|229|3301#
R110|1#
C493|280|4301#
R110|3#
C493|345|3901#
R110|5#
C521|399|3801#
SU521|399|60|undefined#|
```

Stieger and Reips (2010, p. 1490)

Navigation path example



Questionnaire navigation paradata types I

Google

Mouse clicks and mouse coordinates

Mouse clicks and its position can be captured with JavaScript. Excessive mouse movements can be a sign of problems with the question

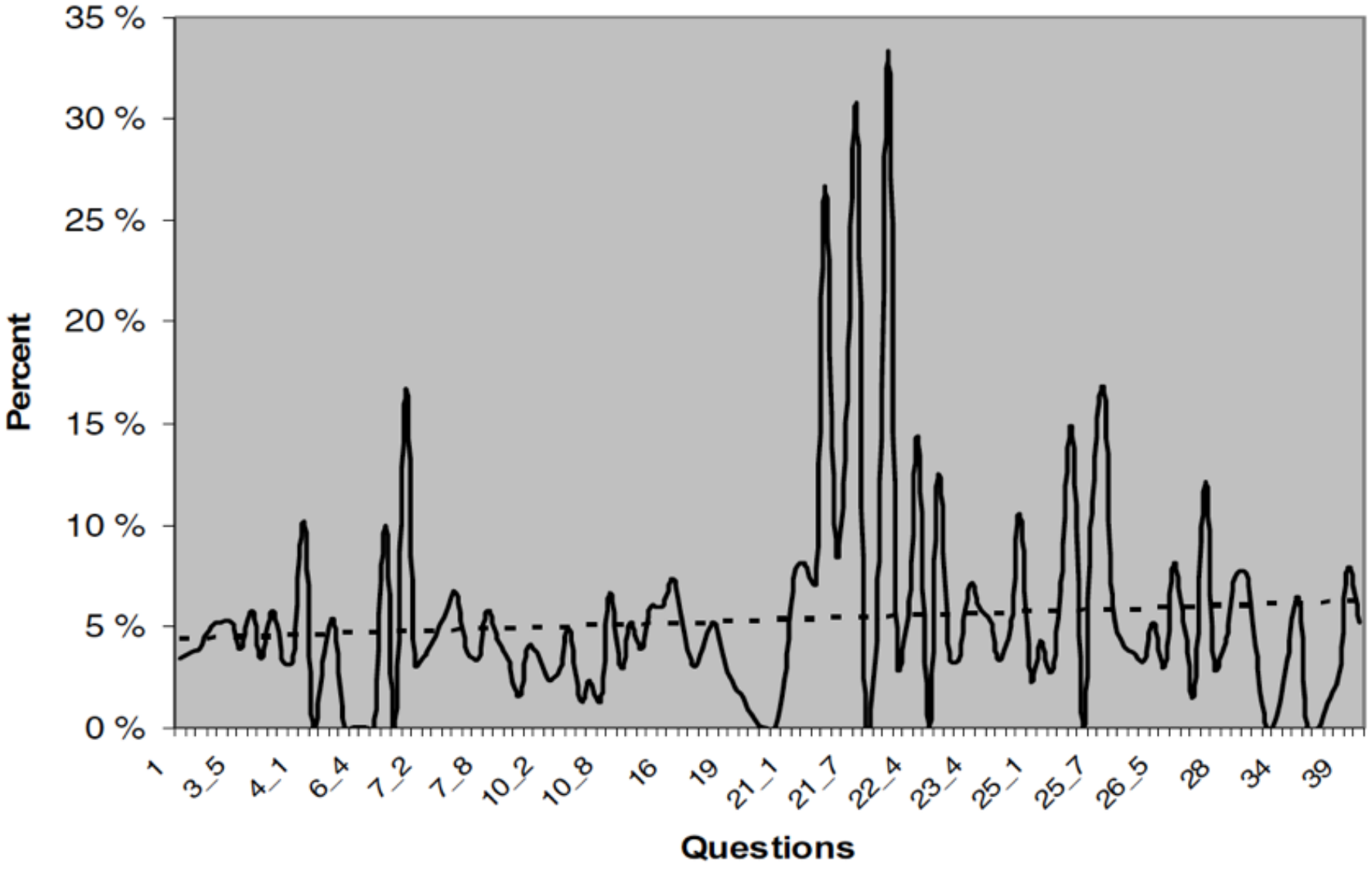
Change of answers

Change of answers is an indicator of potential confusion with a question and can be used to improve questionnaire design

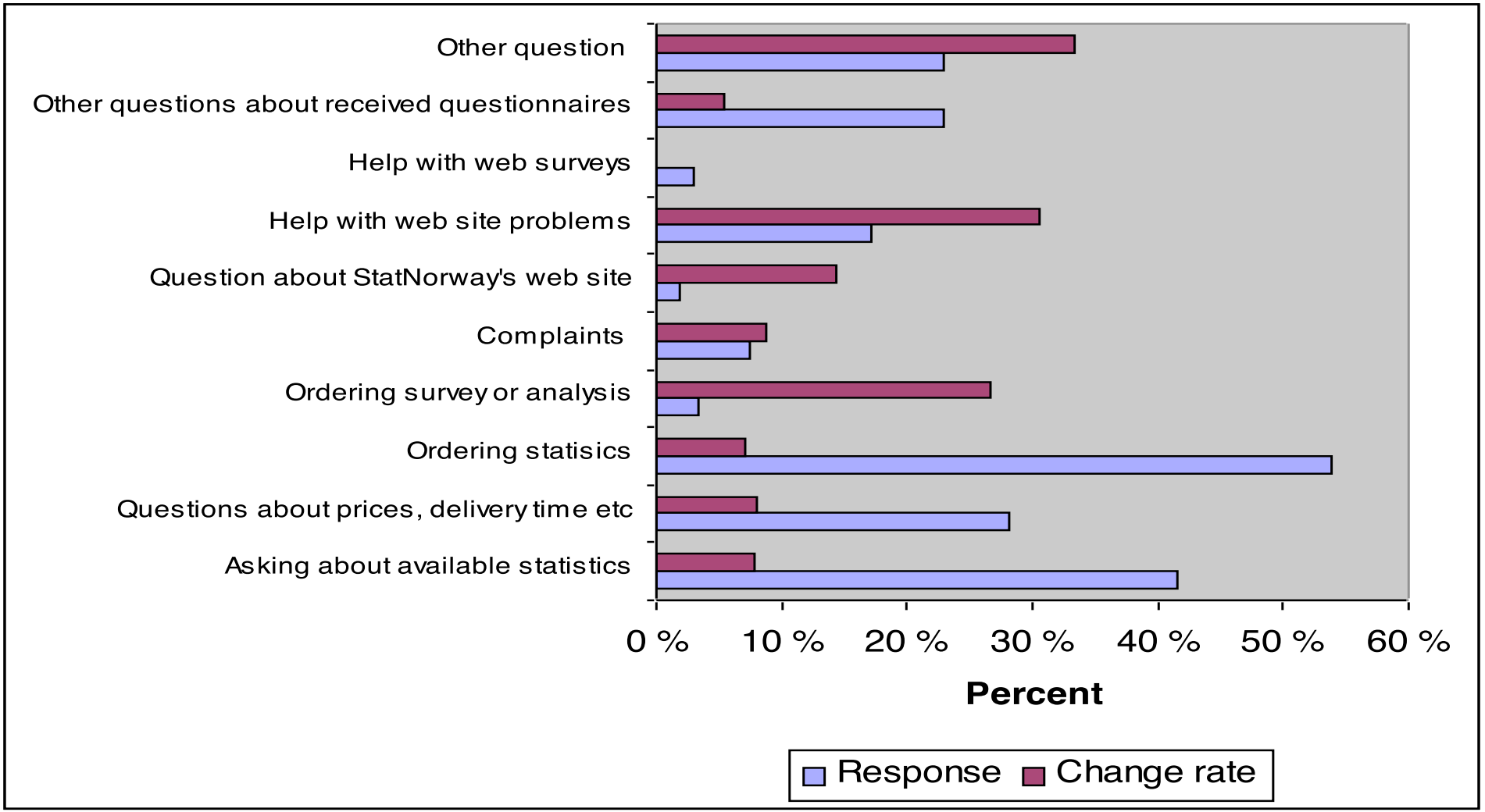
Typing and keystrokes

Typing and keystrokes can create an audit trail for each survey and used to detect unusual behavior both from the respondent side and the interviewer side

Change of answers ex. (Haraldsen et al, 2005)



Response and change rate for a multiple question (Haraldsen et al., 2005)



Question: What was the topic of your last contact? (with statistics Norway)

Questionnaire navigation paradata types II

Google

Order of answering

In a page with multiple questions the order of answering is an indicator on how the respondent reads the questions

Movements across the questionnaire (forward/backward)

If the questionnaire allows going backward or going forward by skipping questions, unusual movements are a symptom of issues with the questionnaire or the respondent

Scrolling

The amount of scrolling depends on the screen size of the device used and on the size of the browser window used by the respondent

Number of appearance of prompts and error messages & clicks on non questions links

- An unusual appearance of prompts or error messages can signify problems in the design of a questionnaire

  Haraldsen (2005) created a quality index formula based on this concept:

$$\text{Quality Index} = 1 - \frac{\sum \frac{\text{Activated errors}}{\text{Possible errors}}}{\text{Number of respondents}}$$

Clicks on non questions links

- A useful indication to check how many respondents actually use of help or definitions placed in the survey instrument

Activated error messages
Possible error messages

Google

Respondent no	∑ rooms let > ∑ lodgings	∑ beds let > ∑ beds available	∑ guest from different nations ≠ ∑ foreign guest	∑ different kinds of lodgings ≠ Total	Illegal numeric number	∑ Norwegian arrivals > Norwegian logdings	∑ rooms let > ∑ rooms available	∑ logding days > ∑ days available	∑ foreign arrivals > foreign arrivals	∑ Unique error messages
1	1			1						2
2			1	2						2
3					7					0
4										0
5			1							1
6			1	2						2
7			1	3				1	2	4
8		1		4						2
9			4	1						2
10			6	6		1				3
11						1				1
12		2				1	2			3
13				2						1
14				1		1				2
	1	3	14	22	7	4	2	1	2	**0.22**

Quality Index = 1-0.22 = 0.78

Haraldsen (2005)

Survey completion paradata:

Google

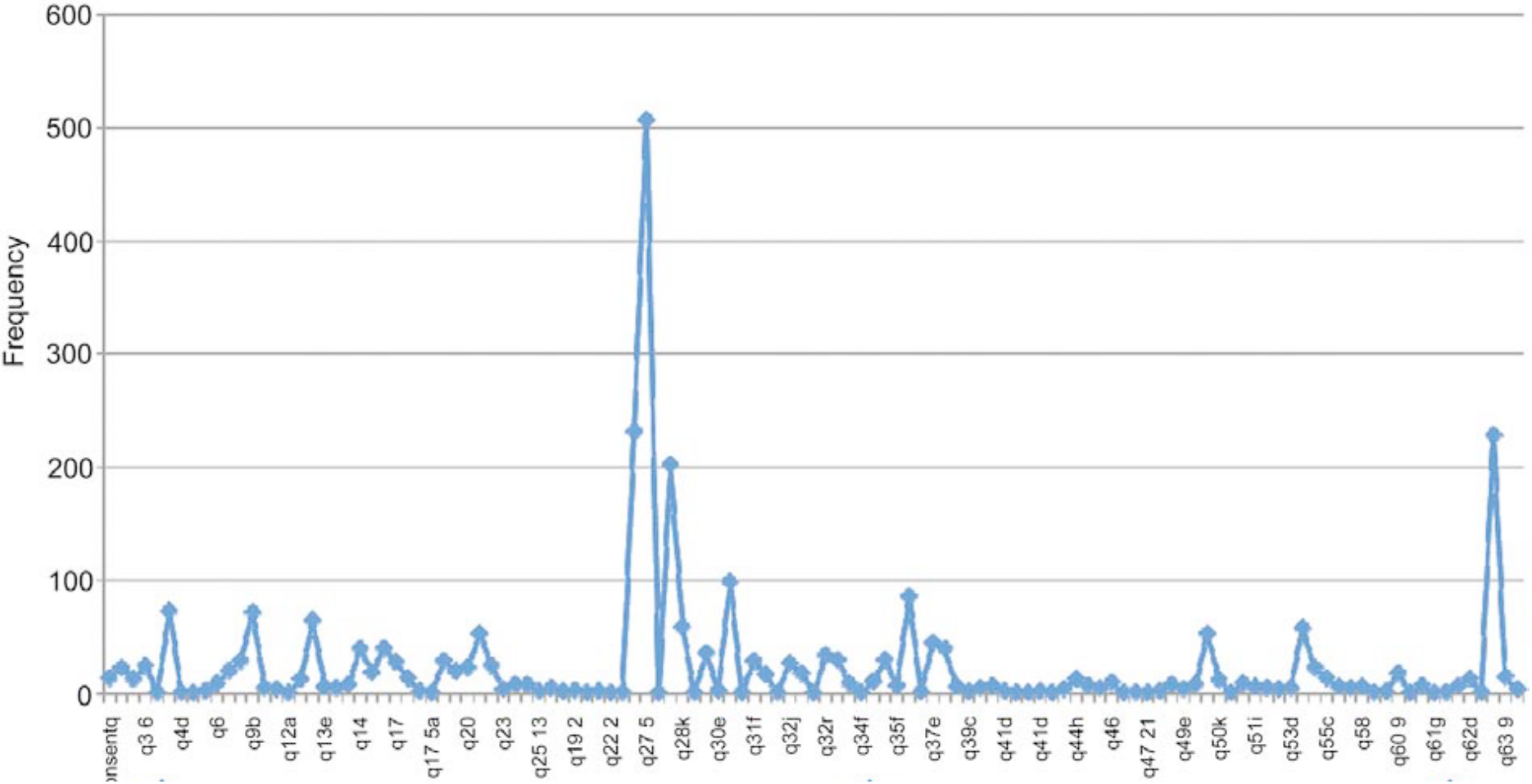
Survey resumed at a later time

Useful to correctly compute response time and interview length. For a long survey it quantifies the amount of respondents who did the survey in one go

Last question answered before dropping off

The last question answered before dropping off is probably the most common type of paradata used in online surveys. This information determines if a survey can be classified as *complete*, *partial,* or *breakoff*

Survey breakoffs by question
(Sakshaug & Crawford, 2010)



Time latency paradata

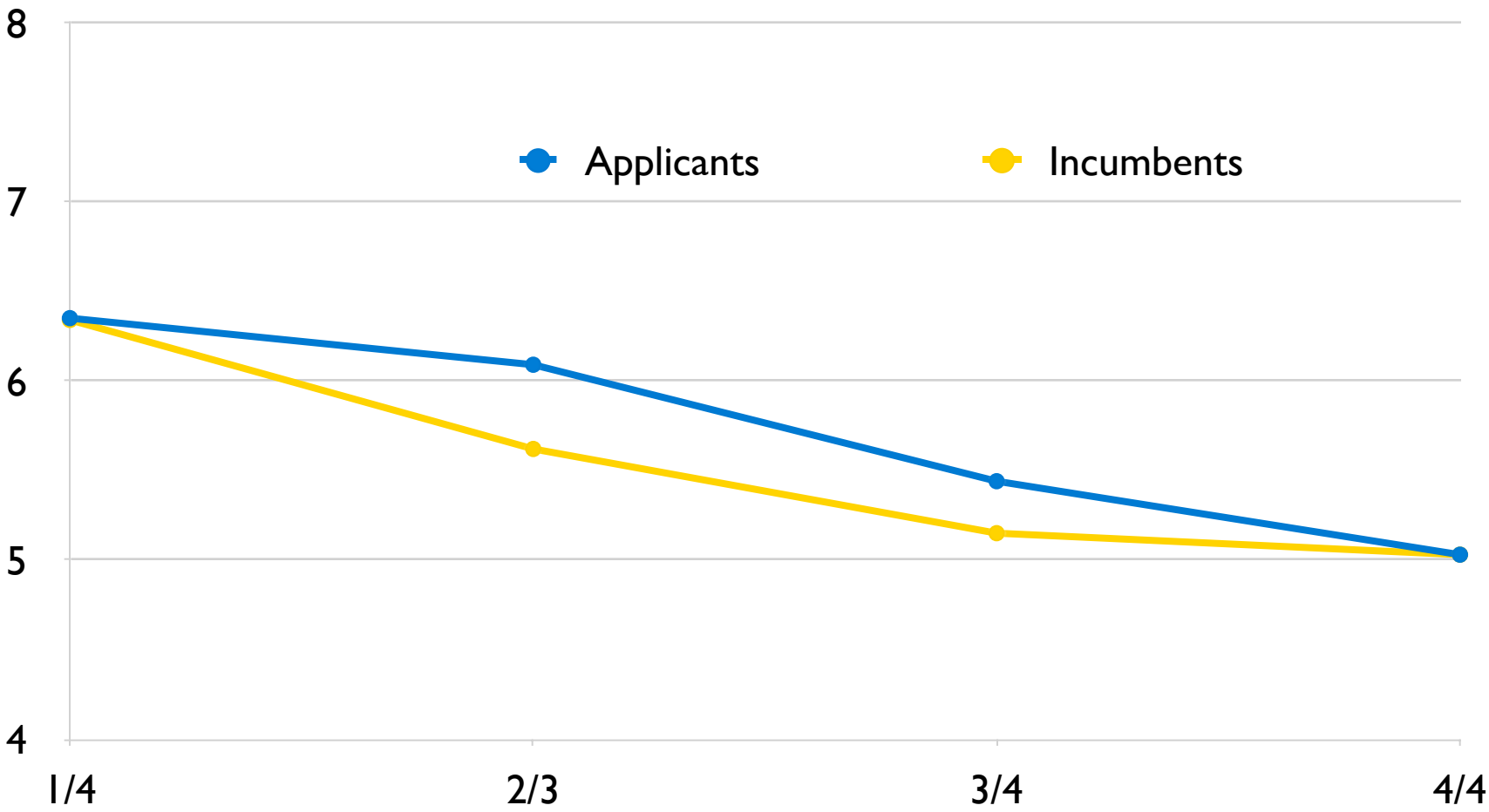
Google

Time spent per question/screen

This is the most published topic in paradata research: time latency information. There are many studies focusing on major themes:

- Attitude strength
- Response uncertainty
- Response error (e.g. speeding)

Time latency and satisfying



Callegaro, Yang, Bhola & Dillman, 2009

Paradata in online panels: longitudinal paradata

Google

Online panels have the luxury to collect historical paradata for each panel member

Examples are:

- Number of email invitations sent
- Percent of surveys qualified
- Percent of survey completed
- Percent of surveys started but not completed
- Topic of the survey
- Panel member history data
- Attrition time

Longitudinal paradata study: Bosnjak, Weyandt & Callegaro (2014)

GESIS Online Pilot Panel (GOPP): Feasibility study (2010-2012) preparing the GESIS Panel (since 2013)

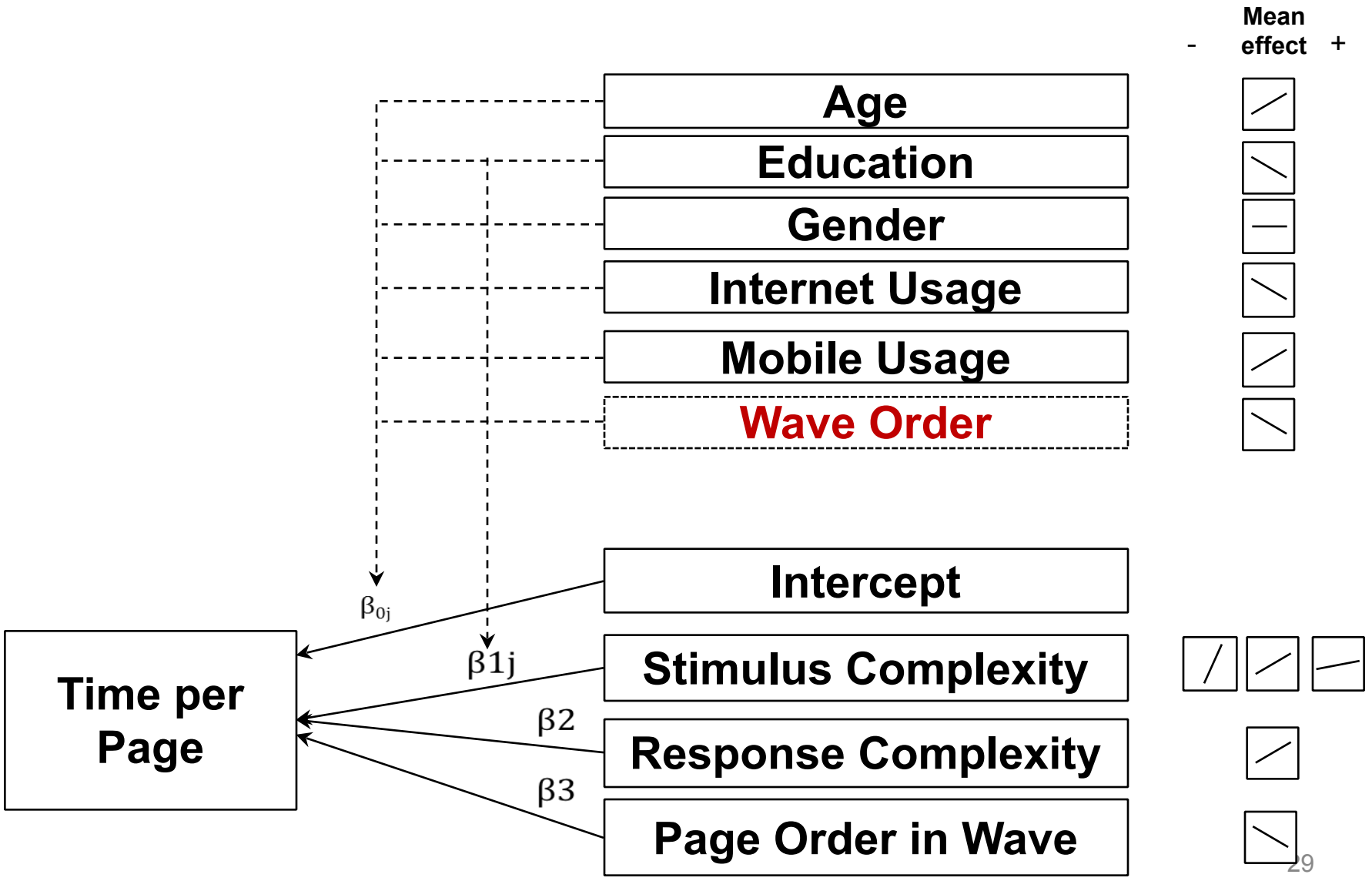
GOPP = Multi-topic panel encompassing 8 survey waves in 2011 and 2012. Several methodological experiments have been conducted.

Subsample used:

Subpopulation of N=1,041 initially active panel members that participated in at least one wave without break-off and have all non-missing values on page time and covariates (N=717)

Dependent variable: time on page measured in second

Study Model: Hypotheses



29

Study results

Multilevel mixed-effects linear regression [Magee's (1990) R^2 = .39]

- Level 1: Page-level characteristics
  Level 2: Respondent and wave characteristics combined

Panel members do not keep the same pace by wave but there is no linear trend as the number of waves increases

The three strongest effects in the model are:

1. Device (Mobile usage = longer time)

2. Change in speed by wave in comparison to first wave

3. Response complexity (more complex question = more time to answer)

Adaptive scripting

Google

- Up to now all paradata are used and analyzed **after** the data collection is concluded
- Adaptive scripting refers to using paradata in **real time** to change the survey experience for the respondent
- Introduced in 2001 by Jeavons, adaptive scripting as a way to mimic what a good interviewer would do, such as:
  - Adapting to the respondent's needs,
  - Encouraging a respondent who is about to give up to complete the survey,
  - Reducing the speed of reading the questions to adjust to the respondent's pace.
- Unfortunately very little has been done: following slides

Edit checks example part1 (Kaczmirek, 2013)

Deutschlandstudie

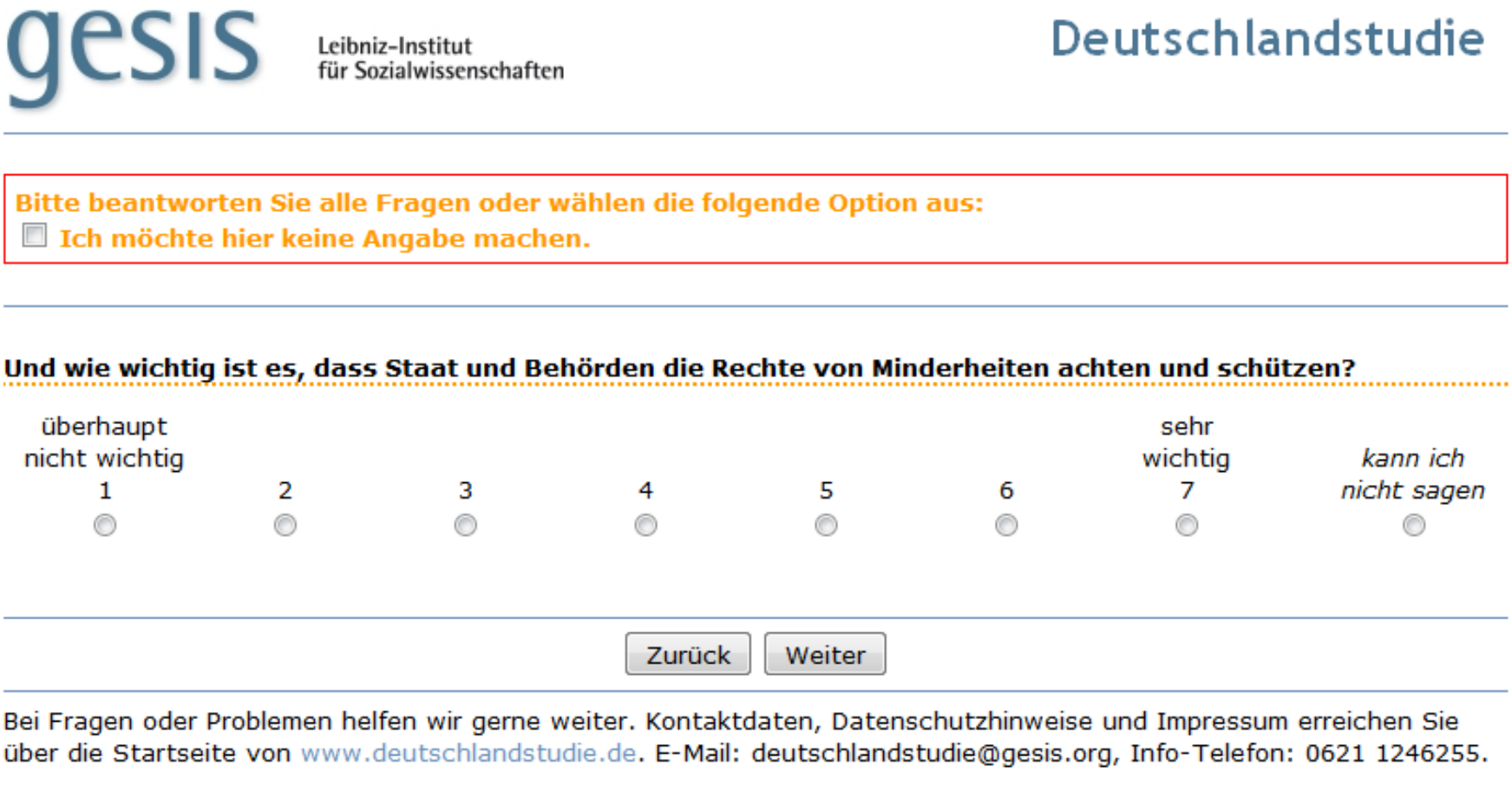
Und wie wichtig ist es, dass Staat und Behörden die Rechte von Minderheiten achten und schützen?

überhaupt nicht wichtig						sehr wichtig	*kann ich nicht sagen*
1	2	3	4	5	6	7	
○	○	○	○	○	○	○	○

Zurück Weiter

Bei Fragen oder Problemen helfen wir gerne weiter. Kontaktdaten, Datenschutzhinweise und Impressum erreichen Sie über die Startseite von www.deutschlandstudie.de. E-Mail: deutschlandstudie@gesis.org, Info-Telefon: 0621 1246255.

Edit checks example part 2 (Kaczmirek, 2013)



Do soft prompts help in reducing item nonresponse?

Google

Paradata on prompts were collected in 4 web surveys on the German Online Panel Project (GOPP)

In all four surveys, edit checks reduced item-nonresponse by an average of 48% (between 14% and 95%)

Conrad and colleagues experiments (2009-2011)

Google

Slowing down respondents

"You seem to have responded very quickly. Please be sure you have given the question on sufficient thought to provide an accurate answer. Do you want to go back and reconsider your answer?"

Major findings:

- Prompt slows respondents
- Prompts reduces straightlining in subsequent grids
- Prompt does not affect breakoffs

Privacy and ethical issues in collecting paradata

Google

Should we tell respondents we are collecting paradata?

What happens when we tell respondents we are collecting paradata and we ask permission to use them?

- 59.5% agreed in the LISS Dutch panel (across manipulations)
- 65.6% agreed in the Knowledge Networks U.S. panel (across manipulations)
- 69.3% agreed in a U.S. volunteer non-probability panel (across manipulations)

(Couper and Singer, 2013, studies done using vignettes)

Conclusions

- The amount of paradata that can be collected seems to grow as the technological capabilities grow and new devices comes to play
- Although paradata can be collected “easily” and at a low cost, researchers should not underestimate the cost of managing and analyzing paradata (Nicolaas, 2011)
- Paradata should not replace other ways of pretesting the questionnaire because it does not answer all the questions
- Paradata analysis is another tool to use in assessing the quality of a survey and in making improvements to the questionnaire and the entire online survey experience

Implication for research

- Collect/ask your vendor to capture paradata
  - Focus on questions that are less tested or where you suspect there might be some issues
  - Ask the vendor to do some pre-processing of paradata
  - Look at paradata closely during the pretest or soft launch phase

References

Chapter 11:

Paradata in Web Surveys