2IMP40 EMPIRICAL METHODS IN SOFTWARE ENGINEERING

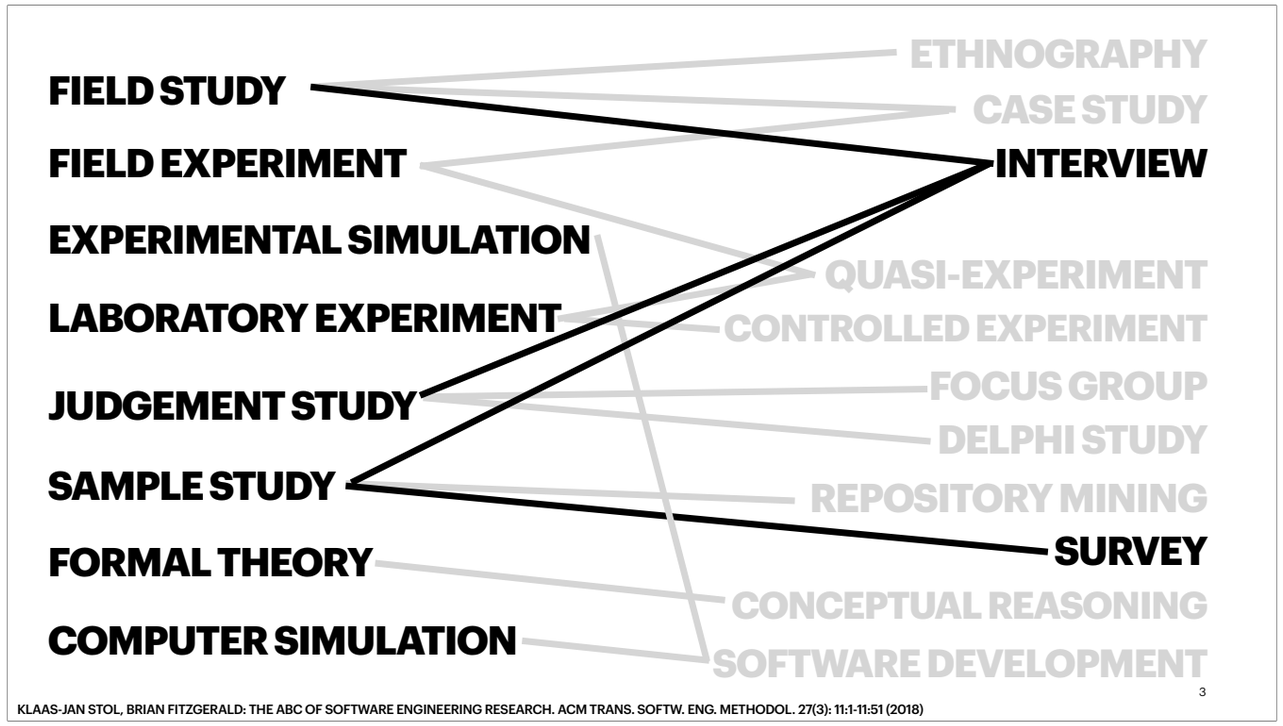
HOW DO WE TALK TO PEOPLE? INTERVIEWS AND SURVEYS

ALEXANDER SEREBRENIK

PLEASE TELL US WHAT YOU THINK

"My only (very minor) critique is to add slide numbers to the presentations (and fix the typo in slide 13...)"

Thank you for providing feedback. RECAP: both Nathan and I will do our best to make this course interesting for you. Please do not wait till the end of the course to provide us feedback.



KLAAS-JAN STOL, BRIAN FITZGERALD: THE ABC OF SOFTWARE ENGINEERING RESEARCH. ACM TRANS. SOFTW. ENG. METHODOL. 27(3): 11:1-11:51 (2018)

Recall the the links merely indicate *typical* methods. For example, you might remember that the fMRI study which is a laboratory experiment also involved interviewing participants. In the same way, interviews can be combined with any other strategy. Similarly, the study of gendered behaviour on GitHub combined repository mining and a survey.

WHY DO WE WANT TO TALK TO PEOPLE?

https://www.nonprofitmarketingguide.com/blog/wp-content/uploads/2017/05/bigstockphoto_Interview_Time_4503972.jpg

Our goal is to conduct a broad survey on how “big software data” impacts engineering teams across different organizations at Microsoft, and how data scientists and other team members coordinate, communicate, and make decisions based on data-driven insights.

We interviewed several formal methods users about the use of formal methods and their impact on various aspects of software engineering including the effects on the company, its products and its development processes as well as pragmatic issues such as scalability, understandability and tool support. The interviews are a first stage of empirical assessment. Future work will investigate some of the issues raised using formal experimentation and case studies.

MIRYUNG KIM, THOMAS ZIMMERMANN, ROBERT DELINE, ANDREW BEGEL: THE EMERGING ROLE OF DATA SCIENTISTS ON SOFTWARE DEVELOPMENT TEAM. INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING, 2016
COLIN F. SNOOK, RACHEL HARRISON PRACTITIONERS' VIEWS ON THE USE OF FORMAL METHODS: AN INDUSTRIAL SURVEY BY STRUCTURED INTERVIEW. INF. SOFTW. TECHNOL. 43(4): 275-283 (2001)

Please read these fragments and see what is common between them.

Our goal is to conduct a **broad survey** on how “big software data” impacts engineering teams across different organizations at Microsoft, and how data scientists and other team members coordinate, communicate, and make decisions based on data-driven insight

Exploratory

about the use of formal
pact on various aspects of software engineering
including the effects on the company, its products and its development processes as well as pragmatic issues such as scalability, understandability and tool support. The interviews are a **first stage** of empirical assessment. Future work will investigate some of the issues raised using formal experimentation and case studies.

MIRYUNG KIM, THOMAS ZIMMERMANN, ROBERT DELINE, ANDREW BEGEL: THE EMERGING ROLE OF DATA SCIENTISTS ON SOFTWARE DEVELOPMENT TEAM. INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING, 2016
COLIN F. SNOOK, RACHEL HARRISON PRACTITIONERS' VIEWS ON THE USE OF FORMAL METHODS: AN INDUSTRIAL SURVEY BY STRUCTURED INTERVIEW. INF. SOFTW. TECHNOL. 43(4): 275-283 (2001)

We observe that 25 out of 26 components contain SSSMs, making up 25.3% of the model base. To **understand** the reasons for this extensive usage we conduct a series of interviews followed by a grounded theory building.

> The goal behind the Interactive Interview is to be able to observe developers actually using a static analysis tool. <...> We asked our participants to **explain** what they are doing out loud so we could get a better understanding of their workflow and thought process.

Our results show no significant differences between working in distributed vs. non-distributed teams, suggesting that Scrum helps alleviate many GSE problems. Our post-course interviews and survey data allows us to **explain** this effect; we found that students over time learned to better self-select tasks with less inter-team dependencies, to communicate more, and to work better in teams.

NAN YANG, PIETER J. L. CUIJPERS, RAMON R. H. SCHIFFELERS, JOHAN LUKKIEN, ALEXANDER SEREBRENIK: PAINTING FLOWERS: REASONS FOR USING SINGLE-STATE STATE MACHINES IN MODEL-DRIVEN ENGINEERING. MSR 2020: 362-373

BRITTANY JOHNSON, YOONKI SONG, EMERSON R. MURPHY-HILL, ROBERT W. BOWDIDGE: WHY DON'T SOFTWARE DEVELOPERS USE STATIC ANALYSIS TOOLS TO FIND BUGS? ICSE 2013: 672-681

MARIA PAASIVAARA, KELLY BLINCOE, CASPER LASSENIUS, DANIELA E. DAMIAN, JYOTI SHEORAN, FRANCIS HARRISON, PRASHANT CHHABRA, AMINAH YUSSUF, VEIKKO ISOTALO: LEARNING GLOBAL AGILE SOFTWARE ENGINEERING USING SAME-SITE AND CROSS-SITE TEAMS. ICSE (2) 2015: 285-294

However, exploration is not the only situation when interviews are useful

We observe that 25 out of 26 components contain SSSMs, making up 25.3% of the model base. To **understand** the reasons for this extensive usage we conduct a series of interviews followed by a grounded theory building.

Th… able to observe
…> We asked our
ud so we could get
ght process.

Explanatory

Our results show no significant differences between working in distributed vs. non-distributed teams, suggesting that Scrum helps alleviate many GSE problems. Our post-course interviews and survey data allows us to **explain** this effect; we found that students over time learned to better self-select tasks with less inter-team dependencies, to communicate more, and to work better in teams.

NAN YANG, PIETER J. L. CUIJPERS, RAMON R. H. SCHIFFELERS, JOHAN LUKKIEN, ALEXANDER SEREBRENIK: PAINTING FLOWERS: REASONS FOR USING SINGLE-STATE STATE MACHINES IN MODEL-DRIVEN ENGINEERING. MSR 2020: 362-373

BRITTANY JOHNSON, YOONKI SONG, EMERSON R. MURPHY-HILL, ROBERT W. BOWDIDGE: WHY DON'T SOFTWARE DEVELOPERS USE STATIC ANALYSIS TOOLS TO FIND BUGS? ICSE 2013: 672-681

MARIA PAASIVAARA, KELLY BLINCOE, CASPER LASSENIUS, DANIELA E. DAMIAN, JYOTI SHEORAN, FRANCIS HARRISON, PRASHANT CHHABRA, AMINAH YUSSUF, VEIKKO ISOTALO: LEARNING GLOBAL AGILE SOFTWARE ENGINEERING USING SAME-SITE AND CROSS-SITE TEAMS. ICSE (2) 2015: 285-294

However, exploration is not the only situation when interviews are useful

QUESTION

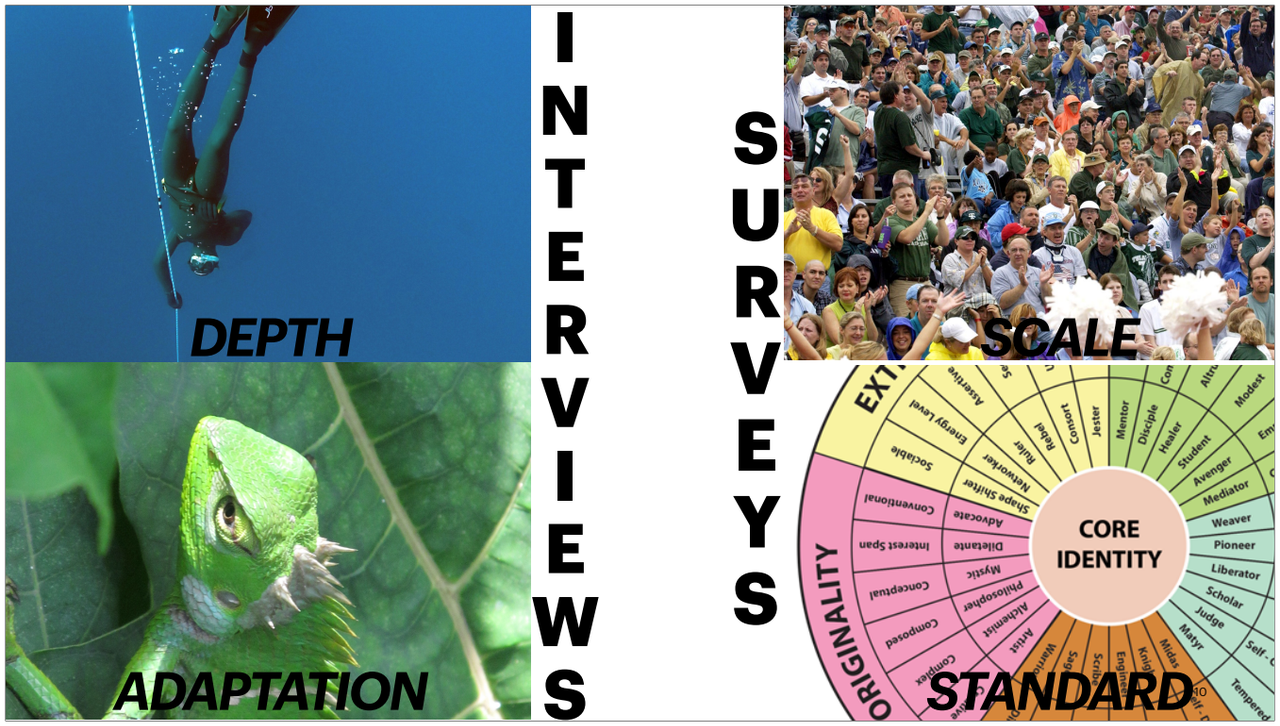
We conducted an interview study with 24 practitioners from 17 companies during 20 interviews. Following a semi-structured approach, we asked for their definition of CSE (*continuous software engineering*), most relevant elements for CSE, their experiences, and plans for further additions to their CSE process.

(A) EXPLORATORY

(B) EXPLANATORY

JAN OLE JOHANSSEN, ANJA KLEEBAUM, BARBARA PAECH, BERND BRUEGGE PRACTITIONERS' EYE ON CONTINUOUS SOFTWARE ENGINEERING: AN INTERVIEW STUDY. ICSSP 2018: 41-50

Continuous software engineering (CSE) bundles activities, such as continuous integration and delivery, to enable continuous learning and improvement by frequently iterating on software increments.
Exploratory.



Interviews allow us to go deeper, ask clarification questions; surveys are often administered using webforms and there is no way to prevent the respondents from giving shallow answers. Moreover, we can adapt the interview to the interviewee, we can them to reflect on what they have said before or on what other interviewees have said. Advantages of the surveys are scale (we can survey many more people than we can ever interview: e.g., we can survey hundreds or thousands of developers while it is rare that more than 50 people are interviewed). This scale would allow us to perform quantitative analysis, tease more fine grained insights, etc. Moreover, surveys allow us to ask standardised questionnaires from psychology grounding our research

Depending on what kind of questions we will ask, interviews come in three flavours: structured, semi-structured and unstructured.

Structured interviews have a very specific goal. Questions are asked exactly as written, no deviations occur. The questions are usually quantitative (yes/no, how many, high/medium/low), the analysis involves statistics and visualisation. Uncommon: survey is a more appropriate instrument

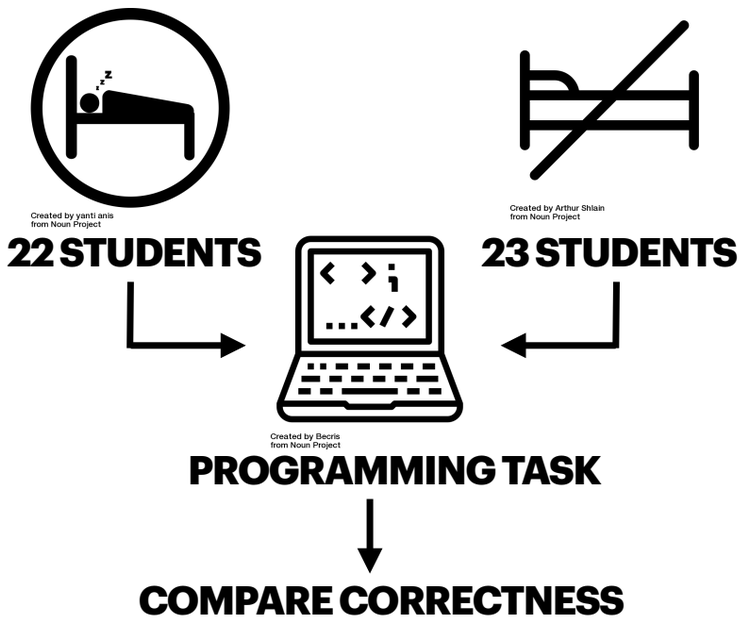
Unstructured interviews do not have predefined questions but a broadly defined topic. Extreme case: no questions asked as not to bias the interviewee. Topic is given and the interviewee talks about the topic. If questions are asked, then they should be as open as possible. Analysis: textual analysis. Uncommon: very costly since different people are likely to diverge greatly.

Semi-structured: A broad enough goal . Questions are asked as written, but more questions can be asked if needed (e.g., unexpected answers). Mixture of closed and open questions. Analysis: statistics and textual analysis. The most common type in empirical software engineering

12

https://cdn.pixabay.com/photo/2012/04/28/18/39/care-43938_1280.png

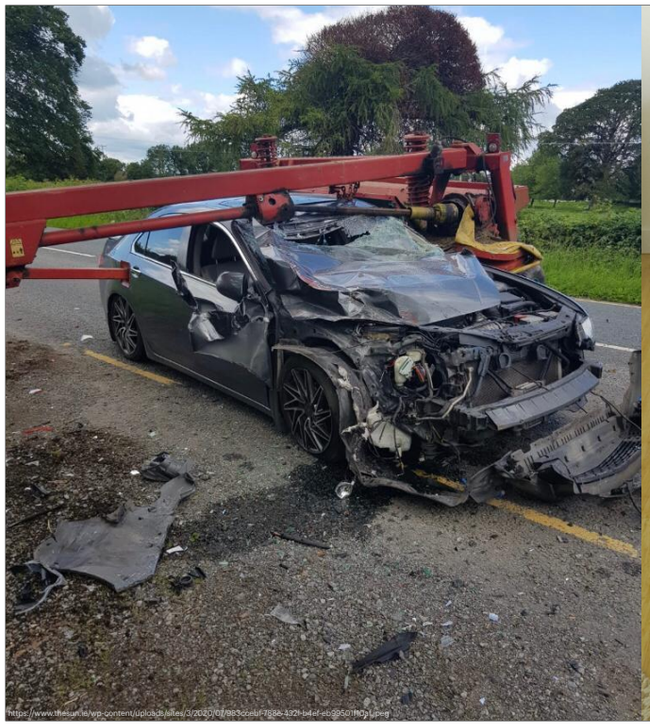
Whether we want to conduct an interview or a survey we need to keep in mind that this study will involve human beings, and this is why our primary concern should be to not harm them.



HOW CAN THIS STUDY HARM PARTICIPANTS?

WHAT WOULD YOU DO TO REDUCE THE LIKELIHOOD OR THE EXTEND OF THIS HARM?

Example of an ethically problematic design (actual paper, but I do not want to shame the authors)


https://www.thesun.ie/wp-content/uploads/sites/3/2020/01/883cceaf-7586-432f-b4ef-eb39501ff0b1.jpeg


https://thecrazyorganizedblog.com/wp-content/uploads/2016/10/how-to-make-a-bed-like-a-hotel-11.jpg

1) Falling asleep at the wheel
2) Offer a night sleep next to the experiment vicinity
Moreover, by now the evidence of negative consequences of sleep deprivation is so abundant that psychology researchers consider this kind of studies unethical and unnecessary.

https://tuenl.sharepoint.com/sites/intranet-ethical-review

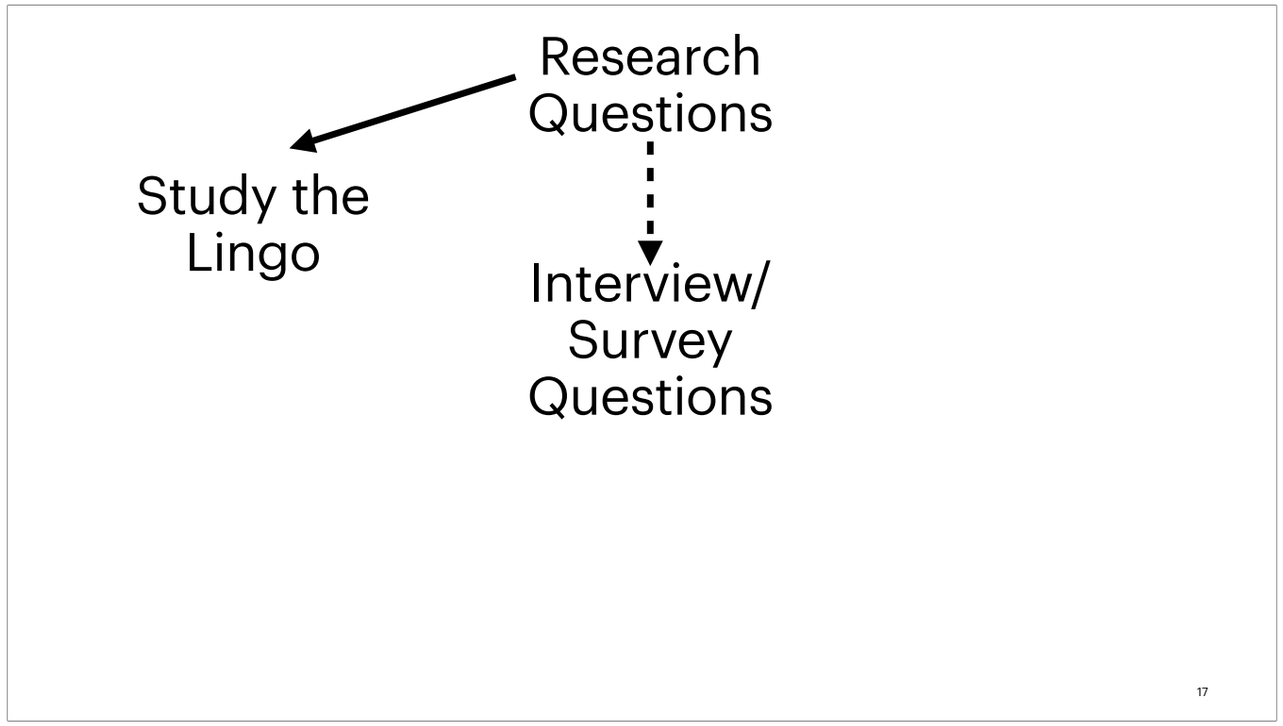
Part 4: Information about the study		
1	What are your main research questions? *Additional explanation: You need to provide at least one clear research question.*	RQs: Lecture 2
2a	Please check the box that indicates the relevant study population *Additional explanation: Please select which persons are eligible for your study.*	☐ Students ☐ General healthy population ☐ General population with specific feature, e.g., pregnancy, specifically ..... ☐ Patients, specifically ..... ☐ Other, specifically .....
2b	Age category of participants	☐ Younger than 12 years of age ☐ Older than 11 and younger than 16 years of age ☐ 16 years or older
3	Description of the research method (select all that applies)	☐ (Semi-structured) interviews ☐ Surveys

Strategies: Lecture 2
Interviews and Surveys: Today

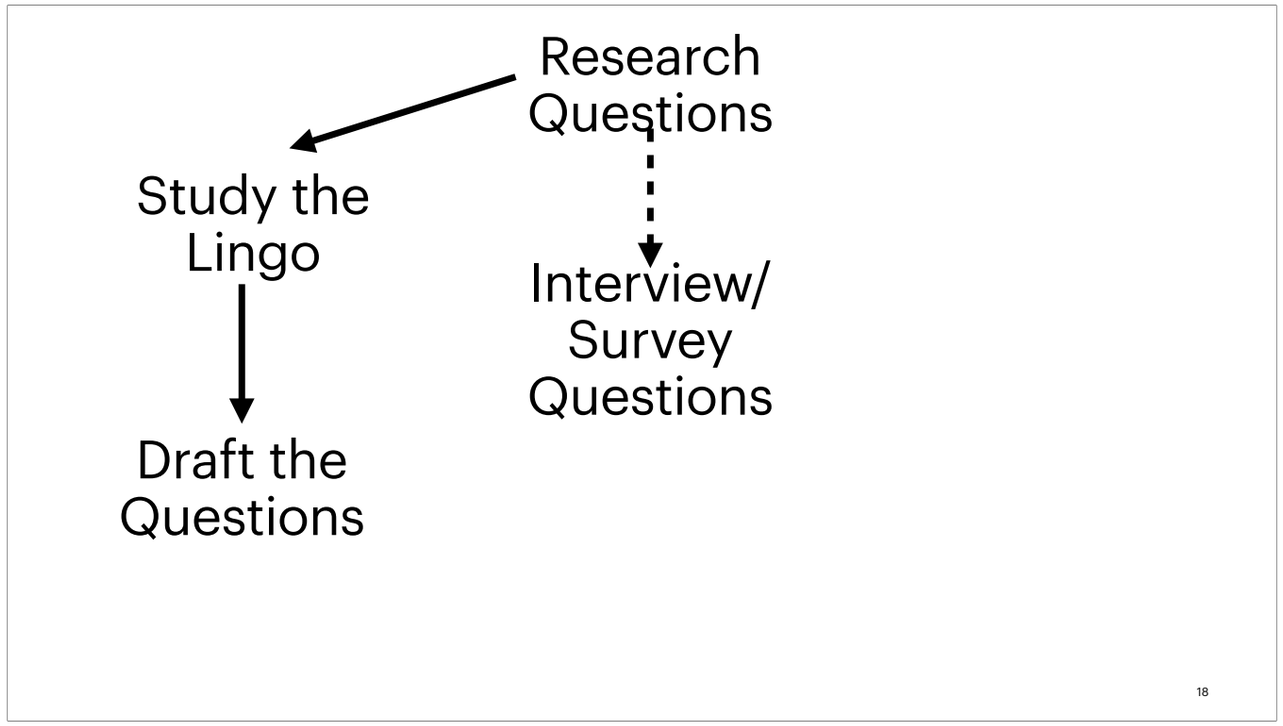
	Additional explanation: Please specify your research method. Note that you need to provide information about the research method in an additional file that you attach to the ERB form. E.g., for interviews you provide the interview questions, for surveys you provide the survey questions, etc.	☐ Group workshops/roundtable discussios ☐ Diary studies ☐ Behavioral observations ☐ Building sensor data ☐ Wearable device (e.g. Fitbit watch, on-skin sensors) ☐ User testing ☐ Pilot study ☐ GPS tracking/location data ☐ Living Lab ☐ Other, namely .....
4	Description of the measurements and/or stimuli/treatments *Additional explanation: Think about your outcome measures and the variables you will be collecting and describe them in a way such that another person understands what the participant will experience. For example: Participants will perform task A and see pictures from database B, and we measure validated Scale 1.*	Interview guide, survey questionnaire
5	Describe and justify the number of participants you need for this study. Also justify the number of observations you need, taking into account the risks and benefits. *Additional explanation: Think about if you need 3 or 30 participants for example, and why? Do they need to provide their input once, or several times, and why? If relevant, specify the duration of the study per participant and the compensation that is needed for the study.*	How many people?
6	Explain why your research is societally important. What benefits and harm to society may result from the study? *Additional explanation: What benefit will the results of your study have to society in general?*	Motivation: Lecture 1
7	Describe the way participants will be recruited *Additional explanation: How will you recruit participants for your study? For example, by using flyers, personal network, panels, etc.* Recruitment	☐ Survey link posted online, e.g., social media platforms ☐ On campus flyers ☐ Personal network ☐ Via a company, namely ..... ☐ Via a hospital, namely ..... ☐ Via an organization ..... ☐ By a Consortium Partner, namely ..... ☐ Other, namely .....
8	Provide a brief statement of the risks you expect for the participants or others involved in the study and explain. Also take into consideration any personal data you may gather and associated privacy issues. *Additional explanation: Risks for the participants can be anything from risk of data breach to risk of safety or well-being (think about stress, extreme emotions, visual or auditory discomfort). Describe these possible risks and describe the way these risks are mitigated.*	Risks

The next question is how many people we want to interview

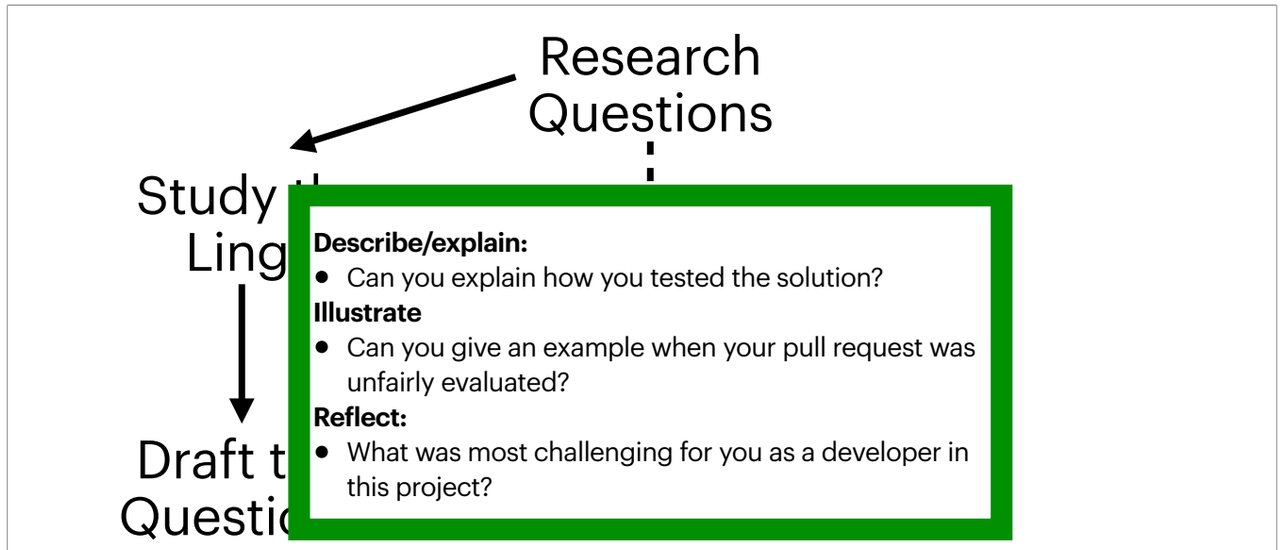
So how do we decide what questions to ask?
If you plan on conducting a survey, a structured with questions or a semi-structured interview, you are not going to ask just questions merely for the sake of asking questions. In research, we have a goal, we want to answer a research question. This means that interview questions should be somehow derived from research questions



Of course, this “somehow” part does not really work. In practice, the process usually is usually more complicated.
The first step is learning the lingo. For example, we wanted to study how logs are used at a large company in Eindhoven region. Intuitively, we could have just asked people about logs. However, at this company has four types of files recording execution-related data, and only one is called “execution logs”. Other types are functional traces, throughput traces and functional data. If we would have asked only about the logs, we would have got only partial answers.

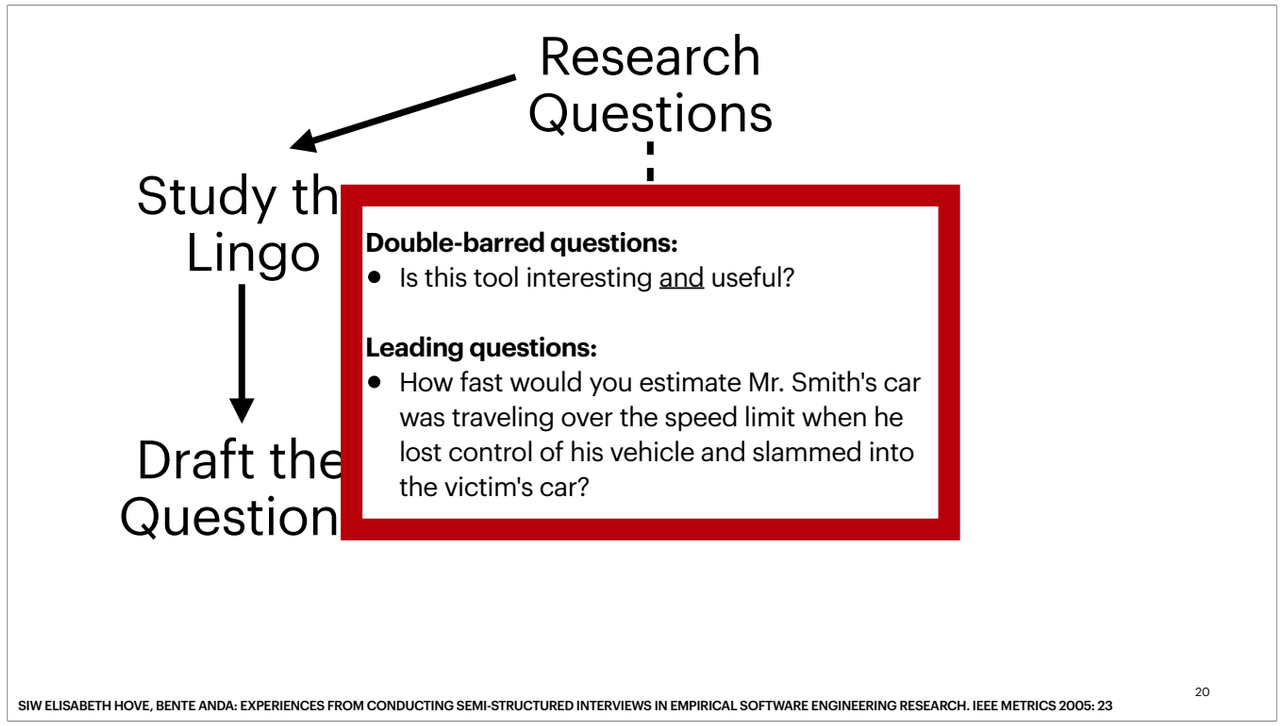


Then the first draft of questions should be prepared. Each question should be motivated: if there is no motivation, there is no need to ask the question. Motivation can come from the research question or from a hypothesis that might associate what we are interested in with one or more contextual phenomena. For example, we know from the research that more experienced and less experienced developers read code in different ways. We conjecture that the same holds for logs. This is why we ask about the interviewee’s experience as part of the interview.



SIW ELISABETH HOVE, BENTE ANDA: EXPERIENCES FROM CONDUCTING SEMI-STRUCTURED INTERVIEWS IN EMPIRICAL SOFTWARE ENGINEERING RESEARCH. IEEE METRICS 2005: 23

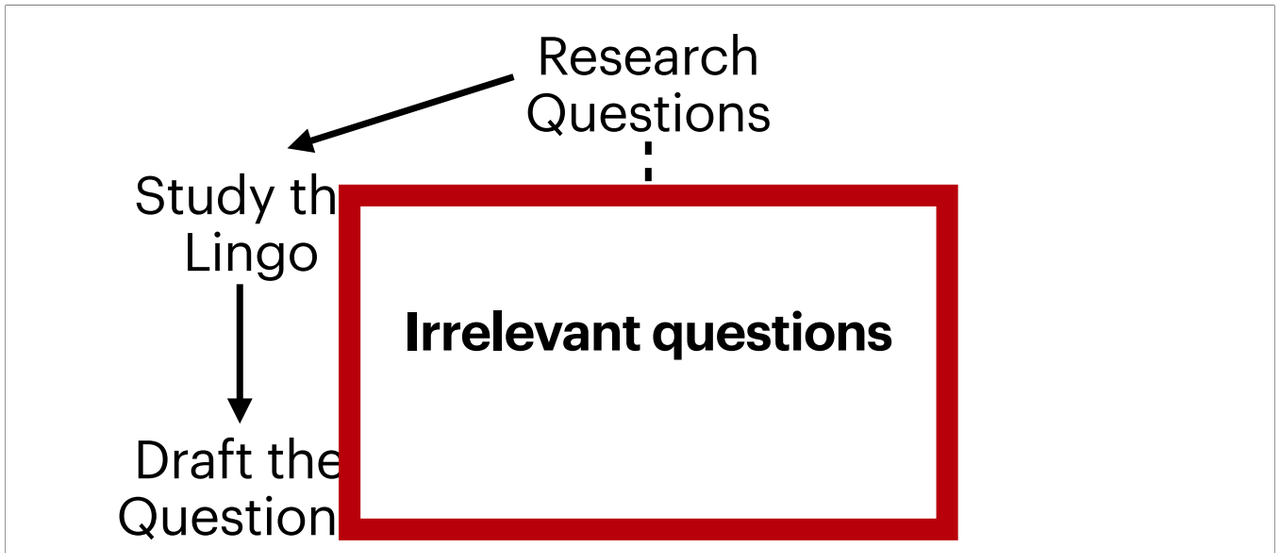
Some questions provide rich information and can be recommended



SIW ELISABETH HOVE, BENTE ANDA: EXPERIENCES FROM CONDUCTING SEMI-STRUCTURED INTERVIEWS IN EMPIRICAL SOFTWARE ENGINEERING RESEARCH. IEEE METRICS 2005: 23

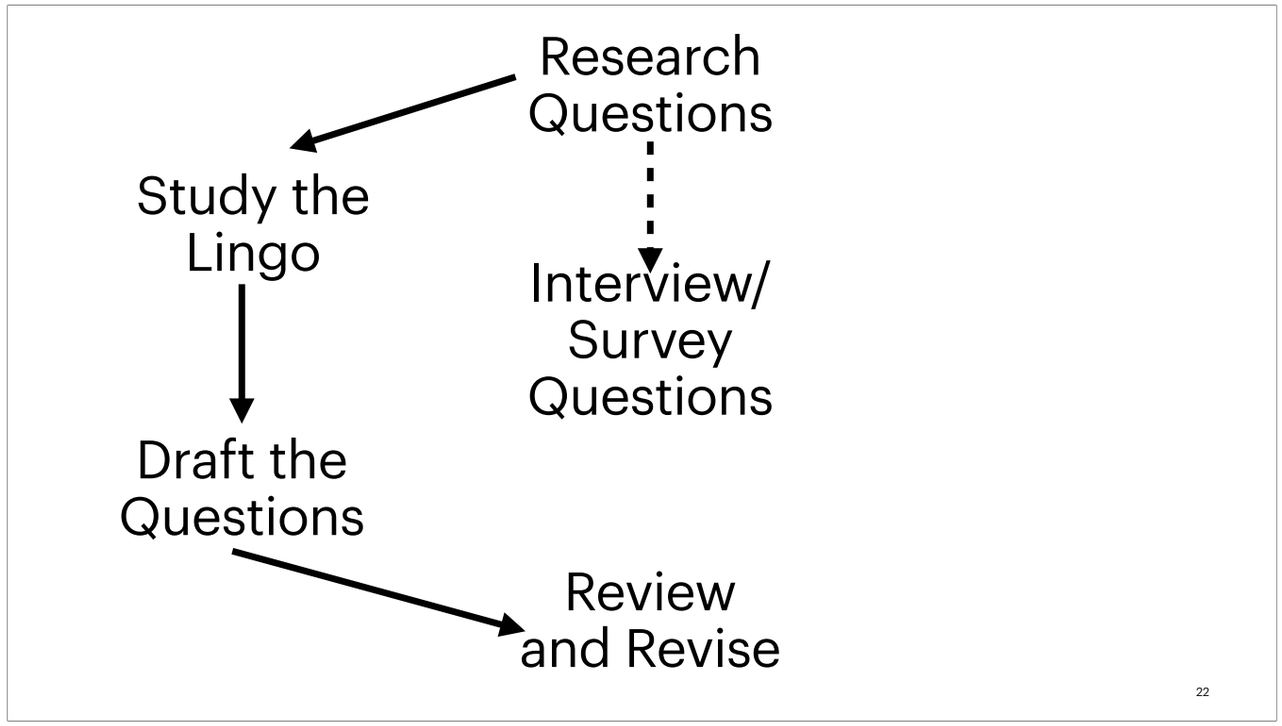
Moreover, not all questions are appropriate! You should avoid leading questions and double-barred questions

Double-barred: what if I agree with one of the options but not another one?
Leading question: is a question that suggests the particular answer or contains the information the examiner is looking to have confirmed

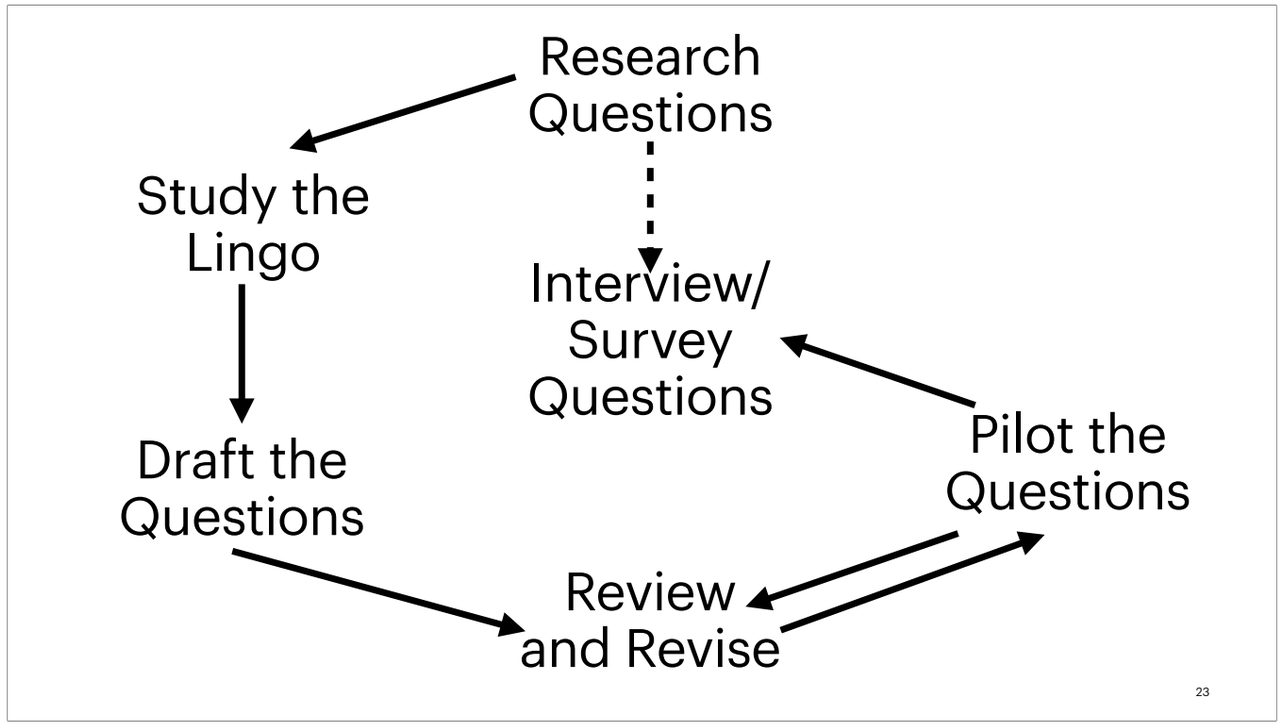


SIW ELISABETH HOVE, BENTE ANDA: EXPERIENCES FROM CONDUCTING SEMI-STRUCTURED INTERVIEWS IN EMPIRICAL SOFTWARE ENGINEERING RESEARCH. IEEE METRICS 2005: 23

Finally, we should not be collecting information that is not really necessary.



The process of polishing questions is usually iterative. During this process you need to ensure that the respondents understand the goal of the study, that the questions are clear and unambiguous and of course, that they do not make the interviewee uncomfortable in any way. AGAIN, if there is no good reason to ask the question, do not ask.



All the steps so far have only involved the researchers, so we cannot really be sure that the questions are unambiguous etc. This is why we need to pilot the questions, ideally with one of the members of the target audience - beware that we cannot use the answers of the pilot participants. This is also why we do not directly go to the pilot stage, and first of all review the interview questions ourselves.

The final version of the interview questions is known as the **interview guide**.

However, before inventing your own questions please check the literature. This is particularly important for surveys but might be important for the interviews as well.

Many surveys have been published and it is very likely that you can reuse some of the questions. For example, if you are interested in studying what kind of development activities make developers happy, you might like to consider a generic Satisfaction with Life questionnaire; how developers understand the source code they are reading, you might consider performing cognitive tests that can measure working memory, executive functioning, and spatial ability as these have been related to program comprehension in the past.

https://tuenl.sharepoint.com/sites/intranet-ethical-review

Part 4: Information about the study		
1	What are your main research questions? *Additional explanation: You need to provide at least one clear research question.*	RQs: Lecture 2
2a	Please check the box that indicates the relevant study population *Additional explanation: Please select which persons are eligible for your study.*	☐ Students ☐ General healthy population ☐ General population with specific feature, e.g., pregnancy, specifically ..... ☐ Patients, specifically ..... ☐ Other, specifically .....
2b	Age category of participants	☐ Younger than 12 years of age ☐ Older than 11 and younger than 16 years of age ☐ 16 years or older
3	Description of the research method (select all that applies)	☐ (Semi-structured) interviews ☐ Surveys

Strategies: Lecture 2
Interviews and Surveys: Today

	Additional explanation: Please specify your research method. Note that you need to provide information about the research method in an additional file that you attach to the ERB form. E.g., for interviews you provide the interview questions, for surveys you provide the survey questions, etc.	☐ Group workshops/roundtable discussios ☐ Diary studies ☐ Behavioral observations ☐ Building sensor data ☐ Wearable device (e.g. Fitbit watch, on-skin sensors) ☐ User testing ☐ Pilot study ☐ GPS tracking/location data ☐ Living Lab ☐ Other, namely .....
4	Description of the measurements and/or stimuli/treatments *Additional explanation: Think about your outcome measures and the variables you will be collecting and describe them in a way such that another person understands what the participant will experience. For example: Participants will perform task A and see pictures from database B, and we measure validated Scale 1.*	Interview guide, survey questionnaire
5	Describe and justify the number of participants you need for this study. Also justify the number of observations you need, taking into account the risks and benefits. *Additional explanation: Think about if you need 3 or 30 participants for example, and why? Do they need to provide their input once, or several times, and why? If relevant, specify the duration of the study per participant and the compensation that is needed for the study.*	How many people?
6	Explain why your research is societally important. What benefits and harm to society may result from the study? *Additional explanation: What benefit will the results of your study have to society in general?*	Motivation: Lecture 1
7	Describe the way participants will be recruited *Additional explanation: How will you recruit participants for your study? For example, by using flyers, personal network, panels, etc.* Recruitment	☐ Survey link posted online, e.g., social media platforms ☐ On campus flyers ☐ Personal network ☐ Via a company, namely ..... ☐ Via a hospital, namely ..... ☐ Via an organization ..... ☐ By a Consortium Partner, namely ..... ☐ Other, namely .....
8	Provide a brief statement of the risks you expect for the participants or others involved in the study and explain. Also take into consideration any personal data you may gather and associated privacy issues. *Additional explanation: Risks for the participants can be anything from risk of data breach to risk of safety or well-being (think about stress, extreme emotions, visual or auditory discomfort). Describe these possible risks and describe the way these risks are mitigated.*	Risks

The next question is how many people we want to interview

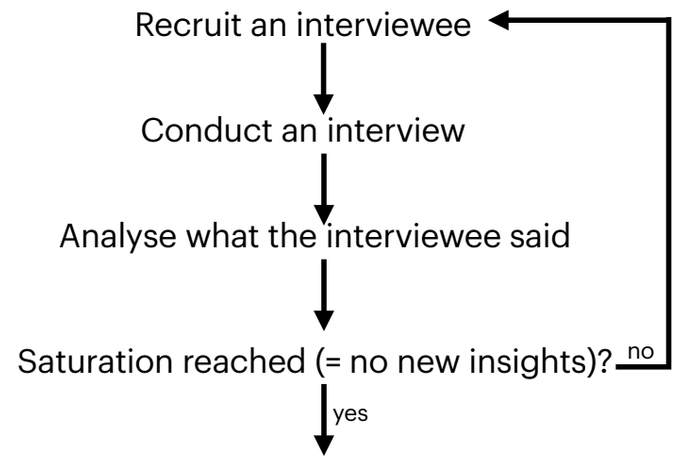
HOW MANY PEOPLE DO WE WANT TO INTERVIEW?

25-35

25-35 is a common number in SE studies. However, many things depend on how do we recruit people. For example, recall from the previous lecture that if we are doing quota sampling, then what is important is all substrata are adequately represented.

What advantages and disadvantages would you expect? We might interview too many people which would constitute an unnecessary burden for them and waste of time for us. We might interview too few people and miss important insights.

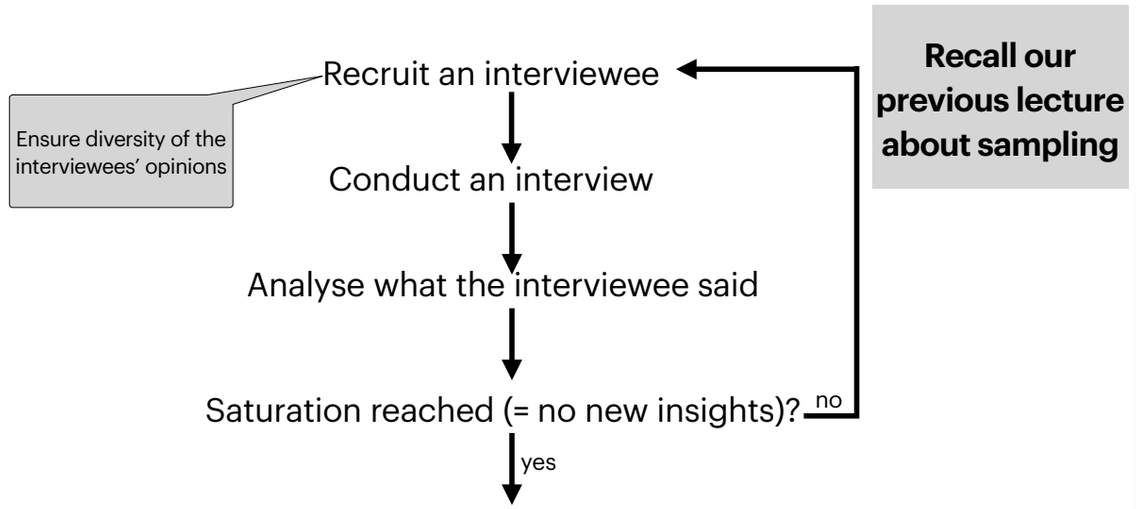
HOW MANY PEOPLE DO WE WANT TO INTERVIEW?



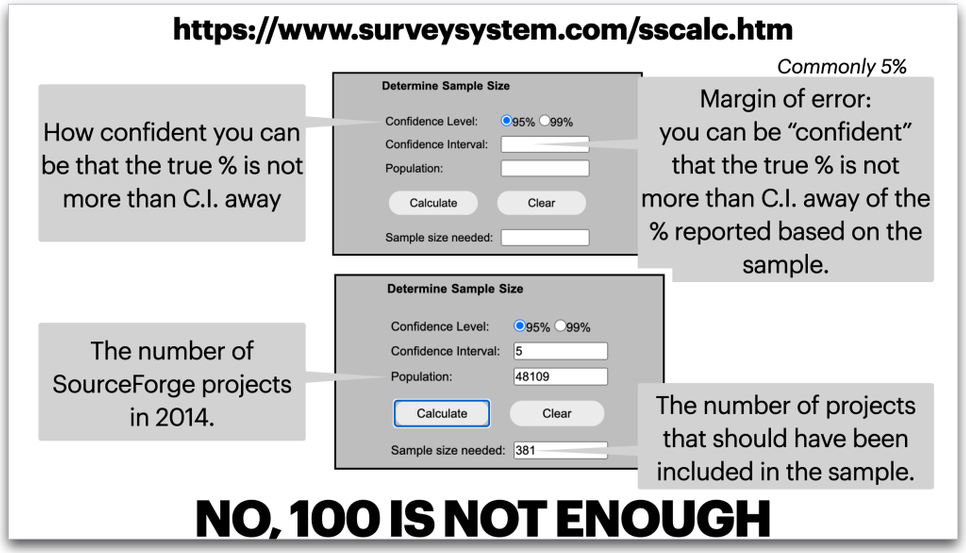
Recall our previous lecture about sampling

An alternative approach would require us to analyse the interviews not after they have been conducted but intermittent with conducting them.

HOW MANY PEOPLE DO WE WANT TO INTERVIEW?



An alternative approach would require us to analyse the interviews not after they have been conducted but intermittent with conducting them. If the sampling strategy is not adequate then the saturation can be reached too early leading to wrong conclusions. For example, if I would have interviewed only CSE students asking about the quality of our educational programs, I could have reached saturation faster but I would have missed the opinions of ES, DSiE and IST students.



What about surveys?

The first step is to use the same computation of the sample size we have discussed in the lecture about sampling to determine the number of projects the authors should have studied. When selecting the parameters you should keep in mind that \

- for the confidence level, 95% and 99% are commonly used.
- for the confidence interval, one usually uses 1% and 5%

$$SS = \frac{Z^2 * p * (1 - p)}{c^2}$$

	Confidence interval 1%	Confidence interval 5%
Confidence level 95%	9604	384.16
Confidence level 99%	16641	665.64

Z is a function of the confidence level, c is the confidence interval. p is the proportion being estimated, in our case it is always 0.5 since it gives the largest sample.
Z = 1.96 for 95% and Z = 2.58 for 99%
Of course, we cannot interview fractions of developers, so we need to round up.
Higher confidence level and smaller confidence intervals increase the sample size

$$SS_f = \frac{SS}{1 + \frac{SS - 1}{pop}}$$

Population	Any	100K	10K	1K	100
Sample (c.l.95%, c.i. 5%)	384.16	382.69	370	277.74	79.5

This is also why the tool takes population size into account. Of course, we need to keep an eye on the pragmatic side of things. If the entire population that we want to survey is 100 people, why bother and sample? We could as easily survey the entire population.

Population	Any	100K	10K	1K	100
Sample (c.l.95%, c.i. 5%)	384.16	382.69	370	277.74	79.5
Response rate 20%	1921	1914	1850	everyone	
Response rate 10%	3842	3827	3700	everyone	

The sample size so far is, of course, the number of answers we would like to receive. Unfortunately, we need to take into account the fact that many people will not respond. In fact, the response rate, i.e., % of respondents among those you have invited, ranges between 10 and 20% in a typical software engineering survey. Unfortunately, developers are being over-surveyed and are increasingly less likely to respond, i.e., the response rate is likely to be even lower than 10%. This means that for smaller populations, sampling does not make sense: we need to contact everyone. This also highlights the importance of making the survey questions as clear as possible and as easy as possible to answer.

https://tuenl.sharepoint.com/sites/intranet-ethical-review

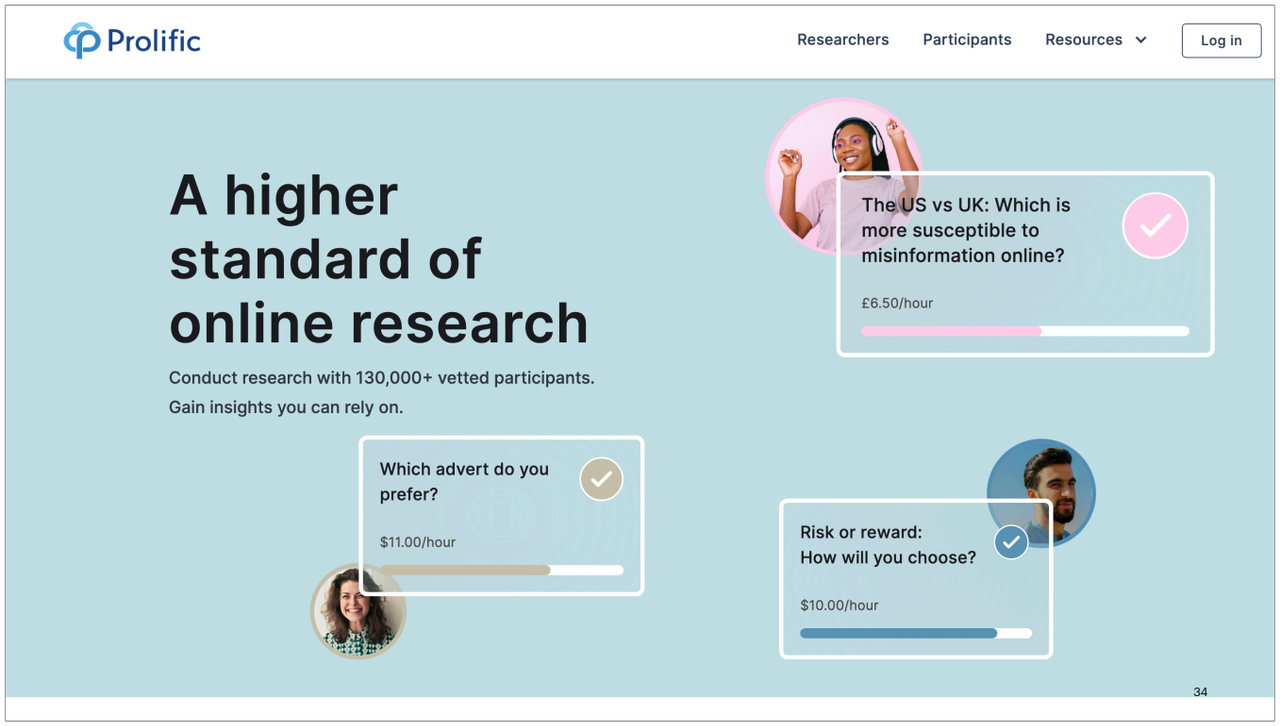
Part 4: Information about the study		
1	What are your main research questions? *Additional explanation: You need to provide at least one clear research question.*	RQs: Lecture 2
2a	Please check the box that indicates the relevant study population *Additional explanation: Please select which persons are eligible for your study.*	☐ Students ☐ General healthy population ☐ General population with specific feature, e.g., pregnancy, specifically ..... ☐ Patients, specifically ..... ☐ Other, specifically .....
2b	Age category of participants	☐ Younger than 12 years of age ☐ Older than 11 and younger than 16 years of age ☐ 16 years or older
3	Description of the research method (select all that applies)	☐ (Semi-structured) interviews ☐ Surveys

Strategies: Lecture 2
Interviews and Surveys: Today

	Additional explanation: Please specify your research method. Note that you need to provide information about the research method in an additional file that you attach to the ERB form. E.g., for interviews you provide the interview questions, for surveys you provide the survey questions, etc.	☐ Group workshops/roundtable discussios ☐ Diary studies ☐ Behavioral observations ☐ Building sensor data ☐ Wearable device (e.g. Fitbit watch, on-skin sensors) ☐ User testing ☐ Pilot study ☐ GPS tracking/location data ☐ Living Lab ☐ Other, namely .....
4	Description of the measurements and/or stimuli/treatments *Additional explanation: Think about your outcome measures and the variables you will be collecting and describe them in a way such that another person understands what the participant will experience. For example: Participants will perform task A and see pictures from database B, and we measure validated Scale 1.*	Interview guide, survey questionnaire
5	Describe and justify the number of participants you need for this study. Also justify the number of observations you need, taking into account the risks and benefits. *Additional explanation: Think about if you need 3 or 30 participants for example, and why? Do they need to provide their input once, or several times, and why? If relevant, specify the duration of the study per participant and the compensation that is needed for the study.*	How many people?
6	Explain why your research is societally important. What benefits and harm to society may result from the study? *Additional explanation: What benefit will the results of your study have to society in general?*	Motivation: Lecture 1
7	Describe the way participants will be recruited *Additional explanation: How will you recruit participants for your study? For example, by using flyers, personal network, panels, etc.*	☐ Survey link posted online, e.g., social media platforms ☐ On campus flyers ☐ Personal network ☐ Via a company, namely ..... ☐ Via a hospital, namely ..... ☐ Via an organization ..... ☐ By a Consortium Partner, namely ..... ☐ Other, namely .....
8	Provide a brief statement of the risks you expect for the participants or others involved in the study and explain. Also take into consideration any personal data you may gather and associated privacy issues. *Additional explanation: Risks for the participants can be anything from risk of data breach to risk of safety or well-being (think about stress, extreme emotions, visual or auditory discomfort). Describe these possible risks and describe the way these risks are mitigated.*	Risks

Recruitment

The next question is how many people we want to interview



Last week we have talked about Prolific. It is, of course, not the only such site.

To: Serebrenik, A. <a.serebrenik@tue.nl>
Subject: A study about human factors on the technical debt management

Dear **Dr Alexander**,

My name is <...> and I am a PhD student at <...> For my thesis, I am examining different aspects of the software development process and the organization that can impact the effectiveness of managing technical debt. This survey is a part of a large research project aiming to propose a technical debt maturity assessment model. <...>

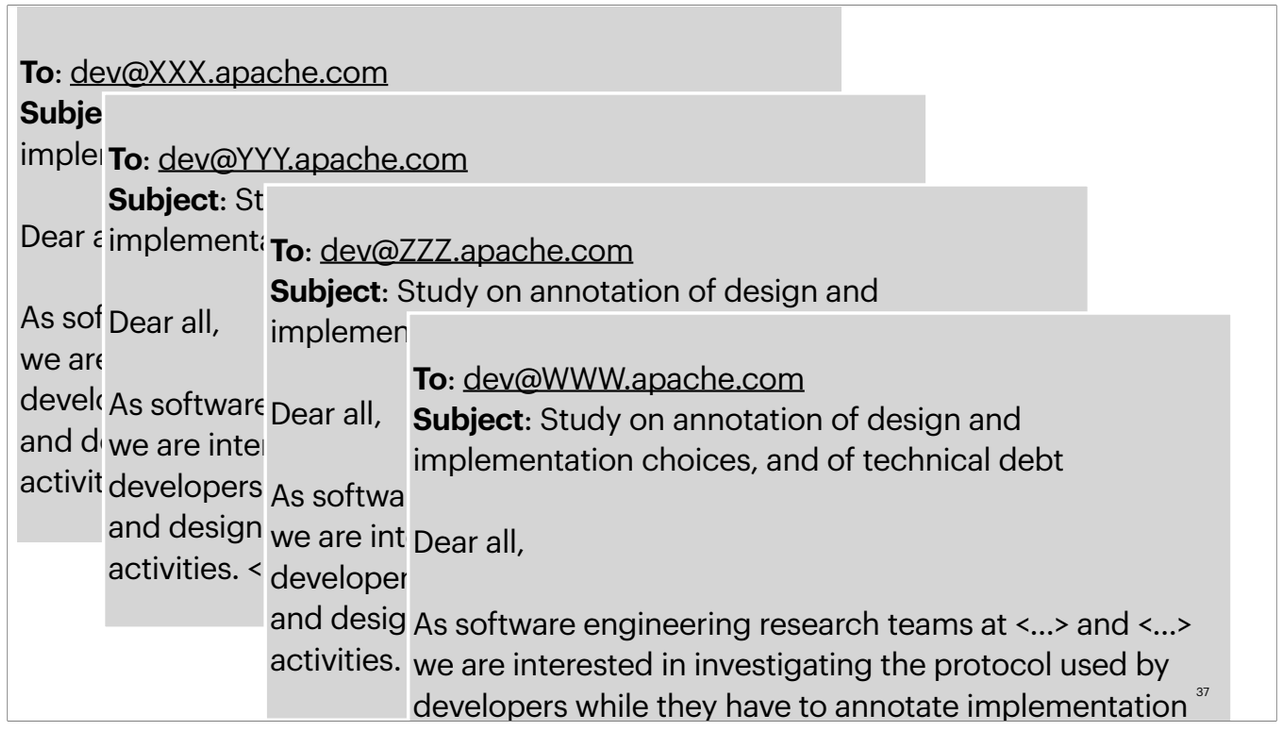
A personalised mail.
+ Feeling appreciated
“why do they ask me?”
SPAM!
GitHub: violation of the Terms of Service
Approval of IT department might be required
Sending thousands of mails…

To: dev@XXX.apache.com
Subject: Study on annotation of design and implementation choices, and of technical debt

Dear all,

As software engineering research teams at <...> and <...> we are interested in investigating the protocol used by developers while they have to annotate implementation and design choices during their normal development activities. <...>

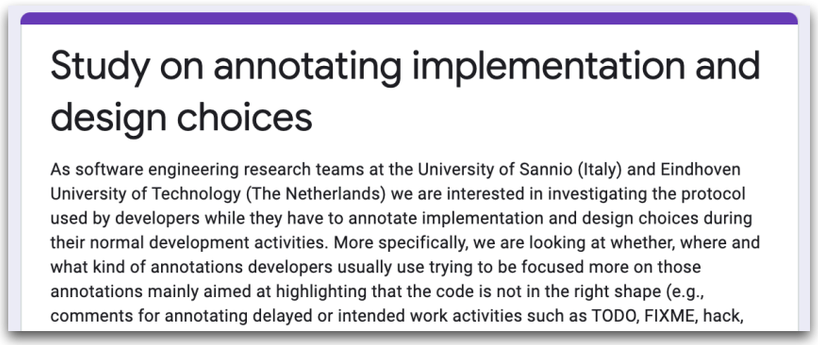
Mailing list, Reddit, FB group
Less personal
Still SPAM (of sorts) but less clear cut
Some channels explicitly prohibit this kind of messages
Ask the admins first
You cannot estimate the non-response
We don’t know how many people receive the message



Recall that we might need to contact thousands of people; usually one mailing list is too small…

If we recruit participants from different projects, we might think that all of them are software developers, but in fact they have different experiences, different expectations, and ultimately different answers to the questions we would like to get answered. Putting everything together can overshadow some groups or lead to wrong conclusions.

Hence, if we recruit participants from different mailing lists/FB groups/channels we need to keep them separate such that we can later on analyse the data and decide whether the groups of responses from different channels are **different**, and should be reported separately, or whether they are the same.



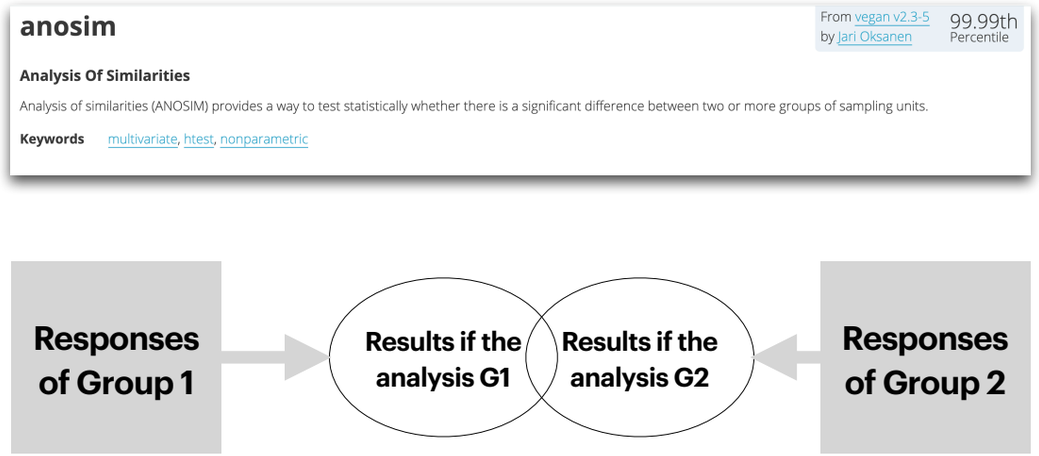
https://forms.gle/u18***

https://forms.gle/Nxd***

https://forms.gle/Pbx***

https://forms.gle/hcj***

How does this work in practice? Copy the same survey questions multiple times such that every project gets its own copy of the survey. In this way we can be sure whether the answers are from the developers of project A or project B.



As I said, this will allow us to analyse the data and decide whether the groups of responses from different channels are different, and should be reported separately, or whether they are the same. If our data is quantitative we can use statistical techniques such as **anosim**. If the data is qualitative, we analyse each one of the groups separately and compare the insights.

You should define the analysis strategy before sending out the surveys!

https://tuenl.sharepoint.com/sites/intranet-ethical-review

Part 4: Information about the study		
1	What are your main research questions? *Additional explanation: You need to provide at least one clear research question.*	RQs: Lecture 2
2a	Please check the box that indicates the relevant study population *Additional explanation: Please select which persons are eligible for your study.*	☐ Students ☐ General healthy population ☐ General population with specific feature, e.g., pregnancy, specifically ..... ☐ Patients, specifically ..... ☐ Other, specifically .....
2b	Age category of participants	☐ Younger than 12 years of age ☐ Older than 11 and younger than 16 years of age ☐ 16 years or older
3	Description of the research method (select all that applies)	☐ (Semi-structured) interviews ☐ Surveys

Strategies: Lecture 2
Interviews and Surveys: Today

	Additional explanation: Please specify your research method. Note that you need to provide information about the research method in an additional file that you attach to the ERB form. E.g., for interviews you provide the interview questions, for surveys you provide the survey questions, etc.	☐ Group workshops/roundtable discussios ☐ Diary studies ☐ Behavioral observations ☐ Building sensor data ☐ Wearable device (e.g. Fitbit watch, on-skin sensors) ☐ User testing ☐ Pilot study ☐ GPS tracking/location data ☐ Living Lab ☐ Other, namely .....
4	Description of the measurements and/or stimuli/treatments *Additional explanation: Think about your outcome measures and the variables you will be collecting and describe them in a way such that another person understands what the participant will experience. For example: Participants will perform task A and see pictures from database B, and we measure validated Scale 1.*	Interview guide, survey questionnaire
5	Describe and justify the number of participants you need for this study. Also justify the number of observations you need, taking into account the risks and benefits. *Additional explanation: Think about if you need 3 or 30 participants for example, and why? Do they need to provide their input once, or several times, and why? If relevant, specify the duration of the study per participant and the compensation that is needed for the study.*	How many people?
6	Explain why your research is societally important. What benefits and harm to society may result from the study? *Additional explanation: What benefit will the results of your study have to society in general?*	Motivation: Lecture 1
7	Describe the way participants will be recruited *Additional explanation: How will you recruit participants for your study? For example, by using flyers, personal network, panels, etc.* Recruitment	☐ Survey link posted online, e.g., social media platforms ☐ On campus flyers ☐ Personal network ☐ Via a company, namely ..... ☐ Via a hospital, namely ..... ☐ Via an organization ..... ☐ By a Consortium Partner, namely ..... ☐ Other, namely .....
8	Provide a brief statement of the risks you expect for the participants or others involved in the study and explain. Also take into consideration any personal data you may gather and associated privacy issues. *Additional explanation: Risks for the participants can be anything from risk of data breach to risk of safety or well-being (think about stress, extreme emotions, visual or auditory discomfort). Describe these possible risks and describe the way these risks are mitigated.*	Risks

The next question is how many people we want to interview

https://upload.wikimedia.org/wikipedia/commons/c/c0/Talk_to_the_hand.jpg

Even if we selected a person for an interview we cannot just interview them. We need to have their consent, and this is not just “yes, yes” but a formal document signed by the interviewee. This is called **informed consent.** Informed consent implies that prospective research participants should be given as much information as might be needed to make an informed decision about whether or not they wish to participate in a study.

Participants should know what they are getting themselves into, what are the possible risks, whether their participation is voluntary, how are they going to be compensated, what happens if they want to withdraw from the study, how is their data going to be collected/accessed/used/destroyed. Data: privacy, but also what happens if the recording of the interview is stolen?!

So why do we bother about consent forms? This is not only an ethical imperative but also a legal requirement of General Data Protection Regulations, in Dutch: Algemene verordening gegevensbescherming (AVG)

LAWFUL BASE FOR DATA COLLECTION

CONSENT

CONTRACT

LEGAL OBLIGATION

VITAL INTEREST

PUBLIC TASK

LEGITIMATE INTEREST

According to GDPR, there are six lawful bases for data collection.
(a) **Consent**: the individual has given clear consent for you to process their personal data for a specific purpose. This is what we have discussed so far. In a way, it is the easiest requirement to satisfy in case of an interview or a survey. One should be mindful however, that consent can be revoked. This means that if an interviewee/survey respondent contacts you and asks you to exclude their data from your study you should do this (unless the data is anonymised and you cannot identify the respondent's contribution).

LAWFUL BASE FOR DATA COLLECTION

✓CONSENT

CONTRACT

LEGAL OBLIGATION

VITAL INTEREST

PUBLIC TASK

LEGITIMATE INTEREST

(b) **Contract**: the processing is necessary for a contract you have with the individual, or because they have asked you to take specific steps before entering into a contract.

LAWFUL BASE FOR DATA COLLECTION

✓CONSENT

~~CONTRACT~~

LEGAL OBLIGATION

VITAL INTEREST

PUBLIC TASK

LEGITIMATE INTEREST

(c) **Legal obligation**: the processing is necessary for you to comply with the law (not including contractual obligations). This is also not applicable for us.

LAWFUL BASE FOR DATA COLLECTION

✓CONSENT

~~CONTRACT~~

~~LEGAL OBLIGATION~~

VITAL INTEREST

PUBLIC TASK

LEGITIMATE INTEREST

(d) **Vital interests**: the processing is necessary to protect someone’s life. We do not really protect lives of humans, or at least I am not aware of empirical SE applications that fit this description.

LAWFUL BASE FOR DATA COLLECTION

✓ CONSENT

~~CONTRACT~~

~~LEGAL OBLIGATION~~

~~VITAL INTEREST~~

PUBLIC TASK

LEGITIMATE INTEREST

(e) **Public task**: the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law. This argument can be used: in general, conducting research or educational activities is part of the tasks of the university. This is also the argument we will put forward when informed consent cannot be obtained, for example when we analyse publicly available data such as GitHub repositories or Stack Overflow questions and answers. However, this is not easy. According to GDPR, the processing must be necessary. If you could reasonably perform your tasks in a less intrusive way, this lawful basis does not apply. This means that if consent can be obtained, it must be obtained!

LAWFUL BASE FOR DATA COLLECTION

✓ CONSENT

~~VITAL INTEREST~~

~~CONTRACT~~

✓ PUBLIC TASK

~~LEGAL OBLIGATION~~

~~LEGITIMATE INTEREST~~

(f) **Legitimate interests**: the processing is necessary for your legitimate interests or the legitimate interests of a third party, unless there is a good reason to protect the individual's personal data which overrides those legitimate interests. (This cannot apply if you are a public authority processing data to perform your official tasks.) This is tricky since this base explicitly excludes public institutions…

(a) **Consent**: the individual has given clear consent for you to process their personal data for a specific purpose. This is what we have discussed so far. In a way, it is the easiest requirement to satisfy in case of an interview.

(e) **Public task**: the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law. This argument can be used: in general, conducting research or educational activities is part of the tasks of the university. This is also the argument we will put forward when informed consent cannot be obtained, for example when we analyse publicly available data such as GitHub repositories or Stack Overflow questions and answers. However, this is not easy. According to GDPR, the processing must be necessary. If you could reasonably perform your tasks in a less intrusive way, this lawful basis does not apply. This means that if consent can be obtained, it must be obtained!

QUESTION

An individual gives a phone number to the website of an attorney to be contacted about a potential case. The attorney has a right to use that phone number and contact the individual based on the following lawful base:

(A) CONSENT

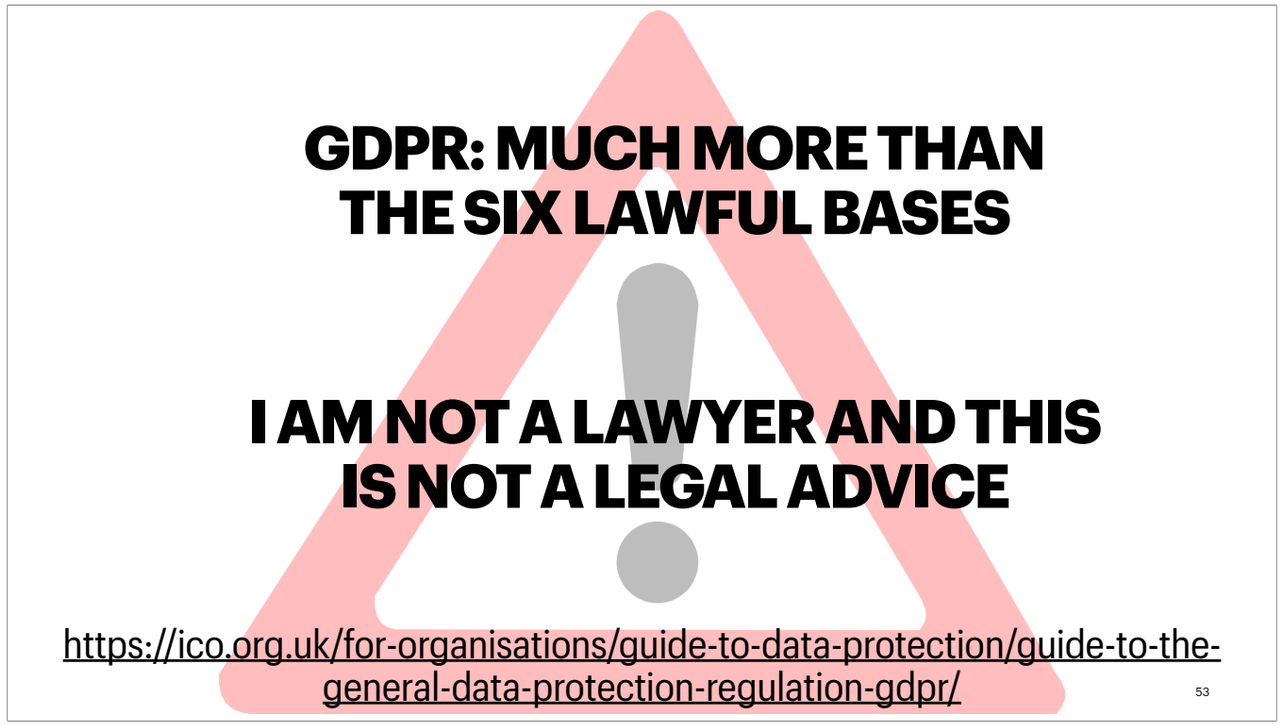
(B) CONTRACT

(C) LEGAL OBLIGATION

(D) PUBLIC TASK

HTTPS://WWW.TERMSFEED.COM/BLOG/GDPR-LAWFUL-BASIS-PROCESSING/

(B) Contract. This is because it is implied that this was the reason why the individual gave out their phone number. This also covers the situation when someone has asked you to take specific steps before entering into a contract.
This is not a legal obligation since there is no legal requirement for an attorney to contact their prospective customers.



Reminder: Dutch: Algemene verordening gegevensbescherming (AVG)
consent, contract, public task, vital interest, legitimate interest or legal requirement

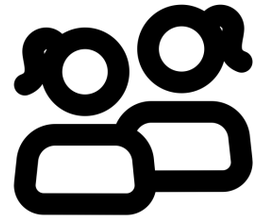
FOCUS: INTERVIEWS

https://libreshot.com/wp-content/uploads/2016/07/microphone.jpg

WHO WILL CONDUCT AN INTERVIEW?

1 OR 2

Created by Gilbert Bages
from Noun Project



Created by Viktor Ostrovsky
from Noun Project

At the first sight the question seems trivial. The researcher should conduct the interview! However, even here there are multiple options.

- One interviewer is easier to arrange. However, it is difficult to ask questions and take notes at the same time - and the interviewer always takes notes, even in presence of audio/video recording! Two interviewers can notice different things and ask different questions. However, they might intimidate interviewees.
- The interviewee should feel comfortable. This means that if there are several researchers they should be able to choose, for example, the one that is located in the timezone they are most comfortable with.
- The third concern that should be taken into account is the similarity between the interviewer and the interviewee. Similarity of previous experiences helps to establish trust and mutual understanding: it is known that when talking about more social subjects women are more likely to disclose more to a woman-interviewer than to a man-interviewer.

SUCCESSFUL INTERVIEWER IS...

Knowledgeable
Critical
Structuring
Remembering
Gentle
Interpreting
Clear
Balanced
Sensitive
Open
Ethically sensitive
Adaptable
Steering

- **Knowledgeable**: is thoroughly familiar with the focus of the interview (pilot interviews to become knowledgeable!). At the same time interviewer should position themselves as a learner rather than as an expert.
- **Structuring**: gives purpose for interview: asks whether interviewee has questions.
- **Clear**: asks simple, easy, short questions; no jargon.
- **Gentle**: lets people finish; gives them time to think; tolerates pauses.
- **Sensitive**: listens attentively to what is said and how it is said; is empathetic in dealing with the interviewee.
- **Ethically sensitive**: is sensitive to the ethical dimension of interviewing, ensuring the interviewee appreciates what the research is its purposes, and that his or her answers will be treated confidentially.
- **Steering**: knows what he or she wants to find out.
- **Critical**: is prepared to challenge what is said—for example, dealing with inconsistencies in interviewees' replies.
- **Remembering**: relates what is said to what has previously been said.
- **Interpreting**: clarifies and extends meanings of interviewees' statements, but without imposing meaning on them.
- **Balanced**: does not talk too much, which may make the interviewee passive, and does not talk too little, which may result in the interviewee feeling they are not talking along the right lines.
- **Open**: responds to what is important to interviewee and is flexible.
- **ADAPTABLE**: each person is different and you have to adapt your interviewing style and behaviour

Alan Bryman, Social Research Methods. 5th Ed. Oxford University Press, 2016

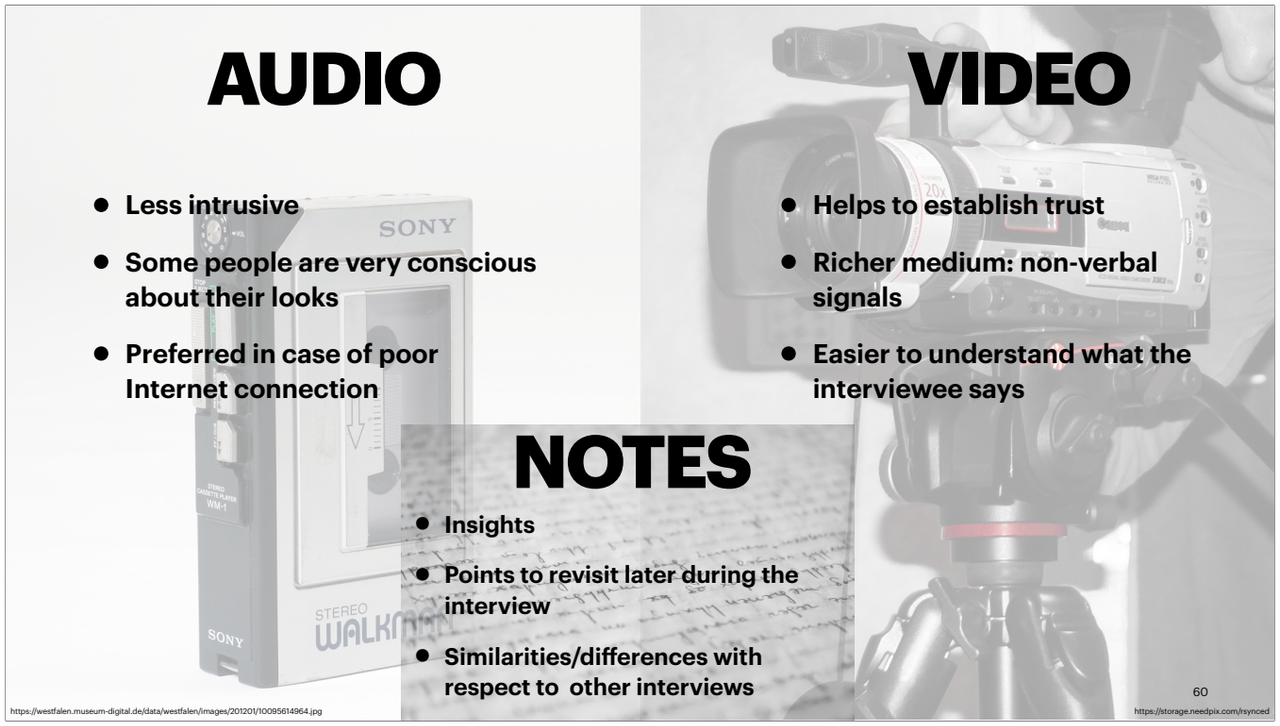
https://www.spacebase.com/en/venue/the-coffee-house-coworking-1961/

Ideal space for an interview? Why/why not?
Glass wall. Computer and copy machine next to this room might attract passers by,
For internal anonymity it should not be obvious to colleagues that an interview has taken place, nor what was mentioned.

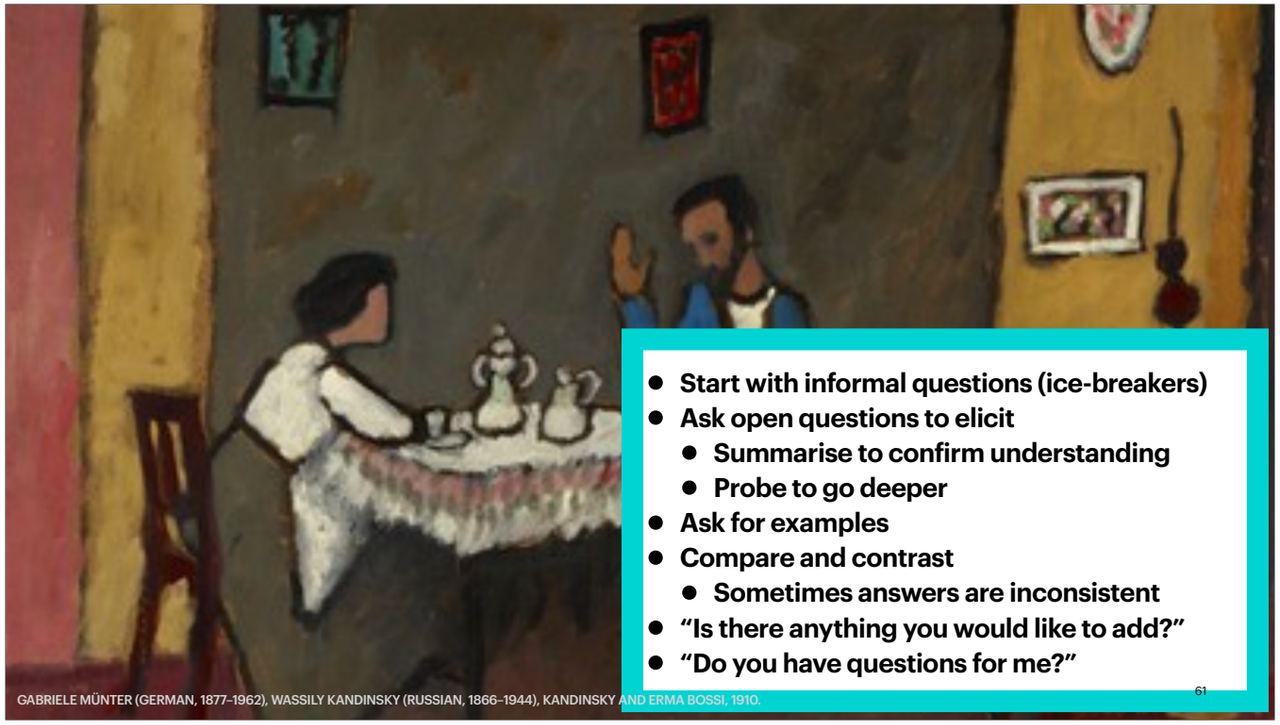
GABRIELE MÜNTER (GERMAN, 1877–1962), WASSILY KANDINSKY (RUSSIAN, 1866–1944), KANDINSKY AND ERMA BOSSI, 1910.

This is not an interview but a painting of Gabriele Münter and Wassily Kandinsky representing a conversation between Kandinsky and another painter, Erma Bossi. We see here several important elements of the interview.
Erma's body language: leaning forward, active listening.
Presumably coffee at the table: this can help creating a more relaxed sphere.
It is a quite space, where nobody else is present. Recall that it is important for interviews (a) to ensure that there are no interruptions, and (b) that the interviewee feels safe to divulge information.
The only thing which is, of course, missing here, is the **recording**!

Recording comes predominantly in two forms: audio and video. Of course, we can also take notes but this is rarely being used as the sole recording mechanism; notes taking is often used in combination with audio/video recording



Recording comes in two forms: audio and video. Of course, we can also take notes but this is rarely being used as the sole recording mechanism; notes taking is often used in combination with audio/video recording

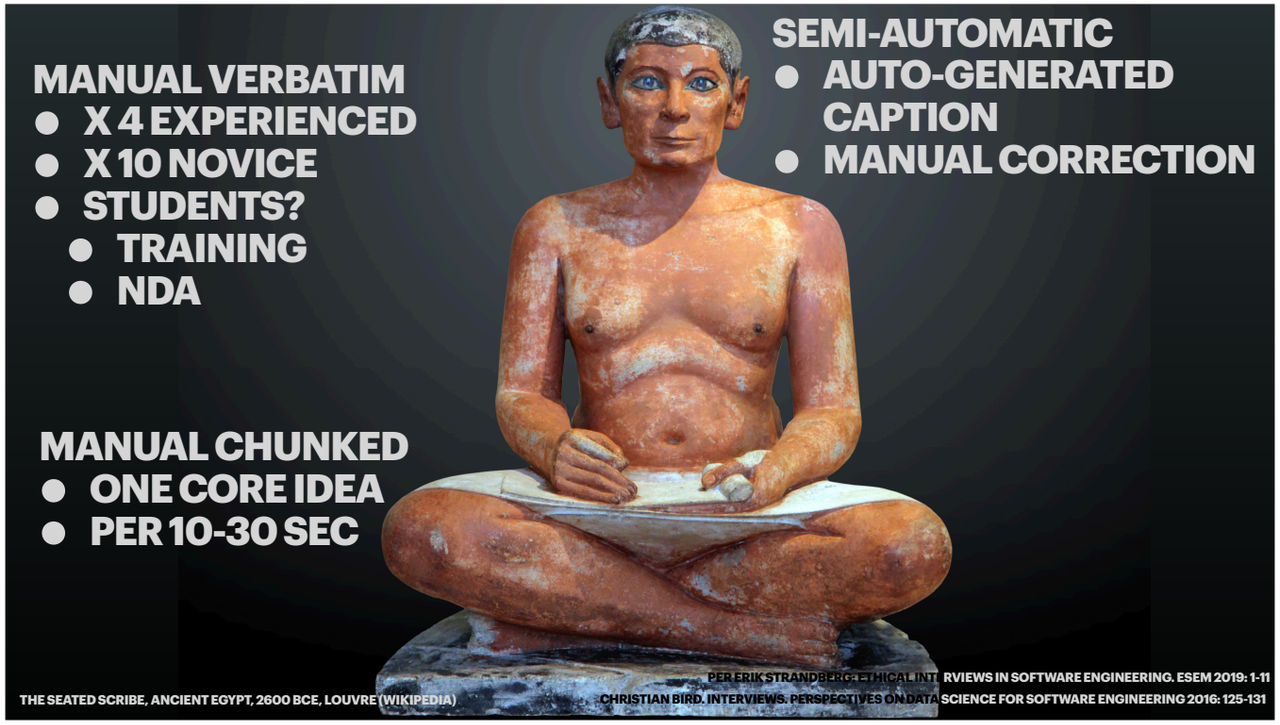


GABRIELE MÜNTER (GERMAN, 1877–1962), WASSILY KANDINSKY (RUSSIAN, 1866–1944), KANDINSKY AND ERMA BOSSI, 1910.

So finally we got our ERB approval and we can start conducting our interviews. During the interview, if it is a semi-structured or an unstructured interview you might consider the following tips.

GABRIELE MÜNTER (GERMAN, 1877–1962), WASSILY KANDINSKY (RUSSIAN, 1866–1944), KANDINSKY AND ERMA BOSSI, 1910.

The product of an interview, of course an audio or a video recording but for the following steps of the analysis we need text, for example, because we want to see whether several interviewees discuss the same problem.



The problem is, of course, that we need to transcribe the video/audio…

However, the transcription often is not limited to literal recording of what the interviewees have said but also to removing parts that can lead to identification of the interviewees. Do not forget that we need to protect privacy of the interviewees and disclose only as little as possible.

The **what** question is easy to answer in generic terms: everything that can lead to identification of the interviewee. However, it is not that easy to know what can this be. Beware that combination of the OSS project name or company name, role (dev/tester) or experience and gender can be enough for de-anonymization. At the same time, reporting gender of the participants is seen as a good practice since it (a) shows that the researchers did not merely study the majority group (men), and (b) gender might affect the variables of interest when it comes to social aspects or program comprehension.

The **how** question is more tricky. The most common strategy is just give people numbers, like P1, …, P35. This to large extent allows one to anonymise as much as possible. However, this also hides information that we might find to be useful, e.g., if we interview developers working with different programming languages we might like to preserve this information and distinguish between, for example, JavaScript-developers and C-developers; or between developers belonging to different groups or teams, e.g., P2.3 is the third interviewee in team 2. We can also consider giving pseudonyms if we want to preserve some more personal information, e.g., Wei can be a pseudonym of a Chinese software developer. All theses techniques assume that every participant has a pseudonym and that this pseudonym is unique.

However, even this is not necessarily the case. Saunders et al. use multiple pseudonyms for the same person when one extract of a transcript will not identify an interviewee, but when the combination of two might. They also mention approaches where multiple interviewees are assigned to one pseudonym in order to create a more representational story.

65

https://upload.wikimedia.org/wikipedia/commons/d/d4/Locked_computer_laptop.jpg

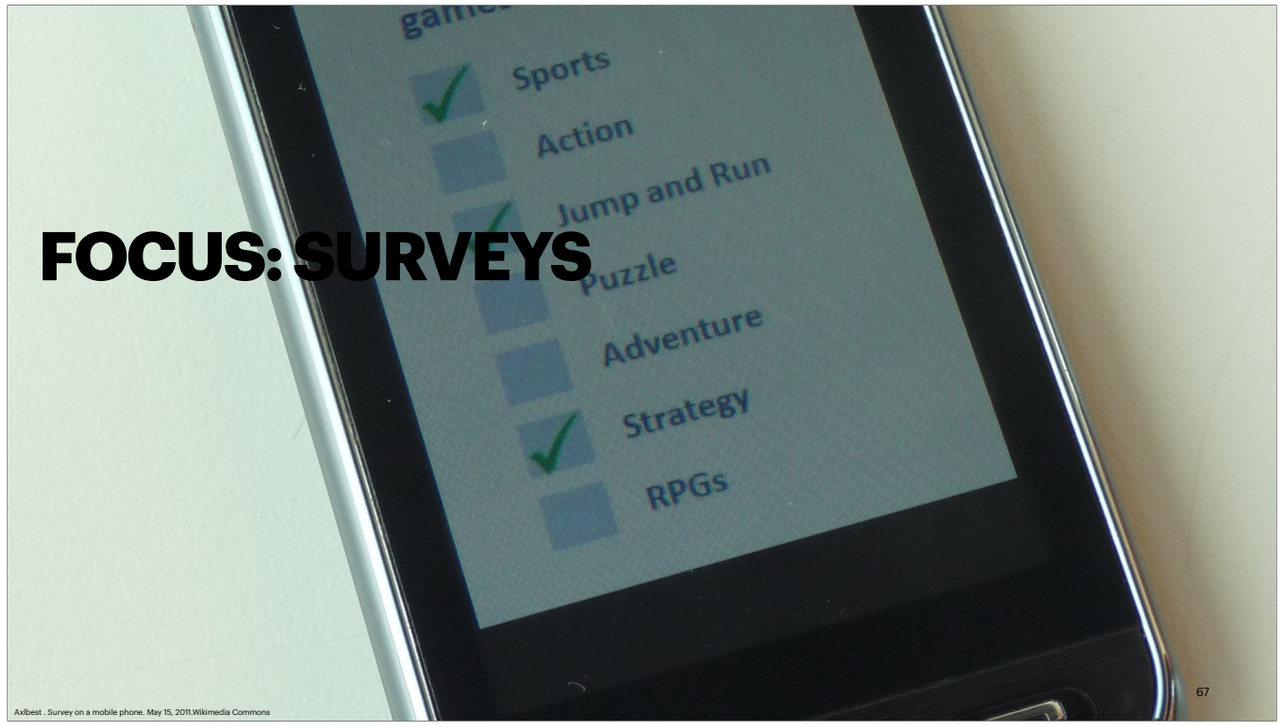
When it comes to storing data we need to clearly separate personally identifiable information from the recordings such that if our storage gets stolen privacy of the participants is not affected.

https://pxhere.com/en/photo/1575599

Finally, the last step of the data collection process in interviewing is confirmation. We would like to check with the interviewees (a) that we did not introduce any mistakes when transcribing the interview or anonymising the data, and (b) whether they would like to elaborate more on certain topics or delete some of the information they have provided. Again, comfort and safety of the interviewee are the primary goal.

Anecdote: in one of the sociological studies I have recently read one of the interviewees has insisted that his interview will has his real name rather than a pseudonym. The interviewer has explained the risks associated with this decision, the interviewee has confirmed that he has understood them and ultimately the publication included the real name of this interviewee.

FOCUS: SURVEYS



Axlbest . Survey on a mobile phone. May 15, 2011.Wikimedia Commons

Focus on surveys

WHAT ARE WE ASKING ABOUT?

DEMOGRAPHICS

FACTS

EXPERIENCES

ATTITUDES

BELIEFS

NORMS

So what are we asking about?
We will talk about demographics in a moment

What (typically)	Question	Answers	Comments
facts	What is your role in the organisation?	[redacted]	Reuse options from the SO/GH surveys
experiences	Please describe a change you have been reviewing that has confused you		Calls for manual analysis of the text
attitudes	Continuous Integration is beneficial for our project		Use a Likert scale
beliefs	How often do you feel confused when reviewing code changes due to lack of documentation?		Anchor the options instead of “frequently”
norms	How many project members believe that allocation of reviewing tasks is fair?		See Cristina Bicchieri “Norms in the Wild”

It is extremely important to be clear what do we want to know. All types of questions can be useful but they will trigger different responses and it is important to understand what we actually want to know.

What (typically)	Question	Answers	Comments
facts	What is your role in the organisation?	Multiple choice	Reuse options from the SO/GH surveys
experiences	Please describe a change you have been reviewing that has confused you	Usually open	Calls for manual analysis of the text
attitudes	Continuous Integration is beneficial for our project	Disagree...Agree	Use a Likert scale
beliefs	How often do you feel confused when reviewing code changes due to lack of documentation?	not at all, less than once a month, once a month, once a week, once a day, and more than once a day	Anchor the options instead of “frequently”
norms	How many project members believe that allocation of reviewing tasks is fair?	Percentage or Likert scale	See Cristina Bicchieri “Norms in the Wild”

It is extremely important to be clear what do we want to know. All types of questions can be useful but they will trigger different responses and it is important to understand what we actually want to know.

Question 1: Tools Used
What tools do you use to search source code? Check all that apply.

grep, fgrep, etc.	[ ]
find or "File Find"	[ ]
editor e.g. vi, emacs, edit	[ ]
integrated development environment e.g. MSDS	[ ]
other	[ ]

Please specify: _____

FACTS

Question 4: Typical Usage Situations
Describe one or more situations when you needed to search source code. What did you use to find it? What were you trying to find? Why did you need to find it?

EXPERIENCES

How useful is it to search source code when:

	Not at all useful				Very useful
doing low-level design?	1	2	3	4	5
writing new code?	1	2	3	4	5
testing?	1	2	3	4	5
understanding old code?	1	2	3	4	5
repairing bugs/defects?	1	2	3	4	5
adding a new feature to old software?	1	2	3	4	5
improving performance?	1	2	3	4	5
inspecting and reviewing code?	1	2	3	4	5
writing documentation?	1	2	3	4	5
maintaining documentation?	1	2	3	4	5

ATTITUDES

Question 7: Time With Source Code Written By Others

Of your total time spent working with source code, what percentage of that time is spent working on source code written by other people?

0-20%	[ ]
21-40%	[ ]
41-60%	[ ]
61-80%	[ ]
81-100%	[ ]

BELIEFS

SUSAN ELLIOTT SIM, CHARLES L. A. CLARKE, RICHARD C. HOLT: ARCHETYPAL SOURCE CODE SEARCHES: A SURVEY OF SOFTWARE DEVELOPERS AND MAINTAINERS. IWPC 1998: 180-187

A survey commonly combines questions of different categories. One usually starts with asking about facts and then moves to attitudes and experiences, and then proceeds to beliefs. Expectations are special kinds of beliefs. Norms are studied less often in SE.

QUESTION

Afzal et al. have conducted a survey of robotics engineers. Their survey included the following question:

How did you use software-based simulation as part of your test automation?

(A) FACTS

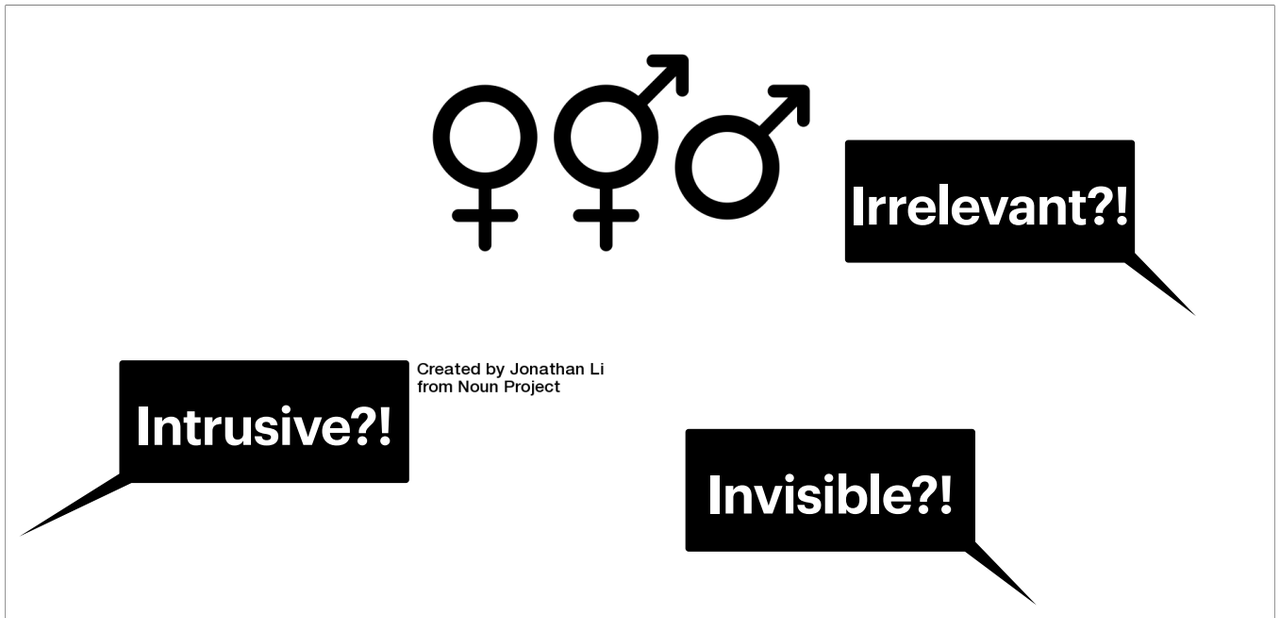
(B) EXPERIENCES

(C) BELIEFS

(D) BELIEFS

DEMOGRAPHICS

The demographics questions tend to be common and shared between different surveys. However, even in this case, asking an appropriate question is not always easy.

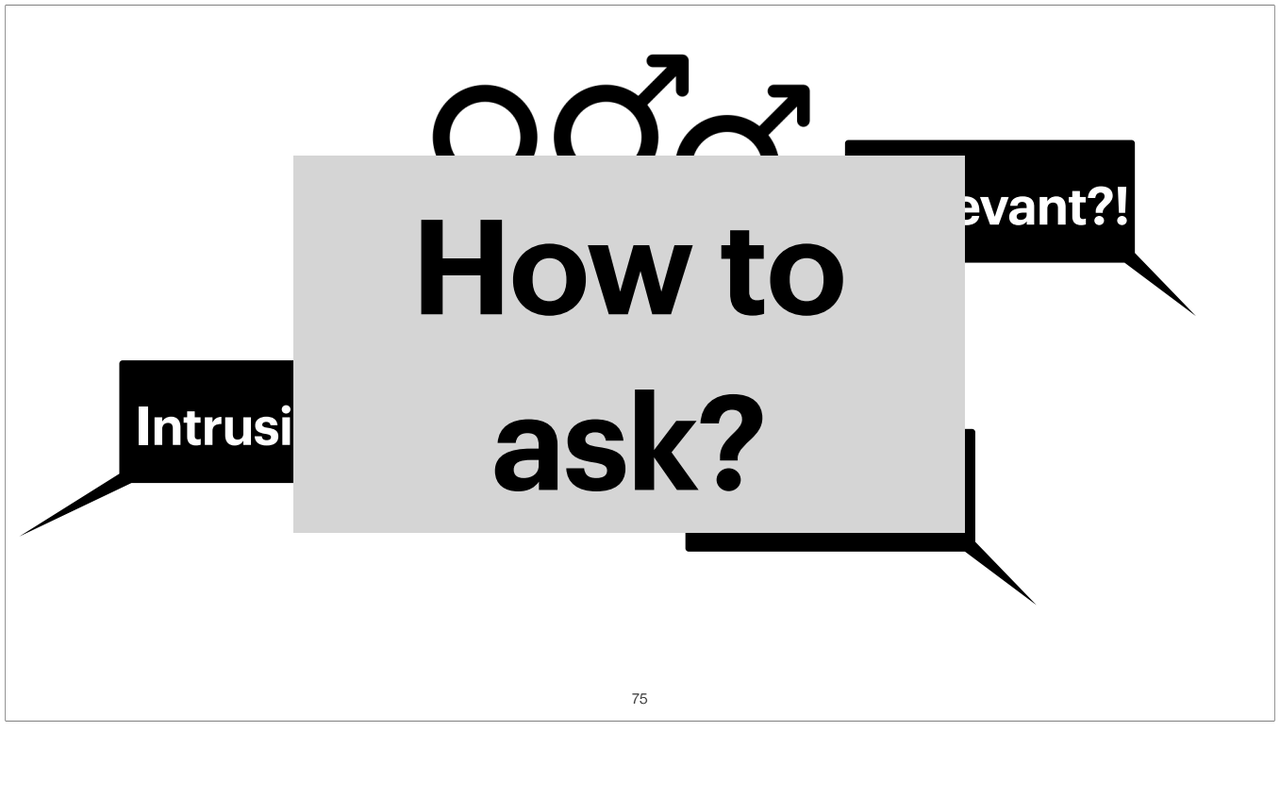


Example of a demographics question: gender. Gender is a very common element in surveys, you have probably seen it many times. However, as with any question we need to know whether we need to ask about gender.

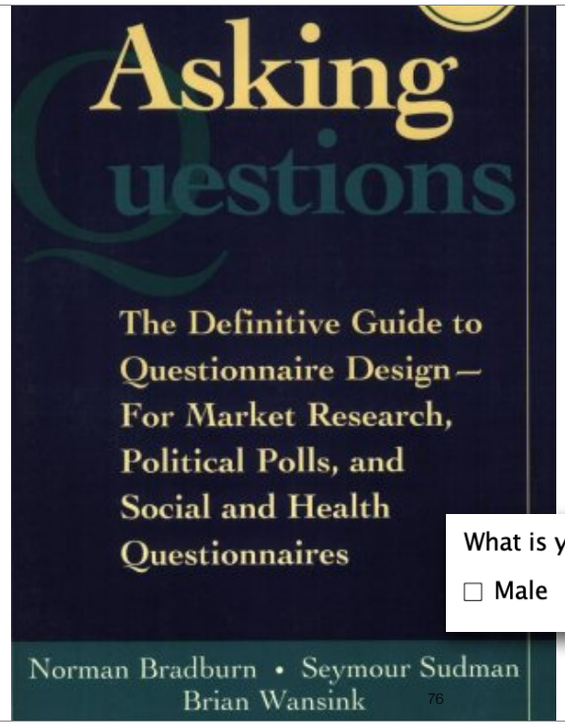
Asking about gender can be perceived as **intrusive**. Don’t forget gender is privacy sensitive and should be treated as such. Respondents might be hiding their gender on purpose, e.g., many women-developers prefer not to disclose their gender due to safety concerns.

This can also be seen as **irrelevant**: for questions related to communication/collaboration between developers or code comprehension one could reasonably argue that one can expect gender-related differences. However, this is not always the case.

Not asking about gender, however, might render women **invisible** overrepresenting the opinions of men. Even if researchers do not expect gender-related differences there still might be some!



Let’s assume that we have decided to ask about gender. It is still not clear how to ask.



2004

What is your/(NAME's) sex?

☐ Male ☐ Female

A highly influential guide on questionnaire design (which I do recommend) published in 2004 recommends the two check boxes. Sex vs gender

1.42%+0.92%

However, recent survey of Stack Overflow (2021) indicates that 1.42% of software developers identify as Non-binary, genderqueer, or gender non-conforming and 0.92% prefer to self describe.

1.42%+0.92%

4%

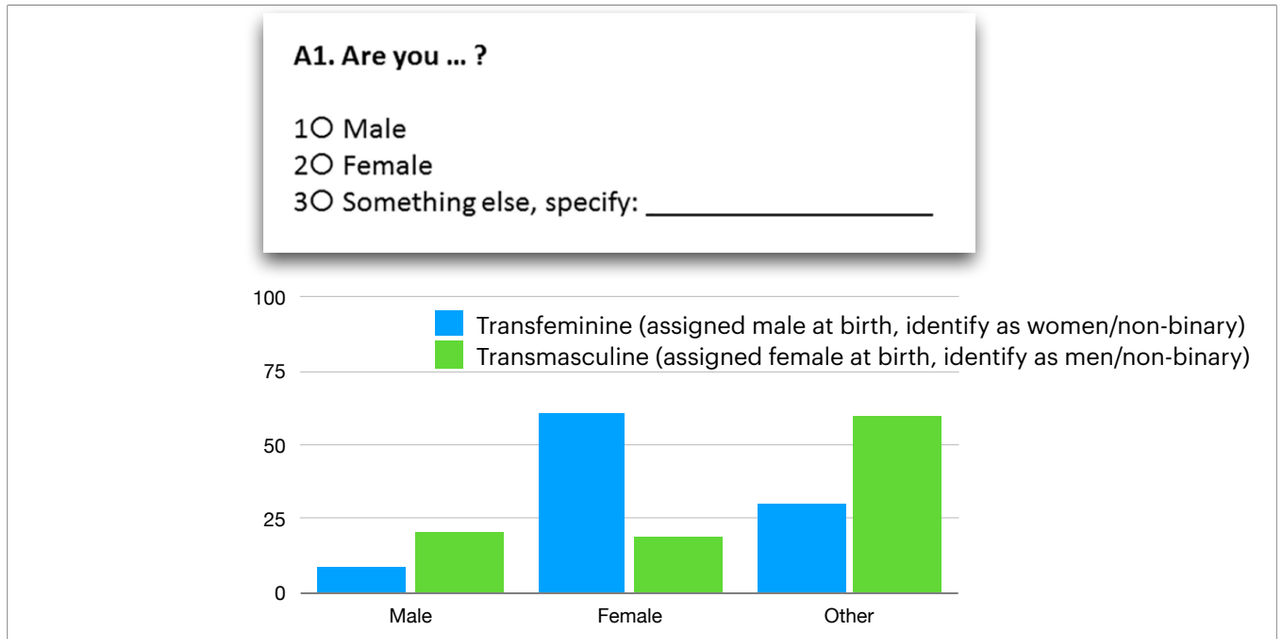
And Stack Overflow is not the only source

A1. Are you … ?

1○ Male
2○ Female
3○ Something else, specify: _____

Bauer GR. Making sure everyone counts: considerations for inclusion, identification, and analysis of transgender and transsexual participants in health surveys. In: Coen S, Banister E, editors. What a difference sex and gender make. Vancouver: Institute of Gender and Health, Canadian Institutes of Health Research; 2012. pp. 59–67.

A better approach has been advocated by Greta Bauer in 2012. Unfortunately, my favourite survey platform, Google forms, does not seem to support this combination of an open answer in a multiple choice question unless it is the “other”.



Bauer GR. Making sure everyone counts: considerations for inclusion, identification, and analysis of transgender and transsexual participants in health surveys. In: Coen S, Banister E, editors. What a difference sex and gender make. Vancouver: Institute of Gender and Health, Canadian Institutes of Health Research; 2012. pp. 59–67.

Bauer GR, Braimoh J, Scheim AI, Dharma C (2017) Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. PLoS ONE 12(5): e0178043.

However, while this question was clear and easily answered by cisgender participants, it did not clearly identify birth-assigned sex or gender identity. In the interviews this item was cognitively taxing for trans interview participants, who tried to figure out exactly what the researchers were asking, and reached different conclusions.

Do you consider yourself to be transgender?
() Yes
() No
() Questioning

Do you consider yourself to be gender non-conforming, gender diverse, gender variant, or gender expansive?
() Yes
() No
() Questioning

Are you intersex?
() Yes
() No
() I don't know

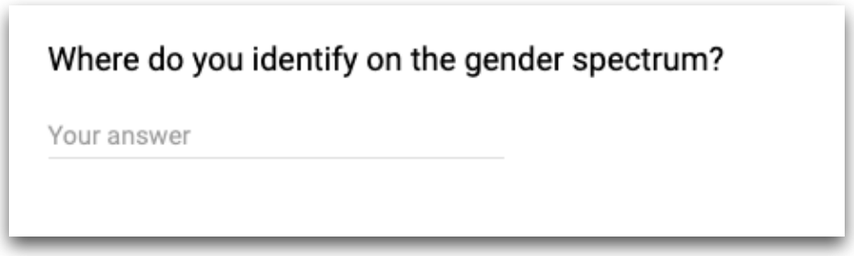
Where do you identify on the gender spectrum (check all that apply)?
[] Woman
[] Demi-girl
[] Man
[] Demi-boy
[] Non-binary
[] Demi-non-binary
[] Genderqueer
[] Genderflux
[] Genderfluid
[] Demi-fluid
[] Demi-gender
[] Bigender
[] Trigender
[] Two-Spirit
[] Multigender/polygender
[] Pangender/omnigender
[] Maxigender
[] Aporagender
[] Intergender
[] Maverique
[] Gender confusion/Gender f*ck
[] Gender indifferent
[] Graygender
[] Agender/genderless
[] Demi-agender
[] Genderless
[] Gender neutral
[] Neutrois
[] Androgynous
[] Androgyne
[] Prefer not to answer
[] Self Identify: _____

Open demographics

https://drnikki.github.io/open-demographics/questions/gender.html

Good news: it separates a question about trans* and about gender non-conforming. Even more good news: woman/man instead of female/male; the former puts more stress on identity as opposed to biology. More good news: the question explicitly refers to gender identity and avoids confusion reported for the survey instrument of Bauer. And even more news: “check all that apply”, i.e., someone can be both woman and non-binary. Bad news: there are too many options and we want to keep surveys (and particularly demographic parts of the surveys) short! Even more, some of these notions might be experienced as confusing or taxing.

Maverique (pronounced mav-reek) is a specific nonbinary gender identity "characterized by autonomy and inner conviction regarding a sense of self that is entirely independent of male/masculinity, female/femininity or anything which derives from the two while still being neither without gender nor of a neutral gender." Maverique is not close to a female or male gender, and is not like a mix of them; the identity goes beyond the entire scope of the gender binary or any identities within and outside of it. **Aporagender** (from Greek apo, apor "separate" + "gender") is a nonbinary gender identity and umbrella term for "a gender separate from male, female, and anything in between (unlike Androgyne) while still having a very strong and specific gendered feeling" (that is, not an absence of gender or agender). **Neutrois** is a non-binary gender identity which is often associated with a "neutral" or "null" gender.



https://www.morgan-klaus.com/sigchi-gender-guidelines

A much better solution according to the HCI Guidelines for Gender Equity and Inclusivity is to ask an open ended question. This might be difficult for us as researchers to process (code) but most of software engineering surveys are relatively small, a couple of hundreds of responses.

QUESTION

Give an example of a common demographic question (other than gender) that similarly to gender might be considered **intrusive** or **irrelevant**, but not asking this question might render the minority group **invisible**?

Pretty much anything, for example age (minority group would be older developers or very young developers)

Experience

How to ask?

Another example of a demographics question: experience. Experience is clearly relevant to how people read code (novices read it line by line, experienced developers look for patterns) and interact with each other.

Do you have experience in Java programming?

◯ Yes ◯ No

How many years of Java programming experience do you have?

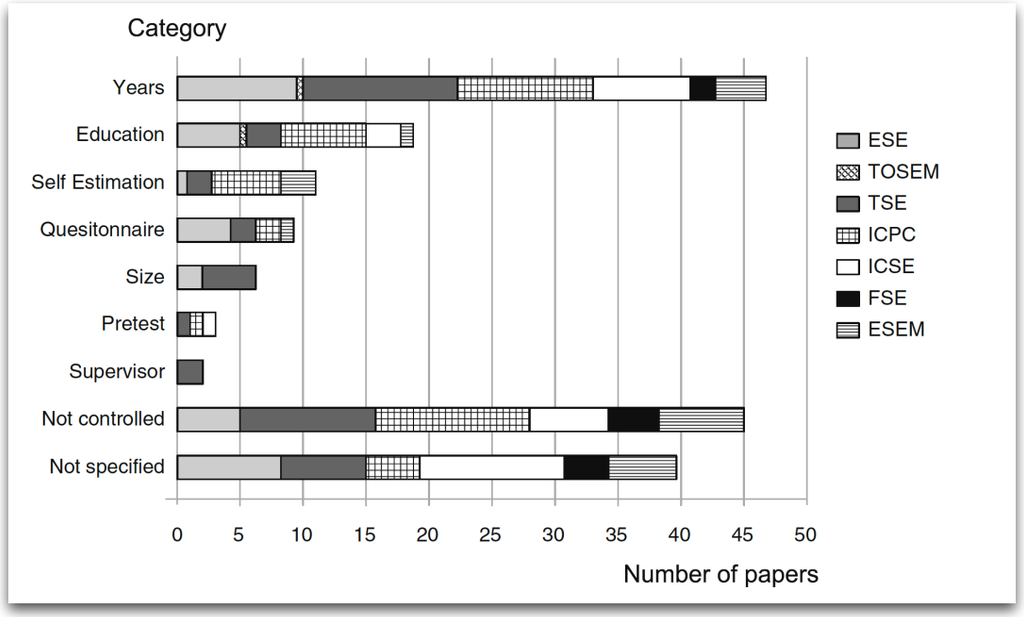
9

How much experience do you have in Java programming?

I got my first Java certificate in 2009, and I have been working with Java at companies since 2011. I am familiar with Struts and Hibernate but not with Google web toolkit.

Inspired by https://www.slideshare.net/mendezfe/surveys-in-software-engineering

Here are three examples of how one can ask about experience. Which one would you prefer?



JANET SIEGMUND, CHRISTIAN KÄSTNER, JÖRG LIEBIG, SVEN APEL, STEFAN HANENBERG: MEASURING AND MODELING PROGRAMMING EXPERIENCE. EMPIR. SOFTW. ENG. 19(5): 1299-1334 (2014)

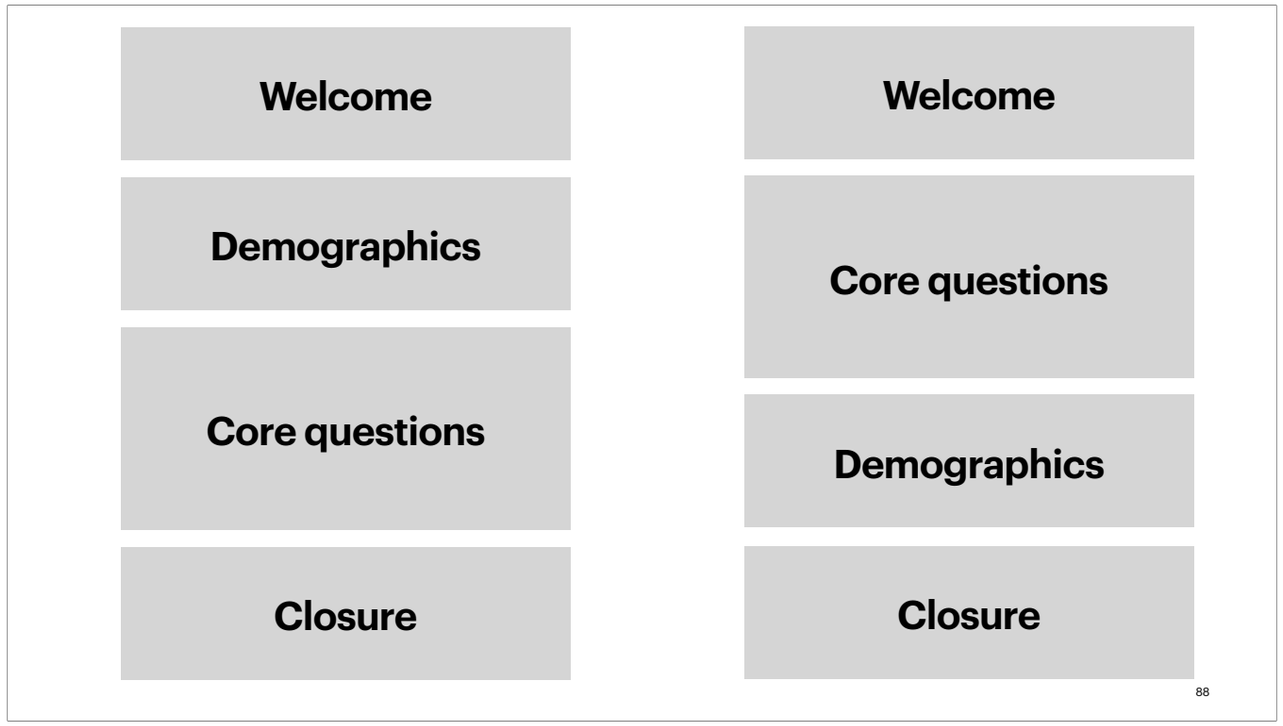
1. In many papers (47), the **years** a participant was programming at all or programming in a company or certain language
2. **Education** includes information such as the level of education (e.g., undergraduate or graduate student) or the grades of courses.
3. In twelve papers, participants were asked to estimate their experience them**selves**. For example, Bunse let his participants estimate their experience on a five-point scale.
4. Some authors applied a **questionnaire**. However, it was not specified what the questionnaire looked like.
5. The size of programs (**LOC**) participants had written.
6. A **pretest** was conducted to assess the participants' programming experience. However, it was not specified in the papers what the pretest looked like.
7. In two papers, in which professional programmers were recruited as participants, the supervising **manager** estimated the experience of participants (Arisholm et al. 2007; Hannay et al. 2010).
8. the authors state that they measured programming experience, but **did not specify how**.
9. programming experience was **not mentioned at all**, which may threaten the validity of the corresponding experiments.

What kind of experience?

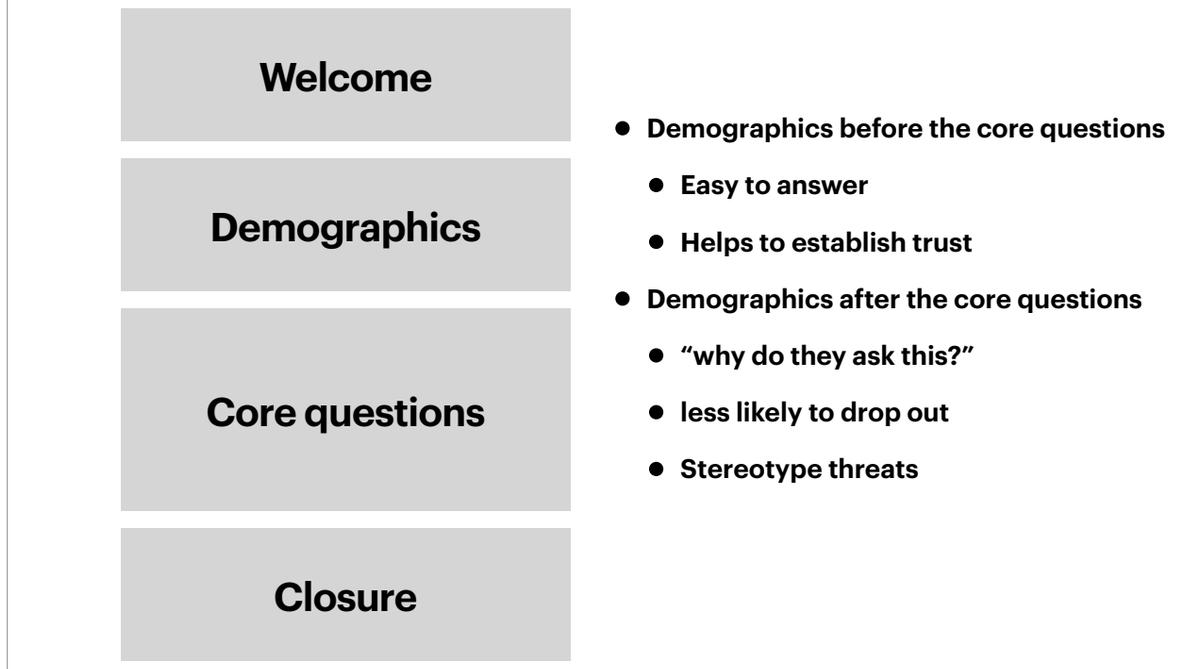
Specific programming language?

Students or practitioners?

Professional experience (large systems, complex problems)?



The next question pertaining to demographics is ***when*** to ask about it: before the core questions or after.
What do you think?



Some of the demographic questions might be experienced as intrusive triggering a negative response from the participants (**why do they ask this**). It is much easier to stop answering the questionnaire if you have not answered that many questions (demographics before) as opposed to if you have answered many questions (demographics after).

Stereotype threats

https://storage.needpix.com/rsynced_images/clock-1010200_1280.jpg

In general, a survey should be short, 5, 10, 15 minutes top! The longer it will take people to fill in the survey, the less responses you will get.

IS THE INTERVIEW GUIDE/SURVEY GOOD ENOUGH?

https://sprudge.com/wp-content/uploads/2018/05/Screen-Shot-2018-05-07-at-3.00.49-PM-1170x777.png

Clear means that the respondents need to understand the question. This is what we have tried to achieve by phrasing the questions differently. Even questions that might appear as basic questions such as “Are you male/female/something else” have been shown to not be clear to participants (in this case, since it was not clear whether the question was referring to gender identity or to sex assigned at birth). In case of interviews, unclear questions can be clarified on the go, but this is not ideal and might be distractive. To check this we need to review the questions, and most important: pilot them, i.e., ask someone from the target group to answer the survey/ hold a mock interview and comment on clarity of the questions. Then we can revise the phrasing if needed.

https://sprudge.com/wp-content/uploads/2018/05/Screen-Shot-2018-05-07-at-3.00.49-PM-1170x777.png

Comprehensive means that the questions capture information relevant for the research and the answer options are as complete as possible. This is where we need to think about existing lists (e.g., professions, company sizes, counties) and instruments. Also here piloting might help but this is not very likely if some answers are expected to be given much more frequently than others.

https://sprudge.com/wp-content/uploads/2018/05/Screen-Shot-2018-05-07-at-3.00.49-PM-1170x777.png

If the survey is not **acceptable**, prospective respondents will not answer the questions. Typical problems related to acceptability are excessive length of the survey, intrusive questions or questions perceived as being irrelevant. In case of interviews, unacceptable questions might lead to participants interrupting interviews (in extreme cases) or giving shallow answers.

https://sprudge.com/wp-content/uploads/2018/05/Screen-Shot-2018-05-07-at-3.00.49-PM-1170x777.png

Reliability means that similar results would have been obtained if the same survey/interviews would have been administered to two similar but different groups. For example, if I would like to study perceptions of bachelor students Computer Science in the Netherlands, I would at least expect the results in Eindhoven, Delft and Twente to be similar. Similar but not identical.

https://sprudge.com/wp-content/uploads/2018/05/Screen-Shot-2018-05-07-at-3.00.49-PM-1170x777.png

Reliability comes in three flavours.

- Test-retest, i.e., how likely is that the person responds in the same way if surveyed twice. Ask **the same question twice** during the pilot and check agreement (Chronbach's alpha or alternatives).
- Inter-rater: to what extent different persons give similar answers. Run **two pilots** with different groups of people and compare the results. Very expensive to perform, so rarely done.
- Inter-coder: in case of open questions, to what extent do the results obtained by different people agree

Validity finally means that the survey instrument measures what it is supposed to measure. For example, asking “are you a female or a male” is not a valid way to ask about gender. Provide a solid argument why what you measure agrees with what you want to measure. This is related to the notion of “construct validity” we will discuss later on.

Trustworthiness: Trustworthiness assesses the validity of our qualitative analyses and conclusions. In our interviews, we used robust labeling with multiple authors and clear guidelines to reduce bias of a single labeler. We reported earlier on rater
reliability for our coding of the survey responses. Our codebook and the anonymized transcripts and answers are available for reliability of our findings

QUESTION

To assess reliability of the survey the researchers conduct **test-retest**, i.e., ask the same question twice during the pilot and check agreement between the responses. After the **test** one of the participants said “Until today, I have never thought much about it, but it’s a very interesting subject”. At this point...

(A) RETESTING IS A GOOD IDEA

(B) RETESTING IS A BAD IDEA

Bad idea. In general, test-retest is not valid

- if variables naturally change over time (between test and retest)
- if answering the questionnaire may change the respondents’ attitudes and hence their answers.
- if respondents remember what they said previously, so they answer the same way in an effort to be consistent (even if new information in the intervening time makes a second, different answer more correct).

We typically conduct surveys and interviews to either explore a new domain or to explain a phenomenon we have observed.
Whether we want to conduct an interview or a survey we need to keep in mind that this study will involve human beings, and this is why our primary concern should be to not harm them. Interview/survey questions should be aligned with the research questions. Finally, we stated the requirements for interviews and surveys.