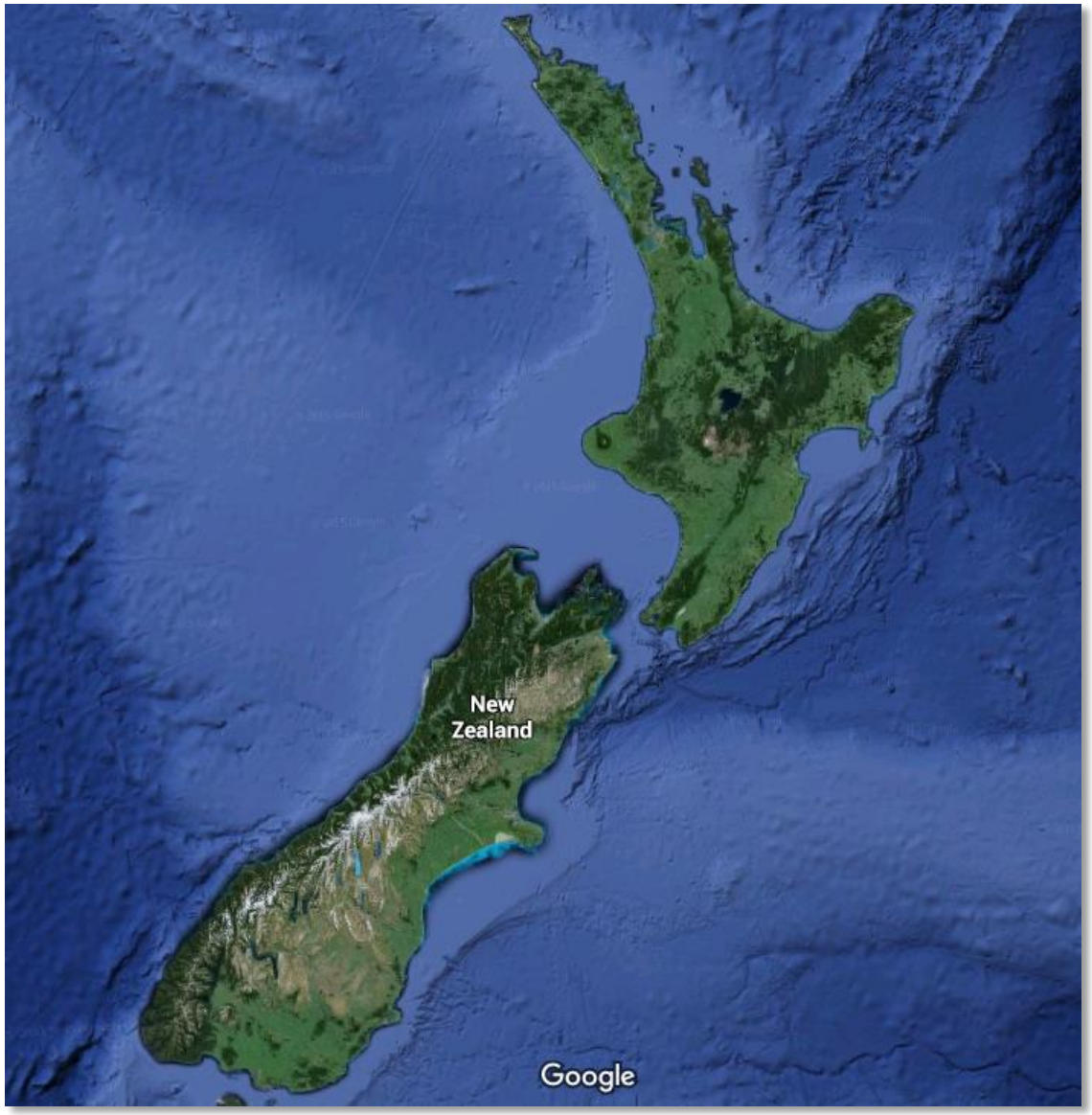
# Reflecting on Empirical Software Architecture Research

Matthias Galster, University of Canterbury, New Zealand

(with Danny Weyns, KU Leuven, Belgium)

# Where I am from




- Why should I submit/attend?
  - A-ranked (CoRE)
  - Full and short papers, impact-to-industry papers
  - Double-blind review
  - Proceedings published by ACM
  - Emerging researchers' forum
  - Keynotes (Claes Wohlin, Blekinge; Gareth Cronin, Xero)

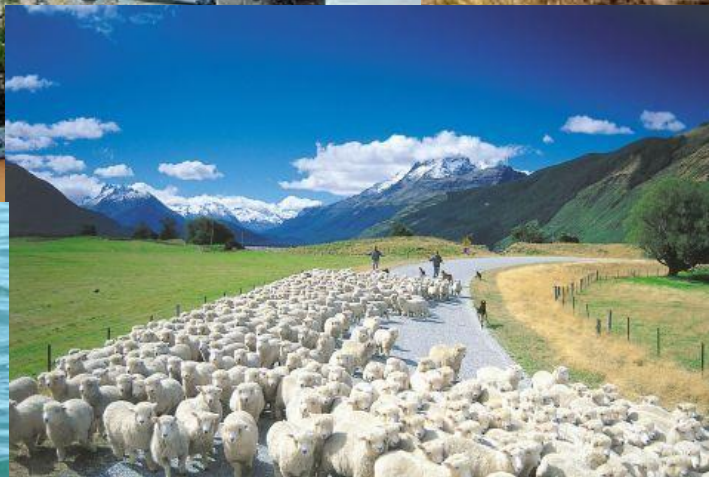




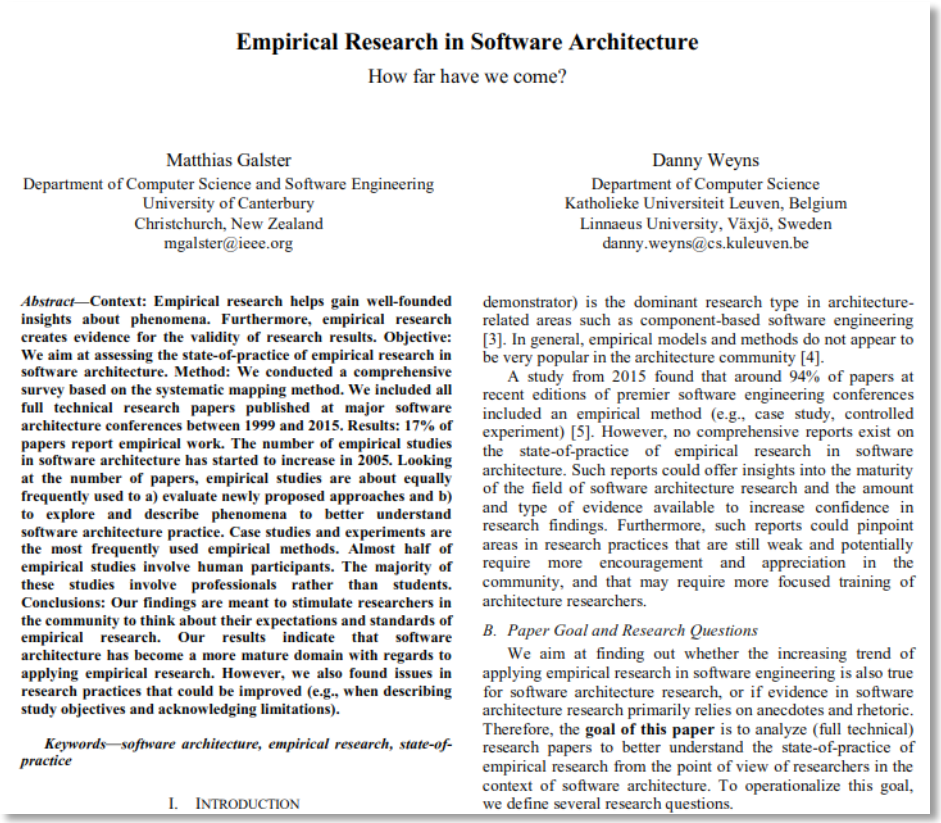








# Reflections



**Empirical Research in Software Architecture**

How far have we come?

Matthias Galster
Department of Computer Science and Software Engineering
University of Canterbury
Christchurch, New Zealand
mgalster@ieee.org

Danny Weyns
Department of Computer Science
Katholieke Universiteit Leuven, Belgium
Linnaeus University, Växjö, Sweden
danny.weyns@cs.kuleuven.be

***Abstract*—Context: Empirical research helps gain well-founded insights about phenomena. Furthermore, empirical research creates evidence for the validity of research results. Objective: We aim at assessing the state-of-practice of empirical research in software architecture. Method: We conducted a comprehensive survey based on the systematic mapping method. We included all full technical research papers published at major software architecture conferences between 1999 and 2015. Results: 17% of papers report empirical work. The number of empirical studies in software architecture has started to increase in 2005. Looking at the number of papers, empirical studies are about equally frequently used to a) evaluate newly proposed approaches and b) to explore and describe phenomena to better understand software architecture practice. Case studies and experiments are the most frequently used empirical methods. Almost half of empirical studies involve human participants. The majority of these studies involve professionals rather than students. Conclusions: Our findings are meant to stimulate researchers in the community to think about their expectations and standards of empirical research. Our results indicate that software architecture has become a more mature domain with regards to applying empirical research. However, we also found issues in research practices that could be improved (e.g., when describing study objectives and acknowledging limitations).**

***Keywords*—software architecture, empirical research, state-of-practice**

I. INTRODUCTION

demonstrator) is the dominant research type in architecture-related areas such as component-based software engineering [3]. In general, empirical models and methods do not appear to be very popular in the architecture community [4].

A study from 2015 found that around 94% of papers at recent editions of premier software engineering conferences included an empirical method (e.g., case study, controlled experiment) [5]. However, no comprehensive reports exist on the state-of-practice of empirical research in software architecture. Such reports could offer insights into the maturity of the field of software architecture research and the amount and type of evidence available to increase confidence in research findings. Furthermore, such reports could pinpoint areas in research practices that are still weak and potentially require more encouragement and appreciation in the community, and that may require more focused training of architecture researchers.

*B. Paper Goal and Research Questions*

We aim at finding out whether the increasing trend of applying empirical research in software engineering is also true for software architecture research, or if evidence in software architecture research primarily relies on anecdotes and rhetoric. Therefore, the **goal of this paper** is to analyze (full technical) research papers to better understand the state-of-practice of empirical research from the point of view of researchers in the context of software architecture. To operationalize this goal, we define several research questions.

What happens at CBSE, ECSA, QoSA, WICSA[1]

All PCs of CBSE, ECSA, QoSA, WICSA[2]

[1]M. Galster and Danny Weyns, "Empirical Research in Software Architecture – How far have we come?" WICSA, 2016
[2]In progress

# Questions

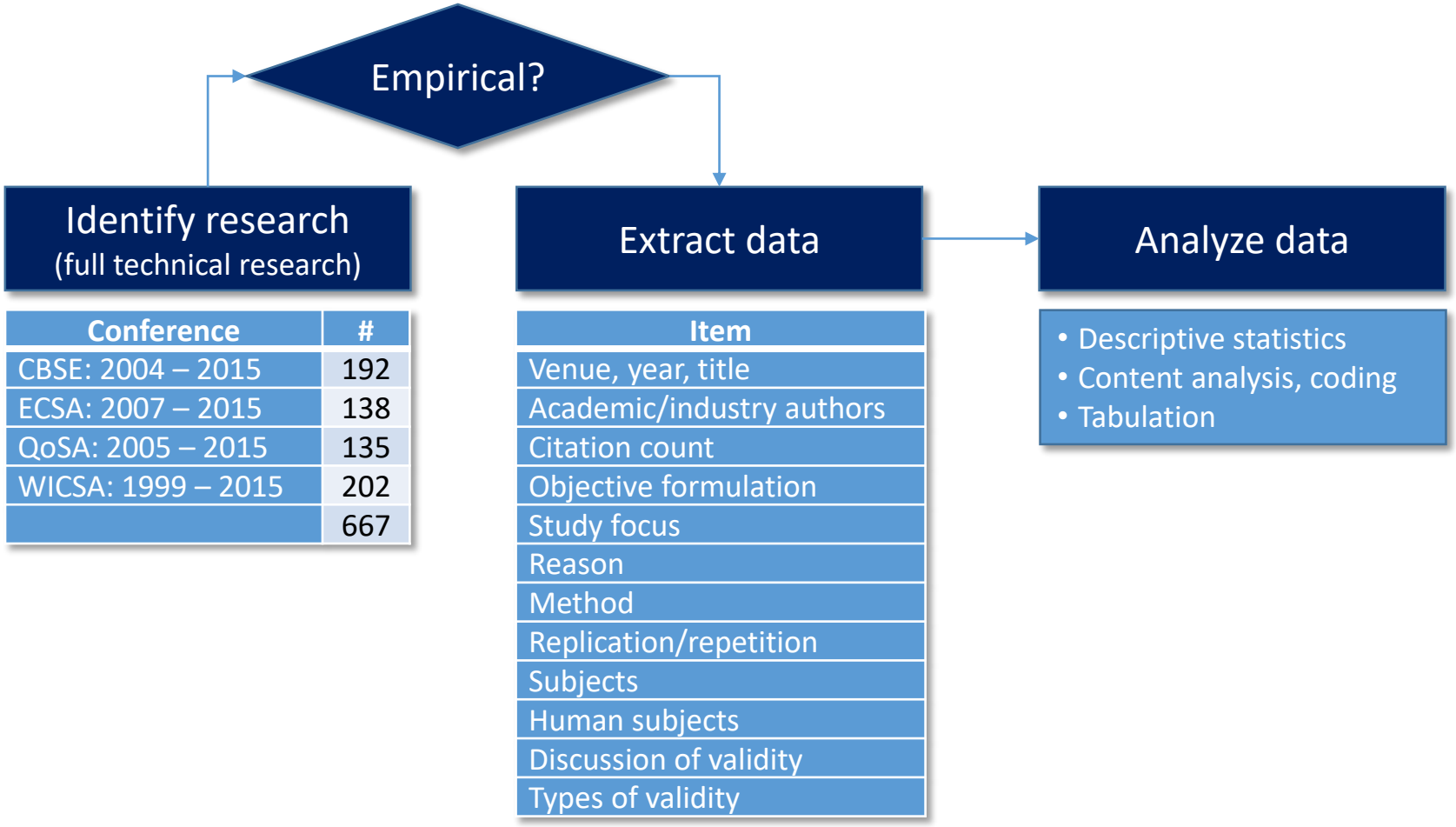
How is empirical research applied in software architecture?

What is the role of human participants in empirical software architecture research?

To what degree does empirical software architecture research acknowledge validity threats?

What are the perceptions of those who conduct/review empirical work?

# Method (part 1)




K. Petersen et al., "Systematic Mapping Studies in Software Engineering," EASE, 2008
K. Petersen et al., "Guidelines for Conducting Systematic Mapping Studies in Software Engineering: An Update," IST, 2015

# Empirical or not

"… seeks to explore, describe, predict, and explain natural, social, or cognitive phenomena by using evidence based on observation or experience. It involves obtaining and interpreting evidence, by, e.g., experimentation, systematic observation, interviews or surveys, or by the careful examination of documents or artifacts."



D. Sjoberg, T. Dyba, and M. Jorgensen, "The Future of Empirical Methods in Software Engineering Research," FOSE, 2007

# Method (part 2)

- Questionnaire-based survey

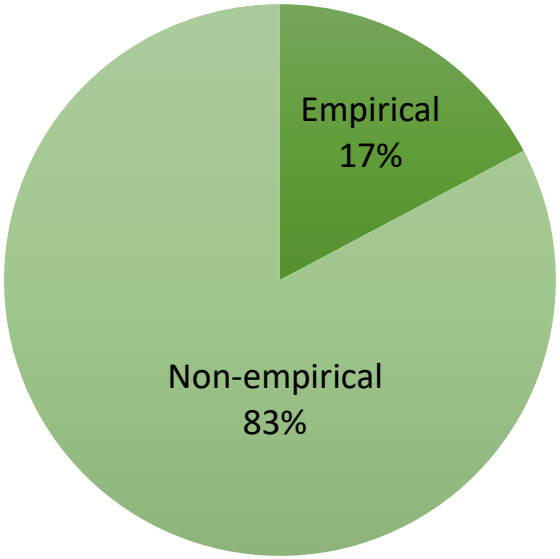
- All members of all PCs of architecture key venues
  - Considered “key players” in the field
  - CBSE, ECSA, QoSA, WICSA
  - Sample size: 455

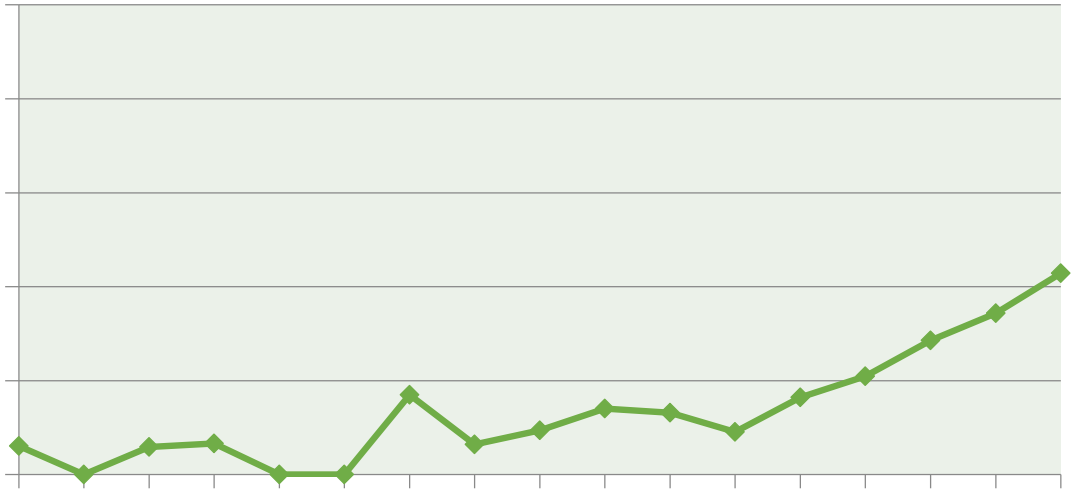
- Responses
  - *N* = 105 (response rate: ~23%)
  - 12% answered as “practitioners”
  - All respondents also review for other venues


OVERVIEW

# What is published?

- Surveyed venues



- Surveyed PCs
  - 31%: never published an “empirical” study
  - [12%: never reviewed an “empirical” study]

# Some quotes

- *"You have to become knowledgeable on systematic studies to appreciate them."*
- *"I learnt more empirical methods, i.e., I also changed my reviews over time. I often see reviews of others which obviously have not learnt empirical methods, that is a pity."*
- *"I am now more biased against such work, because it is usually so contrived. A nice experiment, but a "who cares?" result."*
- "*People who do this empirical research should pair up with a person doing *real* work so that they can understand when they are writing useless drivel*"

# Who writes these papers?

- Ø number of authors
  - Empirical: 3.15 (min: 1, max: 7)
  - Non-empirical: 3.04 (min: 1, max: 12)
- Ø number of authors from academia
  - Empirical: 2.7 (min: 0, max: 7)
  - Non-empirical: 2.5 (min: 0, max: 11)
- Ø number of authors from industry
  - Empirical: 0.5 (min: 0, max: 6)
  - Non-empirical: 0.5 (min: 0; max: 10)

# How “popular” are papers?

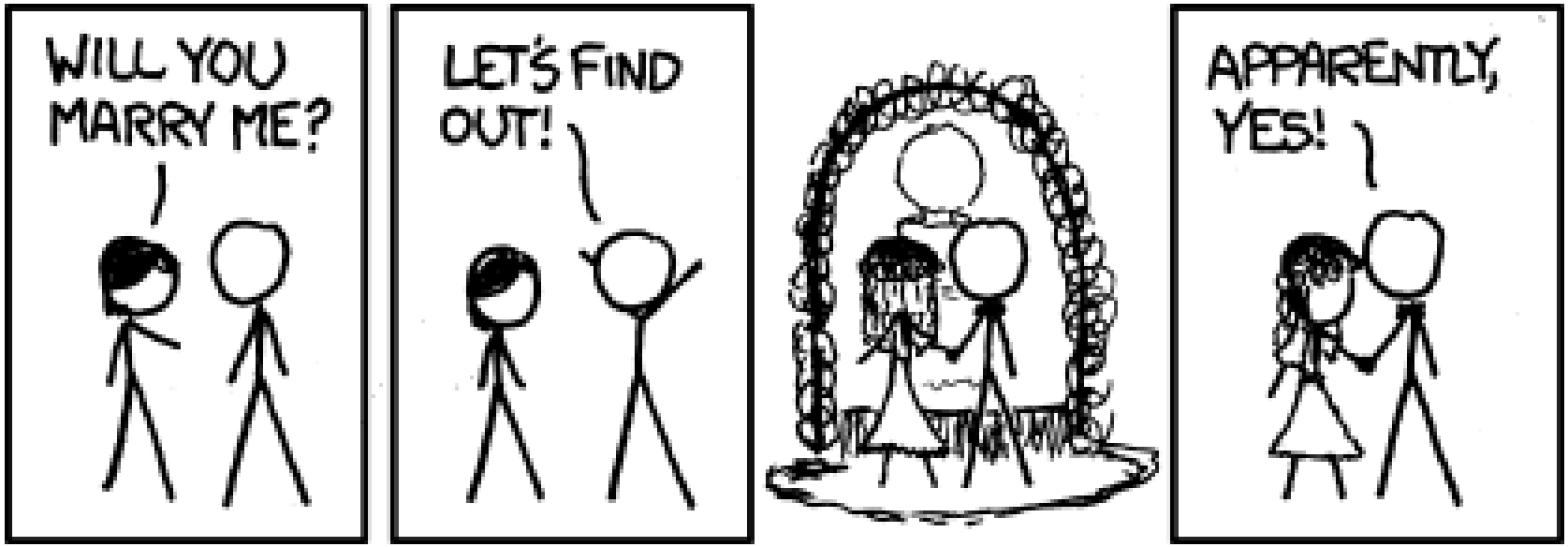
- Ø annual citation count
  - Empirical: 2.8 (min: 0, max: 10.75)
  - Non-empirical: 3.09 (min: 0, max: 50.3)

- Top-100 research papers in software engineering*
  - Min: 21.8
  - Max: 154.2



*V. Garousi and J. Fernandes, "Highly-cited Papers in Software Engineering: The top-100," IST, 2016

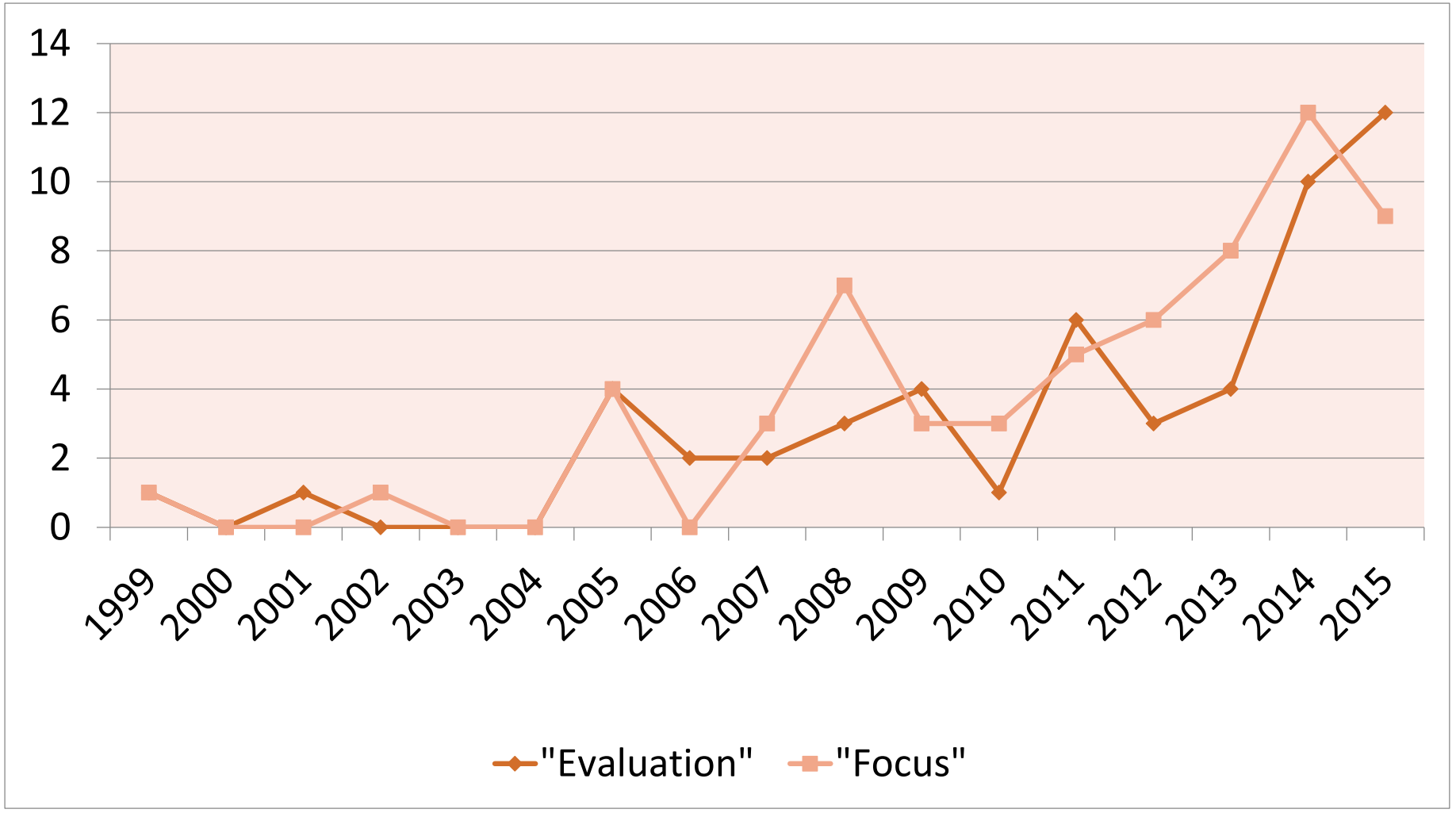
# Q1: how is empirical research applied?



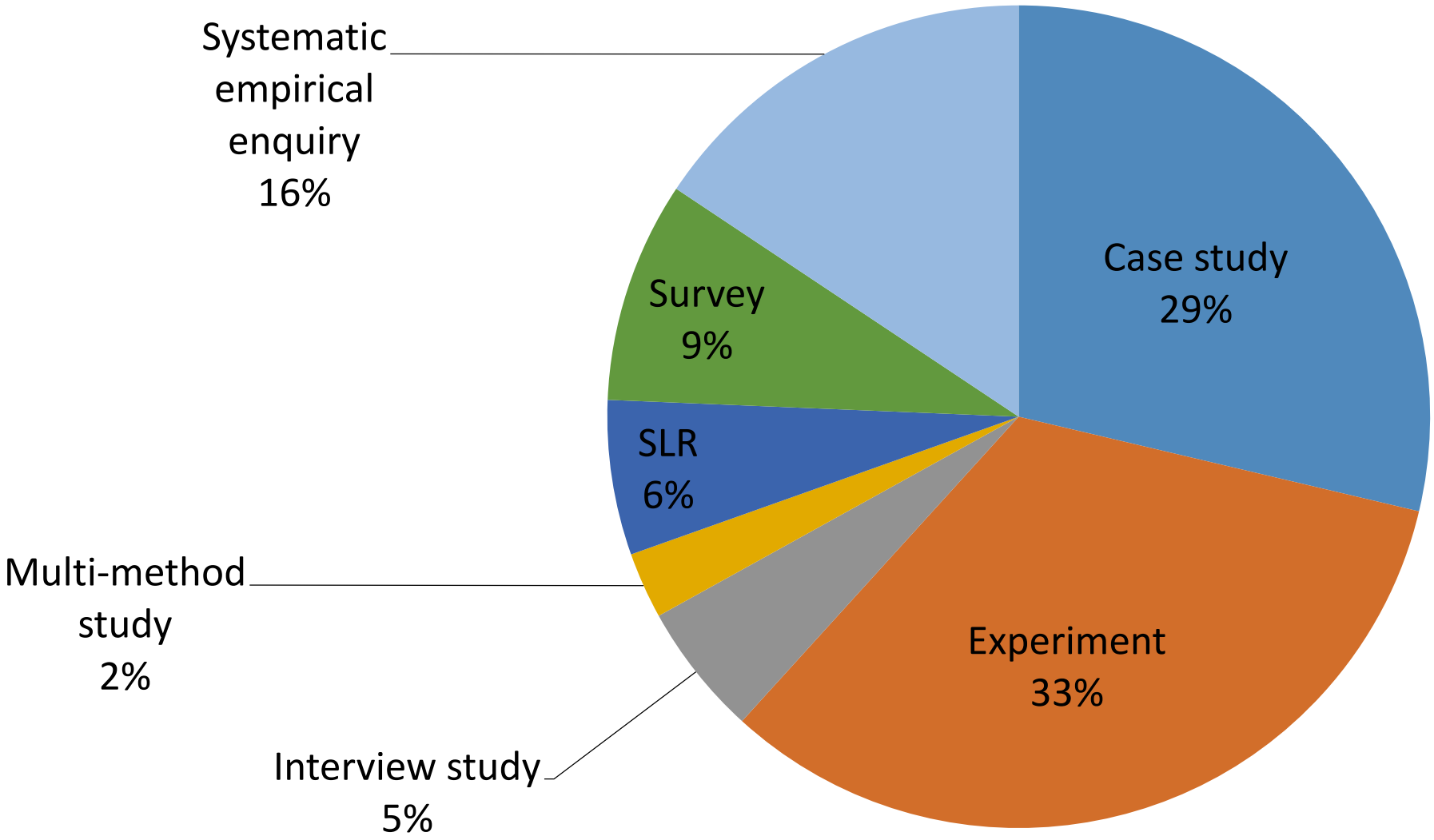



https://xkcd.com/943/

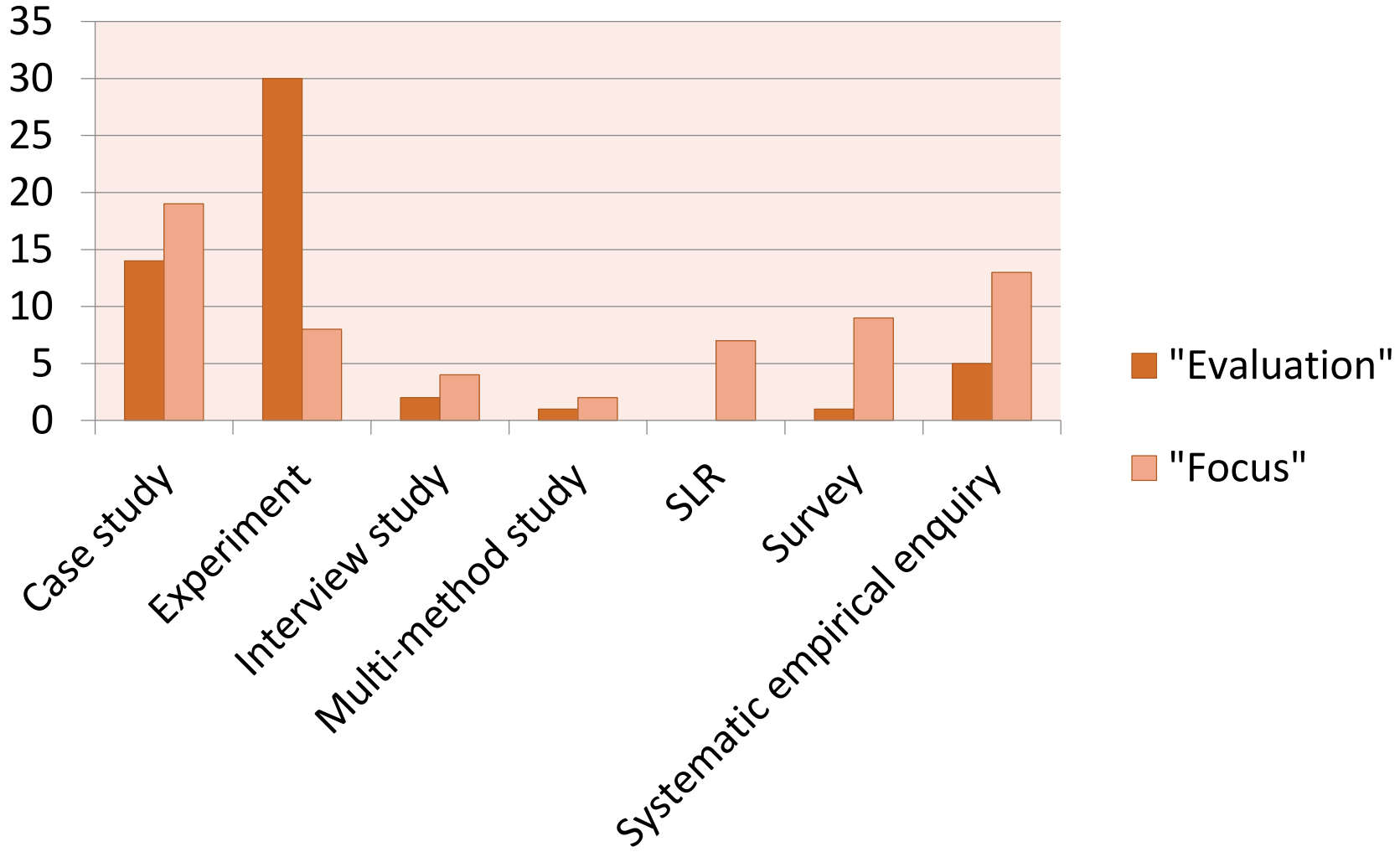
# “Evaluation” or “focus”



# Research methods



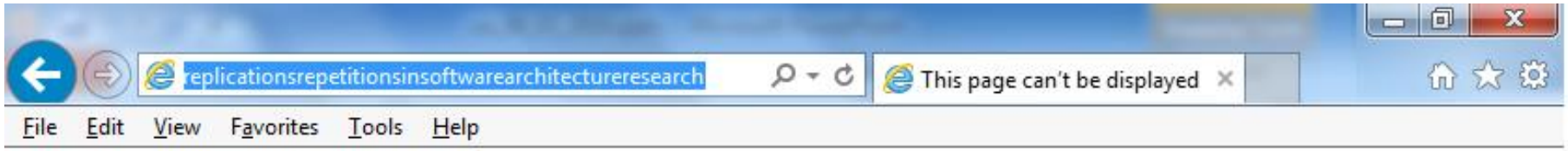
# Reason vs method



# What does the community think?

- 60% indicated no preference of any type of method
- 23%: quantitative research is easier to get accepted
- 21%: quantitative studies are easier to review
- Some quotes
  - "*The only thing that always gets me furious are researchers that 'abuse' one of the methods to claim they showed* [something] *that is out of reach for the specific method*."
  - "*Quantitative stuff is usually bogus. But a lot of it gets published*!"

# Replications and repetitions




## Replications / repetitions can't be displayed

- Make sure replicationsrepetitionsinsoftwarearchitectureresearch is correct.
- Look for the page with your search engine.
- Refresh the page in a few minutes.

Fix connection problems  Fix research problems

# What does the community think?

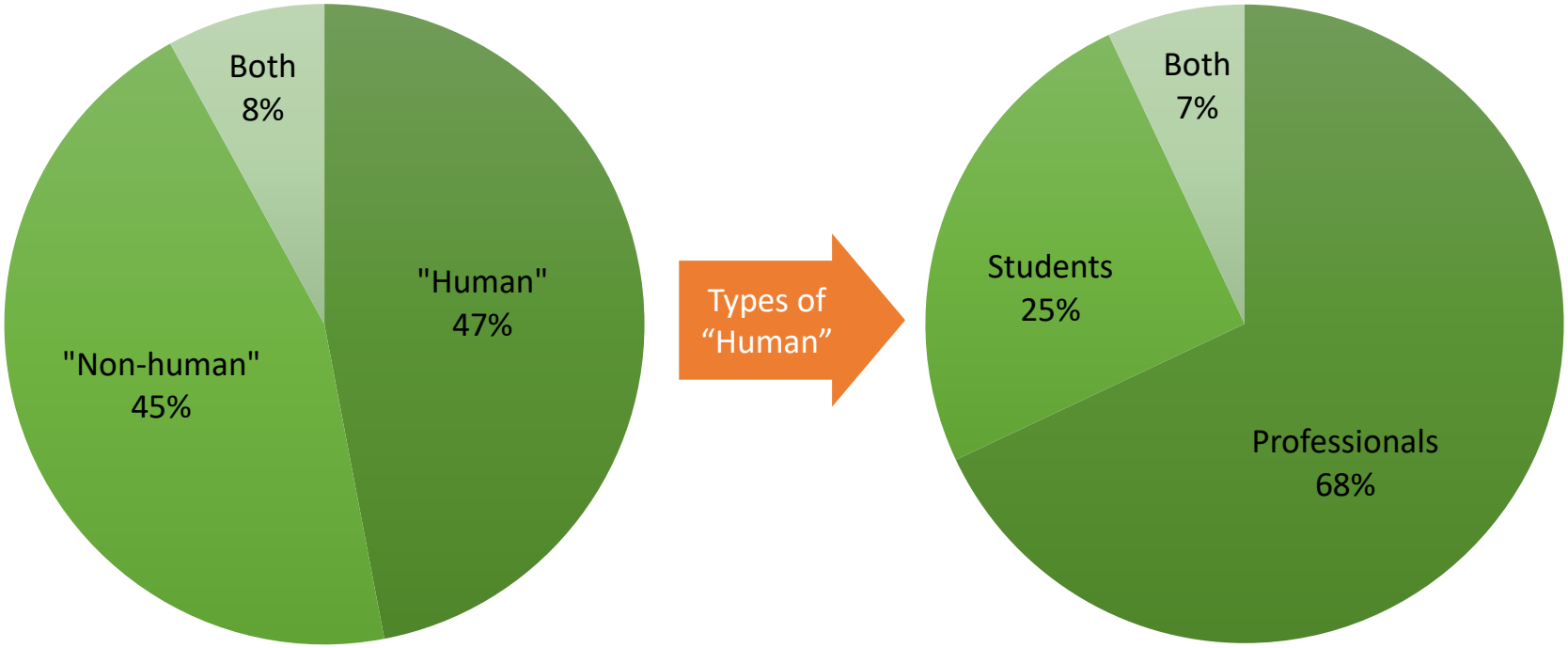
- 61% (strongly) agree that replications advance field
  - But: 76% had never reviewed a replication

- Some quotes
  - *"I think that replication is a fundamental element of empirical studies."*
  - *"Highly doubtful that any useful question in* [software architecture] *can be addressed in a replicated experiment."*
  - *"Well, they* [replications] *have a place....they are part of the fabric serving as the basis of free travel, free food, and all of the other things at a software architecture venue."*

- Paradox?

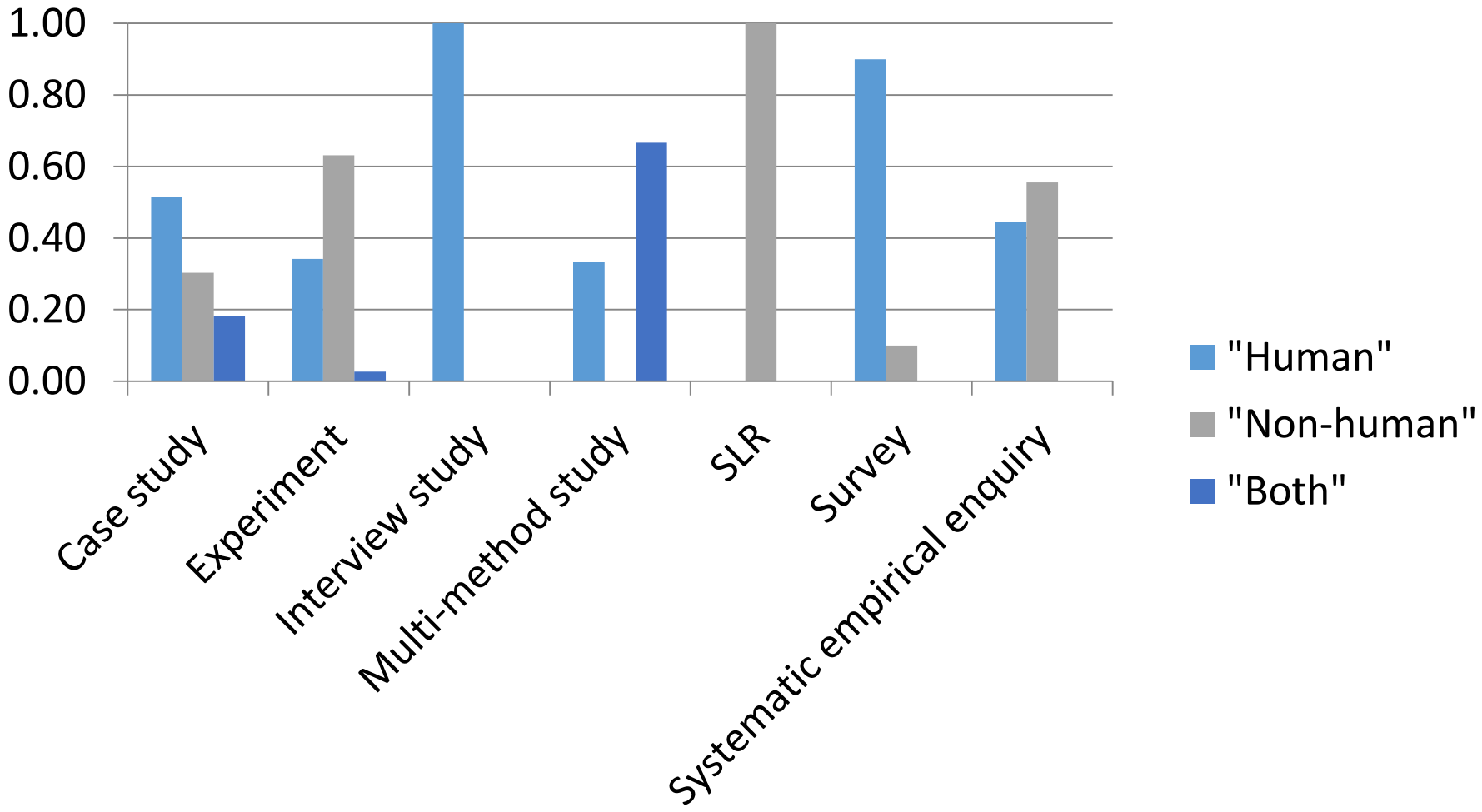
# Q2: role of humans

# “Subjects” in published works

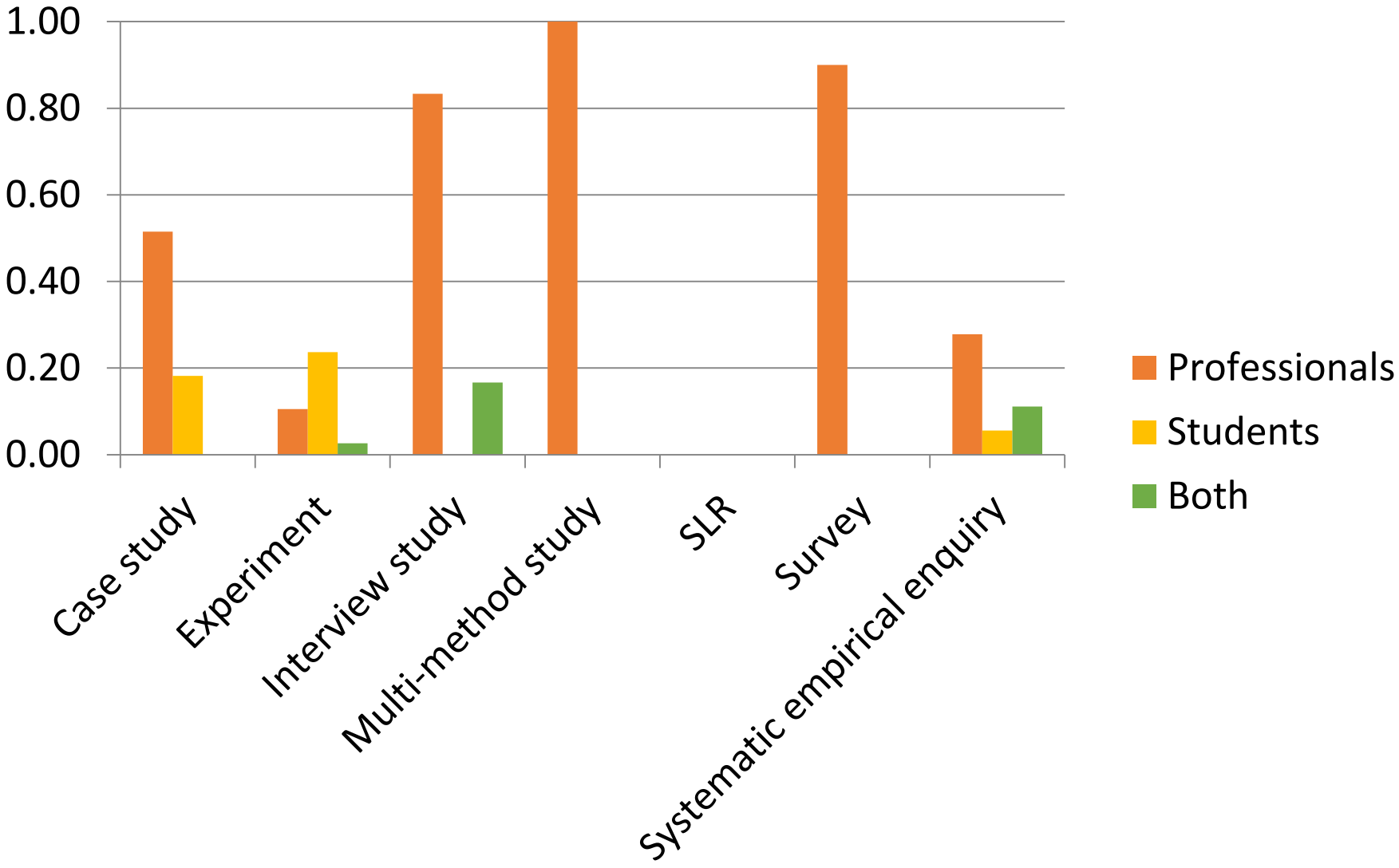



*In comparison – ESEC/FSE, ICSE, ESEM papers: 77% “non-human”, 11% professionals and students, 26% professionals, 44% students
(J. Siegmund et al., "Views on Internal and External Validity in Empirical Software Engineering," ICSE, 2015)

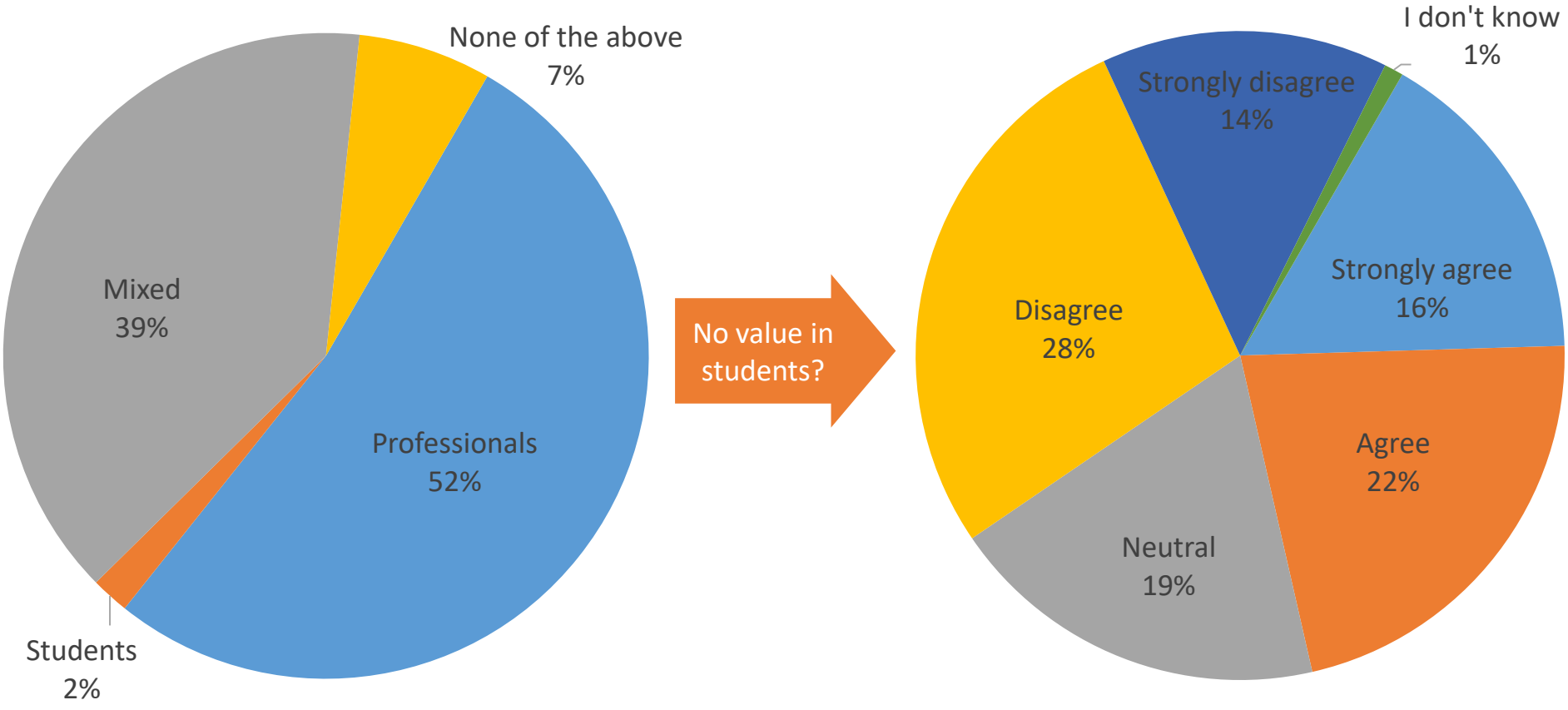
# Participants vs method



# Human participants vs method

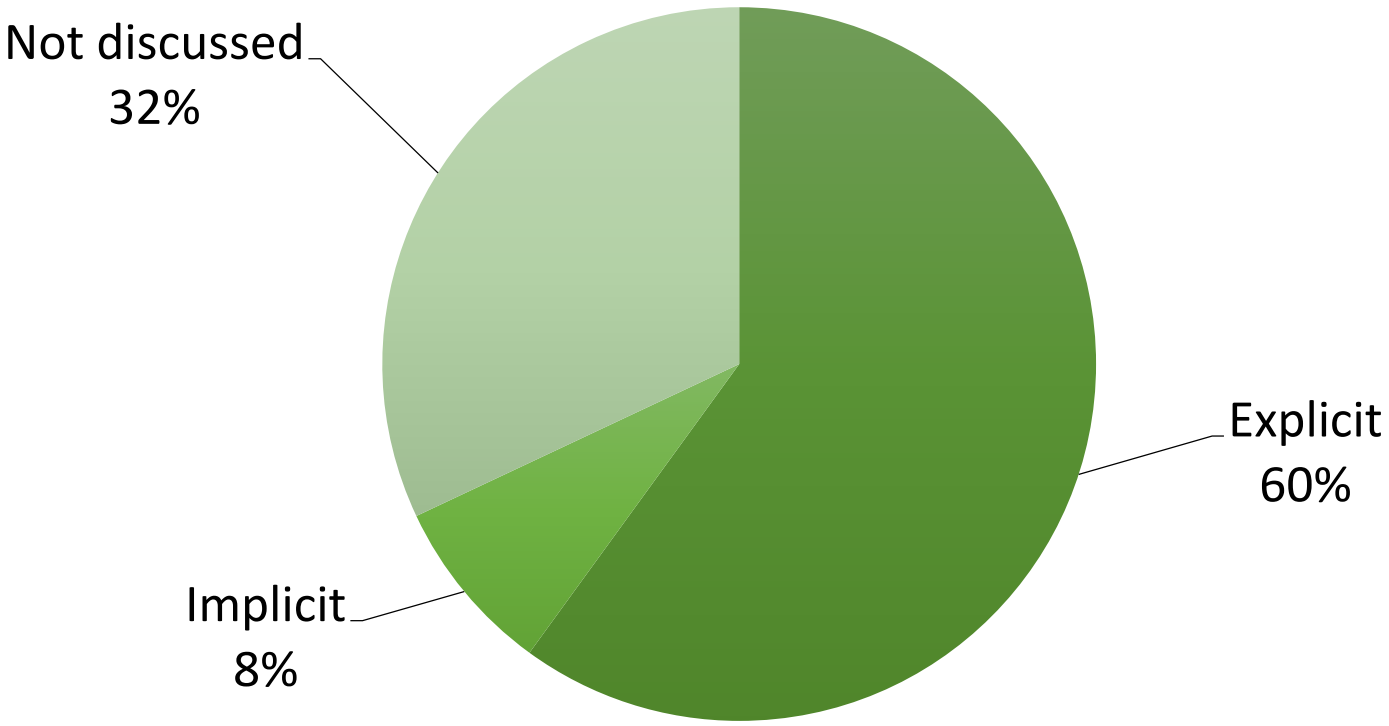


# Preferred types of “human subjects”



# Q3: validity threats

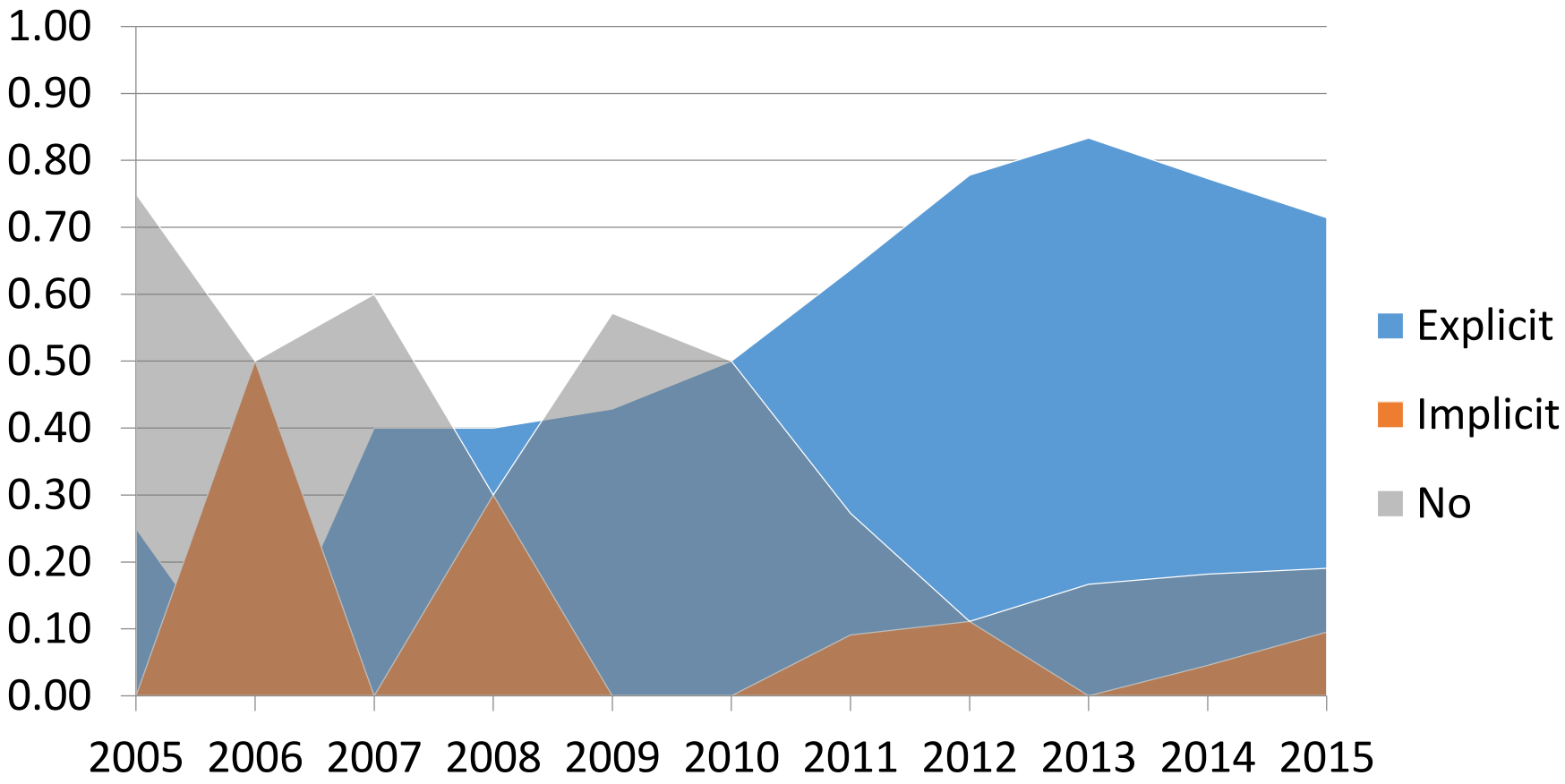
# Threats discussed




- No validity threats in 20.9% of ESEM 2009 papers*
- No validity threats in 46% of ESEC/FSE, ICSE, ESEM papers**

*R. Feldt and A. Magazinius, "Validity Threats in Empirical Software Engineering Research - An Initial Survey," SEKE, 2010
**J. Siegmund et al., "Views on Internal and External Validity in Empirical Software Engineering," ICSE, 2015

# Development over time



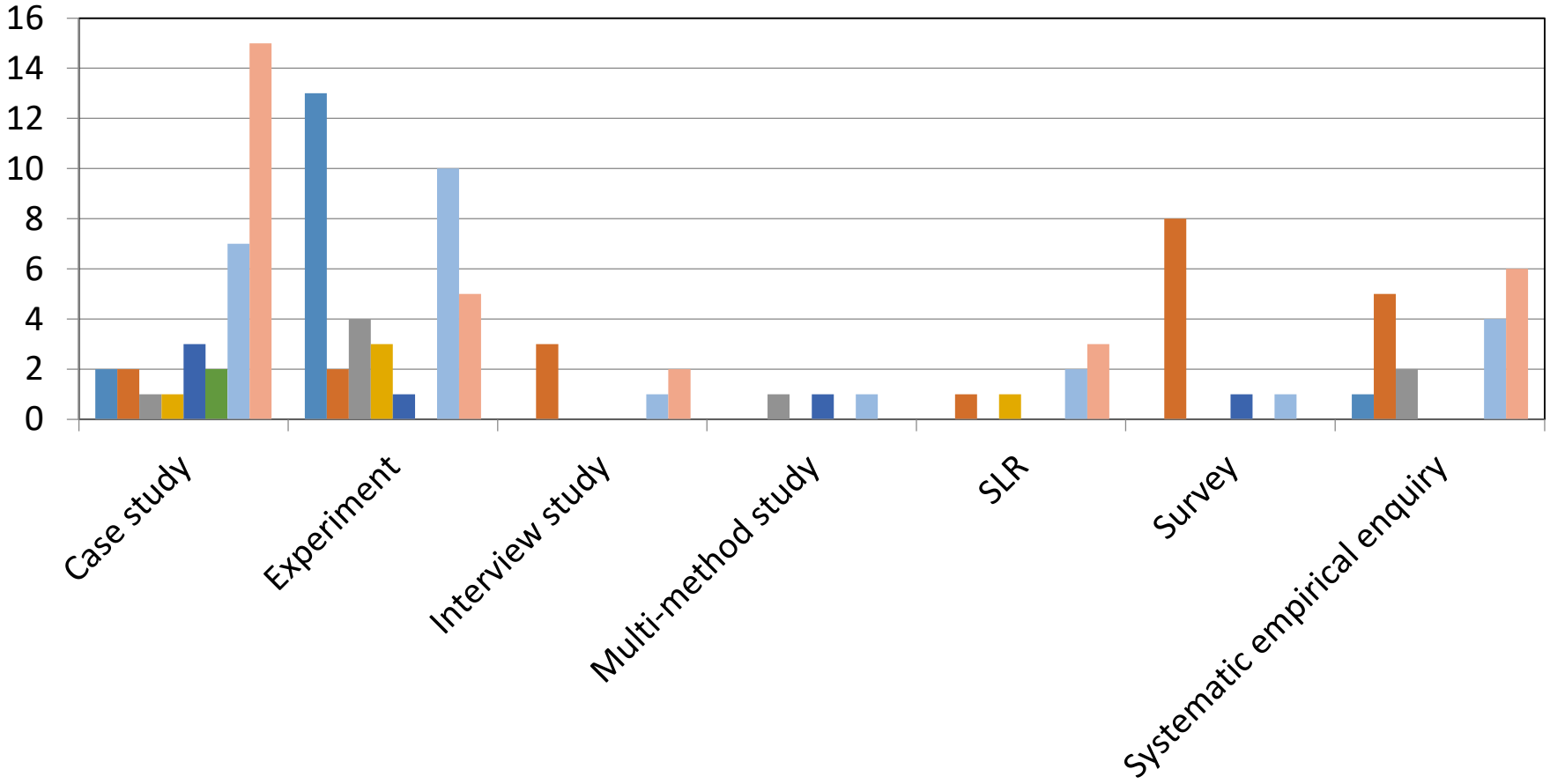
# What does the community think?

- Internal vs external validity
  - 32%: Maximize internal validity
  - 35%: Maximize external validity

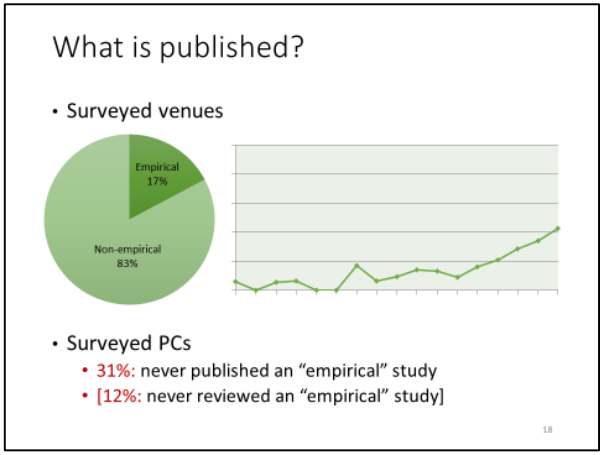
- But overall
  - “It depends…” (obtained from textual comments)

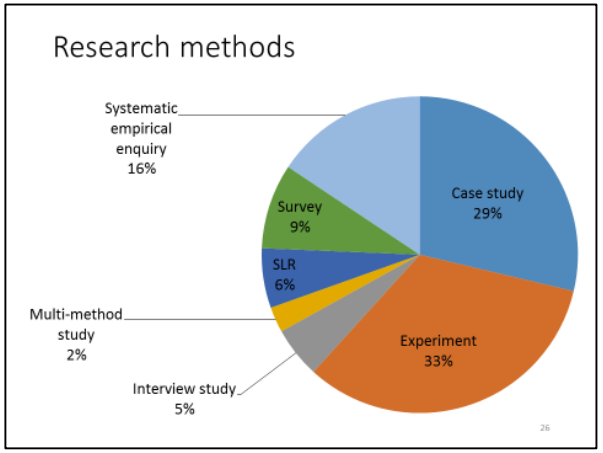
# Further insights and discussions

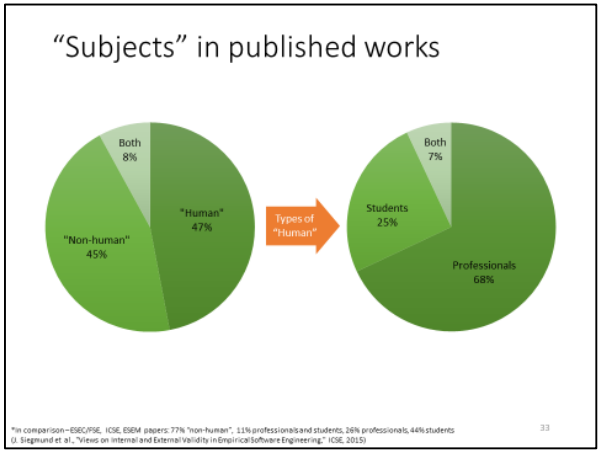
# Research themes

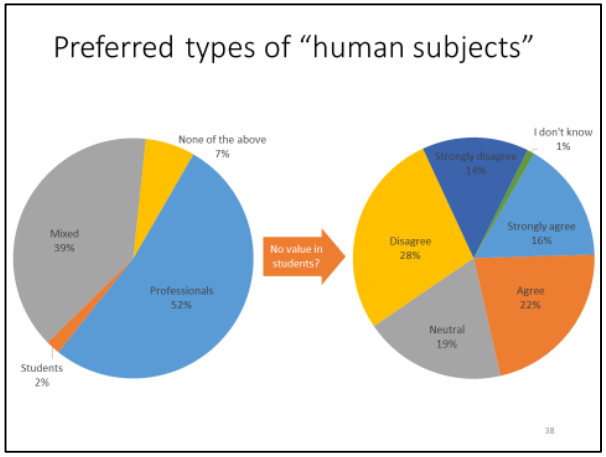


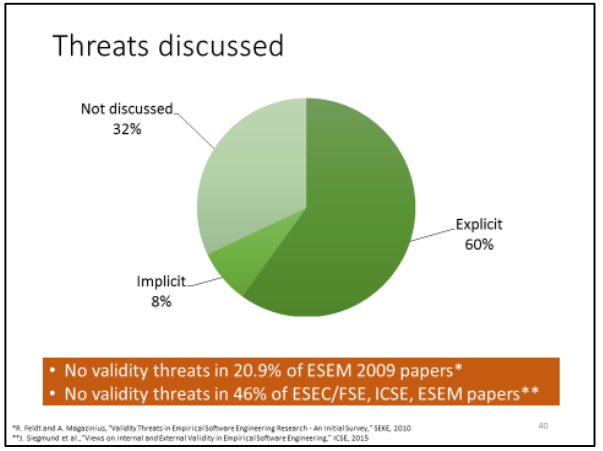
# Summary and conclusions




- Replication paradox?
- Where to next?
  - Examples?
  - Guidelines?
  - Training?

**Contact:** Matthias Galster, University of Canterbury, Christchurch, New Zealand, mgalster@ieee.org