Empirical Software Engineering

ECE1724H S2 (Winter2021)

Shurui Zhou

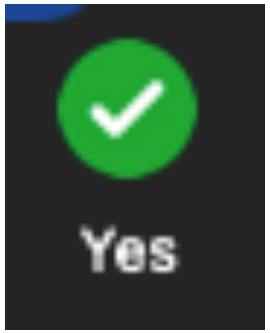
Assistant Professor

shuruiz@ece.utoronto.ca

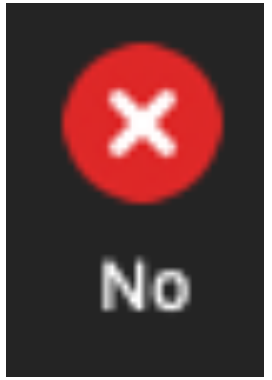
The Edward S. Rogers Sr. Department
of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

Lecture Logistics during a Pandemic

- If you can hear me in zoom, please click



otherwise, click



First of all:

You are not alone!
We are undertaking this new experience together.

This is not normal. We understand.

- Expect:
  - Internet and bandwidth issues
  - Timezone issues
  - Distractions -- parents, siblings, pets
  - Feeling isolated, feeling overwhelmed
  - Many additional sources of stress
  - Hard time dealing with *everything*...

Talk to us about accommodations of any kind!

Simulating in-class Experience

- Discussions and interactions are important. We'll have regular in-class discussions and exercises
- Use chat or "raise hand" feature
- Muted by default, keep camera on if possible
- Attend lecture live, recordings only as backup
- I may call on you
- Contact me for accommodations!

Shurui Zhou

https://www.eecg.utoronto.ca/~shuruiz/
shuruiz@ece.utoronto.ca

Carnegie Mellon University

2014 - 2020 Ph.D.
School of Computer Science
Institute for Software Research

UNIVERSITY OF TORONTO

2020 Fall – Assistant Professor

Research Interests

- Software Engineering (SE)
- SE for AI
- AI for SE
- Collaborative Software Development
- Open Source

Help software developers
to better collaborate

+ Advances in tooling & SE principles
+ Insights from other disciplines
+ Mix a wide range of research methods

Problem → Intervention → Evaluation

Help software developers
to better collaborate

+ Advances in tooling & SE principles
+ Insights from other disciplines
+ Mix a wide range of research methods

Problem → Intervention → Evaluation

Help software developers
to better collaborate

+ Advances in tooling & SE principles
+ Insights from other disciplines
+ Mix a wide range of research methods

Problem → Intervention → Evaluation

Help software developers
to better collaborate

+ Advances in tooling & SE principles
+ Insights from other disciplines
+ Mix a wide range of research methods

Problem → Intervention → Evaluation

Course Goal

- Motivate the need for an empirical basis for research claims
- Better consumers of empirical research results
- See the variety of kinds of methods available, clear idea about when to use them
- Cover the principal empirical methods applicable to human subjects studies
- Prepare students for advanced research:
  - **Learn how to plan, conduct and report on empirical investigations.**
  - **Understand the key steps of a research project:**
    - **formulating research questions,**
    - **theory building,**
    - **data analysis (using both qualitative and quantitative methods),**
    - **building evidence,**
    - **assessing validity,**
    - **publishing.**
- Relate these methods to relevant meta-theories in the philosophy and sociology of science

Learning Goals

- Understand what research designs and research methods are available for empirical research
- Combine research methods in a mixed-methods design
- Collect and analyze qualitative and quantitative data
- Run statistical tests and interpret results
- Build, validate, and interpret regression models
- Draw conclusions from empirical data
- Present results verbally and in writing
- *... and more (see the syllabus doc on Quercus)*

Week	Topic
1	Introduction
2	Literature Review and Theory
3	Interviews
4	Grounded Theory
5	Surveys
6	Introduction to Measurement
7	Experimentation
8	Quasi-experimental Design & Linear Regression
9	Time Series Analysis
10	Mixed-methods
11	Text Mining
12	Social Network Analysis

Topics

- Overarching concerns
  - Epistemic base
  - Human subjects
- Research Designs
  - Case studies
  - Grounded theory
  - Experiments
  - Quasi-experiments
  - Mixed methods
- Data Collection
  - Interviews
  - Surveys
  - Observation/ethnography
  - Archival data
- Data Analysis
  - Coding
  - Network methods
  - Counterfactual causal reasoning

Intended Audience

- This is an advanced course:
  - assumes a strong grasp of the key research questions in your own research area, and that you are already doing independent research
- Focus:
  - How do people use computer technology?
  - How does this technology (re-)shape human activities?
  - How can we apply qualitative and quantitative techniques from the behavioural sciences to help answer these questions?
- The course is aimed at students who:
  - …plan to conduct research that demands some empirical validation
  - …wish to establish an empirical basis for an existing research programe
  - …wish to apply these techniques in related fields (e.g. Cog Sci,)
- Note: we will *not* cover the kinds of experimental techniques used in CS systems areas, nor in medical/biological research
  - Focus is on the relationship between human activity and computer technology

Lectures

Thursday 1:00-4:00 pm EST
(1:10-2pm + 2:10-3pm + 3:10-4pm)

Break: (remind me if I forget!)
Stand/stretch frequently,
Drink water, Try to change location in your room or apartment

- Zoom
- ‘U of T time’ — classes start at 10 minutes past the hour, and wrap at the top of the hour.
- Office hour by appointment

Readings

- Major component is discussion of weekly readings
- Please read the set papers before the seminar
- Students present papers 10-15 min each

Activities

- Read method descriptions / how-tos
- Critique papers that use these methods
- Exercises where you get some experience with some of the methods
- Project on your own research topic
  - Review literature
  - Design and plan a study
  - Write as proposal

Evaluation

- 40% assignments
- 50% research project
  - 10% initial project description (proposal)
  - 2% interim report
  - 8% final presentation
  - 30% final report
- 10% participation and in-class presentations

Class (zoom) etiquette

- You have to be authenticated to enter Zoom
- If you want to ask questions or make comments, use your microphone (or raise your hand)
- Arrive on time
- Add your picture to Zoom (camera use encouraged!)

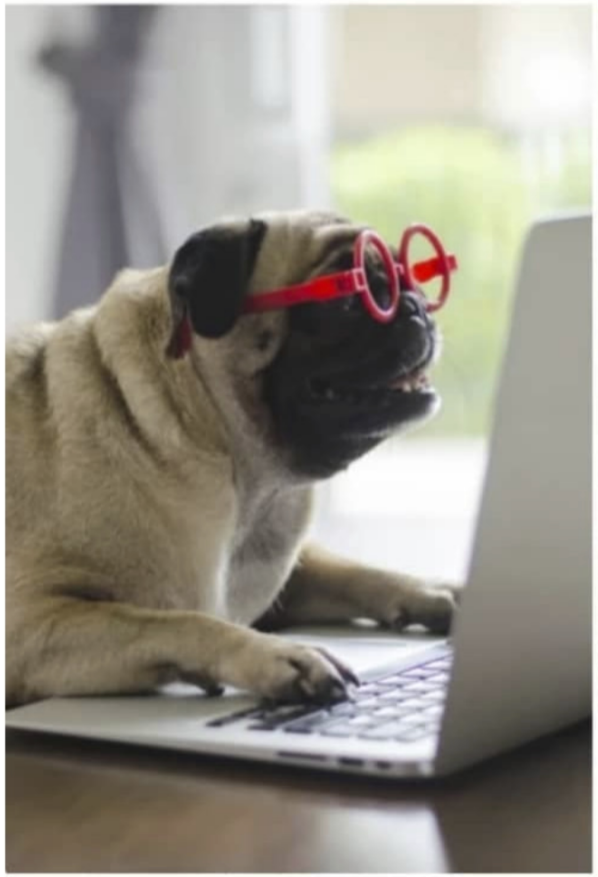
Teachers in normal classes

Stop talking

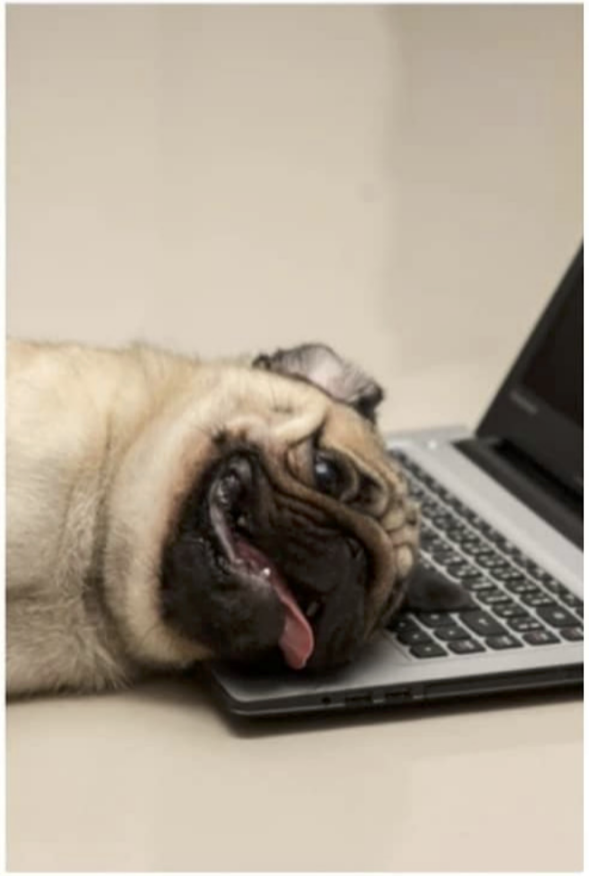
Teachers now

please guys say something

Zoom meeting, with video



Zoom meeting, audio only



https://memes.com/blog/these-hilarious-zoom-memes-are-way-to-real

https://memes.com/blog/these-hilarious-zoom-memes-are-way-to-real

When the teacher calls on you during Zoom and you pretend that the connection is bad

Teaching Assistant

Karthik Mohan

MEng Computer Engineering
karthik.mohan@mail.utoronto.ca
https://karthmnz.github.io/

Disclaimer

- First time teaching + second time online teaching
- 17-803 (2018): Empirical Methods (CMU)

  https://github.com/bvasiles/empirical-methods

  + CSC485 (2020) Empirical SE: Bridging Research and Practice (University of Victoria)

    https://github.com/margaretstorey/EmseUvic2020

  + CSC2130 (2014): Empirical Research Methods for Computer Scientists (UofT)

    http://www.cs.toronto.edu/~sme/CSC2130/index.html

Any Questions?

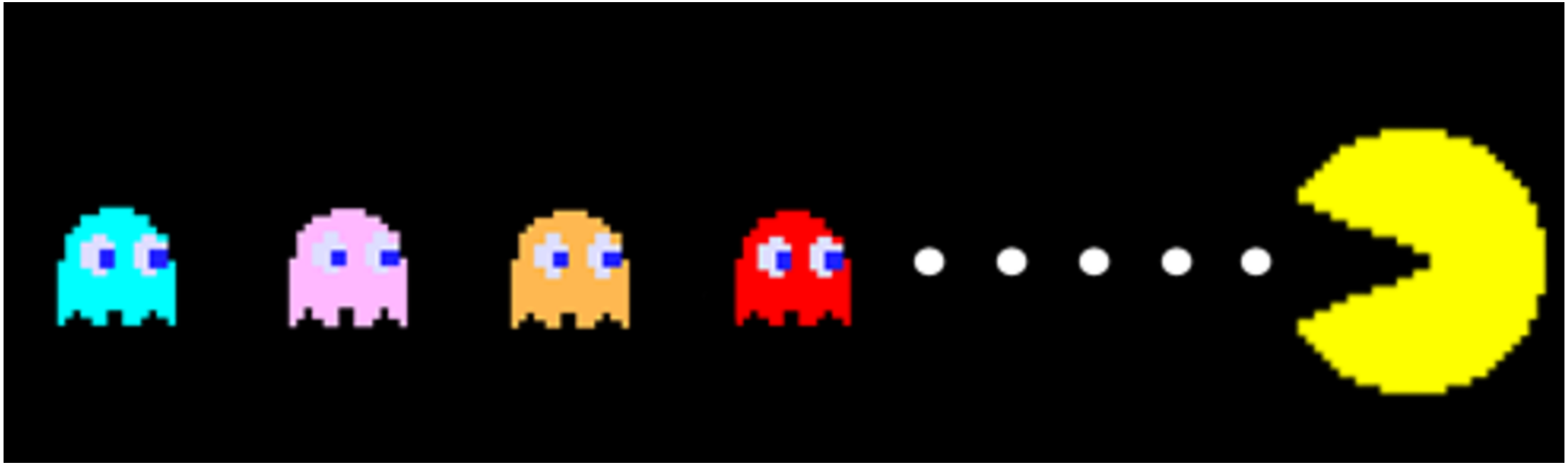
Software Engineering

+

Empirical Methods

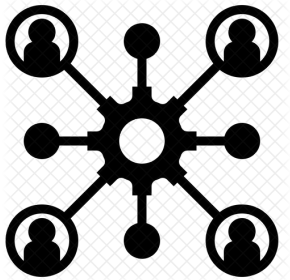
Software is eating the world…. Marc Andreessen

https://a16z.com/2011/08/20/why-software-is-eating-the-world/



(Competing) concerns in SE…

- **Code**: faster, cheaper, more features, more reliable/secure
- **Developers**: more productive, more skilled, happier, better connected
- **Organizations/communities**: attract/retain contributors, encourage a participatory culture, increase value

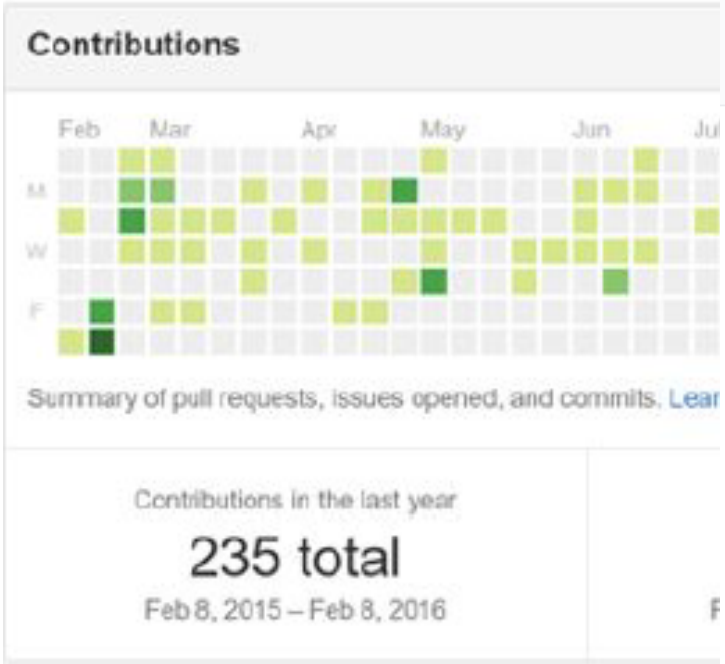


"Measuring programming progress by lines of code is like measuring aircraft building progress by weight."

"In IBM there's a religion in software that says you have to count K-LOCs, ... How big a project is it? ... And IBM wanted to sort of make it the religion about how we got paid. How much money we made off OS 2, how much they did. How many K-LOCs did you do? And we kept trying to convince them - hey, if we have - a developer's got a good idea and he can get something done in 4K-LOCs instead of 20K-LOCs, should we make less money? Because he's made something smaller and faster, less KLOC."

--- Steve Ballmer

https://www.pbs.org/nerds/part2.html



Contributing graphs considered h
https://www.hanselman.com/

isaacs / github

Unwatch

Code | Issues 1.3k | Pull requests 2 | Actions | Projects | Wiki | Security | Insights

Contribution graph can be harmful to contributors #627

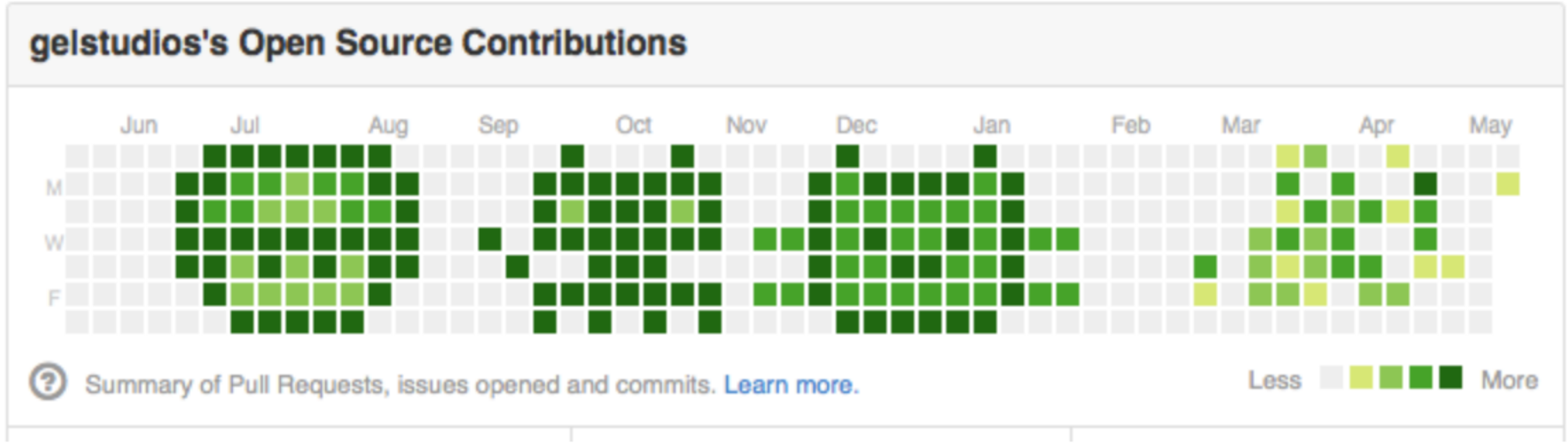
Open **mxsasha** opened this issue on Apr 1, 2016 · 197 comments

mxsasha commented on Apr 1, 2016

A common well-being issue in open-source communities is the tendency of people to over-commit. Many contributors care deeply, at the risk of saying yes too often harming their well-being. Open-source communities are especially at risk, because many contributors work next to a full-time job.

The contribution graph and the statistics on it, prominent on everyone's profile, basically rewards people for doing work on as many different days as possible, generally making more contributions, and making contributions on multiple days in a row without a break.

Stepping away from our work regularly is not only important to uphold high quality work, but also to maintain our well-being. For example, I personally do not generally work in the weekends. That's completely healthy. I take a step back from work and spend time on other things. But in the contribution graph it means I can never make a long streak, even though I do work virtually every day except weekends. So the graph motivates me to work in my weekends as well, and not take breaks. And



Pixel Art

Included "art" from left to right: kitty, oneup, oneup2, hackerschool, octocat, octocat2

https://github.com/gelstudios/gitfiti

Some questions practitioners may care about….

- What is a good architecture to solve problem x? [Devanbu]
- What makes a really awesome programmer? [Software managers]
- How to build a great development team? [Google]
- How is program knowledge distributed? [Naur]
- What is the ideal software engineering process? [Facebook, Microsoft, IBM,…]
- What tools/practices support a participatory development process? [Storey et al.]

Some questions practitioners may care about...

What metrics are the **best predictors of failures**?

If I increase **test coverage**, will that actually increase software quality?

What is the **data quality** level used in empirical studies and how much does it actually matter?

Are there any **metrics that are indicators of failures** in both Open Source and Commercial domains?

I just submitted a **bug report**. Will it be fixed?

Should I be writing **unit tests** in my software project?

How can I tell if a piece of software will have **vulnerabilities**?

Do **cross-cutting concerns** cause defects?

Is strong **code ownership** good or bad for software quality?

Does **Distributed/Global software development** affect quality?

Does **Test Driven Development** (TDD) produce better code in shorter time?

Elon Musk

@elonmusk

Follow

Replying to @timkhiggins

Yes, excessive automation at Tesla was a mistake. To be precise, my mistake. Humans are underrated.

12:54 PM - 13 Apr 2018

7,713 Retweets **40,733** Likes

1.3K 7.7K 41K

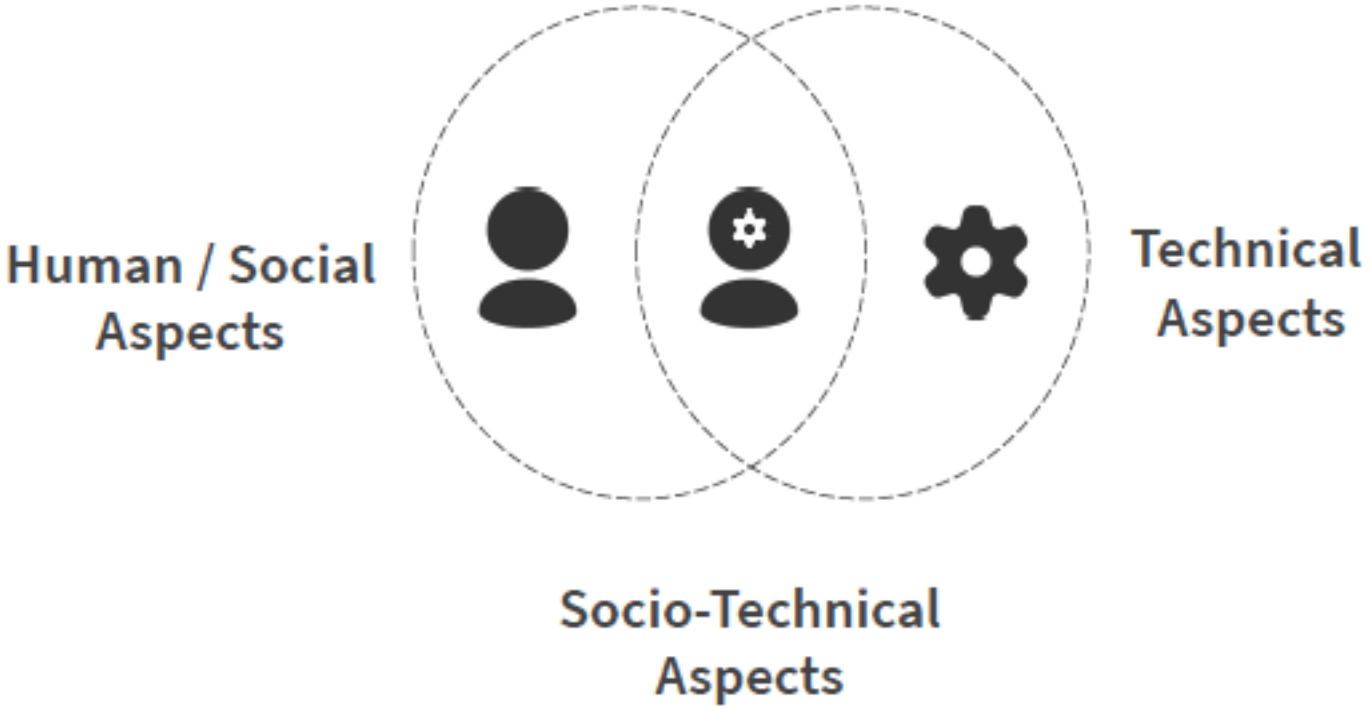
stackoverflow

#1	#2	#3
41.8%	**36.6%**	**36.5%**
Distracting work Environment	Meetings	Non-development work

Developer Study

- https://insights.stackoverflow.com/survey/2019#work-_-greatest-challenges-to-productivity

Software Engineering **Design** Space



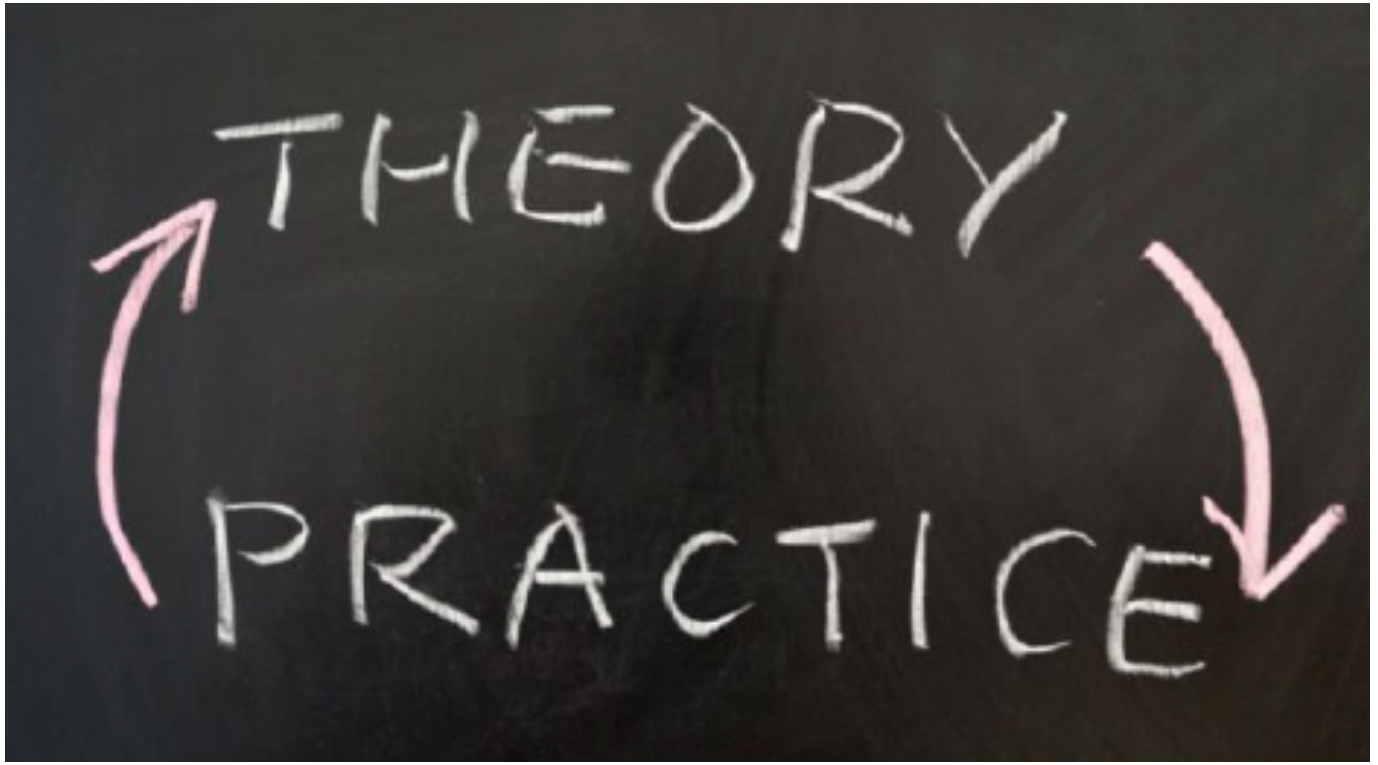
Joint Optimization – Code Review

CodeFlow: Improving the Code Review Process at Microsoft, Czerwonka et al. 2018.

Success practice transfer stories from research

- Automated testing (Facebook)
- Code review tools (Microsoft)
- Software Analytics (Hassan et al.)

Research success?


THEORY
PRACTICE

Dispelling myths in software engineering (or creating new ones?)

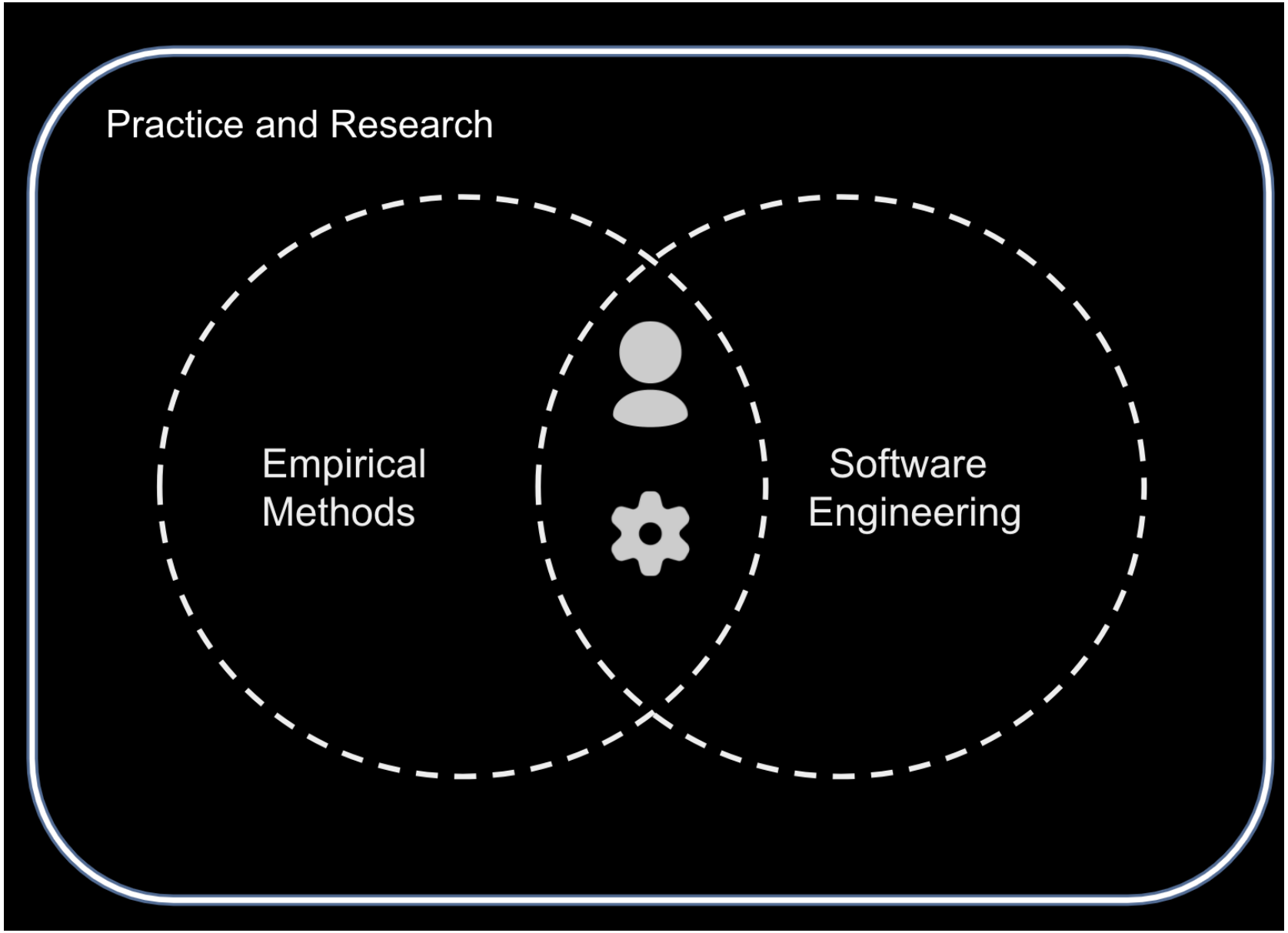
- Does increasing code coverage of testing reduce bugs? No, wasting time testing simple code may increase the presence of bugs! [Mockus et al.]
- Test driven development reduces bugs, but increases time delivering code [Nagappan et al.]
- Geographical distance doesn’t matter much [Bird et al.]
- Code clones do not reduce quality in code [Rahman et al.]

References for previous slide

- A. Mockus, N. Nagappan, and T. Dinh-Trong, “Test coverage and post-verification defects: A multiple case study,” in ESEM, 2009, pp. 291–301. (note see also this reference for a more recent paper on this! https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4915&context=sis_research)
- Nagappan, N., Maximilien, E.M., Bhat, T. et al. Realizing quality improvement through test driven development: results and experiences of four industrial teams. Empir Software Eng 13, 289–302.
- C. Bird, N. Nagappan, P. Devanbu, H. Gall and B. Murphy, "Does distributed development affect software quality?: an empirical case study of windows vista", Communications of the ACM, vol. 52, no. 8, pp. 85-93, 2009.
- M. S. Rahman and C. K. Roy, "On the Relationships Between Stability and Bug-Proneness of Code Clones: An Empirical Study," 2017 IEEE 17th International Working Conference on Source Code Analysis and Manipulation (SCAM), Shanghai, 2017, pp. 131-140.

"Academic software engineering research is a backwater with a tenuous connection to practical software development", Derek Jones

- Lack of industrial relevance (doesn't scale or solve industry problems) [Briand]
- Poor replication of software engineering studies [Menzies et al.]
- Poor actionability (practitioners know which modules are buggy…)
- Perils of mining software repositories [Kaliamvakou, German et al.]
- Lack of focus on human/social aspects [Storey et al.]


Practice and Research
Empirical Methods
Software Engineering

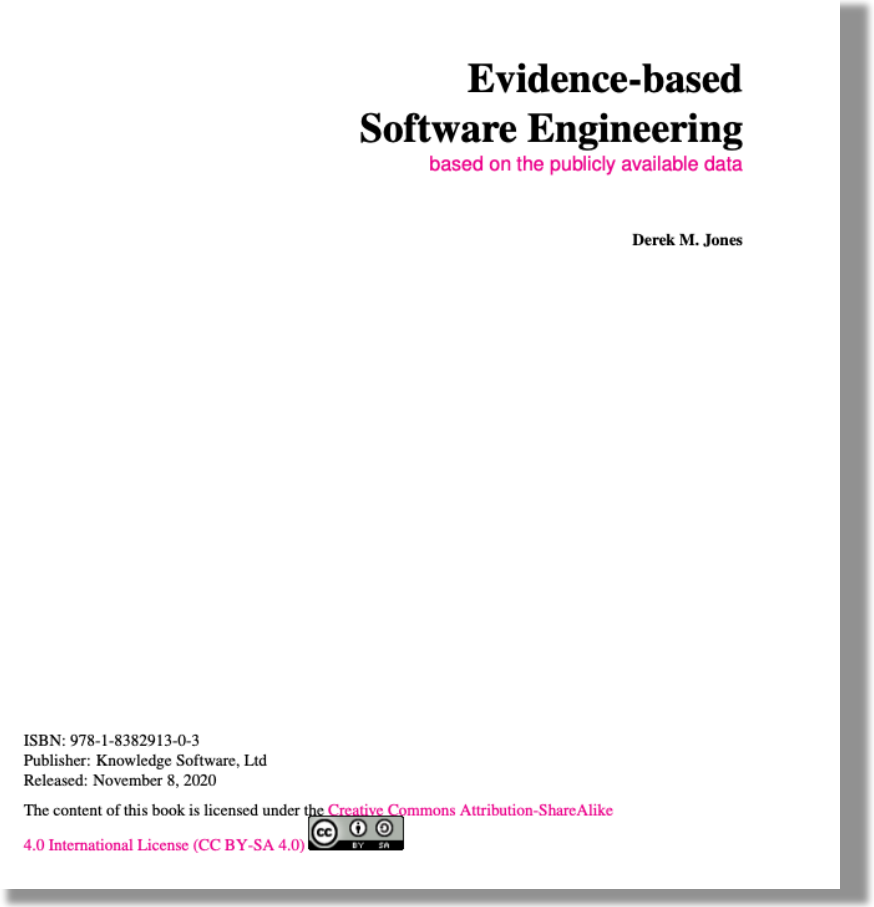
1968 NATO Conference on Software Engineering

- international experts on computer software who agreed on defining best practices for software grounded in the application of engineering.

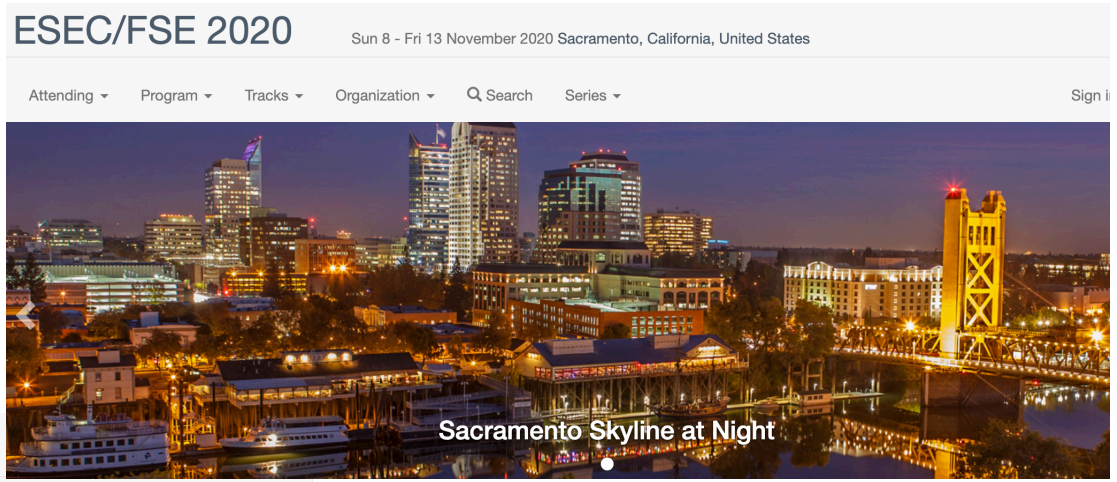
"Academic software engineering research has been a backwater primarily staffed by those interested in theory, with a tenuous connection to practical software development."

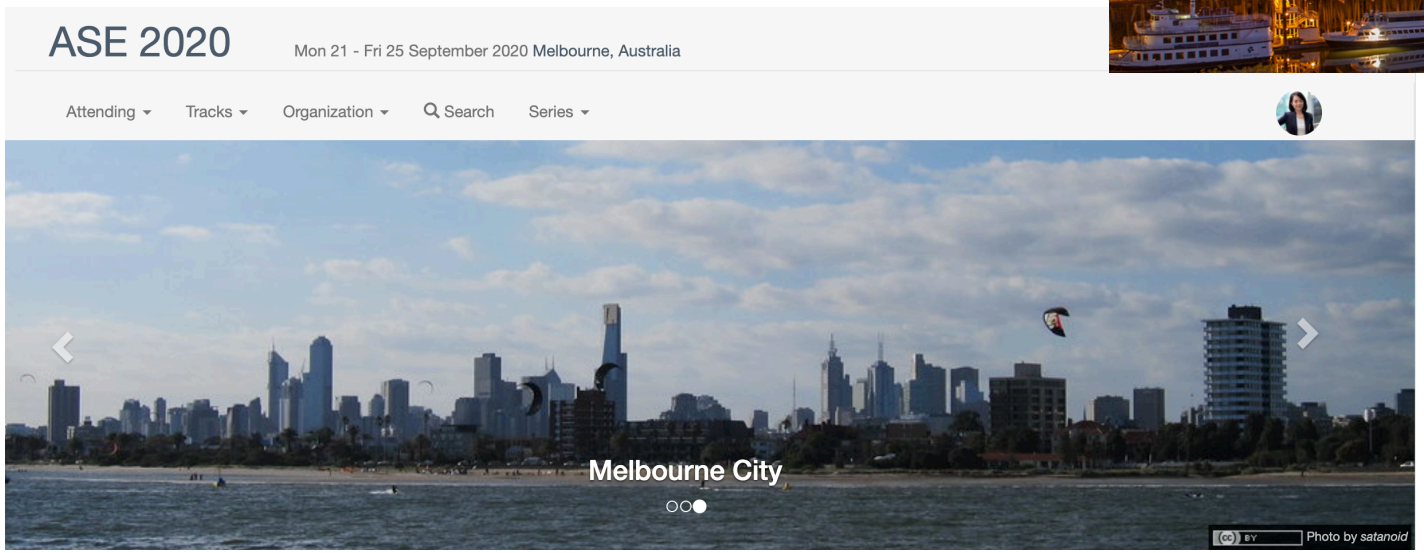
- Lack of **industrial relevance** (doesn't scale or solve industry

  problems) [Briand]
- **Poor replication** of software engineering studies [Menzies et al.]
- **Poor actionability** (practitioners know which modules are buggy…)
- **Perils of mining** software repositories [Kaliamvakou, German et al.]
- Lack of focus on **human/social aspects** [Storey et al.]




ICSE 21
43rd INTERNATIONAL
CONFERENCE ON
SOFTWARE ENGINEERING
MAY 23-29, 2021
MADRID, SPAIN


ESEC/FSE 2020
Sun 8 - Fri 13 November 2020 Sacramento, California, United States
Attending
Program
Tracks
Organization
Search
Series
Sacramento Skyline at Night


ASE 2020
Mon 21 - Fri 25 September 2020 Melbourne, Australia
Attending
Tracks
Organization
Search
Series
Melbourne City
Photo by satanoid

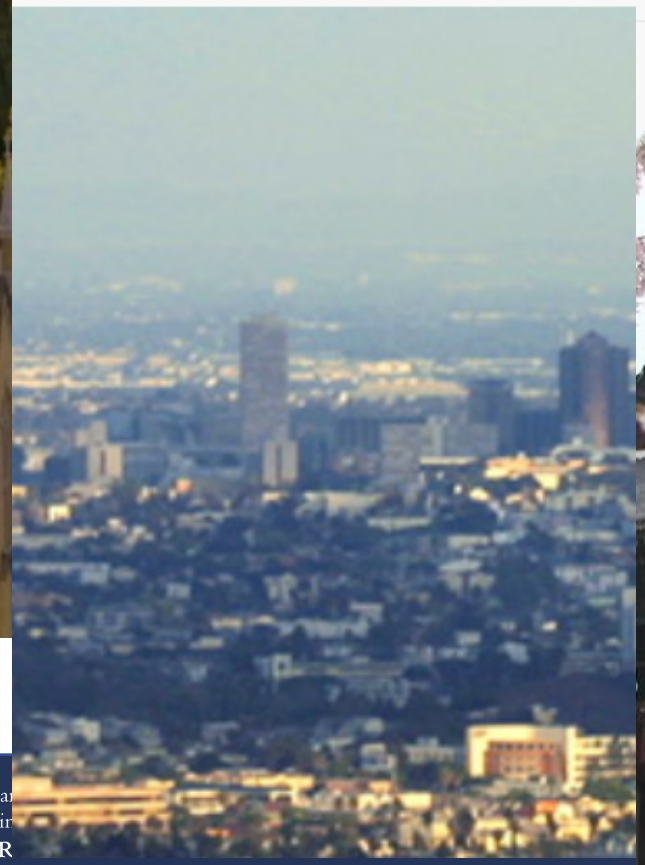
29th IEEE International Requirements Engineering Conference

Notre Dame, South Bend, USA
September 20-24, 2021

ISSTA 2020

Sat 18 - Wed 22 July 2020

Attending ▾ Sponsorship ▾ Progr



MSR 2020

Mon 29 - Tue 30 June 2020

Attending ▾ Travel Support Program ▾ Tracks ▾ Organization ▾ Search Series ▾

Agenda for today

- Introduction
  - Who are you?
  - What's your research?
  - What would make this course valuable to you?
- Why empirical methods?
- Research designs
- Course overview

Let’s go around the “room” for introductions:

Name (preferred name)

What's your (research) background?
What would make this course valuable to you?
One topic you are particularly interested in, if any?

Agenda

- Introduction
  - Who are you?
  - What's your research?
  - What would make this course valuable to you?
- Why empirical methods?
- Research designs
- Course overview

Is this your research plan?

- Step 1: Build a new tool
- Step 2:
- Step 3: Profit

Engineering vs. Science

- Traditional View:

Scientists…
create knowledge
study the world as it is
are trained in scientific method
use explicit knowledge
are thinkers

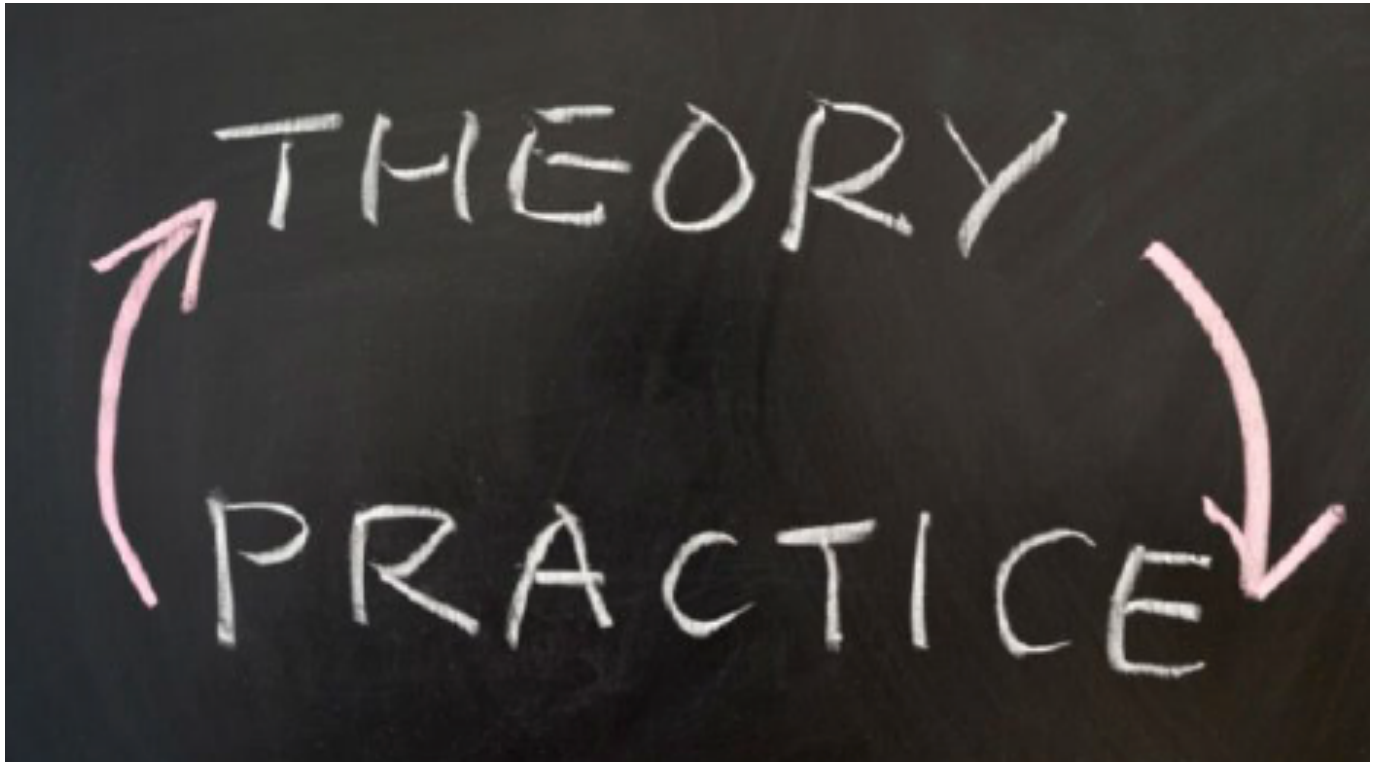
Engineers…
apply that knowledge
seek to change the world
are trained in engineering design
use tacit knowledge
are doers

- More realistic View

Scientists…
create knowledge
are problem-driven
seek to understand and explain
design experiments to test theories
prefer abstract knowledge
but rely on tacit knowledge

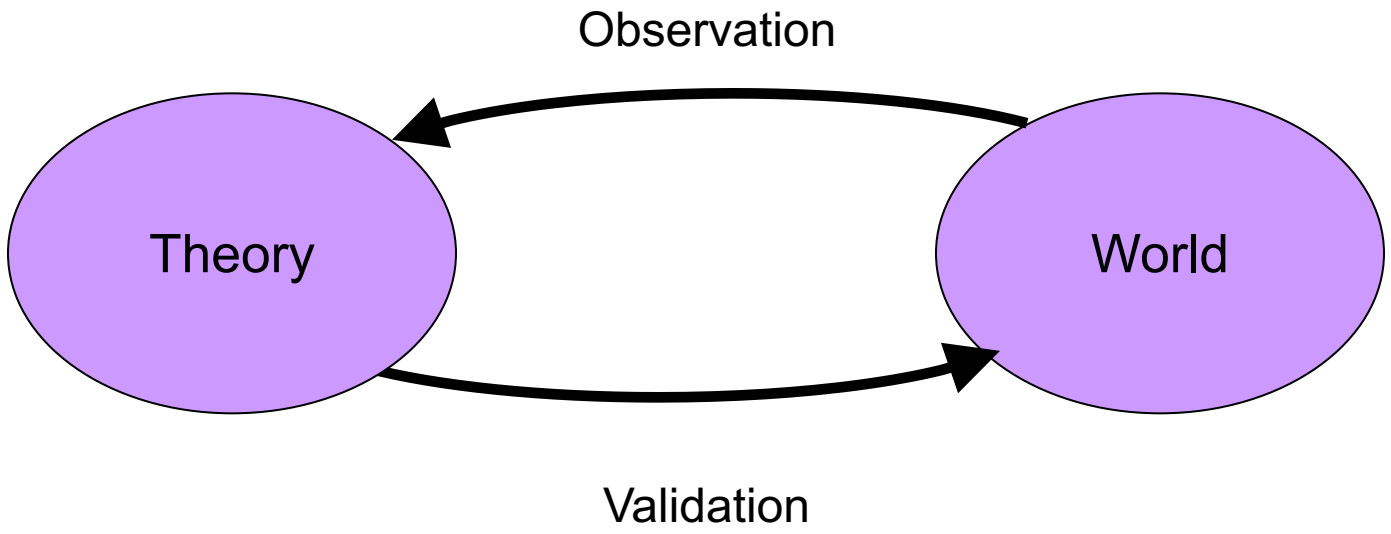
Engineers…
create knowledge
are problem-driven
seek to understand and explain
design devices to test theories
prefer contingent knowledge
but rely on tacit knowledge

Both involve a mix of design and discovery


THEORY
PRACTICE

Scientific Method

- No single “official” scientific method
- Somehow, scientists are supposed to do this:



Observe!


EXAMPLE

Help software developers
to better collaborate

+ Advances in tooling & SE principles
+ Insights from other disciplines
+ Mix a wide range of research methods

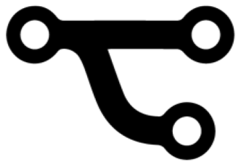
Problem → Intervention → Evaluation

Fork-based Dev. Changed Everything

Fork-based Development

Upstream

Fork/Branch



Fork-based Development

Upstream

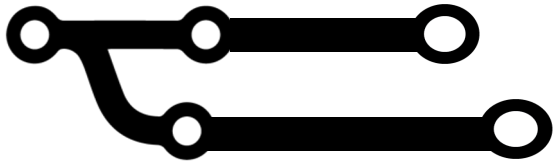
Fork/Branch

Commit

Fork-based Development

Upstream

Fork/Branch

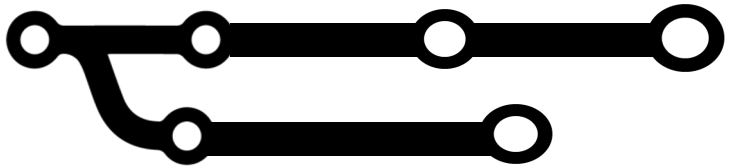


Commit

Fork-based Development

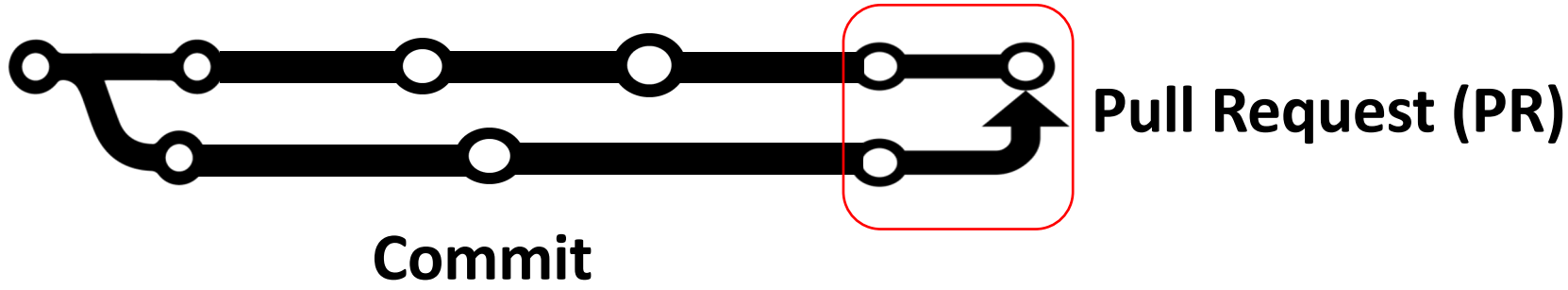
Upstream

Fork/Branch



Commit

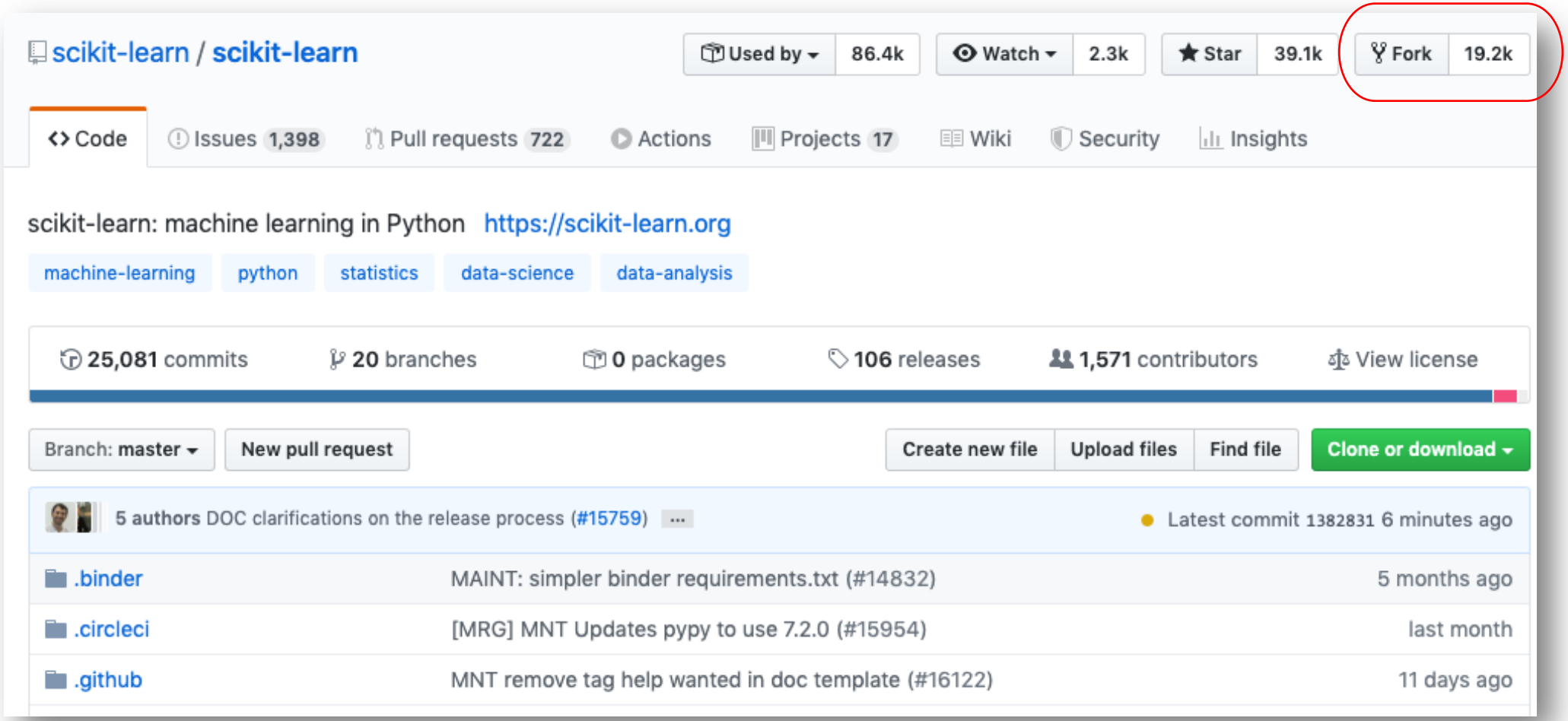
Fork-based Development



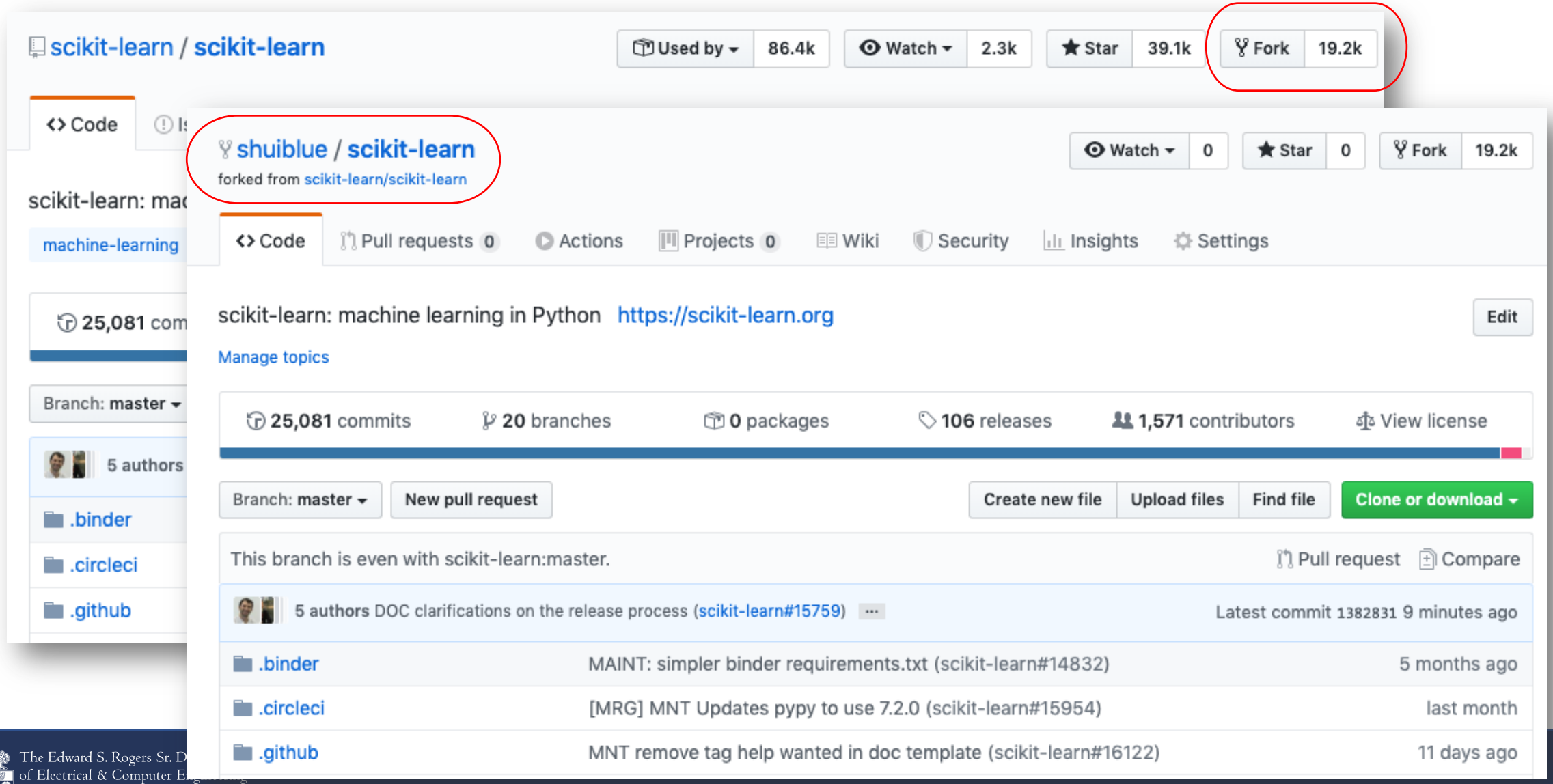
Fork-based / Branch-based / Pull-based Dev.

Pull Request / Merge Request

Fork-based Dev. Lowers Entry Barriers



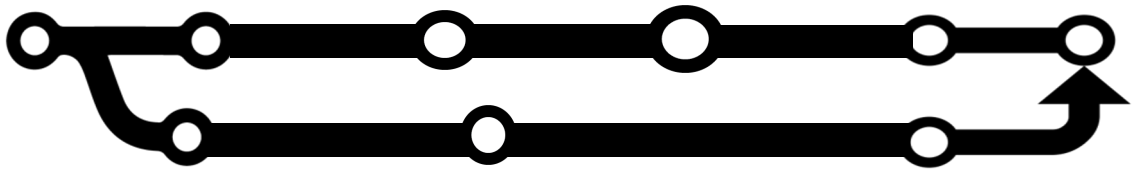
Fork-based Dev. Lowers Entry Barriers



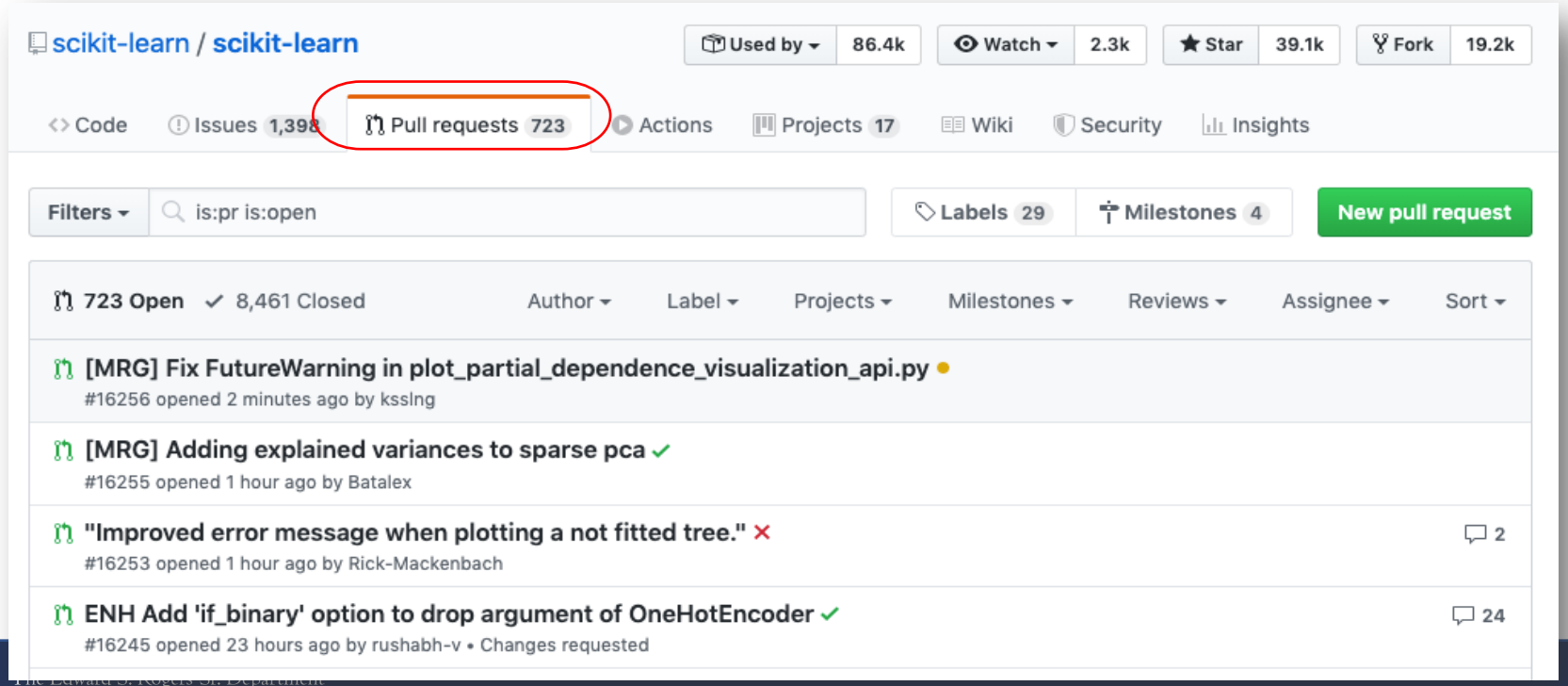
Fork-based Dev. Lowers Entry Barriers

Upstream

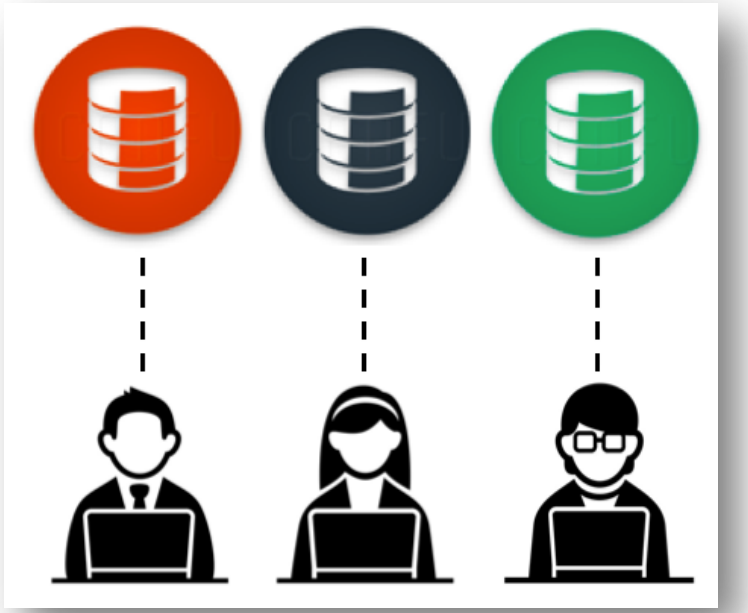
Fork/Branch



Pull Request (PR)



Fork-based Development



Fork-based Dev. Becomes Popular

#Forks	#GitHub Projects
>50	114,120
>500	9164
>1,000	2236
>5,000	198
>10,000	72
>100,000	2

[GHTorrent 2019-06]

Fork-based Dev. Becomes Popular

#Forks	#GitHub Projects
>50	114,120
>500	9164
>1,000	2236
>5,000	198
>10,000	72
>100,000	2

[GHTorrent 2019-06]

Fork-based Dev. Becomes Popular

#Forks	#GitHub Projects
>50	114,120
>500	9164
>1,000	2236
>5,000	198
>10,000	72
>100,000	2

[GHTorrent 2019-06]

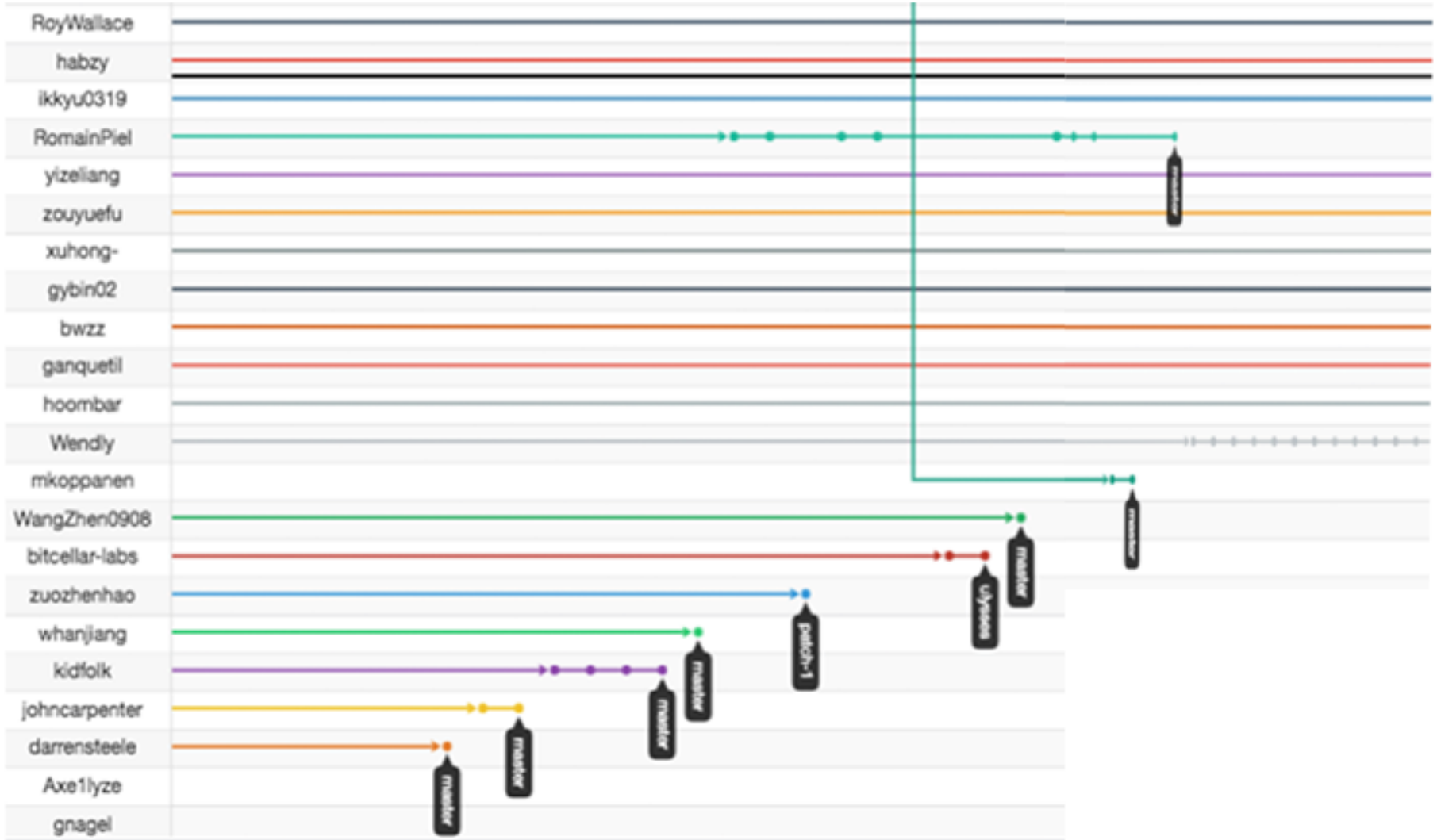
Fork-based Dev. Becomes Popular

Companies

https://github.com/customer-stories?type=enterprise

But

Problem -- Lost Contributions



Problem -- Redundant Development

foosel commented on Aug 22, 2017 — Owner

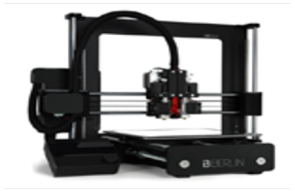
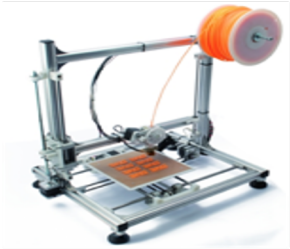
Sorry, but I can't stop laughing right now. I added *exactly* the same kind of functionality yesterday (just with a configurable ambient value and a debug command to also modify it during run time). See `fbcbb3f`

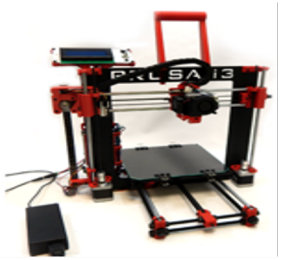
I can't believe this coincidence XD

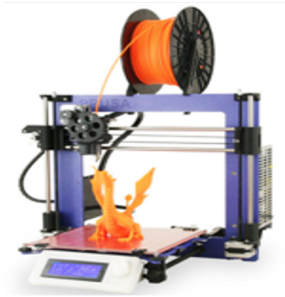
Noiredd commented on Nov 3, 2017 — Member

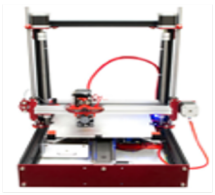
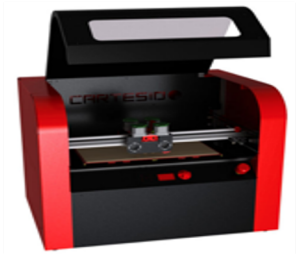
Duplicate of #5869 and #5972, partially also #5879.

Problem -- Fragmented Community

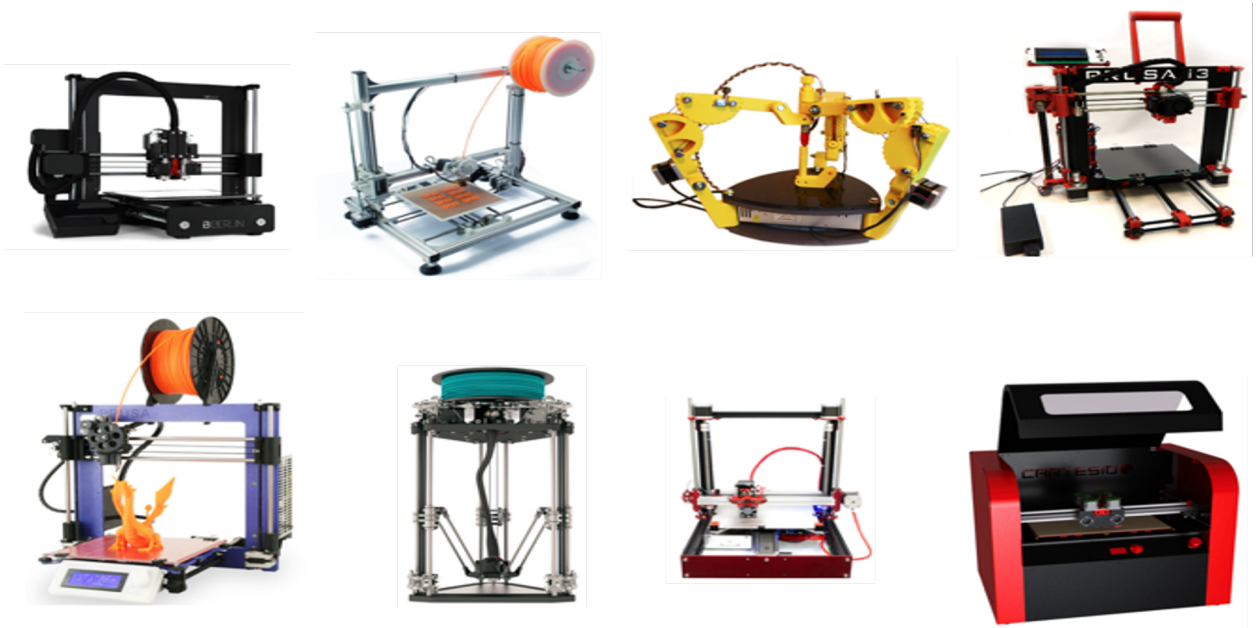







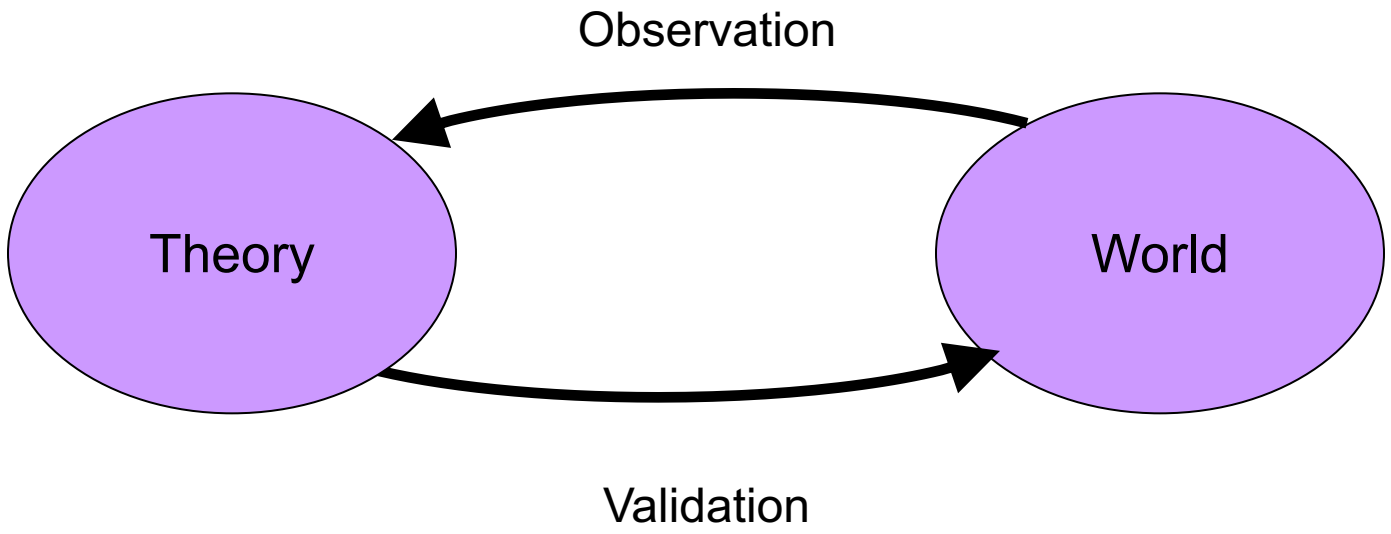



Problem -- Fragmented Community



Scientific Method

- No single “official” scientific method
- Somehow, scientists are supposed to do this:



Problem

Lost Contribution

Redundant Development

Fragmented Community

Problem

- Lost Contribution
- Redundant Development
- Fragmented Community

Is duplicate always bad?

Similar Problems Happen in Industry

It is hard for individual teams to know who is doing what, which features exist elsewhere, and what code changes are made in other forks [1,2].

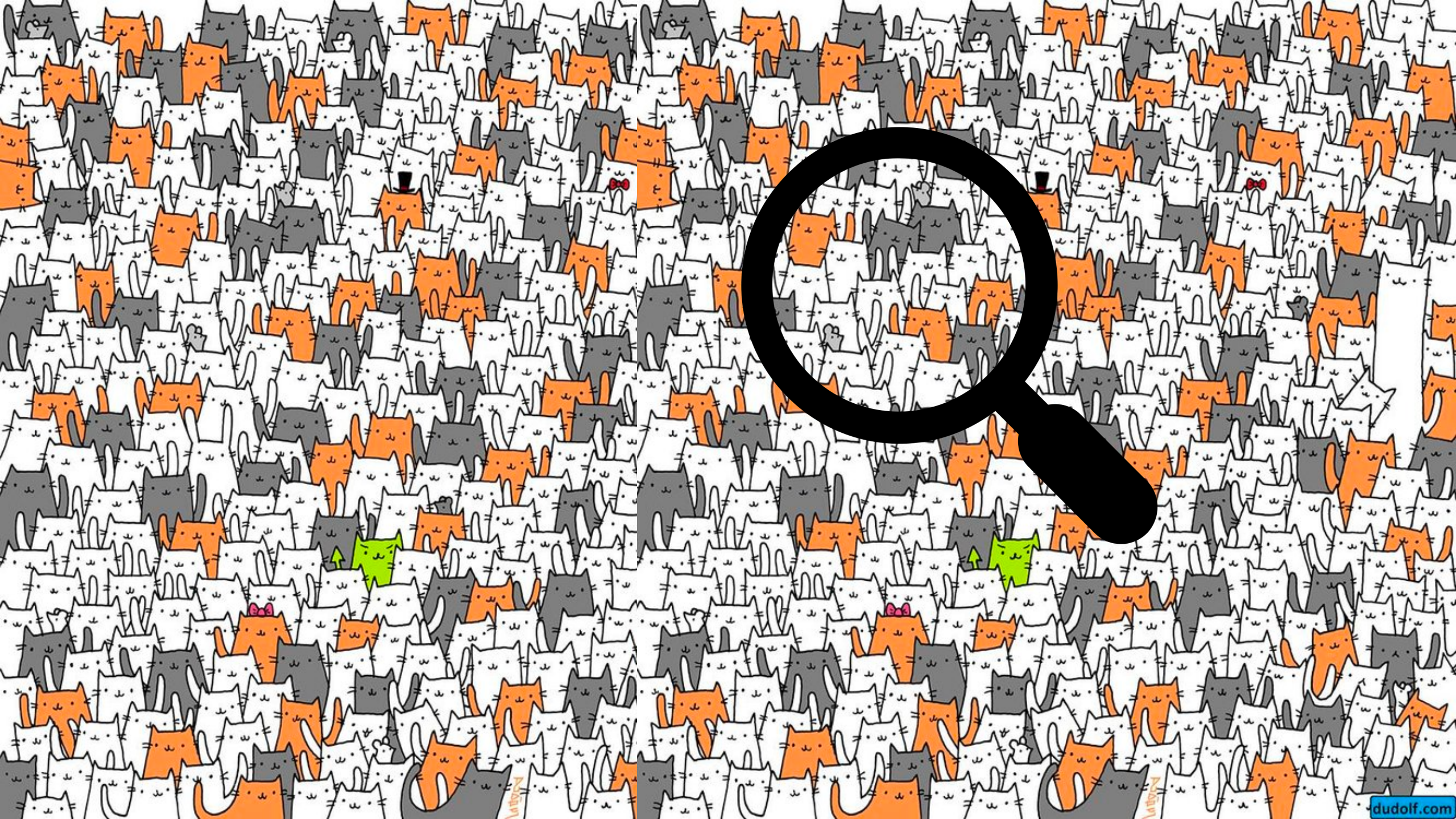
[1] Thorsten Berger, Divya Nair, Ralf Rublack, Joanne M Atlee, Krzysztof Czarnecki, and Andrzej Wąsowski. 2014. Three Cases of Feature-based Variability Modeling in Industry. In Proc. Int'l Conf. Model Driven Engineering Languages and Systems (MoDELS)
[2] Anh Nguyen Duc, Audris Mockus, Randy Hackbarth, and John Palframan. 2014. Forking and Coordination in Multi-platform Development: A Case Study. In Proc. Int'l Symp. Empirical Software Engineering and Measurement (ESEM). ACM

Problem

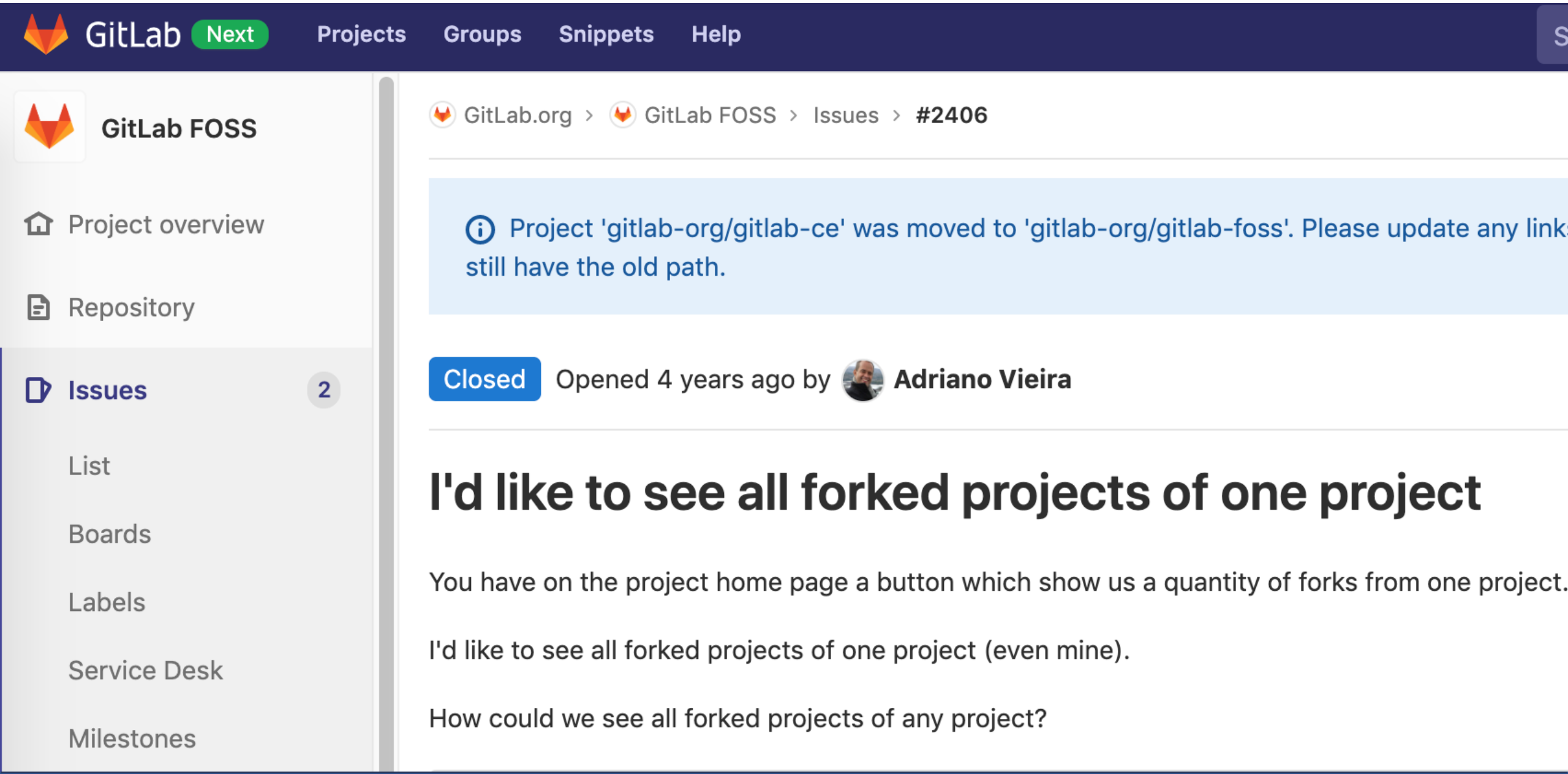
Lost Contribution

Redundant Development

Fragmented Community

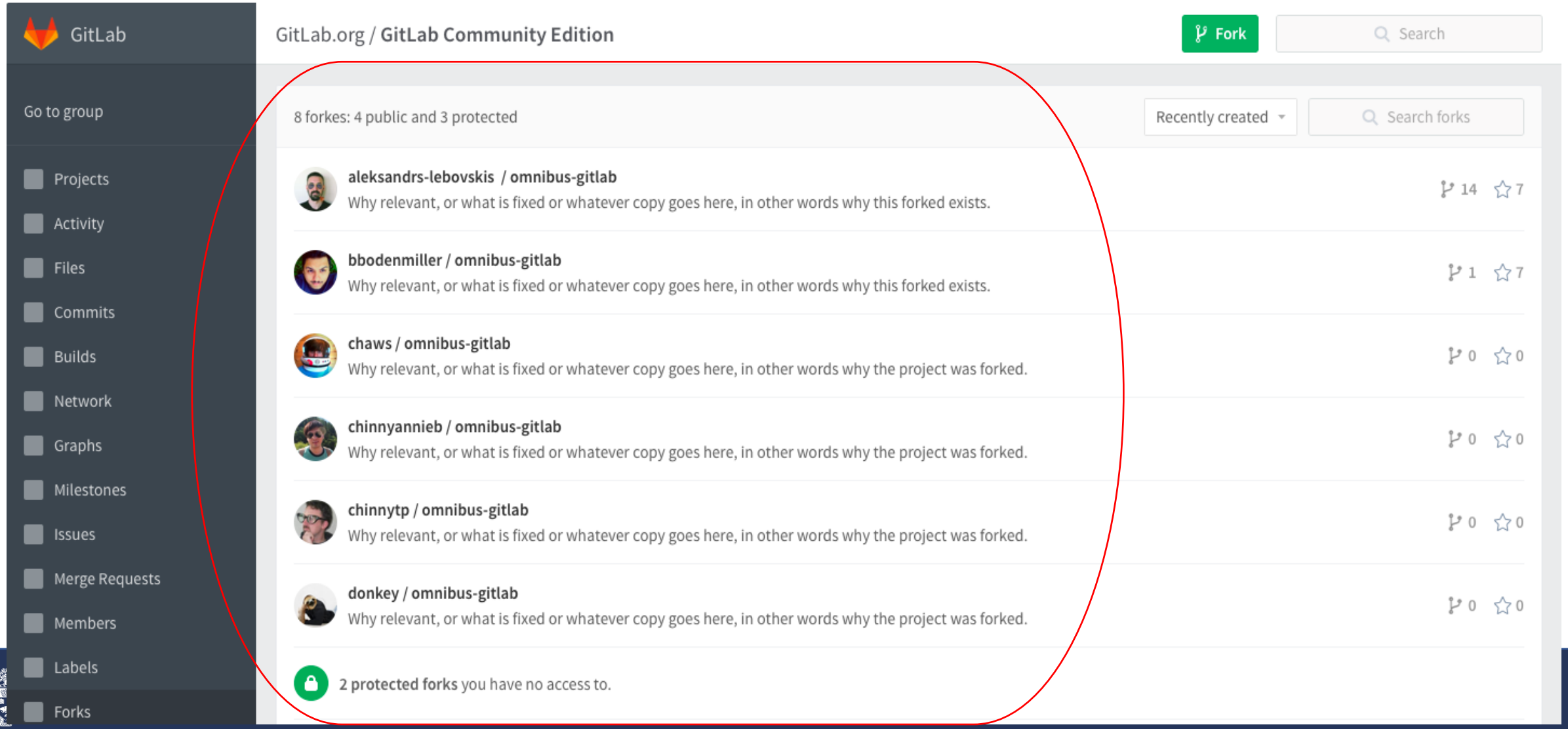

dudolf.com

Problem


GitLab Next
Projects
Groups
Snippets
Help
GitLab FOSS
Project overview
Repository
Issues 2
List
Boards
Labels
Service Desk
Milestones
GitLab.org › GitLab FOSS › Issues › #2406
Project 'gitlab-org/gitlab-ce' was moved to 'gitlab-org/gitlab-foss'. Please update any link
still have the old path.
Closed
Opened 4 years ago by
Adriano Vieira
I'd like to see all forked projects of one project
You have on the project home page a button which show us a quantity of forks from one project.
I'd like to see all forked projects of one project (even mine).
How could we see all forked projects of any project?

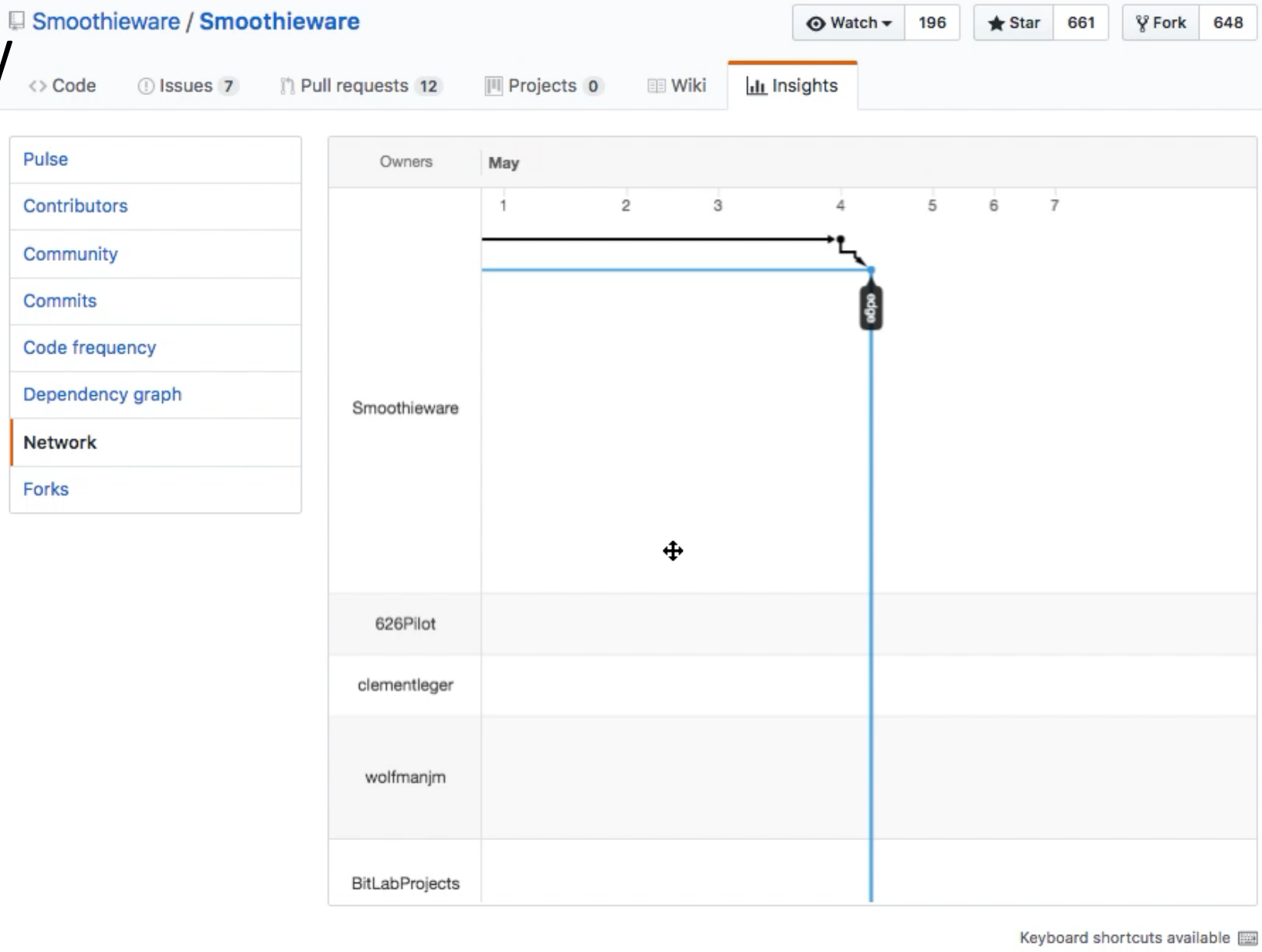
Problem

List of Forks



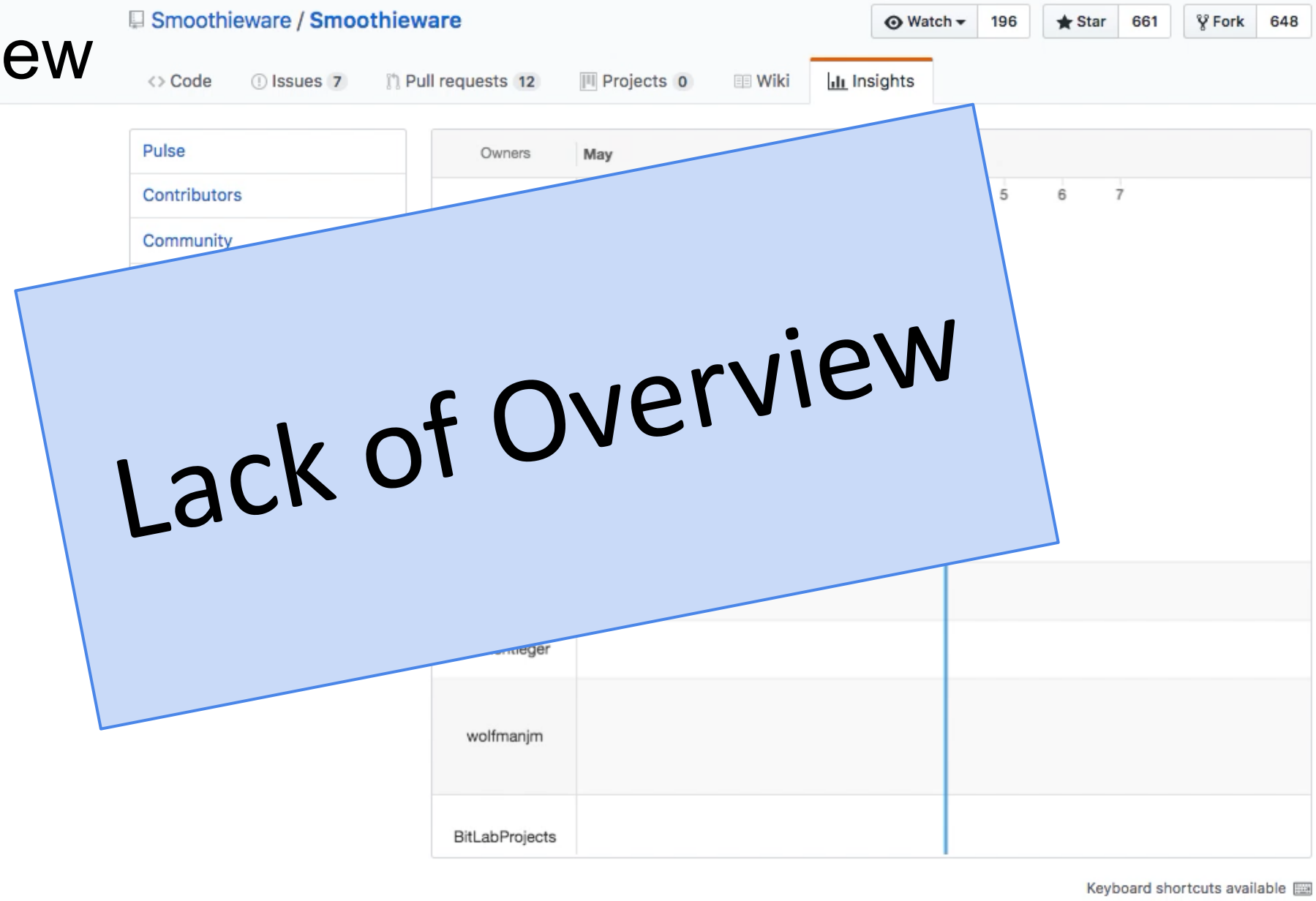
Problem

Network View



Problem

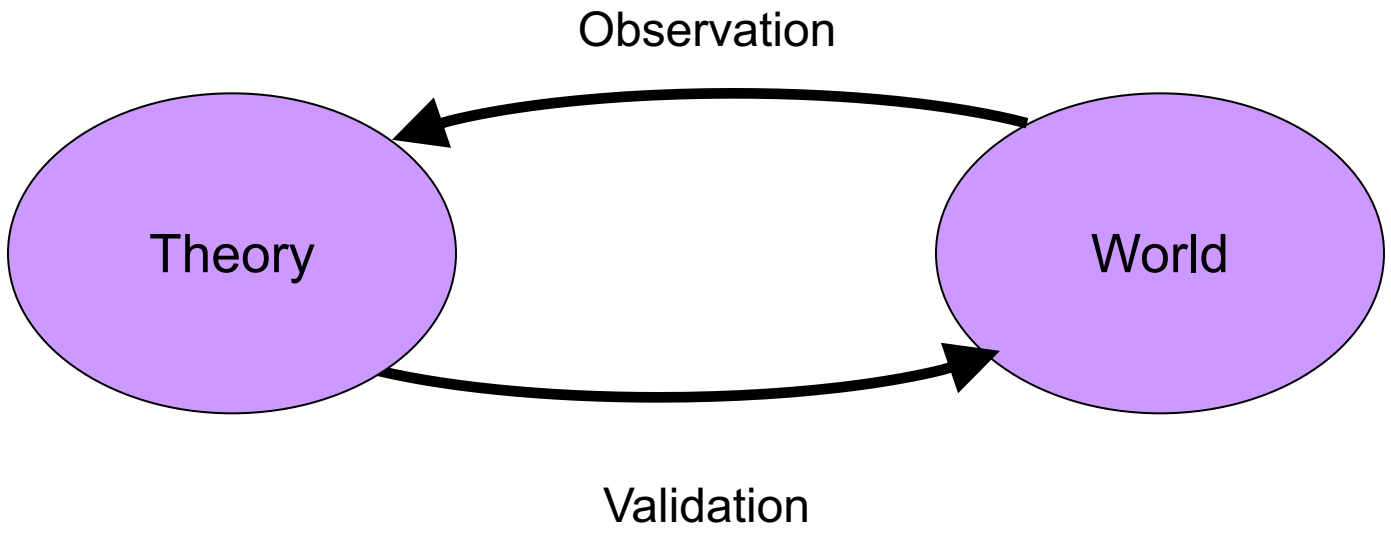
Network View



Lack of Overview

Scientific Method

- No single “official” scientific method
- Somehow, scientists are supposed to do this:



Improving Collaboration Efficiency

Software Dev.

Improving Collaboration Efficiency

Software Dev.

Distributed

Improving Collaboration Efficiency

Software Dev. — **Distributed** — Fork-Based

Improving Collaboration Efficiency

Software Dev.

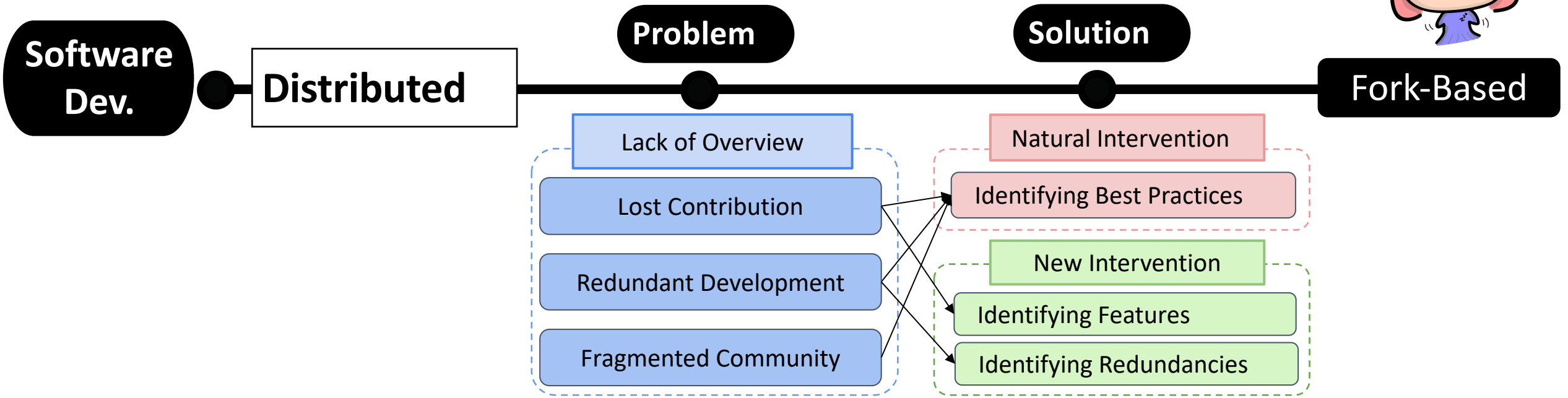
Distributed

Problem

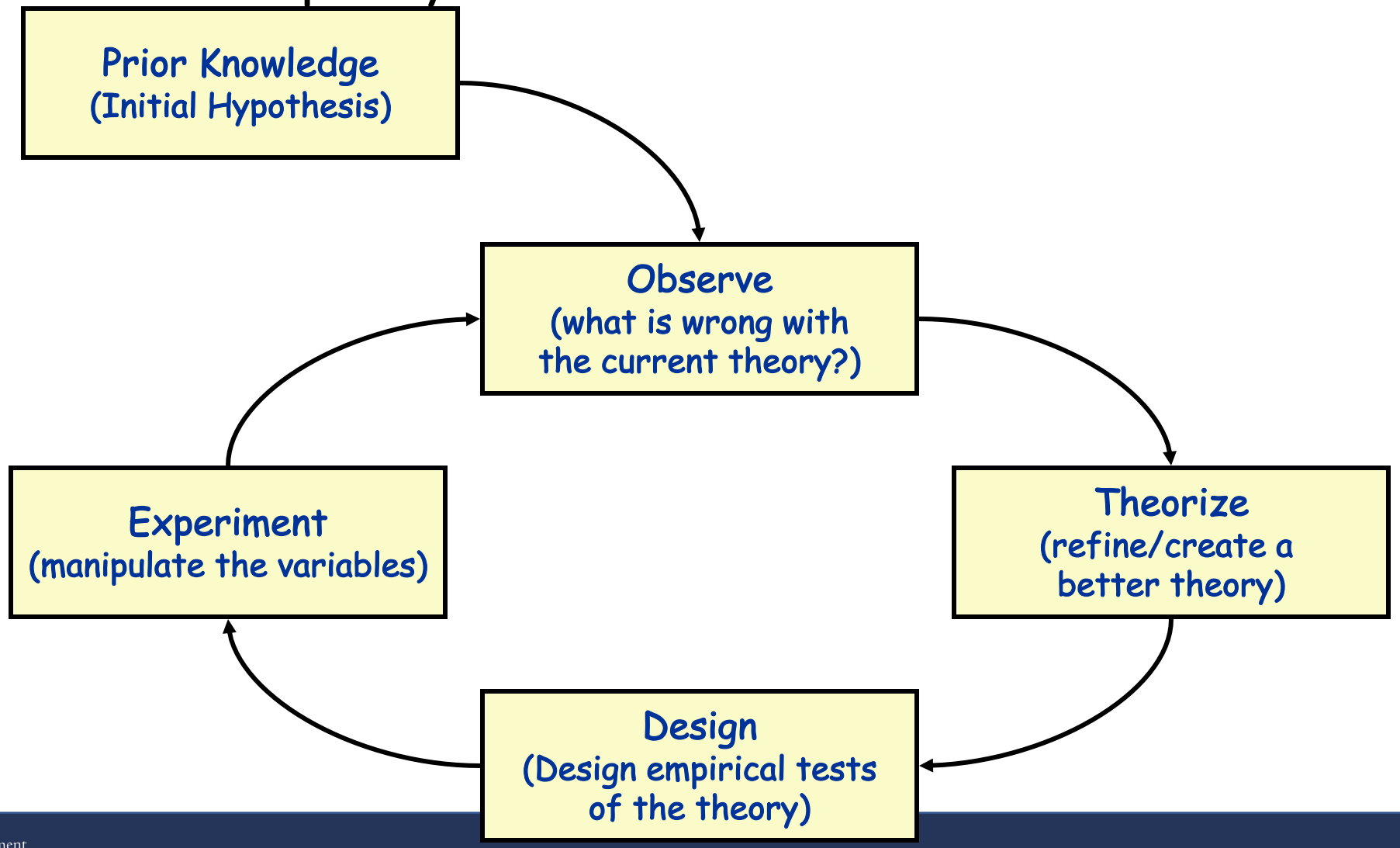
Fork-Based

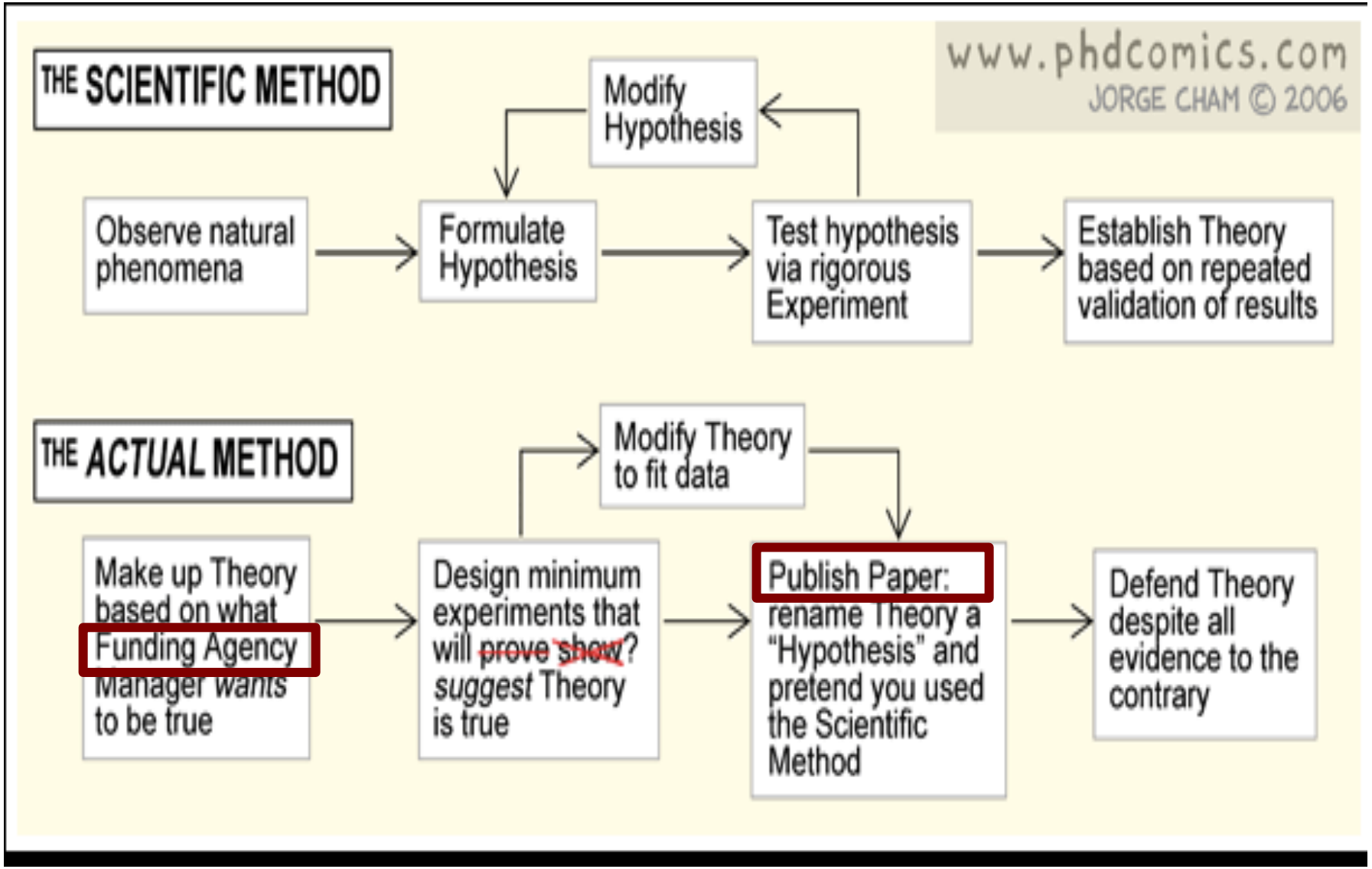
- Lack of Overview
- Lost Contribution
- Redundant Development
- Fragmented Community

Improving Collaboration Efficiency



Scientific Inquiry




THE SCIENTIFIC METHOD
Modify Hypothesis
Observe natural phenomena
Formulate Hypothesis
Test hypothesis via rigorous Experiment
Establish Theory based on repeated validation of results
www.phdcomics.com
JORGE CHAM © 2006
THE ACTUAL METHOD
Modify Theory to fit data
Make up Theory based on what Funding Agency Manager wants to be true
Design minimum experiments that will prove show? suggest Theory is true
Publish Paper: rename Theory a "Hypothesis" and pretend you used the Scientific Method
Defend Theory despite all evidence to the contrary

Observe!

Some Characteristics of Science

- Science seeks to improve our understanding of the world.
- Explanations are based on observations
  - Scientific truths must stand up to empirical scrutiny
  - Sometimes “scientific truth” must be thrown out in the face of new findings
- Theory and observation affect one another:
  - Our perceptions of the world affect how we understand it
  - Our understanding of the world affects how we perceive it
- Creativity is important
  - Theories, hypotheses, experimental designs
  - Search for elegance, simplicity

Empirical Methods

- Why?
  - Accurate perception is hard
  - How to gather evidence and draw conclusions
- There are many methods, no mechanical formulas

A Variety of Reasons for Doing a Study

- Observation
    - E.g., a better understanding of how software engineers/designers/… work
- Identification of problems with state-of-the-art
- Evaluating a new tool/technique
    - E.g., evidence that approach A is better than B

…

How to validate your claims?

A Variety of Types of Questions

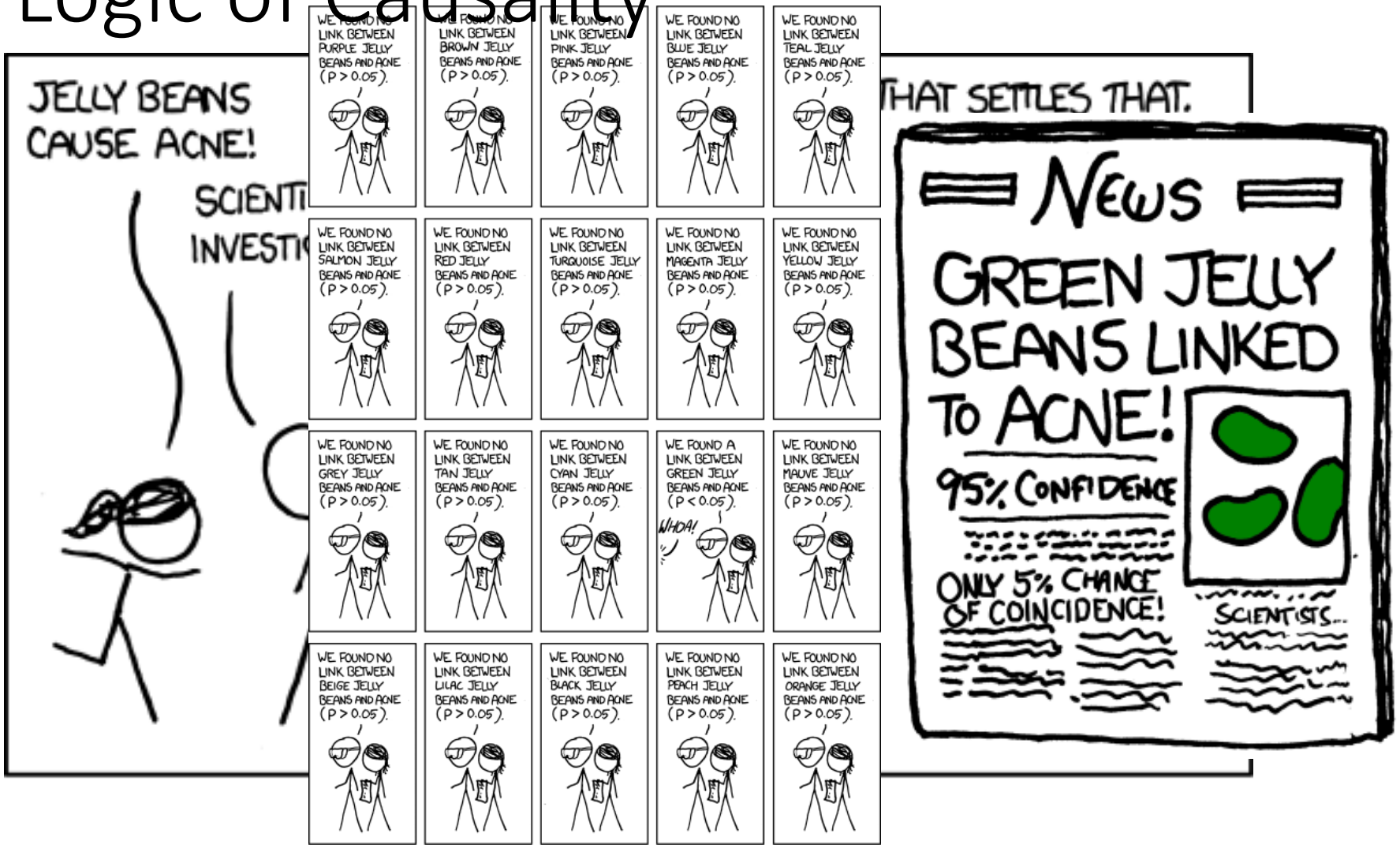
- What's going on here?
- What does an example look like?
- How are several instances same/different?
- What things/events co-occur?
- Do X and Y have a causal relation?
- Does the effect of X on Y vary depending on the value of Z?
- …

How to approach an answer?

Aside: Logic of Causality

- In a sense, the logic of hypothesis testing is flawed
  - If X, then Y (if theory true, hypothesis must be true)
  - Observe that Y
  - Therefore, X
- Example
  - If my technology is effective, people using it will perform better
  - People using it perform better
  - Therefore, my technology is effective
- Fallacy: affirming the consequent
- Problem of confounding – multiple possible causes

Aside: Logic of Causality



Source: http://xkcd.com/882/

Agenda

- Introduction
  - Who are you?
  - What's your research?
  - What would make this course valuable to you?
- Why empirical methods?
- Research designs
- Course overview


EXAMPLE

Meet Stuart Dent

- Name:
  - Stuart Dent (a.k.a. “Stu”)
- Advisor:
  - Prof. Helen Back
- Topic:
  - Merging Stakeholder views in Model Driven Development
- Status:
  - 2 years into his PhD
  - Has built a tool [*Stu-Merge*]
  - Needs an evaluation plan

Stu's Evaluation Plan

Rational software

- Formal Experiment
  - Independent Variable: Stu-Merge vs. Rational Architect (RA)
  - Dependent Variables: Correctness, Speed, Subjective Assessment
  - Task: Merging Class Diagrams from two different stakeholders' models
  - Subjects: Grad Students in SE
  - H_1: "Stu-Merge produces correct merges more often than RA"
  - H_2: "Subjects produce merges faster with Stu-Merge than with RA"
  - H_3: "Subjects prefer using Stu-Merge to RA"
- Results
  - H_1 accepted (strong evidence)
  - H_2 & H_3 rejected
  - Subjects found the tool unintuitive

Threats to Validity

- Construct Validity
  - What do we mean by a merge? What is correctness?
  - 5-point scale for subjective assessment - insufficient discriminatory power
    - (both tools scored very low)
- Internal Validity
  - Confounding variables: Time taken to learn the tool; familiarity
    - Subjects were all familiar with RA, not with Stu-merge
- External Validity
  - Task representativeness
    - class models were of a toy problem
  - Subject representativeness
    - Grad students as sample of what population?
- Theoretical Reliability
  - Researcher bias
    - subjects knew Stu-merge was Stu's own tool

"Measuring programming progress by lines of code is like measuring aircraft building progress by weight."

Threats to Validity

- Construct Validity
  - What do we mean by a merge? What is correctness?
  - 5-point scale for subjective assessment - insufficient discriminatory power
    - (both tools scored very low)
- Internal Validity
  - Confounding variables: Time taken to learn the tool; familiarity
    - Subjects were all familiar with RA, not with Stu-merge
- External Validity
  - Task representativeness
    - class models were of a toy problem
  - Subject representativeness
    - Grad students as sample of what population?
- Theoretical Reliability
  - Researcher bias
    - subjects knew Stu-merge was Stu's own tool

Improving Collaboration Efficiency

Software Dev.

Distributed

Problem

Fork-Based

- Lack of Overview
- Lost Contribution
- Redundant Development
- Fragmented Community

Human-subject Study - Usefulness

Can INFOX help developers to gain a better overview of repository forks?

Feature	ibradypod/phantomjs, last commmit:May 28	LOC
onre.	onresourcerequest, bodi, downloadmultibuffer, qnetworkrepli, respons, qbytearray, data, reply, buffers	28
hea.	header, getcookiestringfromurl, bodi, cookie, get, qurl, qnetworkrepli, respons, url	10
sett.	settings, a, phantomcfg, not, bug, fix, websecurityen, qwebset, setattribute, qwebsettings	2

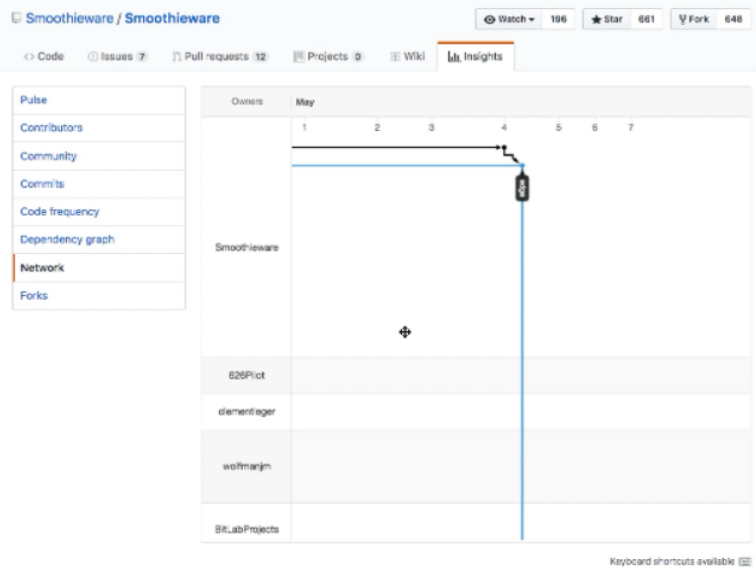
Feature	raff/phantomjs, last commmit: Mar 5	LOC
dow.	download, com, pull, file, ad, support, ariya, http	39
get	get, qt, are, kei_enter, el, mouse, require, clicks, hard, absolute, setfocus, button, coordinates, keypress	29

Feature	ricokahler/phantomjs, last commmit:Feb 2	LOC
readlin	readlin, asyncreadrequest, asyncread, qobject, qstring, readline, data, qvariant, m_file, m_data, file, read	30
uint	uint, tmp_value_, value, tmp, octet, qvariant, data, namesize_, readrawdata, fromvalue	80
frame	frame, bmconsumeok, bmdeliver, method_id, id_enum, bmgetempty, bmreject, bmrecover	29
cono.	conoack, 0x04, consumeoptions, coexclusive, declar_flag, consumeoption, conolocal, conowait, flag	36

Feature	DeviaVir/phantomjs, last commmit: Jan 25, 2016	LOC
allow	allow, set, customwebpag, ratio, m_customwebpag, devicepixelratio, webpage, setdevicepixelratio	7

Problem

Network View



Human-subject Study - Usefulness

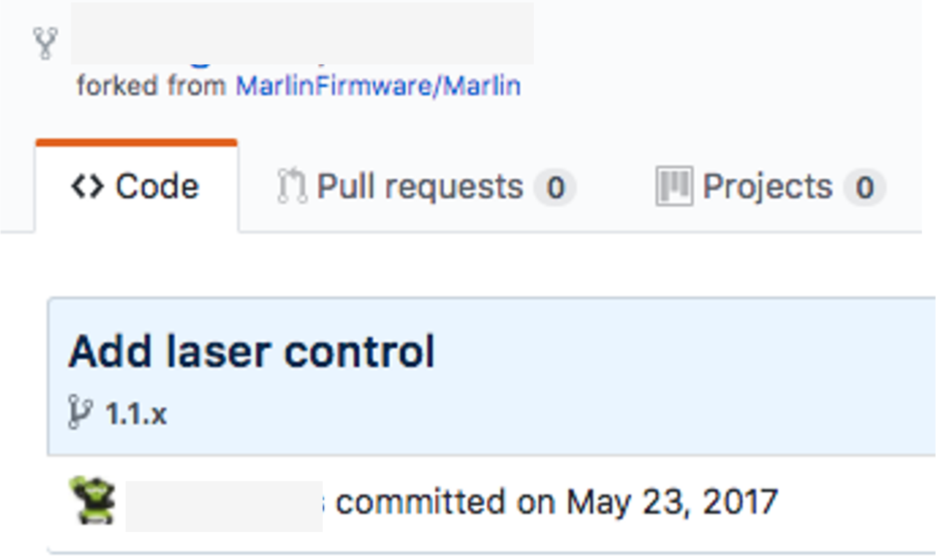
Interesting and Reusable Contribution

P5: "If it is only exists in this fork, then I want to somehow get this fork into my fork."

Human-subject Study - Usefulness

Redundant Development

"It does look like somebody did a very simple one-function.
I think they should use our code, there is great reason to use it."



P3

Experiments as Clinical Trials

Why would we expect it to be better?

Why do we need to know?

What will we do with the answer?

Is drug A better than drug B?

Better at doing what?

Better in what situations?

Better in what way?

You gotta have a theory!

Some Definitions

- A model is an abstract representation of a phenomenon or set of related phenomena
  - Some details included, others excluded
- A theory is a set of statements that explain a set of phenomena
  - Serves to explain and predict
  - Precisely defined terminology
  - Concepts, relationships, causal inferences
  - (operational definitions for theoretical terms)
- A hypothesis is a testable statement derived from a theory
  - A hypothesis is not a theory!


EXAMPLE

Projects are different

VS

- Centralized Mgmt
- Upfront Coordination through Issue Tracker

- De-centralized Mgmt
- No Upfront Coordination

Coordination Mechanism Affects Forking Practices

Centralization makes it easier to coordinate the divisions' product types but more difficult to take advantage of the divisions' private information. [Brandts et al. 2018]

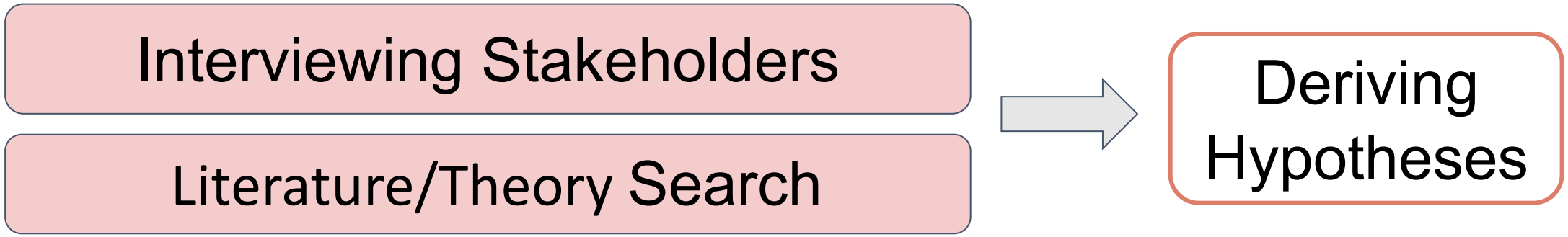
Organizational Theory

+ Advances in tooling & SE principles
+ Insights from other disciplines
+ Mix a wide range of research methods

Research Question

What characteristics and practices of a project associate with efficient forking practices?

Research Method



Derive Hypotheses

Centralized Management ➔ Larger portion of contributing forks

Operationalization - Centralized Management

Measure: $\frac{\text{Number of PRs referring to an Existing Issue}}{\text{All the PRs}}$

Fix issue #13048 - Documentation regarding p-value bootstrapping #14759

Closed achievermina wants to merge 7 commits into scikit-learn:master from achievermina:p_valueBootstrapping

Conversation 9 Commits 7 Checks 11 Files changed 2

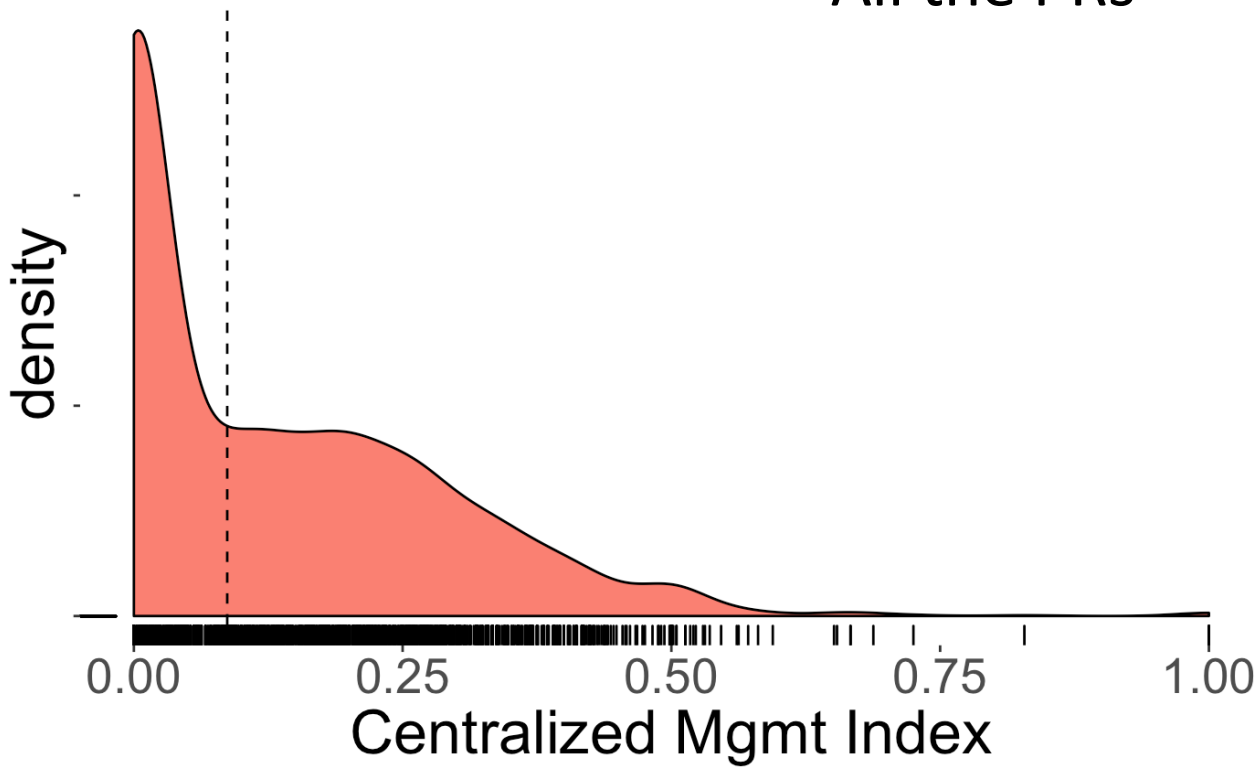
achievermina commented 3 days ago • edited

Issue #13048

Operationalization - Centralized Management

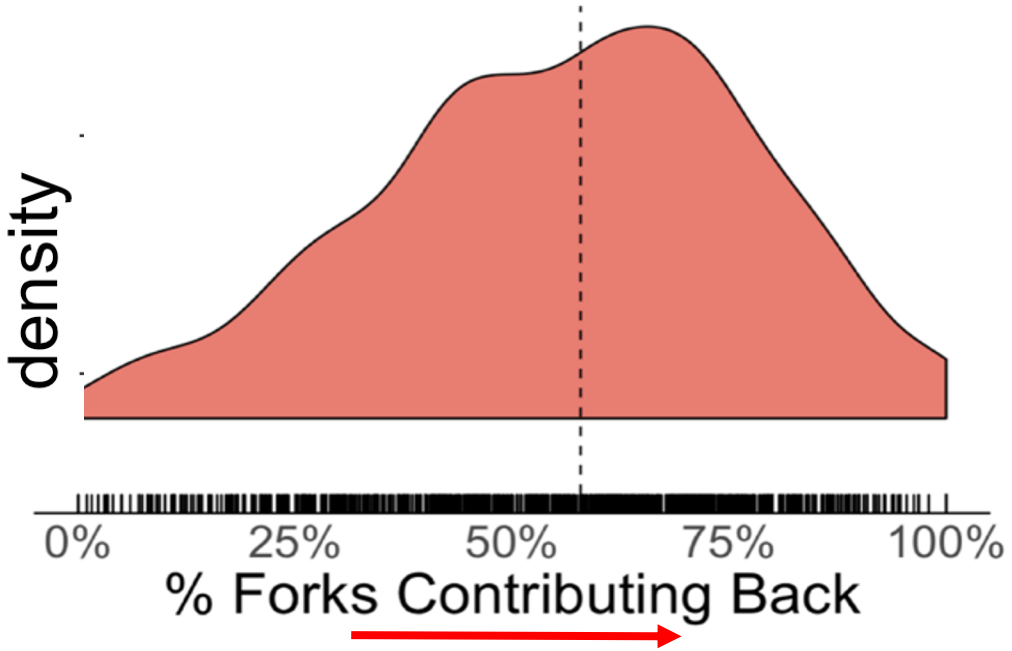
Measure:

$$\frac{\text{Number of PRs referring to an Existing Issue}}{\text{All the PRs}}$$



Operationalization – Contributing Forks

Measure: $\frac{\text{Number of Forks submitted PR(s)}}{\text{All the Active Forks}}$



more efficient

A simpler definition

A (good) Theory is the best explanation of all the available evidence

The Role of Theory Building

- Theories lie at the heart of what it means to do science.
  - Production of generalizable knowledge
- Theory provides orientation for data collection
  - Cannot observe the world without a theoretical perspective
- Theories allow us to compare similar work
  - Theories include precise definition for the key terms
  - Theories provide a rationale for which phenomena to measure
- Theories support analytical generalization
  - Provide a deeper understanding of our empirical results
  - …and hence how they apply more generally
  - Much more powerful than statistical generalization

Stu's Theory

- Background Assumptions
  - Large team projects, models contributed by many actors
  - Models are fragmentary, capture partial views
  - Partial views are inconsistent and incomplete most of the time
- Basic Theory
  - (Brief summary:)
  - Model merging is an exploratory process, in which the aim is to discover intended relationships between views. 'Goodness' of a merge is a subjective judgment. If an attempted merge doesn't seem 'good', many need to change either the models, or the way in which they were mapped together.
  - [Still needs some work]
- Derived Hypotheses
  - Useful merge tools need to represent relationships explicitly
  - Useful merge tools need to be complete (work for any models, even if inconsistent)

What type of question are you asking?

- ➜Existence:
  - Does X exist?
- ➜Description & Classification
  - What is X like?
  - What are its properties?
  - How can it be categorized?
  - How can we measure it?
  - What are its components?
- ➜Descriptive-Comparative
  - How does X differ from Y?
- ➜Frequency and Distribution
  - How often does X occur?
  - What is an average amount of X?
- ➜Descriptive-Process
  - How does X normally work?
  - By what process does X happen?
  - What are the steps as X evolves?
- ➜Relationship
  - Are X and Y related?
  - Do occurrences of X correlate with occurrences of Y?
- ➜Causality
  - Does X cause Y?
  - Does X prevent Y?
  - What causes X?
  - What effect does X have on Y?
- ➜Causality-Comparative
  - Does X cause more Y than does Z?
  - Is X better at preventing Y than is Z?
  - Does X cause more Y than does Z under one condition but not others?
- ➜Design
  - What is an effective way to achieve X?
  - How can we improve X?

What type of question are you asking?

Exploratory

→Existence:
- Does X exist?

→Description & Classification
- What is X like?
- What are its properties?
- How can it be categorized?
- How can we measure it?
- What are its components?

→Descriptive-Comparative
- How does X differ from Y?

Correlation

→Relationship
- Are X and Y related?
- Do occurrences of X correlate with occurrences of Y?

Causal Relationship

→Causality
- Does X cause Y?
- Does X prevent Y?
- What causes Y?
- What effect does X have on Y?

→Causality-Comparative
- Does X cause more Y than does Z?
- Is X better at preventing Y than is Z?
- Does X cause more Y than does Z under one condition but not others?

Baserate

→Frequency and Distribution
- How often does X occur?
- What is an average amount of X?

→Descriptive-Process
- How does X normally work?
- By what process does X happen?
- What are the steps as it evolves?

Design

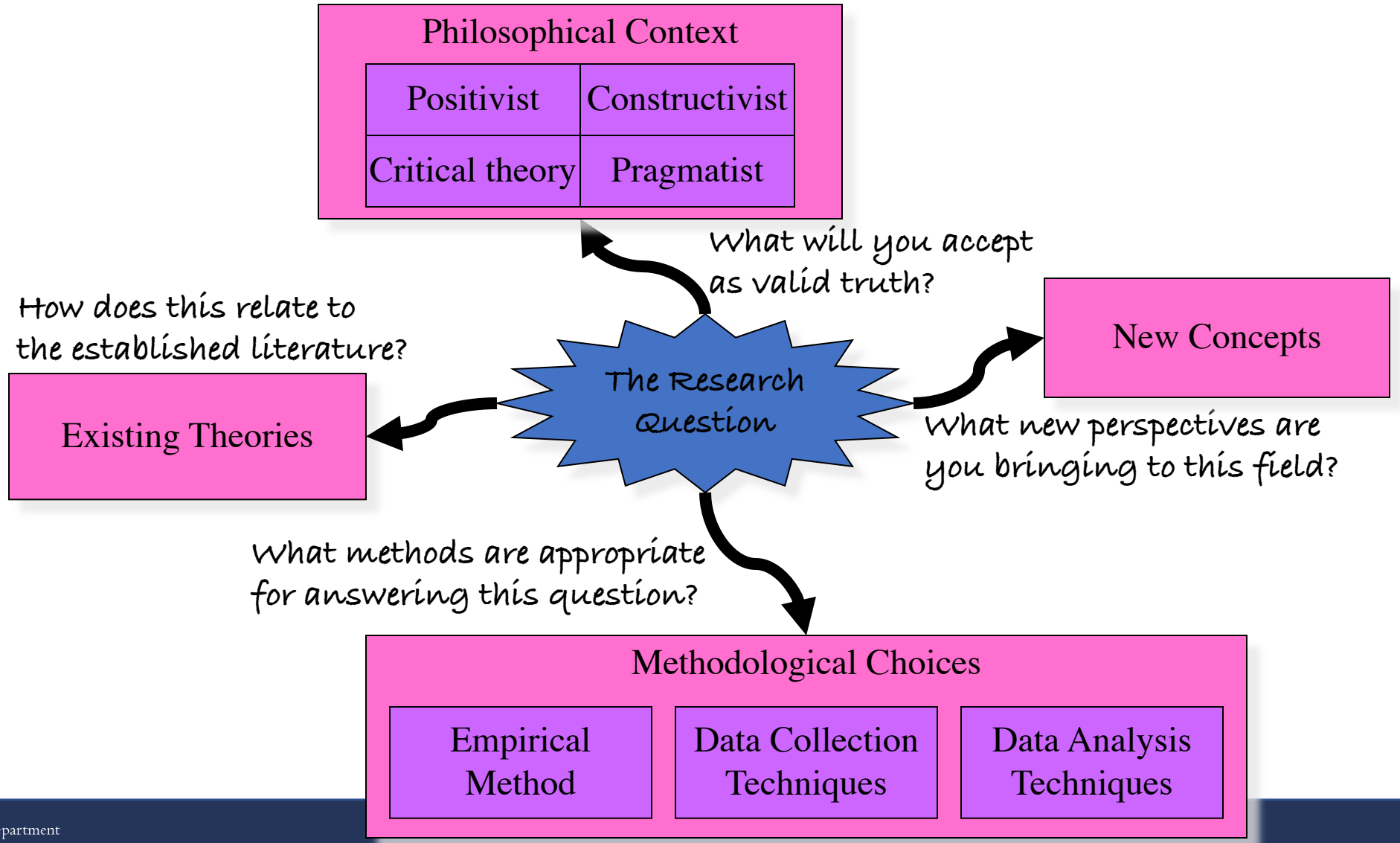
→Design
- What is an effective way to achieve X?
- How can we improve X?

Stu's Research Question(s)

- Existence
  - Does model merging ever happen in practice?
- Description/Classification
  - What are the different types of model merging that occur in practice on large scale systems?
- Descriptive-Comparative
  - How does model merging with explicit representation of relationships differ from model merging without such representation?
- Causality
  - Does an explicit representation of the relationship between models cause developers to explore different ways of merging models?
- Causality-Comparative
  - Does the algebraic representation of relationships in Stu's tool lead developers to explore more than do pointcuts in AOM?

Pick just one for now…

Putting the Question in Context



Many available methods…

Common "in the lab" Methods	Common "in the wild" Methods
Controlled Experiments • Rational Reconstructions • Exemplars • Benchmarks • Simulations	Quasi-Experiments Case Studies Survey Research Ethnographies Action Research

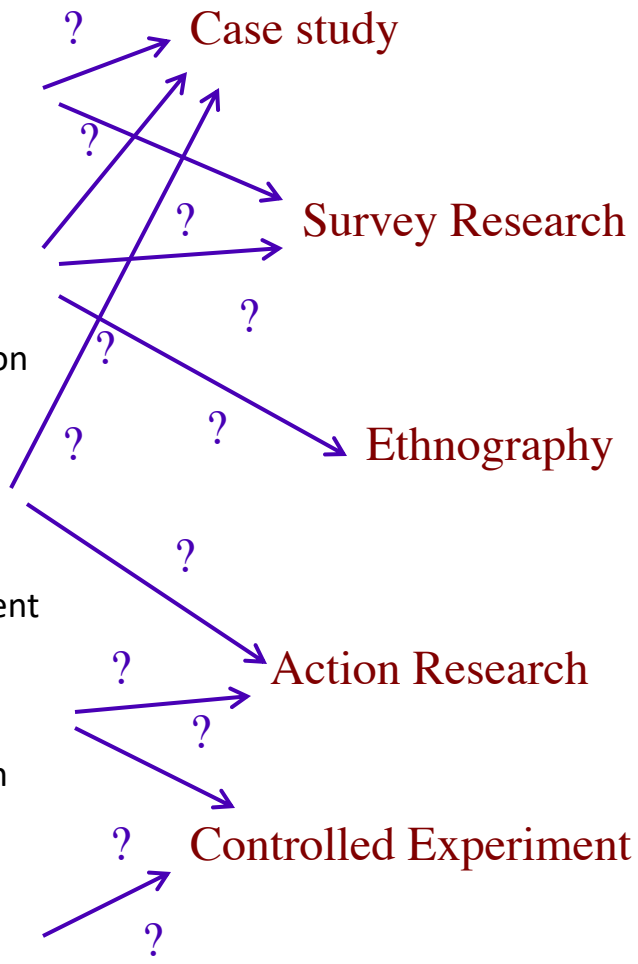
- Artifact/Archive Analysis ("mining"!)

Empirical Methods

- Used in many forms and phases of research
    - Understand problem
    - Current practice
    - Demonstrate utility of solution
- Selection of methods depends on
    - State of knowledge
    - Question researcher is asking
    - Nature of contribution
- **Each method has its own standards and techniques for rigor**

Stu's Method(s) Selection…

- Existence
  - Does model merging ever happen in practice?
- Description/Classification
  - What are the different types of model merging that occur in practice on large scale systems?
- Descriptive-Comparative
  - How does model merging with explicit representation of relationships differ from model merging without such representation?
- Causality
  - Does an explicit representation of the relationship between models cause developers to explore different ways of merging models?
- Causality-Comparative
  - Does the algebraic representation of relationships in Stu's tool lead developers to explore more than do pointcuts in AOM?



Warning

No method is perfect

Don’t get hung up on methodological purity

Pick something and get on with it

Some knowledge is better than none

All Methods are flawed

- E.g. Laboratory Experiments
  - Cannot study large scale software development in the lab!
  - Too many variables to control them all!
- E.g. Case Studies
  - How do we know what's true in one project generalizes to others?
  - Researcher chose what questions to ask, hence biased the study
- E.g. Surveys
  - Self-selection of respondents biases the study
  - Respondents tell you what they think they ought to do, not what they actually do
- …etc...

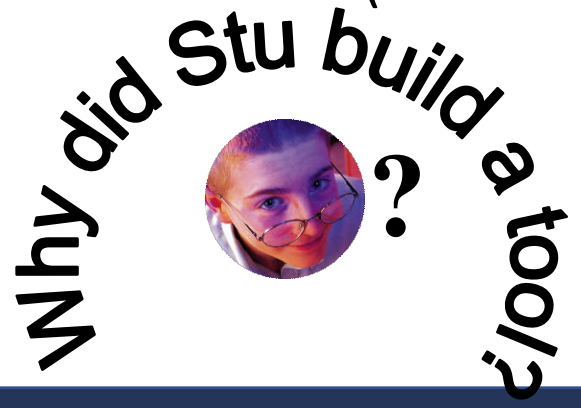
Strategies to overcome weaknesses

- Theory-building
  - Testing a hypothesis is pointless (single flawed study!)…
  - …unless it builds evidence for a clearly stated theory
- Empirical Induction
  - Series of studies over time…
  - Each designed to probe more aspects of the theory
  - …together build evidence for a clearly stated theory
- Mixed Methods Research
  - Use multiple methods to investigate the same research question
  - Each method compensates for the flaws of the others
  - …together build evidence for a clearly stated theory

Okay, but…

Why Build a Tool?

- ➜Build a Tool to Test a Theory
  - ↳Tool is part of the experimental materials needed to conduct your study
- ➜Build a Tool to Develop a Theory
  - ↳Theory emerges as you explore the tool
- ➜Build a Tool to Explain your Theory
  - ↳Theory as a concrete instantiation of (some aspect of) the theory



Agenda

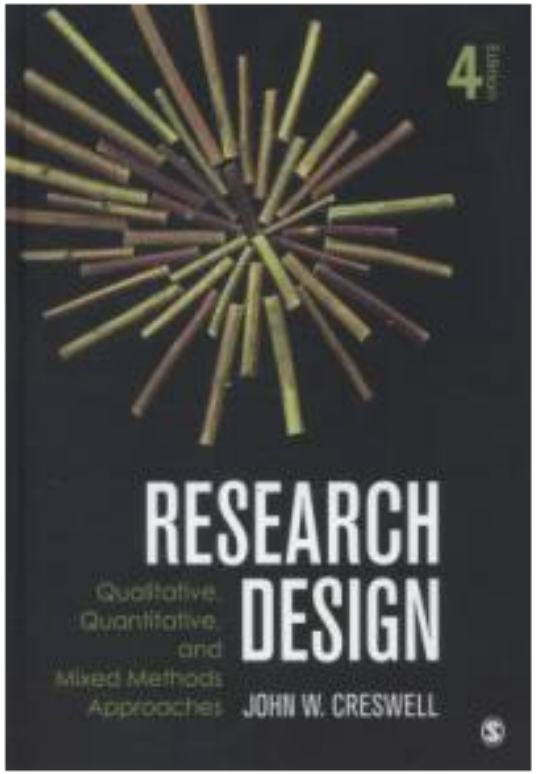
- Introduction
  - Who are you?
  - What's your research?
  - What would make this course valuable to you?
- Why empirical methods?
- Research designs
- Course overview

Help software developers
to better collaborate

+ Advances in tooling & SE principles
+ Insights from other disciplines
+ Mix a wide range of research methods

Problem → Intervention → Evaluation

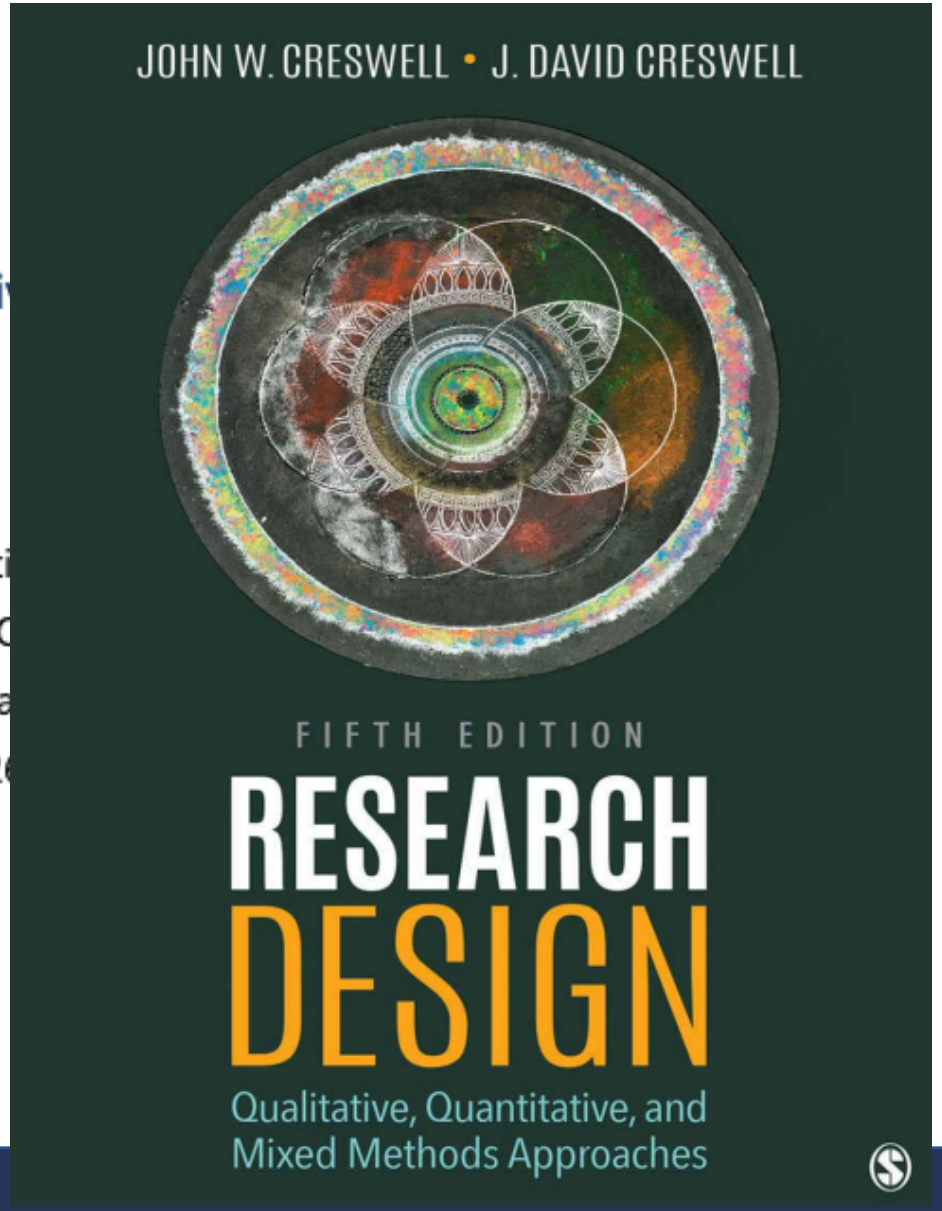
Textbook



Research design : qualitati
approaches /

John W. Creswell.

edition	Fourth editi
imprint	Thousand C
description	xxix, 273 pa
ISBN	978145222
format(s)	Book



Professionalism

- Being a professional means you should work well with others
- The best professionals are those who make those around them better
- If you feel someone is not treating you or someone else in a professional manner, you have two options:
  - If you feel you have the standing to do so, speak up!
  - Reach out to the course staff, and we will meet with you privately to discuss it, as well as preserve your anonymity

Academic Honesty

- In a nutshell: do not copy, do not lie, do not share or publicly release your solutions
- If you feel overwhelmed or stressed, please come and talk to us

Before next lecture

- Assignment 1
- Sign up for presentations