The Greatest Challenge

Joachim Parrow
Bertinoro 2014

The slides for this talk is a subset of the slides for my invited talk at Discotec 2014. I here include all of them.

The Right Stuff -
failure is not an option

This is a public copy of the slides for my invited plenary talk at DisCoTec, Berlin, June 6th 2014.

The Right Stuff

Apollo 13 launch, April 11 1970

A book by Tom Wolfe (1979) and a movie by Philip Kaufmann (1983) about the fine qualities of the early astronauts.

Coolness in the face of danger

"Failure is not an option"

Gene Kranz, flight director Apollo 13

The Right Stuff

”Failure is not an option”

Gene Kranz, flight director Apollo 13

That stuff is not quite right!

Only, in reality he never said that!

It was attributed to him in order to market the movie *Apollo 13* (1995)

The Right Stuff

This talk will not be about spacecrafts

nor about fine qualities of astronauts

= **stuff** that is **right**!

It will be about **correctness of artifacts**

The Right Stuff - failure is not an option

Joachim Parrow, Uppsala University

we = *theoretical computer scientists*

= *our theorems*

What are the dangers that our stuff is not right?
How can we make sure that it is right?

The Right Stuff - failure is not an option

Joachim Parrow, Uppsala University

- The Stuff in science
- The Stuff in theoretical computer science
- The psi experience: how I get **my** Stuff right

The Stuff in Science

Are there reasons to worry?

The Economist

OCTOBER 19TH–25TH 2013

Economist.com

Britain's angry white men

How to do a nuclear deal with Iran

Investment tips from Nobel economists

Junk bonds are back

The meaning of Sachin Tendulkar

HOW SCIENCE GOES WRONG

YES!

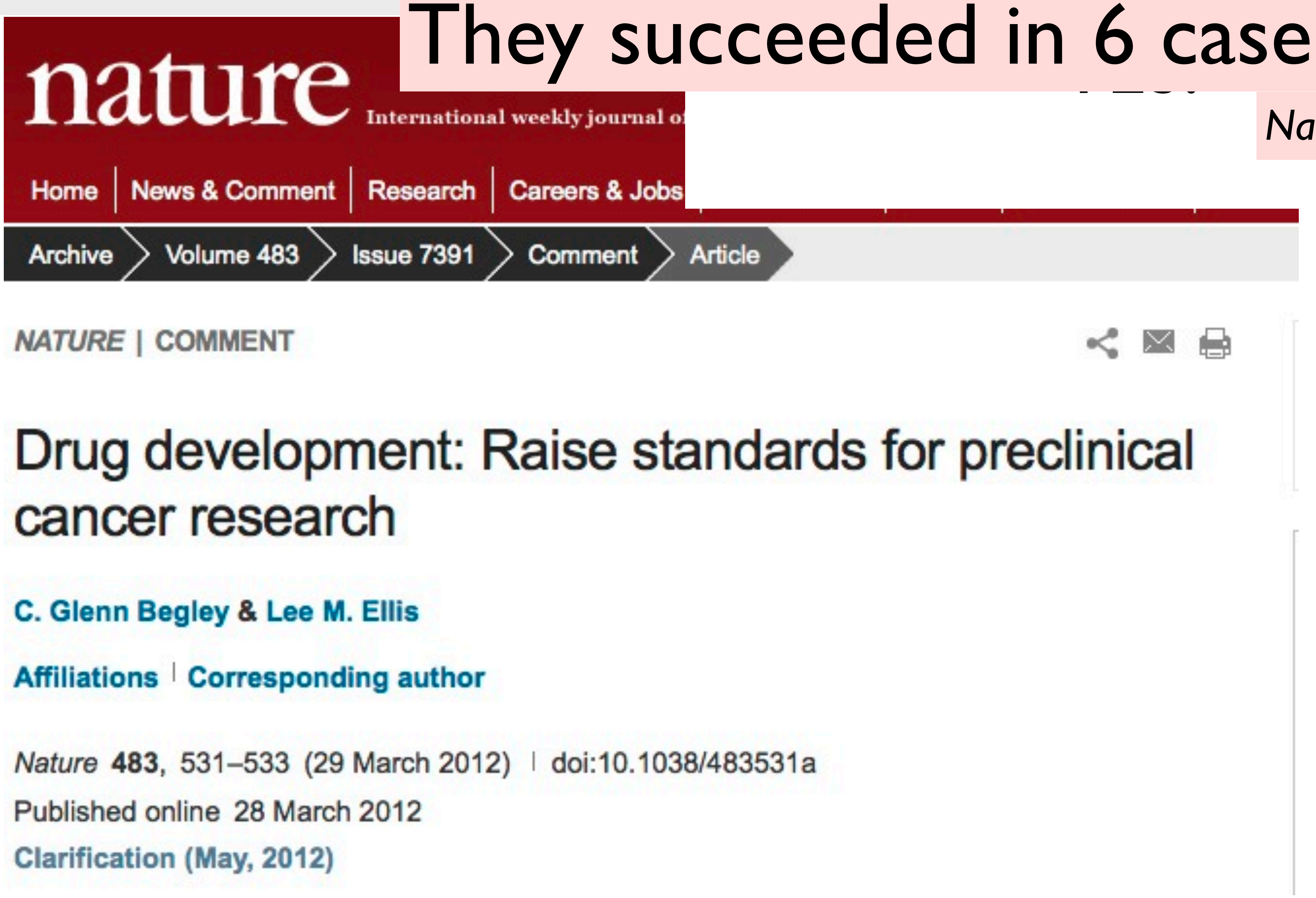
Biotechnology VC rule of thumb: half of published research cannot be replicated.

Amgen tried to replicate 53 landmark results in cancer research

Are there reasons to worry?

They succeeded in 6 cases (=11%)

Nature, March 2012



NATURE | COMMENT

Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Affiliations | Corresponding author

Nature **483**, 531–533 (29 March 2012) | doi:10.1038/483531a

Published online 28 March 2012

Clarification (May, 2012)

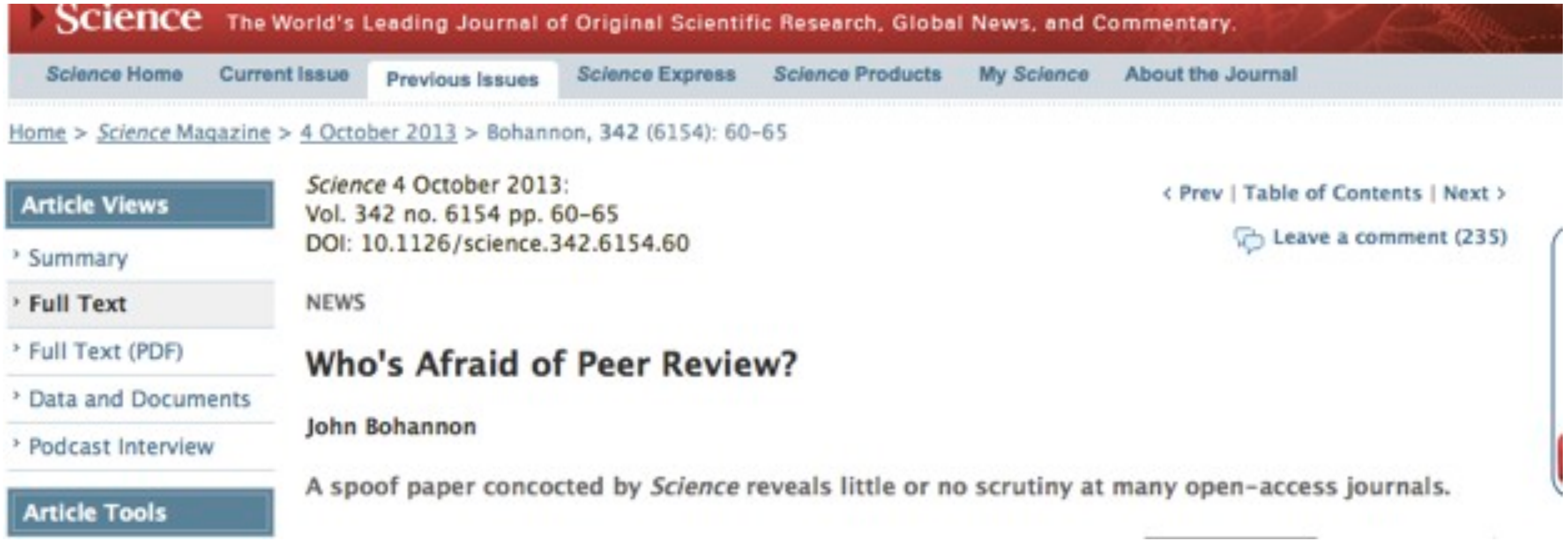
Why ?

Publish or Perish

- Need to publish a lot
- Need to publish quickly
- High rewards for publications
- No penalty for getting things wrong

Shoddy peer reviews

- 157 out of 304 journals accepted a bogus paper (*Bohannon, Science 2013*)



Science 4 October 2013:
Vol. 342 no. 6154 pp. 60-65
DOI: 10.1126/science.342.6154.60

NEWS

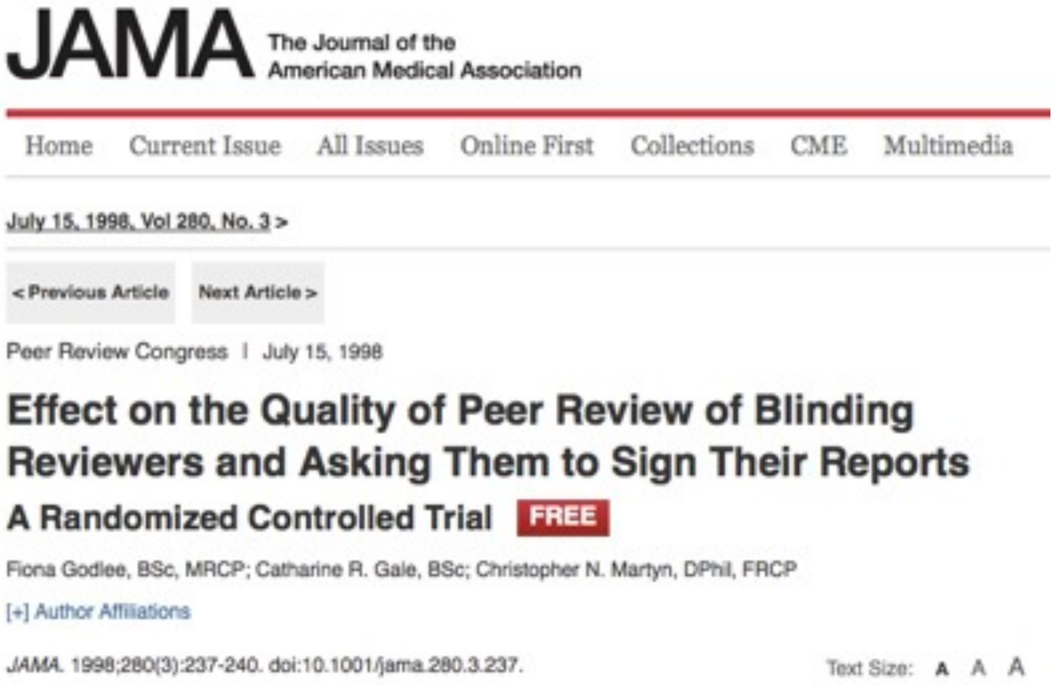
Who's Afraid of Peer Review?

John Bohannon

A spoof paper concocted by *Science* reveals little or no scrutiny at many open-access journals.

Shoddy peer reviews

- 157 out of 304 journals accepted a bogus paper (*Bohannon, Science 2013*)
- *British Medical Journal* referees spotted less than 25% of planted mistakes (*Godlee et all, J. American Medical Association 1998*)



July 15, 1998, Vol 280, No. 3 >

Peer Review Congress | July 15, 1998

Effect on the Quality of Peer Review of Blinding Reviewers and Asking Them to Sign Their Reports

A Randomized Controlled Trial

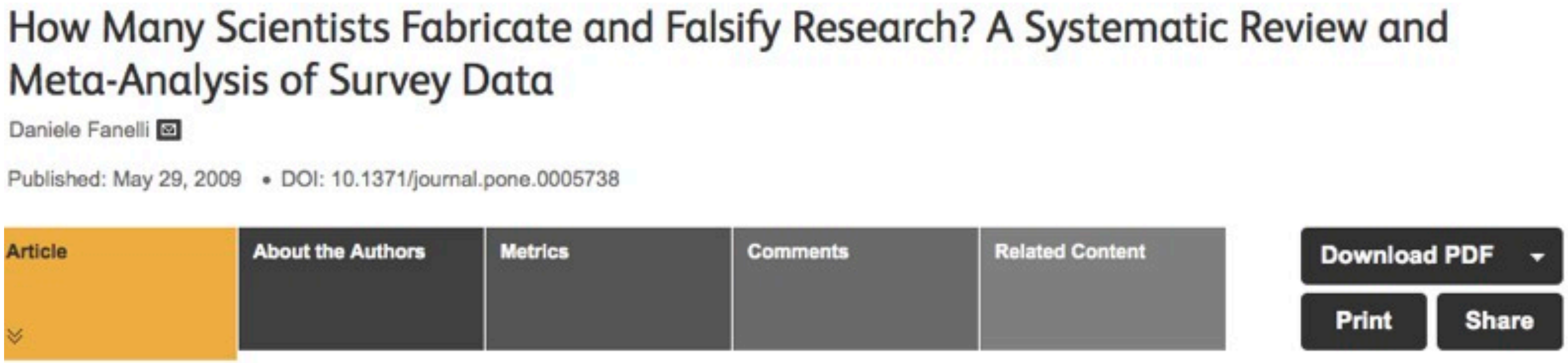
Fiona Godlee, BSc, MRCP; Catharine R. Gale, BSc; Christopher N. Martyn, DPhil, FRCP

JAMA. 1998;280(3):237-240. doi:10.1001/jama.280.3.237.

Fraud

Fanelli, *Plos One* 2009
Summarizes 18 studies 1988-2005

- 2% admit to falsifying data



How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data

Daniele Fanelli

Published: May 29, 2009 • DOI: 10.1371/journal.pone.0005738

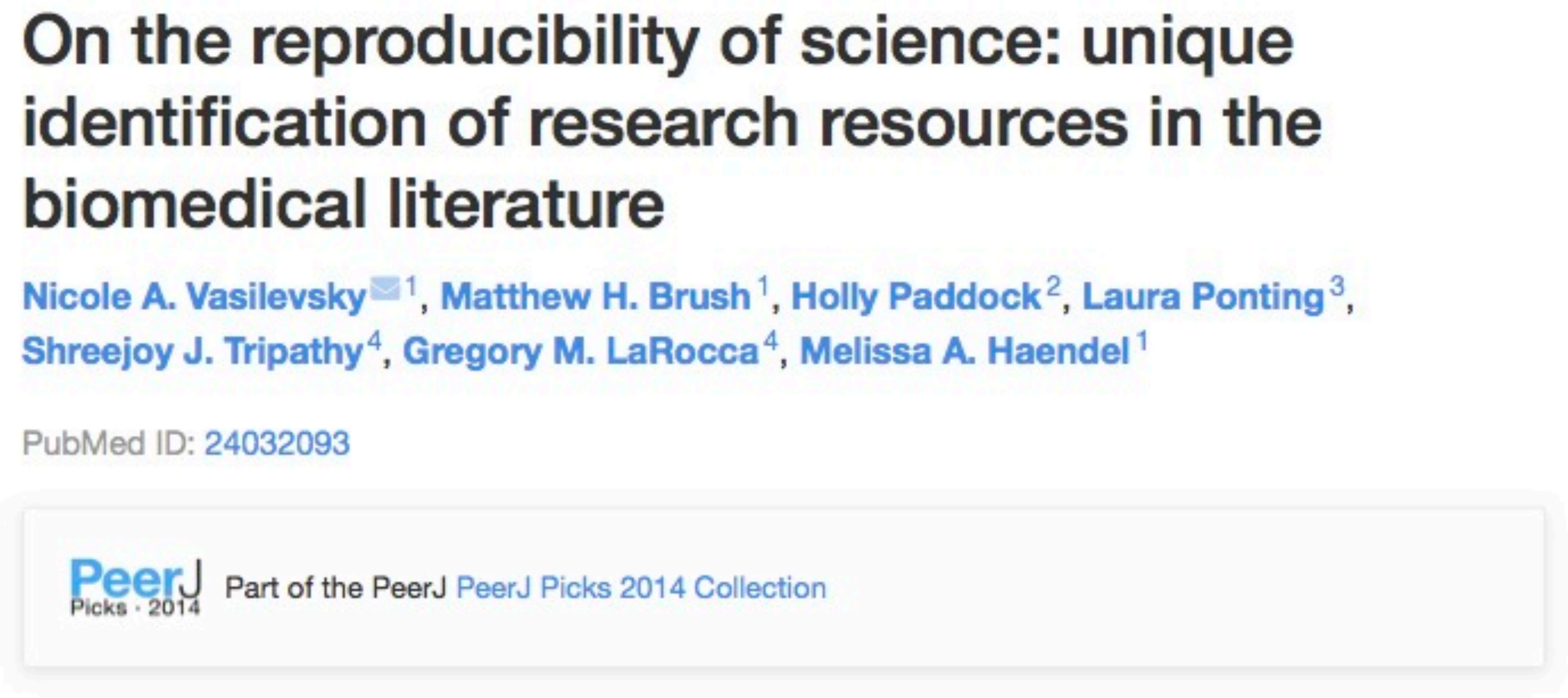
Fraud

Fanelli, *Plos One* 2009
Summarizes 18 studies 1988-2005

- 2% admit to falsifying data
- 14% claim to know colleagues who do
- 33% admit to questionable research practice
- 72% claim to know colleagues who do

Irreproducibility

- In 238 papers from 84 journals 2012-2013, 54% of resources were not identified (Vasilevsky et al, *PeerJ* 2013)



On the reproducibility of science: unique identification of research resources in the biomedical literature

Nicole A. Vasilevsky[1], Matthew H. Brush[1], Holly Paddock[2], Laura Ponting[3], Shreejoy J. Tripathy[4], Gregory M. LaRocca[4], Melissa A. Haendel[1]

PubMed ID: 24032093

PeerJ Picks · 2014 Part of the PeerJ PeerJ Picks 2014 Collection

Irreproducibility

- In 238 papers from 84 journals 2012-2013, 54% of resources were not identified (Vasilevsky et al, *PeerJ* 2013)
- Does not vary with impact factor!
- Reproducing results is a lot of work for very little gain.

Chance

- Experiment with sampled data: a risk that the samples are **a fluke**
- **False negative**: fail to establish a result
- **False positive**: establish an incorrect result

Hypotheses

- Never experiment at random! Always try to support or reject a **hypothesis**, that some interesting property holds
- Compared to the **null hypothesis** = no interesting property holds

p-value

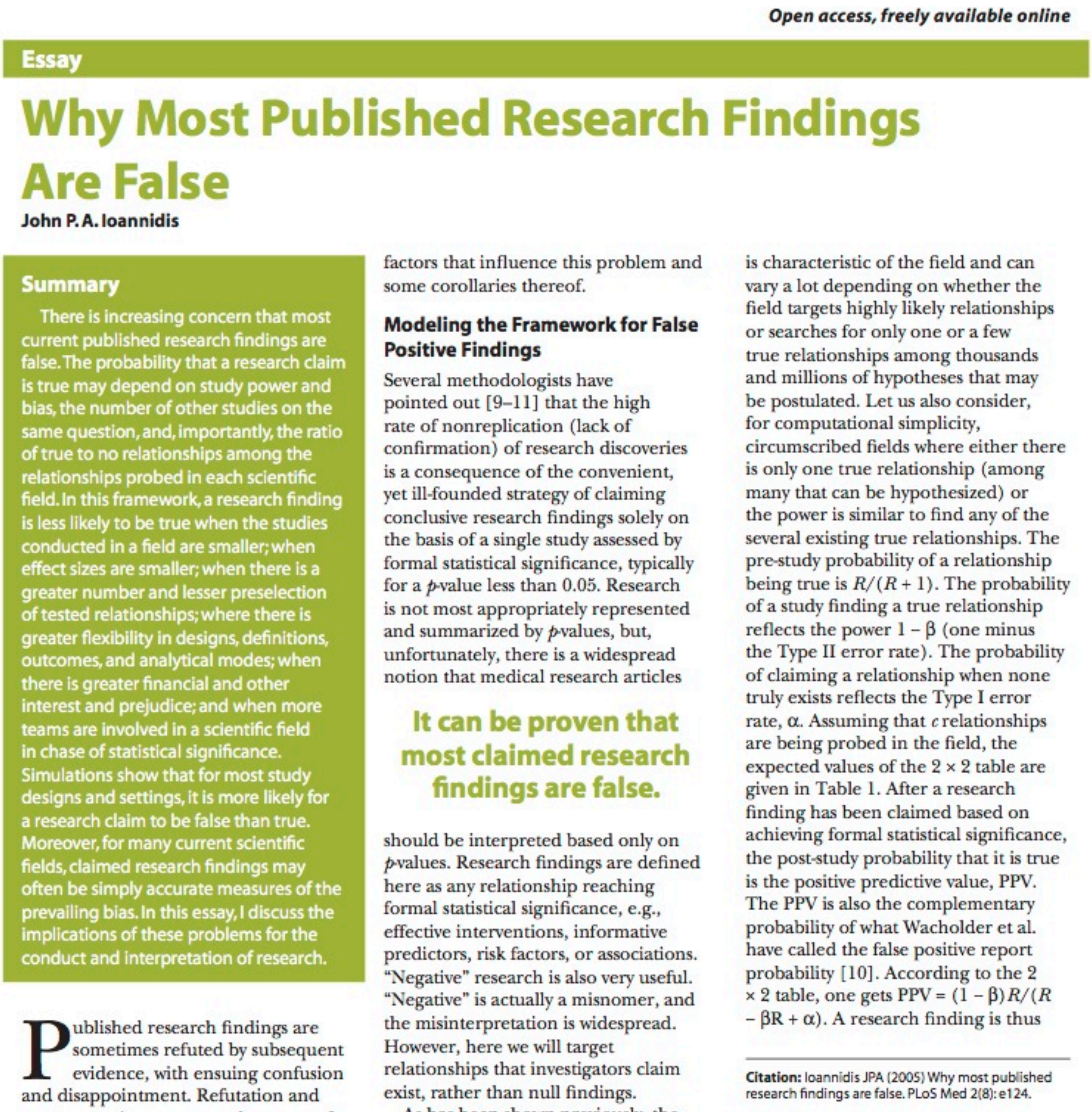
- **Outcome** of an experiment: can be because of **a fluke**, assuming the **null hypothesis**
- The probability of this = the **p-value**
- Small p-value => reject null hypothesis

p-value

- **Example**: a coin is **fair or biased**. Null hypothesis = fair coin.
- Five tosses gets **five heads**
- Assuming null hypothesis: probability 1/32 ≈ **3%**
- I believe the coin is not fair

p-value

- Area standard: p-value of **5%** is enough to reject the null hypothesis.
- **Q: So, because of this, what proportion of the published results will be false?**



Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

It can be proven that most claimed research findings are false.

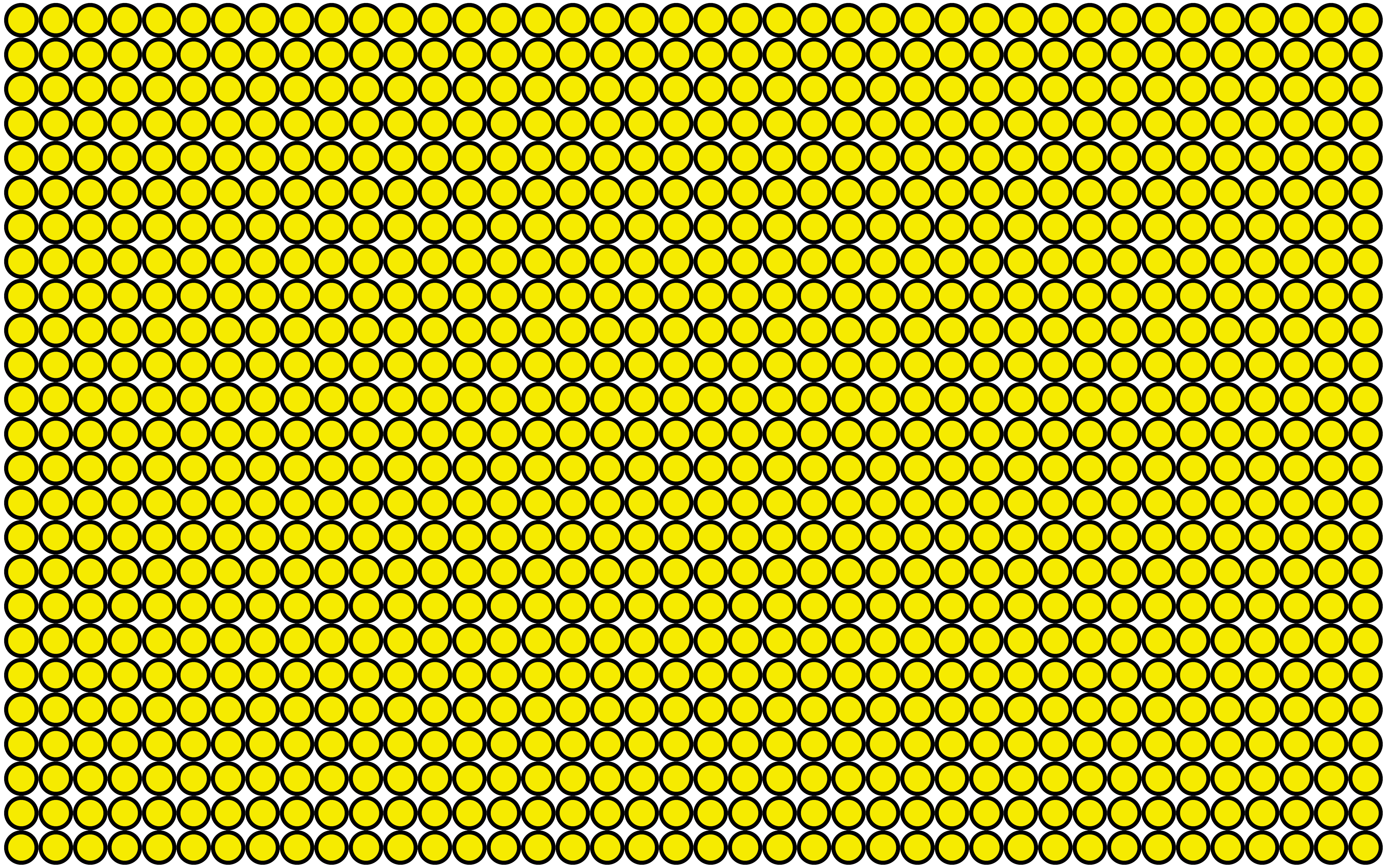
is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1-\beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2 × 2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2 × 2 table, one gets PPV = $(1-\beta)R/(R-\beta R+\alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

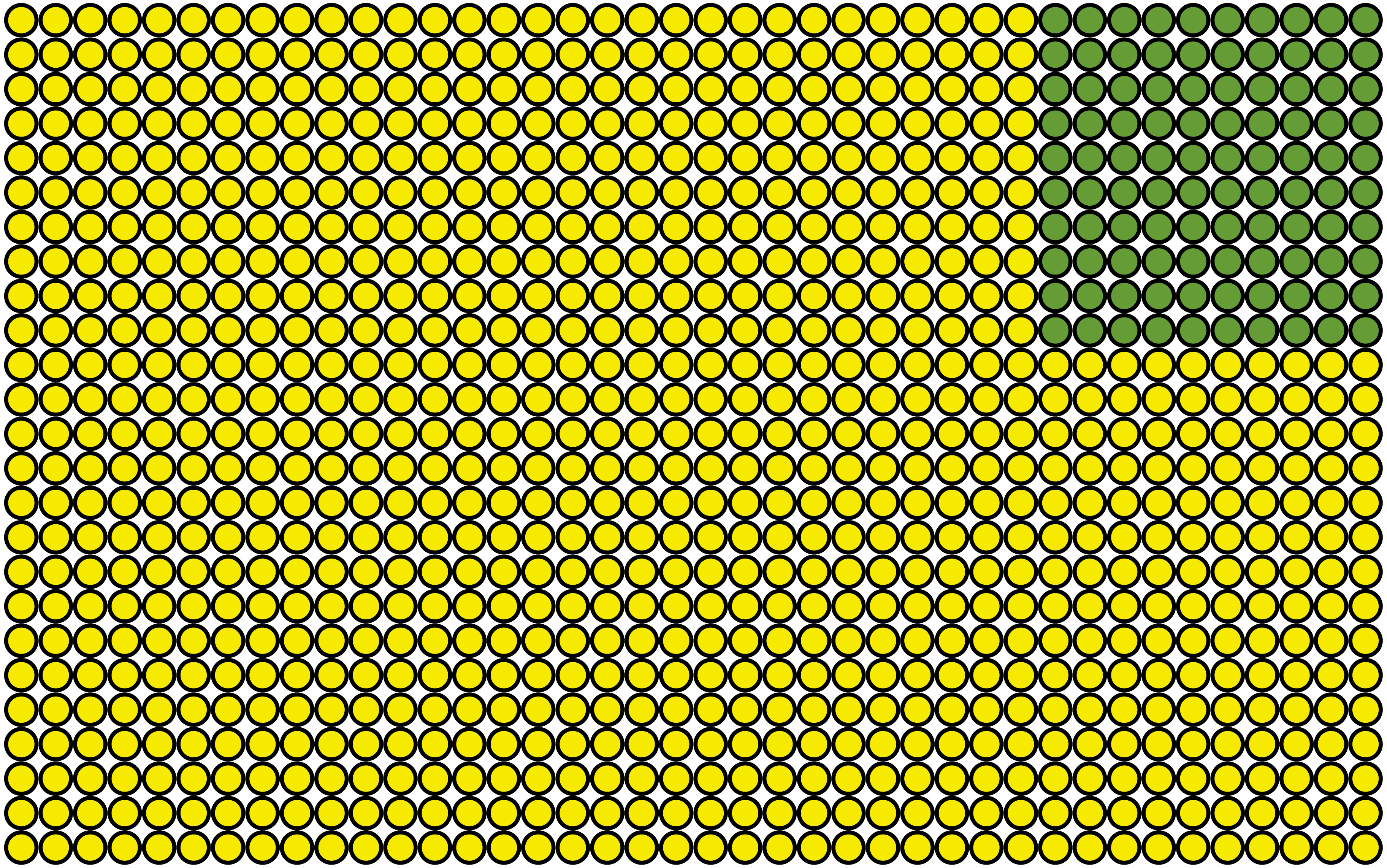
False hypotheses

- Out of all hypotheses tested, what proportion is **actually** true?
- Depends heavily on the field
- Reasonable overall assumption: **0.1** (one out of ten hypotheses is actually true)

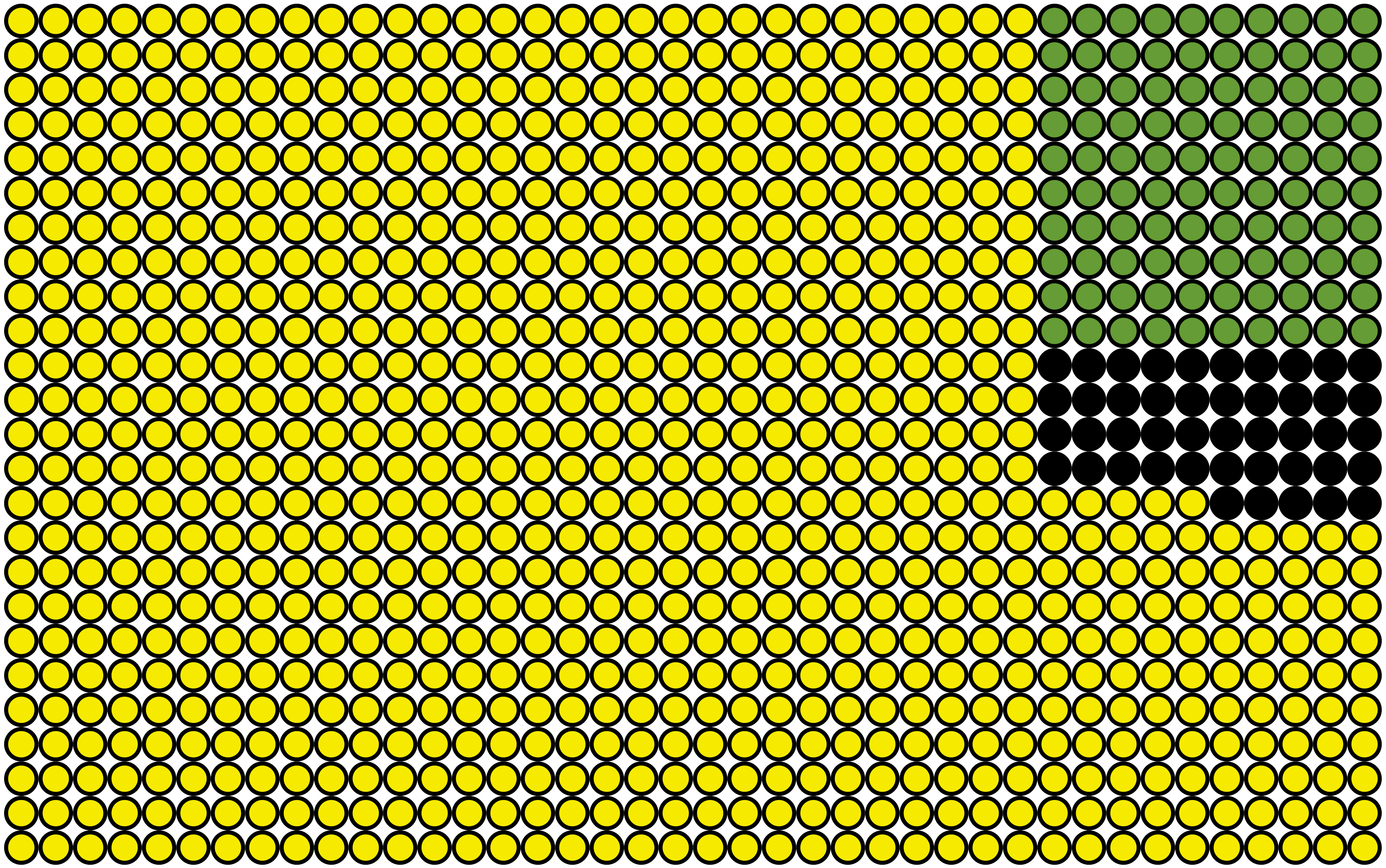
One thousand hypotheses tested



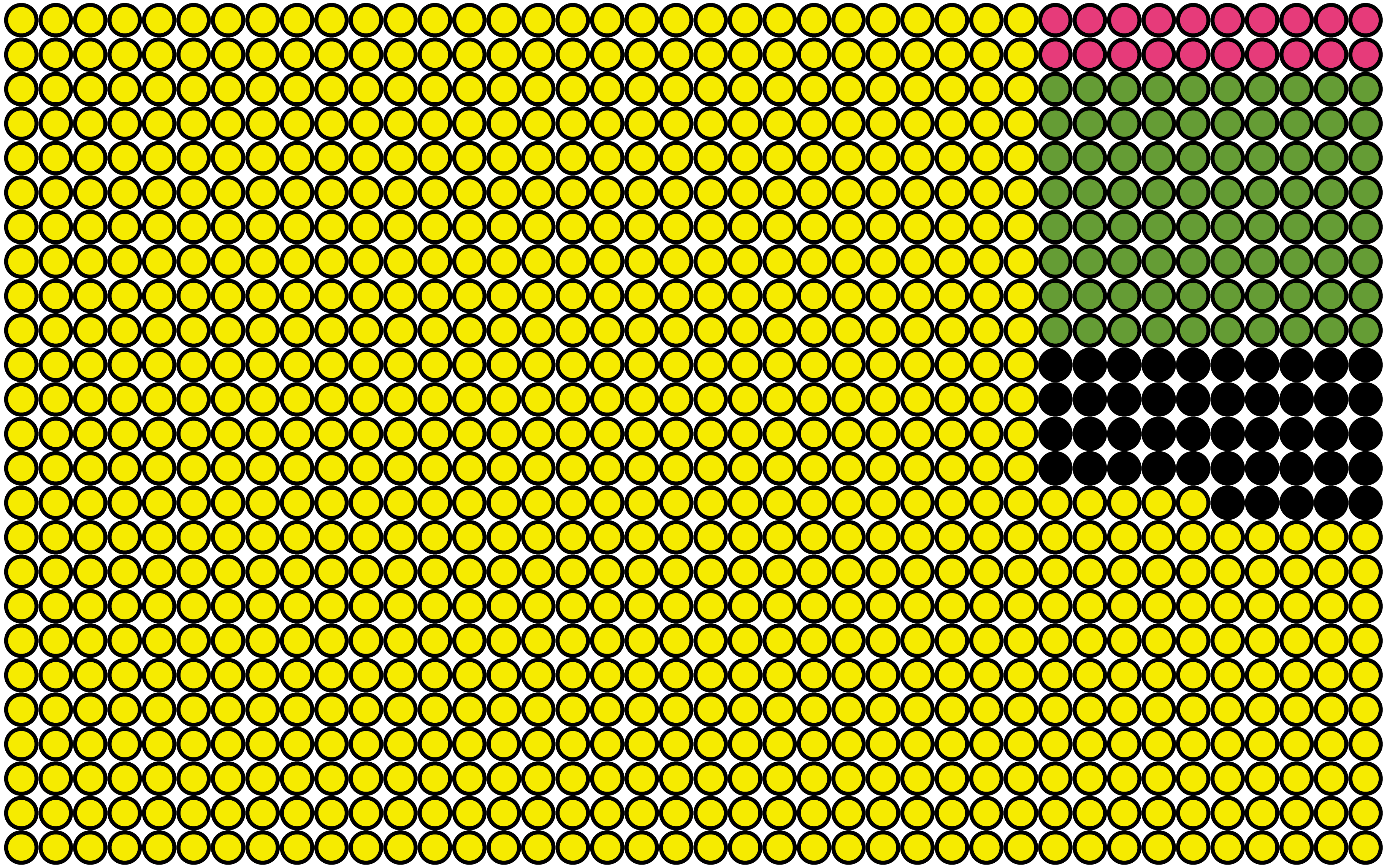
One hundred of them are actually true



900 x 0.05 = 45 are erroneously found to be true



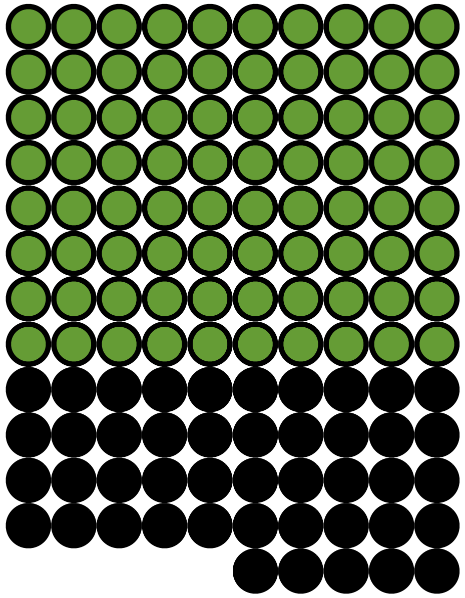
False negatives: typically at least 20%



What we publish as true:

80 things that are actually true

45 things that are actually false



36% of published ”truths” are false

Corollaries

Increased likelihood of study being wrong if

- The number of attempts is large
- The flexibility in designs, definitions etc is large
- The topic is hot
- etc

The Stuff in Theoretical Computer Science

Do we have any of

- Publish or Perish?
- Shoddy peer reviews?
- Fraud?
- Irreproducibility?
- Chance?

What about the p-values?

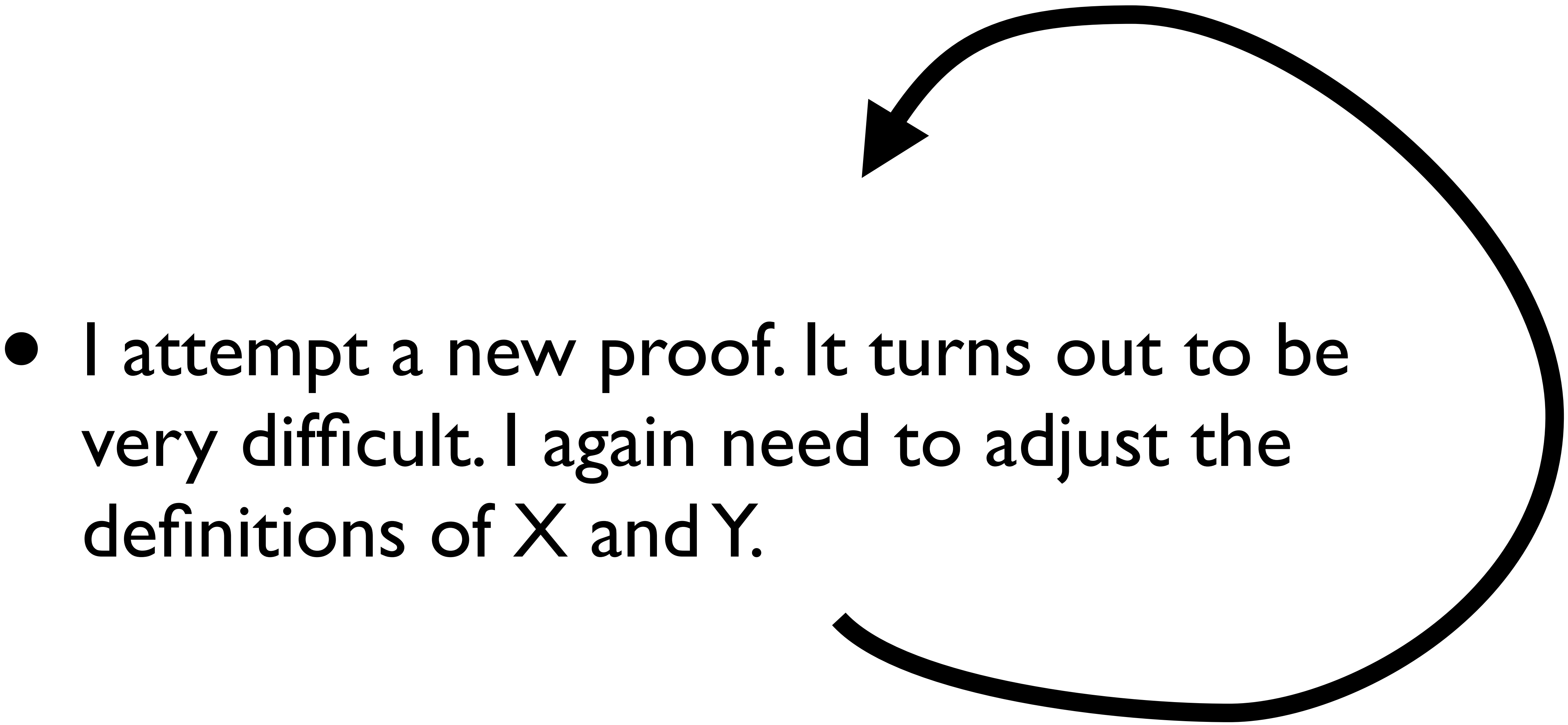
- No p-values! A theorem is either proven or not!
- But, we do occasionally have errors in proofs.
- With what frequency will we produce a proof with an error in it?

What about the hypotheses?

- No hypotheses!
- But, we do have **conjectures** that we try to prove.
- How often do we try to establish conjectures that are not true?

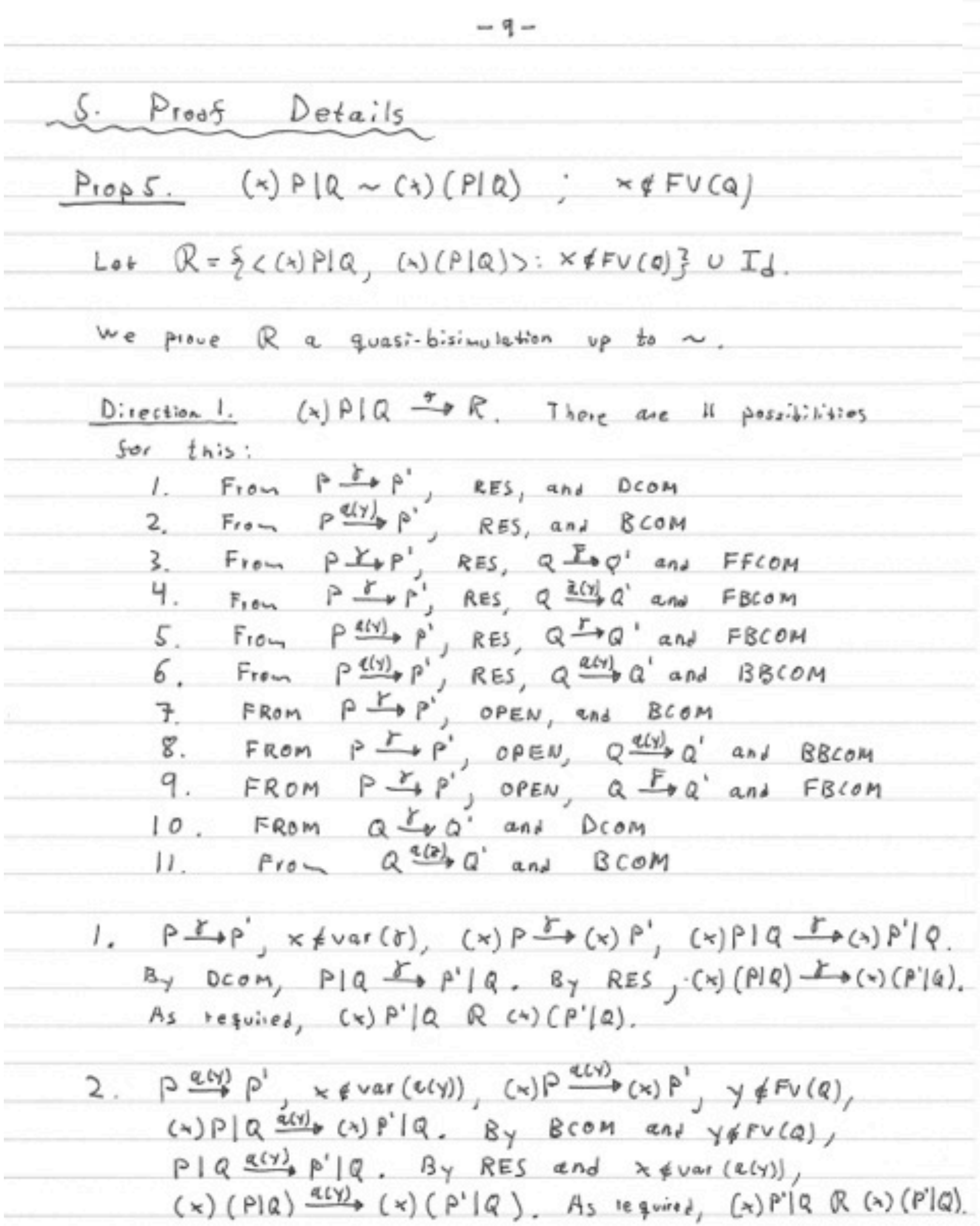
My typical day at work

- My hunch: objects of kind X satisfy property Y.
- X and Y are complicated (= several pages of definitions) and apt to change.
- I attempt a proof. It turns out to be very difficult. I need to adjust the definitions of X and Y.



- I attempt a new proof. It turns out to be very difficult. I again need to adjust the definitions of X and Y.

From the pi-calculus proof archive (1987): first ever proof of scope extension law!



-9-

5. Proof Details

Prop 5. $(x)P|Q \sim (x)(P|Q)$; $x \notin FV(Q)$

Let $\mathcal{R} = \{\langle (x)P|Q, (x)(P|Q)\rangle : x \notin FV(Q)\} \cup Id$.

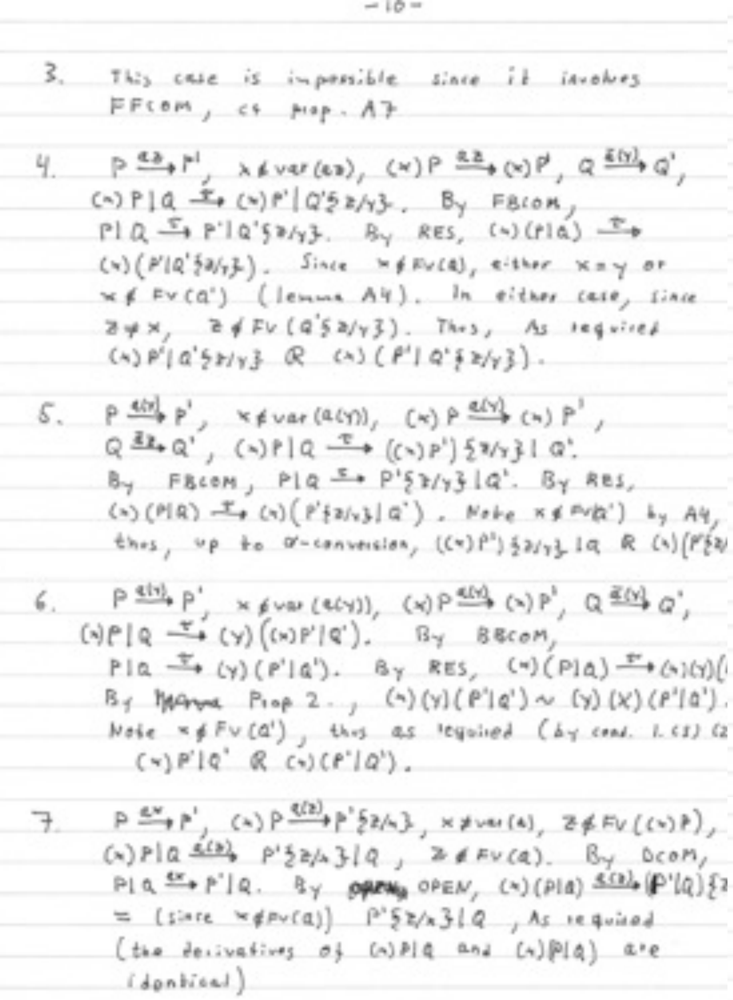
We prove $\mathcal{R}$ a quasi-bisimulation up to $\sim$.

Direction 1. $(x)P|Q \xrightarrow{\alpha} R$. There are 11 possibilities for this:

1. From $P \xrightarrow{\gamma} P'$, RES, and DCOM
2. From $P \xrightarrow{a(y)} P'$, RES, and BCOM
3. From $P \xrightarrow{\gamma} P'$, RES, $Q \xrightarrow{\bar{\gamma}} Q'$ and FFCOM
4. From $P \xrightarrow{\gamma} P'$, RES, $Q \xrightarrow{\bar{a}(y)} Q'$ and FBCOM
5. From $P \xrightarrow{a(y)} P'$, RES, $Q \xrightarrow{\gamma} Q'$ and FBCOM
6. From $P \xrightarrow{a(y)} P'$, RES, $Q \xrightarrow{\bar{a}(y)} Q'$ and BBCOM
7. FROM $P \xrightarrow{\gamma} P'$, OPEN, and BCOM
8. FROM $P \xrightarrow{\gamma} P'$, OPEN, $Q \xrightarrow{a(y)} Q'$ and BBCOM
9. FROM $P \xrightarrow{\gamma} P'$, OPEN, $Q \xrightarrow{\bar{\gamma}} Q'$ and FBCOM
10. FROM $Q \xrightarrow{\gamma} Q'$ and DCOM
11. From $Q \xrightarrow{a(z)} Q'$ and BCOM

1. $P \xrightarrow{\gamma} P'$, $x \notin var(\gamma)$, $(x)P \xrightarrow{\gamma} (x)P'$, $(x)P|Q \xrightarrow{\gamma} (x)P'|Q$. By DCOM, $P|Q \xrightarrow{\gamma} P'|Q$. By RES, $(x)(P|Q) \xrightarrow{\gamma} (x)(P'|Q)$. As required, $(x)P'|Q \;\mathcal{R}\; (x)(P'|Q)$.

2. $P \xrightarrow{a(y)} P'$, $x \notin var(a(y))$, $(x)P \xrightarrow{a(y)} (x)P'$, $y \notin FV(Q)$, $(x)P|Q \xrightarrow{a(y)} (x)P'|Q$. By BCOM and $y \notin FV(Q)$, $P|Q \xrightarrow{a(y)} P'|Q$. By RES and $x \notin var(a(y))$, $(x)(P|Q) \xrightarrow{a(y)} (x)(P'|Q)$. As required, $(x)P'|Q \;\mathcal{R}\; (x)(P'|Q)$.



-10-

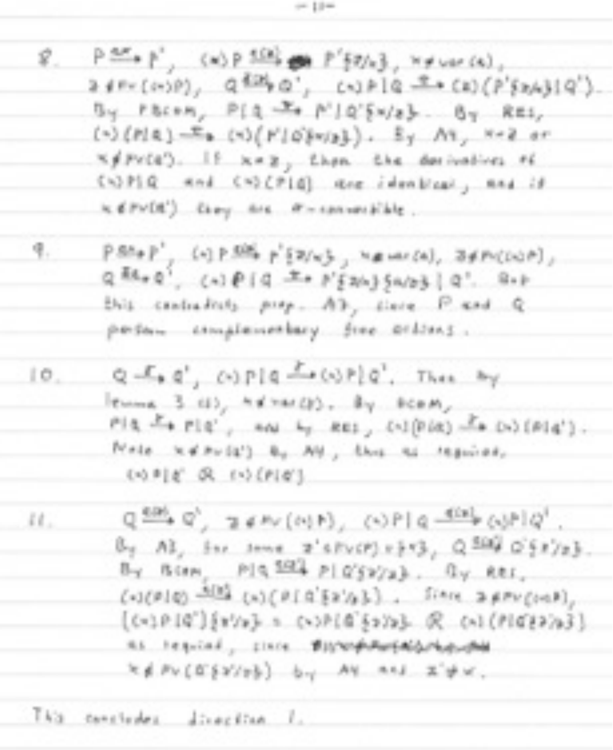
3. This case is impossible since it involves FFCOM, cf prop. A7

4. $P \xrightarrow{az} P'$, $x \notin var(az)$, $(x)P \xrightarrow{az} (x)P'$, $Q \xrightarrow{\bar{a}(y)} Q'$, $(x)P|Q \xrightarrow{\tau} (x)P'|Q'\{z/y\}$. By FBCOM, $P|Q \xrightarrow{\tau} P'|Q'\{z/y\}$. By RES, $(x)(P|Q) \xrightarrow{\tau} (x)(P'|Q'\{z/y\})$. Since $x \notin FV(Q)$, either $x = y$ or $x \notin FV(Q')$ (lemma A4). In either case, since $z \neq x$, $z \notin FV(Q'\{z/y\})$. Thus, As required $(x)P'|Q'\{z/y\} \;\mathcal{R}\; (x)(P'|Q'\{z/y\})$.

5. $P \xrightarrow{a(y)} P'$, $x \notin var(a(y))$, $(x)P \xrightarrow{a(y)} (x)P'$, $Q \xrightarrow{\bar{a}z} Q'$, $(x)P|Q \xrightarrow{\tau} ((x)P')\{z/y\}|Q'$. By FBCOM, $P|Q \xrightarrow{\tau} P'\{z/y\}|Q'$. By RES, $(x)(P|Q) \xrightarrow{\tau} (x)(P'\{z/y\}|Q')$. Note $x \notin FV(Q')$ by A4, thus, up to α-conversion, $((x)P')\{z/y\}|Q \;\mathcal{R}\; (x)(P'\{z/$

6. $P \xrightarrow{a(y)} P'$, $x \notin var(a(y))$, $(x)P \xrightarrow{a(y)} (x)P'$, $Q \xrightarrow{\bar{a}(y)} Q'$, $(x)P|Q \xrightarrow{\tau} (y)((x)P'|Q')$. By BBCOM, $P|Q \xrightarrow{\tau} (y)(P'|Q')$. By RES, $(x)(P|Q) \xrightarrow{\tau} (x)(y)($ By ~~[illegible]~~ Prop 2., $(x)(y)(P'|Q') \sim (y)(x)(P'|Q')$. Note $x \notin FV(Q')$, thus as required (by cond. 1.(3) (2 $(x)P'|Q' \;\mathcal{R}\; (x)(P'|Q')$.

7. $P \xrightarrow{ax} P'$, $(x)P \xrightarrow{a(z)} P'\{z/x\}$, $x \neq var(a)$, $z \notin FV((x)P)$, $(x)P|Q \xrightarrow{a(z)} P'\{z/x\}|Q$, $z \notin FV(Q)$. By DCOM, $P|Q \xrightarrow{ax} P'|Q$. By ~~OPEN~~ OPEN, $(x)(P|Q) \xrightarrow{a(z)} (P'|Q)\{z$ $=$ (since $x \notin FV(Q)$) $P'\{z/x\}|Q$, As required (the derivatives of $(x)P|Q$ and $(x)(P|Q)$ are identical)



-11-

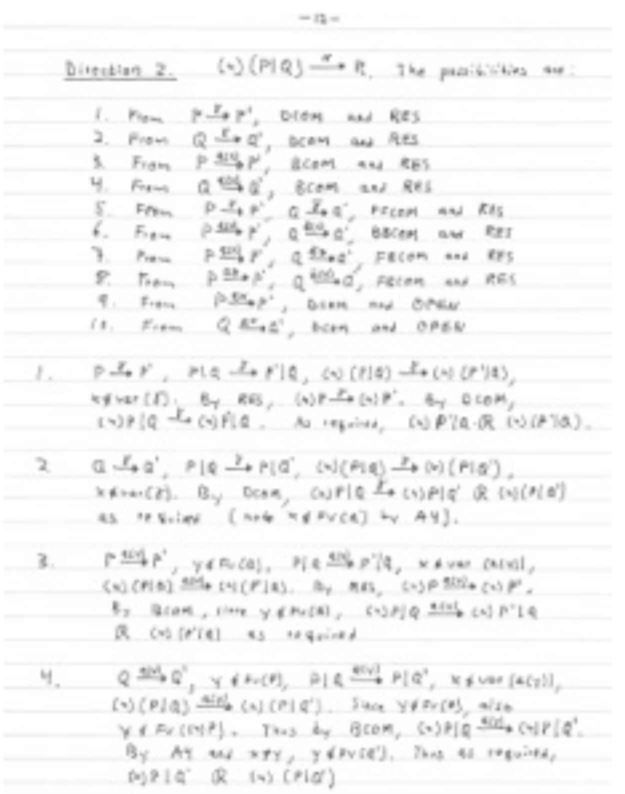
8. $P \xrightarrow{ax} P'$, $(x)P \xrightarrow{a(z)} P'\{z/x\}$, $x \neq var(a)$, $z \notin FV((x)P)$, $Q \xrightarrow{\bar{a}(z)} Q'$, $(x)P|Q \xrightarrow{\tau} (z)(P'\{z/x\}|Q')$. By FBCOM, $P|Q \xrightarrow{\tau} P'|Q'\{x/z\}$. By RES, $(x)(P|Q) \xrightarrow{\tau} (x)(P'|Q'\{x/z\})$. By A4, $x = z$ or $x \notin FV(Q')$. If $x = z$, then the derivatives of $(x)P|Q$ and $(x)(P|Q)$ are identical, and if $x \notin FV(Q')$ they are α-convertible.

9. $P \xrightarrow{ax} P'$, $(x)P \xrightarrow{a(z)} P'\{z/x\}$, $x \neq var(a)$, $z \notin FV((x)P)$, $Q \xrightarrow{\bar{a}x} Q'$, $(x)P|Q \xrightarrow{\tau} P'\{z/x\}\{x/z\}|Q'$. But this contradicts prop. A7, since P and Q perform complementary free actions.

10. $Q \xrightarrow{\gamma} Q'$, $(x)P|Q \xrightarrow{\gamma} (x)P|Q'$. Then by lemma 3 (i), $x \notin var(\gamma)$. By DCOM, $P|Q \xrightarrow{\gamma} P|Q'$, and by RES, $(x)(P|Q) \xrightarrow{\gamma} (x)(P|Q')$. Note $x \notin FV(Q')$ by A4, thus as required, $(x)P|Q' \;\mathcal{R}\; (x)(P|Q')$

11. $Q \xrightarrow{a(z)} Q'$, $z \notin FV((x)P)$, $(x)P|Q \xrightarrow{a(z)} (x)P|Q'$. By A3, for some $z' \notin FV(P) \cup \{x\}$, $Q \xrightarrow{a(z')} Q'\{z'/z\}$. By BCOM, $P|Q \xrightarrow{a(z')} P|Q'\{z'/z\}$. By RES, $(x)(P|Q) \xrightarrow{a(z')} (x)(P|Q'\{z'/z\})$. Since $z \notin FV((x)P)$, $[(x)P|Q']\{z'/z\} = (x)P|Q'\{z'/z\} \;\mathcal{R}\; (x)(P|Q'\{z'/z\})$ as required, since ~~[illegible]~~ $x \notin FV(Q'\{z'/z\})$ by A4 and $z' \neq x$.

This concludes direction 1.



-12-

Direction 2. $(x)(P|Q) \xrightarrow{\alpha} R$. The possibilities are:

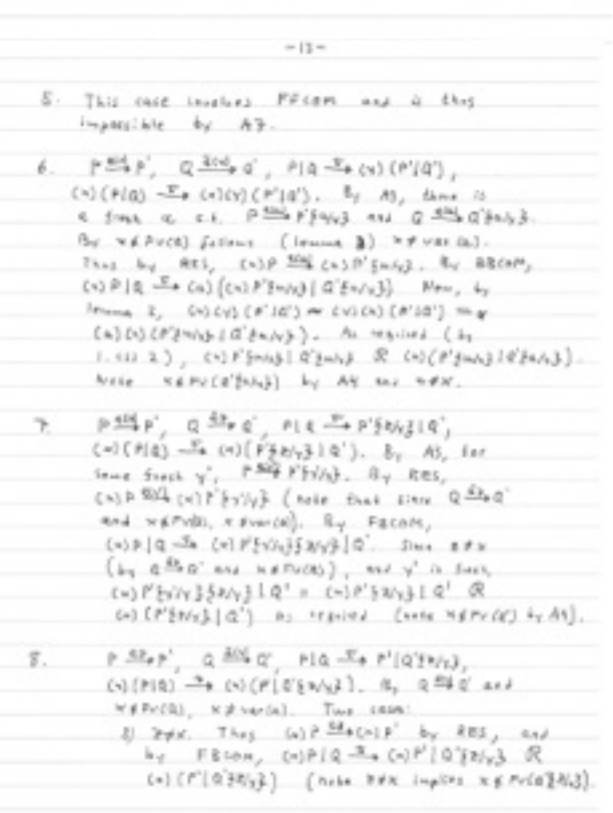
1. From $P \xrightarrow{\gamma} P'$, DCOM and RES
2. From $Q \xrightarrow{\gamma} Q'$, DCOM and RES
3. From $P \xrightarrow{a(y)} P'$, BCOM and RES
4. From $Q \xrightarrow{a(y)} Q'$, BCOM and RES
5. From $P \xrightarrow{\gamma} P'$, $Q \xrightarrow{\bar{\gamma}} Q'$, FFCOM and RES
6. From $P \xrightarrow{a(y)} P'$, $Q \xrightarrow{\bar{a}(y)} Q'$, BBCOM and RES
7. From $P \xrightarrow{a(y)} P'$, $Q \xrightarrow{\bar{a}z} Q'$, FBCOM and RES
8. From $P \xrightarrow{az} P'$, $Q \xrightarrow{\bar{a}(y)} Q'$, FBCOM and RES
9. From $P \xrightarrow{ax} P'$, DCOM and OPEN
10. From $Q \xrightarrow{ax} Q'$, DCOM and OPEN

1. $P \xrightarrow{\gamma} P'$, $P|Q \xrightarrow{\gamma} P'|Q$, $(x)(P|Q) \xrightarrow{\gamma} (x)(P'|Q)$, $x \notin var(\gamma)$. By RES, $(x)P \xrightarrow{\gamma} (x)P'$. By DCOM, $(x)P|Q \xrightarrow{\gamma} (x)P'|Q$. As required, $(x)P'|Q \;\mathcal{R}\; (x)(P'|Q)$.

2. $Q \xrightarrow{\gamma} Q'$, $P|Q \xrightarrow{\gamma} P|Q'$, $(x)(P|Q) \xrightarrow{\gamma} (x)(P|Q')$, $x \notin var(\gamma)$. By DCOM, $(x)P|Q \xrightarrow{\gamma} (x)P|Q' \;\mathcal{R}\; (x)(P|Q')$ as required (note $x \notin FV(Q')$ by A4).

3. $P \xrightarrow{a(y)} P'$, $y \notin FV(Q)$, $P|Q \xrightarrow{a(y)} P'|Q$, $x \notin var(a(y))$, $(x)(P|Q) \xrightarrow{a(y)} (x)(P'|Q)$. By RES, $(x)P \xrightarrow{a(y)} (x)P'$. By BCOM, since $y \notin FV(Q)$, $(x)P|Q \xrightarrow{a(y)} (x)P'|Q$ $\mathcal{R}$ $(x)(P'|Q)$ as required

4. $Q \xrightarrow{a(y)} Q'$, $y \notin FV(P)$, $P|Q \xrightarrow{a(y)} P|Q'$, $x \notin var(a(y))$, $(x)(P|Q) \xrightarrow{a(y)} (x)(P|Q')$. Since $y \notin FV(P)$, also $y \notin FV((x)P)$. Thus by BCOM, $(x)P|Q \xrightarrow{a(y)} (x)P|Q'$. By A4 and $x \neq y$, $y \notin FV(Q')$. Thus as required, $(x)P|Q'$ $\mathcal{R}$ $(x)(P|Q')$



-13-

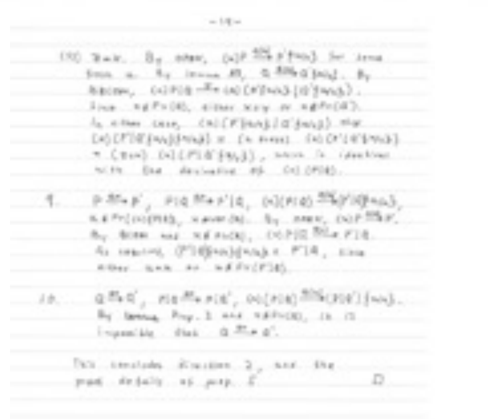
5. This case involves FFCOM and is thus impossible by A7.

6. $P \xrightarrow{a(y)} P'$, $Q \xrightarrow{\bar{a}(y)} Q'$, $P|Q \xrightarrow{\tau} (y)(P'|Q')$, $(x)(P|Q) \xrightarrow{\tau} (x)(y)(P'|Q')$. By A5, there is a fresh u s.t. $P \xrightarrow{a(u)} P'\{u/y\}$ and $Q \xrightarrow{\bar{a}(u)} Q'\{u/y\}$. By $x \notin FV(Q)$ follows (lemma 3) $x \neq var(a)$. Thus by RES, $(x)P \xrightarrow{a(u)} (x)P'\{u/y\}$. By BBCOM, $(x)P|Q \xrightarrow{\tau} (u)((x)P'\{u/y\}|Q'\{u/y\})$. Now, by lemma 2, $(x)(y)(P'|Q') \sim (y)(x)(P'|Q') \equiv_\alpha (u)(x)(P'\{u/y\}|Q'\{u/y\})$. As required (by 1.(ii) 2), $(x)P'\{u/y\}|Q'\{u/y\}$ $\mathcal{R}$ $(x)(P'\{u/y\}|Q'\{u/y\})$. Note $x \notin FV(Q'\{u/y\})$ by A4 and $u \neq x$.

7. $P \xrightarrow{a(y)} P'$, $Q \xrightarrow{\bar{a}z} Q'$, $P|Q \xrightarrow{\tau} P'\{z/y\}|Q'$, $(x)(P|Q) \xrightarrow{\tau} (x)(P'\{z/y\}|Q')$. By A5, for some fresh y', $P \xrightarrow{a(y')} P'\{y'/y\}$. By RES, $(x)P \xrightarrow{a(y')} (x)P'\{y'/y\}$ (note that since $Q \xrightarrow{\bar{a}z} Q'$ and $x \notin FV(Q)$, $x \neq var(a)$). By FBCOM, $(x)P|Q \xrightarrow{\tau} (x)P'\{y'/y\}\{z/y'\}|Q'$. Since $z \neq x$ (by $Q \xrightarrow{\bar{a}z} Q'$ and $x \notin FV(Q)$), and y' is fresh, $(x)P'\{y'/y\}\{z/y'\}|Q' = (x)P'\{z/y\}|Q'$ $\mathcal{R}$ $(x)(P'\{z/y\}|Q')$ as required (note $x \notin FV(Q')$ by A4).

8. $P \xrightarrow{az} P'$, $Q \xrightarrow{\bar{a}(y)} Q'$, $P|Q \xrightarrow{\tau} P'|Q'\{z/y\}$, $(x)(P|Q) \xrightarrow{\tau} (x)(P'|Q'\{z/y\})$. By $Q \xrightarrow{\bar{a}(y)} Q'$ and $x \notin FV(Q)$, $x \neq var(a)$. Two cases:

(i) $z \neq x$. Thus $(x)P \xrightarrow{az} (x)P'$ by RES, and by FBCOM, $(x)P|Q \xrightarrow{\tau} (x)P'|Q'\{z/y\}$ $\mathcal{R}$ $(x)(P'|Q'\{z/y\})$ (note $z \neq x$ implies $x \notin FV(Q'\{z/y\})$).



-14-

(ii) $z = x$. By OPEN, $(x)P \xrightarrow{a(u)} P'\{u/x\}$ for some fresh u. By lemma A5, $Q \xrightarrow{\bar{a}(u)} Q'\{u/y\}$. By BBCOM, $(x)P|Q \xrightarrow{\tau} (u)(P'\{u/x\}|Q'\{u/y\})$. Since $x \notin FV(Q)$, either $x = y$ or $x \notin FV(Q')$. In either case, $(u)(P'\{u/x\}|Q'\{u/y\})$ [illegible] $=$ (α conv.) [illegible] $=$ ($z=x$) $(x)(P'|Q'\{z/y\})$, which is identical with the derivative of $(x)(P|Q)$.

9. $P \xrightarrow{ax} P'$, $P|Q \xrightarrow{ax} P'|Q$, $(x)(P|Q) \xrightarrow{a(u)} (P'|Q)\{u/x\}$, [illegible]. By OPEN, $(x)P \xrightarrow{a(u)} P'\{u/x\}$. By BCOM and $x \notin FV(Q)$, [illegible]. As required, [illegible], since either [illegible] or $x \notin FV(P'|Q)$.

10. $Q \xrightarrow{ax} Q'$, $P|Q \xrightarrow{ax} P|Q'$, $(x)(P|Q) \xrightarrow{a(u)} (P|Q')\{u/x\}$. By lemma Prop. 1 and $x \notin FV(Q)$, it is impossible that $Q \xrightarrow{ax} Q'$.

This concludes direction 2, and the proof details of prop. 5. □

Time passes, and eventually...

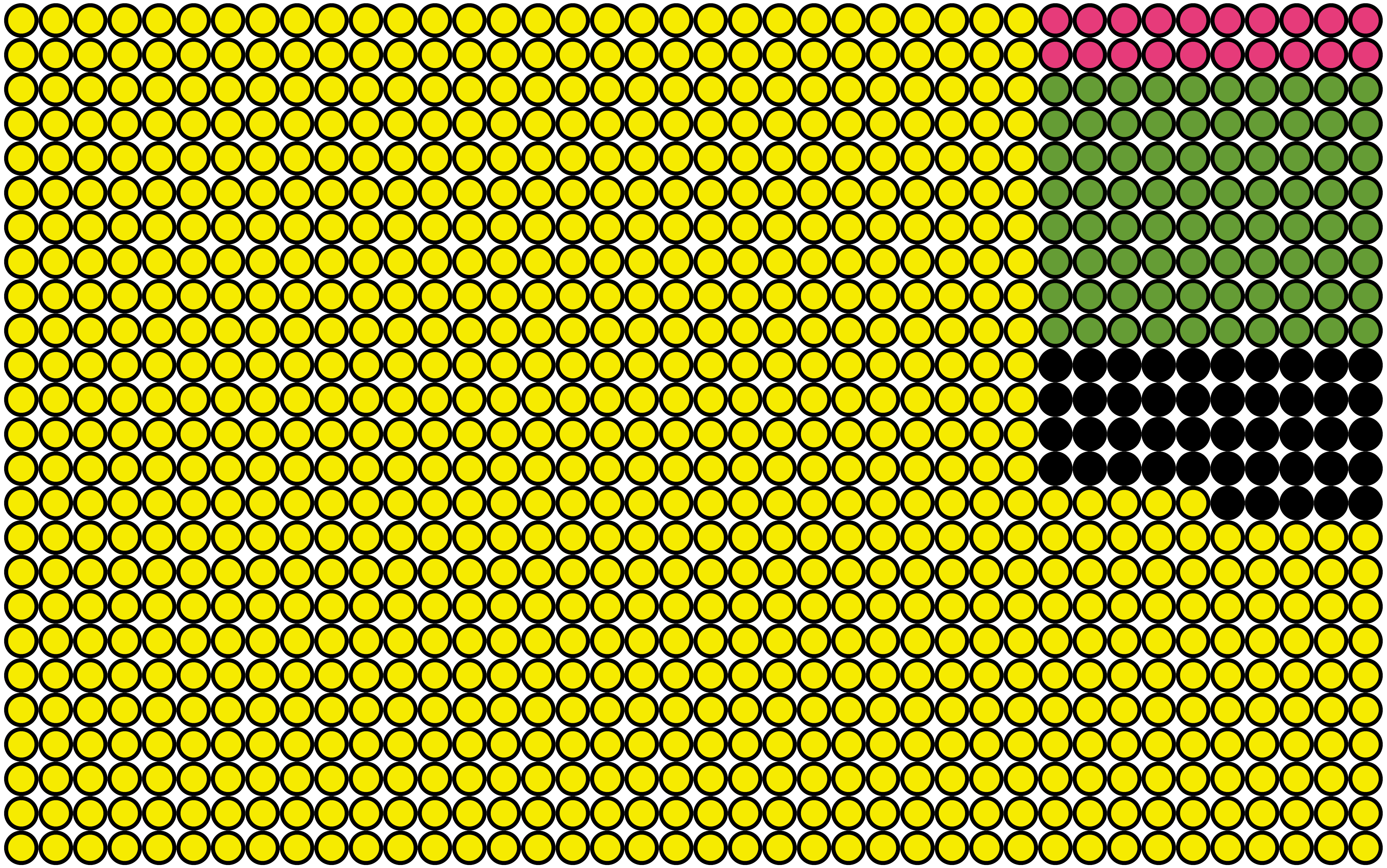
- I attempt a new proof. It succeeds! Now I can publish!

standard research practice:
Discovering exactly what to prove
in parallel with proving it

Time passes, and eventually...

I spend much more time trying to prove things that are false than proving things that are true.

Caveat: As opposed to the situation in life sciences, we cannot yet quantify the figures.



Things I try to prove

Things I fail to prove

Things I manage to prove

Things I prove but wrongly

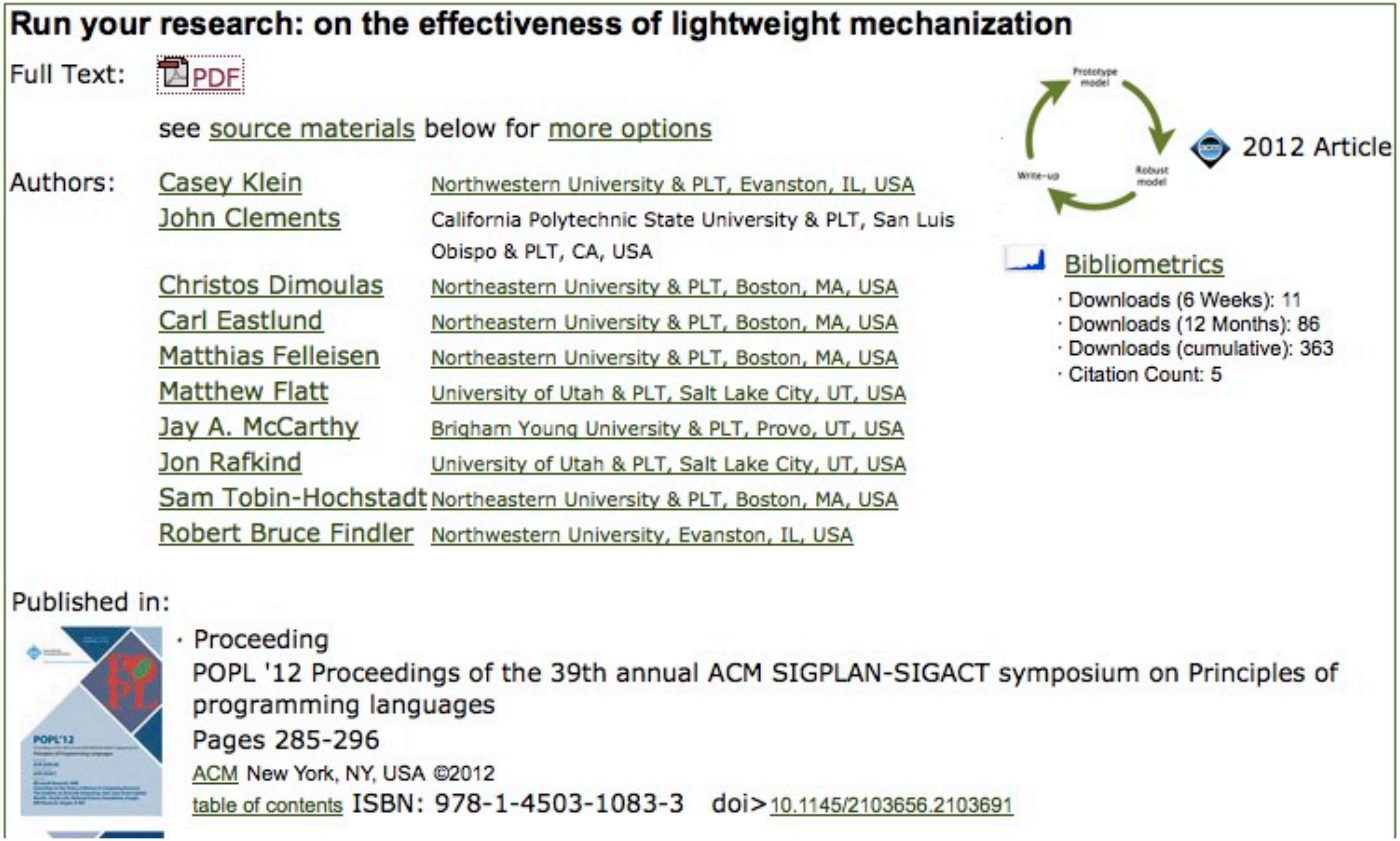
How bad is it?

Anecdotal: My personal experience

- Several results published in my immediate area in major conferences the last years
- Serious error in the statement or proof of a theorem
- Many are well cited and used
- One of them is my own

Run your research

Klein et al, POPL 2012



Run your research: on the effectiveness of lightweight mechanization

Full Text: PDF

see source materials below for more options

2012 Article

Authors:

Casey Klein	Northwestern University & PLT, Evanston, IL, USA
John Clements	California Polytechnic State University & PLT, San Luis Obispo & PLT, CA, USA
Christos Dimoulas	Northeastern University & PLT, Boston, MA, USA
Carl Eastlund	Northeastern University & PLT, Boston, MA, USA
Matthias Felleisen	Northeastern University & PLT, Boston, MA, USA
Matthew Flatt	University of Utah & PLT, Salt Lake City, UT, USA
Jay A. McCarthy	Brigham Young University & PLT, Provo, UT, USA
Jon Rafkind	University of Utah & PLT, Salt Lake City, UT, USA
Sam Tobin-Hochstadt	Northeastern University & PLT, Boston, MA, USA
Robert Bruce Findler	Northwestern University, Evanston, IL, USA

Bibliometrics

- Downloads (6 Weeks): 11
- Downloads (12 Months): 86
- Downloads (cumulative): 363
- Citation Count: 5

Published in:

- Proceeding
  POPL '12 Proceedings of the 39th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages
  Pages 285-296
  ACM New York, NY, USA ©2012
  table of contents ISBN: 978-1-4503-1083-3 doi>10.1145/2103656.2103691

Run your research

Klein et al, POPL 2012

- Investigates 9 papers from a major conference
- Selection criterion: suitable for formalisation in Redex (high level executable functional modelling language)
- Result: found serious mistakes in **all papers**
- Formalisation effort **less** than the effort to understand the papers

Errors in examples (results verified in Coq)

Mistake in translating Agda code to the paper

Decidability result false

Optimization applied also when unsound

False main theorem

Abstract machine uses unbounded resources

Program transformation undefined in presence of constants

Missing constructor definitions for some datatypes

Assumed decomposition lemma does not hold

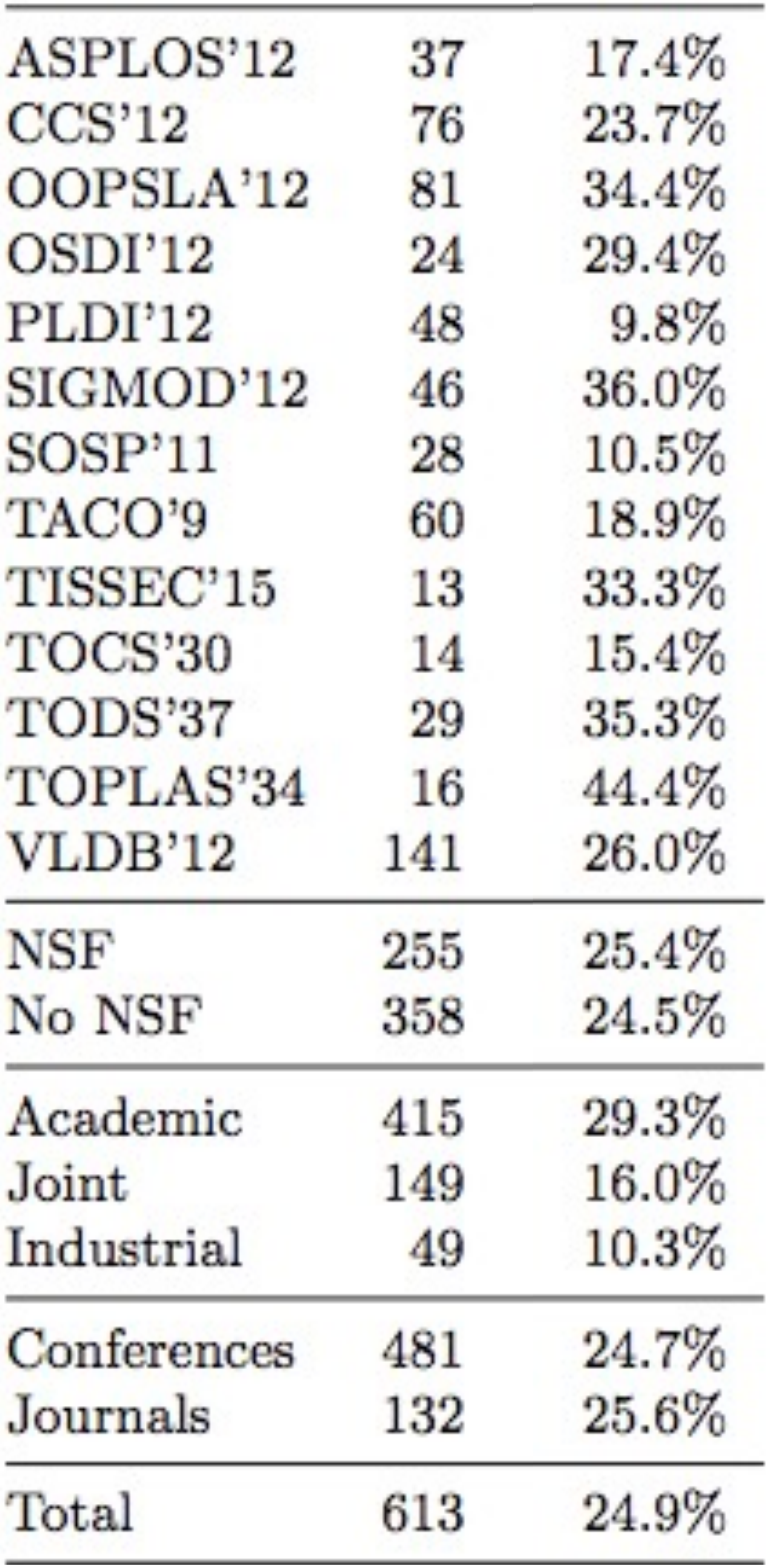
Measuring Reproducibility in Computer Systems Research

http://reproducibility.cs.arizona.edu/tr.pdf

Collberg et al, Univ. Arizona
March 2014

Examines reproducibility of tool performances

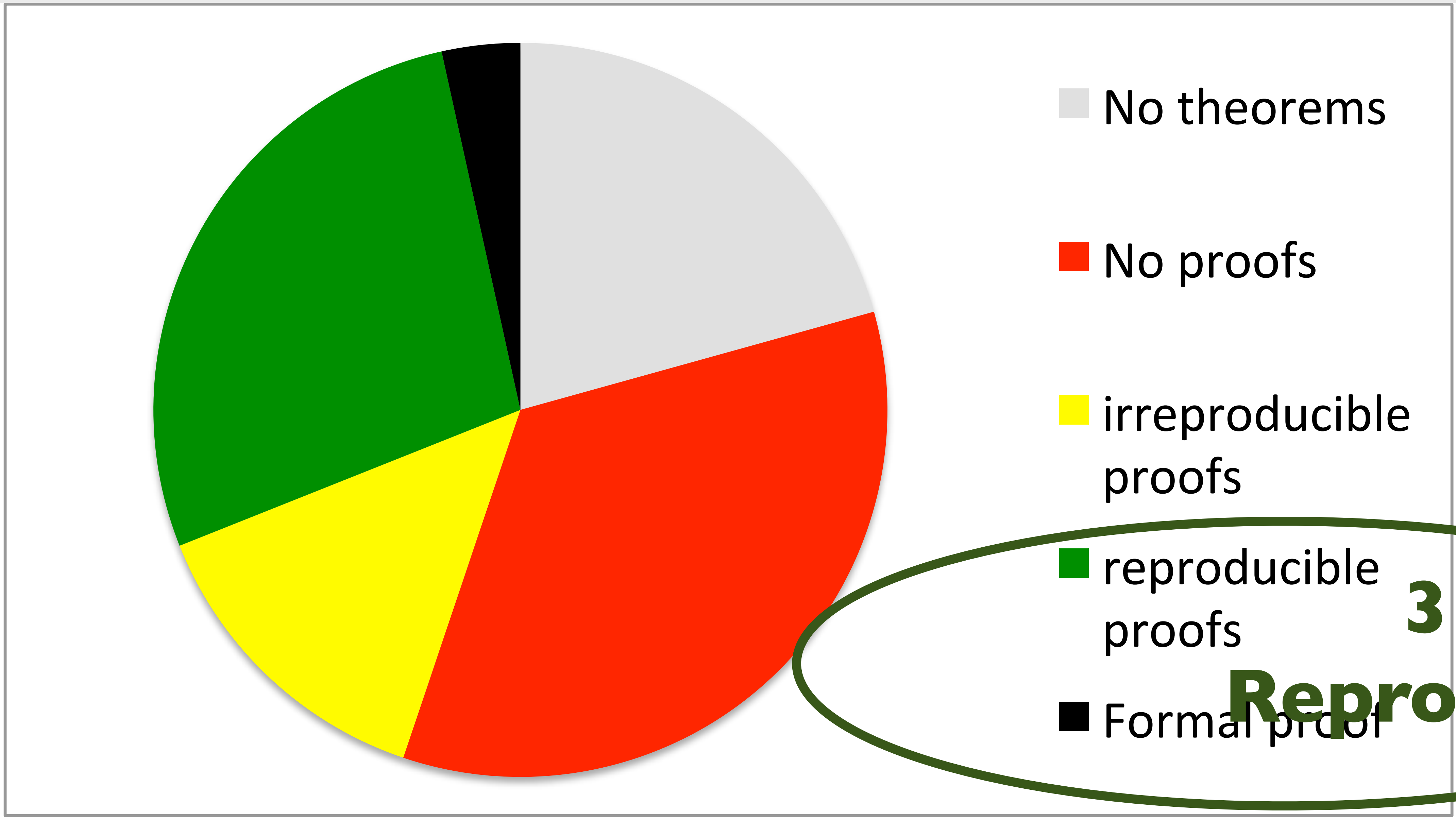
25% out of 613 tools could be built and run



	Papers	%reproducible
ASPLOS'12	37	17.4%
CCS'12	76	23.7%
OOPSLA'12	81	34.4%
OSDI'12	24	29.4%
PLDI'12	48	9.8%
SIGMOD'12	46	36.0%
SOSP'11	28	10.5%
TACO'9	60	18.9%
TISSEC'15	13	33.3%
TOCS'30	14	15.4%
TODS'37	29	35.3%
TOPLAS'34	16	44.4%
VLDB'12	141	26.0%
NSF	255	25.4%
No NSF	358	24.5%
Academic	415	29.3%
Joint	149	16.0%
Industrial	49	10.3%
Conferences	481	24.7%
Journals	132	25.6%
Total	613	24.9%

Reproducible proofs?

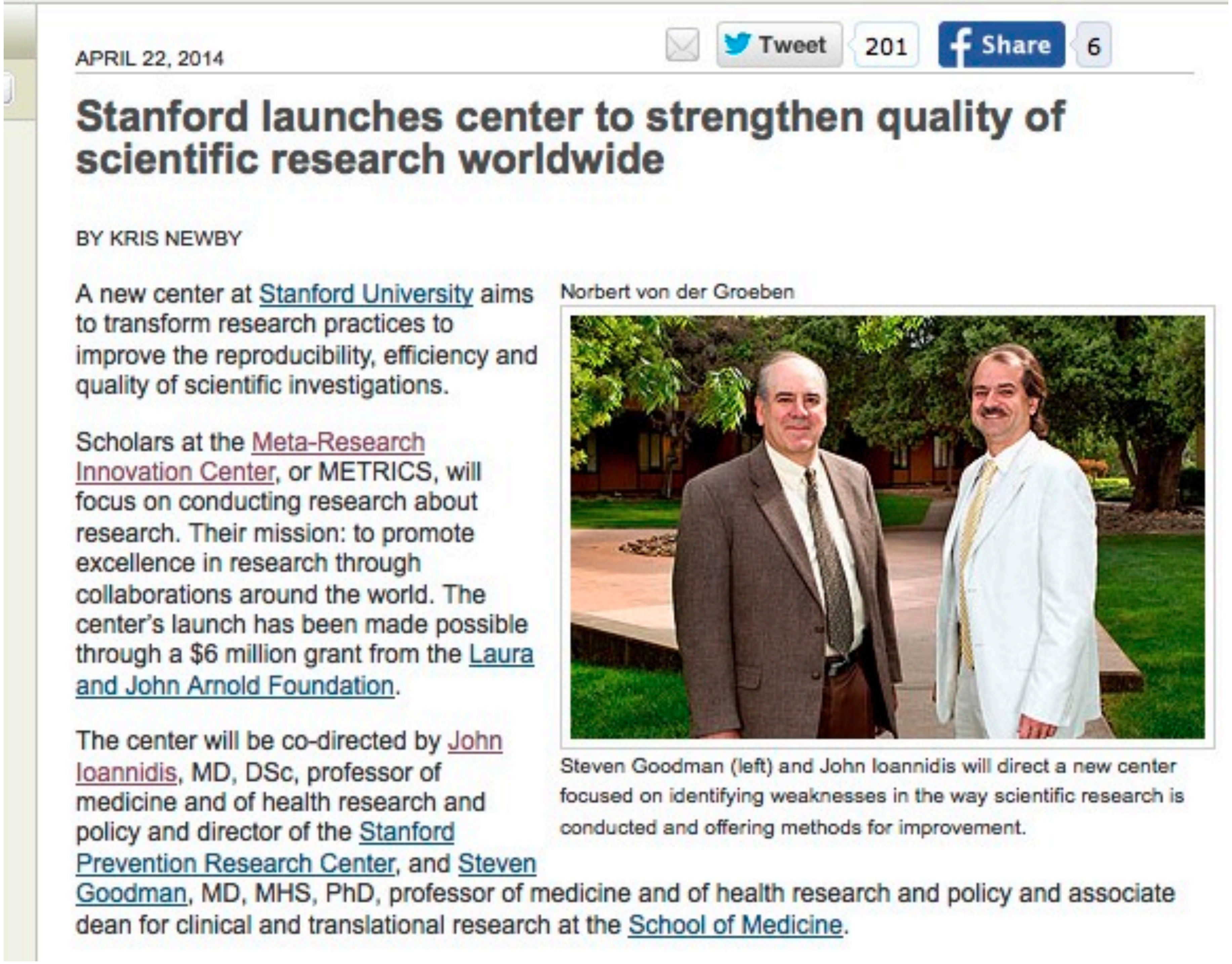
My own quick investigation of all 29 papers in ESOP 2014



31% Reproducible

Doing the Right Stuff

So what can we do?



APRIL 22, 2014

Stanford launches center to strengthen quality of scientific research worldwide

BY KRIS NEWBY

A new center at Stanford University aims to transform research practices to improve the reproducibility, efficiency and quality of scientific investigations.

Scholars at the Meta-Research Innovation Center, or METRICS, will focus on conducting research about research. Their mission: to promote excellence in research through collaborations around the world. The center's launch has been made possible through a $6 million grant from the Laura and John Arnold Foundation.

Norbert von der Groeben

Steven Goodman (left) and John Ioannidis will direct a new center focused on identifying weaknesses in the way scientific research is conducted and offering methods for improvement.

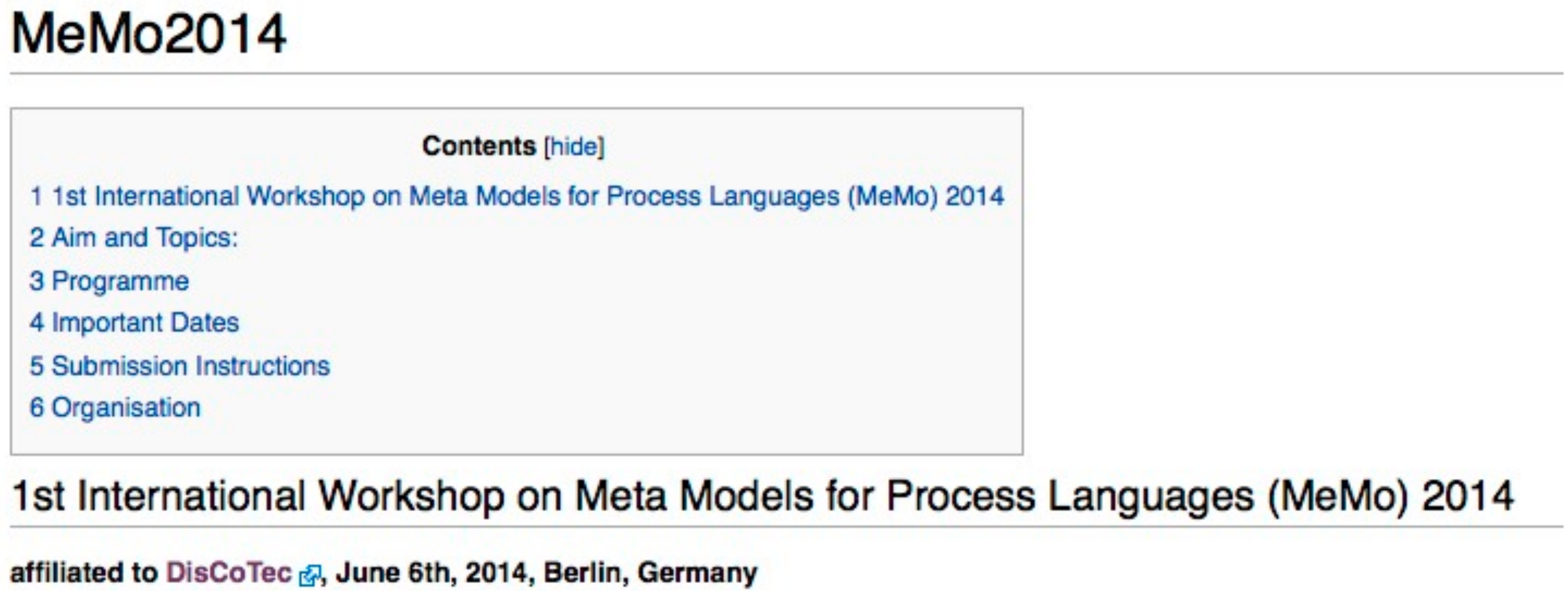
The center will be co-directed by John Ioannidis, MD, DSc, professor of medicine and of health research and policy and director of the Stanford Prevention Research Center, and Steven Goodman, MD, MHS, PhD, professor of medicine and of health research and policy and associate dean for clinical and translational research at the School of Medicine.

Structural changes

- More recognition for thorough results, less publish and perish
- More recognition for re-proving old results
- Better paid reviewers with more time
- Ignore results without full proofs

Meta models

Come to MeMo2014 tomorrow
to learn about meta models



MeMo2014

Contents [hide]

1 1st International Workshop on Meta Models for Process Languages (MeMo) 2014
2 Aim and Topics:
3 Programme
4 Important Dates
5 Submission Instructions
6 Organisation

1st International Workshop on Meta Models for Process Languages (MeMo) 2014

affiliated to DisCoTec, June 6th, 2014, Berlin, Germany

Get your stuff right

- Be careful in proofs.
- Write out all details
- Make available and have someone check

Use a theorem prover

- A tool to help you find and check proofs
- Better nomenclature: interactive proof assistant
- Much more usable today than ten years ago

Psi - calculi framework

- A **meta model** for process calculi
- Developed in Uppsala since 2008
- 2-6 persons working on it
- (Come to MeMo tomorrow to learn more)

The psi experience

- **Using Isabelle/Nominal to verify theory**
  - **What are the benefits?**
  - **What are the costs?**

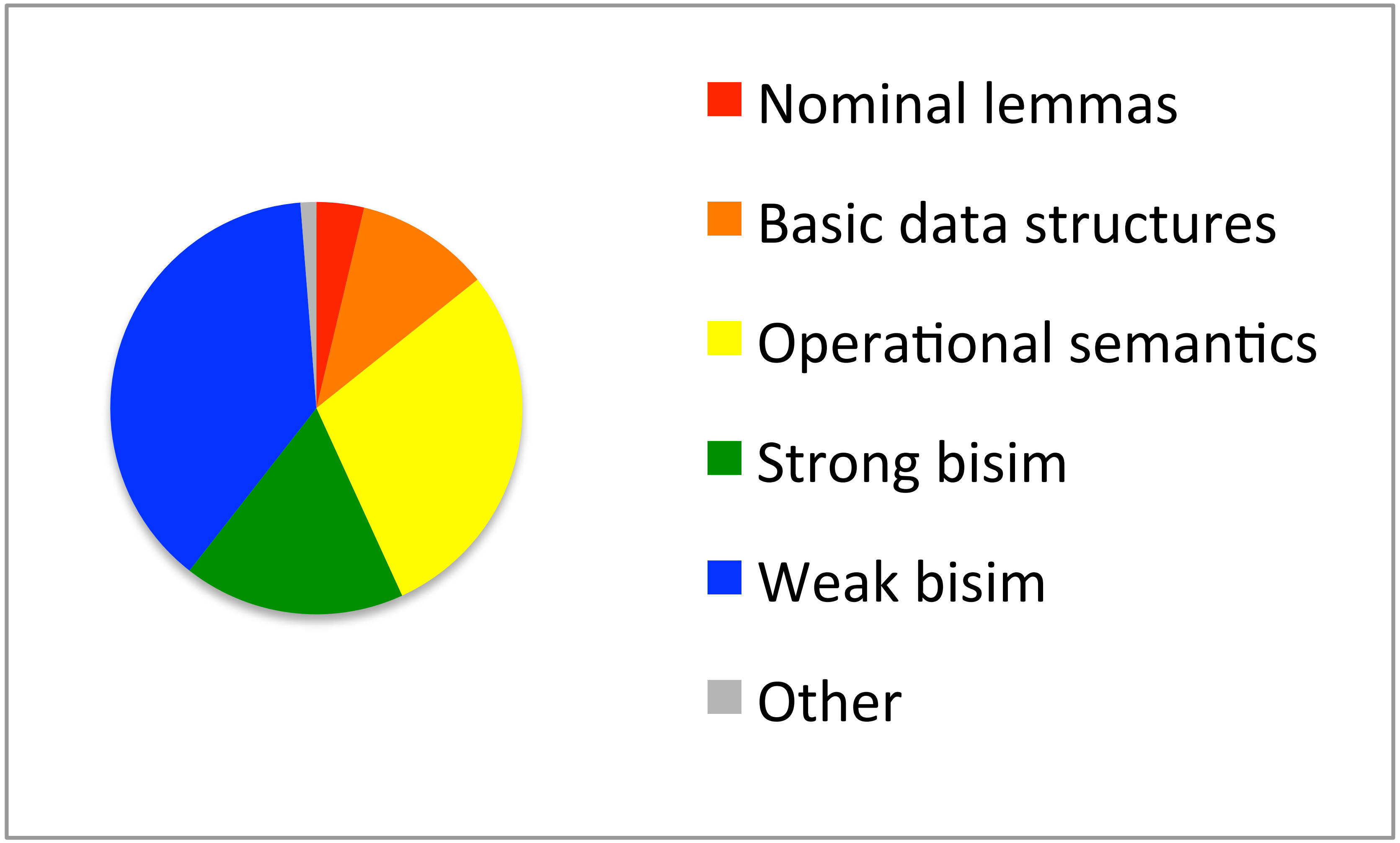
Using a theorem prover

Benefit 1: **Certainty** (no false assertions)
Benefit 2: Good proof **structure** (clarity of arguments)

Formalisation during development, not post hoc:

Benefit 3: **Flexibility** (easy to change details)
Benefit 4: **Generality** (keep track of assumptions)

Our proof archive, 2010



~32 KLoC

Example: case rule

$$\frac{\Psi \vdash \varphi_i}{\Psi \;\rhd\; \mathbf{case}\ \widetilde{\varphi} : \widetilde{P} \;\xrightarrow{\tau}\; P_i}$$

change to

$$\frac{\Psi \;\rhd\; P_i \;\xrightarrow{\alpha}\; P' \qquad \Psi \vdash \varphi_i}{\Psi \;\rhd\; \mathbf{case}\,\widetilde{\varphi} : \widetilde{P} \;\xrightarrow{\alpha}\; P'}$$

does this matter?

Example: Higher-order rule

$$\frac{\Psi \vdash M \Leftarrow P \qquad \Psi \ \rhd \ P \ \xrightarrow{\alpha} \ P'}{\Psi \ \rhd \ \mathbf{run}\, M \ \xrightarrow{\alpha} \ P'}$$

Now re-prove all the theory!

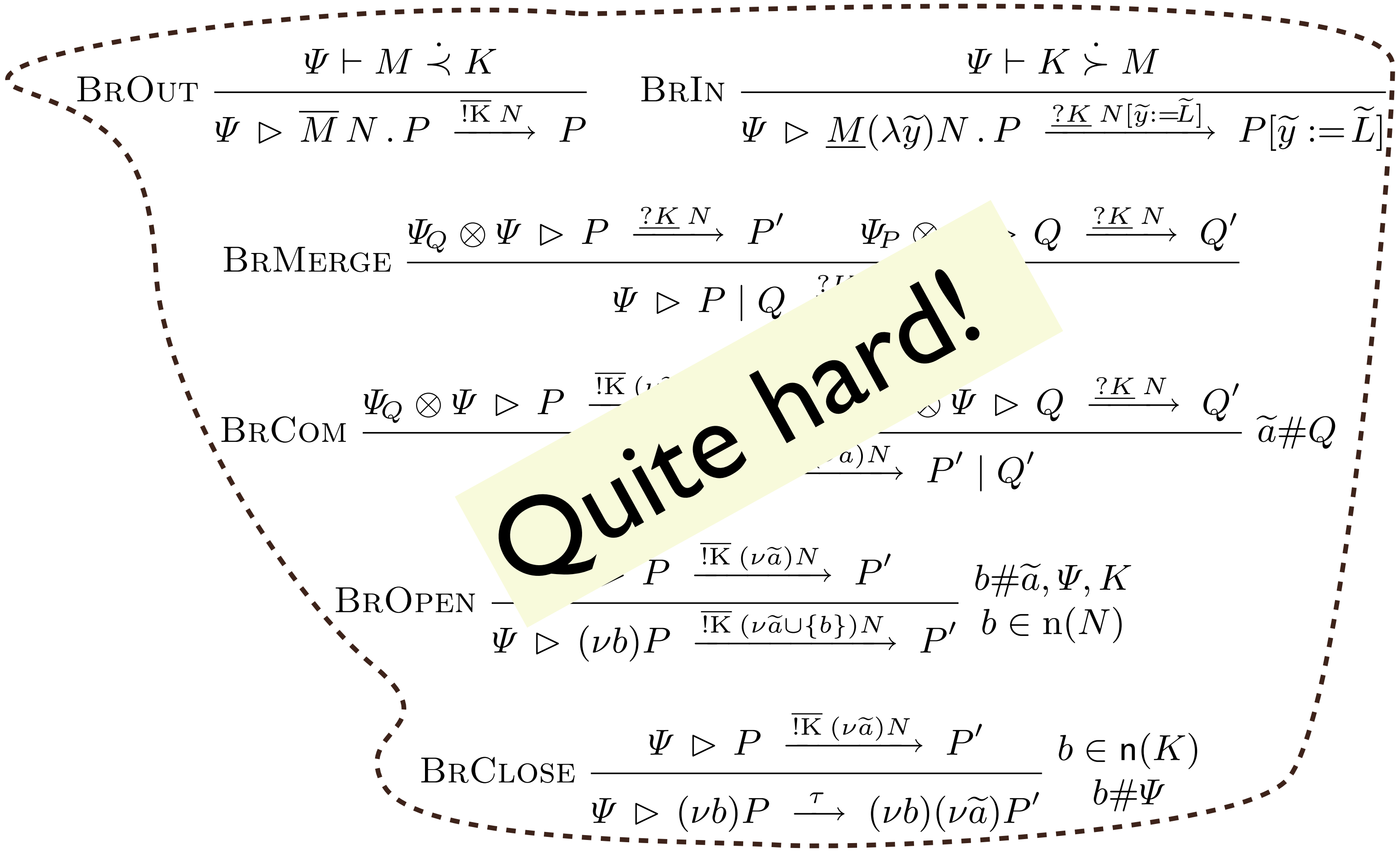
With Isabelle: took a day and a night

Example: Broadcast

One transmission : many listeners

Channels with dynamic connectivity

Six new semantic rules, two new kinds of action



$$\textsc{BrOut}\ \frac{\Psi \vdash M \mathrel{\dot{\prec}} K}{\Psi \rhd \overline{M}\,N\,.\,P \xrightarrow{\overline{!\mathrm{K}}\ N} P} \qquad \textsc{BrIn}\ \frac{\Psi \vdash K \mathrel{\dot{\succ}} M}{\Psi \rhd \underline{M}(\lambda\widetilde{y})N\,.\,P \xrightarrow{\underline{?K}\ N[\widetilde{y}:=\widetilde{L}]} P[\widetilde{y}:=\widetilde{L}]}$$

$$\textsc{BrMerge}\ \frac{\Psi_Q \otimes \Psi \rhd P \xrightarrow{\underline{?K}\ N} P' \qquad \Psi_P \otimes \quad Q \xrightarrow{\underline{?K}\ N} Q'}{\Psi \rhd P \mid Q \quad ?}$$

$$\textsc{BrCom}\ \frac{\Psi_Q \otimes \Psi \rhd P \xrightarrow{\overline{!\mathrm{K}}\ (} \qquad \otimes \Psi \rhd Q \xrightarrow{\underline{?K}\ N} Q'}{\xrightarrow{\quad a)N} P' \mid Q'}\ \widetilde{a}\#Q$$

$$\textsc{BrOpen}\ \frac{P \xrightarrow{\overline{!\mathrm{K}}\ (\nu\widetilde{a})N} P'}{\Psi \rhd (\nu b)P \xrightarrow{\overline{!\mathrm{K}}\ (\nu\widetilde{a}\cup\{b\})N} P'}\ \begin{array}{l} b\#\widetilde{a},\Psi,K \\ b \in \mathrm{n}(N) \end{array}$$

$$\textsc{BrClose}\ \frac{\Psi \rhd P \xrightarrow{\overline{!\mathrm{K}}\ (\nu\widetilde{a})N} P'}{\Psi \rhd (\nu b)P \xrightarrow{\tau} (\nu b)(\nu\widetilde{a})P'}\ \begin{array}{c} b \in \mathsf{n}(K) \\ b\#\Psi \end{array}$$

Quite hard!

Example: HO broadcast

Combining broadcast and higher order

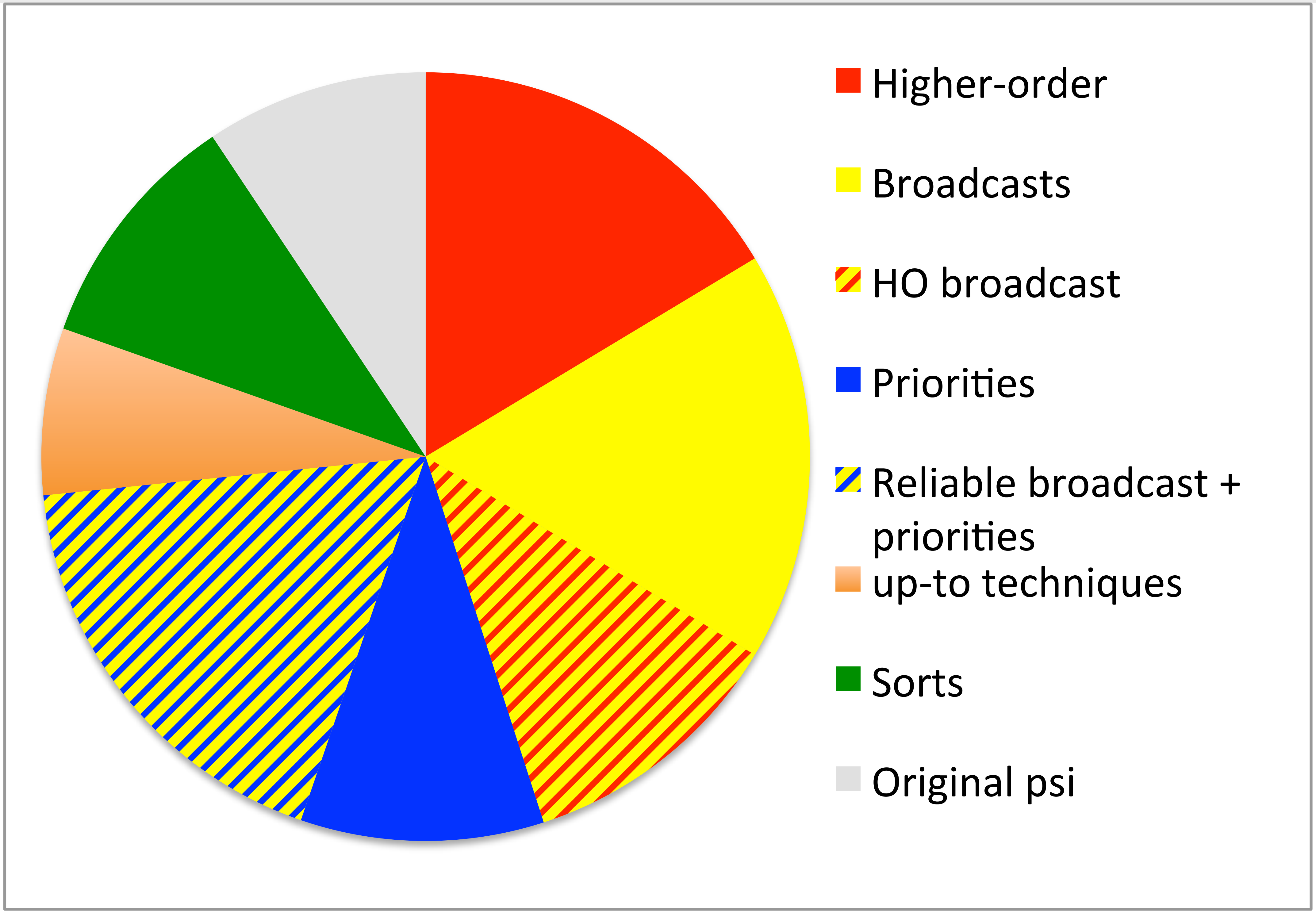
”These extensions don’t interact” (wild handwaving)

With Isabelle, took half a day and a cup of tea

Experiences

- Facilitated continuous development
- Absolutely necessary to gain confidence
- Main error source: theorem formulation
- Isabelle/Nominal itself is developing...

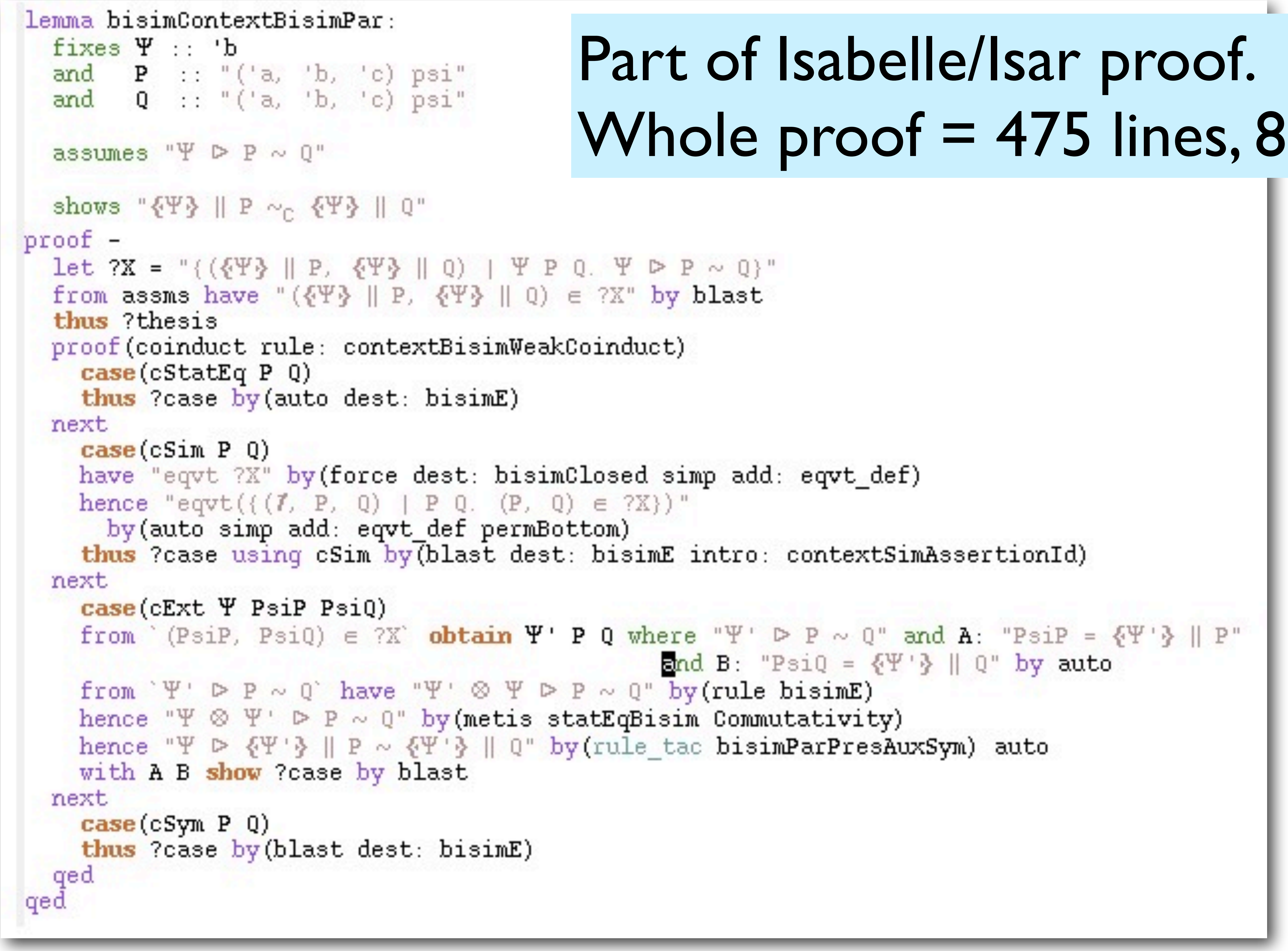
Our proof archive, 2013



342 KLoC

What about the cost?

Part of Isabelle/Isar proof.
Whole proof = 475 lines, 8h work



```
lemma bisimContextBisimPar:
  fixes Ψ :: 'b
  and   P  :: "('a, 'b, 'c) psi"
  and   Q  :: "('a, 'b, 'c) psi"

  assumes "Ψ ⊳ P ∼ Q"

  shows "⦃Ψ⦄ ∥ P ∼c ⦃Ψ⦄ ∥ Q"
proof -
  let ?X = "{(⦃Ψ⦄ ∥ P, ⦃Ψ⦄ ∥ Q) | Ψ P Q. Ψ ⊳ P ∼ Q}"
  from assms have "(⦃Ψ⦄ ∥ P, ⦃Ψ⦄ ∥ Q) ∈ ?X" by blast
  thus ?thesis
  proof(coinduct rule: contextBisimWeakCoinduct)
    case(cStatEq P Q)
    thus ?case by(auto dest: bisimE)
  next
    case(cSim P Q)
    have "eqvt ?X" by(force dest: bisimClosed simp add: eqvt_def)
    hence "eqvt({(𝟭, P, Q) | P Q. (P, Q) ∈ ?X})"
      by(auto simp add: eqvt_def permBottom)
    thus ?case using cSim by(blast dest: bisimE intro: contextSimAssertionId)
  next
    case(cExt Ψ PsiP PsiQ)
    from `(PsiP, PsiQ) ∈ ?X` obtain Ψ' P Q where "Ψ' ⊳ P ∼ Q" and A: "PsiP = ⦃Ψ'⦄ ∥ P"
                                                  and B: "PsiQ = ⦃Ψ'⦄ ∥ Q" by auto
    from `Ψ' ⊳ P ∼ Q` have "Ψ' ⊗ Ψ ⊳ P ∼ Q" by(rule bisimE)
    hence "Ψ ⊗ Ψ' ⊳ P ∼ Q" by(metis statEqBisim Commutativity)
    hence "Ψ ⊳ ⦃Ψ'⦄ ∥ P ∼ ⦃Ψ'⦄ ∥ Q" by(rule_tac bisimParPresAuxSym) auto
    with A B show ?case by blast
  next
    case(cSym P Q)
    thus ?case by(blast dest: bisimE)
  qed
qed
```

Part of corresponding manual proof. From our email archive. Whole proof = 70 lines, 2h work

A binary relation R on agents is an MJbisim if R(P,Q) implies

1. F(P)=F(Q) (static equiv)
2. R(Q,P)
3. Forall Psi. R({Psi}|P, {Psi}|Q)
4. Forall a s.t. bn(a)#Q. P -a-> P' => Q-a->Q' and R(P',Q')

(here transitions without assertion means bottom assertion)

Conjecture 1.
a) Psi |> P -a-> P' implies {Psi}|P -a-> {Psi}|P'.
b) {Psi}|P -a-> T implies exists P'. T = {Psi}|P' and Psi |> P -a-> P'
Proof: For a: by the PAR rule and F({Psi})=Psi. For b: case analysis on deirivation of {Psi}|P-a->T, and here only PAR can be used. Details are left as an exercise for the reader :)

Conjecture 2.
{Psi}|{Psi'} ~ {Psi+Psi'}
Proof: Directly from definitions. Obvious :)

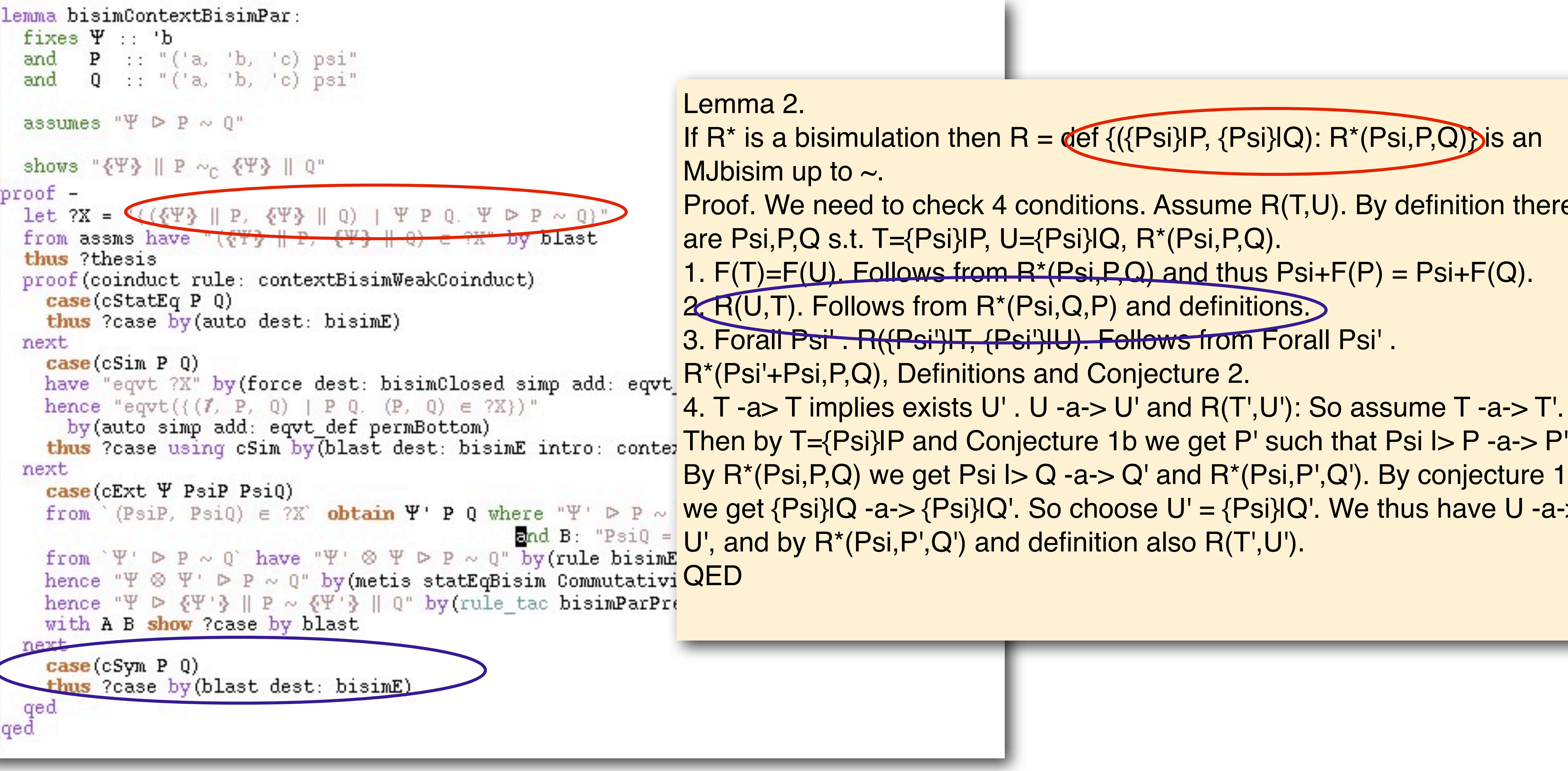
Conjecture 3. If R is an MJbisim up to ~ and R(P,Q) then there is an MJbisim R' such that R'(P,Q)
Proof: By intimidation :)

Lemma 1
If R is an MJbisim then R* =def {(Psi, P,Q): R({Psi}|P, {Psi}|Q)} is a bisimulation up to ~
Proof. We need to check 4 conditions. Assume R*(Psi,P.Q). Then R({Psi}|P, {Psi}|Q).
1. Psi + F(P) = Psi + F(Q). Follows from F({Psi}|P) = F({Psi}|Q).
2. R*(Psi,Q,P). Follows from R({Psi}|Q, {Psi}|P).
3. All Psi' . R*(Psi+Psi', P, Q). Follows from All Psi' . R({Psi'}|{Psi}|P, {Psi'}|{Psi}|Q), and Conjecture 2. Note that here we probably need associativity.
4. Psi |> P-a-> P' implies exists Q' . Psi |> Q-a-> Q' and R(Psi,P',Q'). So assume Psi |> P-a-> P'. Then by Conjecture 1a {Psi}|P -a-> {Psi}|P'. By Condition 4 on MJbisim and R({Psi}|P, {Psi}|Q)
{Psi}|Q -a-> T with R({Psi}|P',T). Conjecture 1b then gives that there exists a Q' such that T = {Psi}|Q' and {Psi} |> Q -a-> Q'. Also R({Psi}|P',{Psi}|Q') by definition implies R*(Psi,P',Q'), as requ
QED

Lemma 2.
If R* is a bisimulation then R = def {({Psi}|P, {Psi}|Q): R*(Psi,P,Q)} is an MJbisim up to ~.
Proof. We need to check 4 conditions. Assume R(T,U). By definition there are Psi,P,Q s.t. T={Psi}|P, U={Psi}|Q, R*(Psi,P,Q).
1. F(T)=F(U). Follows from R*(Psi,P,Q) and thus Psi+F(P) = Psi+F(Q).
2. R(U,T). Follows from R*(Psi,Q,P) and definitions.
3. Forall Psi' . R({Psi'}|T, {Psi'}|U). Follows from Forall Psi' . R*(Psi'+Psi,P,Q), Definitions and Conjecture 2.
4. T -a> T implies exists U' . U -a-> U' and R(T',U'): So assume T -a-> T'. Then by T={Psi}|P and Conjecture 1b we get P' such that Psi |> P -a-> P'. By R*(Psi,P,Q) we get Psi |> Q -a-> Q' and
R*(Psi,P',Q'). By conjecture 1a we get {Psi}|Q -a-> {Psi}|Q'. So choose U' = {Psi}|Q'. We thus have U -a-> U', and by R*(Psi,P',Q') and definition also R(T',U').
QED

Corollary
P ~ Q iff there exists an MJbisim R such that R(P,Q)
Proof.
=>: Suppose P ~ Q. Then there is a bisimulation R* such that R*(bot,P,Q). Define R as in Lemma 2, using this R*. It follows that R is an MJbisim and R(0|P, 0|Q), and therefore R U {(P,Q)} is
MJ-bisimulation up to ~. By Conjecture 3 there is than an MJbisim as required.
<=: Suppose R is an MJ-bisimulation up to ~ and R(P,Q). Then R(0|P, 0|Q). By Conjecture 3 there is an MJbisim R' such that R'(0|P,0|Q). So by Lemma 1 there is a bisimulation (up to ~) R* su
R*(bot,P,Q), which implies P~Q.

Structure vs Syntax



```
lemma bisimContextBisimPar:
  fixes Ψ :: 'b
  and   P  :: "('a, 'b, 'c) psi"
  and   Q  :: "('a, 'b, 'c) psi"

  assumes "Ψ ⊳ P ∼ Q"

  shows "{Ψ} || P ∼c {Ψ} || Q"
proof -
  let ?X = "{({Ψ} || P, {Ψ} || Q) | Ψ P Q. Ψ ⊳ P ∼ Q}"
  from assms have "({Ψ} || P, {Ψ} || Q) ∈ ?X" by blast
  thus ?thesis
  proof(coinduct rule: contextBisimWeakCoinduct)
    case(cStatEq P Q)
    thus ?case by(auto dest: bisimE)
  next
    case(cSim P Q)
    have "eqvt ?X" by(force dest: bisimClosed simp add: eqvt_
    hence "eqvt({(𝟙, P, Q) | P Q. (P, Q) ∈ ?X})"
      by(auto simp add: eqvt_def permBottom)
    thus ?case using cSim by(blast dest: bisimE intro: contex
  next
    case(cExt Ψ PsiP PsiQ)
    from `(PsiP, PsiQ) ∈ ?X` obtain Ψ' P Q where "Ψ' ⊳ P ∼
                                              and B: "PsiQ =
    from `Ψ' ⊳ P ∼ Q` have "Ψ' ⊗ Ψ ⊳ P ∼ Q" by(rule bisimE
    hence "Ψ ⊗ Ψ' ⊳ P ∼ Q" by(metis statEqBisim Commutativi
    hence "Ψ ⊳ {Ψ'} || P ∼ {Ψ'} || Q" by(rule_tac bisimParPre
    with A B show ?case by blast
  next
    case(cSym P Q)
    thus ?case by(blast dest: bisimE)
  qed
qed
```

Lemma 2.
If R* is a bisimulation then R = def {({Psi}|P, {Psi}|Q): R*(Psi,P,Q)} is an MJbisim up to ~.
Proof. We need to check 4 conditions. Assume R(T,U). By definition there are Psi,P,Q s.t. T={Psi}|P, U={Psi}|Q, R*(Psi,P,Q).
1. F(T)=F(U). Follows from R*(Psi,P,Q) and thus Psi+F(P) = Psi+F(Q).
2. R(U,T). Follows from R*(Psi,Q,P) and definitions.
3. Forall Psi' . R({Psi'}|T, {Psi'}|U). Follows from Forall Psi' . R*(Psi'+Psi,P,Q), Definitions and Conjecture 2.
4. T -a> T implies exists U' . U -a-> U' and R(T',U'): So assume T -a-> T'. Then by T={Psi}|P and Conjecture 1b we get P' such that Psi |> P -a-> P' By R*(Psi,P,Q) we get Psi |> Q -a-> Q' and R*(Psi,P',Q'). By conjecture 1 we get {Psi}|Q -a-> {Psi}|Q'. So choose U' = {Psi}|Q'. We thus have U -a- U', and by R*(Psi,P',Q') and definition also R(T',U').
QED

The cost?

One measure of effort: **”manhours”**

This particular proof:

Isabelle effort is four times the manual proof

In general

This factor varies wildly

The cost?

One measure of effort: **”manhours”**

Theory development is not exclusively
- not even mainly -
about writing down proofs.

So the factor is not so important.

The cost!

Study of time spent by 4 persons over 25 months on developing the Psi framework

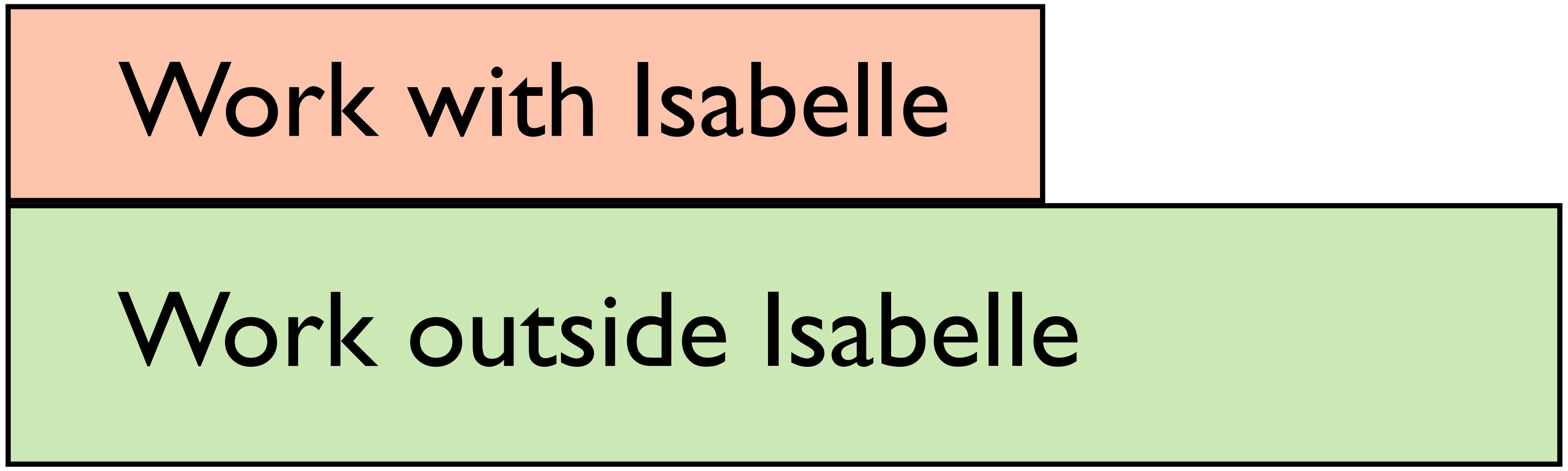
1/3 of the effort went into Isabelle formalisation

2/3 of the results have been fully formalised

The cost!

1/3 of the effort went into Isabelle formalisation

2/3 of the results have been fully formalised



Work with Isabelle

Work outside Isabelle

The Right Stuff

Apollo 13 landing, April 17 1970

A lecture by Joachim Parrow (2014) about the fine qualities of contemporary computer science

Correctness in the face of complications

”Failure is not an option”

Our motto, from now on!

Thank you!

Addendum: references

How Science Goes Wrong. *The Economist*, 2013 Oct 19th.

Begley, C. Glenn, and Lee M. Ellis. "Drug development: Raise standards for preclinical cancer research." *Nature* 483.7391 (2012): 531-533

Bohannon, John. "Who's Afraid of Peer Review?." *Science* 342.6154 (2013): 60-65.

Godlee, Fiona, Catharine R. Gale, and Christopher N. Martyn. "Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial." *Jama* 280.3 (1998): 237-240.

Fanelli, Daniele. "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data." *PloS one* 4.5 (2009): e5738.

Vasilevsky, Nicole A., et al. "On the reproducibility of science: unique identification of research resources in the biomedical literature." *PeerJ* 1 (2013): e148.

Ioannidis, John PA. "Why most published research findings are false." *PLoS medicine* 2.8 (2005): e124.

Klein, Casey, et al. Run your research: on the effectiveness of lightweight mechanization. In: *ACM SIGPLAN Notices* (Vol. 47, No. 1). ACM, 2012. p. 285-296.

Collberg, Christian et al. "Measuring Reproducibility in Computer Systems Research." Tech. Report, Univ. Arizona, March 2014. http://reproducibility.cs.arizona.edu/

Newby, Kris. ”Stanford launches center to strengthen quality of scientific research worldwide”. April 22, 2014. http://med.stanford.edu/ism/2014/april/metrics.html

Material on the research on psi-calculi and associated formal proofs can be found at http://www.it.uu.se/research/group/mobility/