LABEX MILYON
UNIVERSITÉ DE LYON

Papers have bugs — what is to be done?

Lê Thành Dũng (Tito) Nguyễn
ÉNS Lyon – postdoc fellowship from *LabEx MILyon, also an Undone CS sponsor*
joint with **Enka Blanchard**
CNRS / LAMIH, Univ. Polytechnique Hauts-de-France / Centre Internet et Société

Undone Computer Science, Nantes, 7 II 2024

- An ironically unrigorous rant about my frustration with one aspect of "definition-theorem-proof" computer science research
  - Not covered here: why is such research worth doing?
- A proposal mostly by Enka Blanchard (at the end);
  they are motivated by more socially relevant research
  - E. B., Fabrizio Li Vigni & Pablo Rauzy (2022) detail the flaws in a paper promoting blockchain-based electronic voting
    `https://hal.science/hal-03741811`
- Trying to be short $\longrightarrow$ time for discussion

No definitive truths in science?

Obvious fact: X is claimed in a published paper $\not\Rightarrow$ X is true

Humanities / social science: debate between several opposing viewpoints

Natural sciences: experiments may not replicate; theories may be overturned

No definitive truths in science?

Obvious fact: X is claimed in a published paper $\not\Rightarrow$ X is true

Humanities / social science: debate between several opposing viewpoints

Natural sciences: experiments may not replicate; theories may be overturned

Formal sciences (math, theoretical comp. sci., ...):

a valid proof should stay valid forever (as long as you accept the axioms)

- usual standards: any published theorem can be taken for granted in one's own work...

No definitive truths in science?

Obvious fact: X is claimed in a published paper $\not\Rightarrow$ X is true

Humanities / social science: debate between several opposing viewpoints

Natural sciences: experiments may not replicate; theories may be overturned

Formal sciences (math, theoretical comp. sci., ...):

a valid proof should stay valid forever (as long as you accept the axioms)

- usual standards: any published theorem can be taken for granted in one's own work...
- but "everybody knows" the literature is **full of mistakes**!

Automata theory anecdote 1

Autobóz problem-solving camp, Polish countryside, 1 week in September 2022

Day 1 several open problems proposed by participants

Day 3 "Our group managed to solve problem X using published paper Y!"

Automata theory anecdote 1

Autobóz problem-solving camp, Polish countryside, 1 week in September 2022

Day 1 several open problems proposed by participants

Day 3 "Our group managed to solve problem X using published paper Y!"

Day 4 "On closer examination, turns out the theorem we were using from paper Y was wrong..."

Automata theory anecdote 1

Autobóz problem-solving camp, Polish countryside, 1 week in September 2022

Day 1 several open problems proposed by participants

Day 3 "Our group managed to solve problem X using published paper Y!"

Day 4 "On closer examination, turns out the theorem we were using from paper Y was wrong..."

Today (Feb 2024) This error is (as far as I can tell) *not publicly mentioned anywhere* even though some of the coauthors of Y attended the camp

Automata theory anecdote 1

Autobóz problem-solving camp, Polish countryside, 1 week in September 2022

Day 1 several open problems proposed by participants

Day 3 "Our group managed to solve problem X using published paper Y!"

Day 4 "On closer examination, turns out the theorem we were using from paper Y was wrong..."

Today (Feb 2024) This error is (as far as I can tell) *not publicly mentioned anywhere* even though some of the coauthors of Y attended the camp

When such a situation lasts for long enough, it's called *folklore*

$\longrightarrow$ especially annoying when you're a beginner in the field

Automata theory anecdote 2

2020 Main theorem of paper at *Logic in Computer Science* (LICS):
"pebble transducers can be minimized" (whatever that means)

Automata theory anecdote 2

2020 Main theorem of paper at *Logic in Computer Science* (LICS):
"pebble transducers can be minimized" (whatever that means)

early 2021 "a variant of pebble transducers can be minimized" [N., Noûs & Pradic]
reused parts of LICS'20 paper; didn't understand some other parts...

Automata theory anecdote 2

2020 Main theorem of paper at *Logic in Computer Science* (LICS):
"pebble transducers can be minimized" (whatever that means)

early 2021 "a variant of pebble transducers can be minimized" [N., Noûs & Pradic]
reused parts of LICS'20 paper; didn't understand some other parts...

summer 2021 "wait, this simple pebble transducer doesn't look minimizable"
[Bojańczyk, Douéneau-Tabot, Kiefer, N. & Pradic]

Automata theory anecdote 2

2020 Main theorem of paper at *Logic in Computer Science* (LICS):
"pebble transducers can be minimized" (whatever that means)

early 2021 "a variant of pebble transducers can be minimized" [N., Noûs & Pradic]
reused parts of LICS'20 paper; didn't understand some other parts...

summer 2021 "wait, this simple pebble transducer doesn't look minimizable"
[Bojańczyk, Douéneau-Tabot, Kiefer, N. & Pradic]

early 2023 2 papers with 3 proofs of pebble *non*-minimization appear on arXiv

Automata theory anecdote 2

2020 Main theorem of paper at *Logic in Computer Science* (LICS):
"pebble transducers can be minimized" (whatever that means)

early 2021 "a variant of pebble transducers can be minimized" [N., Noûs & Pradic]
reused parts of LICS'20 paper; didn't understand some other parts...

summer 2021 "wait, this simple pebble transducer doesn't look minimizable"
[Bojańczyk, Douéneau-Tabot, Kiefer, N. & Pradic]

early 2023 2 papers with 3 proofs of pebble *non*-minimization appear on arXiv

"But... but... but... LICS is an A*-ranked conference!"

Automata theory anecdote 2

2020 Main theorem of paper at *Logic in Computer Science* (LICS):
"pebble transducers can be minimized" (whatever that means)

early 2021 "a variant of pebble transducers can be minimized" [N., Noûs & Pradic]
reused parts of LICS'20 paper; didn't understand some other parts...

summer 2021 "wait, this simple pebble transducer doesn't look minimizable"
[Bojańczyk, Douéneau-Tabot, Kiefer, N. & Pradic]

early 2023 2 papers with 3 proofs of pebble *non*-minimization appear on arXiv

"But... but... but... LICS is an A*-ranked conference!"
well, perhaps that's actually the problem?

The mandatory rant about conference proceedings

Author's side: can't miss the once-a-year [insert prestigious conf. here] deadline

$\rightsquigarrow$ rush paper writing process, cut corners

$\rightsquigarrow$ sloppy proofs, maybe not fully honest

The mandatory rant about conference proceedings

Author's side: can't miss the once-a-year [insert prestigious conf. here] deadline

$\rightsquigarrow$ rush paper writing process, cut corners

$\rightsquigarrow$ sloppy proofs, maybe not fully honest

Program committee side: strict deadline for reviews

+ get a mountain of highly technical submissions at once

$\rightsquigarrow$ no time to carefully check proofs

The mandatory rant about conference proceedings

Author's side: can't miss the once-a-year [insert prestigious conf. here] deadline

$\rightsquigarrow$ rush paper writing process, cut corners

$\rightsquigarrow$ sloppy proofs, maybe not fully honest

Program committee side: strict deadline for reviews

+ get a mountain of highly technical submissions at once

$\rightsquigarrow$ no time to carefully check proofs

(not to speak of the harms of mandatory in-person attendance...)

The mandatory rant about conference proceedings

Author's side: can't miss the once-a-year [insert prestigious conf. here] deadline
$\rightsquigarrow$ rush paper writing process, cut corners
$\rightsquigarrow$ sloppy proofs, maybe not fully honest

Program committee side: strict deadline for reviews
+ get a mountain of highly technical submissions at once
$\rightsquigarrow$ no time to carefully check proofs

(not to speak of the harms of mandatory in-person attendance...)

Other fields have their own issues, e.g. journal impact factor sometimes negatively correlated with quality [Dougherty & Horne 2022]

Even pure math has its issues

Pure mathematicians publish in journals & don't care about impact factors...

Even pure math has its issues

Pure mathematicians publish in journals & don't care about impact factors...

but they also run into lots of bugs in papers!

$\rightarrow$ see any recent talk by K. Buzzard for examples

e.g. both A and not(A) published in *Annals of Mathematics*, neither retracted

Even pure math has its issues

Pure mathematicians publish in journals & don't care about impact factors...

but they also run into lots of bugs in papers!

$\rightarrow$ see any recent talk by K. Buzzard for examples

e.g. both A and not(A) published in *Annals of Mathematics*, neither retracted

V. Voevodsky on an error found in 2000

"Starting from 1993, multiple groups of mathematicians studied my paper at seminars and used it in their work and none of them noticed the mistake. [...] A technical argument by a trusted author, which is hard to check and looks similar to arguments known to be correct, is hardly ever checked in detail."

Even pure math has its issues

Pure mathematicians publish in journals & don't care about impact factors...
but they also run into lots of bugs in papers!

$\rightarrow$ see any recent talk by K. Buzzard for examples
e.g. both A and not(A) published in *Annals of Mathematics*, neither retracted

V. Voevodsky on an error found in 2000

"Starting from 1993, multiple groups of mathematicians studied my paper at seminars and used it in their work and none of them noticed the mistake. [...] A technical argument by a trusted author, which is hard to check and looks similar to arguments known to be correct, is hardly ever checked in detail."

Voevodsky and Buzzard both suggest using computer-assisted *formal proofs*

Formal proofs are no panacea

Formal proofs offer strong guarantees; routinely used e.g. in programming languages theory (for tedious proofs: “induction on syntax w/ 25 cases…”).

Formal proofs are no panacea

Formal proofs offer strong guarantees; routinely used e.g. in programming languages theory (for tedious proofs: "induction on syntax w/ 25 cases..."). But:

- They can be very time-consuming

Formal proofs are no panacea

Formal proofs offer strong guarantees; routinely used e.g. in programming languages theory (for tedious proofs: "induction on syntax w/ 25 cases..."). But:

- They can be very time-consuming
- No guarantee that "what is proved" = "what we (informally) want"

(orthogonal issue to human-language vs formal proofs)

Formal proofs are no panacea

Formal proofs offer strong guarantees; routinely used e.g. in programming languages theory (for tedious proofs: "induction on syntax w/ 25 cases..."). But:

- They can be very time-consuming
- No guarantee that "what is proved" = "what we (informally) want"

(orthogonal issue to human-language vs formal proofs)

$\rightsquigarrow$ important issue in direct applications of formal sciences, e.g. cryptography

(cf. Koblitz & Menezes *Another Look at Provable Security* paper series)

(more conference-proceedings-bashing in N. Koblitz's 2007 *Notices of the AMS* column)

Formal proofs are no panacea

Formal proofs offer strong guarantees; routinely used e.g. in programming languages theory (for tedious proofs: "induction on syntax w/ 25 cases..."). But:

- They can be very time-consuming
- No guarantee that "what is proved" = "what we (informally) want"

(orthogonal issue to human-language vs formal proofs)

$\rightsquigarrow$ important issue in direct applications of formal sciences, e.g. cryptography

(cf. Koblitz & Menezes *Another Look at Provable Security* paper series)

(more conference-proceedings-bashing in N. Koblitz's 2007 *Notices of the AMS* column)

$\rightsquigarrow$ finding bugs in proofs is not enough for computer science,

we also want *methodological critiques*

To recap until now...

- Many systemic incentives for bad science
- But in any case, ruling out bugs *a priori* is unrealistic
  $\longrightarrow$ need for *a posteriori* **correction mechanisms**

To recap until now...

- Many systemic incentives for bad science
- But in any case, ruling out bugs *a priori* is unrealistic
  $\longrightarrow$ need for *a posteriori* **correction mechanisms**

- Unwritten folklore

To recap until now...

- Many systemic incentives for bad science
- But in any case, ruling out bugs *a priori* is unrealistic
  $\longrightarrow$ need for *a posteriori* **correction mechanisms**

- Unwritten folklore
- Errata: rarely used for journal papers or preprint servers
  are there any for conference proceedings??

To recap until now...

- Many systemic incentives for bad science
- But in any case, ruling out bugs *a priori* is unrealistic
  $\longrightarrow$ need for *a posteriori* **correction mechanisms**

- Unwritten folklore
- Errata: rarely used for journal papers or preprint servers
  are there any for conference proceedings??
  existence of refutation $\not\Longrightarrow$ original paper amended or retracted
  stops being cited

To recap until now...

- Many systemic incentives for bad science
- But in any case, ruling out bugs *a priori* is unrealistic
  $\longrightarrow$ need for *a posteriori* **correction mechanisms**

- Unwritten folklore
- Errata: rarely used for journal papers or preprint servers
  are there any for conference proceedings??
  existence of refutation $\not\Longrightarrow$ original paper amended or retracted
  stops being cited
- Public reviews (OpenReview, PubPeer, ...): not well-established in comp. sci.

To recap until now...

- Many systemic incentives for bad science
- But in any case, ruling out bugs *a priori* is unrealistic
  $\longrightarrow$ need for *a posteriori* **correction mechanisms**

- Unwritten folklore
- Errata: rarely used for journal papers or preprint servers
  are there any for conference proceedings??
  existence of refutation $\not\Longrightarrow$ original paper amended or retracted
  stops being cited
- Public reviews (OpenReview, PubPeer, ...): not well-established in comp. sci.
  "I have an *officially peer-reviewed* publication with a DOI,
  you're just slandering me on some random website"

How to lend “academic value” to refutations?

How to lend "academic value" to refutations?

Proposal: a journal of critiques and refutations

- Would count as official publication $\rightsquigarrow$ academic currency
  (compatibility with current incentive structures)
- Actual editorial process w/ peer-review $\rightsquigarrow$ guarantee of seriousness

How to lend "academic value" to refutations?

Proposal: a journal of critiques and refutations

- Would count as official publication $\rightsquigarrow$ academic currency
  (compatibility with current incentive structures)
- Actual editorial process w/ peer-review $\rightsquigarrow$ guarantee of seriousness

Several remaining questions (**feedback welcome!**)

- Scope? Question of "critical mass" $\Rightarrow$ no disciplinary boundaries at first?

How to lend "academic value" to refutations?

Proposal: a journal of critiques and refutations

- Would count as official publication $\rightsquigarrow$ academic currency
  (compatibility with current incentive structures)
- Actual editorial process w/ peer-review $\rightsquigarrow$ guarantee of seriousness

Several remaining questions (**feedback welcome!**)

- Scope? Question of "critical mass" $\Rightarrow$ no disciplinary boundaries at first?
  Severity of flaws? Level of expertise required for rebuttals?

How to lend "academic value" to refutations?

Proposal: a journal of critiques and refutations

- Would count as official publication $\rightsquigarrow$ academic currency
  (compatibility with current incentive structures)
- Actual editorial process w/ peer-review $\rightsquigarrow$ guarantee of seriousness

Several remaining questions (**feedback welcome!**)

- Scope? Question of "critical mass" $\Rightarrow$ no disciplinary boundaries at first?
  Severity of flaws? Level of expertise required for rebuttals?
- Anonymity option vs accountability?

Proposal: a journal of critiques and refutations

- Would count as official publication $\rightsquigarrow$ academic currency
  (compatibility with current incentive structures)
- Actual editorial process w/ peer-review $\rightsquigarrow$ guarantee of seriousness

Several remaining questions (**feedback welcome!**)

- Scope? Question of "critical mass" $\Rightarrow$ no disciplinary boundaries at first?
  Severity of flaws? Level of expertise required for rebuttals?
- Anonymity option vs accountability?