Reproducible Research Methodologies and New Publication Models

Panel discussion

Lieven Eeckhout

Ghent University, Belgium

HiPEAC – Adapt workshop

January 22, 2014

“Essentially, all models are wrong,
but some are useful”

George E. P. Box

Statistician (1919 – 2013)

"Essentially, all research papers are flawed, but some are useful"

It's all about proposing new ideas!

But we need experimental evaluation

- Support claims
- **Provide insight**
- NOT to provide the final performance figure
  - No single academic simulator is truly cycle-accurate!

BUT…

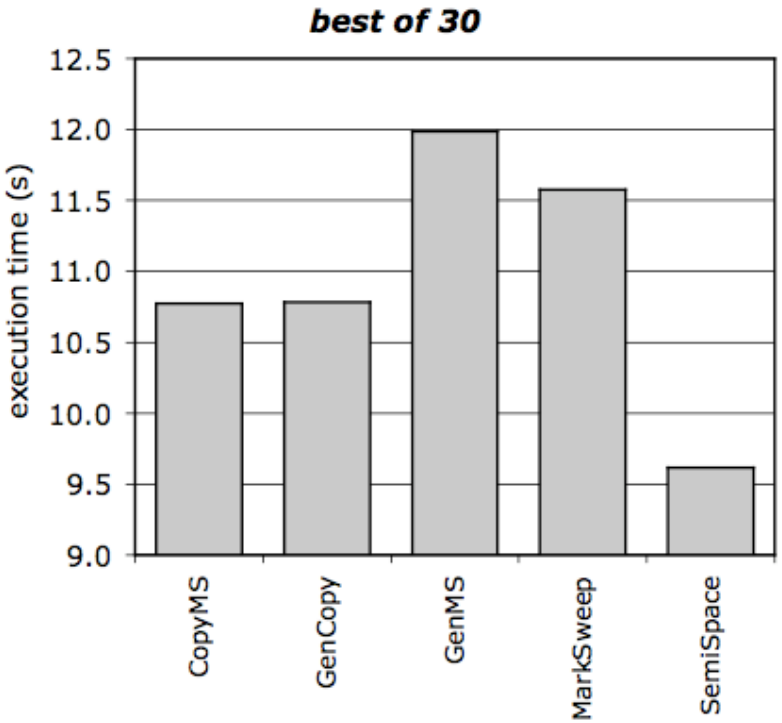
- This doesn't mean we shouldn't be careful about our experimental setup, methodology, metrics, and how we analyze the results!
  - On the contrary!!
- AND we should be even more careful when translating results into insights/claims/ conclusions!

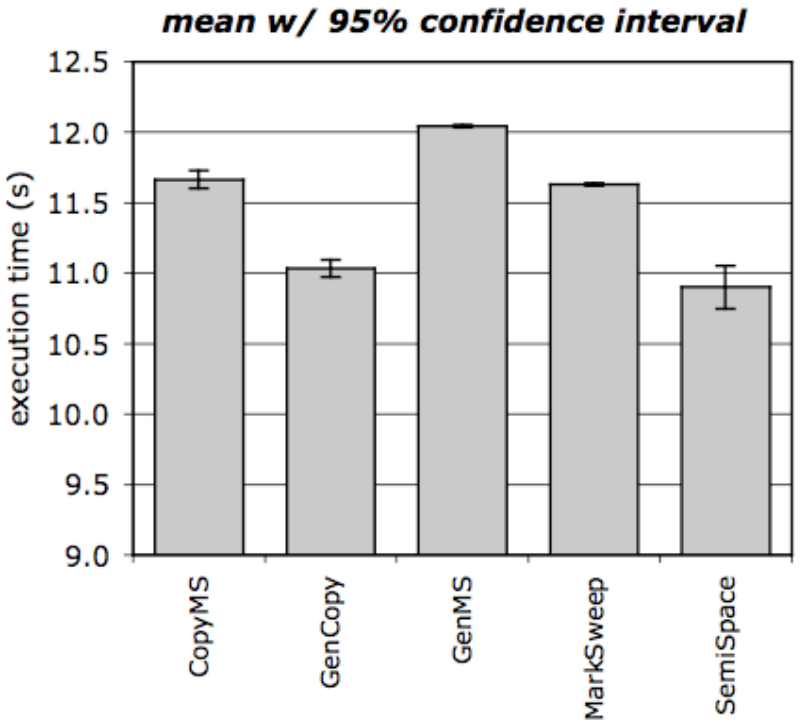
- But this is very difficult and error-prone!
  - If we mess it up, this may lead to incorrect conclusions, suboptimal designs, …

Very easy to screw up!

- Experimental design
  - Simulator/system configuration
    - Hundreds of parameters
  - Workloads: benchmarks, inputs, settings (start-up vs. steady-state; heap size), representative samples
- Data analysis
  - Appropriate metrics: multi-threaded / multi-program workloads; energy efficiency
  - Non-determinism

How to deal with non-determinism?

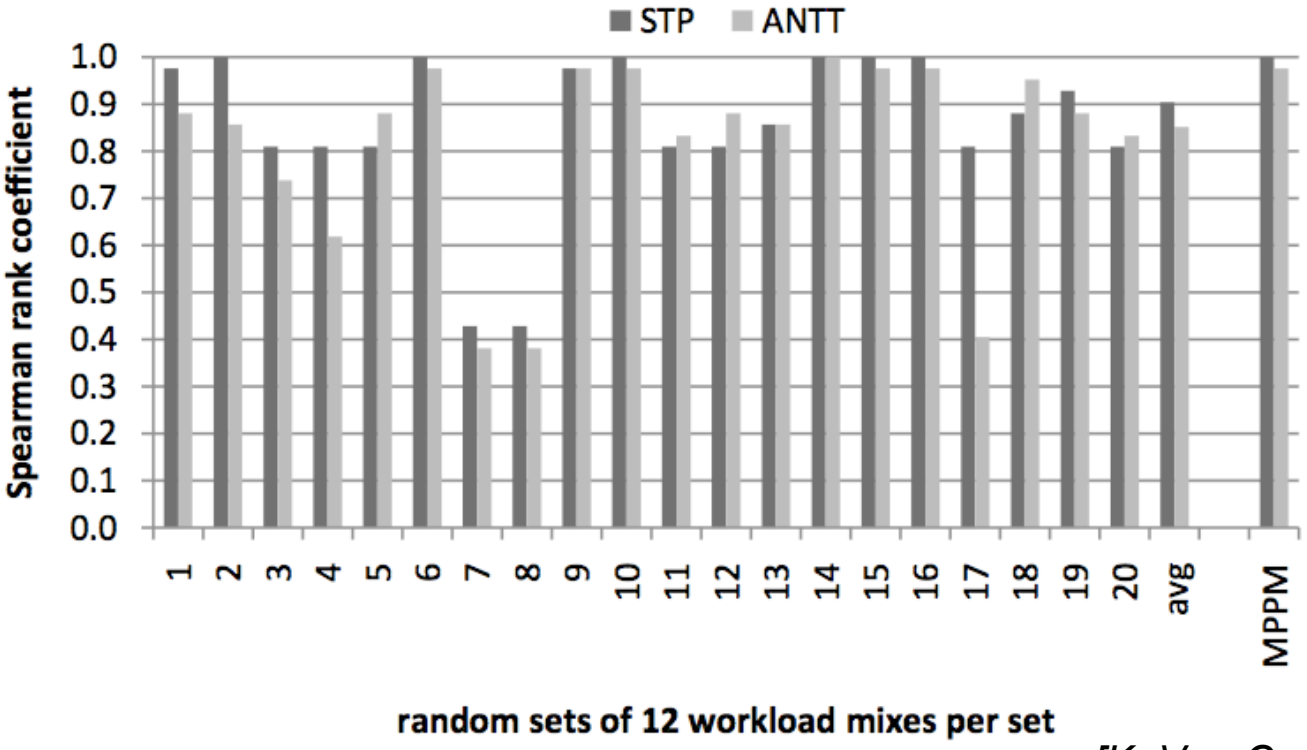




[A. Georges et al., OOPSLA 2007]

Multi-core workload selection

	size	*assoc*	*latency*
config #1	512KB	8	16
config #2	512KB	16	20
config #3	1MB	8	18
config #4	1MB	16	22
config #5	2MB	8	20
config #6	2MB	16	24



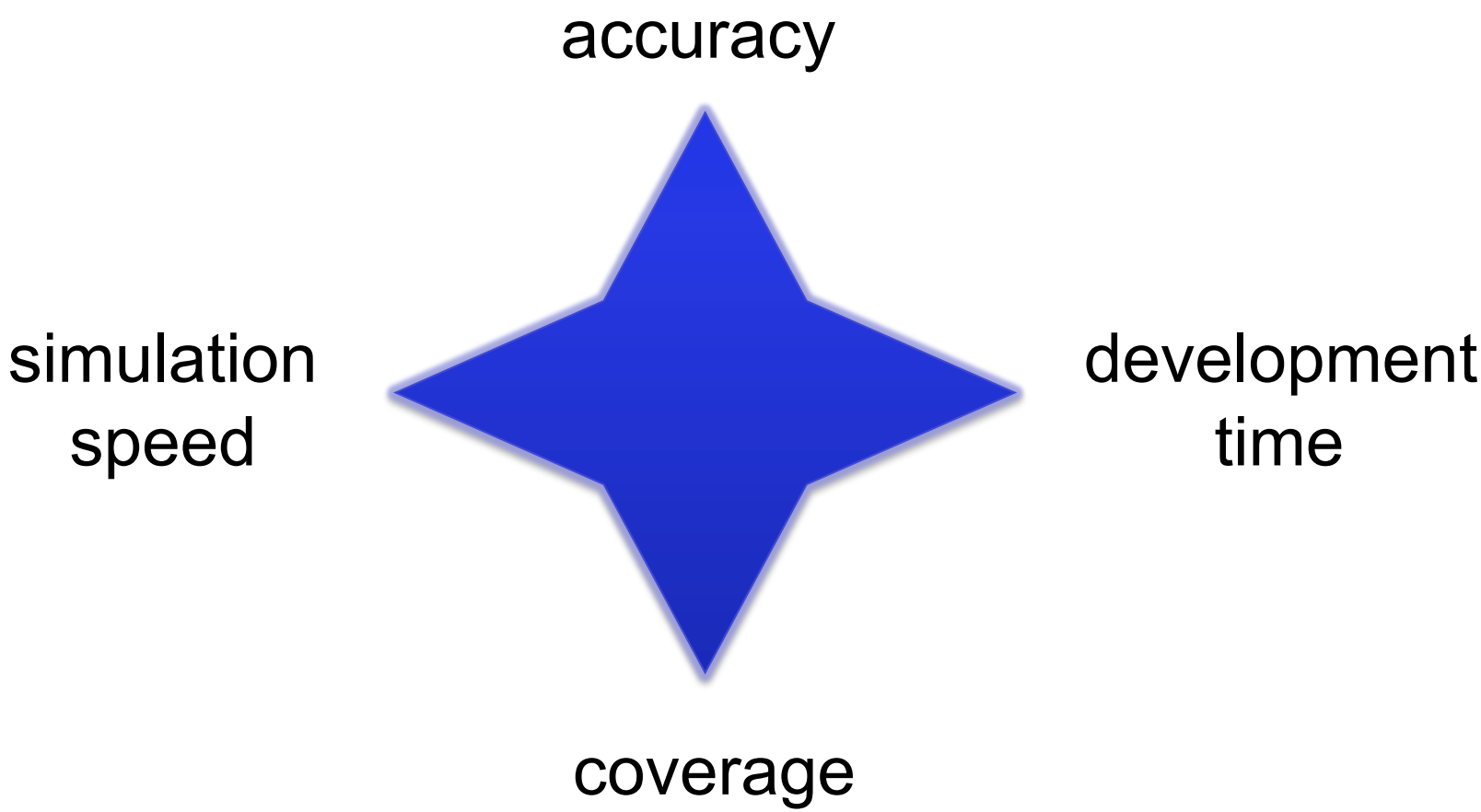
[K. Van Craeynest et al., IISWC 2011]

People have been advocating for a common platform

- Arguments:
  - Reproducibility of research results
  - Leverage community effort

- I don't quite believe in this approach
  - Unified platform would involve too much overhead in the interfaces
  - No single tool can potentially serve all needs

Building/picking a simulation environment is a trade-off



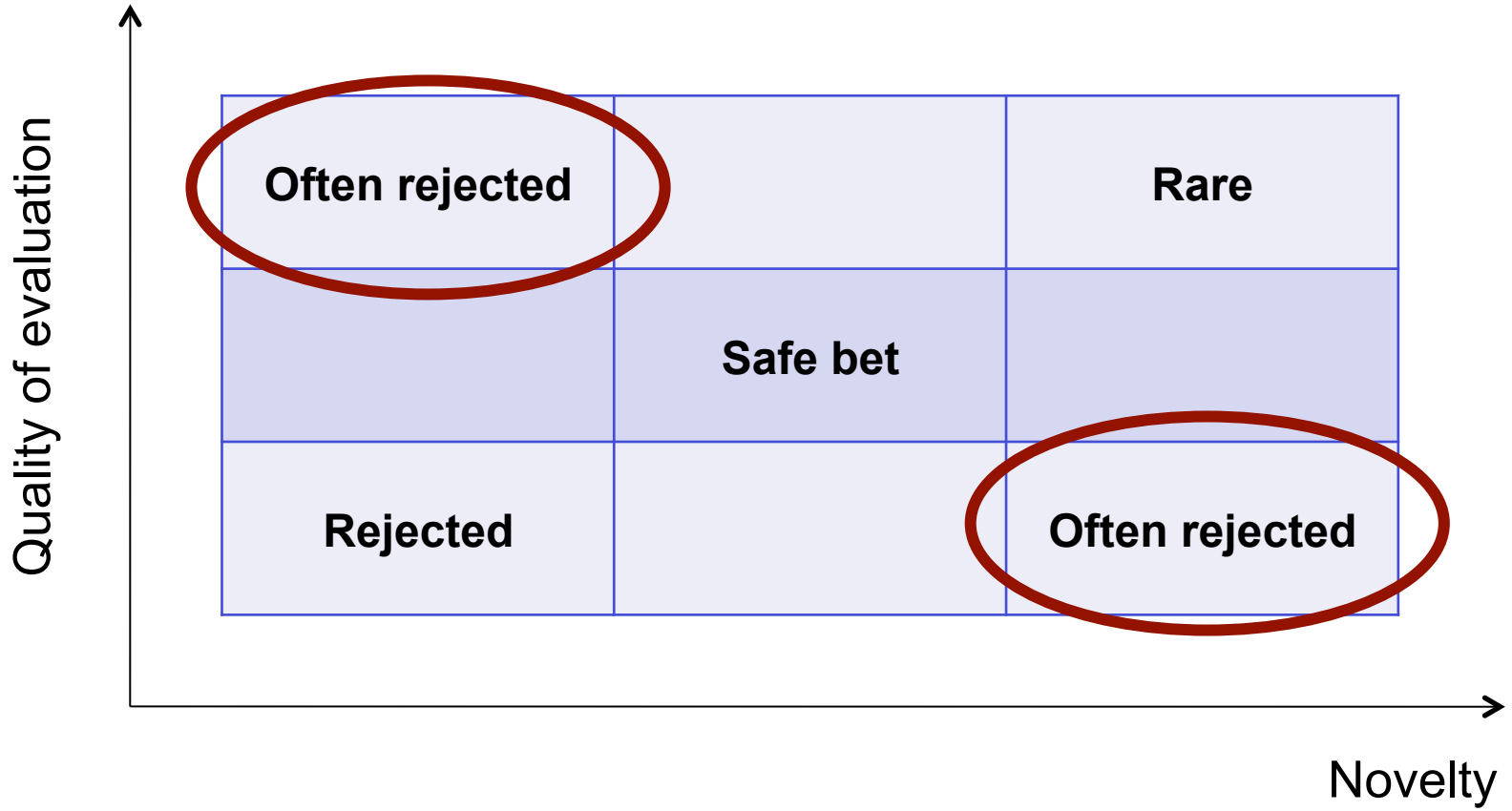
Encouraging steps towards reproducibility

- OOPSLA & ECOOP 2013
- Artifact should be
  - **consistent** with the paper,
  - as **complete** as possible,
  - **well documented**, and
  - **easy to reuse**, facilitating further research.

- Publicly-released tools
  - Community should be a little more receptive to tools papers in our top venues

Which papers get accepted?

As a community, we should start accepting more low-novelty, high-evaluation and high-novelty, low-evaluation papers



[P. Sweeney, A. Diwan, S. Blackburn, M. Hauswirth]

Publication models

conference versus journal(-first)

- Highly-selective conferences
  - ISCA, MICRO, HPCA, ASPLOS, PLDI, OOPSLA
  - Very rewarding to publish in
- Managing randomness
  - Double-blind review
  - 5 to 6 reviews per paper
  - Physical PC meeting
    - Lively discussions
    - Set a common ‘bar’ for acceptance
  - Striving to reach consensus

Final thoughts

- Remember we're producing results to gain insight!
- Be careful wrt experimental design, data analysis, and translating results into conclusions/claims
  - Use the appropriate tool and setup for the job

- For the community:
  - Be open-minded to high-novelty, low-evaluation and low-novelty, high-evaluation papers
  - Better reward tools papers
  - Keep on improving the review process