

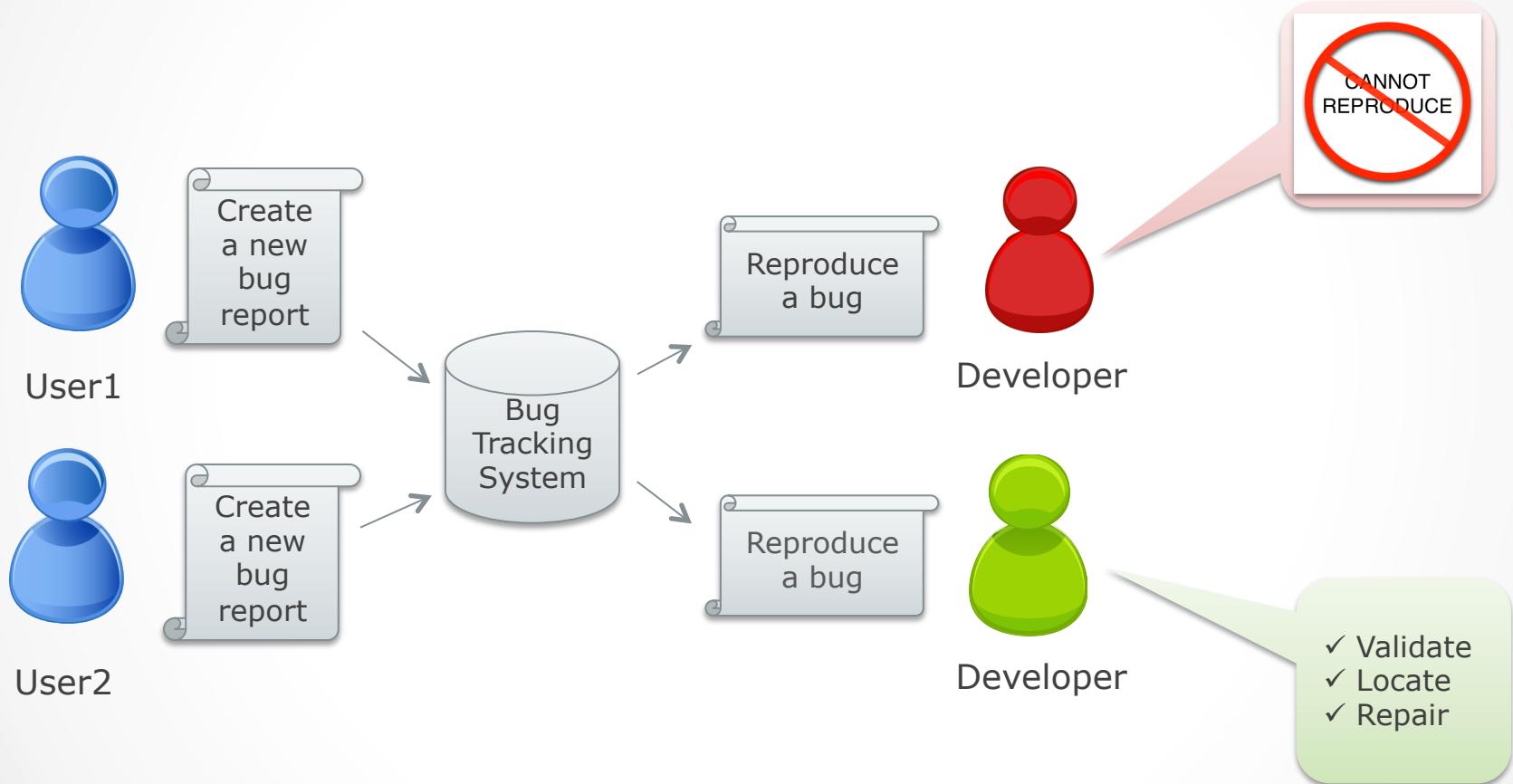
Works For Me!

Characterizing Non-reproducible Bug Reports

Mona Erfani Joorabchi, Mehdi Mirzaaghaei, Ali Mesbah



Bug report should help developers to find/fix a software fault.



Motivation

Security Vulnerability Facebook

"Unfortunately your report [...] did not have enough technical information for us to take action on it. We cannot respond to reports which do not contain enough detail to allow us to reproduce an issue."

The screenshot shows Mark Zuckerberg's Facebook profile page at <https://www.facebook.com/zuck>. The profile picture is a smiling Mark Zuckerberg. The main post is a link shared by Khalil, with a timestamp of "about a minute ago". The message reads:

Dear Mark Zuckerberg,

First sorry for breaking your privacy and post to your wall , i has no other choice to make after all the reports i sent to Facebook team .

My name is KHALIL, from Palestine .

See More

A green circle highlights the comment from Khalil.



No studies on Non-Reproducible (NR) bugs

NR Definition



Other Terminologies for NR:

Cannot Reproduce [Jira], Works For Me [Bugzilla], Works on My Machine

Definition 1 A Non-Reproducible (NR) bug is one that cannot be reproduced **based on the information** provided in the bug report. All attempts at reproducing the issue have been **futile**, and reading the **system's code** provides no clues as to **why** the described behavior would occur.

Four Research Questions

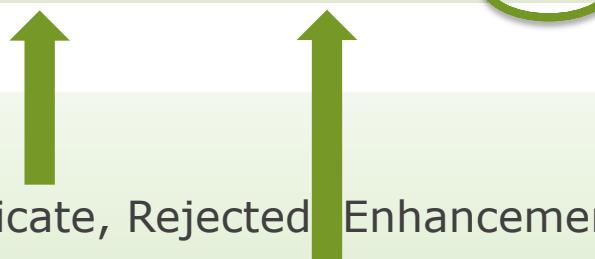
- ① How **prevalent** are NR bug reports? Are NR bug reports **treated** differently than other bug reports?
- ② Why can NR bug reports not be reproduced? What are the most common **cause categories**?
- ③ Which **resolution transition patterns** are common in NR bug reports?
- ④ What portion of NR bug reports is **fixed** eventually? Were they mislabeled initially? What cause categories do they belong to?

Bug Repository Selection

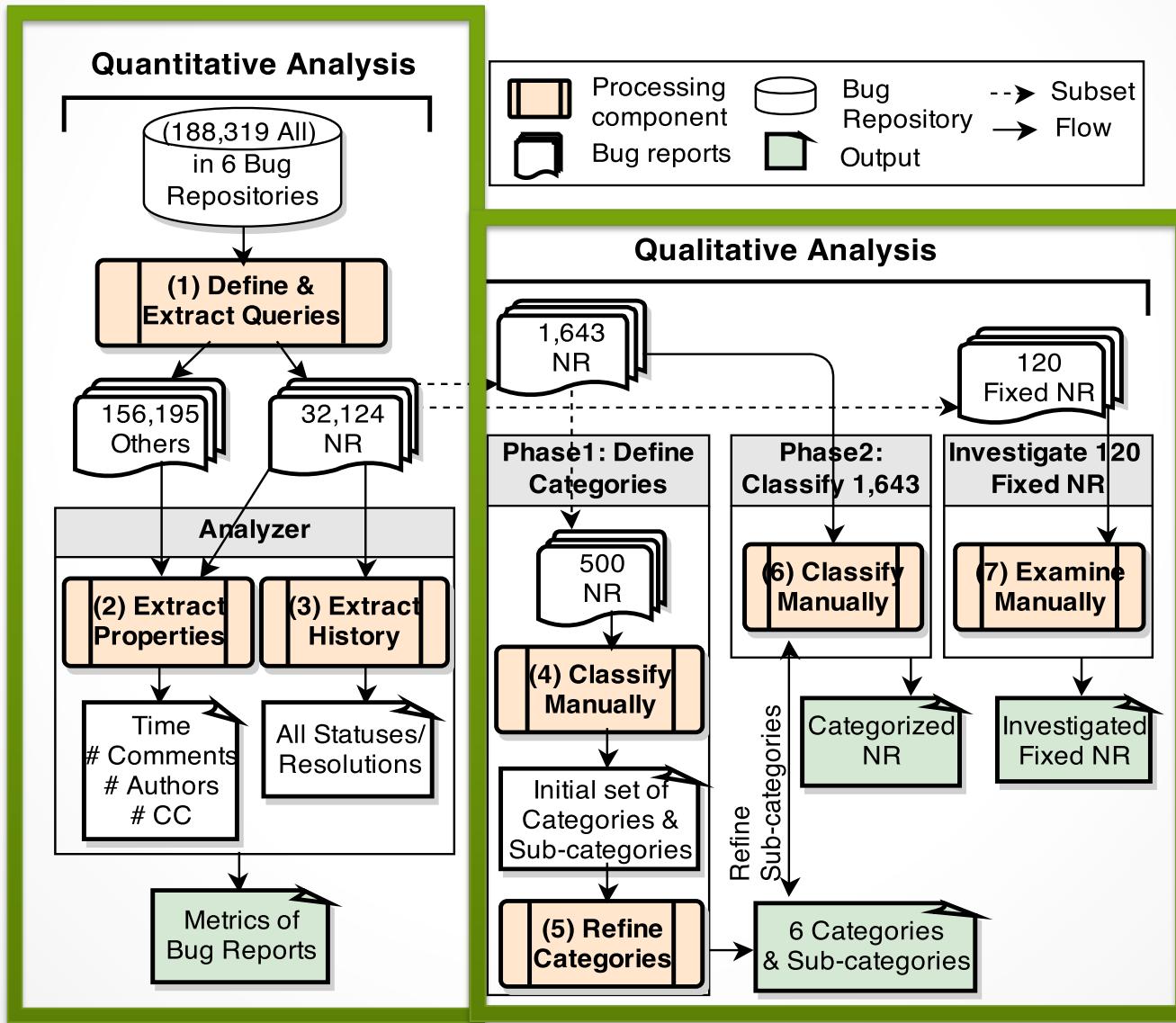
6 bug repositories

Product/Component	#All Bugs	#NR Bugs	NR (%)	FixedNR (%)
Firefox	65,408	18,516	28%	1%
Eclipse/Platform	65,475	8,180	13%	4%
MediaWiki	9,335	1,125	12%	9%
Moodle	22,175	2,503	11%	5%
FirefoxAndroid	7,902	1,148	15%	3%
Proprietary	18,024	643	4%	17%
	188,319	32,124	17%	3%

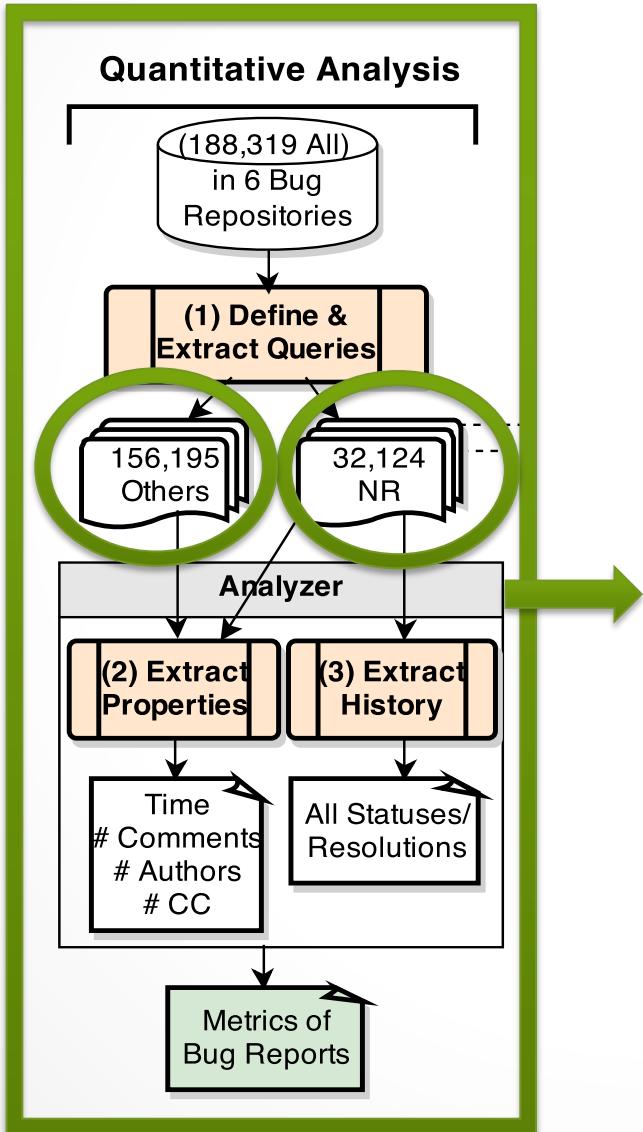
Filtering steps:

- 
- ① Exclude Invalid, Duplicate, Rejected Enhancement, Feedback, Unconfirmed
 - ② Apply **Changed to/from** operator on resolution field (bug reports that **resolved as NR at least once in their life-cycles.**)
 - ③ Extract data in XML format

Methodology: Mixed-methods Approach



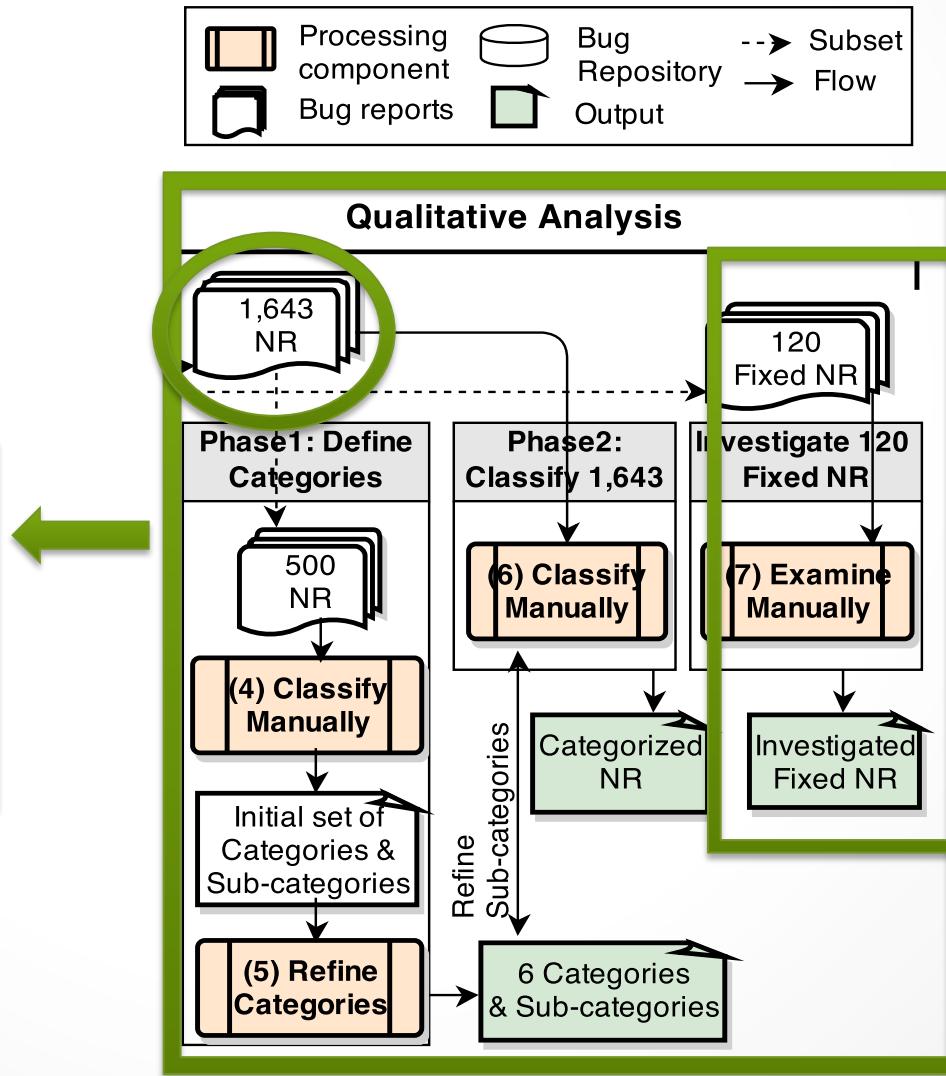
Methodology: Mixed-methods Approach



NR-Bug-Analyzer tool:

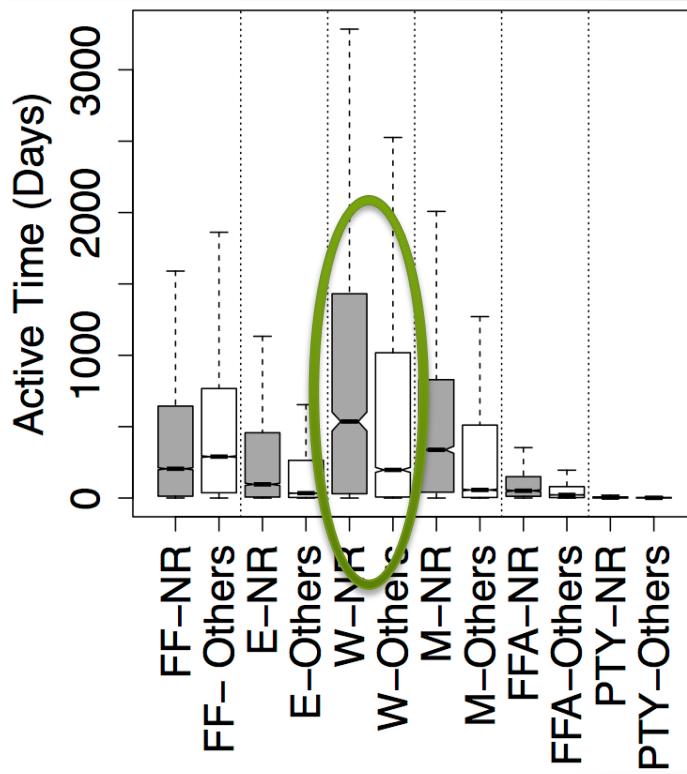
- ✓ Active Time (Creation – Last Update)
- ✓ Number of Unique Authors
- ✓ Number of Comments
- ✓ Number of CCs/Watchers
- ✓ Historical Status/Resolution Changes
(parses HTML source of each NR bug)

Methodology: Mixed-methods Approach



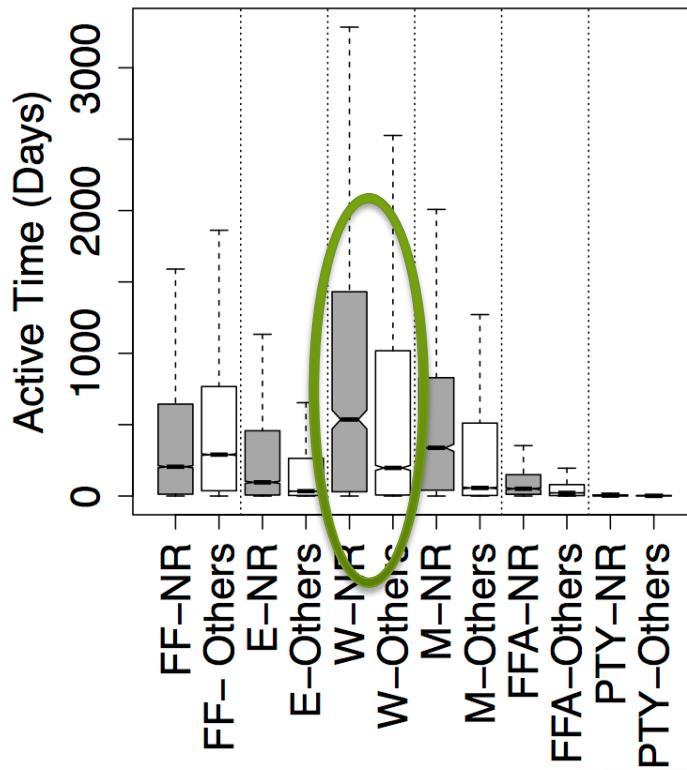
- ✓ Randomly selected
- ✓ Bug's description, tester/developer discussions/comments, historical data
- ✓ Infer 6 NR categories
- ✓ Investigate Fixed NR

RQ1: NR vs Others



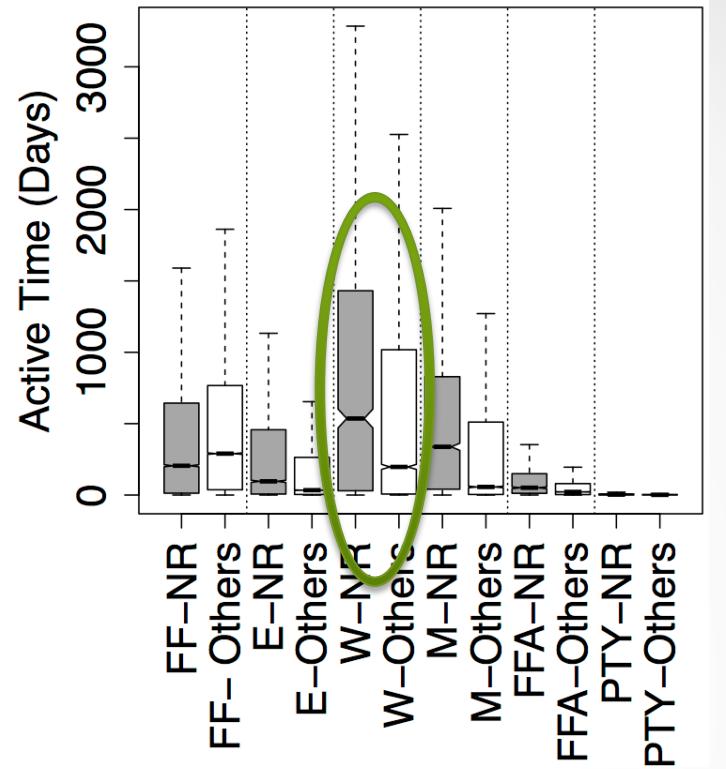
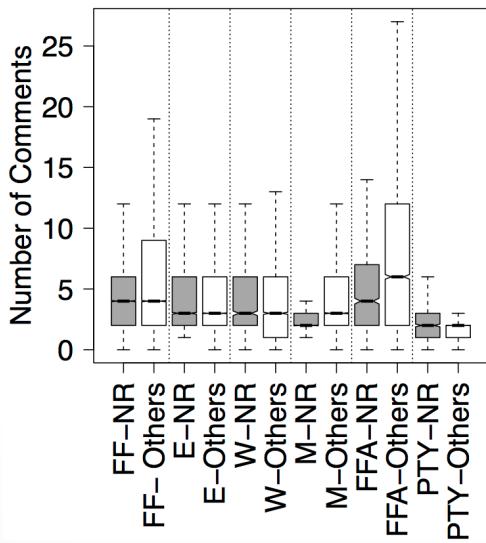
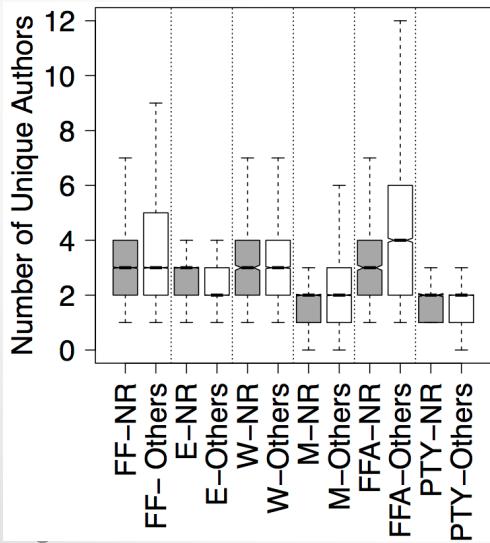
RQ1: NR vs Others

- ✓ Active 3 months longer



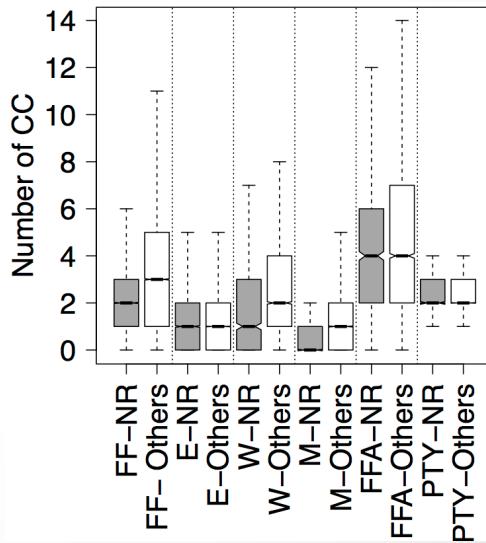
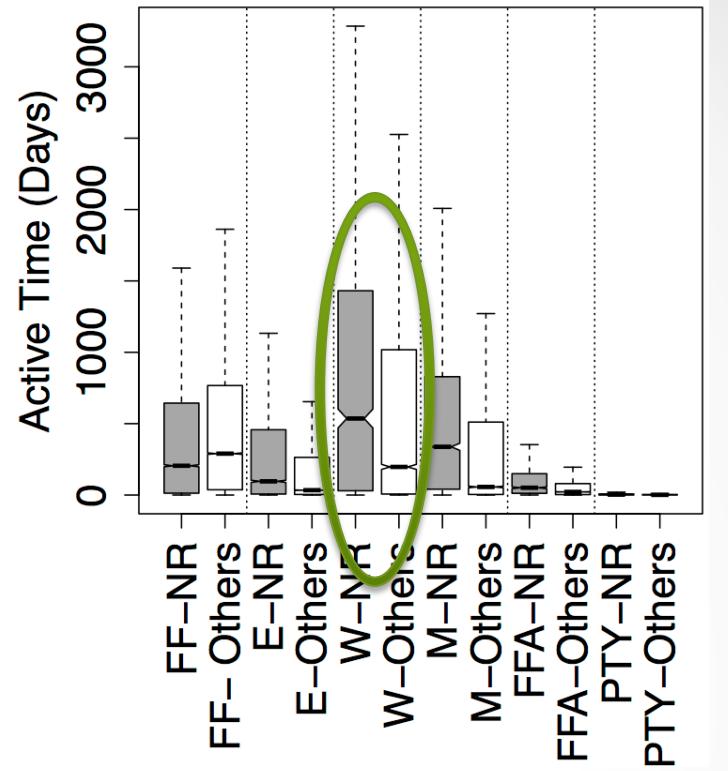
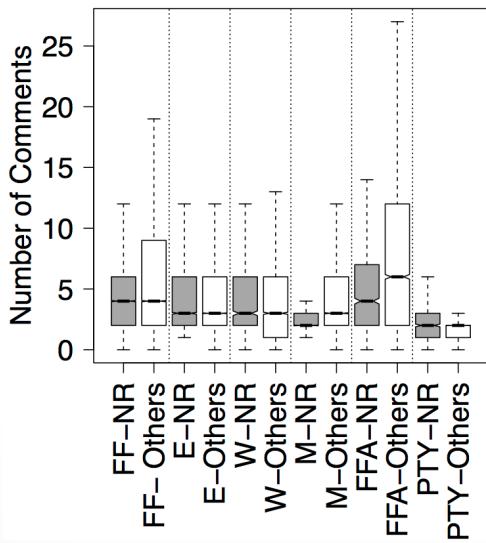
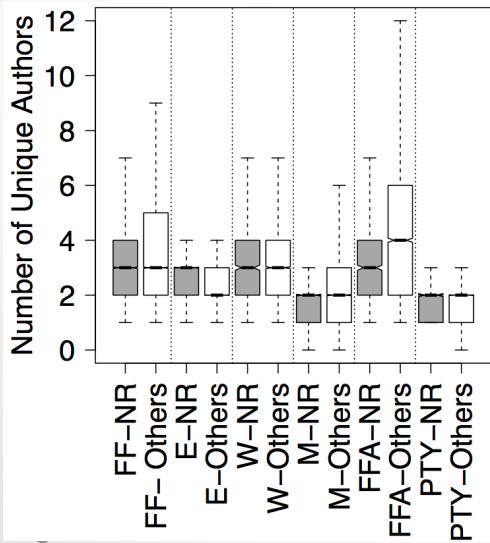
RQ1: NR vs Others

- ✓ Active 3 months longer

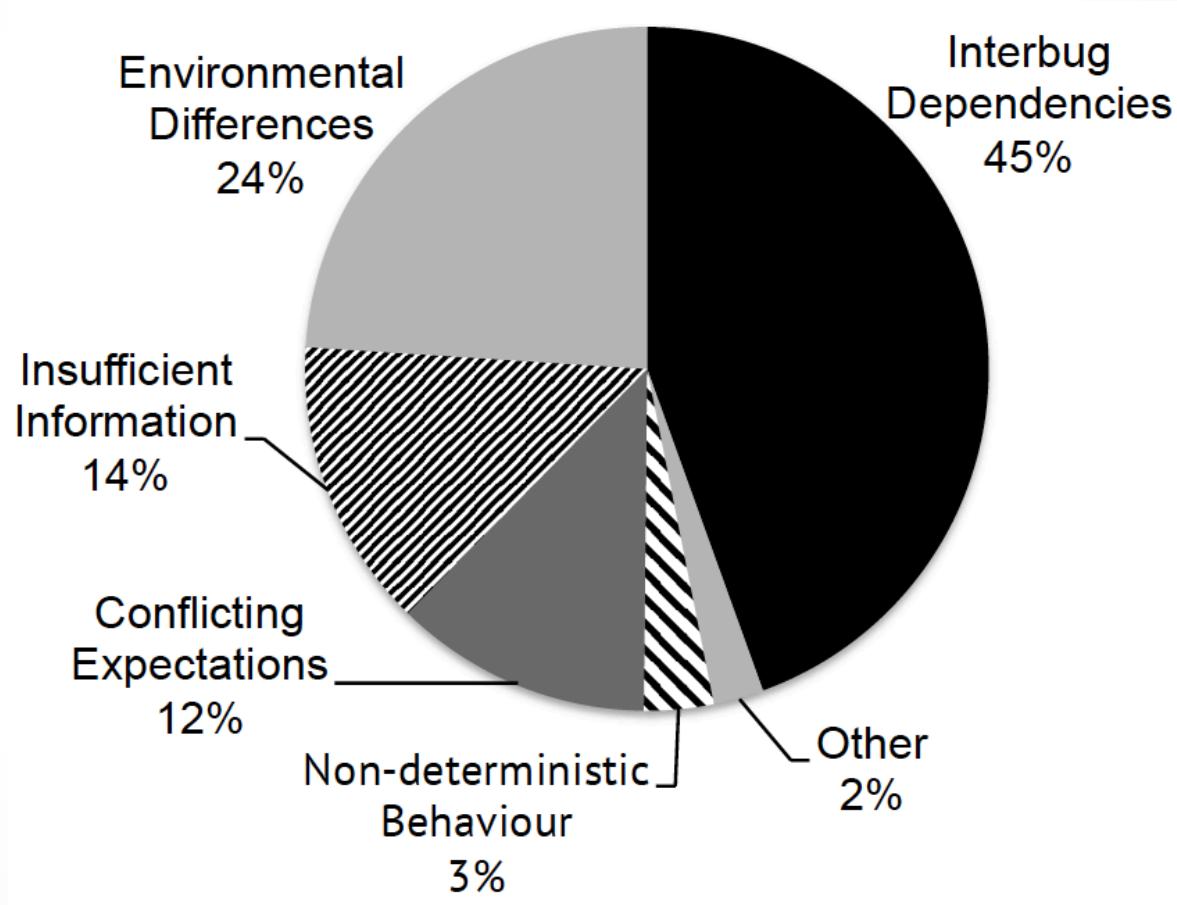


RQ1: NR vs Others

- ✓ Active 3 months longer
- ✓ Require as much attention as the rest

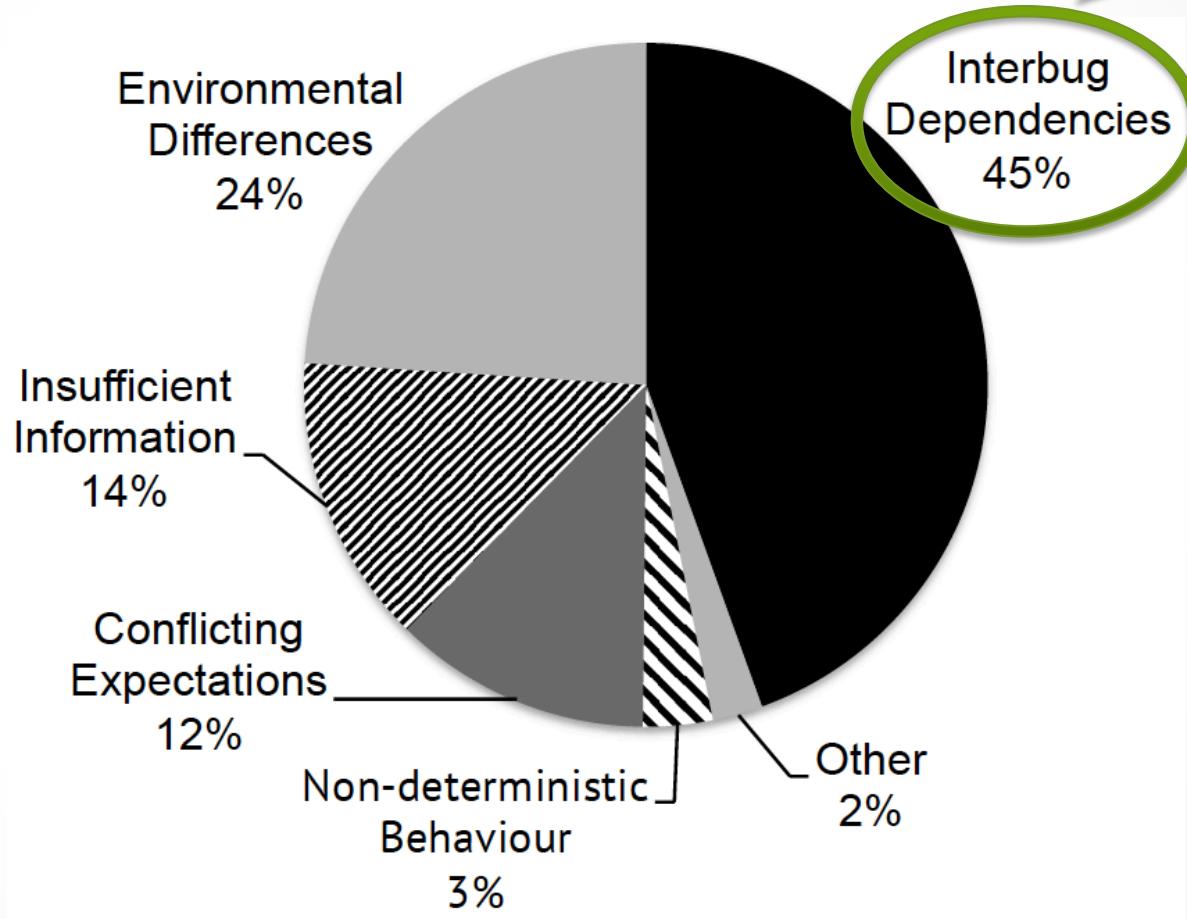


RQ2: Cause Categories

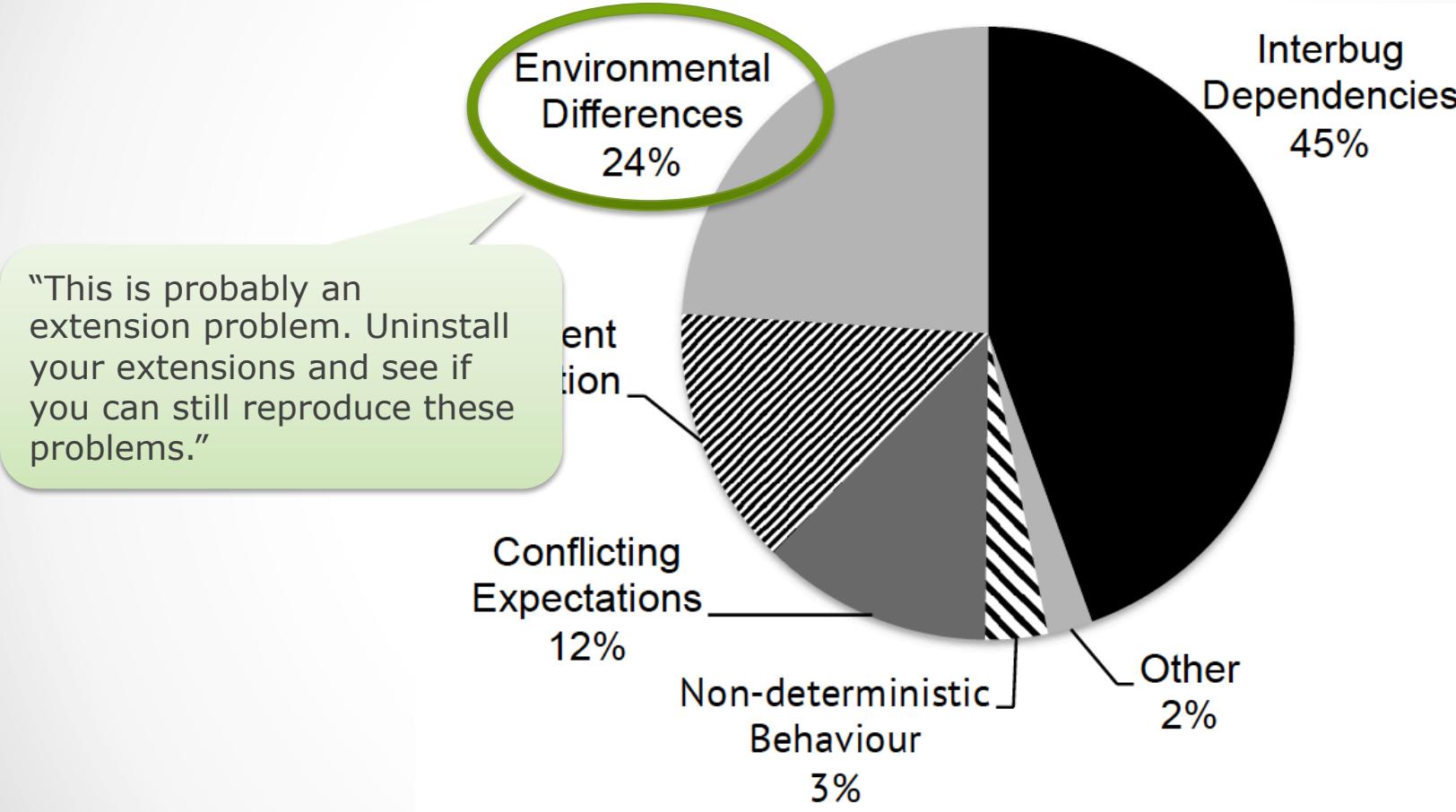


"It is now working with Firefox 15.0.1. I believe it was fixed by the patches to #780543 and #788600 [...]."

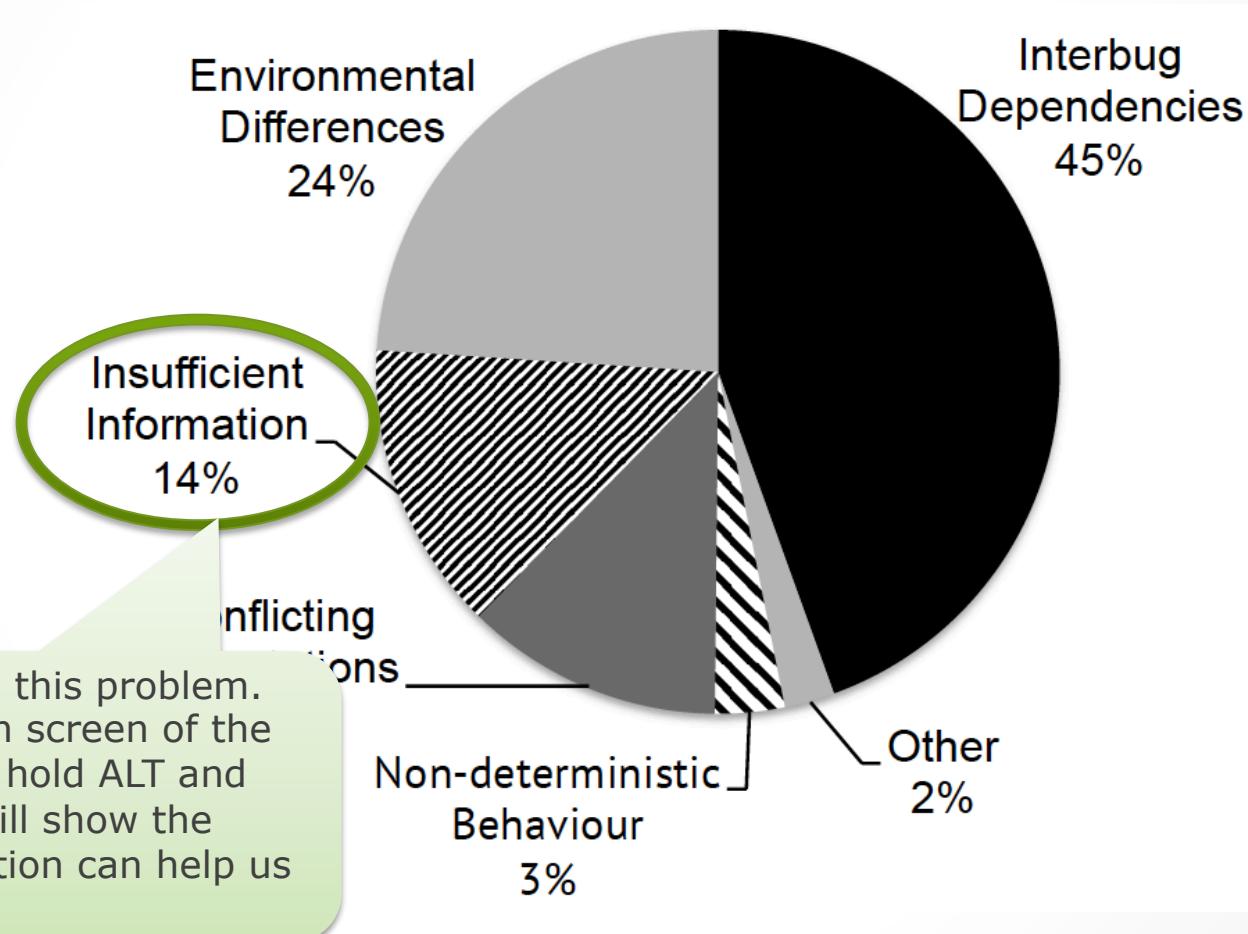
RQ2: Cause Categories



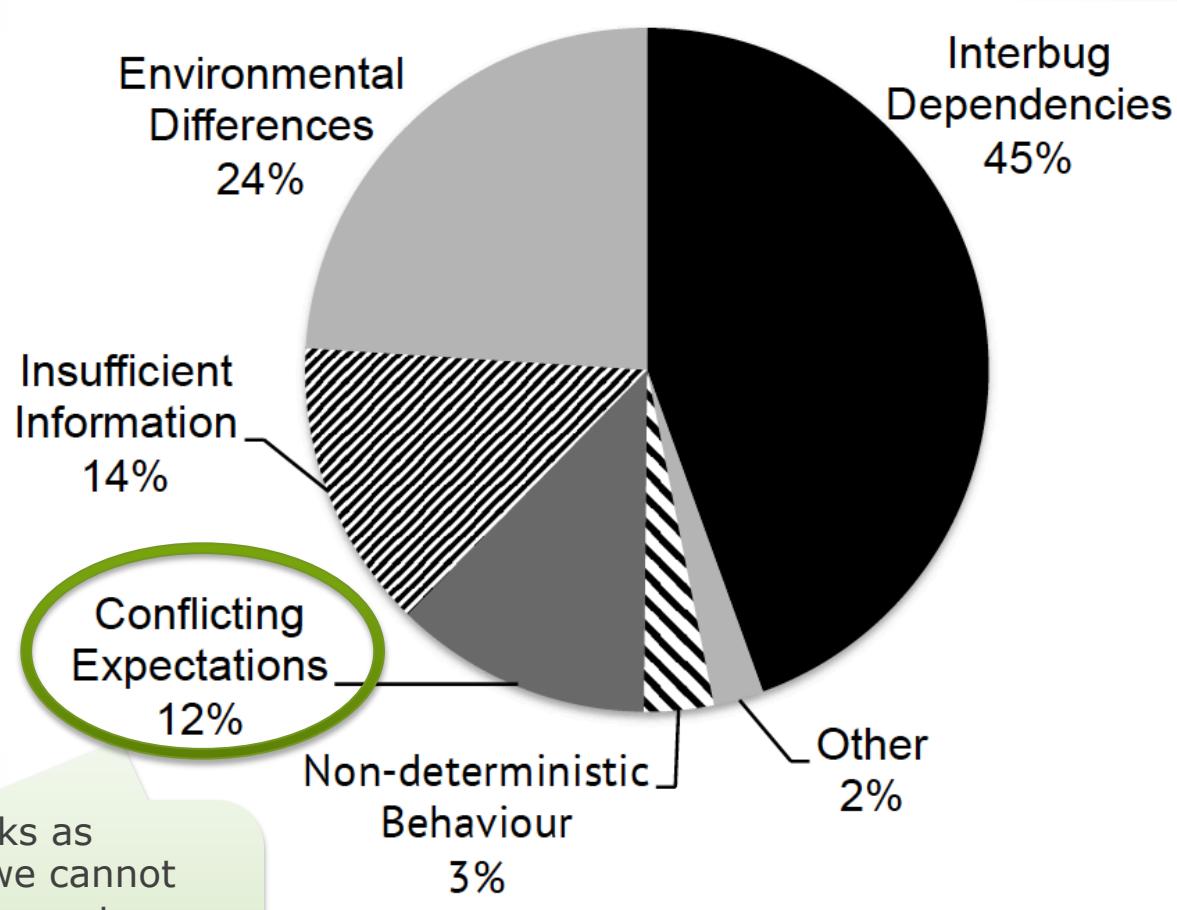
RQ2: Cause Categories



RQ2: Cause Categories

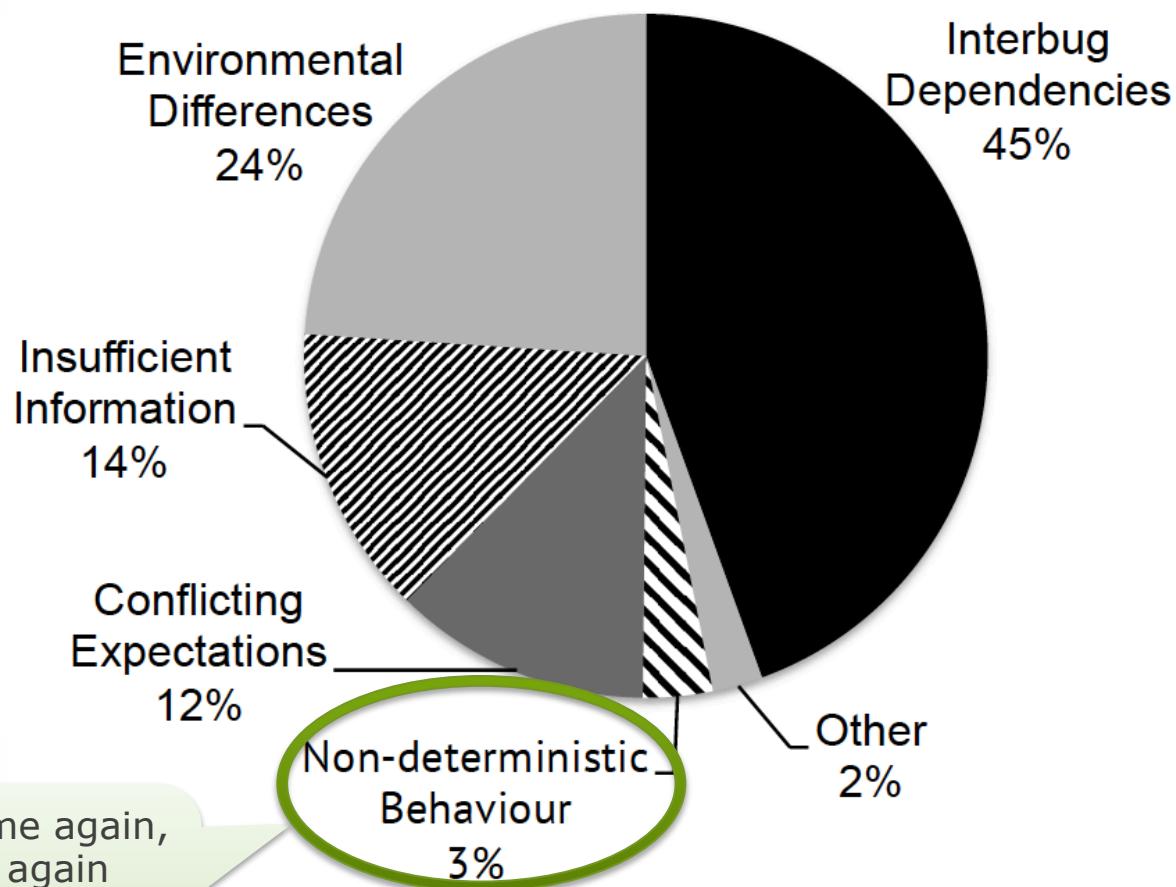


RQ2: Cause Categories



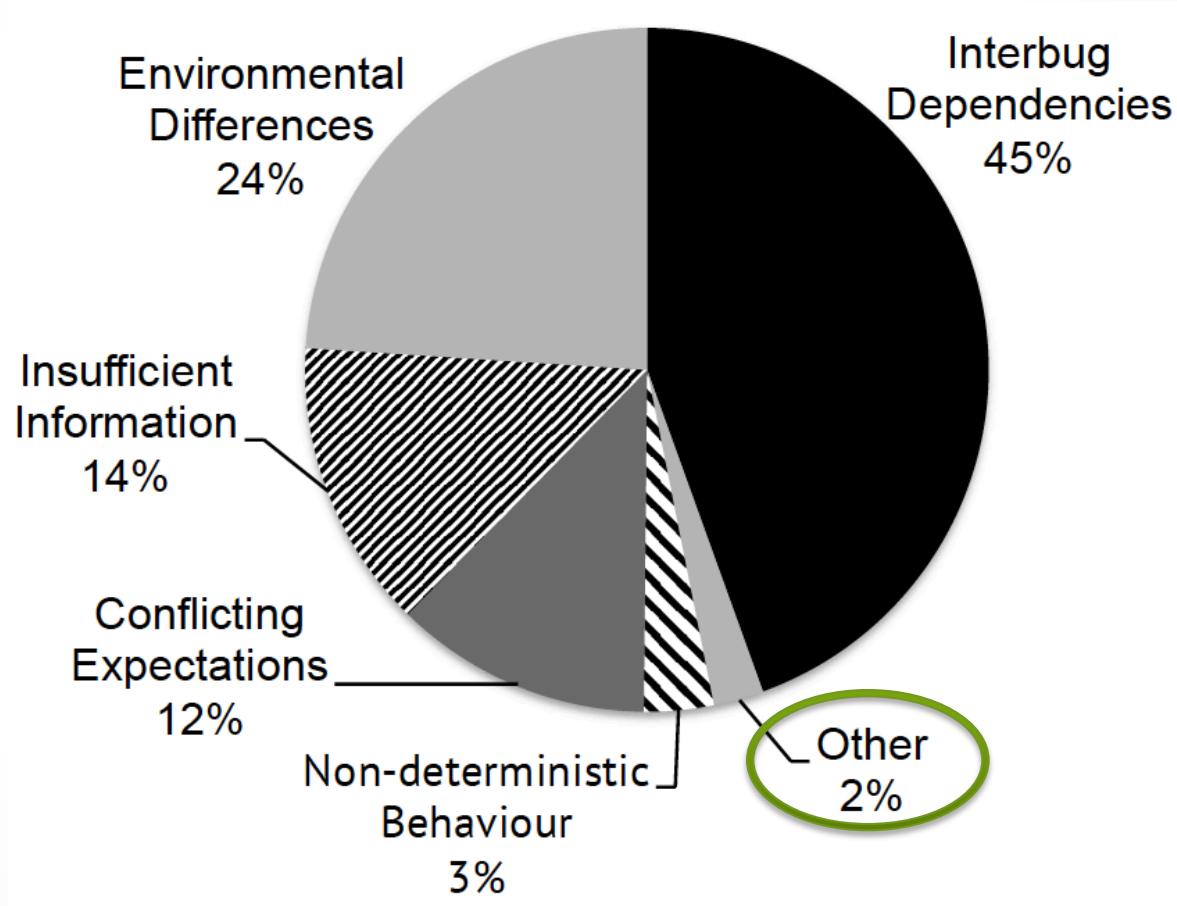
D: "PDE Schema works as designed [...] Since we cannot tell when you want to use tags and when you want to use reserved chars as-is, you need to [...]."

RQ2: Cause Categories



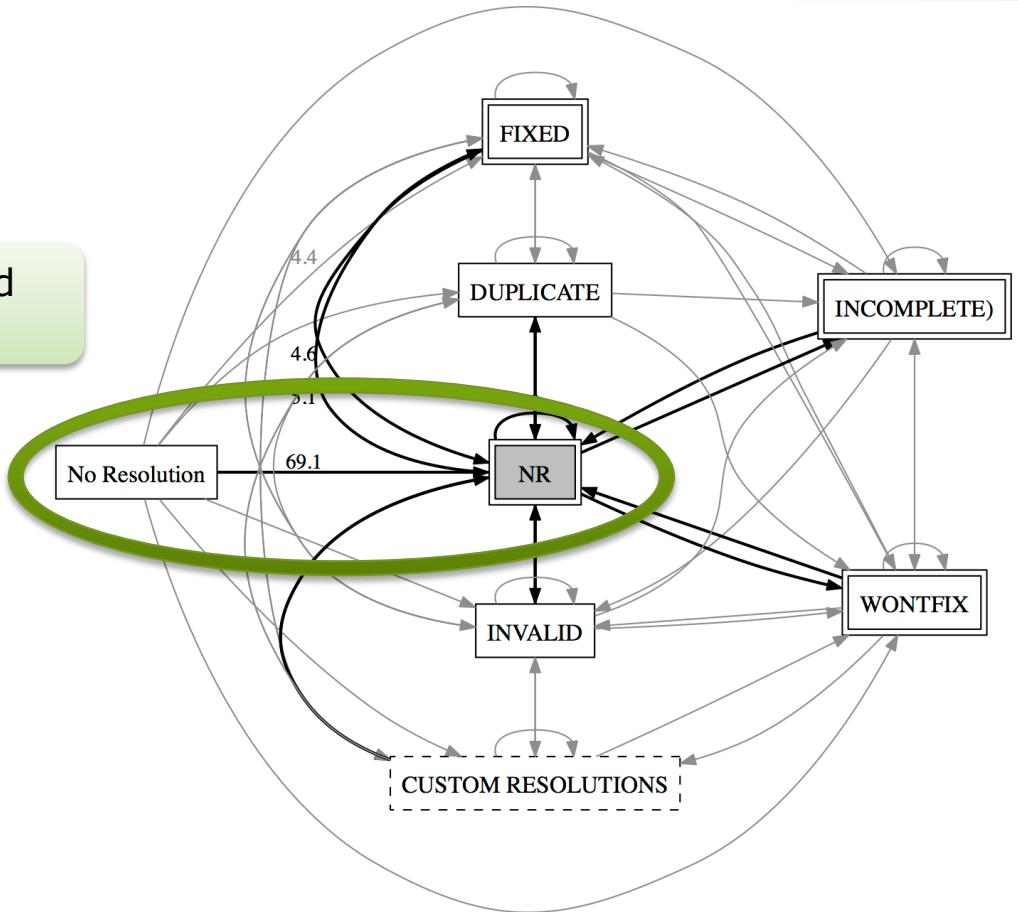
"This happened for me again, and then went away again (started working). It seems there is an intermittent problem."

RQ2: Cause Categories



RQ3: Common Transition Patterns

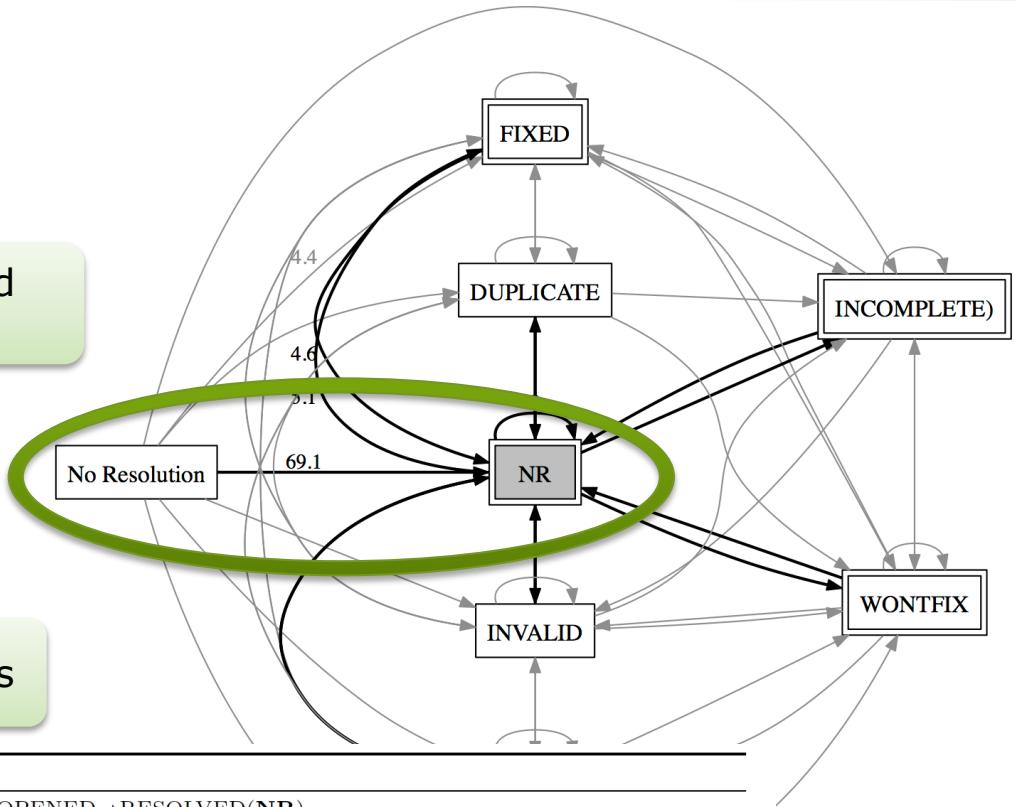
68% of NR resolved
(New/Open) -> NR



RQ3: Common Transition Patterns

68% of NR resolved
(New/Open) -> NR

32% exhibit many resolution transitions



| Status (Resolution)

- 1 NEW → RESOLVED(NR) → REOPENED → RESOLVED(NR) → REOPENED → RESOLVED(NR)
- 2 NEW → RESOLVED(NR) → REOPENED → ASSIGNED → RESOLVED(FIXED) → REOPENED → RESOLVED(FIXED)
- 3 NEW → RESOLVED(FIXED) → REOPENED → RESOLVED(WONTFIX) → RESOLVED(NR)
- 4 NEW → RESOLVED(FIXED) → REOPENED → RESOLVED(NR) → REOPENED → NEW → RESOLVED(WONTFIX)
- 5 NEW → RESOLVED(FIXED) → REOPENED → RESOLVED(NR) → REOPENED → RESOLVED(FIXED)
- 6 UNCONFIRMED → NEW → RESOLVED(FIXED) → REOPENED → RESOLVED(FIXED) → REOPENED → RESOLVED(INVALID) → REOPENED → RESOLVED(NR) → REOPENED → RESOLVED(FIXED)
- 7 NEW → ASSIGNED → NEW → RESOLVED(NR) → REOPENED → ASSIGNED → RESOLVED(FIXED) → VERIFIED
- 8 NEW → ASSIGNED → RESOLVED(NR) → REOPENED → ASSIGNED → RESOLVED(LATER) → REOPENED → NEW → ASSIGNED → NEW → ASSIGNED → RESOLVED(FIXED)
- 9 UNCONFIRMED → RESOLVED(INCOMPLETE) → UNCONFIRMED → RESOLVED(INCOMPLETE) → RESOLVED(FIXED) → RESOLVED(NR)

RQ4: Fixed NR Bugs

NR = **17%** of all bug reports

Fixed NR = **3%** marked Fixed + **45%** implicitly fixed (~Half of NR **mislabeled**)

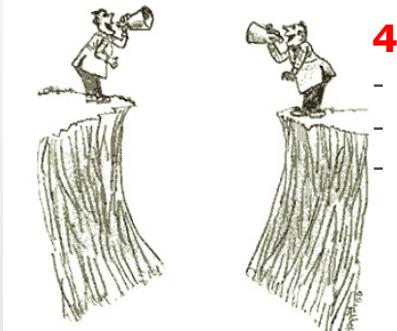
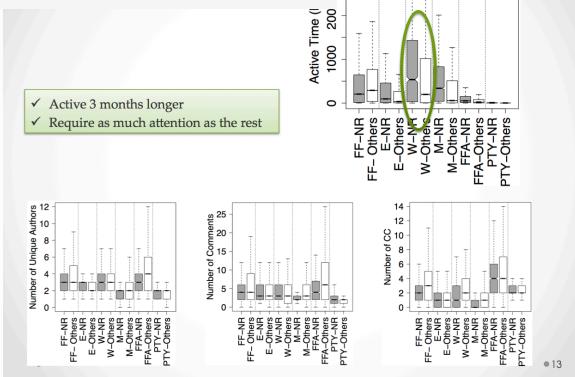
Mislabeling?!

- Inconsistencies around the usage of **Fixed** & **NR** resolutions.
- NR resolution ~ **irrelevant**, **unimportant**, **fixed** (different from formal definition)
- Available resolutions do not cover all possible scenarios

Fixed NR: “**Insufficient Information**”, “**Environmental Differences**”, “**Conflicting Expectations**”

Conclusion

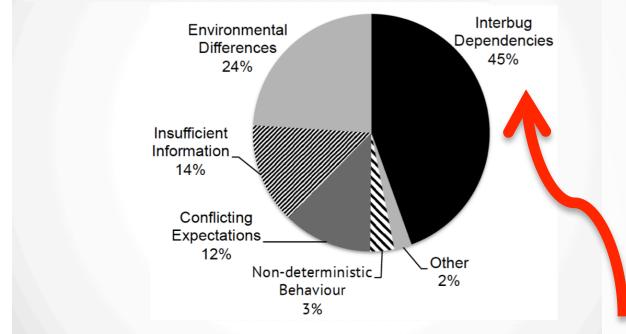
1) Quantitative Study



4) Communication Issues

- “Insufficient Information”
 - “Conflicting Expectations”
 - “Environmental Differences”

2) Qualitative Study



3) Mislabeling

5) NR Resolution Fan in/out

