

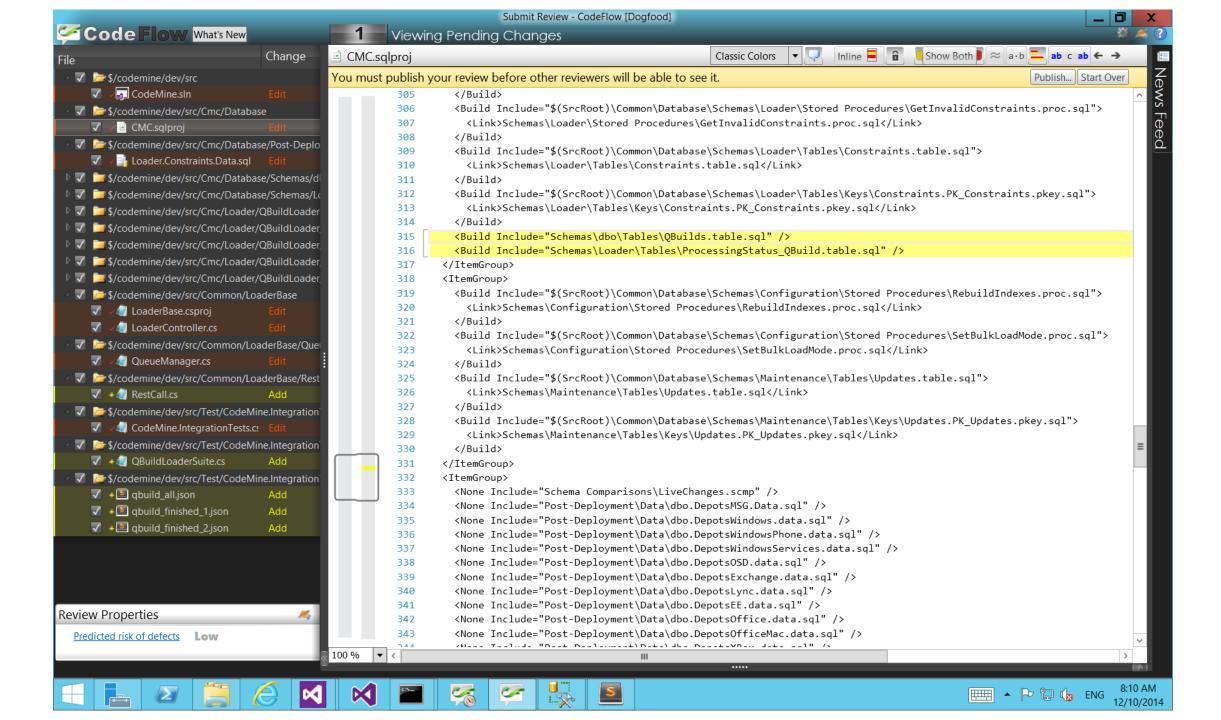


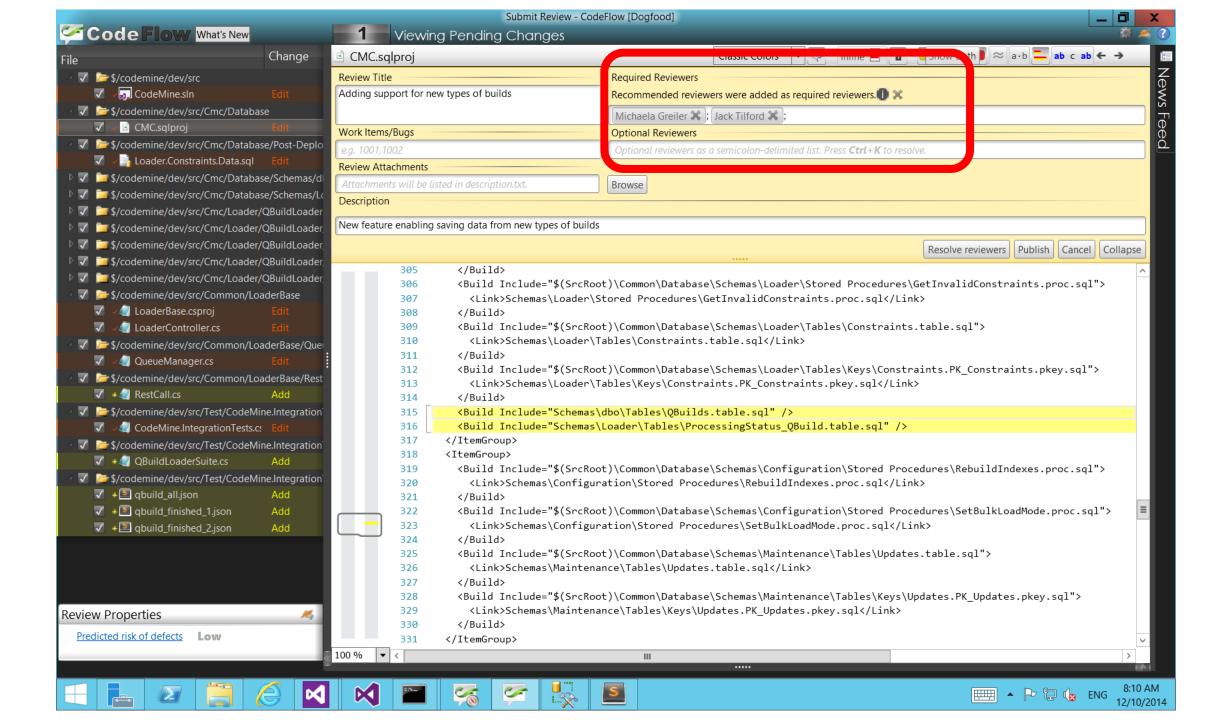
## Code Reviews Do Not Find Bugs

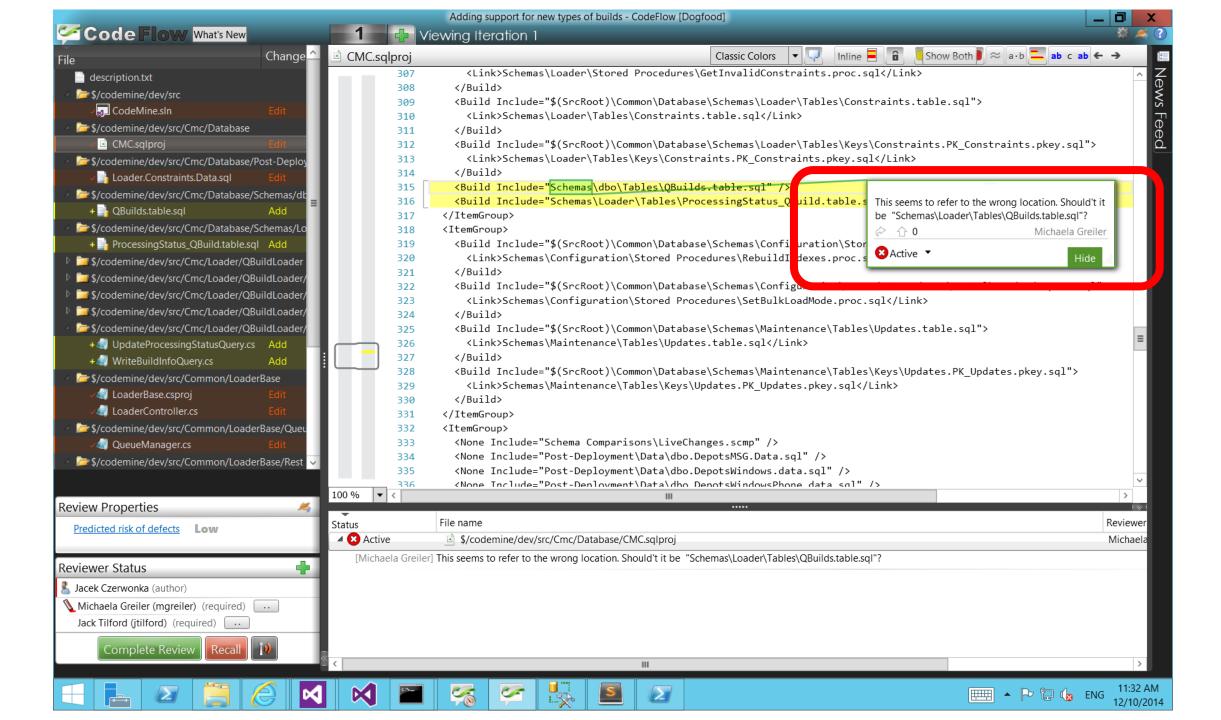
How the Current Code Review Best Practice Slows Us Down

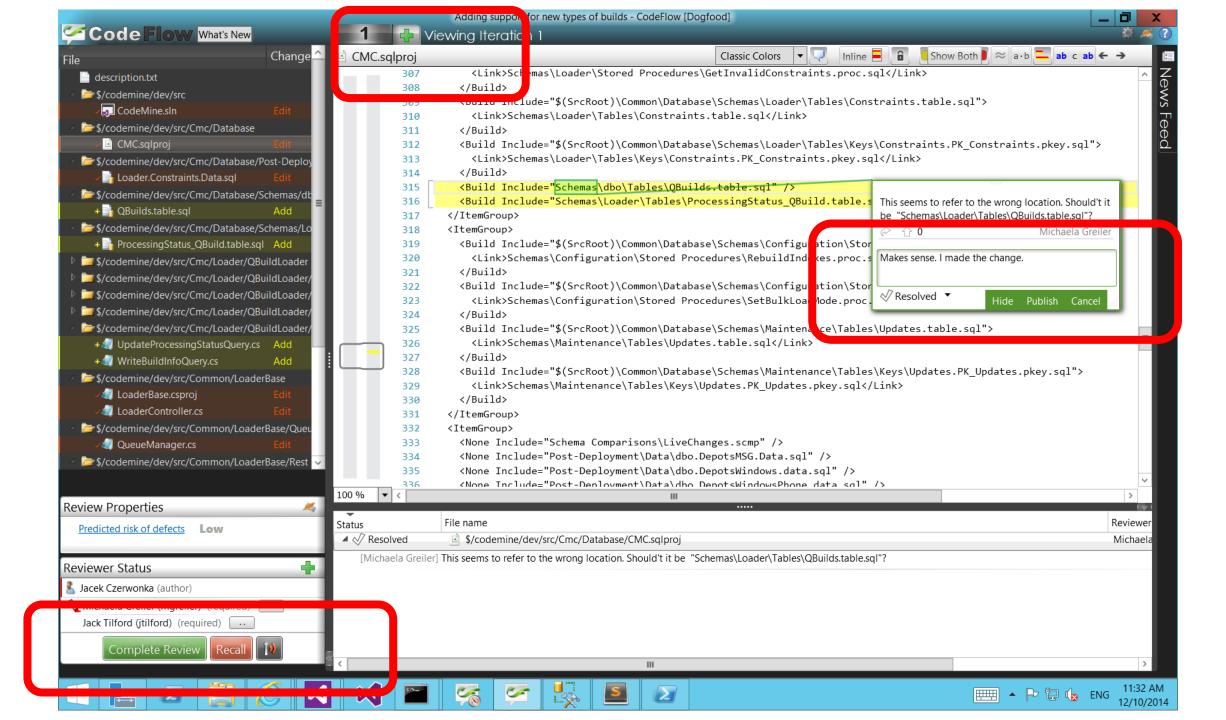
<u>Jacek Czerwonka</u>, Michaela Greiler, Jack Tilford { jacekcz, mgreiler, jtilford } @ microsoft.com
Tools for Software Engineers team, Microsoft Corp.

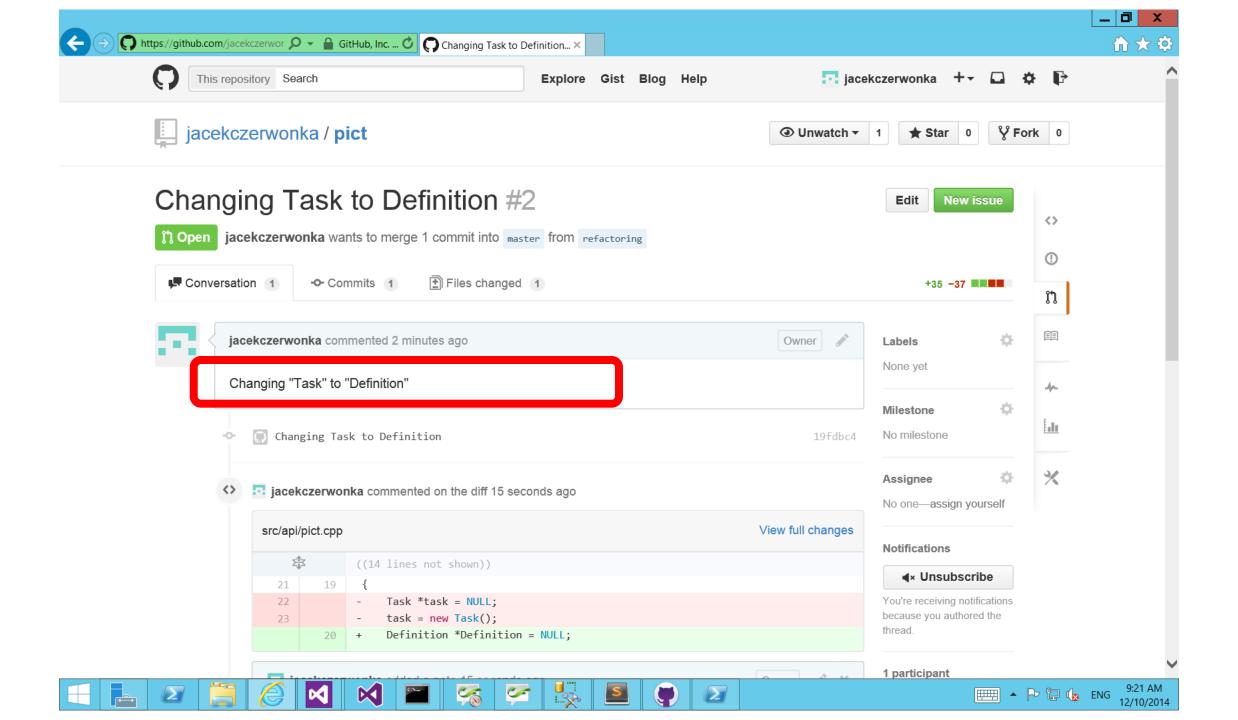
(in collaboration with Christian Bird, Microsoft Research)

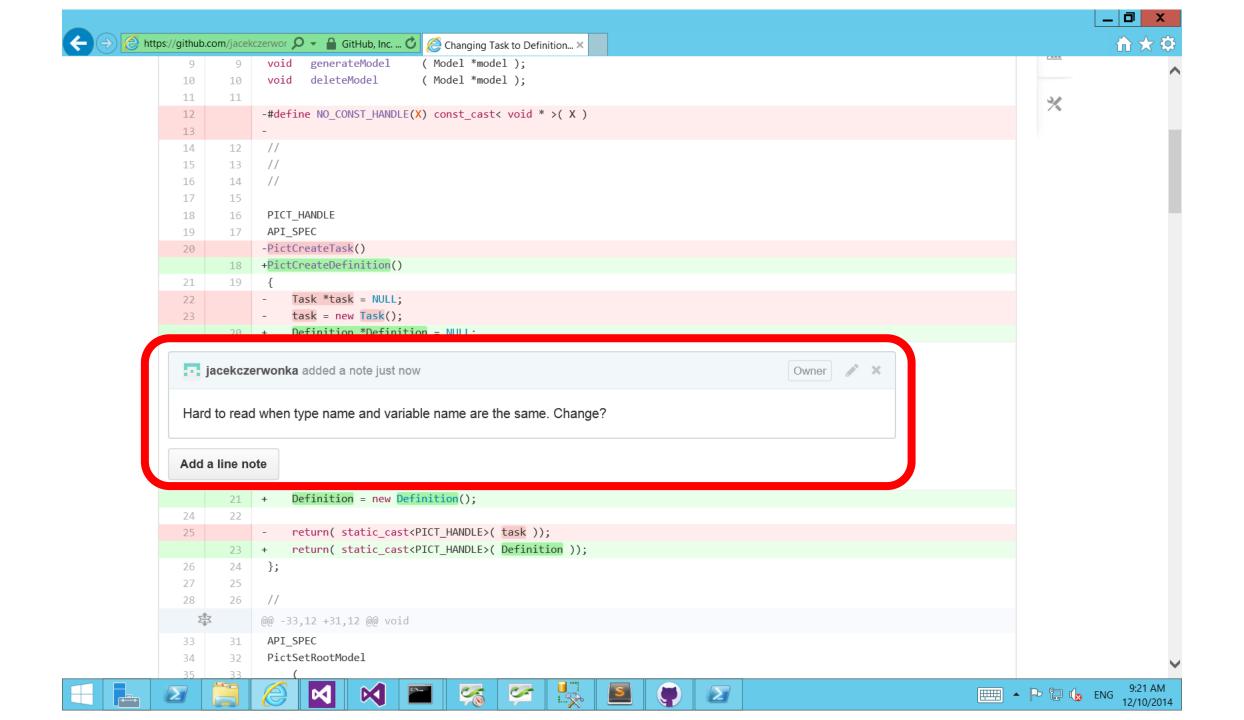


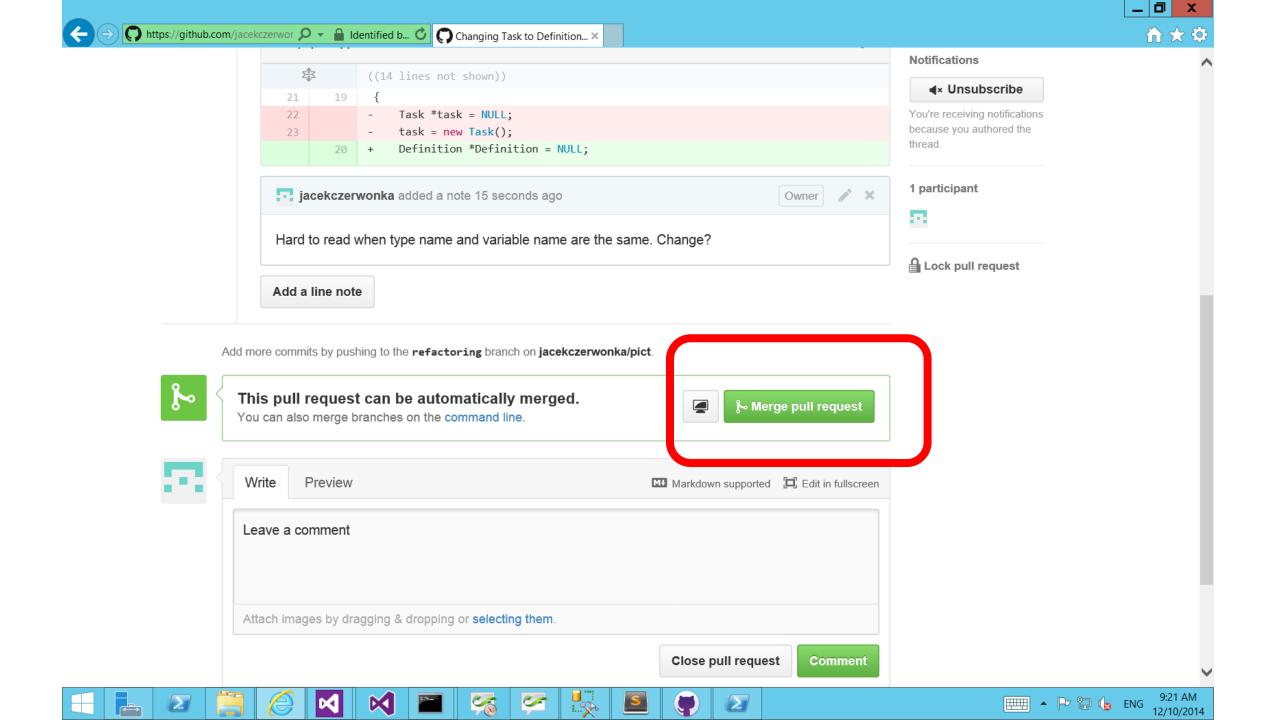












# The goals of code reviewing...



## Why Code Review?

Find defects
Improve maintainability
Share knowledge
Broadcast progress

It is all about the conversation

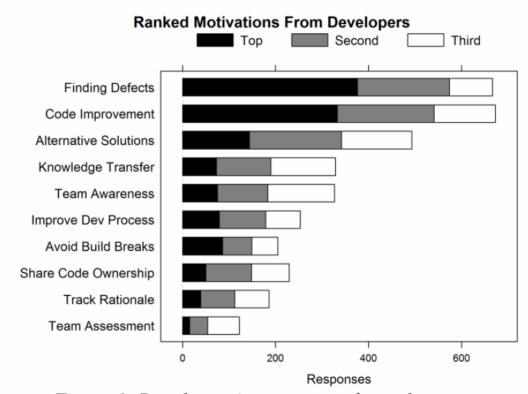
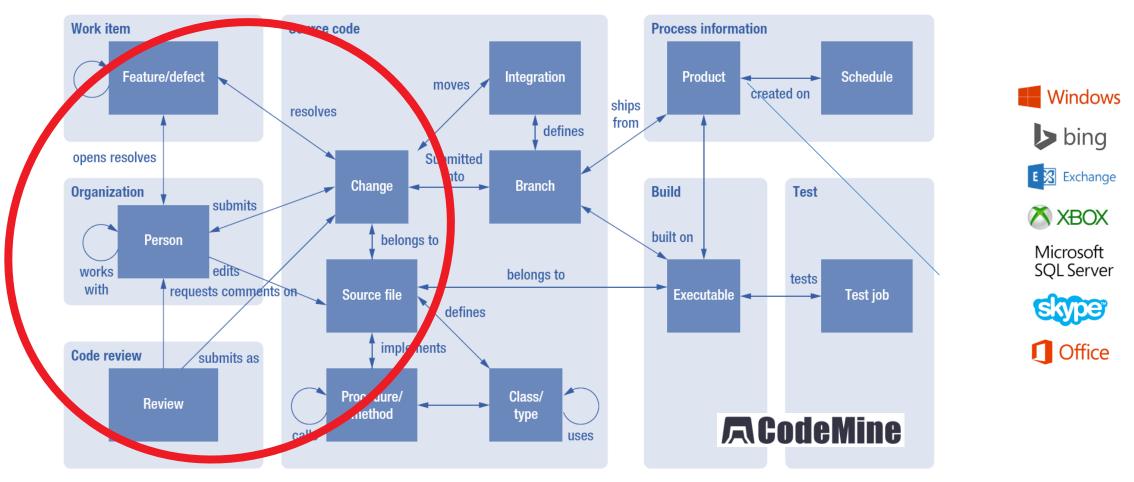


Figure 3. Developers' motivations for code review.

# What do we achieve in practice?



#### **Engineering Data Collection**



Jacek Czerwonka, Nachiappan Nagappan, Wolfram Schulte, Brendan Murphy, CODEMINE: Building a Software Development Data Analytics Platform at Microsoft, IEEE Software, vol. 30, no. 4, pp. 64-71, July-Aug., 2013.

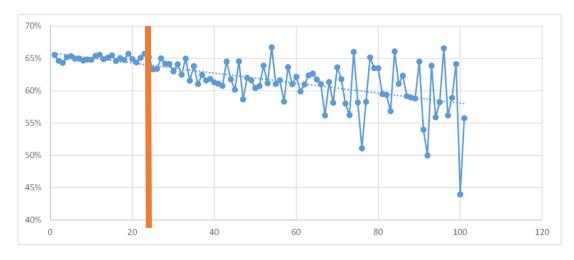
## Study: Comments Classification

Category	Types of issues included	Frequency	
Documentation	Comments, naming, style	22.3%	
Organization of code	Modularity, location of artifacts, new class, duplicate code, size of methods	15.9%	~50% of all
Solution approach	Alternate algorithm or data structure	8.5%	
Validation	Lack of or improper validation	7.2%	
Visual representation	Beautification, indentation, blank linkes	6.4%	
False positive	Not an real issue	4.6%	
Defect	Incorrect implemenation or missing functionality	2.6%	
Logical	Control flow or logic issues	2.3%	15% of all
Support	Configuration support systems or libraries	2.0%	
Interface	Interactions with other components	1.5%	
Resources	Resource initialization, manipulation, and release	1.3%	
Timing	Thread synchronization, races	0.3%	
Other		25.1%	

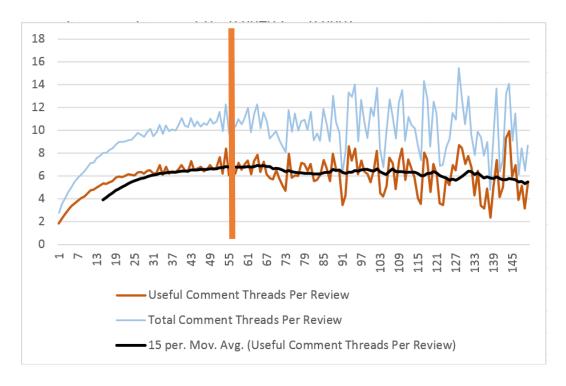
## Study: Code Review Usefulness

Category	Types of issues included	Frequency
Documentation	Comments, naming, style	22.3%
Organization of code	Modularity, location of artifacts, new class, duplicate code, size of methods	15.9%
Solution approach	Alternate algorithm or data structure	8.5%
Validation	Lack of or improper validation	7.2%
Visual representation	Beautification, indentation, blank linkes	6.4%
False positive	Not an real issue	4.6%
Defect	Incorrect implemenation or missing functionality	2.6%
Logical	Control flow or logic issues	2.3%
Support	Configuration support systems or libraries	2.0%
Interface	Interactions with other components	1.5%
Resources	Resource initialization, manipulation, and release	1.3%
Timing	Thread synchronization, races	0.3%
Other		25.1%

#### Smaller Reviews Are Better



Comment usefulness ratio vs. number of files in a reviewed change



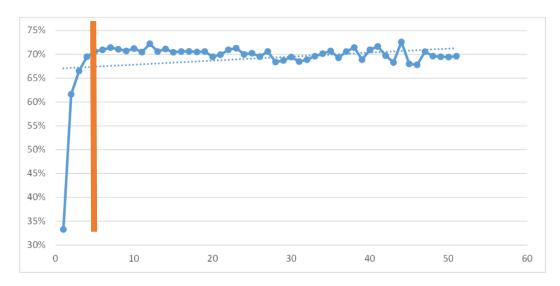
Anecdotally: smaller reviews are "better"

From data: <=20 files implies usefulness stability and predictability

Absolute number of useful comments grows with size of review until 25-30 files, steady until 55-65 and then starts going down

By: Amiangshu Bosu (U of Alabama), Michaela Greiler (TSE), Christian Bird (Microsoft Research Redmond)

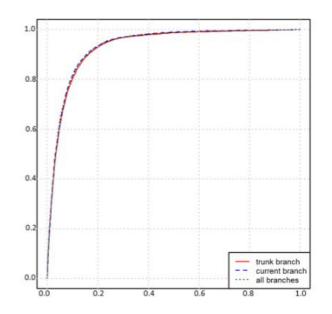
### Relevant Experience Makes for Better Reviewers

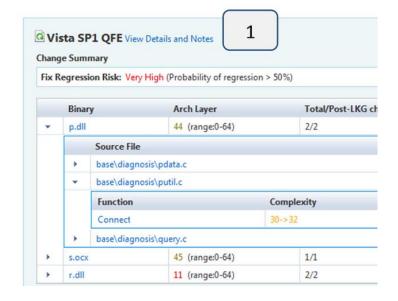


Reviewer's comment usefulness vs. number of previous reviews on a changed file

Reviewers with prior experience with the changed file produce much more useful feedback New reviewers learn fast but need 6-12 months to be as productive as the rest of the team

## Risk of Defects In a Change Can Be Predicted





Prior success with large-scale defect prediction

Expose risk prediction in code review to change the reviewer behavior

Predicting Risk of Pre-Release Code Changes with CheckinMentor, A. Tarvo, N. Nagappan, T. Zimmermann, T. Bhat, J. Czerwonka

CRANE: Failure Prediction, Change Analysis and Test Prioritization in Practice - Experiences from Windows, J. Czerwonka, R. Das, N. Nagappan, A. Tarvo, A. Teterev

# Improving the tools

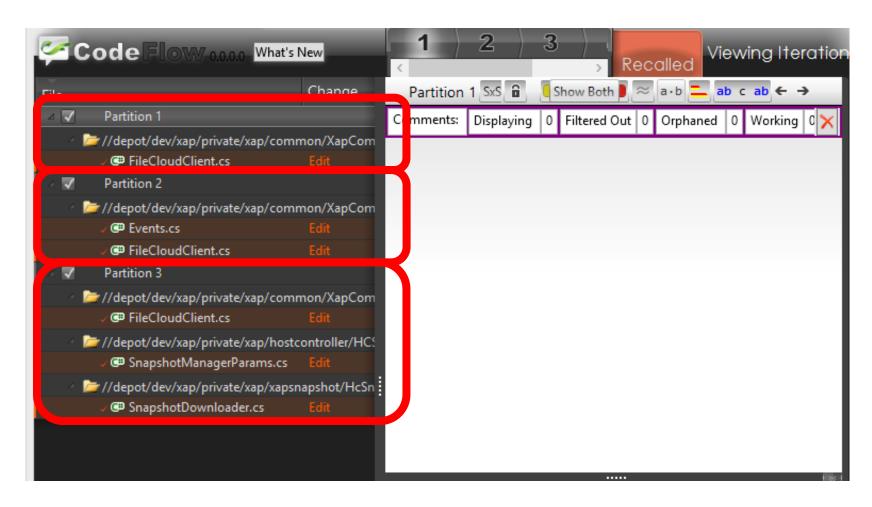


#### Feature #1: Reviewer Recommendations

- Find potential reviewers based on their previous history with the code
- Consider number of changes and time since last activity
- Default is two reviewers based on most common practice and usefulness data

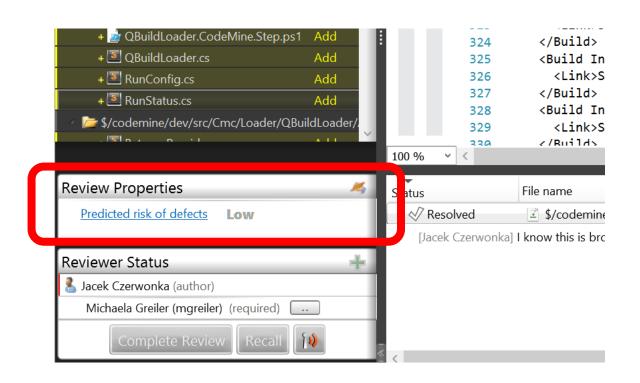


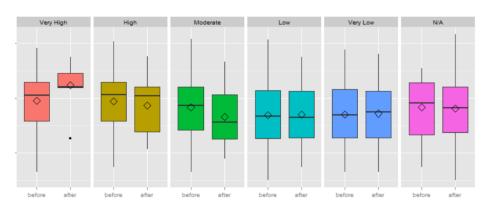
## Feature #2: Change Decomposition



By: Shuvendu Lahiri, Mike Barnett, Christian Bird, Jack Tilford (Microsoft Research Redmond and TSE)

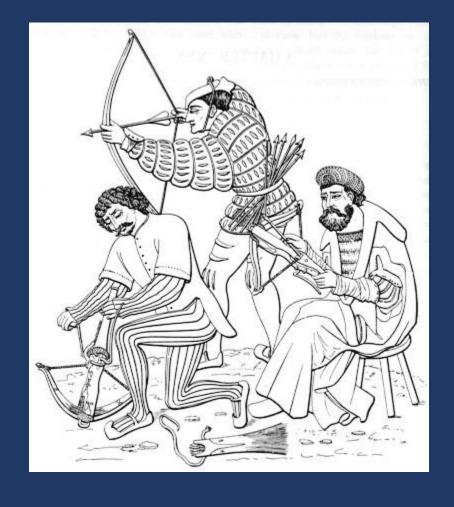
#### Feature #3: Change Risk Prediction



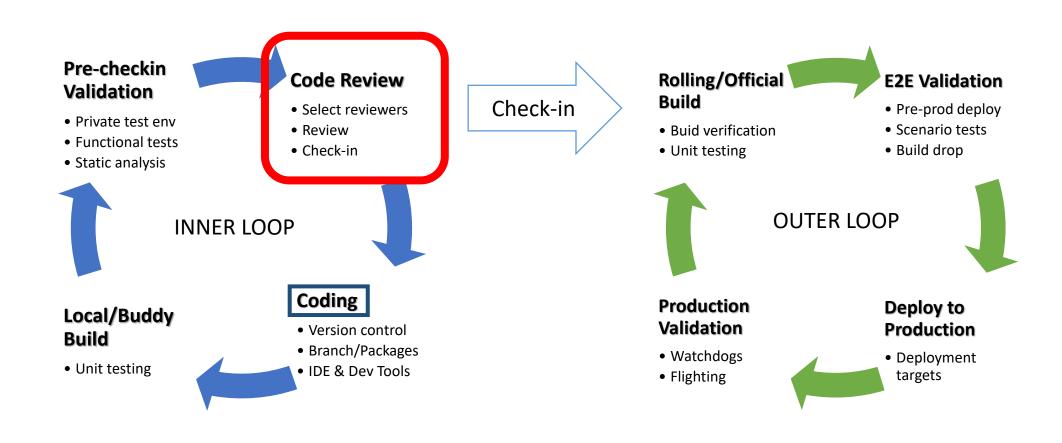


Time to first sign-off before and after enabling "change risk" feature

# Improving the workflow



## Code Reviews in Engineering Workflow



## Study: Comments Classification

Category	Types of issues included	Frequency	
Documentation	Comments, naming, style	22.3%	
Organization of code	Modularity, location of artifacts, new class, duplicate code, size of methods	15.9%	~50% of all
Solution approach	Alternate algorithm or data structure	8.5%	
Validation	Lack of or improper validation	7.2%	
Visual representation	Beautification, indentation, blank linkes	6.4%	
False positive	Not an real issue	4.6%	
Defect	Incorrect implemenation or missing functionality	2.6%	
Logical	Control flow or logic issues	2.3%	15% of all
Support	Configuration support systems or libraries	2.0%	
Interface	Interactions with other components	1.5%	
Resources	Resource initialization, manipulation, and release	1.3%	
Timing	Thread synchronization, races	0.3%	
Other		25.1%	

#### Code Reviewing: It Takes Time and Effort

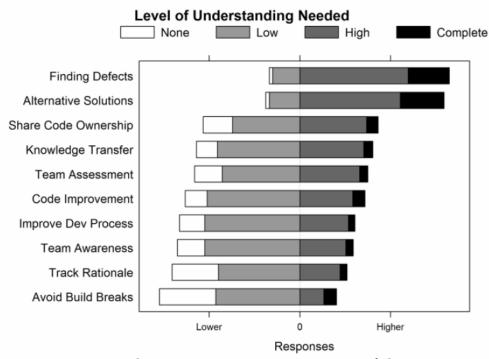


Figure 5. Developers' responses in surveys of the amount of code understanding for code review outcomes.

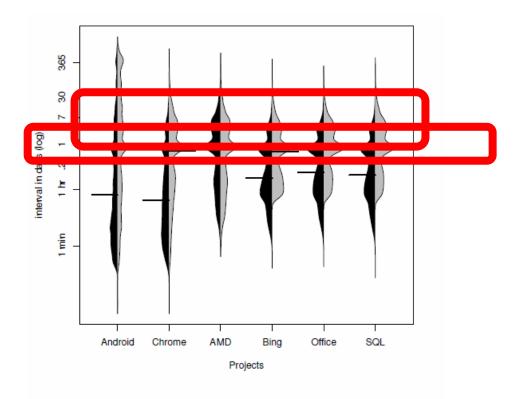


Figure 1: First Response on left (we do not have first response data for AMD) and Full interval on right

Peter C. Rigby, and Christian Bird. *Convergent contemporary software peer review practices*. In Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering, 2013, ESEC/FSE 2013, ACM, pp. 202–212

### The Importance of Speed

**Topic:** General

**Feedback:** Please implement a "virtual whip" feature to ping people who are dragging their feet on a review.

ReviewID: xyz-a723ea9fdc14440586656f9

Project: test

SourceControl: 123456

#### Effects of delaying:

- Process stalls; impact on dependents
- Costly context switches for engineers

#### Code Reviewing is a Social Process

Waiting is not just due to lack of time:

I'm expected to participate but I'm not quite sure how. I'll wait until someone else starts.

There are a lot of outstanding comments already. I'll wait until the next version.

Code reviewing in social context

Reviewing can be uncomfortable for authors <u>and reviewers</u>

Team's hierarchy influence the outcome

How many engineers does it take to do a proper code review?

Sometimes reviewers are added as a courtesy

## Code Reviewing is Social

#### **Review completed**

Submitted in changelist:3561857

Open in: [CodeFlow] [Dogfood] [Browser] [Visual Studio]

Author:

Reviewer	Туре	Status	
Newton Sandar Joes Sandar	Required	SignedOff	
Vinod St.	Required	SignedOff	
Erik Walland Balling	Required	SignedOff	
Yancho (Anna)	Required	Started	
Hamed Estate Charge	Required	NotStarted	
CodeFlow Listener (to)	Optional	SignedOff	
Q Build Team - Internal (	Optional	InviteOnly	

#### Description:

Fix Bug 168324. Zero build outputs get Cached and persist through sessions causing build failures

#### Code Reviews Are Not Free

#### **Potential Benefits**

- More maintainable designs
- More consistent code base
- Knowledge sharing
- Better awareness of changes
- Additional defects found

#### **Potential Costs**

- Time spent by reviewers
- Time spent by author addressing feedback
- Time spent by change waiting in process

Do we need to apply the same verification criteria to all changes?

### Opportunities for Optimization

#### **Automate tasks**

• Syntax, style conformance, static analysis

#### Make better use of people and resources

- When is it required to have a senior engineer?
- Since we mostly find maintainability issues, who is best at that?
- How to best prioritize work for reviewers?

#### **Reorder steps**

Code review after submission not before?

#### **Eliminate steps**

Auto-submit after required reviewers signed off?

#### Make reviewing optional

- Does every change need the same level of reviewing?
- Which types of changes do not benefit little from code reviewing?

Significant workflow changes require better understanding of how and when code reviews provide value

#### Summary

Code reviews do not find as many bugs as you may think (they are still very useful) There is a discrepancy between what developers aim for and what the process does You can't use code reviews as a replacement for other verification techniques Code reviewing is costly as often the longest and most variable part of code integration Code reviewing is a skill that needs time to hone Social aspects of code reviewing make it a complex process Data can influence changes to tools but verify the outcome Applying the same process to all commits is wasteful but to early to make changes To improve further we need to understand the process even more



jacekcz@microsoft.com @jacekcz

Mission: "Enable Microsoft to accelerate its software development"

TSE contributes to and innovates on major parts of the engineering system

TSE runs many of the engineering services and works with all engineering teams

TSE collaborates closely with Microsoft Research teams

Collaboration with academia: visiting researchers and PhD interns (10 this summer)



Wolfram Schulte



Michaela Greiler



Kim Herzig



Suresh Thummalapenta



Nikolai Tillmann



Trevor Carnahan

http://research.microsoft.com/tse