

# Confusion Detection in Code Reviews



Felipe Ebert



Fernando Castor

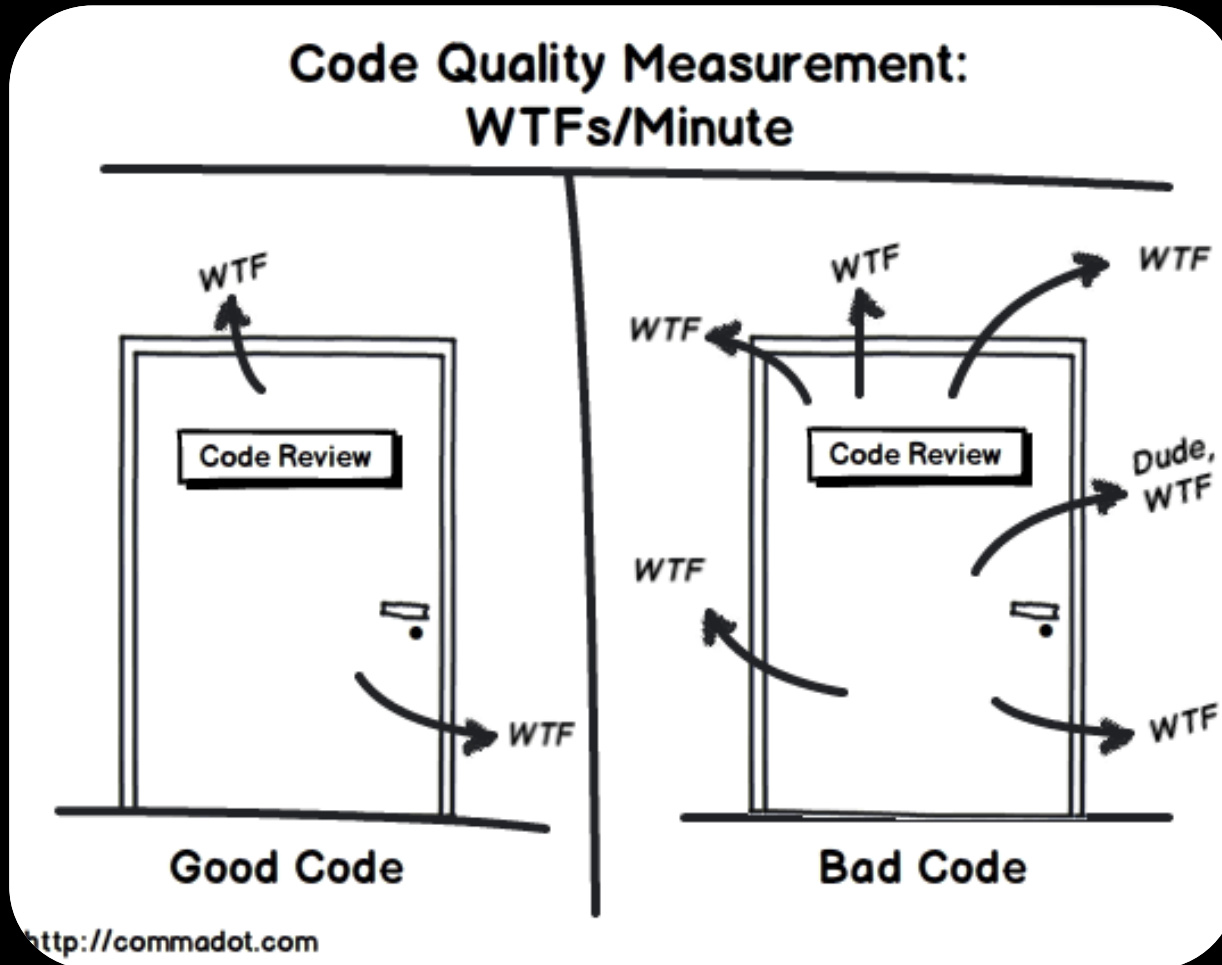


Nicole Novielli



Alexander Serebrenik

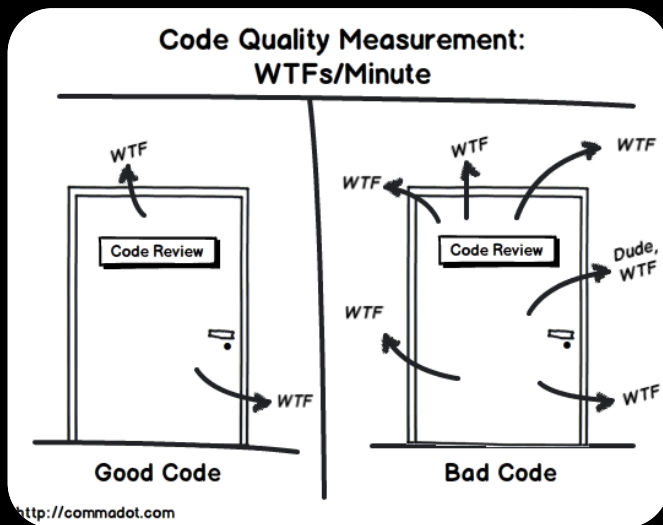
# Why confusion?!



# What is confusion?!

*“a situation in which people are  
**uncertain** about what to do or are  
**unable to understand**  
something clearly”*





## Patch Set 2: Code-Review+2

Though ***I don't really understand*** why *ValueObject* moved to runtime...

<https://android-review.googlesource.com/110347>

## Patch Set 1:

***What's the context?*** Is this fixing/improving existing code? Could you use the assembler tests for it?

<https://android-review.googlesource.com/140403>

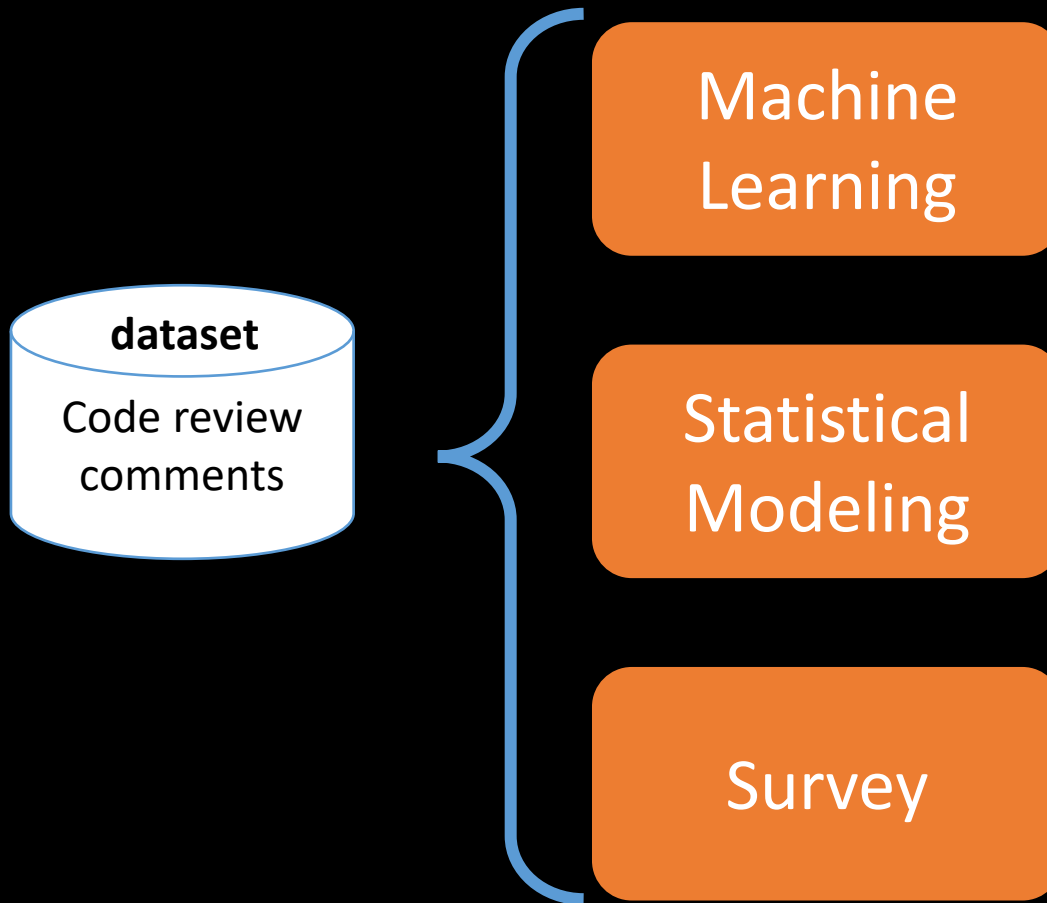
***why do you need any pixels here?*** as I understand, *nullptr* could be OK here, as this is an output, not input texture

<https://android-review.googlesource.com/291770>



To understand the  
**reasons** and **consequences**  
of **confusion** in code reviews

# Methodology



Provide the code documentation

Reviewers

*Patch Set 2: Code-Review+2*

*Though **I don't really understand** why ValueObject moved to runtime...*

Guidelines with best practices on coding and submitting for review

Authors

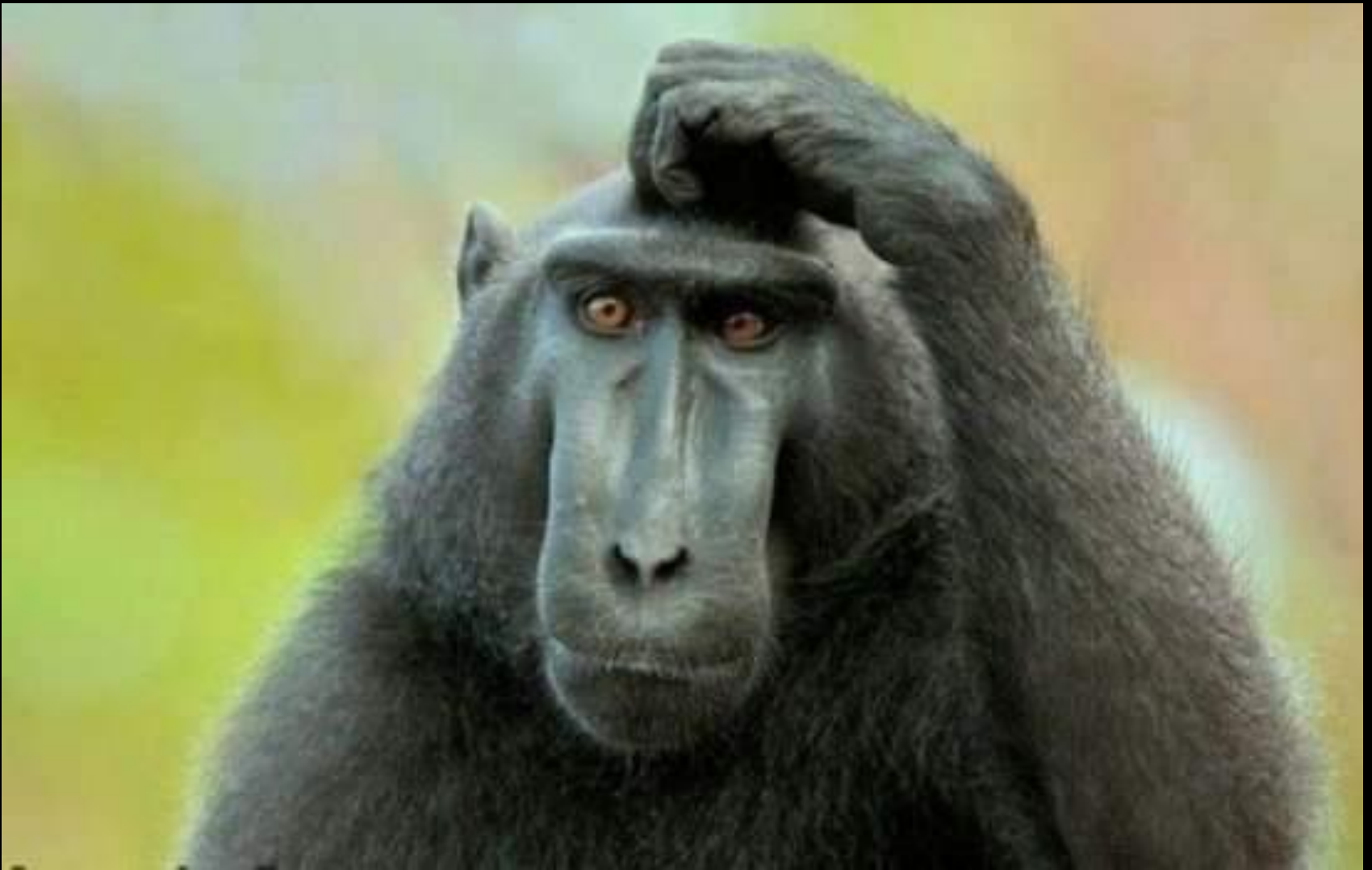
*Patch Set 1:*

***What's the context?** Is this fixing/improving existing code? Could you use the assembler tests for it?*

Provide other parts of the code

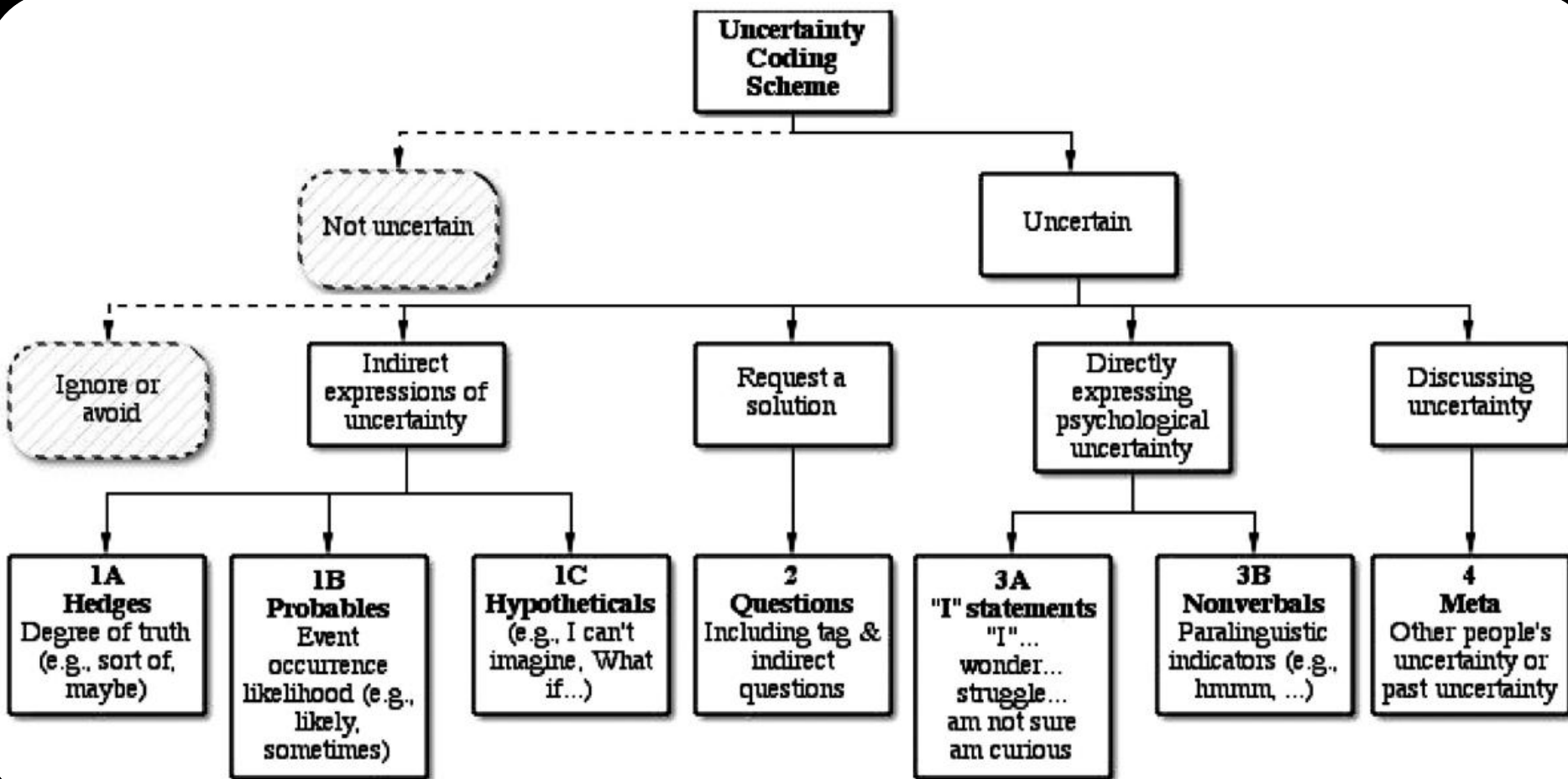
Reviewers

***why do you need any pixels here?** as I understand, nullptr could be OK here, as this is an output, not input texture*



How do we identify and measure confusion?





Michelle E. Jordan et al., "Expressing uncertainty in computer-mediated discourse: Language as a marker of intellectual work," *Discourse Processes*, vol. 49, no. 8, pp. 660–692, 2012.

## Initial Data

android

### comments

660,845 GC

232,471 IC

140,006

code reviews

GC – General Comment

IC – Inline Comment

## Initial Data

android

comments

660,845 GC

232,471 IC

140,006

code reviews



## Filtering

Confusion  
Framework



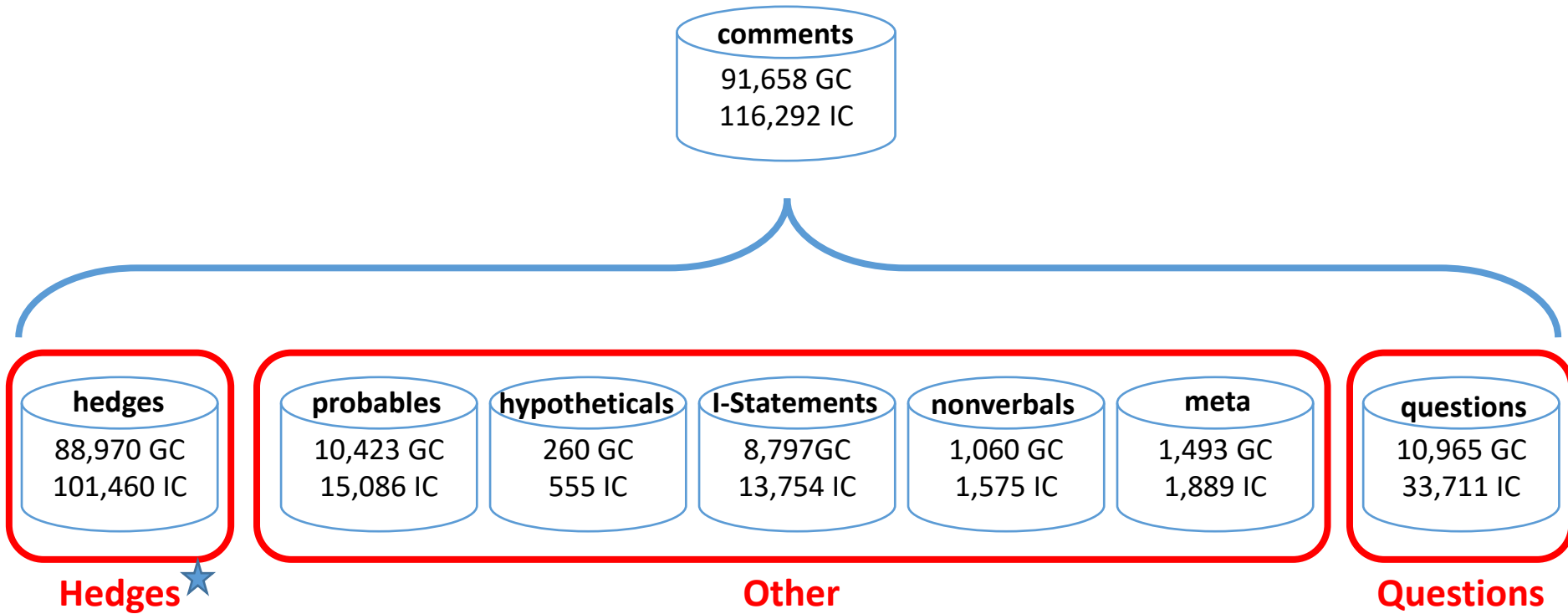
comments

91,658 GC

116,292 IC

# Filtering

## Confusion Framework



## Initial Data

android

comments

660,845 GC  
232,471 IC

140,006  
code reviews



## Filtering

Confusion  
Framework



comments

91,658 GC  
116,292 IC

***Maybe** write a comment with the  
XML format here*

no confusion!

*Patch Set 1: **Could** anyone submit  
this?*

no confusion!

*Patch Set 5: Svet: **Could** you please  
review?*

no confusion!

## Initial Data

android

comments

660,845 GC  
232,471 IC

140,006  
code reviews

## Annotation of Confusion

hedges

400 GC  
400 IC



Annotation  
of  
Confusion

- 4 raters
- $K(GC) = .59$
- $K(IC) = .49$

## Filtering

Confusion  
Framework



comments

91,658 GC  
116,292 IC



## Initial Data

android

### comments

660,845 GC  
232,471 IC

140,006  
code reviews

## Annotation of Confusion

### hedges

400 GC  
400 IC

### Annotation of Confusion

- 4 raters
- $K(GC) = .59$
- $K(IC) = .49$

IC: lower Fleiss' kappa!!!

## Filtering

### Confusion Framework

### comments

91,658 GC  
116,292 IC

## Initial Data

**android**

**comments**

660,845 GC  
232,471 IC

140,006  
code reviews

## Annotation of Confusion

**hedges**

400 GC  
400 IC

Annotation  
of  
Confusion

- 4 raters
- $K(GC) = .59$
- $K(IC) = .49$

## Filtering

Confusion  
Framework

**comments**

91,658 GC  
116,292 IC

## Gold Standard

**comments**

396 GC  
396 IC

Confusion comments:

- 72 GC (18%)
- 84 IC (21%)
- 4 GC and 4 IC discarded



12



### Precision

OneR

	P	R	F
GC	.875	.194	.318
IC	.615	.095	.165

### Recall

Multinomial  
Naive Bayes

	P	R	F
GC	.209	.944	.342
IC	.234	.988	.378

### Precision and Recall

JRip

	P	R	F
GC	.696	.542	.609
IC	.434	.583	.497

Logistic



### Precision

OneR

	P	R	F
GC	.875	.194	.318
IC	.615	.095	.165

### Recall

Multinomial  
Naive Bayes

	P	R	F
GC	.209	.944	.342
IC	.234	.988	.378

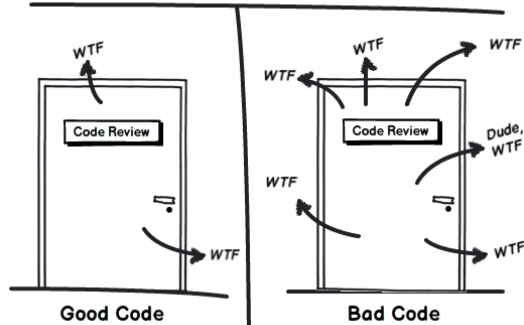
### Precision and Recall

JRip

Logistic

	P	R	F
GC	.696	.542	.609
IC	.434	.583	.497

### Code Quality Measurement: WTFs/Minute

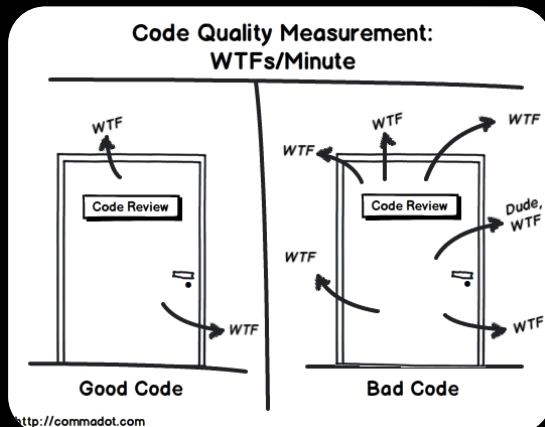


<http://commadot.com>

## Inline comment

*Do you really want a Java string here?  
A ModifiedUTF8 one not enough?*

confusion!



## Inline comment

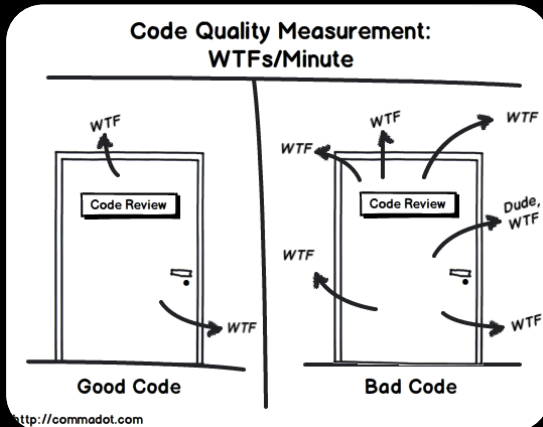
*Do you really want a Java string here?  
A ModifiedUTF8 one not enough?*

confusion!

## Inline comment

*Maybe write a comment with the XML  
format here*

no confusion!



## Inline comment

*Do you really want a Java string here?  
A ModifiedUTF8 one not enough?*

confusion!

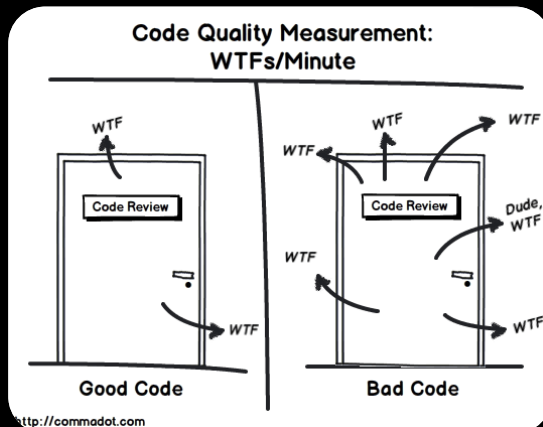
## Inline comment

*Maybe write a comment with the XML  
format here*

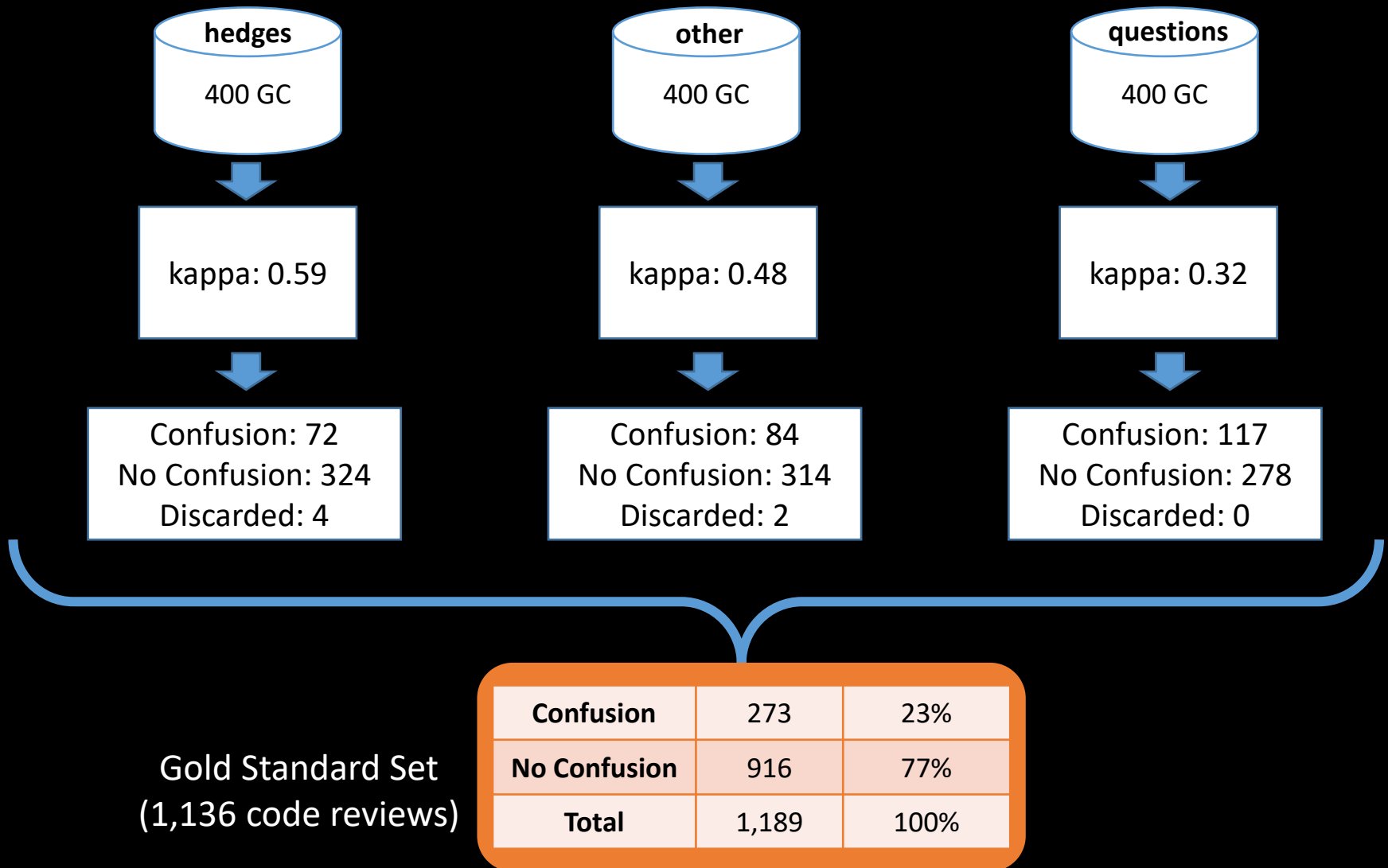
no confusion!

## Future work

- Other categories + new classifiers
- Statistical modeling
- Surveys

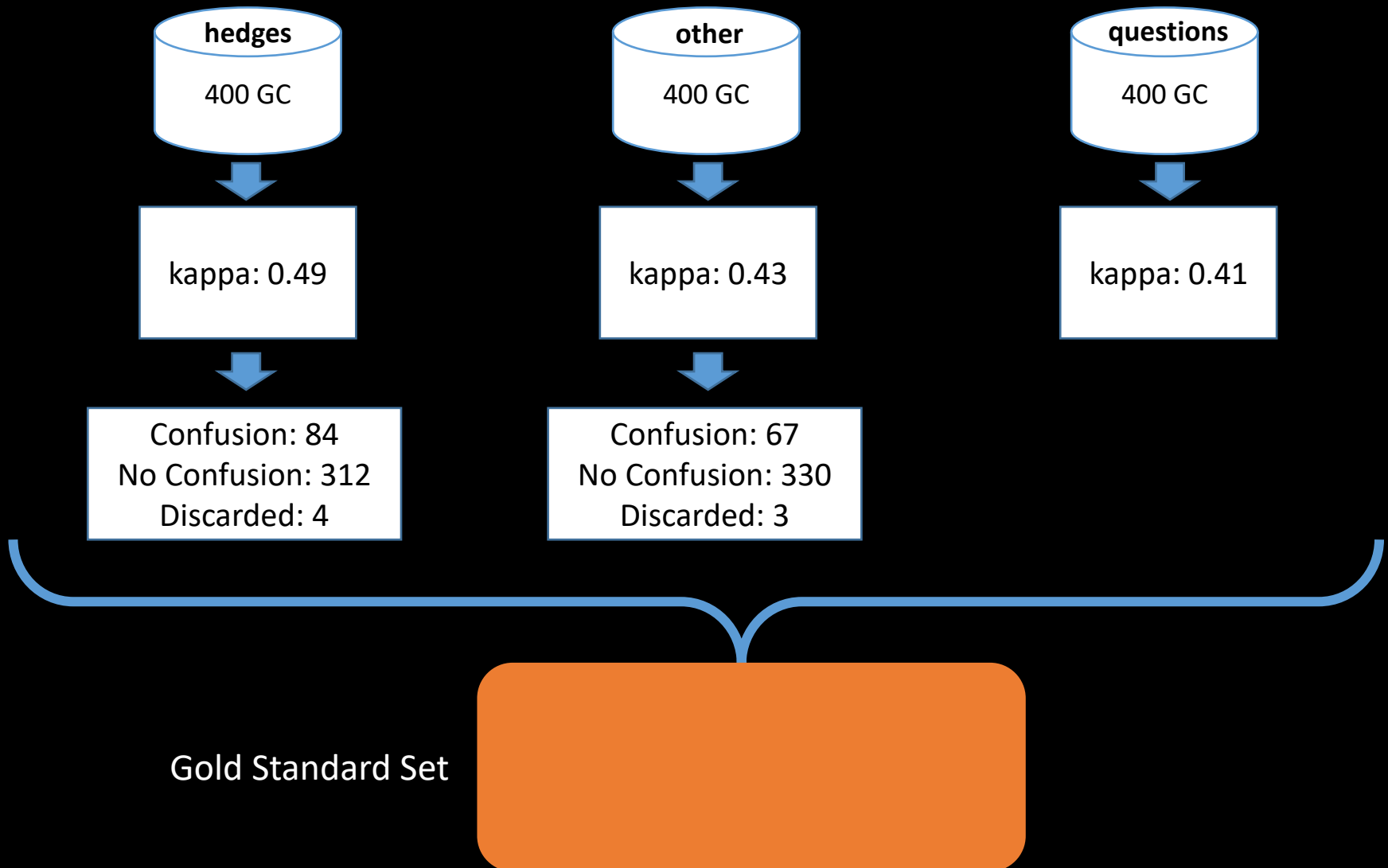


# Manual Annotation - GC





# Manual Annotation - IC



# Survey

## Confusion in Code Reviews Survey

Welcome to the Confusion in Code Reviews survey.

In this study we aim at understanding the reasons why developers get confused when performing code reviews and the impact of this confusion. By identifying and classifying those reasons we want to make the code review more efficient.

We believe that developers can benefit from this study by learning causes of confusion and trying to avoid them in the code changes they submit for review. We also think static analysis tools can be expanded so as to provide early feedback on code changes that might be hard to understand for reviewers.

Your participation is voluntary and confidential. We do not record any identifying information. If you agree to participate, you will be asked about experiences related to code reviews. Participation in this study is expected to take about 20 minutes of your time. You might withdraw at any time.

This survey is conducted by a joint team of computer science researchers from Federal University of Pernambuco, Brazil (Felipe Ebert <[fe@cin.ufpe.br](mailto:fe@cin.ufpe.br)> and Fernando Castor <[fjclf@cin.ufpe.br](mailto:fjclf@cin.ufpe.br)>), Eindhoven University of Technology, The Netherlands (Alexander Serebrenik <[a.serebrenik@tue.nl](mailto:a.serebrenik@tue.nl)>) and University of Bari, Italy (Nicole Novielli <[nicole.novielli@uniba.it](mailto:nicole.novielli@uniba.it)>).

We thank you in advance for your participation in this study. Individual responses cannot be traced back to an individual respondent. We plan to include the results of this survey in a scientific publication. Should you be interested in being informed about the outcome of this study or any resulting publication, you will be provided an opportunity to indicate this and provide us with your email address.

If you have any additional comments, please feel free to use the text box at the end, or to contact us directly.

\* Required

### ELECTRONIC CONSENT \*

Please select your choice below. Selecting the "yes" option below indicates that: i) you have read and understood the above information, ii) you voluntarily agree to participate, and iii) you are at least 18 years old. If you do not wish to participate in the research study, please decline.

- Emails sent: 4,645
- Deliverable: 3,765
- Undeliverable: 880
- Responses: 17 (0.4%)

# Survey

- **How often** did you feel confused...?
  - Author
  - Reviewer
- **What** usually makes you confused...?
- What is the **impact** of confusion...?
- What do you usually do to **overcome** confusion...?

# What usually makes you confused when you are reviewing code changes?

- *The code change is too big*
- *I am not familiar with the component*
- *I don't have enough time to get into the code to understand it*

# What usually makes you confused when you are reviewing code changes?

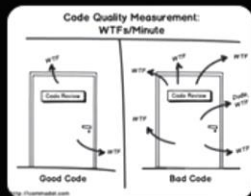
- *The code change is too big*
- *I am not familiar with the component*
- *I don't have enough time to get into the code to understand it*
- *I have too much chaos and noise going on around me in this open office work environment*

# How does the confusion you experience as a reviewer **impact** code review?

- *Code reviews take longer*
- *There's additional back and forth*
- *I can't give a high quality code review*
- *I need to study the code, documentation*

# How does the confusion you experience as a reviewer **impact** code review?

- *Code reviews take longer*
  - *There's additional back and forth*
  - *I can't give a high quality code review*
  - *I need to study the code, documentation*
- 
- *It pissed me off*
  - *Frustration*



#### Patch Set 2: Code-Review+2

Though I don't really understand why ValueObject moved to runtime ...

<https://android-review.googlesource.com/110347>

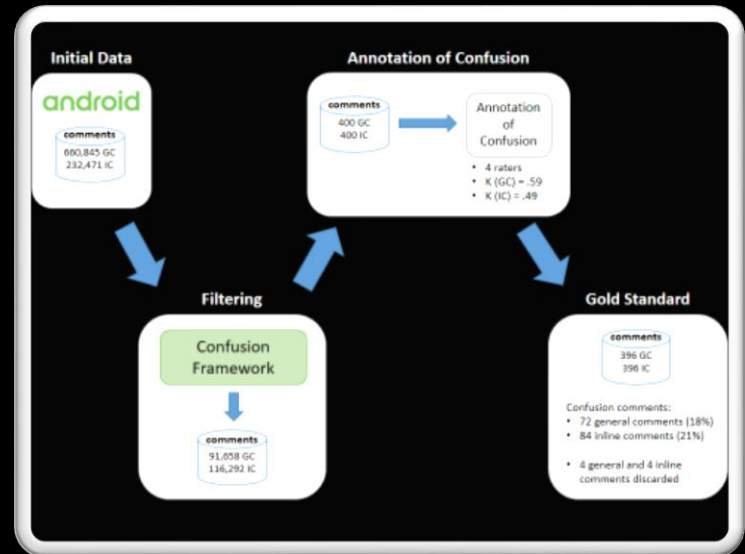
#### Patch Set 1:

What's the context? Is this fixing/improving existing code? Could you use the assembler tests for it?

<https://android-review.googlesource.com/140403>

why do you need any pixels here? as I understand, nullptr could be OK here, as this is an output, not input texture

<https://android-review.googlesource.com/291770>



Classifier	Class	Baseline + 3		
		Precision	Recall	F-Measure
OneR	Confusion	.875	.194	.439
	No Confusion	.847	.994	.915
Multinomial Naive Bayes	Confusion	.209	.94	.342
	No Confusion	.943	.204	.335
JRIp	Confusion	.696	.542	.609
	No Confusion	.903	.948	.925

GC vs IC

Classifier	Class	Baseline + 3			Baseline + hedges		
		Precision	Recall	F-Measure	Precision	Recall	F-Measure
OneR	Confusion	-	-	-	.615	.095	.165
	No Confusion	-	-	-	.802	.984	.883
Multinomial Naive Bayes	Confusion	.234	.988	.378	-	-	-
	No Confusion	.976	.128	.227	-	-	-
Logistic	Confusion	.434	.583	.497	-	-	-
	No Confusion	.876	.795	.834	-	-	-

