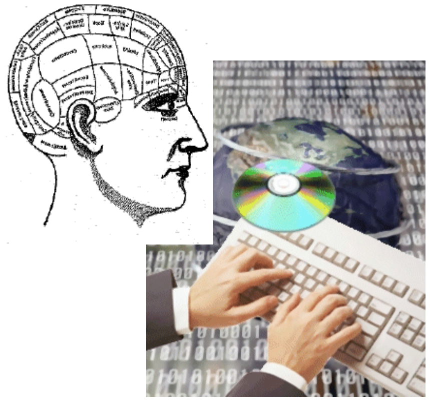
Soft Lab Software Research Laboratory

PROMISE 2010



Empirical Analyses of the Factors Affecting Confirmation Bias and the Effects of Confirmation Bias on Software Developer/Tester Performance

Gül Çalıklı[1] and Ayşe Bener[2]

Bogazici University[1] and Ryerson University[2]

Department of Computer Engineering[1], Ted Rogers School of Information Technology Management[2]

Istanbul, Turkey[1], Toronto, Canada[2]

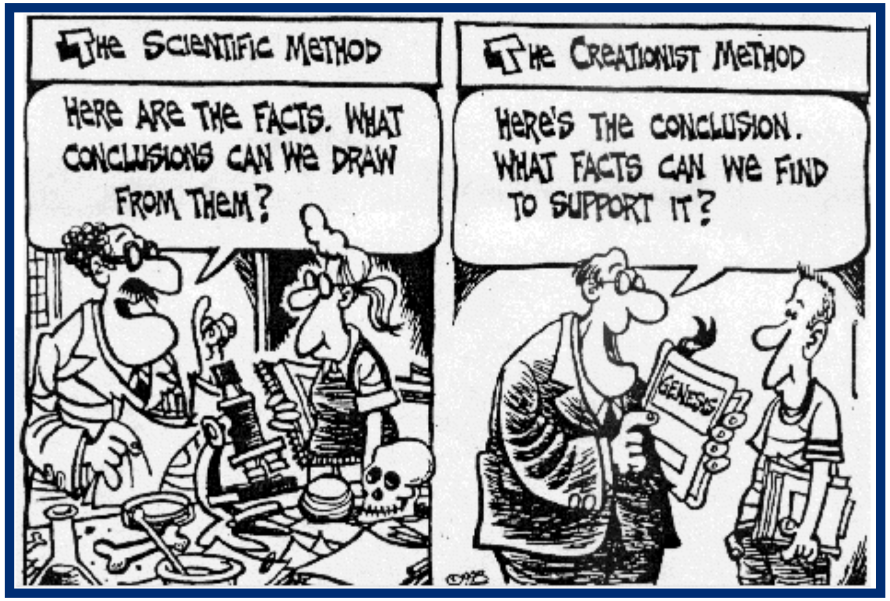
gul. calikli@boun.edu.tr[1], ayse.bener@ryerson.edu[2]

OUTLINE

- Introduction and Motivation
- Research Questions & Contributions
- Methodology for Data Extraction
- Metrics
- Dataset
- Analyses and Results
- Threats to Validity
- Conclusion and Future Work

INTRODUCTION AND MOTIVATION

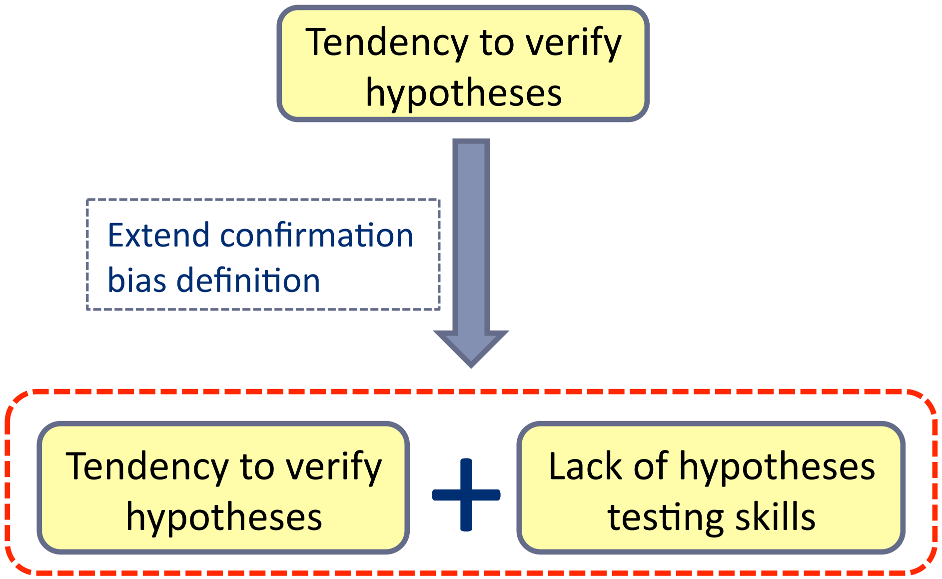
- Definition (*cognitive psychology*):
  - **confirmation bias** is defined as the tendency of people to seek evidence that could <u>verify</u> their theories rather than seeking for evidence that could falsify them.
- Relationship between confirmation bias and software testing:
  - During all levels of ***software testing*** the attempt should be to <u>*fail the code*</u> to reduce software defect density.



CONFIRMATION BIAS → Leads to → increase in software DEFECT DENSITY

INTRODUCTION AND MOTIVATION

- What is known as "***hypotheses testing skills & methods***" in cognitive psychology should also be employed while testing software. As a result of a systematic aproach, the ability to detect errors can be maximized.
- Definition (*software development/testing*):
  - We extend the definition of ***confirmation bias*** to include the lack of ***hypotheses testing skills & methods*** as well as the tendency to verify hypotheses.



RESEARCH QUESTIONS

- How is confirmation bias affected by factors such as "experience" and "education"?

- How is the performance of software developers affected by confirmation bias?

- How is the performance of software testers affected by confirmation bias?

CONTRIBUTIONS OF THE WORK

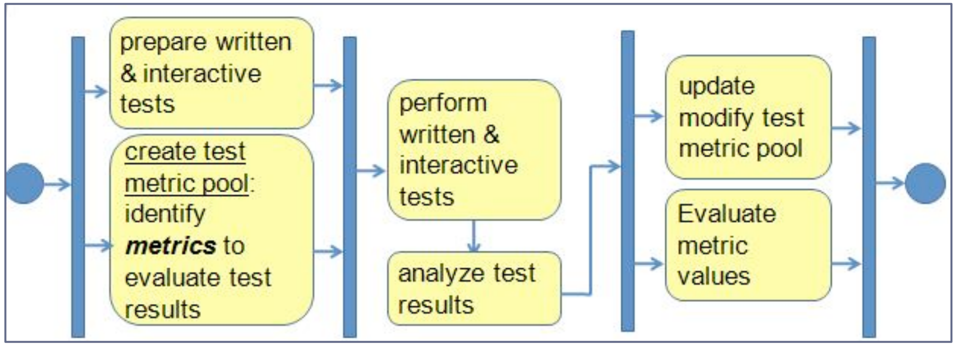
1. Construct a methodology to quantify/measure confirmation bias levels of software developers/testers

2. Analyze the effect of the factors such as experience and education on confirmation bias

3. Analyze the effect of confirmation bias on software testers

4. Analyze the effect of confirmation bias on software developers

DATA EXTRACTION

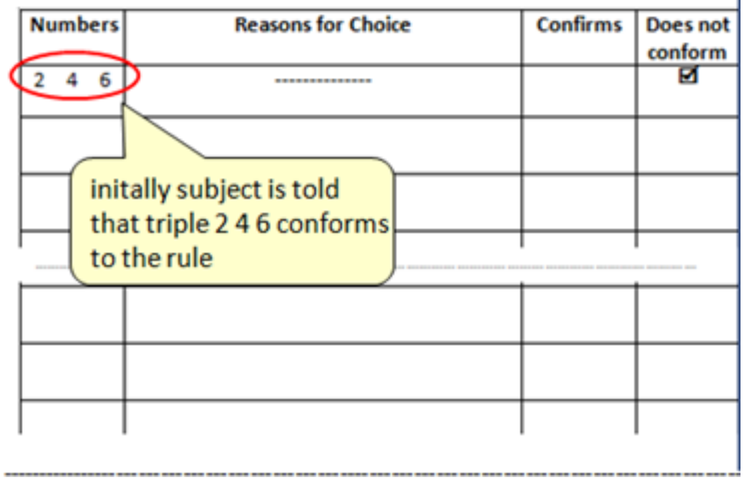


Written Test is based on Wason's Selection Task **.

Information about Written Test content

Question Type	No. of Questions
Abstract Questions	8
Thematic Questions	6
SW development/ testing questions	8
TOTAL	**22**

Interactive Test is based on Wason's Rule Discovery Task *. Subject is given a record sheet and the goal is to discover a rule applying to triples of numbers.



METRICS

Interactive Test Metrics	
Abbreviation	**Explanation**
$Ind_{Elim/Enum}$	Elimination/Enumeration index by Wason.
T_I	Interactive test duration (minutes)
F^{IR}	Immediate rule announcement frequency
avg_L^{IR}	Total number of rule announcements in a series, where no instances are given in between rule announcements
avg_F^{RR}	Average frequencies of reason repetition/reformulation
N_A	Total number of rule discovery attempts.
Written Test Metrics	
Abbreviation	**Explanation**
S_{ABS}	Score in abstract questions
S_{Th}	Score in thematic questions
T_{Th+ABS}	Duration it took to solve abstract and thematic questions (minutes)
S_{SW}	Score in questions with software development/testing theme
T_{SW}	Duration it took to solve questions with software development/testing theme (minutes)

DATASET

- **Group 1**
  - 28 Computer Engineering Graduate Students of Bogazici University
  - 14 of them have mininum 3 years of software development experience in various companies.

- **Group 2**
  - 28 software developers/testers working in a large scale telecommunication company in Europe
  - 28 subjects = 12 developers + 16 testers
  - Subjects have only undergraduate degree in Computer Engineering and related fields.

ANALYSES AND RESULTS

➢ Analysis of the Factors Affecting Confirmation Bias

1. ***Effect of Experience on Confirmation Bias***

➢ Independent variable: Experience in software development/ testing

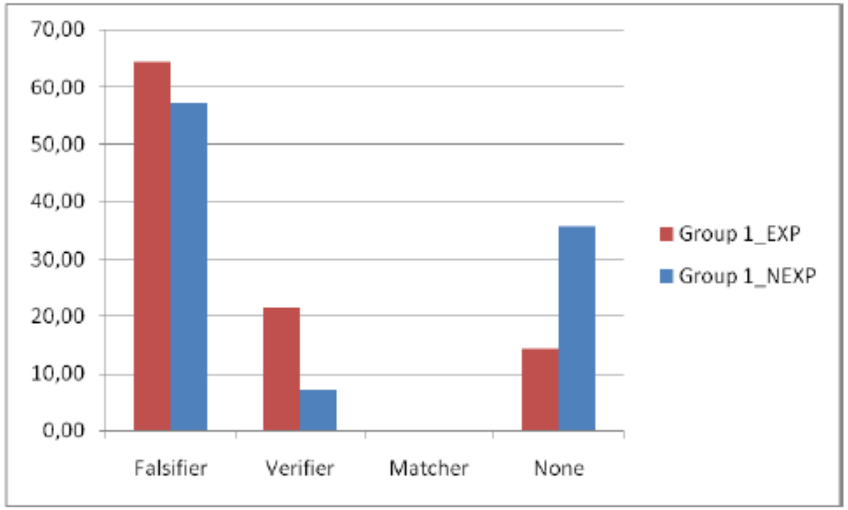
➢ Dealing with Possible Confounding Factors: Statistical analyses were performed within members of Group 1 and Group 2 separately.

ANALYSES AND RESULTS

➢Analysis of the Factors Affecting Confirmation Bias

➢**1.** Effect of Experience on Confirmation Bias (within Group 1)

Distribution of Verifiers, Falsifiers and Matchers in members of subgroups Group1 _EXP & Group1_NEXP according to Reich and Ruth's Method



Results of bootstrapped Kolmogorov-Smirnov Test among Group 1_EXP and Group 1_NEXP

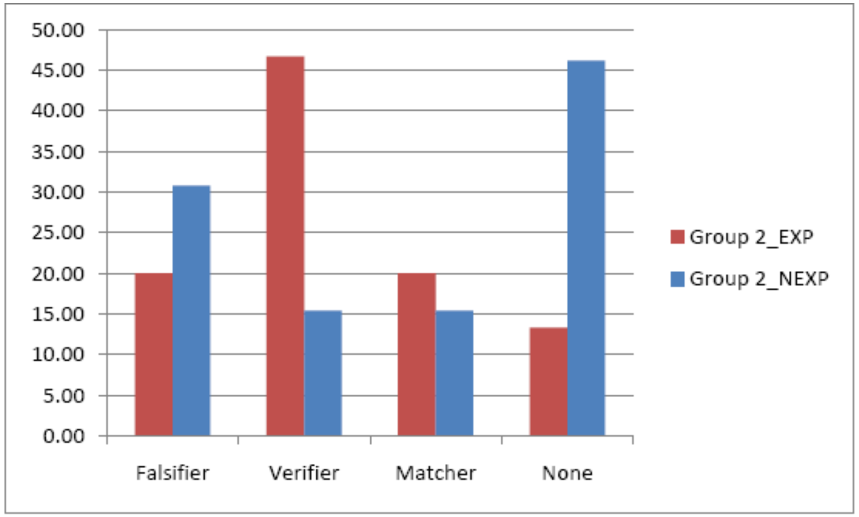
	Group 1_{EXP}	Group 1_{NEXP}	p-value
$Ind_{Elin/Enum}$	1.3029	0.6538	0.3775
T_I	8.6429	6.6923	0.1310
F^{IR}	0.1429	0.6124	0.1150
avg_L^{IR}	0.1429	0.2692	0.2205
avg_F^{RR}	0.6786	0.9746	0.5455
N_A	1.7857	2.6154	0.5365
S_{ABS}	0.5914	0.6350	0.7195
S_{Th}	0.8823	0.9064	0.5405
T_{Th+ABS}	15.0714	12.7857	0.2740
S_{SW}	0.8308	0.7186	0.0010
T_{SW}	11.1429	12.3571	0.2865

ANALYSES AND RESULTS

➢Analysis of the Factors Affecting Confirmation Bias

➢**1.** Effect of Experience on Confirmation Bias (within Group 2)

Distribution of Verifiers, Falsifiers and Matchers in members of subgroups Group2 _EXP & Group2_NEXP according to Reich and Ruth's Method



Results of bootstrapped Kolmogorov-Smirnov Test among Group 2_EXP and Group 2_NEXP

	Group 2_{EXP}	Group 2_{NEXP}	p-value
$Ind_{Elin/Enum}$	1.11	1.12	0.6899
T_I	18.06	16.59	0.3874
F^{IR}	1.00	0.67	1.0000
avg_L^{IR}	0.55	0.53	1.0000
avg_F^{RR}	1.17	0.80	0.8644
N_A	3.61	2.18	0.1170
S_{ABS}	0.19	0.13	0.3874
S_{Th}	0.72	0.71	0.9313
T_{Th+ABS}	18.12	14.5	0.2336
S_{SW}	0.46	0.53	0.9303
T_{SW}	17.59	14.41	0.3874

ANALYSES AND RESULTS

➢ Analysis of the Factors Affecting Confirmation Bias

2. Effect of Education on Confirmation Bias

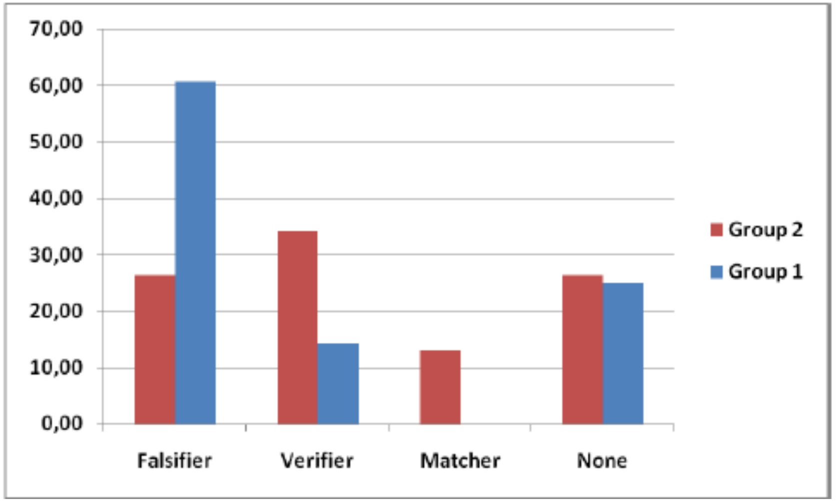
- Independent variable: Education
- Dealing with Possible Confounding Factors:
  - Age : Both groups own similar age profiles (i.e. 25-32 years old)
  - Experience: Analyses within Group1 and within Group 2 as well as 4 more groups from our repository resulted in no statistically significant difference with respect to experience.
  - Software Development Domain: Members of Group 1 with experience used to work for companies who develop software within the same domain and with similar methodologies.

ANALYSES AND RESULTS

➢Analysis of the Factors Affecting Confirmation Bias

➢2. Effect of Education on Confirmation Bias

Distribution of Verifiers, Falsifiers and Matchers in Groups 1 and Group 2 according to Reich and Ruth's Method



Results of bootstrapped Kolmogorov-Smirnov Test among Group 1 and Group 2

	Group1	Group2	p-value
$Ind_{Elim/Enum}$	**1.04**	**1.59**	**0.0039**
T_I	**18.14**	**11.56**	**0.0351**
F^{IR}	0.85	0.19	0.0535
avg_L^{IR}	**0.54**	**0.11**	**0.0535**
N_A	**3.07**	**1.78**	**0.0221**
S_{ABS}	0.27	0.43	0.4285
avg_F^{RR}	1.01	1.61	0.5001
S_{Th}	0.73	0.71	1.0000
T_{Th+ABS}	16.36	16.54	0.6826
S_{SW}	**0.51**	**0.79**	**0.0008**
T_{SW}	15.97	11.54	0.0319

ANALYSES AND RESULTS

➢Analysis of the Effect of Confirmation Bias on Software Development

- ➢ An analysis is performed among developers of Group 2
- ➢ Group 2 members are responsible from development of a customer services package
- ➢ 10 releases of the software that were developed and tested between May 2009 and November 2009 are analysed
- ➢ For each release, each file is categorized as defected or not defected based on the testing phase results
- ➢ For each file which is found to be defected during a testing phase, developers who updated/created that file before that testing phase are held responsible →
  - ➢ Each file is updated/created by a group of one or more developers → <u>**TOTAL:**</u> **124** developer groups.

ANALYSES AND RESULTS

➢ Analysis of the Effect of Confirmation Bias on Software Development

➢ Multilinear regression modeling is used to find the relation between defect rate and confirmation bias.

➢ Response Variable: Defect density for each developer group

- ➢ We defined *defect density* for each developer group as the ratio of the total number of defected files created/updated by that group to the total number of files that group created/updated.

➢ Independent Variables:

$$X_1 = \overline{Ind}_{E\lim/Enum}^{1...N} \qquad X_2 = Ind_{E\lim/Enum}^{1...N}$$

$$X_3 = T_I^{1...N} \qquad X_4 = \left(F^{IR}\right)^{1...N}$$

$$X_5 = \left(avg\_L^{IR}\right)^{1...N} \qquad X_6 = \left(avg\_F^{RR}\right)^{1...N}$$

$$X_9 = S_{Th}^{1...N} \qquad X_{10} = T_{Th+ABS}^{1...N}$$

$$X_{11} = S_{SW}^{1...N} \qquad X_{12} = T_{SW}^{1...N}$$

$$X_{13} = TestSeverity_{avg}^{1...N}$$

$$X_{14} = TestSeverity_{\min}^{1...N}$$

$$X_{15} = TestSeverity_{\max}^{1...N}$$

$$X_{avg} = \begin{bmatrix} \dfrac{\sum_{i=1}^{N} X_1^i}{N} & \cdots & \dfrac{\sum_{i=1}^{N} X_{15}^i}{N} \end{bmatrix}$$

ANALYSES AND RESULTS

➢ Results for the Analysis of the Effect of Confirmation Bias on Software Development

➢ PCA (Principal Component Analysis): β_2, β_3 , β_4 , β_5, β_6 are linear combinations of the metrics and confirmation bias related values.

➢ $R^2 = 0.4477$

➢ $R_{adj}^2 = 0.4243$ → %42 of the variability in defect density is explained by the parameters β_i , i = 2,...,6

Coefficient	Coefficient Value	Confidence Interval	p value
β_1	6.5669	6.0569 - 7.0688	1.0791E-12
β_2	0.2696	0.0507 - 0.4896	0.0162
β_3	-0.1472	-0.4809 - 1.1866	0.3843
β_4	1.4814	1.0971 - 1.8657	6.543E-12
β_5	0.6248	0.0496 - 1.2000	0.0335
β_6	-1.2697	-1.9005 - -0.6309	1.167E-4

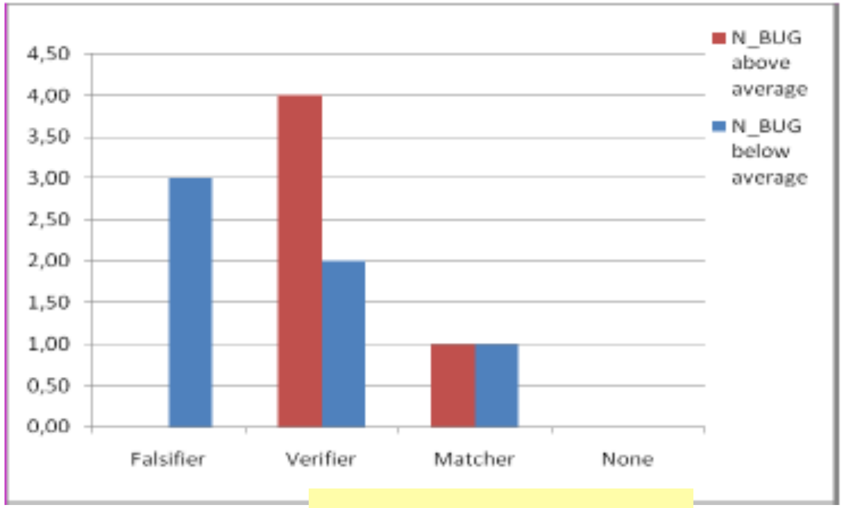
The values of regression coefficients with their confidence intervals ($\alpha = 0.05$)

ANALYSES AND RESULTS

➢Analysis of the Effect of Confirmation Bias on Software Tester Performance

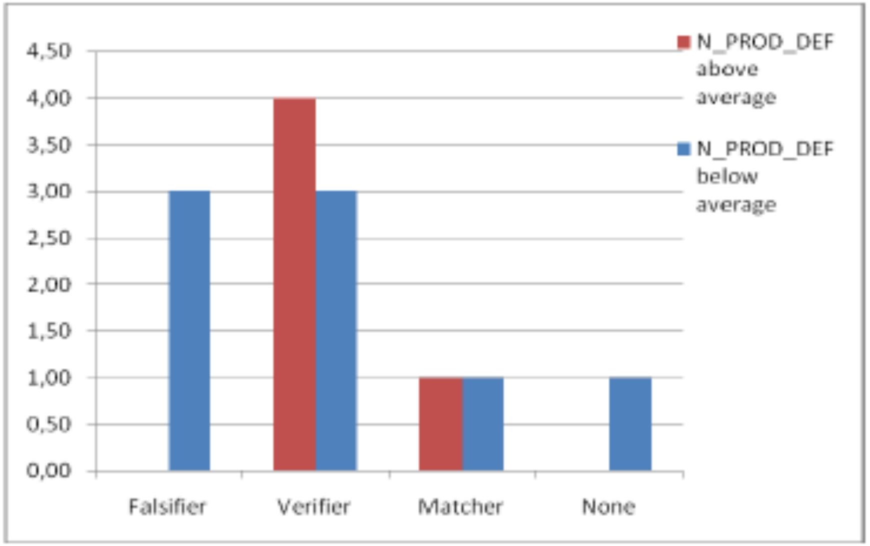
➢Two metrics for tester performance:

- ➢ N_{BUG}: Total number of bugs reported by each tester
- ➢ N_{PROD_DEF}: Total number of production defects that are overlooked by each tester during testing phase.



More Falsifiers & less Verifiers among testers with N_{BUG} value below average.

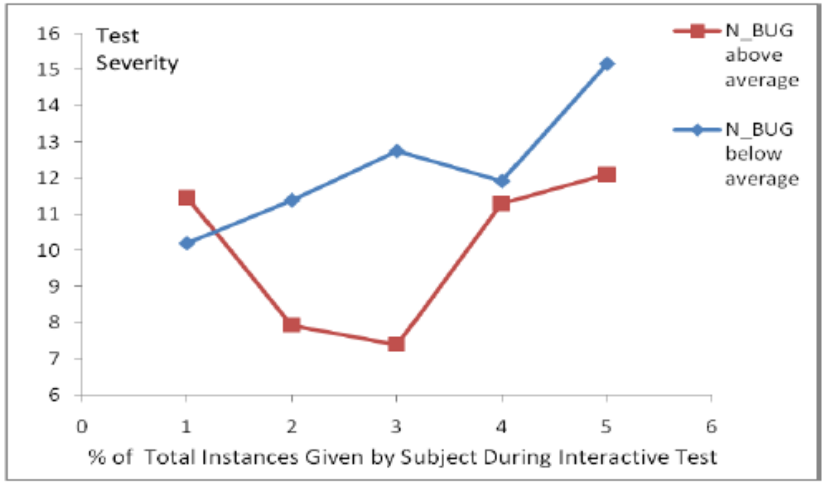
More Falsifiers & less verifiers among testers with N_{PROD_DEF} value below average.



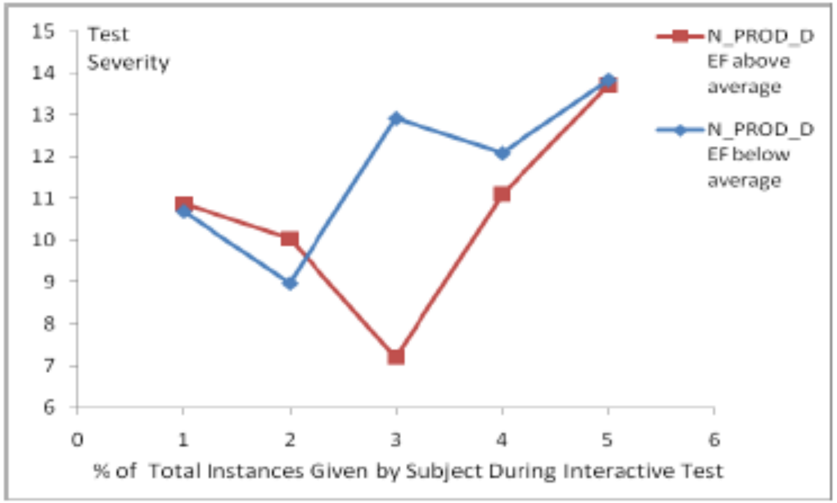
ANALYSES AND RESULTS

➢Analysis of the Effect of Confirmation Bias on Software Tester Performance

Testers who report bugs above average are less likely to follow a testing strategy



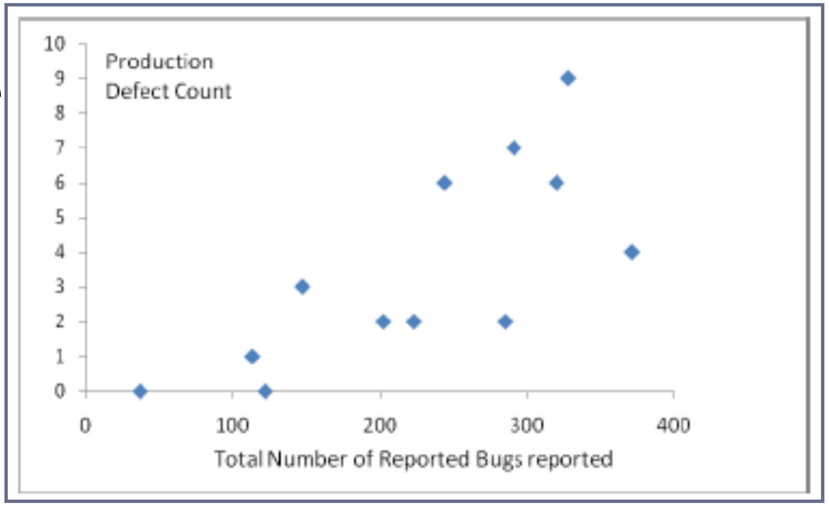
Vincent curves for test severities of testers with N_{PROD_DEF} values below & above average



ANALYSES AND RESULTS

➢Results for the Analysis of the Effect of Confirmation Bias on Software Tester Performance

➢Testers who report bugs below average are also the testers who cause less production defects → Spearman correlation value between N_{BUG} and N_{PROD_DEF} is 0.8234.



➢ Possible Reasons:

- ➢ Testers who report bugs might be assigned codes with high defect density.
- ➢ Bugs are not classified according to their severities → HENCE, large number of reported bugs does not necessarily mean that a significant portion of severe bugs have been detected.

THREATS TO VALIDITY

- Internal Validity:
  - Written & Interactive Tests are performed within a week for both groups and within any of the groups there was no event in between the confirmation bias tests that can affect subjects' performance.
  - However, problem may arise due to experimental conditions (i.e. Stress factor of company workers compared to graduate students)
- Construct Validity:
  - In order to avoid under-representing the construct,we used more than one single dependent & independent variables (i.e. we extracted metrics from written and interactive tests)
- External Validity:
  - Two datasets used & we continue to expand our dataset.
- Statistical Validity:
  - Bootstrapped Kolmogorov-Smirnov test used to validate our results statistically as we do not have prior knowledge about metric value distributions which are discontinuous.

CONCLUSIONS AND FUTURE WORK

- CONCLUSIONS:
- Confirmation bias levels of individuals who have been trained in logical reasoning and mathematical proof techniques are significantly lower.
- A significant effect of experience in software development/testing has not been observed. This implies that training in organizations is focused on tasks rather than personal skills.
- High levels of defect rates introduced by software developers are directly related to confirmation bias.
- High levels of confirmation bias among software testers are very likely to result in an increase in the number of production defects.

- FUTURE WORK:
- We plan to expand our dataset and replicate this study in other software development companies.
- We shall construct software defect prediction models that use confirmation bias metrics as people related metrics in addition to product and process metrics. It is highly probable that confirmation bias metrics will improve defect prediction performance.

THANK YOU

ANY QUESTIONS?