Even the Very Wise Cannot See All Ends: Many Facets of the Test Oracle Problem

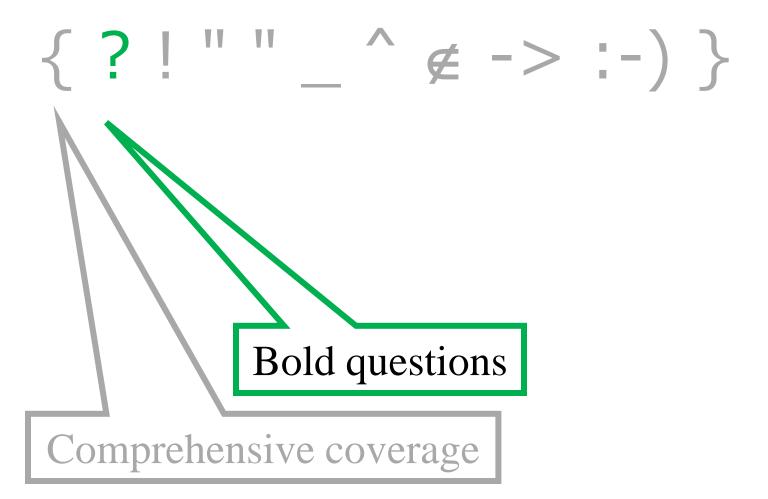


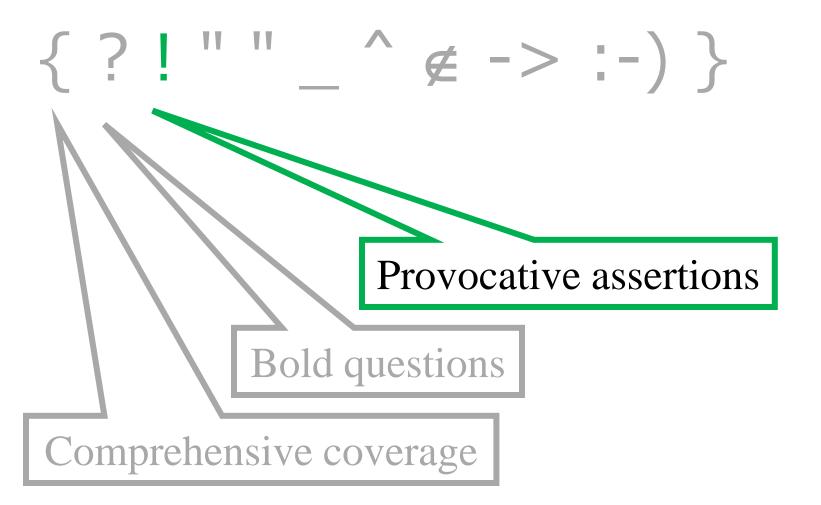
Prof. T.H. Tse

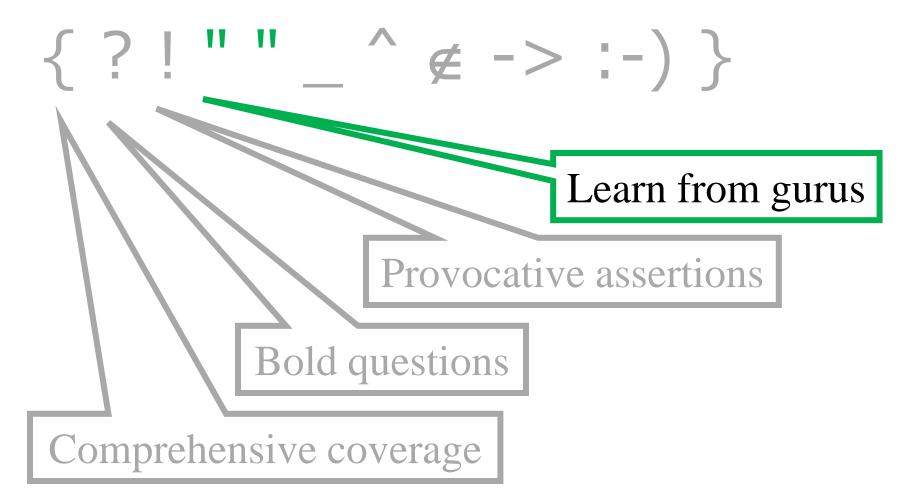
The University of Hong Kong Pokfulam, Hong Kong

{ ? ! " " _ ^ ∉ -> :-) }

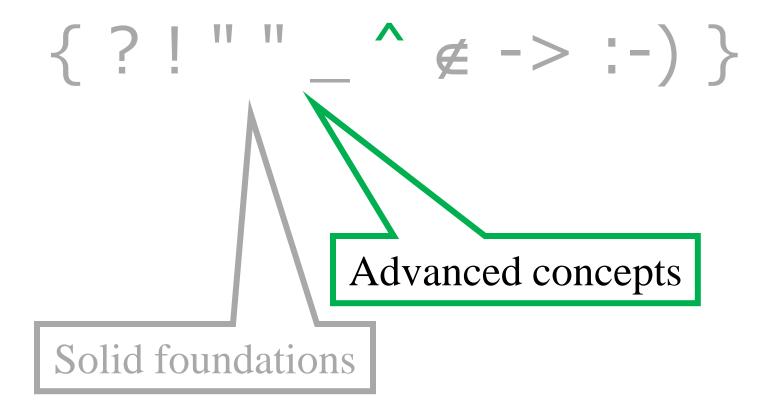
$\left\{ \begin{array}{c} ? \\ ! \\ " \\ - \end{array} \right\} \stackrel{}{\not \in} \left\{ \begin{array}{c} - \\ - \\ - \\ \end{array} \right\}$ Comprehensive coverage

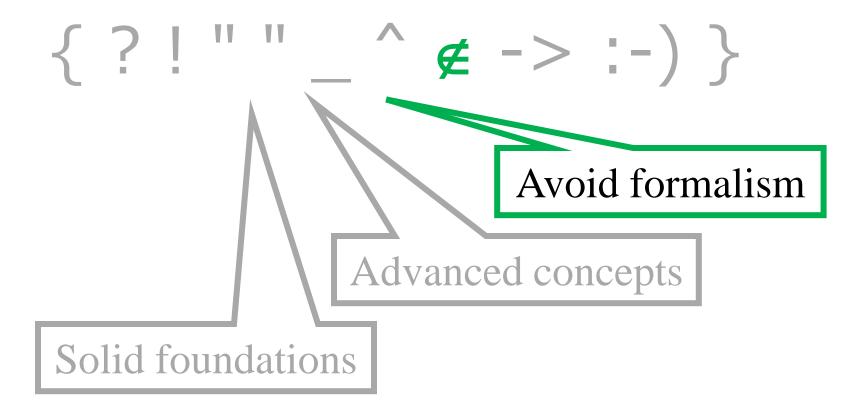


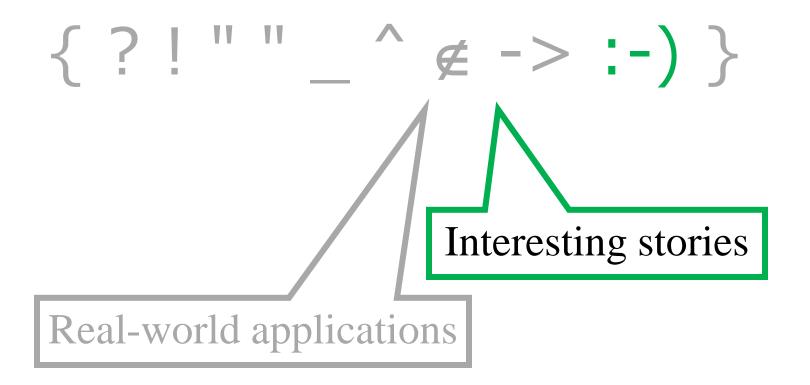


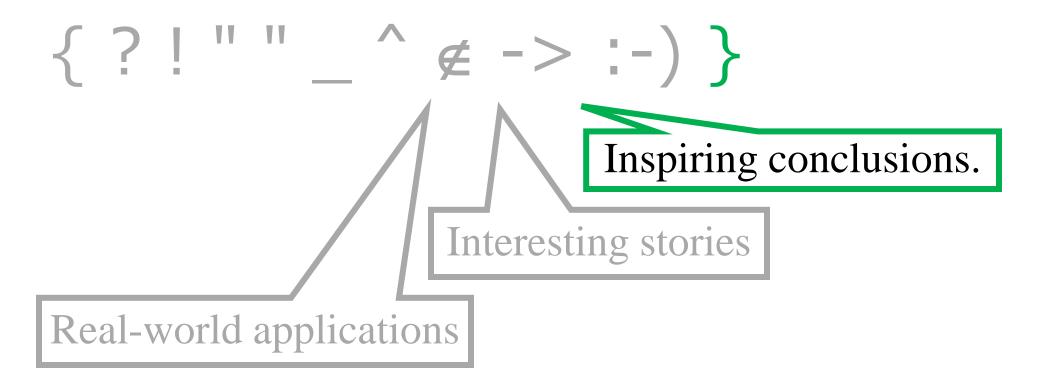


$\left\{ \begin{array}{c} ? \\ ! \\ \\ \end{array} \right\} \stackrel{}{=} \left\{ \begin{array}{c} - \\ - \\ \end{array} \right\}$ Solid foundations









- Formula for a successful keynote speech
- A successful keynote speech.

Background

- Background
- Many facets of the test oracle problem

- Background
- Many facets of the test oracle problem
 - Expected outcome
 - = actual execution result

- Background
- Many facets of the test oracle problem
 - Expected outcome
 - = actual execution result
 - Expected outcome
 actual execution result

- Background
- Many facets of the test oracle problem
 - Expected outcome
 - = actual execution result
 - Expected outcome
 actual execution result
 - Expected outcome
 - = actual execution result

- Background
- Many facets of the test oracle problem
 - Expected outcome
 - = actual execution result
 - Expected outcome
 actual execution result
 - Expected outcome

= actual execution result

Jungle of proposals

- Background
- Many facets of the test oracle problem
 - Expected outcome
 - = actual execution result
 - Expected outcome
 actual execution result
 - Expected outcome

- Jungle of proposals
- Empirical studies?

- Background
- Many facets of the test oracle problem
 - Expected outcome
 - = actual execution result
 - Expected outcome
 actual execution result
 - Expected outcome

- Jungle of proposals
- Empirical studies?
- What do other researchers do?

- Background
- Many facets of the test oracle problem
 - Expected outcome
 - = actual execution result
 - Expected outcome
 actual execution result
 - Expected outcome

- Jungle of proposals
- Empirical studies?
- What do other researchers do?
- Trim the tree or tame the forest? •

- Background
 Many facets of thest oracle problem
 Expected outcome
 actual execution result
 - Expected outcome
 actual execution result
 - Expected outcome

- Jungle of proposals
- Empirical studies?
- What do other researchers do?
- Trim the tree or tame the forest?



• Testing of object-oriented software

- Black and White [ACM TOSEM 1998]
- TACCLE [ACM TOSEM 2001]
- VITAMIN [Communications of the ACM 2007]
- Equality to Equals and Unequals [*IEEE TSE* 2013].



• Testing pervasive software

- [COMPSAC 2004 best paper]
- [*FSE* 2006]
- [*ICSE* 2008].



Testing services computing

- [*ICSE* 2008b]
- [*FSE* 2009]
- [WWW 2009]
- [IEEE TSC 2015 spotlight paper]
- [*IEEE TSC* 2015b].



Testing based on formal specifications

• Tabular Expressions [*IEEE TSE* 2011]



- *Testing based on formal specifications* Tabular Expressions [*IEEE TSE* 2011]
- *Testing based on informal specifications* CHOC'LATE [*IEEE TSE* 2003]
 [*Communications of the ACM* 2010]
 - DESSERT [*IEEE TSE* 2012].



Spectrum-based fault localization

- [COMPSAC 2008 best paper]
- [*FSE* 2009b]
- [IEEE Computer 2012]



Spectrum-based fault localization

- COMPSAC 2008 best paper]
- [*FSE* 2009b]
- [IEEE Computer 2012]

Debugging of concurrent systems

- [Information Sciences 2012]
- [ISSTA 2012].



Integration of testing, debugging, and proving

- [COMPSAC 2009 best paper]
- [QSIC 2011 best paper]
- [*IEEE TSE* 2011b].



Dr W.K. Chan, Associate Professor, City University of Hong Kong

Dr Zhenyu Zhang, Associate Professor, Institute of Software, Chinese Academy of Sciences

Dr Bo Jiang, Associate Professor, Beihang University

Dr Lijun Mei, IBM Research — China

Dr Ke Zhai, Goldman Sachs.



Dr W.K. Chan, Associate Professor, City University of Hong Kong

> Dr Yan Cai, Associate Professor, Institute of Software, Chinese Academy of Sciences.

Dr Zhenyu Zhang, Associate Professor, Institute of Software, Chinese Academy of Sciences

Dr Bo Jiang, Associate Professor, Beihang University

Dr Lijun Mei, IBM Research — China

Dr Ke Zhai, Goldman Sachs

- Background
- Many facets of the test oracle problem
 - Expected outcome
 - = actual execution res
 - Expected outcome
 actual execution result
 - Expected outcome

- Jungle of proposals
- Empirical studies?
- What do other researchers do?
- Trim the tree or tame the forest?

The Test Oracle Problem

Even the Very Wise Cannot See All Ends: Many Facets of the Test Oracle Problem



Prof. T.H. Tse The University of Hong Kong Pokfulam, Hong Kong

Present 20 years of work in 45 minutes

Many Facets of the Test Oracle Problem

Even the Very Wise Cannot See All Ends

Many Facets of the Test Oracle Problem



Prof. T.H. Tse The University of Hong Kong Pokfulam, Hong Kong

J.R.R. Tolkien The Lord of the Rings



Many Facets of the Test Oracle Problem

Even the Very Wise Cannot See All Ends

Many Facets of the Test Oracle Problem



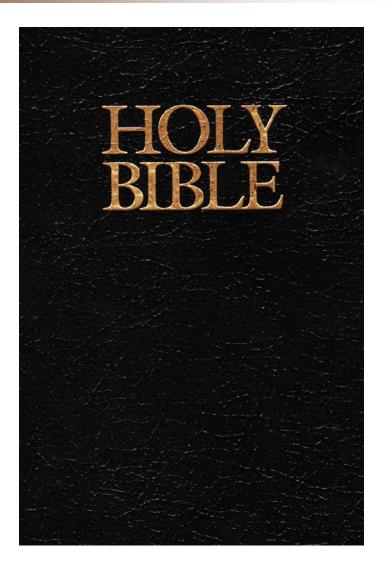
Prof. T.H. Tse The University of Hong Kong Pokfulam, Hong Kong J.R.R. Tolkien The Lord of the Rings



- Rawlinson and Bosworth Professor of Anglo-Saxon, University of Oxford (1925–1945)
- Merton Professor of English Language and Literature, University of Oxford (1945–1959).

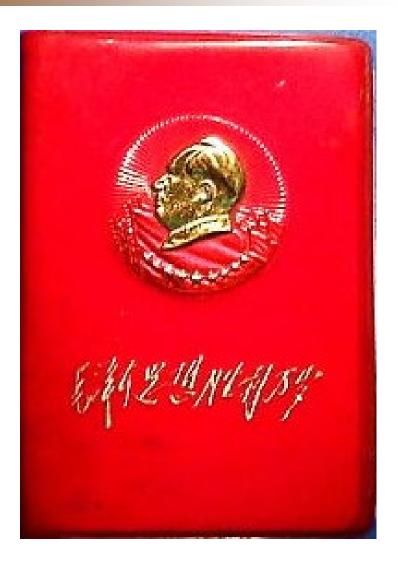
Oracle

 A message supposedly from God's inspiration, usually given by a priest



Oracle

- A message supposedly from God's inspiration, usually given by a priest
- An utterance of deep import or wisdom; an opinion or declaration regarded as authoritative and infallible; undeniable truth.



Oracle

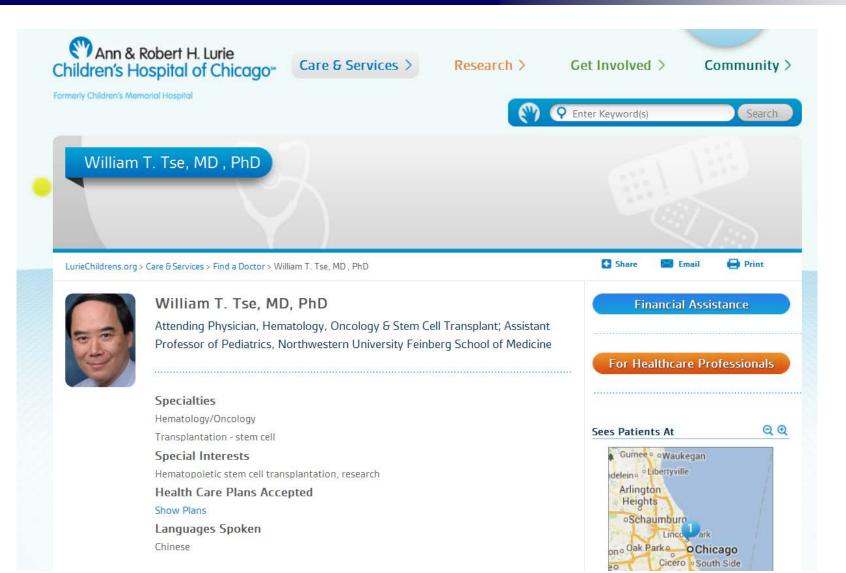
- A message supposedly from God's inspiration, usually given by a priest
- An utterance of deep import or wisdom; an opinion or declaration regarded as authoritative and infallible; undeniable truth.



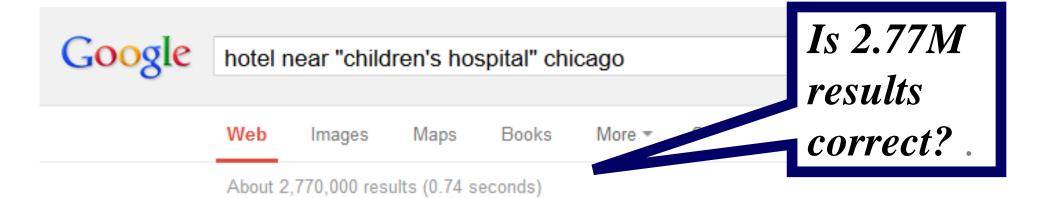
Test Oracle

- ◆ A *test oracle* is a mechanism to check whether
 - expected outcome according to the specification
 = actual result of executing the implementation.

Real-Life Search Engine Example



Real-Life Search Engine Example



Hotels near Lurie Children's Hospital of Chicago Chicago

www.hotels-rates.com > ... > Illinois Hotels > Chicago Illinois Hotels [] Hotels 1 - 25 of 75 - Browse and book hotels near Hotels in Lurie Children's Hospital of Chicago Chicago. An online hotel reservation directory with instant email ...

Hotel Discounts - University of Chicago Comer Children's Hospital www.uchicagokidshospital.org > Visiting Us -Chicago area hotels that offer discounts to Comer Children's Hospital patients and ... House near the University of Chicago Medicine Comer Children's Hospital

Many Facets of the Test Oracle Problem

Challenge 1: Expected outcome = actual execution result

sin 0.9876



sin 0.9876

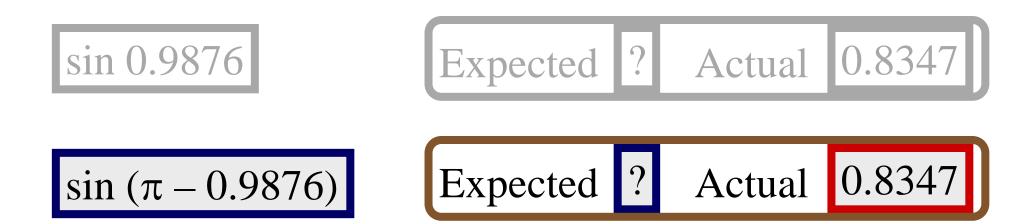


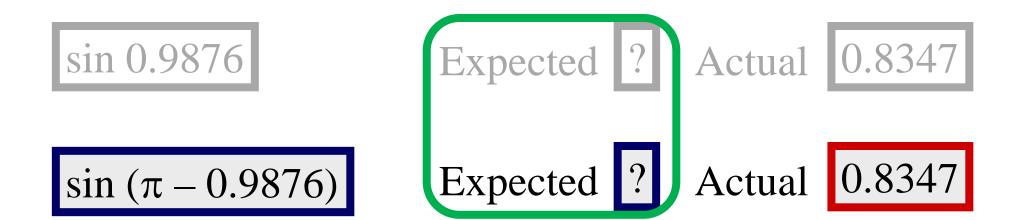
 Cannot be verified because we do not know what to expect

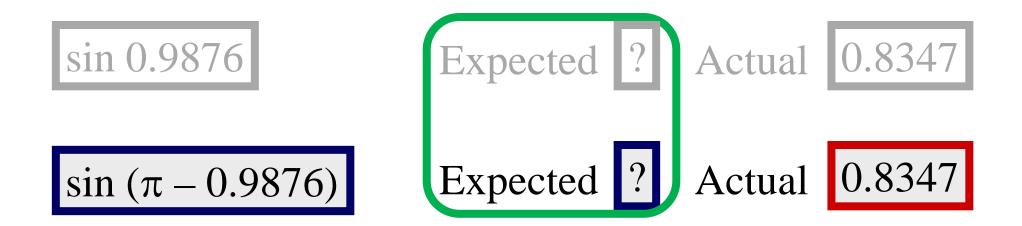
- Cannot be verified because we do not know what to expect
- Take a follow-up test case:

sin 0.9876

sin (
$$\pi$$
 – 0.9876) Expected ? Actual 0.8347

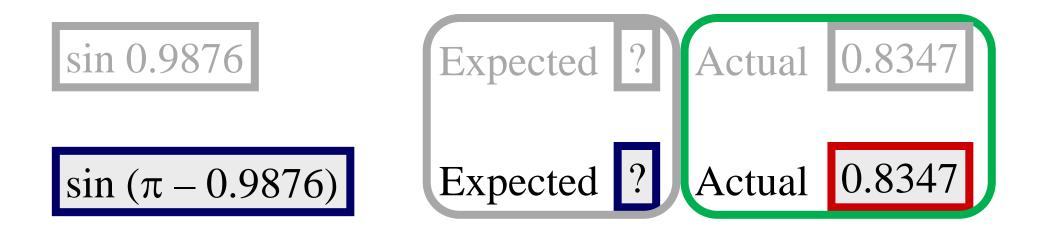






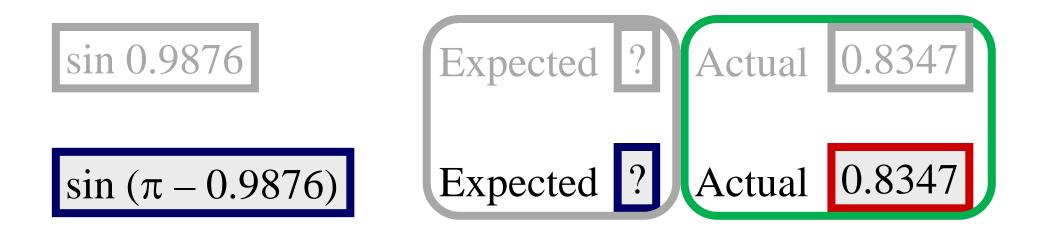
Expected metamorphic relation

 $\sin 0.9876 = \sin (\pi - 0.9876)$



Expected metamorphic relation

 $\sin 0.9876 = \sin (\pi - 0.9876)$



Expected metamorphic relation

Actual relation

 $\sin 0.9876 = \sin (\pi - 0.9876)$

$$0.8347 = 0.8347$$

Real-Life Search Engine Example Apply Metamorphic Testing



Metamorphic Relation: More refined search should produce fewer number of entries.



Real-Life Search Engine Example



Ann and Robert H. Lurie Children's Hospital of Chicago in ...



查看此网页的中文翻译, 请点击 <u>翻译此页</u> Lurie Children's Hospital of Chicago in Chicago, IL is ranked national ly in 10 pediatric specialties. Ann and Robert H... health.usnews.com/best... ▼ - <u>百度快照</u>

The University of Chicago Medicine Comer Children's Hospital

查看此网页的中文翻译, 请点击 <u>翻译此页</u> The University of Chicago Medicine Comer Children's Hospital is a state-of-the-art children's hospital. Our teams of pediatric experts use advanced ... www.uchicagokids**hospit**... ▼ - <u>百度快照</u> - <u>评价</u>

Refine to **Real-Life Search Engine** "children's hospital of chicago" hotel near "children's hospital of chicago" 网页 贴吧 音乐 图片 视频 地图 更多» 新闻 知道 文库 More entries 百度为您找到相关结果约9,790,000个 *indicate failure* 您可以**仅**查看:英文结果 in search engine. Ann & Robert H. Lurie Children's Hospital of Chicag 查看此网页的中文翻译, 请点击 翻译此页 Lurie Children's, formerly Children's Memorial Hospital, is one of the top pediatric providers in the Midwest, treating Chicago's kids with the highest... www.luriechildrens.org/ - 百度快照 - 评价

Ann & Robert H. Lurie Children's Hospital of Chicago Jobs

查看此网页的中文翻译,请点击 翻译此页

Lurie Children's Hospital of Chicago is a unique place where children, families, physicians an

d staff come together to make sure our patients get the ...

luriechildrensjobs.sil... ▼ - 百度快照 - <u>评价</u>

Real-Life Search Engine Example Human Issues

- Microsoft Research selected our project for a Virtual Earth award
- Google asked us to submit a bug report
- Each of my brothers has 2 doctoral degrees
- Or I have only one doctoral degree.

Metamorphic Testing Other Real-World Applications

- Services computing
- Ubiquitous computing
- Concurrent systems
- Graphic applications
- Numerical programs.

Metamorphic Testing Recent Track Records

Compilers

 Based on metamorphic testing, an "equivalence modulo inputs" technique identified 147 unique faults in GCC and LLVM using one metamorphic relation

Metamorphic Testing Recent Track Records

Compilers

 Based on metamorphic testing, an "equivalence modulo inputs" technique identified 147 unique faults in GCC and LLVM using one metamorphic relation

Siemens Suite

 Detected 3 new faults in Siemens suite after its long history of test case studies.

Many Facets of the Test Oracle Problem

*Challenge 2:*Expected outcome = actual execution result

Many Facets of the Test Oracle Problem

Challenge 2:

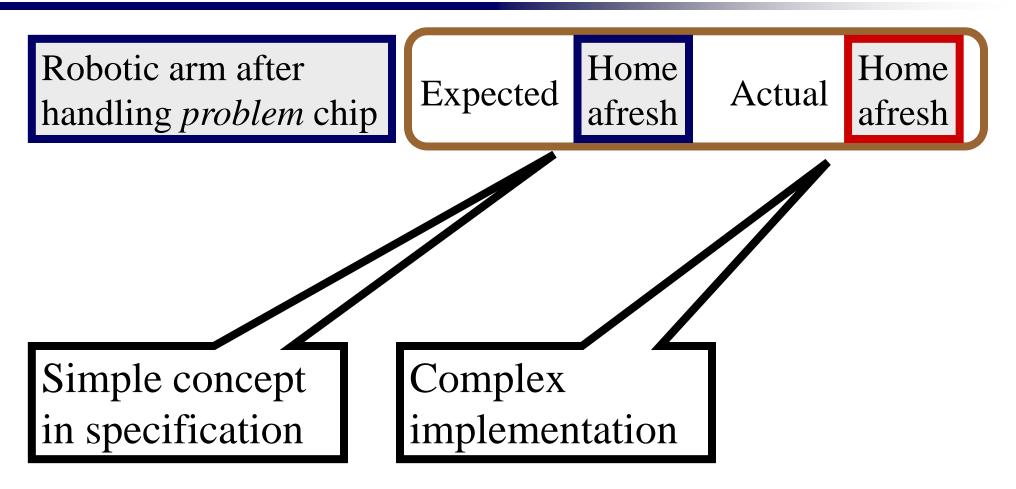
- Expected outcome = actual execution result
- Consider the testing of real-life object-oriented software.

Testing of Object-Oriented Software Our Real-Life Experience

 Technology-transfer project for ASM, the world's largest supplier of assembly and packaging equipment for the semiconductor industry.



Testing of Object-Oriented Software



Many Facets of the Test Oracle Problem

Challenge 2:

- Expected outcome = actual execution result
- Consider the testing of real-life object-oriented software

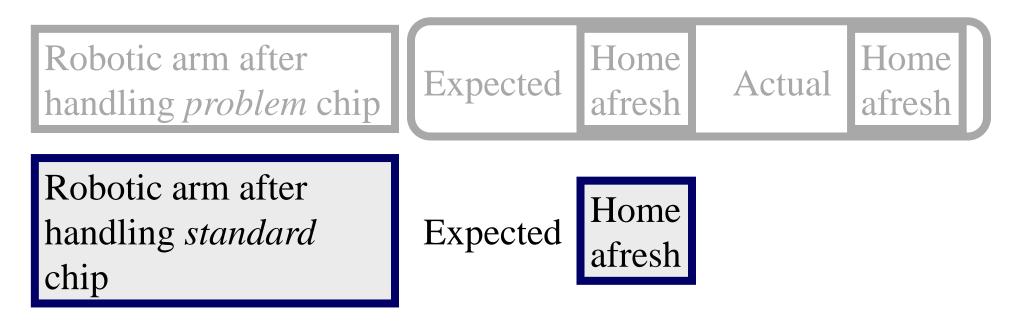
Cannot define "=" at two different levels of abstraction.

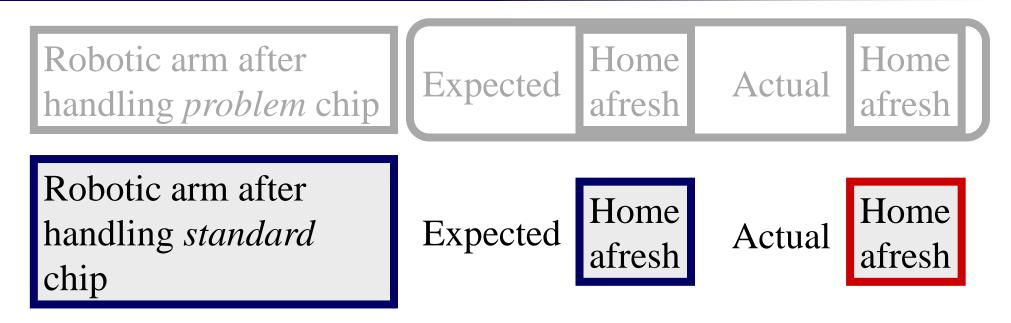


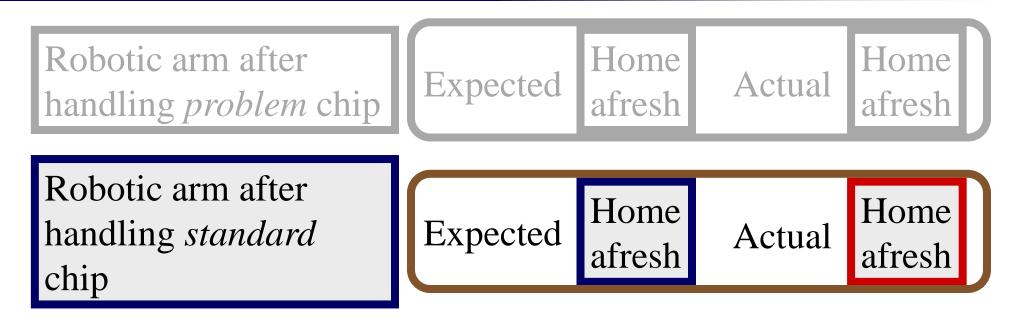
Robotic arm after handling *problem* chip Expected Home afresh

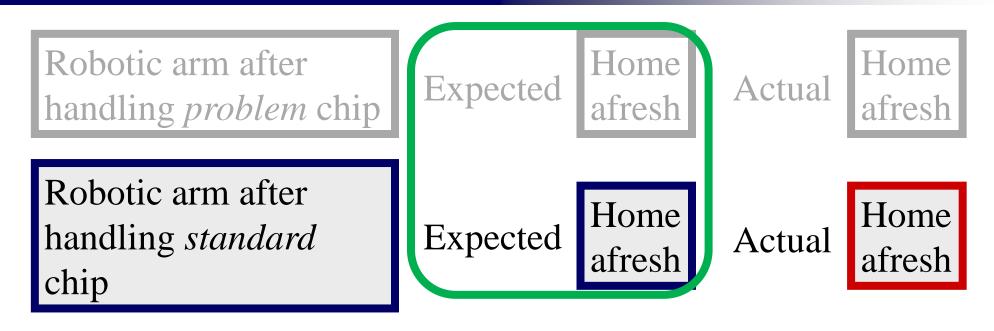
Robotic arm after handling *standard* chip Home

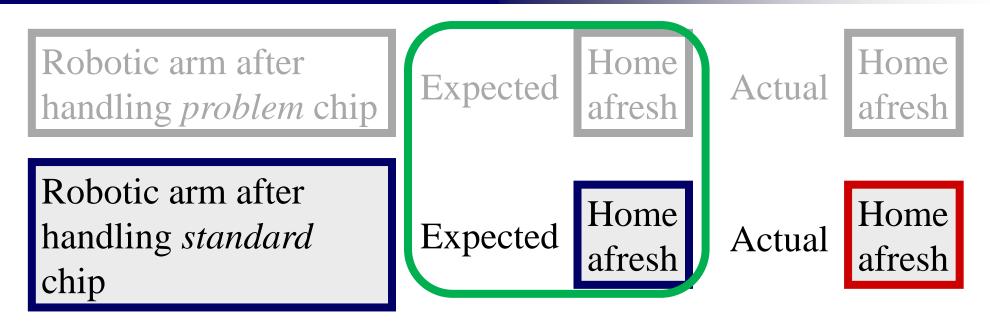
Actual







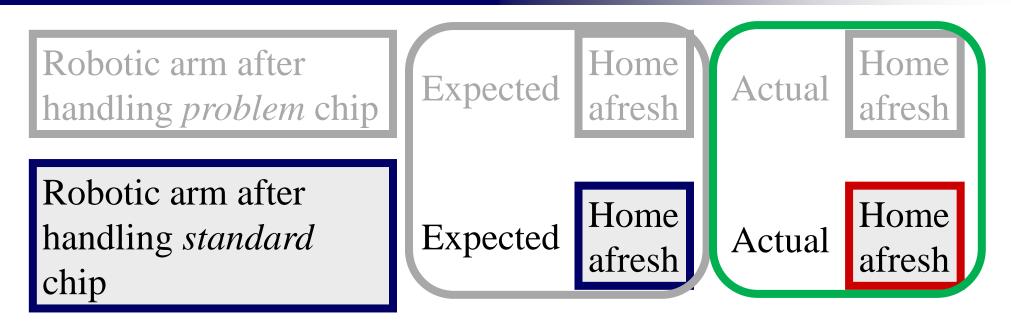




Expected relation

Equivalent

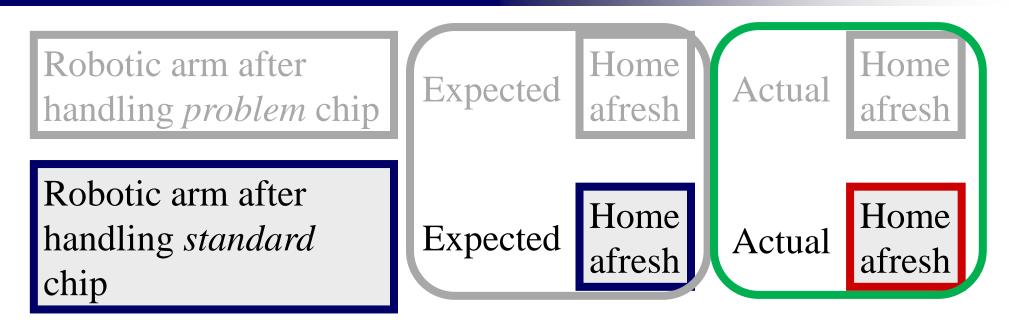
Testing of Object-Oriented Software Mimic Metamorphic Testing?



Expected relation

Equivalent

Testing of Object-Oriented Software Mimic Metamorphic Testing?



Expected relation

Actual relation

Equivalent

Equivalent

Target of OO Software Testing

- An implementation *P* is *correct with respect to the specification Sp* if and only if
 - For any pair of *equivalent* sequences of operations in *Sp*, the actual objects resulting from *P* must be *equivalent*.

Many Facets of the Test Oracle Problem

	י ט י ט <i>I</i>	•							thke	eyExp1.docx - M	Vicrosoft Word	đ								- 0
9	Home Insert	Page Layout	References	Mailings	Review V	iew Devel	oper													
	∦ Cut I≧ Copy I Format Painter	Times New Roman B I U - ab						╪ × . 2. ¶ ≡ . <u>^</u> . ⊞ .		AaBbCcDc 1 No Spacing			AaBt		AaBbCcDd Subtle Emp		AaBbCcDc Intense Emp	AaBbCcDa Quote	Change Styles *	라 Find ~ ⁸ ac Replace 장 Select ~
	Clipboard 🕞		Font		Gi.		Paragraph	n 5						Styles					6	Editing
]	1			· <u>1</u> ·			2		• 3 • •		• • 4			5 · ·		6		 7	<u>ki i i</u>	

T.H. Tse is an Honorary Professor in Computer Science at The University of Hong Kong.

n 9- C 1	r 🔍 🖛							thke	yExp2.docx - N	/licrosoft Word	I									- 0	p
Home Inser	t Page Layout	References Ma	ailings Review	View	Developer																
Cut Copy Paste Format Painter	Times New Roman B I U - al	~ 27.5 ~ ,	A A A A A					AaBbCcDc 1 Normal	AaBbCcDc 11 No Spacing		AaBbCcD Heading 2	AaBt	AaBbCcDa Subtitle	AaBbCcDd Subtle Emp		AaBbCcDc Intense Emp			← A ← Change Styles ←	and Find → abac Replace A Select	ace
Clipboard	i l	Font		6	Pa	aragraph	Gi.						Styles						5	Editing	9
	· · · · ·	and the second sec	1 1 1 1	1.1	• • 2			3 · ·		• • 4			5 1 1	1997 - 1997 - 1997 1997 - 1997 - 1997 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 19	• • 6		1	· 7 · ·	Δ. · · ·		

T.H. Tse is an Honorary Professor in Computer Science at The University of Hong Kong. He is a Steering Committee Chair of the IEEE International Conference on Software Quality, Reliability & Security.

thkevExp3.docx - Microsoft Word

1 Normal 1 No Spacing Heading 1

AaBbCcDe AaBbCcDe AaBbCcI AaBbCcD AaBt AaBbCcDe AaBbCcDe

Title

Subtitle

Styles

Subtle Emp... Emphasis

Heading 2

T.H. Tse is an Honorary Professor in Computer
Science at The University of Hong King. He is
a Steering Committee Chair of the IEEE
International Conference on Software Quality,
Reliability & Security. He was on the search
committee for the editor-in-chief of IEEE
Transactions on Software Engineering in 2013.

") - U I) **-**

Paste

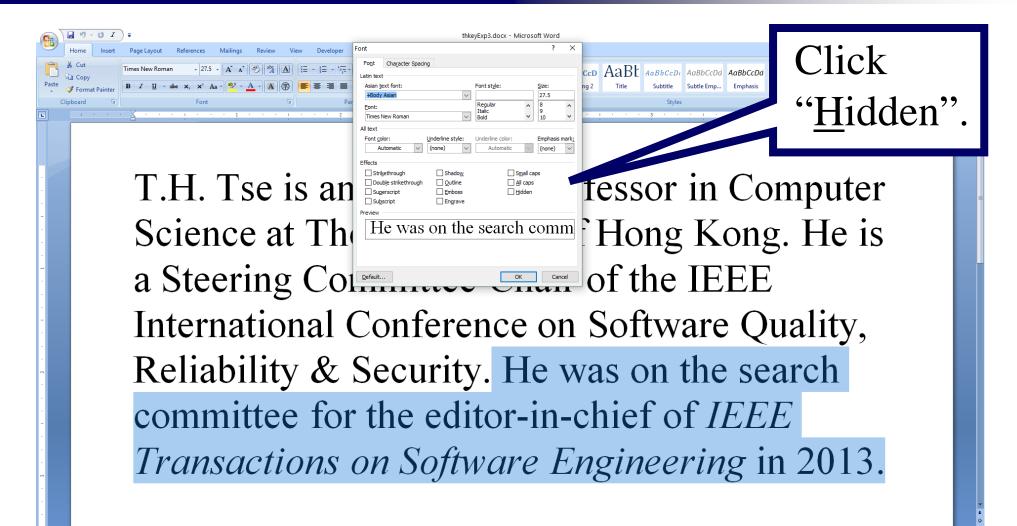
Clipboard

Page Layout References Mailings Review View Developer

Times New Roman 🔹 27.5 👻 🗛 🖓 🕸 🗛 🗄 🛨 🗄 🐨 🛱 🛣 🖓 📲

Call this

object O.



	I 9 - C I	₹ thkeyExp2.docx - Microsoft Word	
Paste	Home Insert	Page Layout References Mailings Review View Developer Times New Roman 277.5 A A D R A B C D A B C D A B C D A B C D A B C D A B C D A A B C D C D A A B C D A A A B C D A A A B C D A A A B C D A A A B C D A A A A C D A A A A A A A A A A A A	
		T.H. Tse is an Honorary Professor in Coputer Science at The University of Hong King. He is a Steering Committee Chair of the IEEE International Conference on Software Quality, Reliability & Security.	

What is Object Equivalence? Output from Object H

T.H. Tse is an Honorary Professor in ComputerScience at The University of Hong Kong.He is a Steering Committee Chair of the IEEEInternational Conference on Software Quality,Reliability & Security.

thkevExp3.docx - Microsoft Word Take Page Layout References Mailings Review View Developer AaBbCcDc AaBbCcDc AaBbCcI AaBbCcD AaBt AaBbCcDc AaBbCcDd AaBbCcDd AaBbCcDd AaBbCcDd 1 Normal 1 No Spacing Heading 1 Subtle Emp... Emphasis Intense Emp Heading 2 Subtitle object O again. T.H. Tse is an Honorary Professor in *C* Science at The University of Hong Vong. He is a Steering Committee Chair of the IEEE International Conference on Software Quality, Reliability & Security. He was on the search committee for the editor-in-chief of *IEEE* Transactions on Software Engineering in 2013.

thkevExp3.docx - Microsoft Word

1 No Spacing Heading 1

Heading 2

Title

AaBbCcDc AaBbCcDc AaBbCcI

1 Normal

AaBbCcD AaBt AaBbCcDi AaBbCcDd AaBbCcDd AaBbCcDc AaB

Subtle Emp...

Subtitle

Style

Emphasis Intense Emp...

T.H. Tse is an Honorary Professor in Computer				
Science at The University of Hong Vong. He is				
a Steering Committee Chair of the IEEE				
International Conference on Software Quality,				
Reliability & Security. He was on the search				
committee for the editor-in-chief of IEEE				
Transactions on Software Engineering in 2013.				

I) - (≥ I) =

Paste Cop

Clipboard

Times New Roman

Page Layout References Mailings Review View Developer

- 27.5 - A A 🕺 🥙 🖄 A 三・三・三・窪 建 🛣 😣 ¶

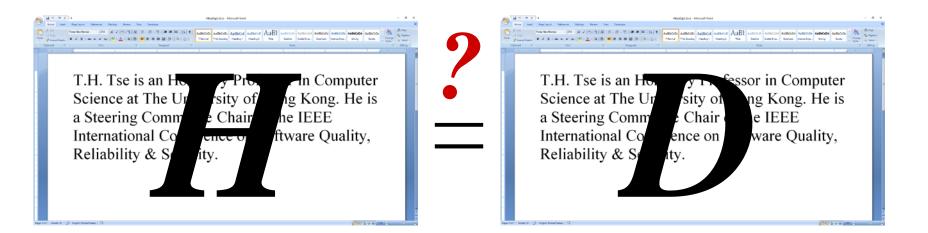
Delete.

	9-	(° I	→ thkeyExp2.docx - Microsoft Word	
Paste	Home	Insert t Painter	Page Layout References Mailings Review View Developer Times New Roman 27.5 A <td< th=""><th></th></td<>	
			Font 5 Paragraph 5 Styles Object D	
			T.H. Tse is an Honorary Professor in Coputer Science at The University of Hong King. He is a Steering Committee Chair of the IEEE International Conference on Software Quality, Reliability & Security.	

What is Object Equivalence? Output from Object D

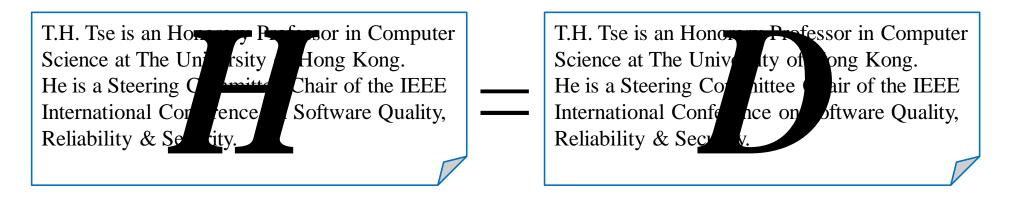
T.H. Tse is an Honorary Professor in ComputerScience at The University of Hong Kong.He is a Steering Committee Chair of the IEEEInternational Conference on Software Quality,Reliability & Security.

What is Object Equivalence? Attributive Equivalence



 Two objects will be *attributively equivalent* if they have the exactly the same *visible attributes*

What is Object Equivalence? Attributive Equivalence

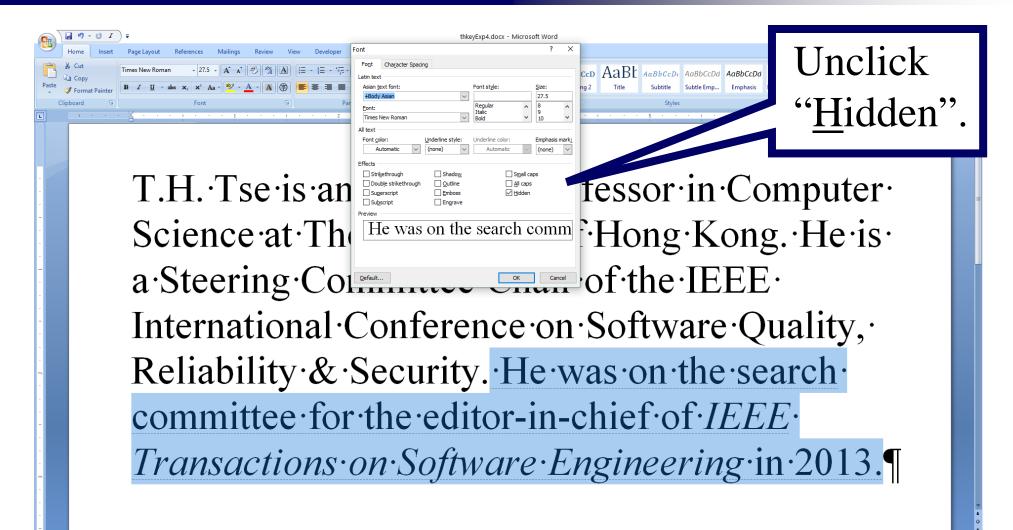


- Simple to test
- *But* the definition is too weak to be useful ...
 Why?.

3	Image: Second	
aste	rome inset page Layout References mainings Review View Developer	Take
C	Clipboard Font Font <th>object</th>	object
	T.H. Tse is an Honorary Professor in Com	H again
	Science at The University of Hong Kong	(with
	a Steering Committee Chair of the	hidden
	International Conference on Service Qua	text).
	Reliability & Security.	

Image: Second	
Home Inset Page Layout References Mailings Review View Developer → Cut → Cut → Copy → Cut → Cu	
Paste	strong

T.H. Tse is an Honorary Professor in Computer Science at The University of Hong Kong. He is a Steering Committee Chair of the IEEE International Conference on Software Quality, Reliability & Security.



thkeyExp3.docx - Microsoft Word					
Home Insert Page Layout References Mailings Review View Developer # Cut Times New Roman • 27.5 • A * A * *** *** #** A # AaBbCcDt AaBbCcDt	We get an				
Format Parameter Styles Clipboard Styles I I	object				
TIL Tas is an Honorery Drofossor in Co	that				
T.H. Tse is an Honorary Professor in Co Science at The University of Hong Kong	$h \rho h \sigma \eta \rho \sigma$				
a Steering Committee Chair of the	like the				
International Conference on Sec	original				
Reliability & Security. He was on the se	-				
committee for the editor-in-chief of IEE	E				
Transactions on Software Engineering in 2013.					

	Image: Image	Take
Paste	T.H. Tse is an Honorary Professor in Com Science at The University of Hong Kons a Steering Committee Chair of the P International Conference on Science Of the P	object D again (with deleted
	Reliability & Security.	

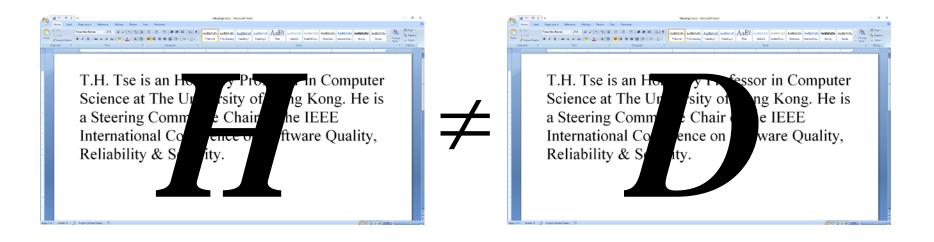
Image: Second	
Home Inset Page Layout References Mailings Review View Developer → Cut → Cut → Copy → Cut → Cu	
Paste	strong

T.H. Tse is an Honorary Professor in Computer Science at The University of Hong Kong. He is a Steering Committee Chair of the IEEE International Conference on Software Quality, Reliability & Security.

	🚽 🔊 - 😈 I	♥ thkeyExp2.docx - Microsoft Word	
Paste	Home Insert Cut Copy Format Painter	Page Layout References Mailings Review View Developer Times New Roman + 27.5 + A + 29 + A + 27 + 27 + A + 27 + 27 + A + 28 + 27 + 10 + 28 + 10 + 28 + 10 + 28 + 10 + 28 + 10 + 28 + 10 + 28 + 10 + 28 + 10<	We do not
L	Clipboard 🕞	Font I Paragraph I Styles •	get an
		T.H.·Tse·is·an·Honorary·Professor·in·Co	object that
-		Science at The University of Hong Ko	behaves
- - -		a·Steering·Committee·Chair·of·the	like the
-		International Conference on Servare Q	original <i>O</i> .
		Reliability & Security.	

What is Object Equivalence? Observational Equivalence

 Two objects will be *observationally equivalent* if they have the exactly the same *visible attributes* and *behavior*

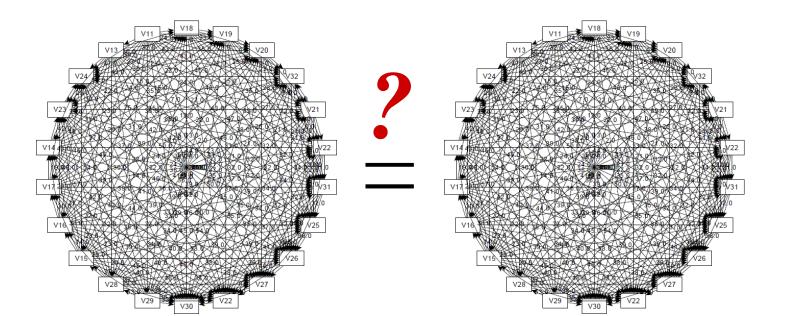


What is Object Equivalence? Observational Equivalence

- Ideal criterion in object-oriented software testing
- - Need to check an infinite combination of possible behavior for every test case.

UML State Machine Not Truly Suitable for OO Specification

- Cannot distinguish between visible and hidden attributes and behavior
- Cannot distinguish between attributive and observational equivalence



- A *ground term* is a sequence of operations
 - show(HKU QRS hide(TSE))
- An *axiom* is a rule that specifies the refinement of a term
 - hide(D D') = hide(D) hide(D')
 - show(D D') = show(D) show(D')
 - show(hide(D)) = show(D)
 - show(W) = W.

 A ground term may be transformed into another using axioms as left-to-right rewrite rules, until it reaches a *normal form*

 A ground term may be transformed into another using axioms as left-to-right rewrite rules, until it reaches a *normal form*

 A ground term may be transformed into another using axioms as left-to-right rewrite rules, until it reaches a *normal form*

show(HKU QRS hide(TSE))

→ show(HKU) show(QRS) show(hide(TSE))

$$show(D D') = show(D) show(D')$$

 A ground term may be transformed into another using axioms as left-to-right rewrite rules, until it reaches a *normal form*

- > show(HKU) show(QRS) show(hide(TSE))
- \rightarrow show(HKU) show(QRS) show(TSE)

 A ground term may be transformed into another using axioms as left-to-right rewrite rules, until it reaches a *normal form*

- show(HKU) show(QRS) show(hide(TSE))
- → show(HKU) show(QRS) show(TSE)
- \rightarrow HKU QRS TSE



 A ground term may be transformed into another using axioms as left-to-right rewrite rules, until it reaches a *normal form*

- show(HKU) show(QRS) show(hide(TSE))
- show(HKU) show(QRS) show(TSE)
- \rightarrow HKU QRS TSE



Algebraic Specification Denotational Semantics

- Mathematical meaning
- Like simultaneous equations

2x + 3y = 45x + 6y = 7

- "=" means "equals"
- The order of the equations is not important.

Algebraic Specification Operational Semantics

- How the statements should be executed
- Like C programs
 j = 1;
 j = j + 2;
 - Replace the variable on the left-hand side by the value on the right-hand side.

Canonical Algebraic Specification

 An algebraic specification is *canonical* if every sequence of rewrites of a ground term produces a unique normal form

Canonical Algebraic Specification

- An algebraic specification is *canonical* if every sequence of rewrites of a ground term produces a unique normal form
 - show(HKU) show(QRS) show(hide(TSE))
 - \rightarrow show(HKU) show(QRS) show(TSE)
 - \rightarrow HKU QRS TSE

Canonical Algebraic Specification

 An algebraic specification is *canonical* if every sequence of rewrites of a ground term produces a unique normal form

show(HKU) show(QRS) show(hide(TSE))

- → show(HKU) show(QRS) show(TSE)
- \rightarrow HKU QRS TSE

show(HKU) show(QRS) show(hide(TSE))

- → HKU QRS show(hide(TSE))
- \rightarrow HKU QRS TSE

Canonical Algebraic Specification

• For a *canonical* specification, the operational semantics agrees with the denotational semantics.

Target of OO Software Testing According to Real-World Practice

• An implementation *P* is *correct with respect to the specification Sp* if and only if

Target of OO Software Testing According to Real-World Practice

- An implementation *P* is *correct with respect to the specification Sp* if and only if
 - *P* satisfies the set (*AE*) of all *attributively equivalent* pairs of ground terms in *Sp*

For example, **hide(TSE)** and **delete(TSE)** are attributively equivalent.

- An implementation *P* is *correct with respect to the specification Sp* if and only if:
 - Equivalence Criterion
 P satisfies the set (OE) of all observationally equivalent pairs of ground terms in Sp

- An implementation *P* is *correct with respect to the specification Sp* if and only if:
 - *Equivalence Criterion P* satisfies the set (*OE*) of all *observationally equivalent* pairs of ground terms in *Sp*
 - Non-Equivalence Criterion

P satisfies the set (*OE'*) of all *observationally nonequivalent* pairs of ground terms in *Sp*.

For example, **hide(TSE)** and **delete(TSE)** are observationally nonequivalent.

Targets of OO Software Testing

AE	?	<i>OE U OE'</i>
Real-world short-cut, but too weak to be useful	A jungle of proposals by various researchers.	Real-world correctness, but <i>very very very very</i> <i>very very very very very</i> difficult to test.

Target of OO Software Testing

According to Bernot et al. (1991), Le Gall and Arnould (1996), Machado (1998, 2000), Machado and Sannella (2002), Aiguier et al. (2006)

- An implementation *P* is *correct with respect to the specification Sp* if and only if
 - *P* satisfies the set (*GI*) of all *ground instances* of every axiom in *Sp*

Target of OO Software Testing

According to Doong and Frankl (1994), Gaudel (1995), Zhu (2003)

- An implementation *P* is *correct with respect to the specification Sp* if and only if:
 - Equivalence Criterion

P satisfies the set (RP) of all "equivalent" ground pairs such that one can be rewritten to the other using axioms in Sp as left-to-write rewrite rules

Target of OO Software Testing

According to Doong and Frankl (1994), Gaudel (1995), Zhu (2003)

- An implementation *P* is *correct with respect to the specification Sp* if and only if:
 - **Equivalence** Criterion

P satisfies the set (*RP*) of all "equivalent" ground pairs such that one can be rewritten to the other using axioms in *Sp* as left-to-write rewrite rules

Non-Equivalence Criterion

P satisfies the set (*RP'*) of all "nonequivalent" ground pairs.

- An implementation *P* is *correct with respect to the specification Sp* if and only if
 - *P* satisfies the set (*FP*) of all *fundamental pairs* in *Sp*

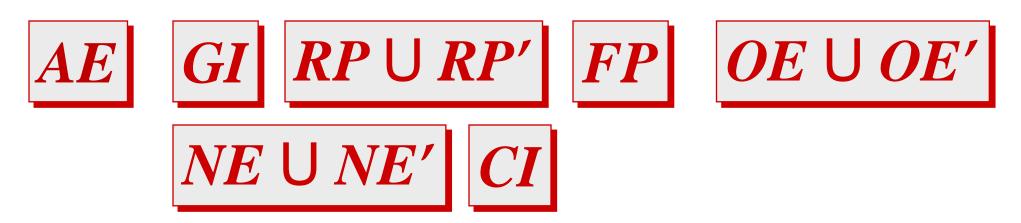
- An implementation *P* is *correct with respect to the specification Sp* if and only if:
 - Equivalence Criterion
 P satisfies the set (NE) of all normally equivalent ground pairs in Sp

- An implementation *P* is *correct with respect to the specification Sp* if and only if:
 - Equivalence Criterion
 P satisfies the set (NE) of all normally equivalent ground pairs in Sp
 - Non-Equivalence Criterion
 P satisfies the set (NE') of all normally nonequivalent ground pairs in Sp

Target of OO Software Testing *According to Aiguier et al. (2006)*

- An implementation *P* is *correct with respect to the specification Sp* if and only if
 - *P* satisfies the set (*CI*) of all ground instances of every axiom in *Sp* that contains creators or constructors only.

Targets of OO Software Testing



• Comparisons through empirical studies?

- Research based on experimentation or observation to answer a specific question or to test a hypothesis
- Undue emphasis in software engineering, even in first-tier publication venues.

- Statistical significance does not mean research significance
 - A hypothesis may be very probable simply because it tells us nothing, or very little



Sir Karl Popper, Professor of Logic and Scientific Method, London School of Economics (1949–1969)

- Statistical significance does not mean research significance
 - A hypothesis may be very probable simply because it tells us nothing, or very little
 - A high degree of probability is therefore not an indication of "goodness"



Sir Karl Popper, Professor of Logic and Scientific Method, London School of Economics (1949–1969)

- Statistical significance does not mean research significance
 - A hypothesis may be very probable simply because it tells us nothing, or very little
 - A high degree of probability is therefore not an indication of "goodness"
 - It may be merely a symptom of low informative content.



Sir Karl Popper, Professor of Logic and Scientific Method, London School of Economics (1949–1969)

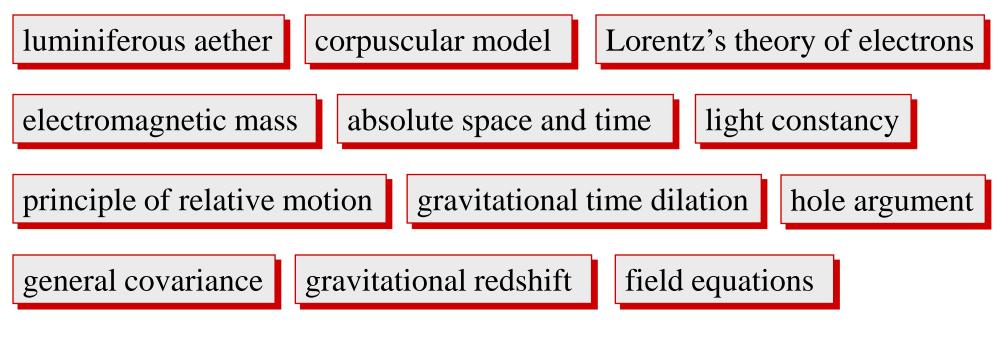
Presentation Outline

- Background
- Many facets of the test oracle problem
 - Expected outcome
 - = actual execution result
 - Expected outcome
 actual execution result
 - Expected outcome

= actual execution result

- Jungle of proposals
- Empirical studies?
- What do other researchers do?
- Trim the tree of tame the fore.

What Do Other Researchers Do? General Relativity



• A jungle of proposals

What Do Other Researchers Do? General Relativity

So many people today — and even professional scientists — seem to me like somebody who has seen thousands of trees but has never seen a forest.



Albert Einstein *The Collected Papers of Albert Einstein* Princeton University Press

What Do Other Researchers Do? **General Relativity**

Trim the individual trees? Tame the forest!

What Do Others Do? General Relativity						
luminiferous aether corpuscular model Lorentz's theory of electrons						
electromagnetic mass absolute space and time light constancy						
principle of relative motion gravitational time dilation hole argument						
general covariance gravitational redshift field equations						

• A jungle of proposals

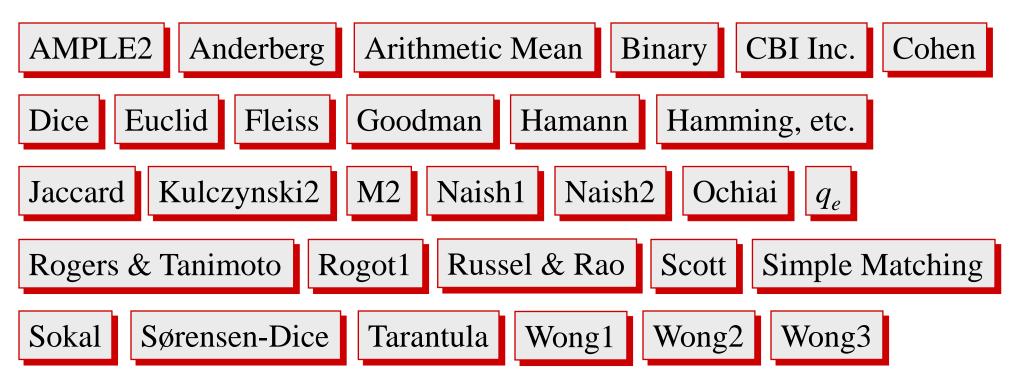
space time curvature



Albert Einstein

What Do Other Researchers Do? Spectrum-Based Fault Localization

Risk evaluation formulas



What Do Other Reseachers Do? Spectrum-Based Fault Localization

- Another jungle of proposals
- Comparisons through empirical studies in *IEEE TSE*, *ICSE*, and *FSE*.

What Do Other Researchers Do? **Spectrum-Based Fault Localization**

Trim the individual trees? Tame the forest!

What Do Others Do? **Spectrum-Based Fault Localization**

٠	Risk	evaluation	formulas
---	------	------------	----------

AMPLE2	Anderberg	Arithmetic Mean	Binary	CBI Inc.	Cohen		
Dice Euclid Fleiss Goodman Hamann Hamming, etc.							
Jaccard Kulczynski2 M2 Naish1 Naish2 Ochiai q_e							
Rogers & Tanimoto Rogot1 Russel & Rao Scott Simple Matching							
Sokal Sørensen-Dice Tarantula Wong1 Wong2 Wong3							



T.Y. Chen and Team

Spectrum-Based Fault Localization *According to T.Y. Chen and Team (2013)*

- Theoretical framework to compare risk evaluation formulas for single-fault programs
- No single formula can outperform the rest
- Among the formulas under study, only five are "maximal"
- Most of best-known formulas are not among them.

Test Case Generation Techniques *According to T.Y. Chen and Team (2008)*

- Prove that no test case generation technique can be better than random testing by more than 50%
 - in the absence of further information on possible locations of failure-causing inputs
- Adaptive random testing is close to this theoretic limit.

Partition Testing Techniques According to T.Y. Chen and Team (2000)

 Prove that proposed proportional sampling strategy is the only partition testing technique that ensures probability of finding at least one failure is no lower than random testing for any program.

Targets of OO Software Testing

Trim the individual trees?
Tame the forest!

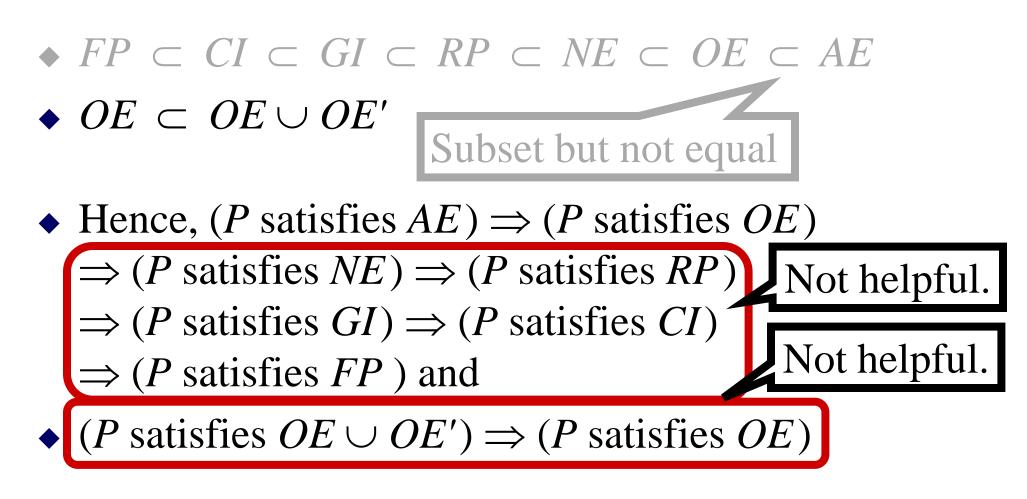


• Comparisons through empirical studies?

Toward Ultimate Target of OO Software Testing Tame the Forest

• $FP \subset CI \subset GI \subset RP \subset NE \subset OE \subset AE$ Subset but not equal

Toward Ultimate Target of OO Software Testing Tame the Forest



Ultimate Target of OO Software Testing Tame the Forest

- Have we made full use of algebraic specifications?
- Given a *canonical* specification Sp with *proper imports* and a *complete* implementation P
 - (*P* satisfies *AE*)

Ultimate Target of OO Software Testing Tame the Forest

- Have we made full use of algebraic specifications?
- Given a *canonical* specification Sp with *proper imports* and a *complete* implementation P
 - (*P* satisfies *AE*)
 ⇔ (*P* satisfies *OE*)

- Have we made full use of algebraic specifications?
- Given a *canonical* specification Sp with *proper imports* and a *complete* implementation P
 - (P satisfies AE)
 ⇔ (P satisfies OE) ⇔ (P satisfies OE')

- Have we made full use of algebraic specifications?
- Given a *canonical* specification Sp with *proper imports* and a *complete* implementation P
 - (P satisfies AE)
 ⇔ (P satisfies OE) ⇔ (P satisfies OE')
 ⇔ (P satisfies OE ∪ OE')

- Have we made full use of algebraic specifications?
- Given a *canonical* specification Sp with *proper imports* and a *complete* implementation P
 - (*P* satisfies *AE*)
 - $\Leftrightarrow (P \text{ satisfies } OE) \Leftrightarrow (P \text{ satisfies } OE')$ $\Leftrightarrow (P \text{ satisfies } OE \cup OE')$
 - $\Leftrightarrow (P \text{ satisfies } NE) \Leftrightarrow (P \text{ satisfies } RP)$ $\Leftrightarrow (P \text{ satisfies } GI) \Leftrightarrow (P \text{ satisfies } CI)$ $\Leftrightarrow (P \text{ satisfies } FP).$

In short

 Given a *canonical* specification Sp with *proper imports* and a *complete* implementation P

• (*P* satisfies AE) \Leftrightarrow (*P* satisfies $OE \cup OE'$)

Real-world short-cut considered too weak to be useful Real-world correctness considered too difficult to test.

Many Facets of the Test Oracle Problem

Challenge 4:

• Expected outcome = actual execution result

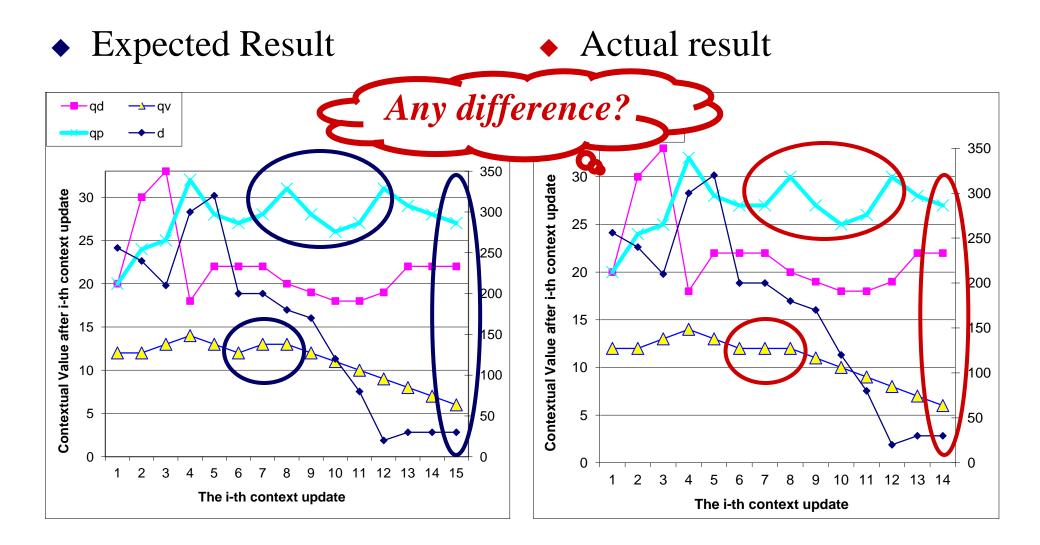
- Computing everywhere and at any time
- Applications operate in dynamic environment

- Computing everywhere and at any time
- Applications operate in dynamic environment
- Expected outcome = actual execution result

- Computing everywhere and at any time
- Applications operate in dynamic environment
- Expected outcome = actual execution result
- When does a test case end?

- Computing everywhere and at any time
- Applications operate in dynamic environment
- Expected outcome = actual execution result
- When does a test case end?
 - Middleware remains active and environment context keeps changing!

Ubiquitous Computing Expected and Actual Context Trends



- Computing everywhere and at any time
- Applications operate in dynamic environment
- Expected outcome = actual execution result
- When does a test case end?
 - Middleware remains active and environment context keeps changing!
- Identify *checkpoints* where the system is momentarily stable.

• Many facets of the test oracle problem

Many facets of the test oracle problem

Expected outcome = actual execution result

Metamorphic testing

Many facets of

metamorphic testing.

• Many facets of the test oracle problem

- Expected outcome = actual execution result
- Expected outcome = actual execution result

Algebraic specifications

Many facets of algebraic specifications.

Many facets of the test oracle problem

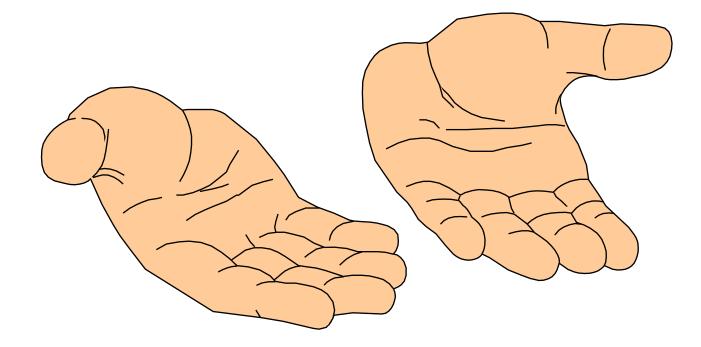
- Expected outcome = actual execution result
- Expected outcome = actual execution result

Expected outcome = actual execution result

Many facets of ubiquitous computing.

- Jungle of problems
- Jungle of proposals
- Empirical studies are just an exploratory first step rather than the ultimate goal
- Tame the forest rather than trimming individual trees.

Your Comments are Welcome





Load the next file please