User Studies of Principled Model Finder Output

Natasha Danas[1]
Tim Nelson[1]
Lane Harrison[2]
Shriram Krishnamurthi[1]
Daniel J. Dougherty[2]

[1] Brown University [2] Worcester Polytechnic Institute

Model Finding

Declarative Spec → Model

for example...

Mutex Algorithm → Possible Deadlocks

Network Configuration → Suspicious Packet Traces

UML Class Diagram → Object Diagrams

What makes a good example?

Aluminum: Principled Scenario Exploration through Minimality

Tim Nelson[1], Salman Saghafi[1], Daniel J. Dougherty[1], Kathi Fisler[1], Shriram Krishnamurthi[2]

Exploring Theories with a Model-Finding Assistant*

Salman Saghafi, Ryan Danas, and Daniel J. Dougherty

Target oriented relational model finding

Alcino Cunha, Nuno Macedo, and Tiago Guimarães

How can one example shed light on all possible examples?

The Power of "Why" and "Why Not": Enriching Scenario Exploration with Provenance

Tim Nelson
Brown University, USA

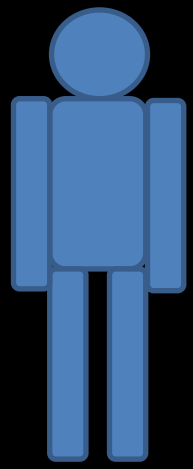
Natasha Danas
Brown University, USA

Daniel J. Dougherty
Worcester Polytechnic Institute, USA

Shriram Krishnamurthi
Brown University, USA

Something is missing

Declarative Spec → Model



Apply HCI to evaluate FM!

(not the same as FM to help HCI)

Where to find participants?

FM conferences?

```
abstract sig Target { }
sig Addr extends Target { }
sig Name extends Target { }
one sig Book {
      entry: set Name,
      target: entry->set Target
}

fact { all n:Name | n in Book.entry }
fact { all e:Book.entry | one Book.target[e] }
fact { no  n:Name | n in (n.^(Book.target)) }
```

Where to find participants?

FM conferences?

- Not everyone knows Alloy;
- low participation rates, population
- one shot per ~year

Students learning Alloy?

"volunteers" hard to attract, so do as part of class:
Brown's Logic for Systems

Class: minimality vs. maximality

Designed into mandatory lab:
Modeling reference-counting garbage collection

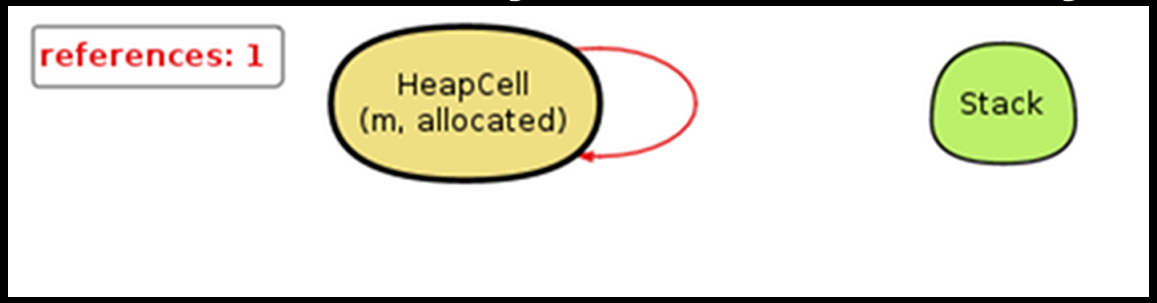
Two groups: shown minimal (or maximal) failures.
(60 total participants; 35 min, 25 max)

Can they:

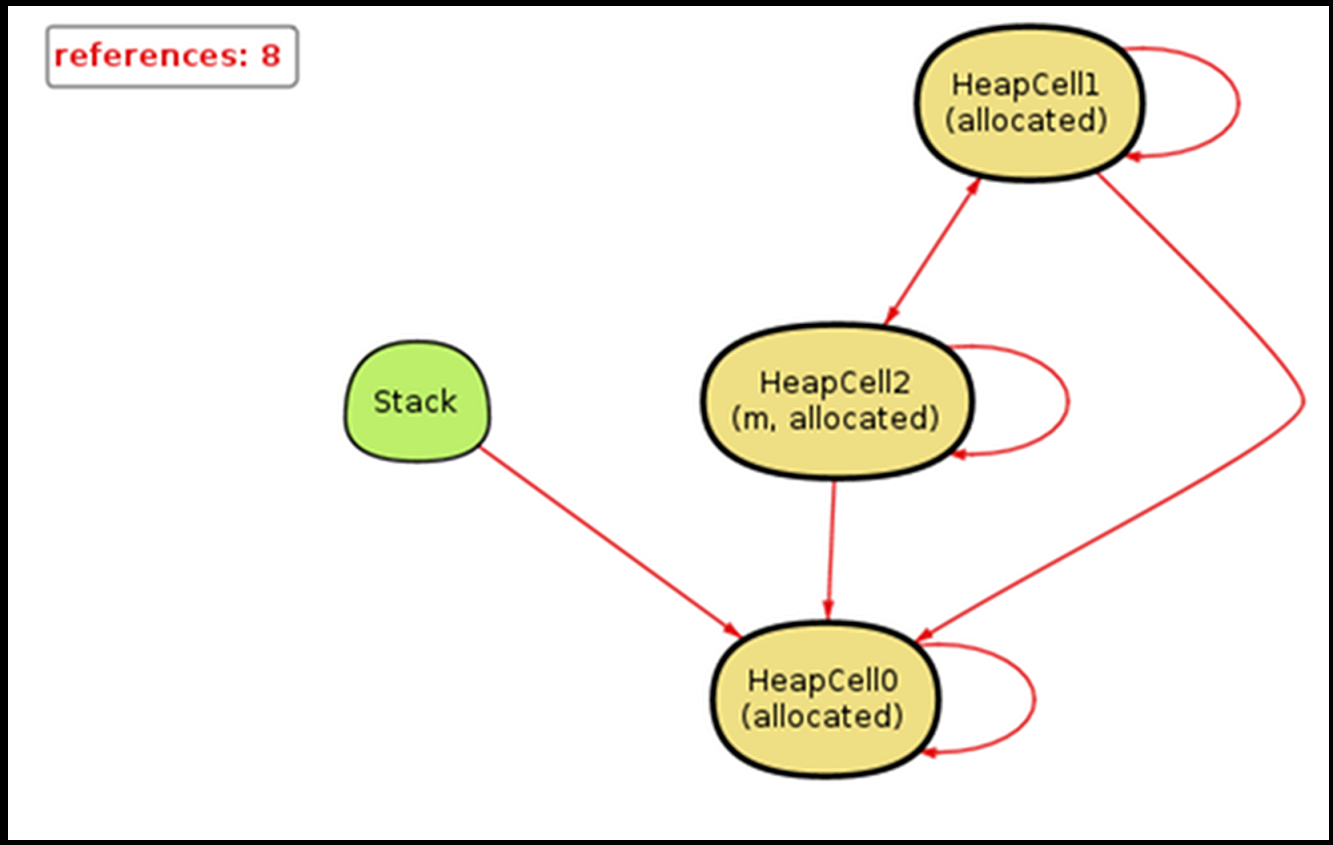
1. identify the collection problem (cyclic references) and
2. write a constraint characterizing it?

Class: minimality vs. maximality

Min:



Max:



Class: minimality vs. maximality

```
all s: State, m: HeapCell | …
```

Incorrect answers include trivial unsat or begging the question:

```
all s: State, m: HeapCell |
    s != s
```

Expected correct answer:

```
all s: State, m: HeapCell |
    m not in m.^(s.references)
```

Class: minimality vs. maximality

Significant frustration in both groups:

MIN: 35 – 28 = 7 left after dropout

MAX: 25 – 10 = 15 left after dropout

Expected something non-min, non-max

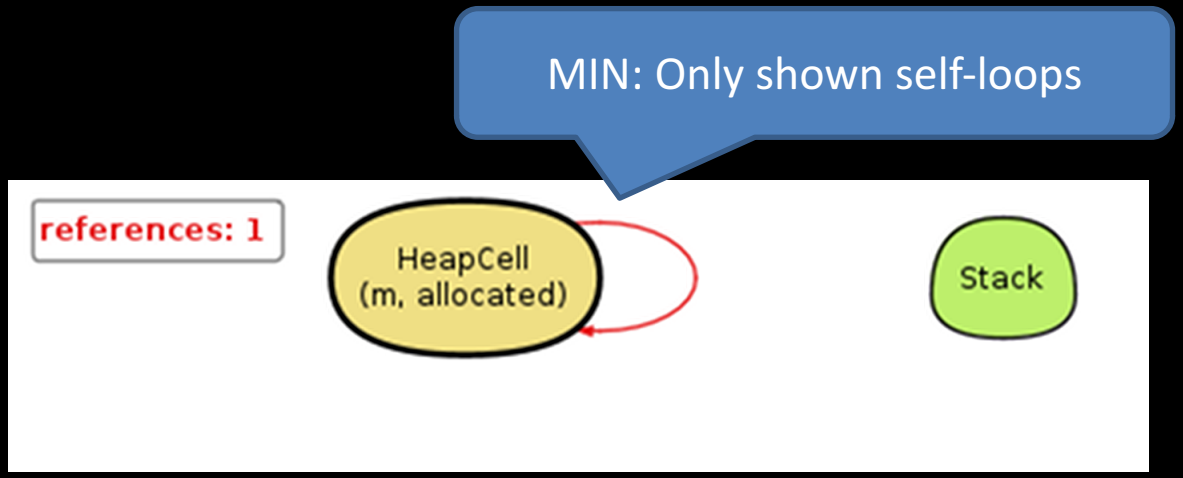
MIN: 3 correct, 3 self-loop only, 1 other incorrect.
MAX: 2 correct, 5 self-loop only, 8 other incorrect.

Both principled forms did badly!

("other incorrect" = begging the question, trivial, etc.)

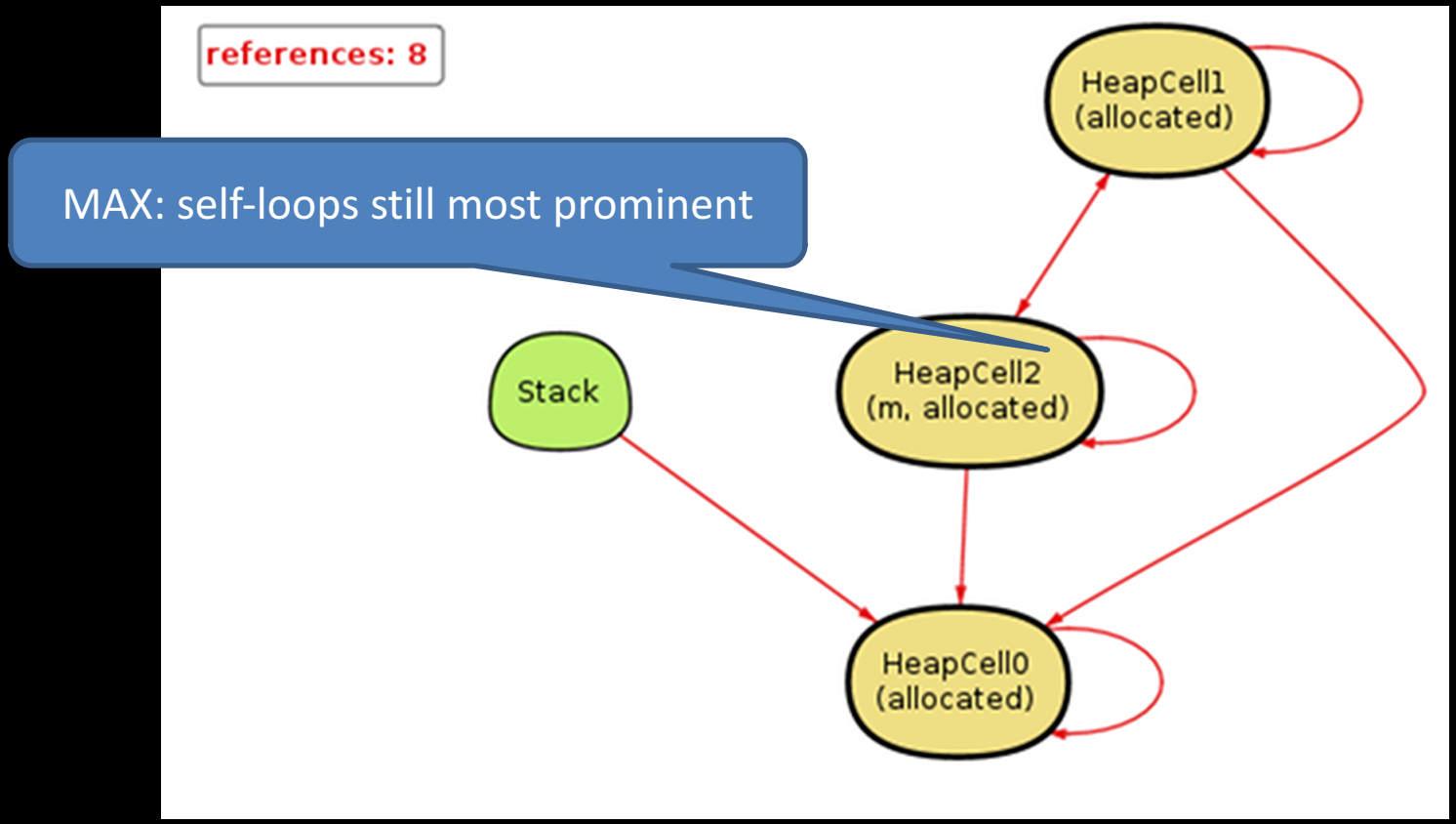
Class: minimality vs. maximality

MIN: 3 correct, 3 self-loop only, 1 other bug.



Class: minimality vs. maximality

MAX: 2 correct, 5 self-loop only, 8 other bugs.



Weaknesses of class

Timing limited to semester
Sample size limited to class

Expensive to iterate
Need to design to avoid interfering with pedagogy
+ avoid higher grades for confirming our hypotheses

No way to avoid re-using participants if multiple studies
(priming hard to avoid)

Worse: they want to please us!

Where else can we go?

Crowdsourcing: Challenges

Not many Alloy users on Turk. May be technical, but non-expert.
(same problems as FM conferences, writ much larger)

Training phase

Some are dishonest, bots, or trolls.

Filtering

May have no knowledge of (e.g.) garbage collection.

Basic spec domains, English language

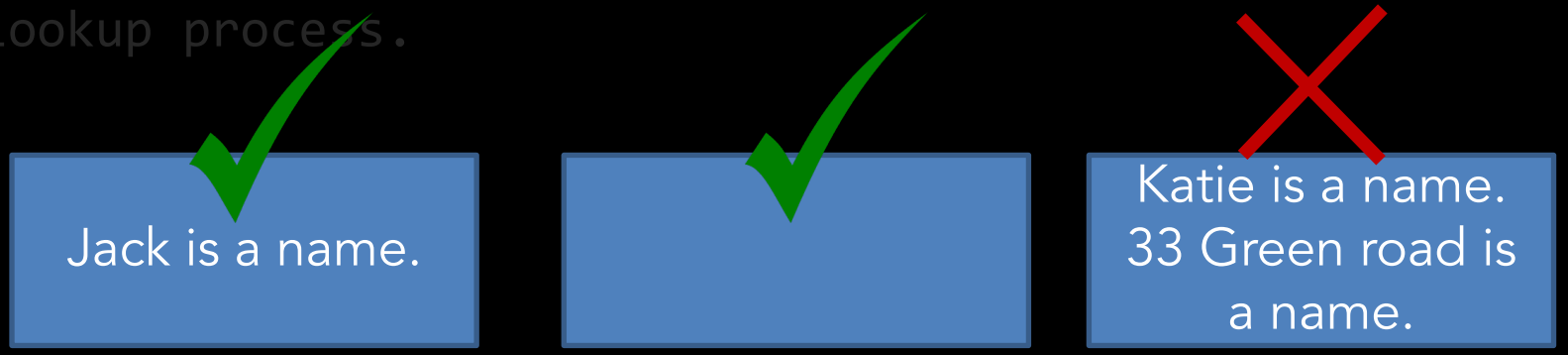
Address books map a finite set of names to targets.

A target is either a name (like ‘Bob’) or an address (like ‘102 Rose Way’).

Every name is listed in the book.

Every name maps to exactly one target.

No name is listed in its own lookup. To look up a name, one finds the respective listing; if the mapped target is not an address, one continues the lookup process.



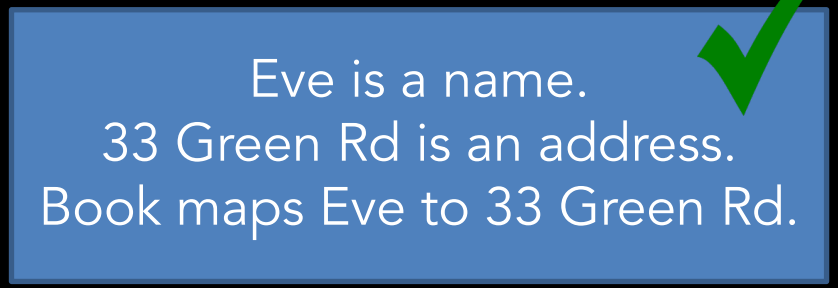
Address books map a finite set of names to targets.

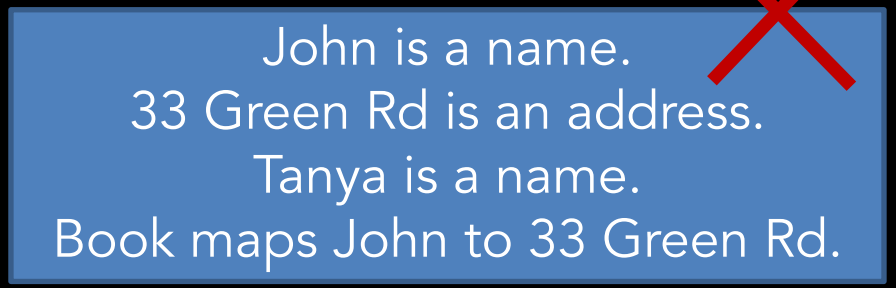
A target is either a name (like ‘Bob’) or an address (like ‘102 Rose Way’).

Every name is listed in the book.

Every name maps to exactly one target.

No name is listed in its own lookup. To look up a name, one finds the respective listing; if the mapped target is not an address, one continues the lookup process.





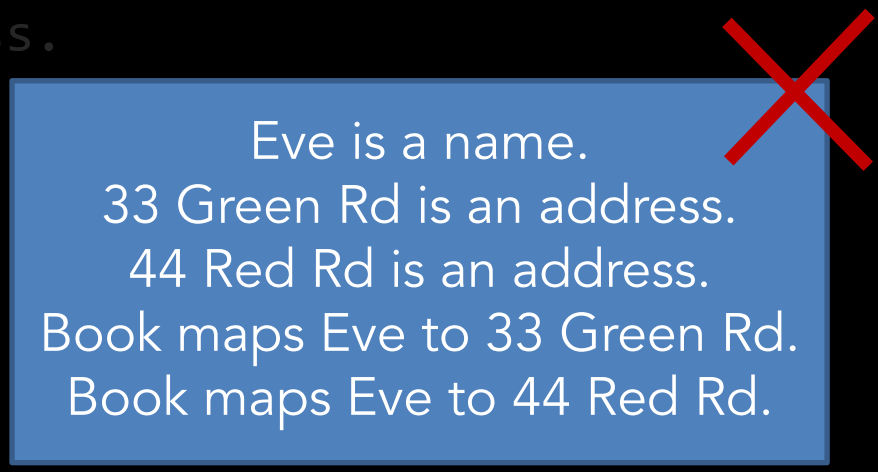
Address books map a finite set of names to targets.

A target is either a name (like ‘Bob’) or an address (like ‘102 Rose Way’).

Every name is listed in the book.

Every name maps to exactly one target.

No name is listed in its own lookup. To look up a name, one finds the respective listing; if the mapped target is not an address, one continues the lookup process.



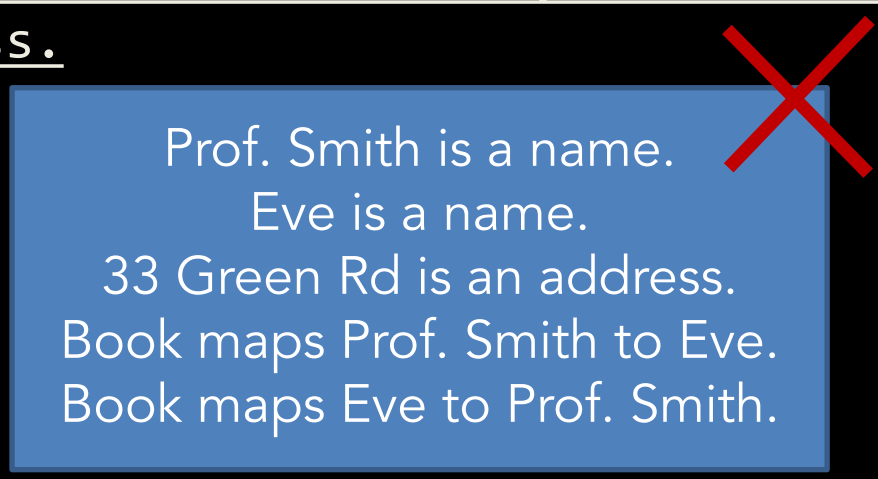
Address books map a finite set of names to targets.

A target is either a name (like ‘Bob’) or an address (like ‘102 Rose Way’).

Every name is listed in the book.

Every name maps to exactly one target.

No name is listed in its own lookup. To look up a name, one finds the respective listing; if the mapped target is not an address, one continues the lookup process.



Turk: Training phase

Recruited 320 participants
192 were above bar

Living wage: 0.15 US$ per minute
(No more than $1.50 to train each person)

Exercise!

Turk: Study phase

Different experiment: evaluate effectiveness of two highlighting output modes: unsat cores and provenance

The Power of "Why" and "Why Not": Enriching Scenario Exploration with Provenance

Tim Nelson
Brown University, USA

Natasha Danas
Brown University, USA

Daniel J. Dougherty
Worcester Polytechnic Institute, USA

Shriram Krishnamurthi
Brown University, USA

Turk: Study phase

Address books map a finite set of names to targets.

A target is either a name (like ‘Bob’) or an address (like ‘102 Rose Way’).

Every name is listed in the book.

Every name maps to **exactly one** target.

No name is listed in its own lookup. To look up a name, one finds the respective listing; if the mapped target is not an address, one continues the lookup process.

Core Highlight

Turk: Study phase

Address books map a finite set of names to targets.

A target is either a name (like ‘Bob’) or an address (like ‘102 Rose Way’).

Every name is listed in the book.

Every name maps to **exactly one** target.

No name is listed in its own lookup. To look up a name, one finds the respective listing; if the mapped target is not an address, one continues the lookup process.

Provenance Highlight

Turk: Results

Address book		Grade book	
Proof Output Type	# Correct	Proof Output Type	# Correct
unsat core	9 / 49 (18%)	unsat core	23 / 53 (43%)
provenance	25 / 46 (54%)	provenance	32 / 44 (73%)

Table 4. Comparing Proof Output Effects on Crowd Workers

For both address book and grade book, provenance helped localize ($p < 0.01$ both); medium effect size (~0.3)

46% did not select the constraint provenance highlighted!

Do user studies:
you may be surprised!

tn@cs.brown.edu