Sledgehammer Hell
The Day after Judgment

Jasmin C. Blanchette
TU München

Larry Paulson

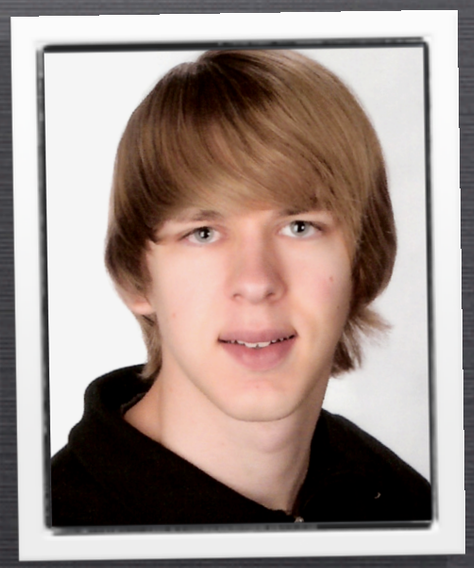
Jia Meng

Kong Susanto

Claire Quigley

Markus Wenzel

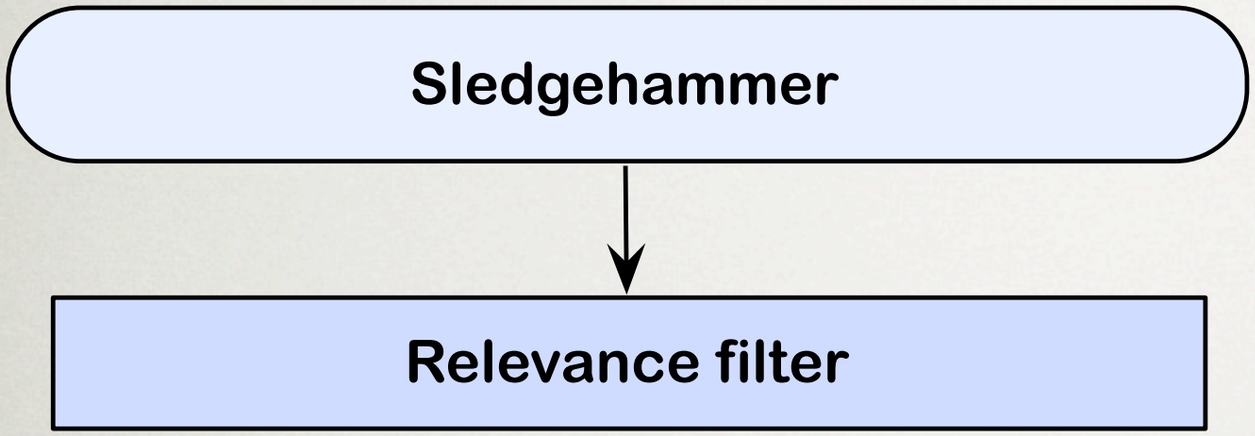
Fabian Immler

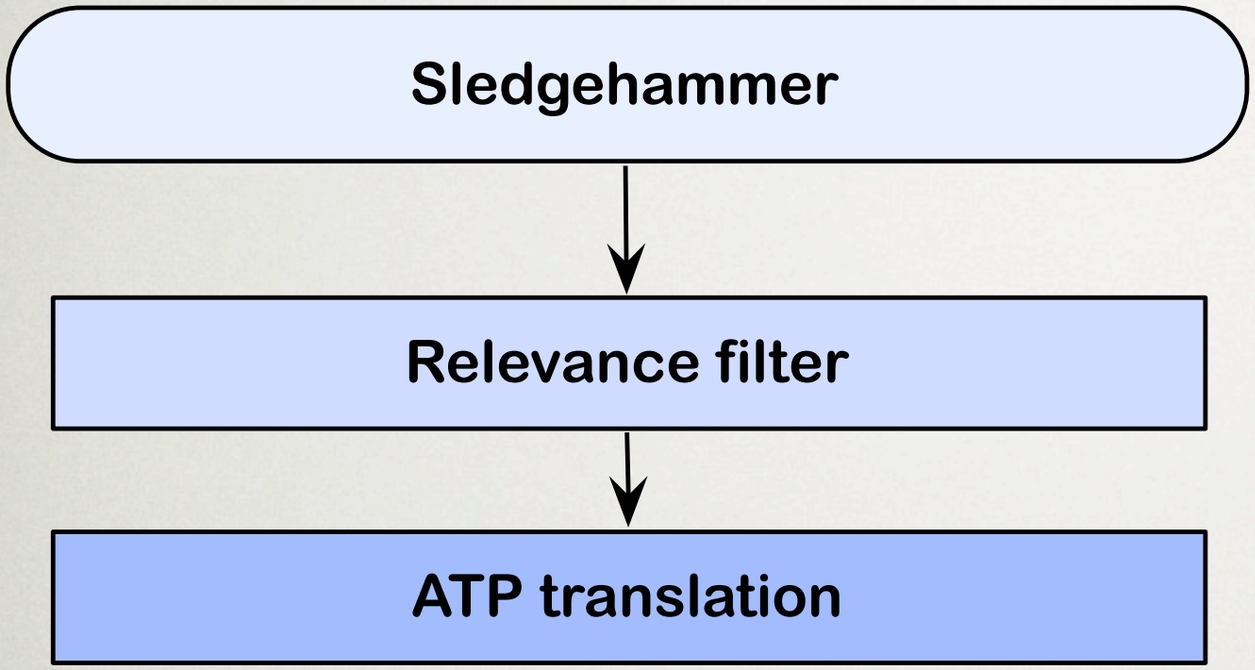


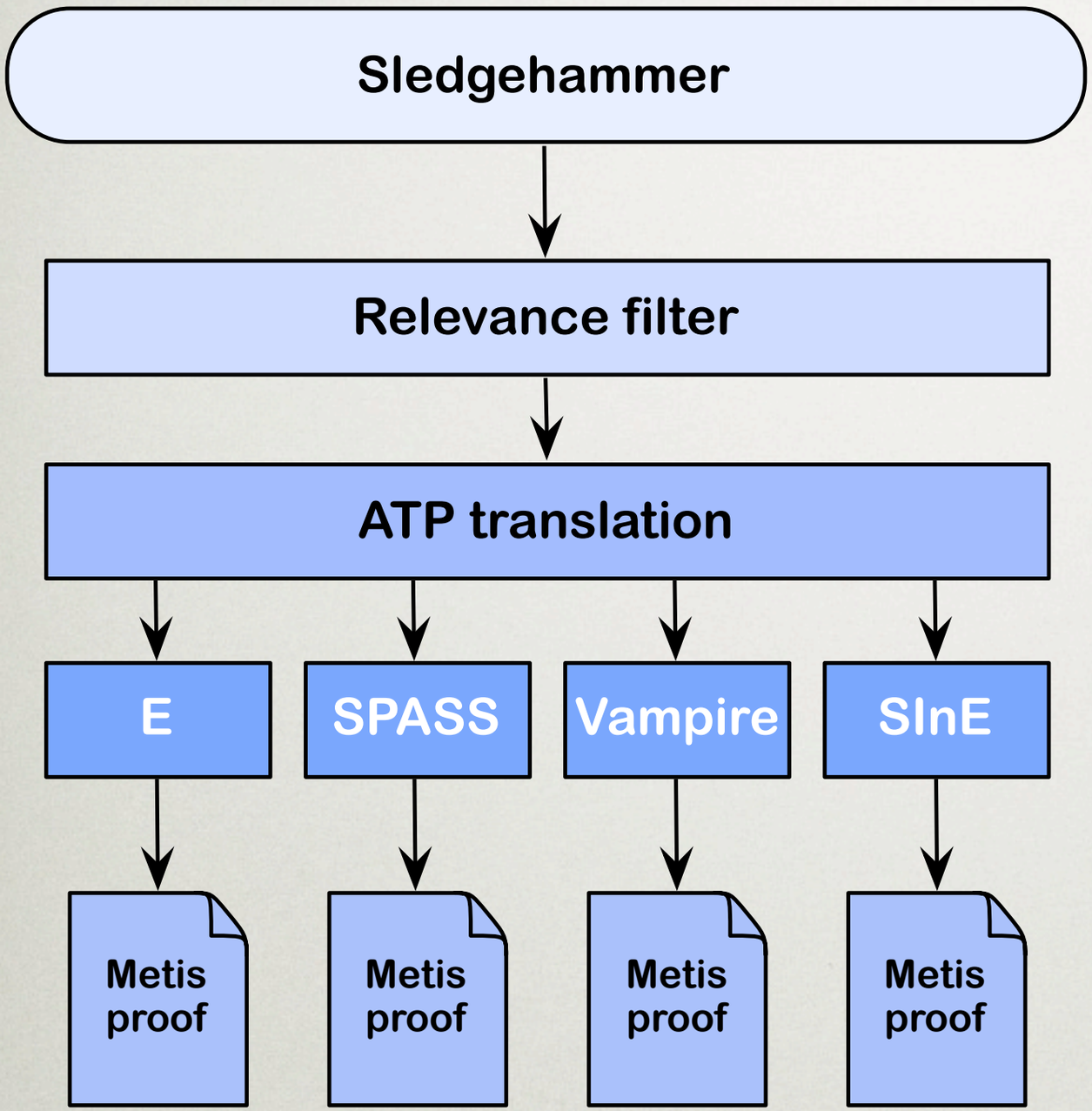
Philipp Meyer

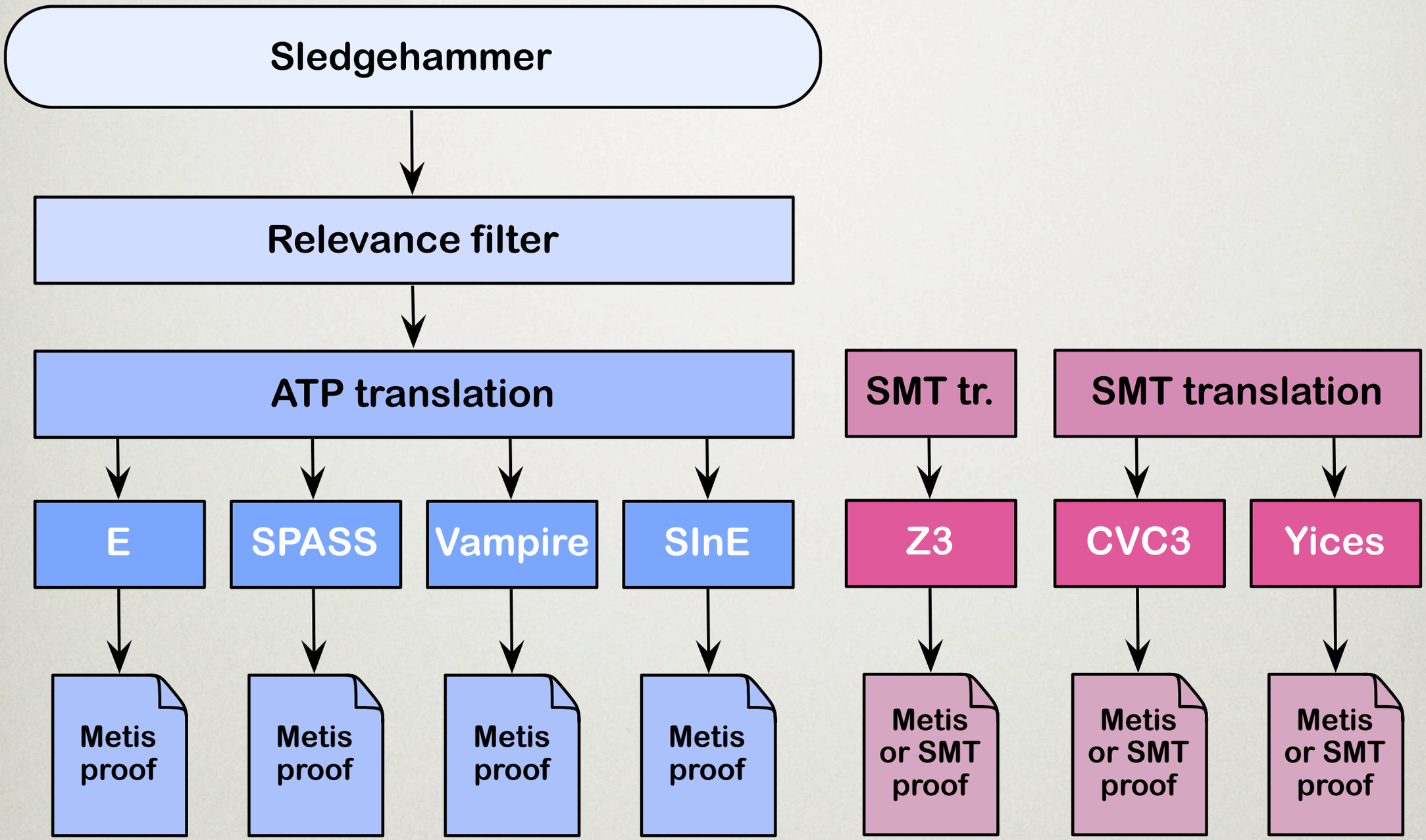
Sascha Böhme

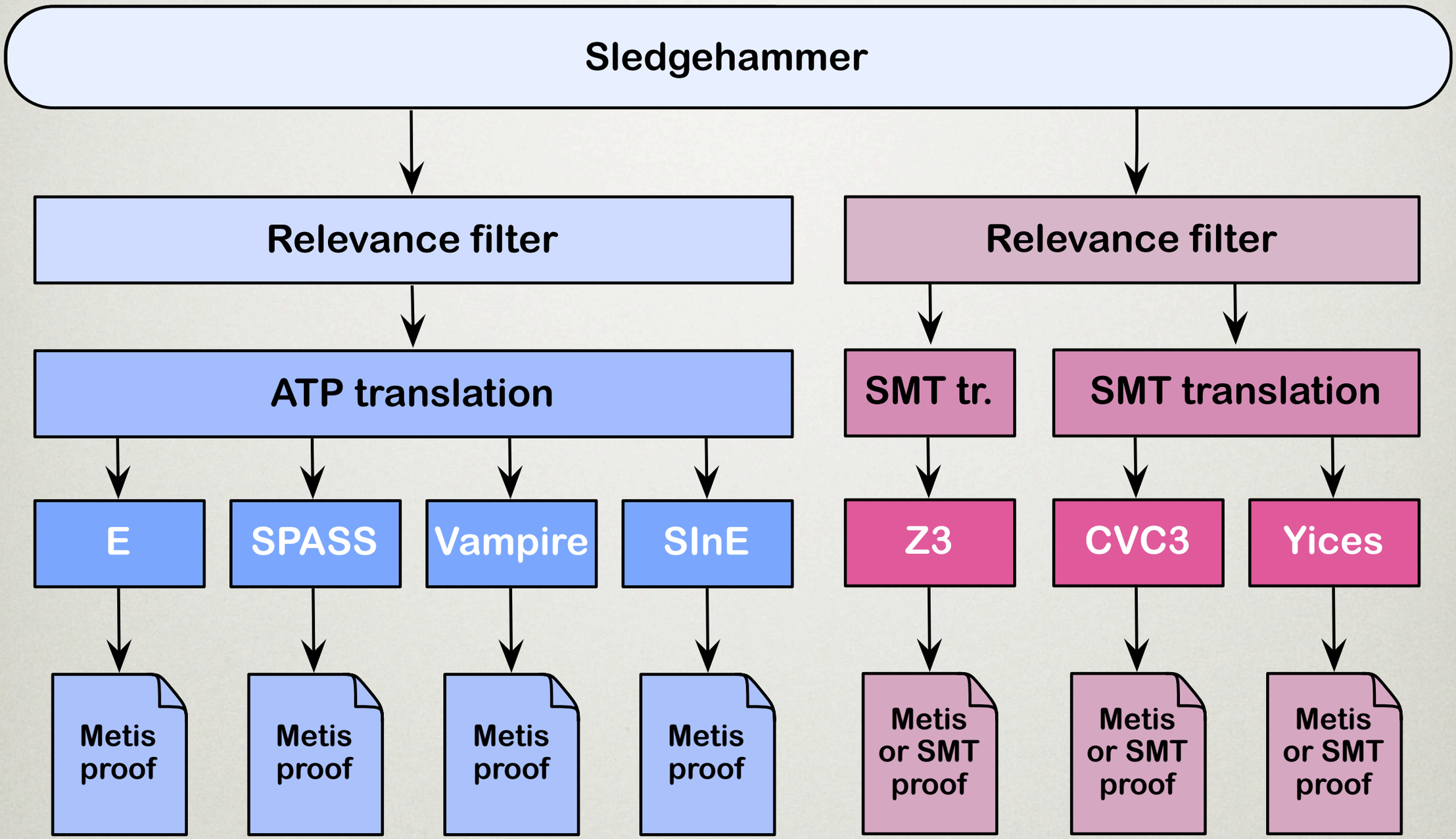
Sledgehammer


Sledgehammer
Relevance filter


Sledgehammer
Relevance filter
ATP translation


Sledgehammer
Relevance filter
ATP translation
E
SPASS
Vampire
SInE
Metis proof
Metis proof
Metis proof
Metis proof


Sledgehammer
Relevance filter
ATP translation
SMT tr.
SMT translation
E
SPASS
Vampire
SInE
Z3
CVC3
Yices
Metis proof
Metis proof
Metis proof
Metis proof
Metis or SMT proof
Metis or SMT proof
Metis or SMT proof


Sledgehammer
Relevance filter
Relevance filter
ATP translation
SMT tr.
SMT translation
E
SPASS
Vampire
SInE
Z3
CVC3
Yices
Metis proof
Metis proof
Metis proof
Metis proof
Metis or SMT proof
Metis or SMT proof
Metis or SMT proof

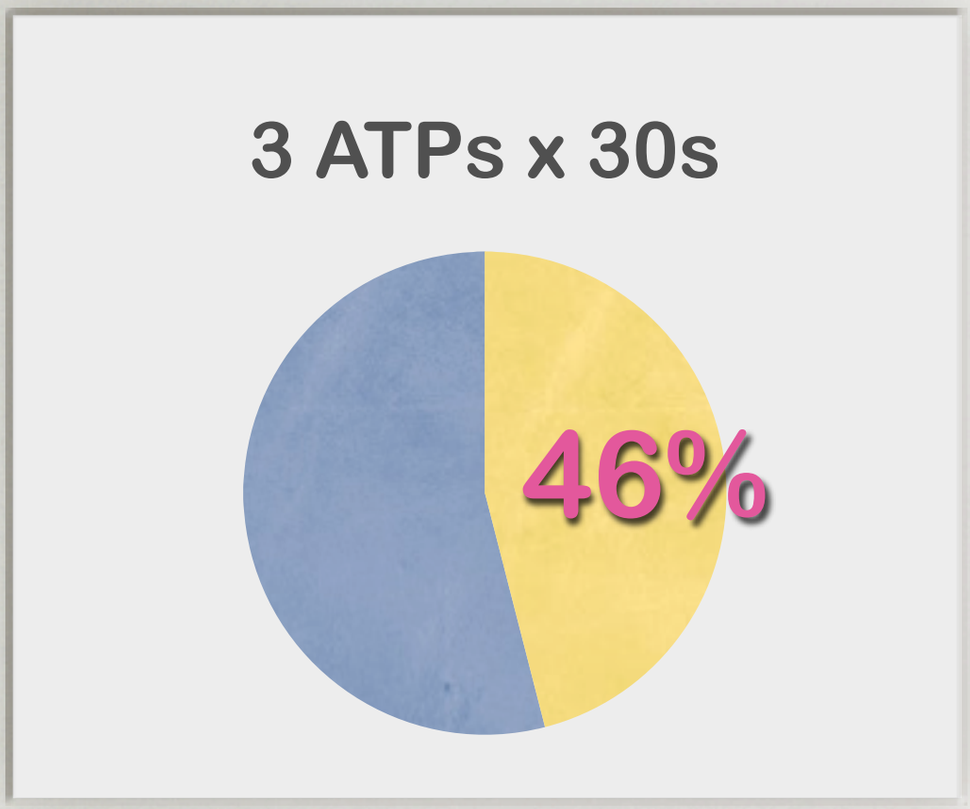
```
rev [a, b] = [b, a]
```

```
rev [a, b] = [b, a]
```

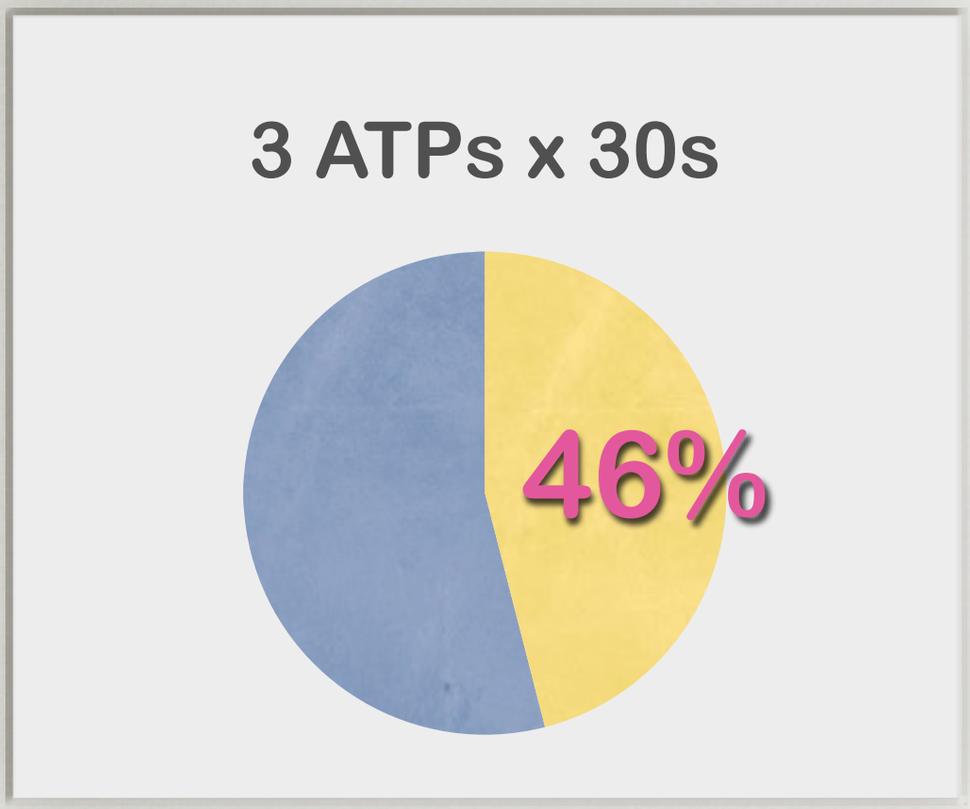
```
by (metis Cons_eq_appendI
          eq_Nil_appendI
          rev.simps(2)
          rev_singleton_conv)
```

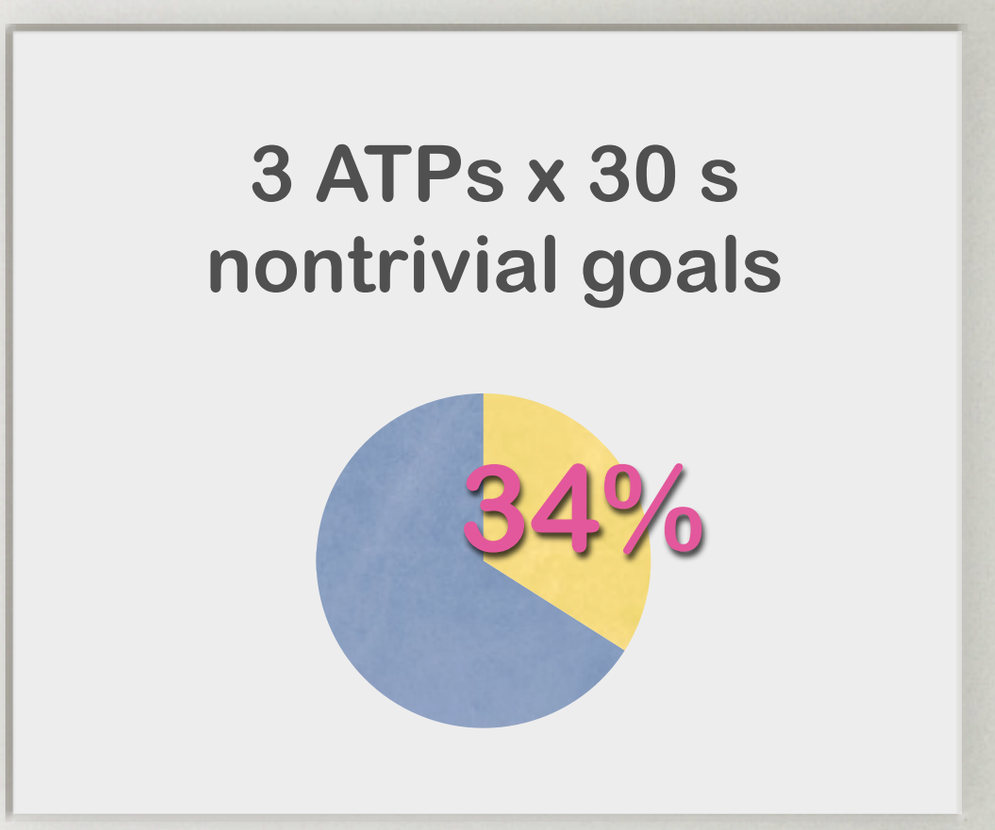
2010

2010

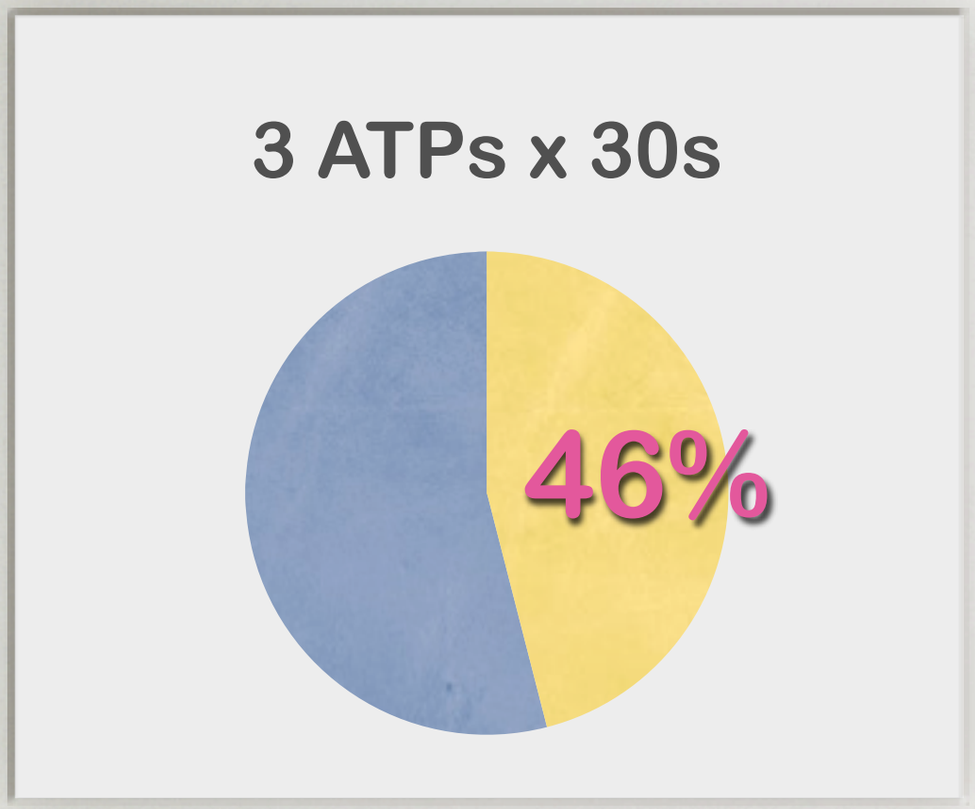


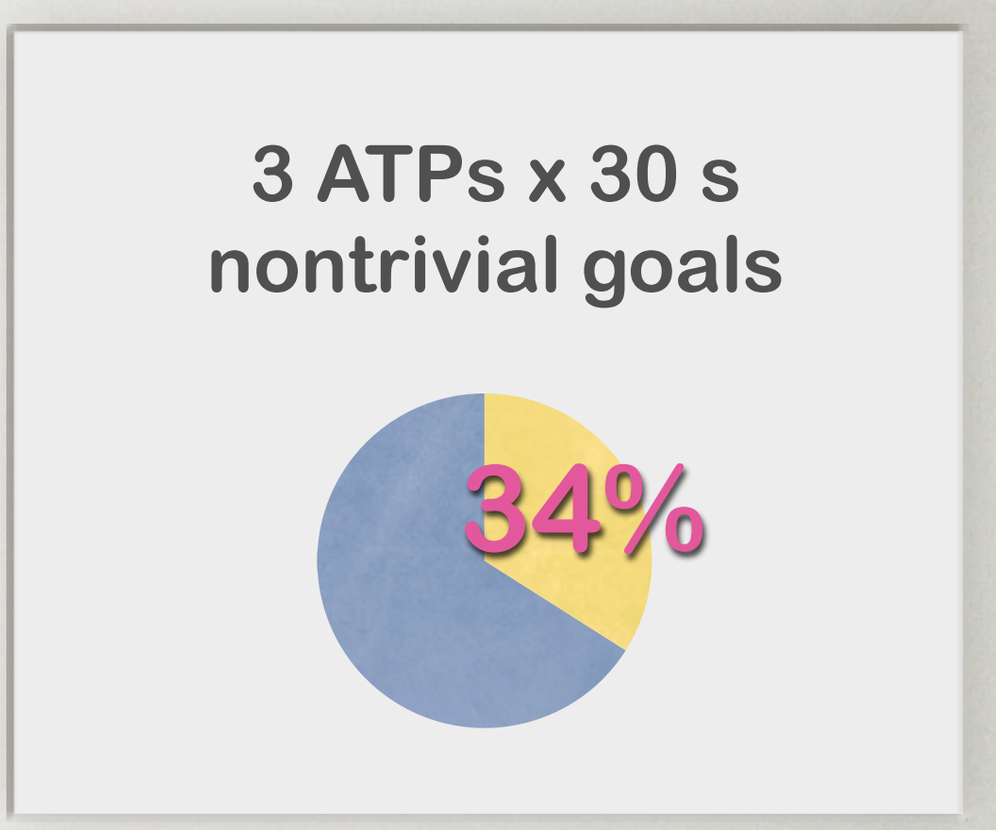
2010



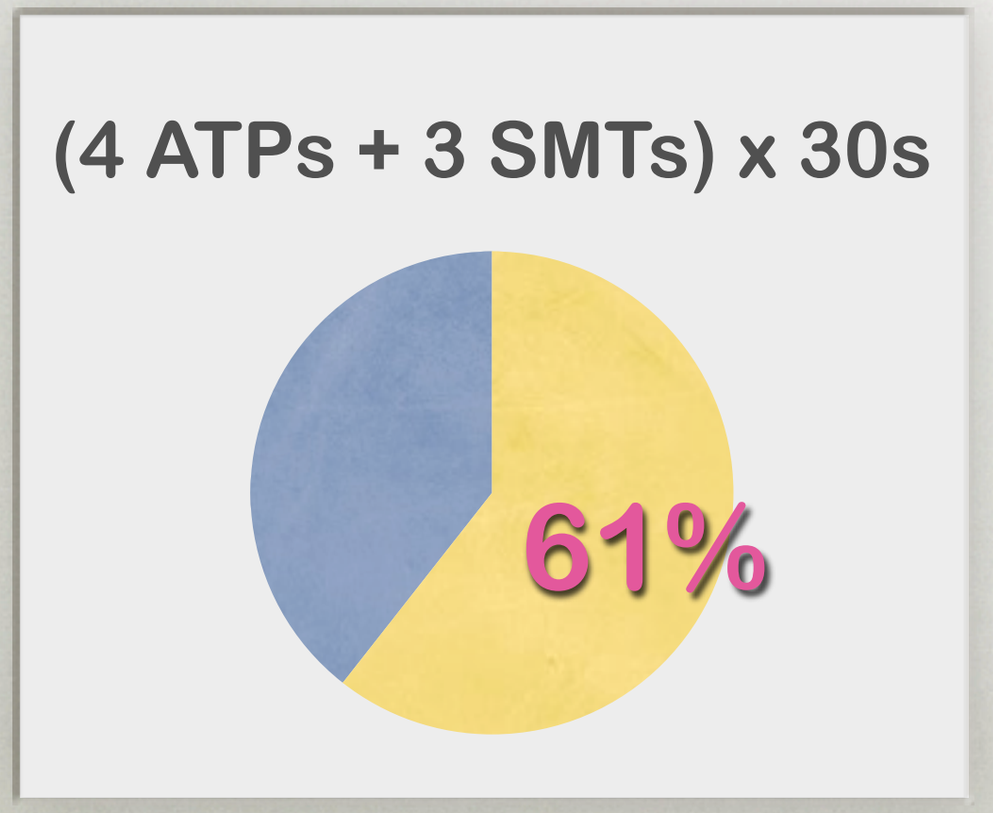


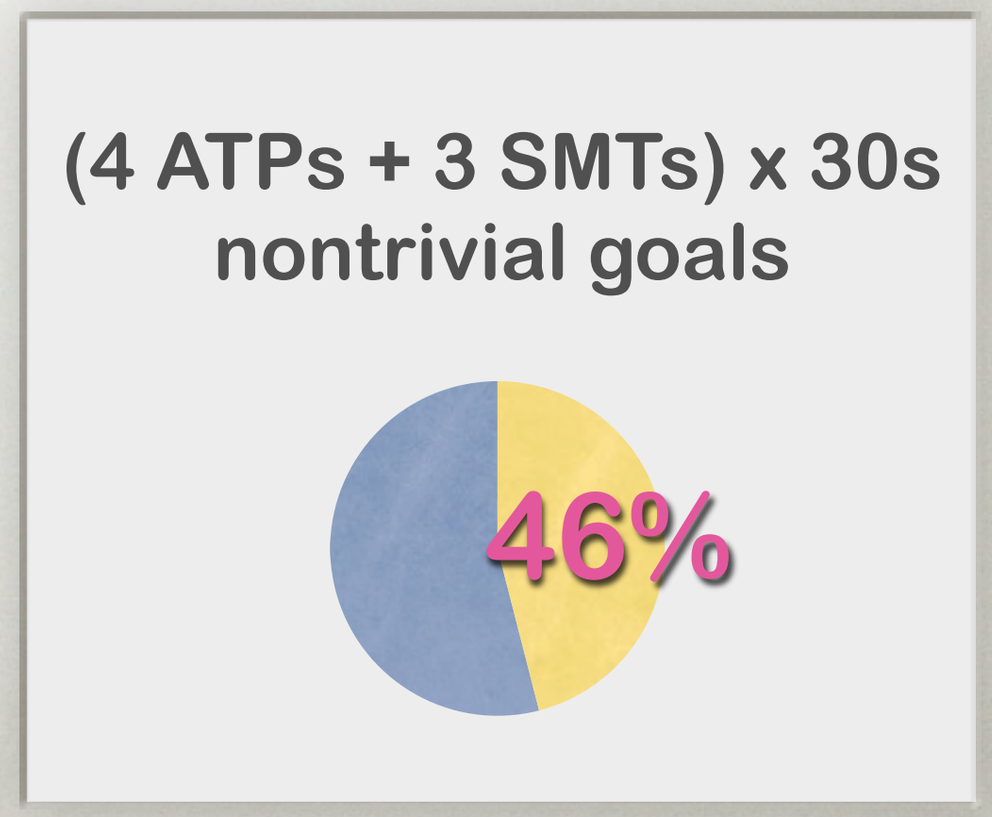
2010


3 ATPs x 30s
46%


3 ATPs x 30 s
nontrivial goals
34%

2011


(4 ATPs + 3 SMTs) x 30s
61%


(4 ATPs + 3 SMTs) x 30s
nontrivial goals
46%

Issue #1:
Too Many Facts

Issue #2:
Encoding Overhead

Issue #3:
Large Terms

Issue #1:

Too Many Facts

Conjecture: … c … d … e …

Conjecture: … c … d … e …

Lemma 1: … c … f …

Lemma 2: … f … g …

Lemma 3: … g … h …

Conjecture: … c … d … e …

Lemma 1: … c … f … ✓

Lemma 2: … f … g …

Lemma 3: … g … h …

Conjecture: … c … d … e …

Lemma 1: … c … f … ✓

Lemma 2: … f … g … ✓

Lemma 3: … g … h …

Conjecture: … c … d … e …

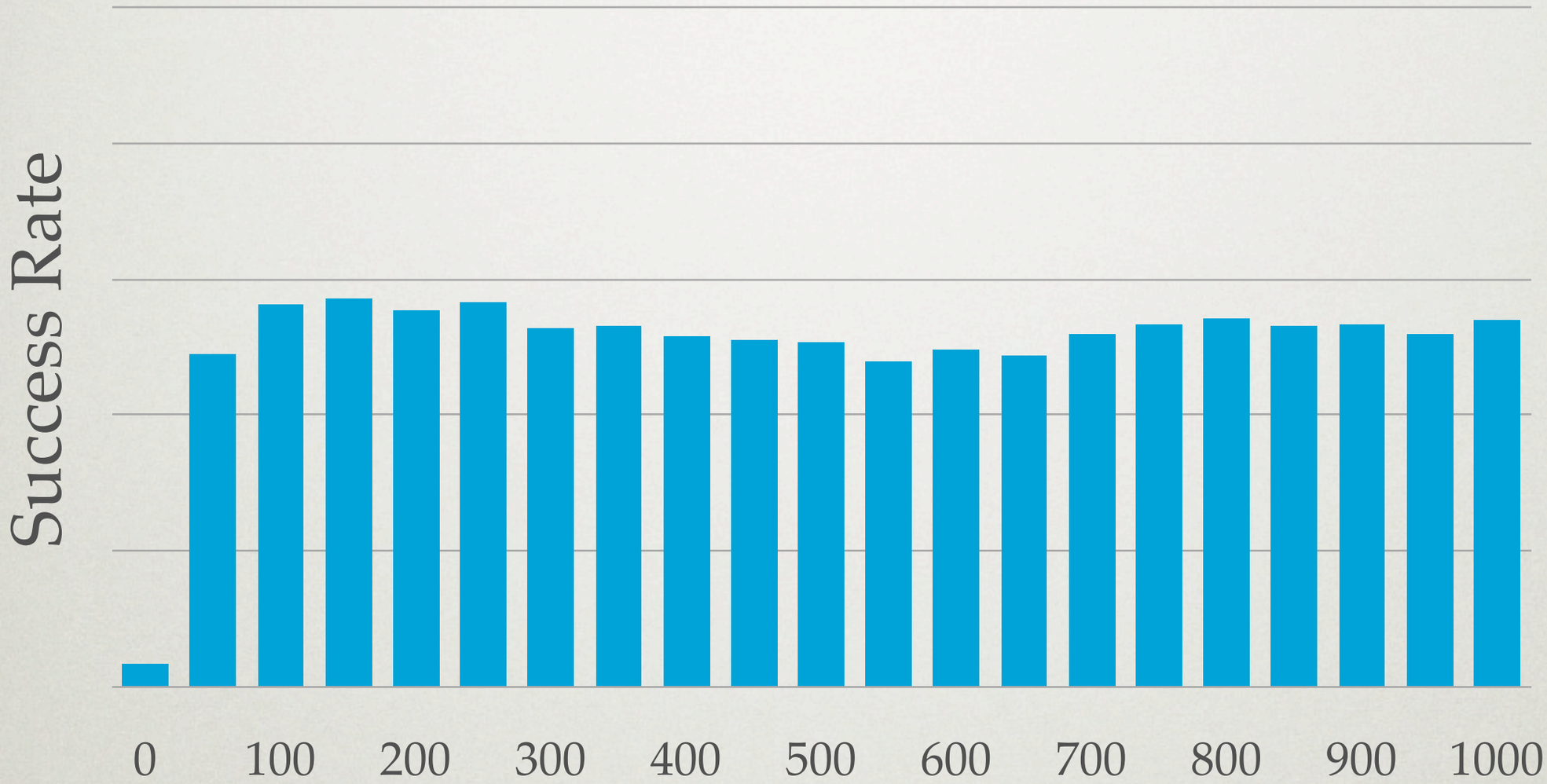
Lemma 1: … c … f … ✓

Lemma 2: … f … g … ✓

Lemma 3: … g … h … ✓

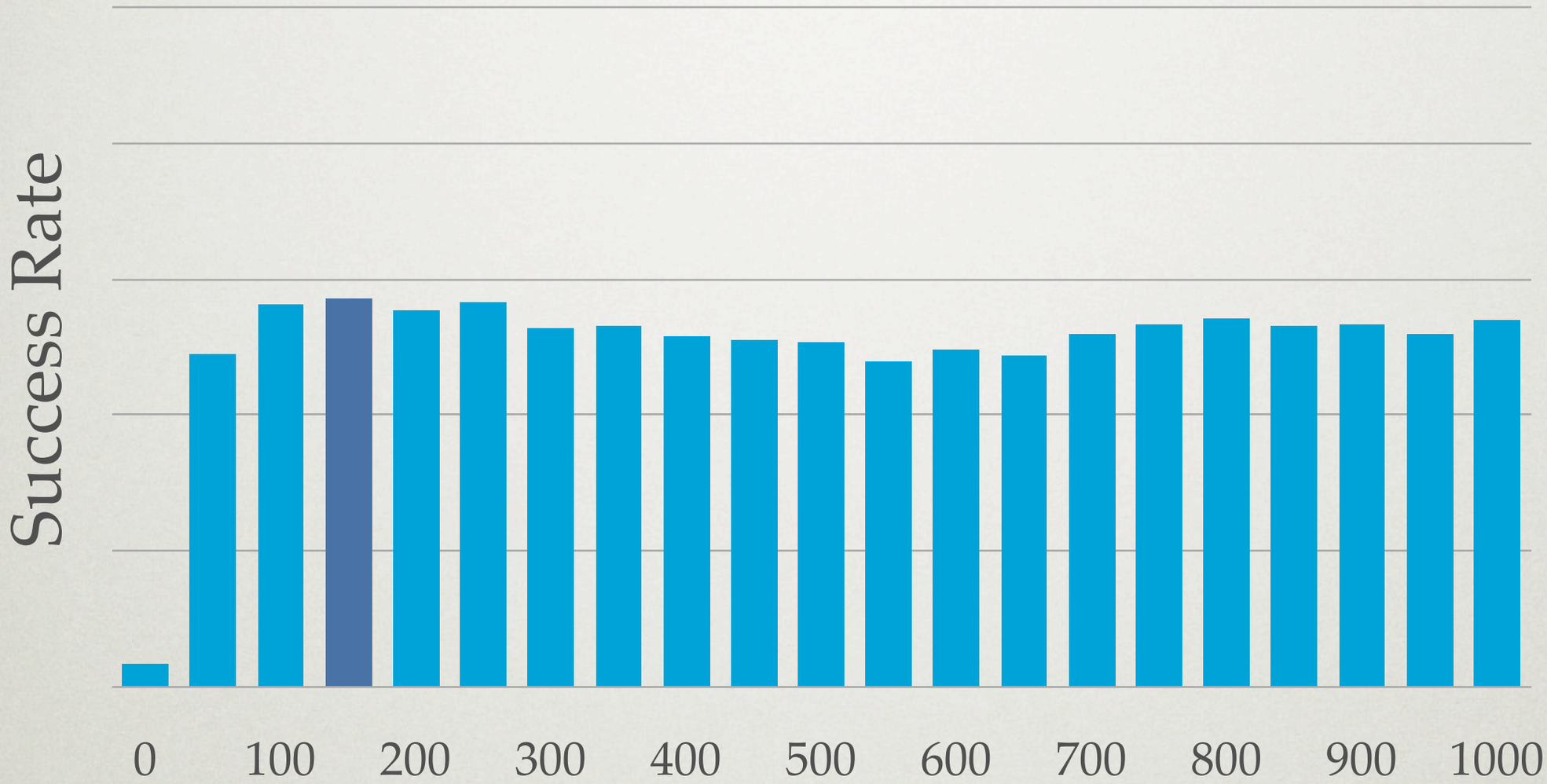
The More Facts...

E 1.2



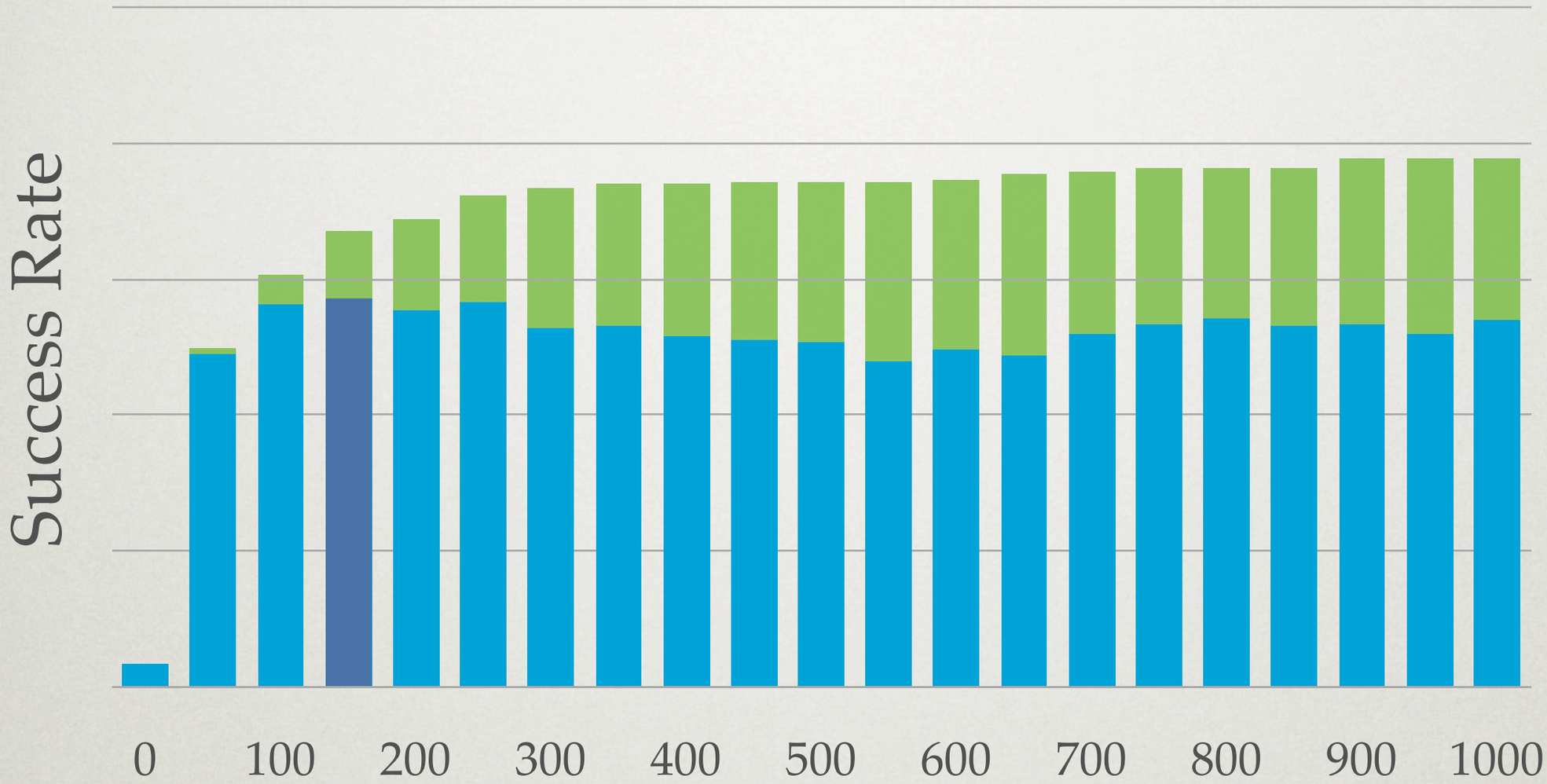
The More Facts...

E 1.2



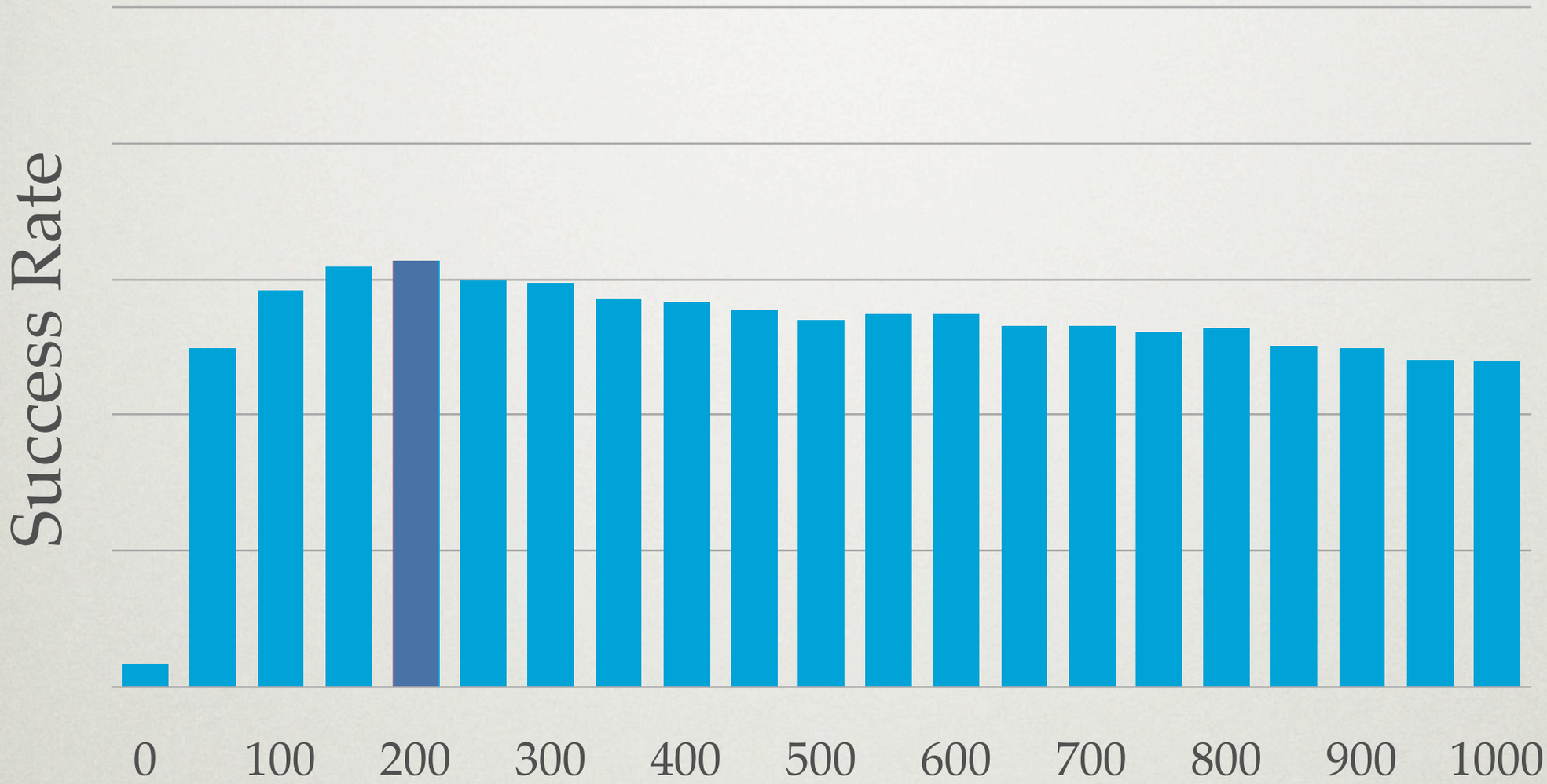
The More Facts...

E 1.2



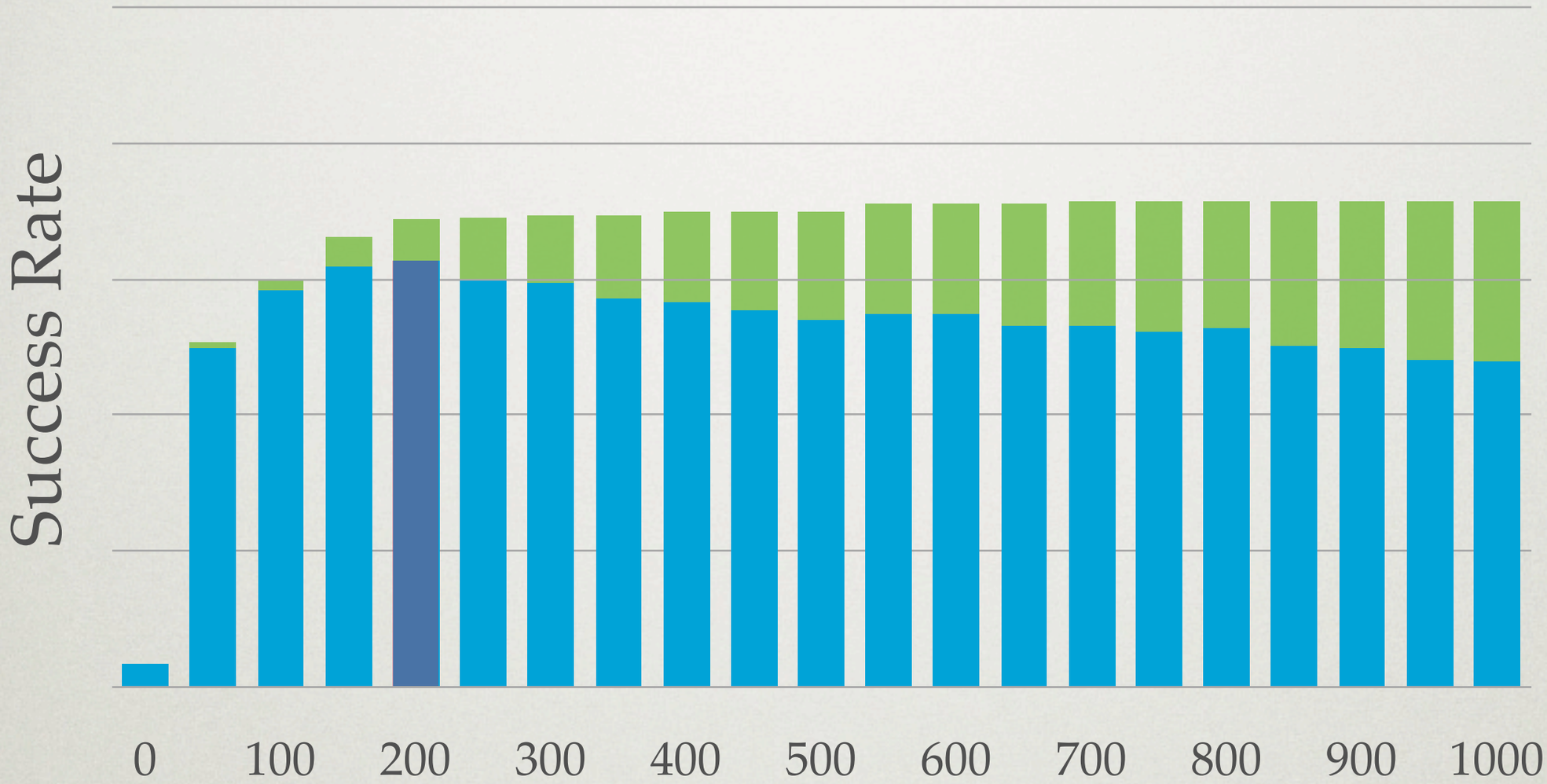
The More Facts...

SPASS 3.7



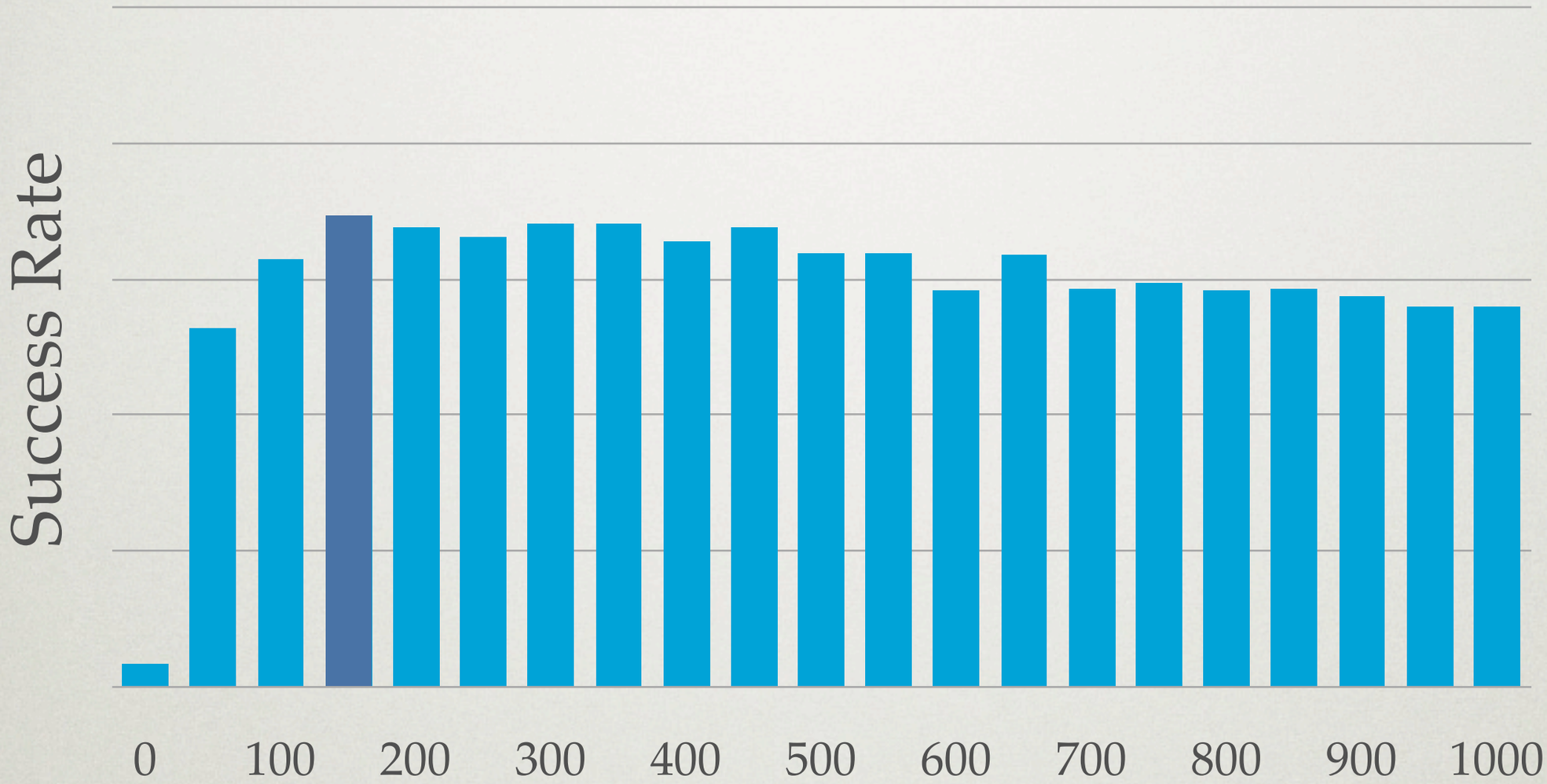
The More Facts...

SPASS 3.7



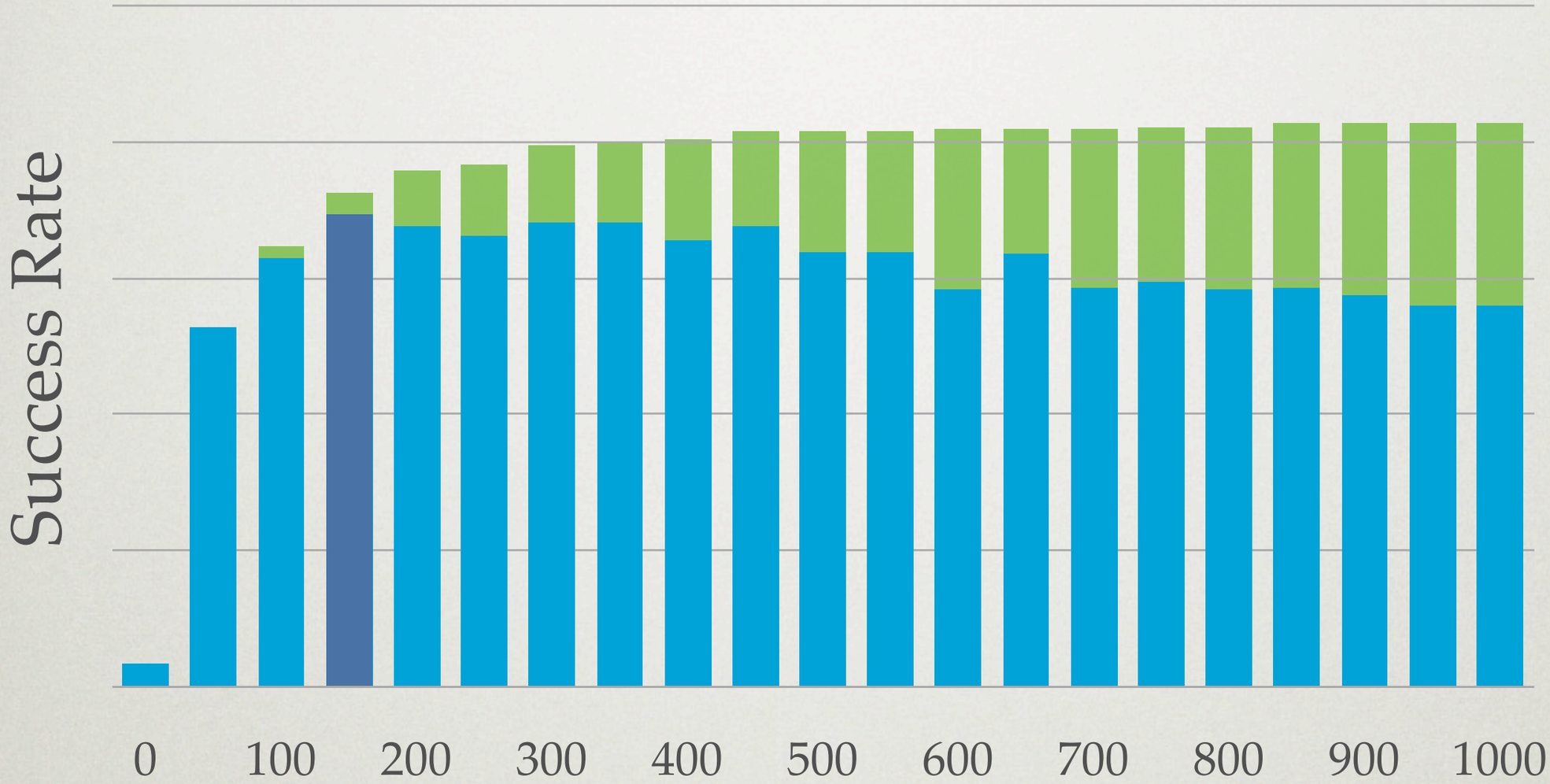
The More Facts...

Vampire 1.0



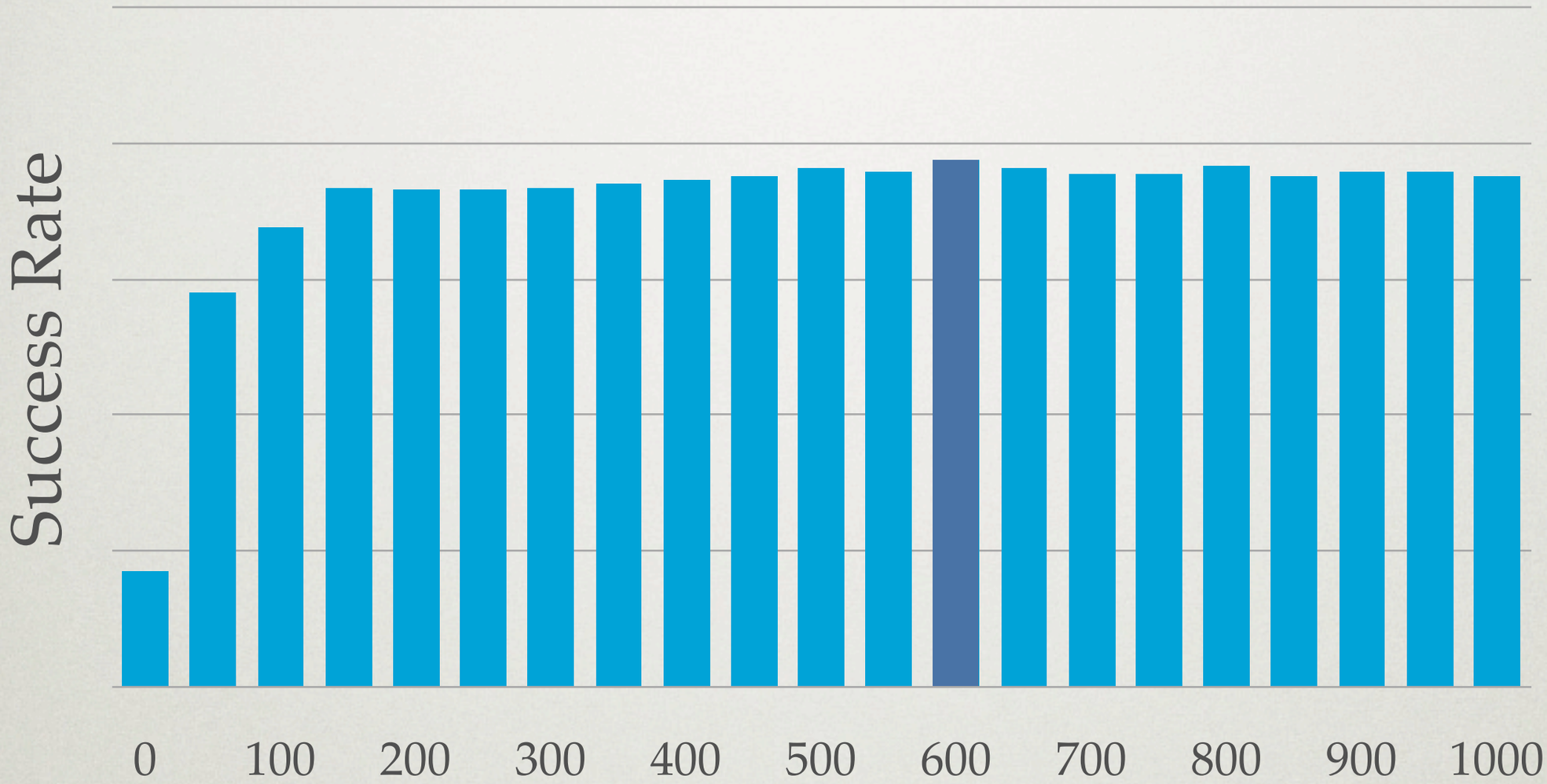
The More Facts...

Vampire 1.0



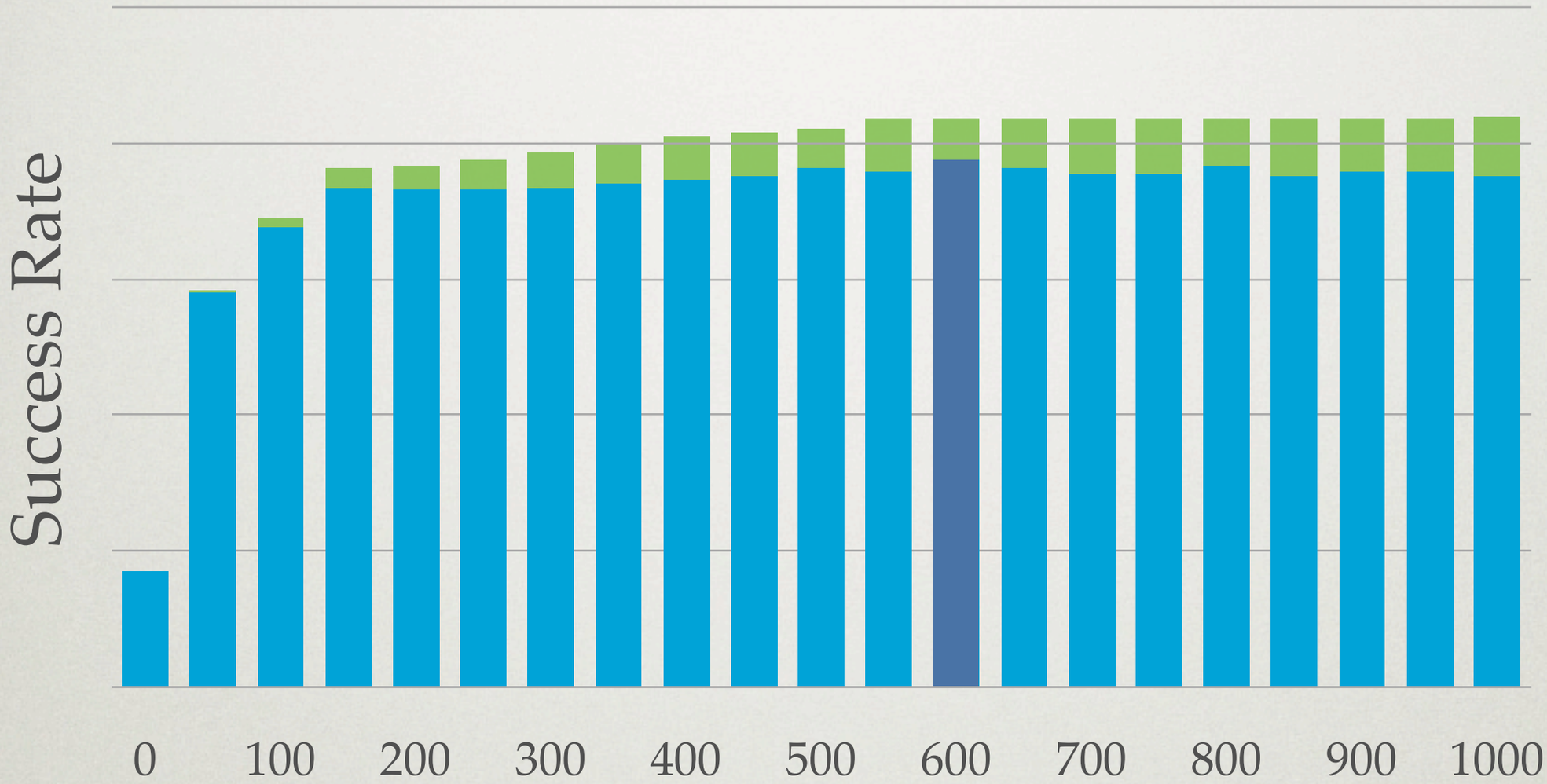
The More Facts...

Z3 2.15



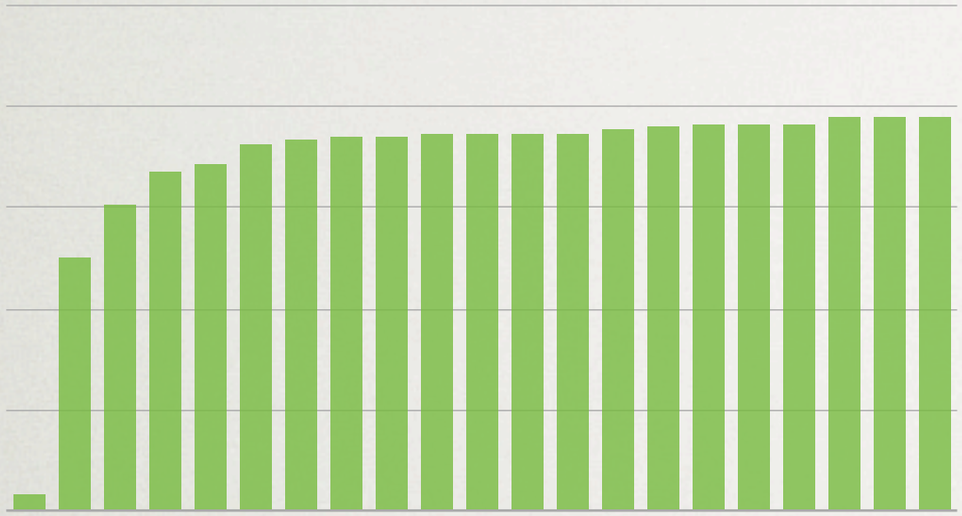
The More Facts...

Z3 2.15

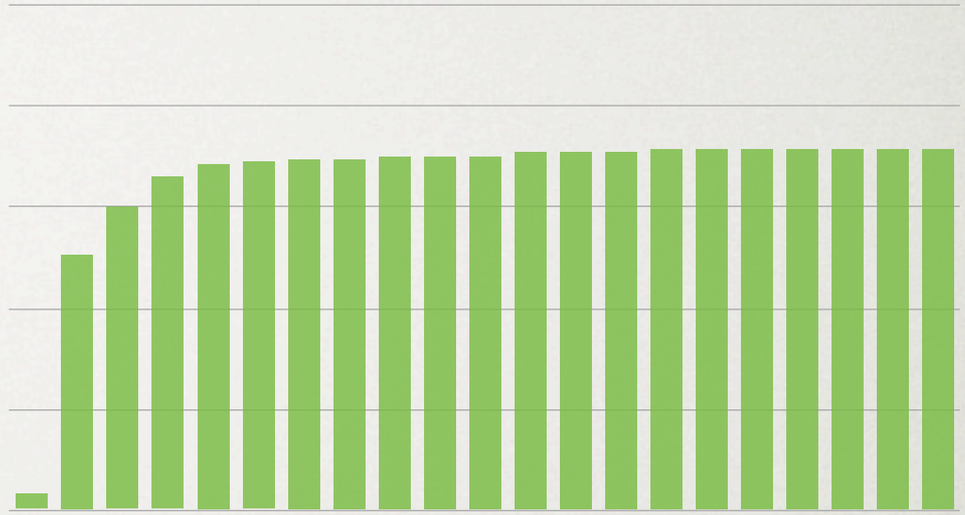


The More Facts...

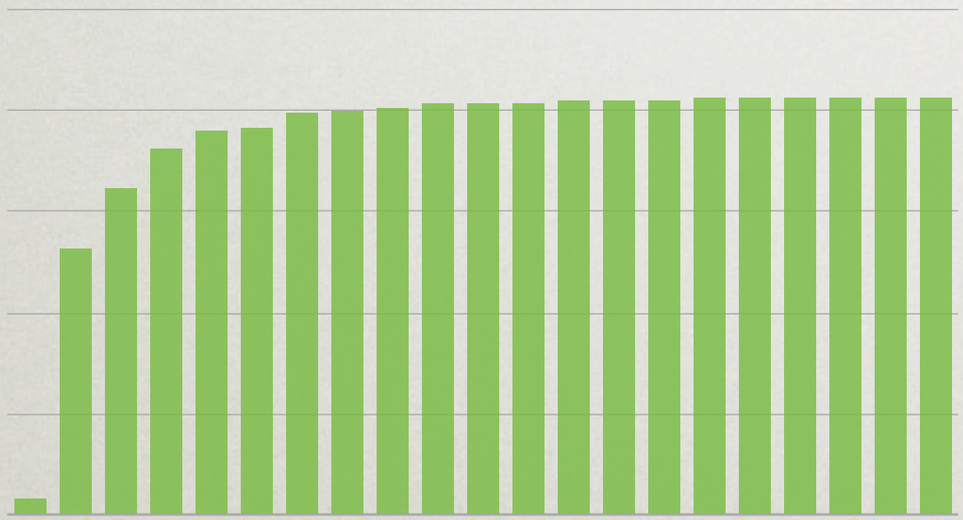
E



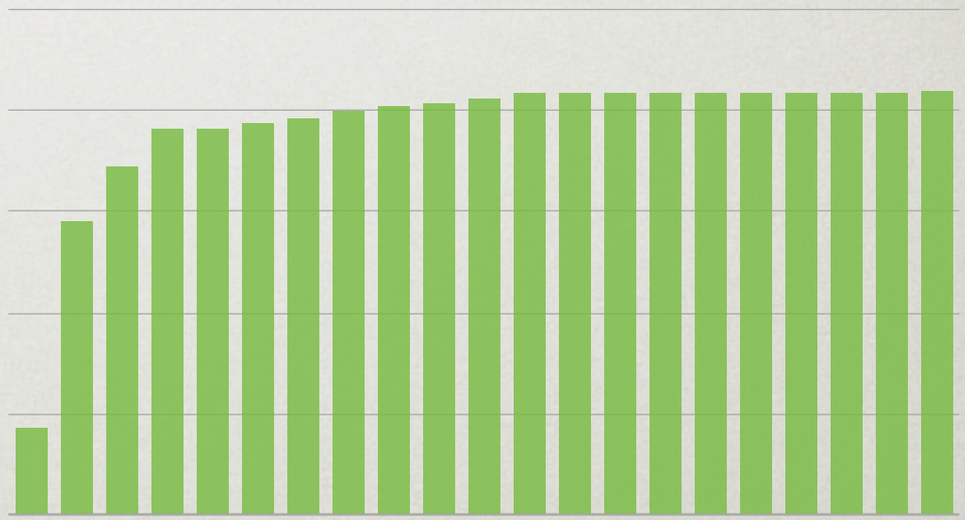
SPASS



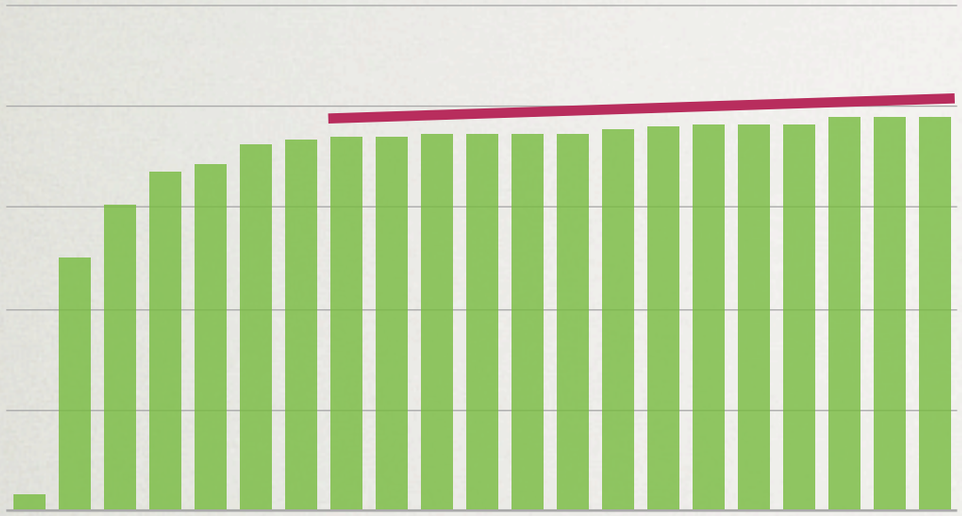
Vampire



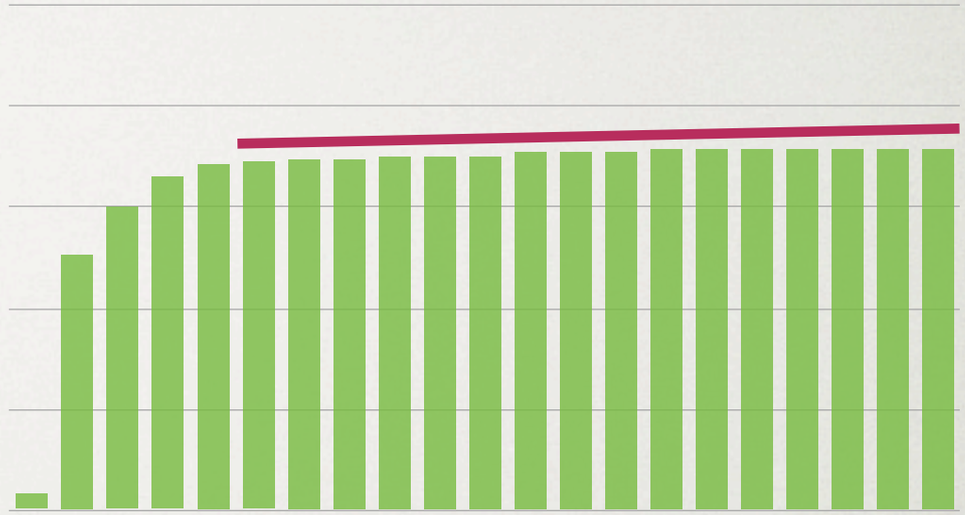
Z3



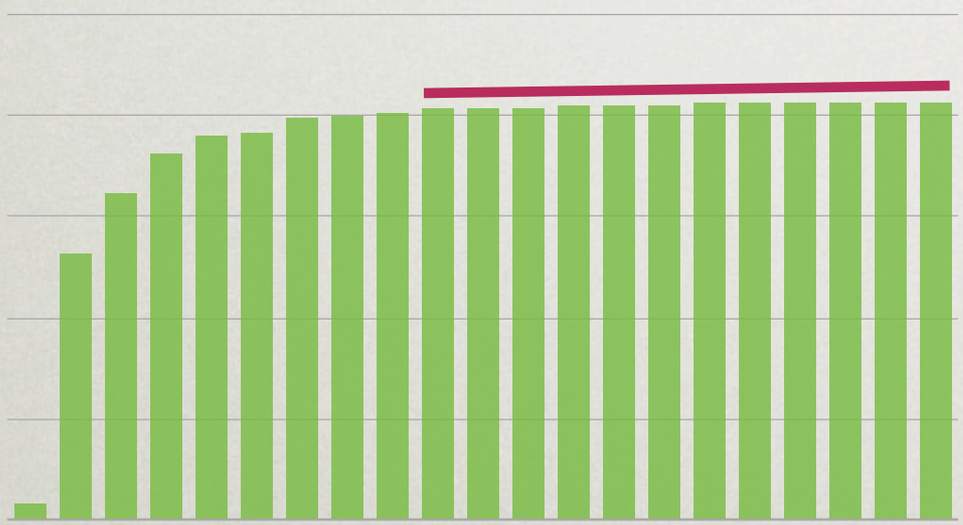
E



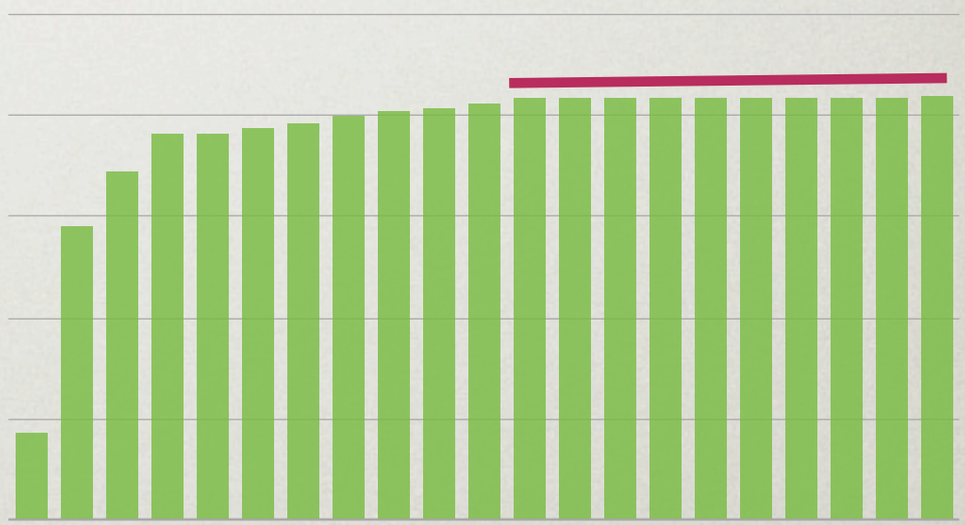
SPASS



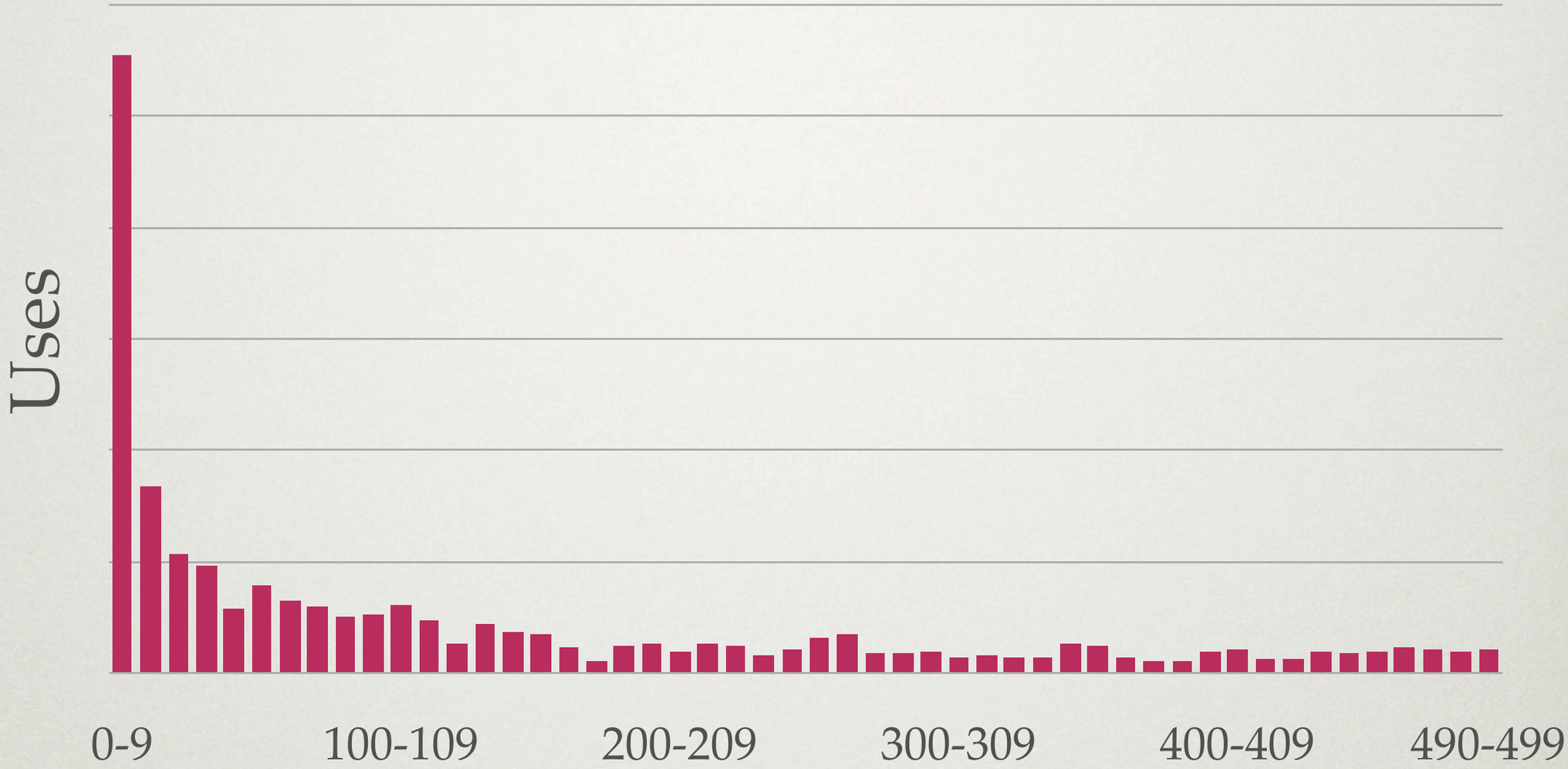
Vampire



Z3



How Effective is the Relevance Filter?



How Effective is the Relevance Filter?

Experiments

Z3 Weights

Experiments

Z3 Weights ..... +0.4 pp

Experiments

Z3 Weights ..... +0.4 pp

Z3 Triggers

Experiments

Z3 Weights ..... +0.4 pp

Z3 Triggers ..... +0.1 pp

Experiments

Z3 Weights ..... +0.4 pp

Z3 Triggers ..... +0.1 pp

Z3 “Slicing”

Experiments

Z3 Weights ..... +0.4 pp

Z3 Triggers ..... +0.1 pp

Z3 “Slicing” ..... +2.1 pp

Experiments

Z3 Weights ..... +0.4 pp

Z3 Triggers ..... +0.1 pp

Z3 “Slicing” ..... +2.1 pp

E Weights

Experiments

Z3 Weights ..... +0.4 pp

Z3 Triggers ..... +0.1 pp

Z3 “Slicing” ..... +2.1 pp

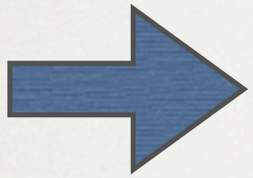
E Weights ..... +1.4 pp

Experiments

Issue #2: Encoding Overhead

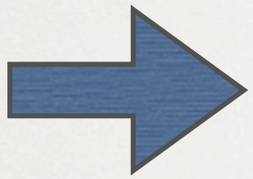
HOL to FOL

Application Operator

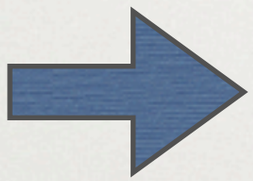
suc(N)  app(suc, N)

HOL to FOL

Application Operator

suc(N)  app(suc, N)

Type Information

suc(N)  ti(suc(ti(N, nat)), nat)

HOL to FOL

Application Operator

suc(N) → app(suc, N)

Type Information

suc(N) → ti(suc(ti(N, nat)), nat)

→ ti(app(ti(suc, fun(nat, nat)),
ti(N, nat)), nat)

HOL to FOL

Problem Size

	No App.	App.
No Types	32K	41K
Types	60K	107K

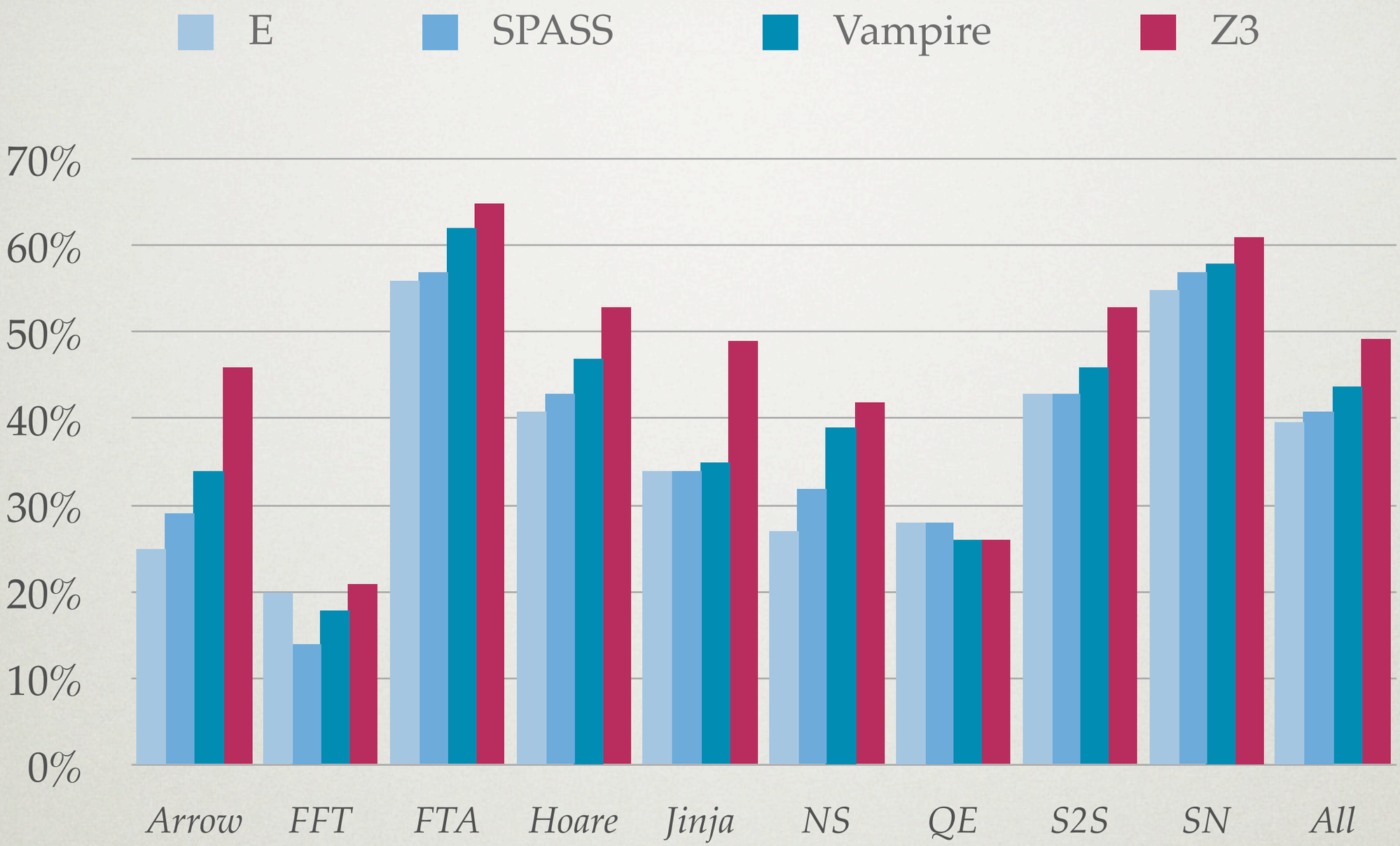
Problem Size

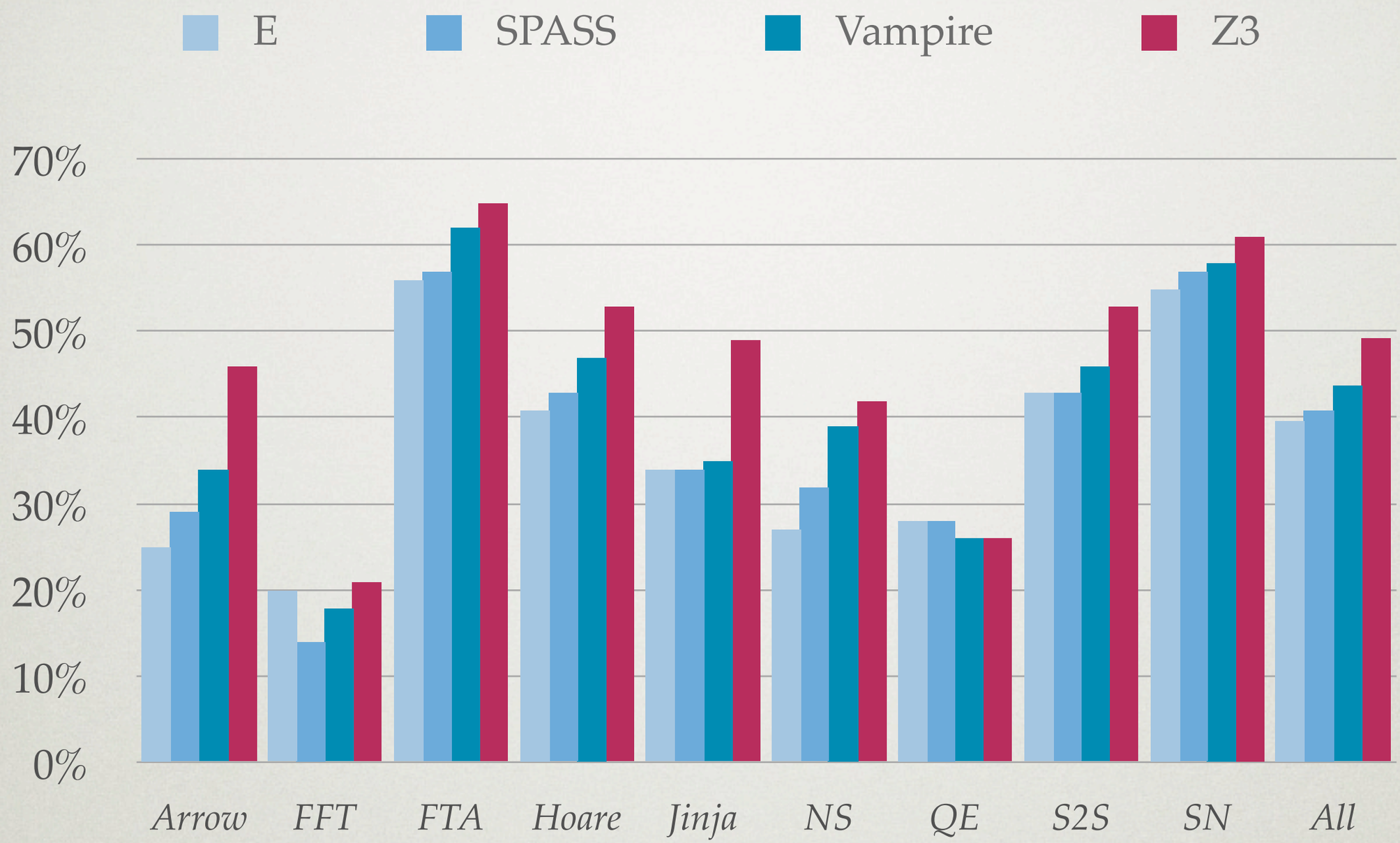
	No App.	App.
No Types	32K	41K
Types	60K	107K

Solving Time (E 1.2)

	No App.	App.
No Types	1 s	15 s
Types	172 s	???? s

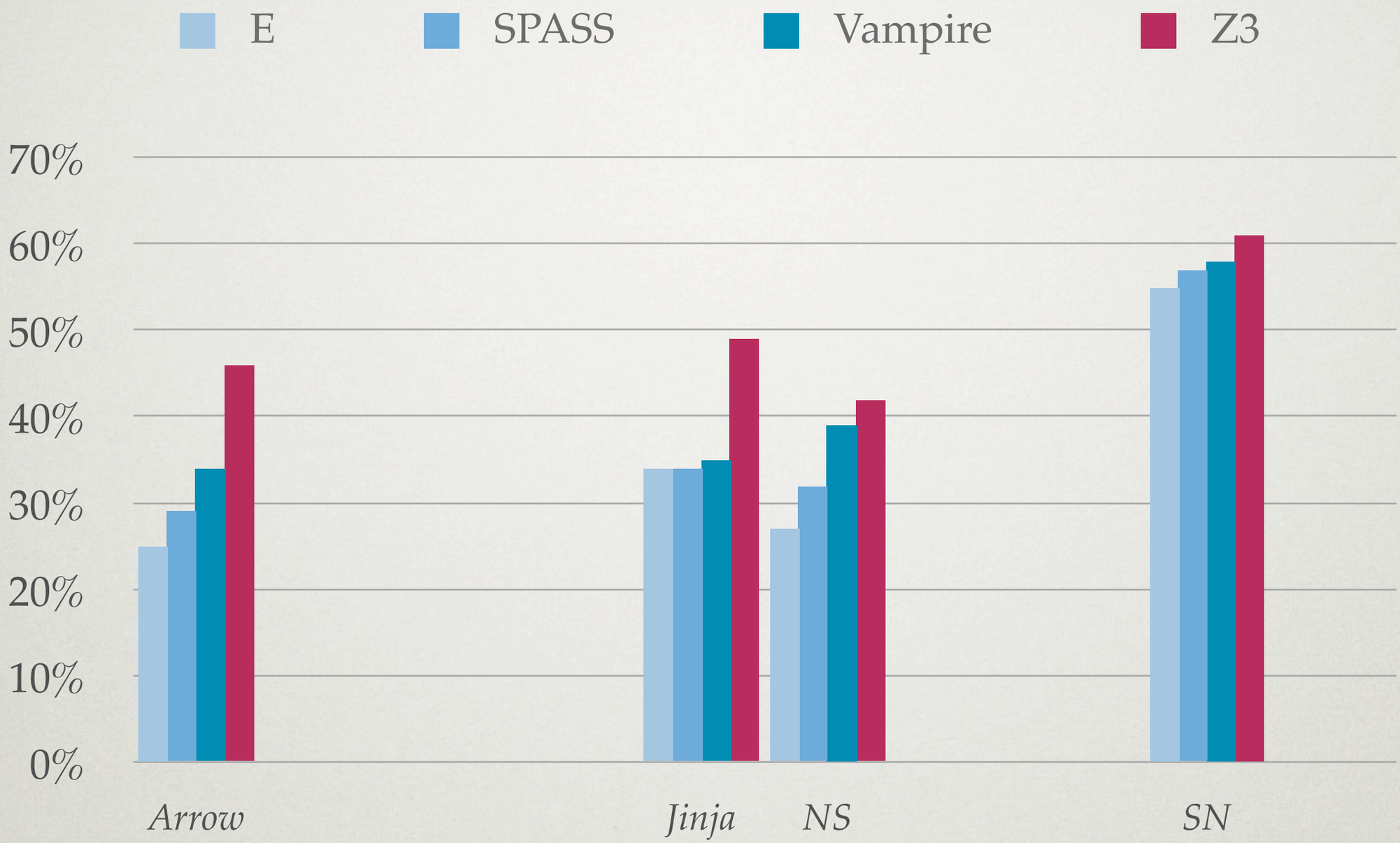
E SPASS Vampire Z3


E
SPASS
Vampire
Z3
70%
60%
50%
40%
30%
20%
10%
0%
Arrow
FFT
FTA
Hoare
Jinja
NS
QE
S2S
SN
All



Claim: Sorts Rock

(Types)



Claim: Sorts Rock

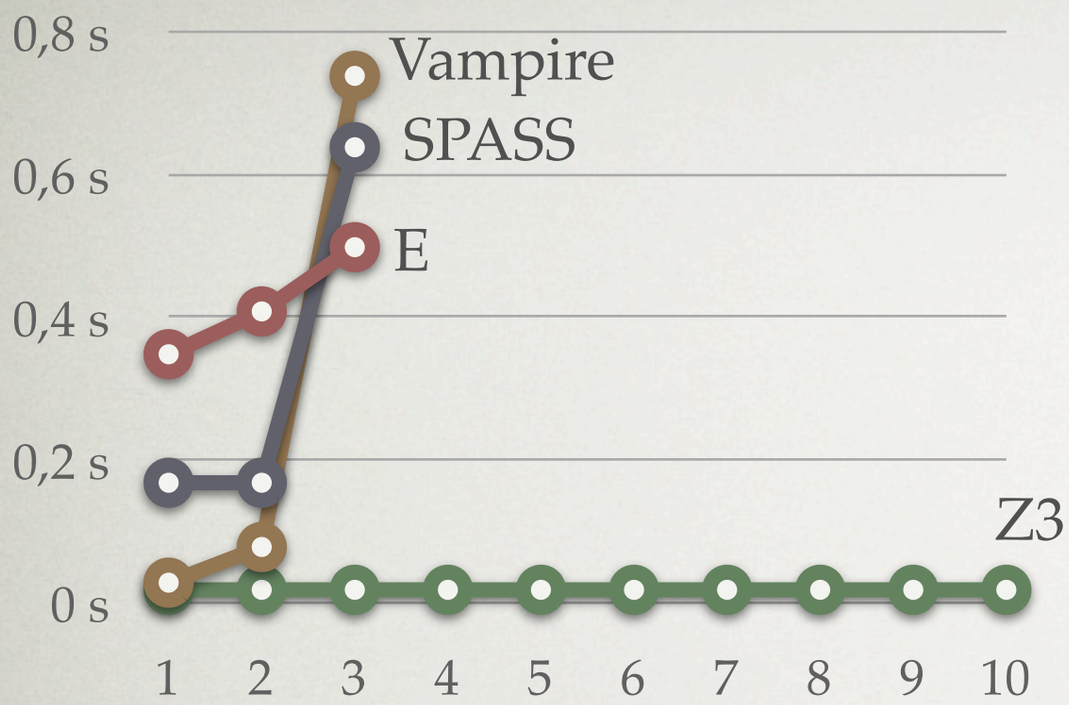
(Types)

Issue #3: Large Terms

$1 + \ldots + 1 = (\mathbf{n}::\mathrm{nat})$

$\mathrm{rev}\ [x_1, \ldots, x_{\mathbf{n}}] = [x_{\mathbf{n}}, \ldots, x_1]$

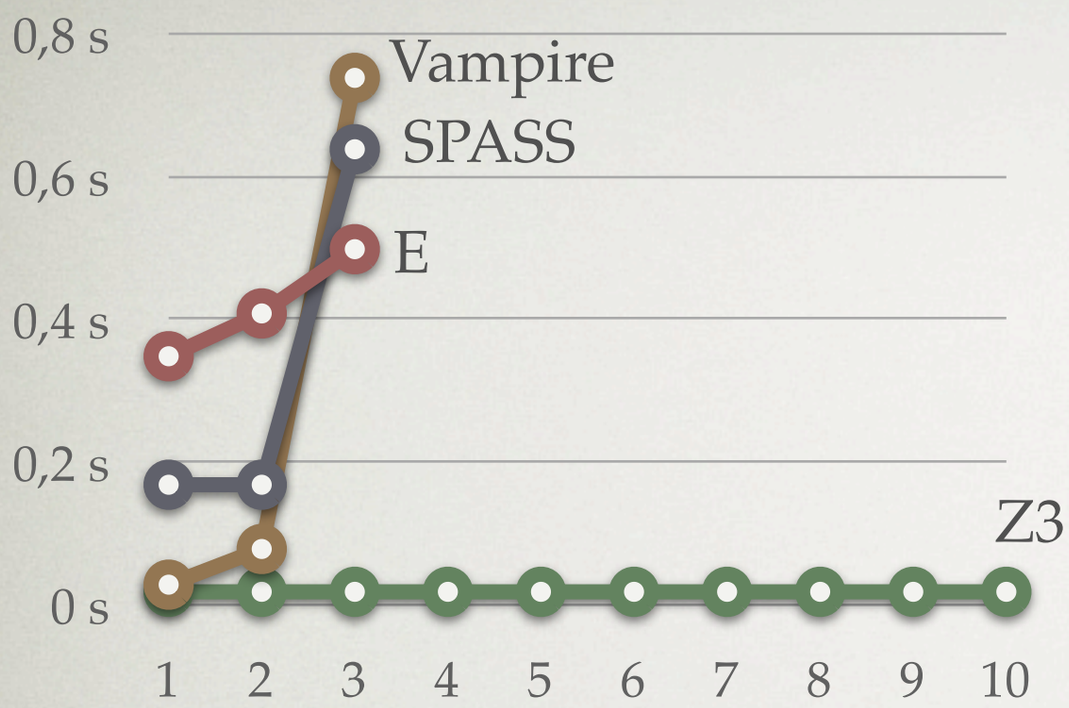
$\mathrm{map}\ f\ [x_1, \ldots, x_{\mathbf{n}}] = [f\ x_1, \ldots, f\ x_{\mathbf{n}}]$



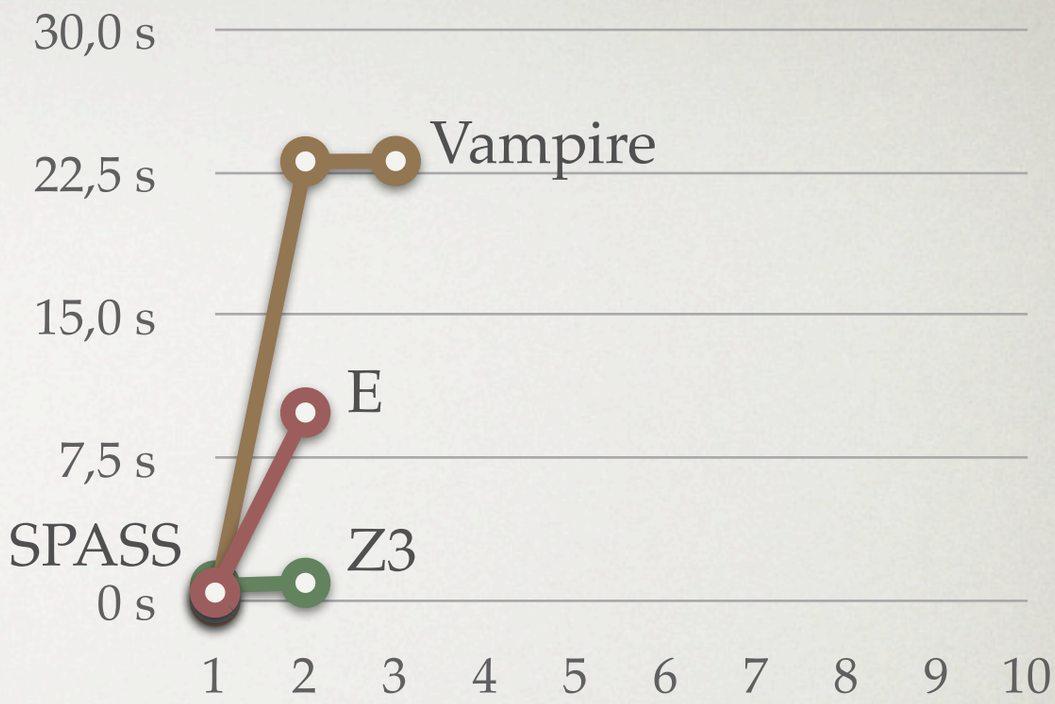
1 + ... + 1 = (**n**::nat)

rev [x_1, ..., $x_{\mathbf{n}}$] = [$x_{\mathbf{n}}$, ..., x_1]

map f [x_1, ..., $x_{\mathbf{n}}$] = [f x_1, ..., f $x_{\mathbf{n}}$]

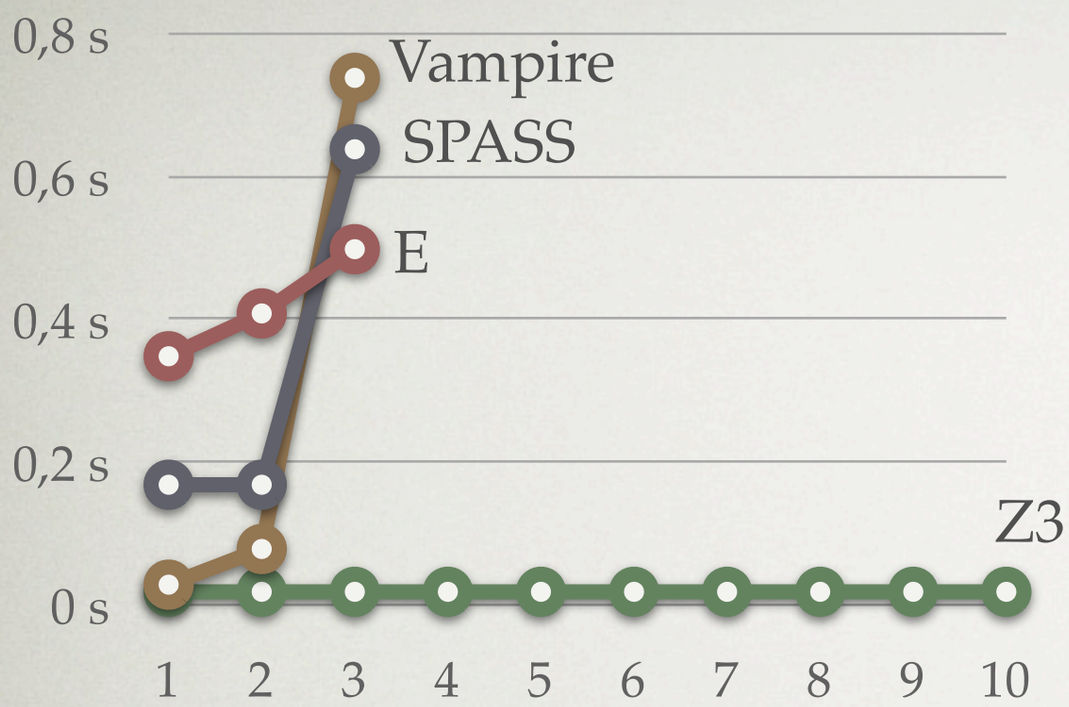


1 + ... + 1 = (**n**::nat)

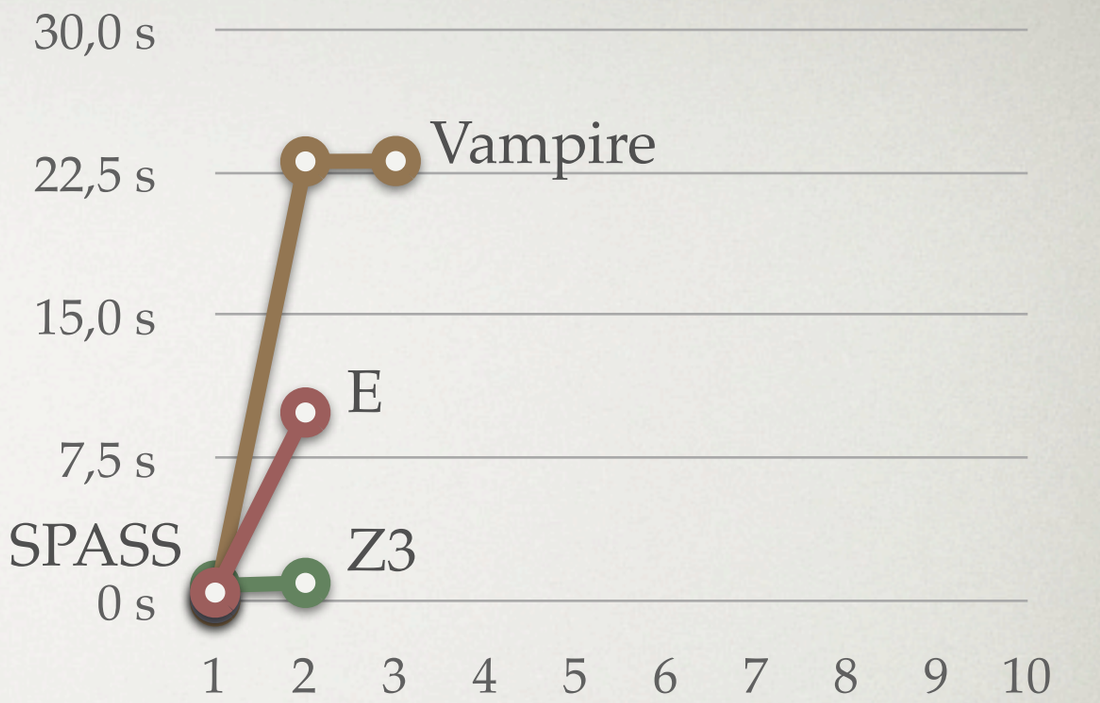


rev [x_1, ..., $x_{\mathbf{n}}$] = [$x_{\mathbf{n}}$, ..., x_1]

map f [x_1, ..., $x_{\mathbf{n}}$] = [f x_1, ..., f $x_{\mathbf{n}}$]

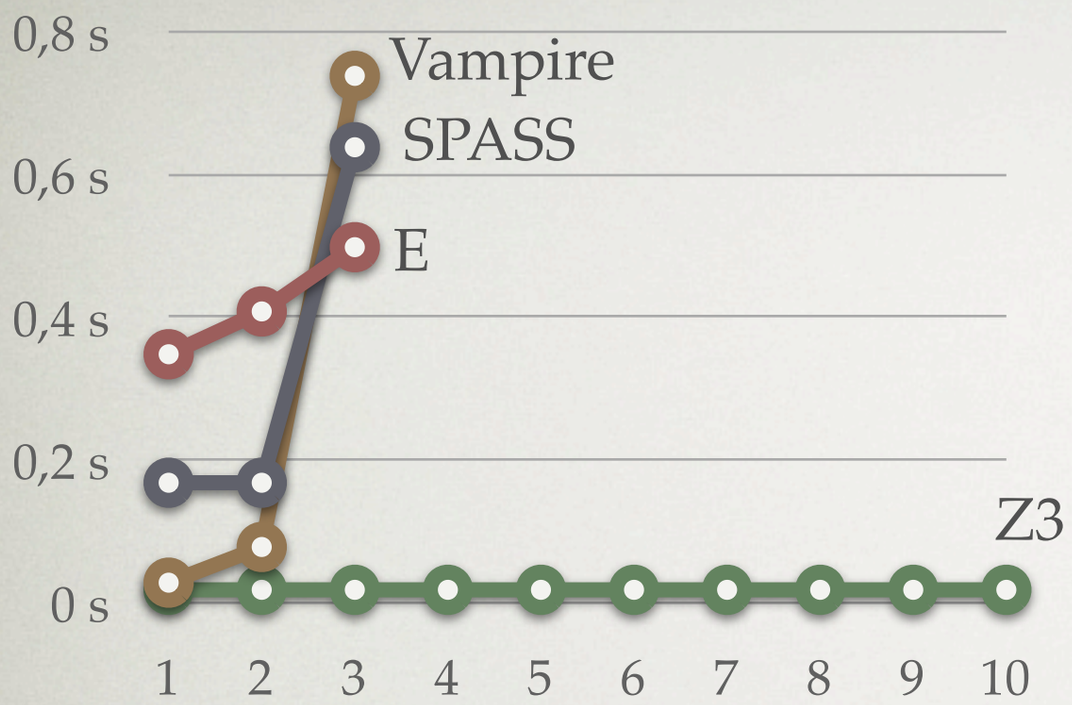


1 + ... + 1 = (**n**::nat)

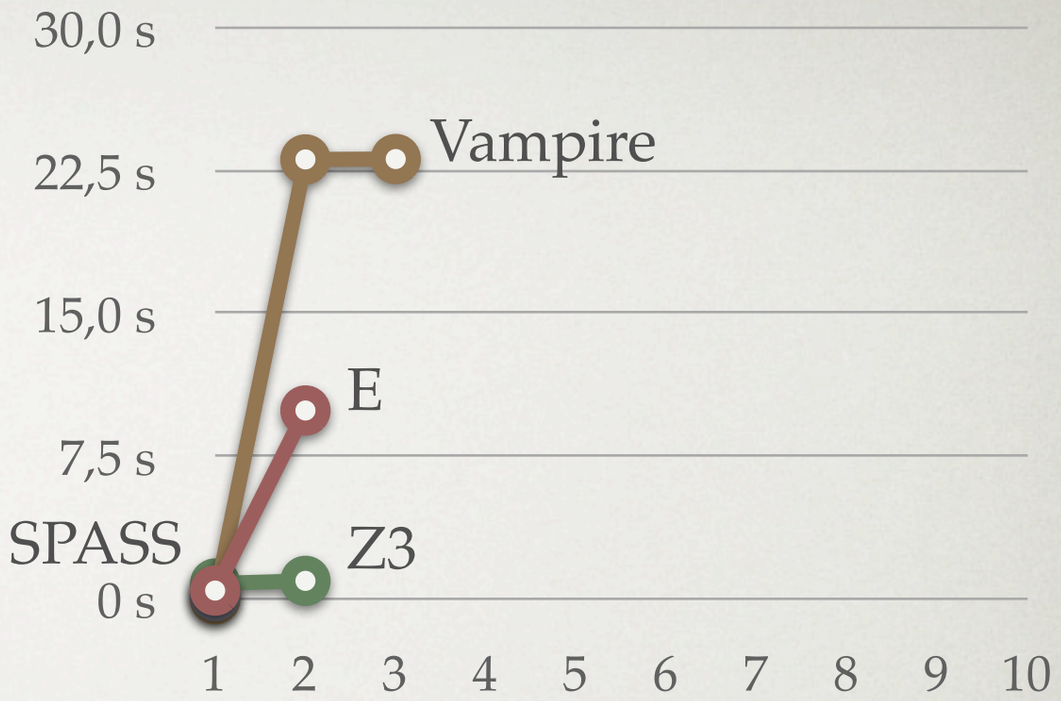


rev [x_1, ..., x_n] = [x_n, ..., x_1]

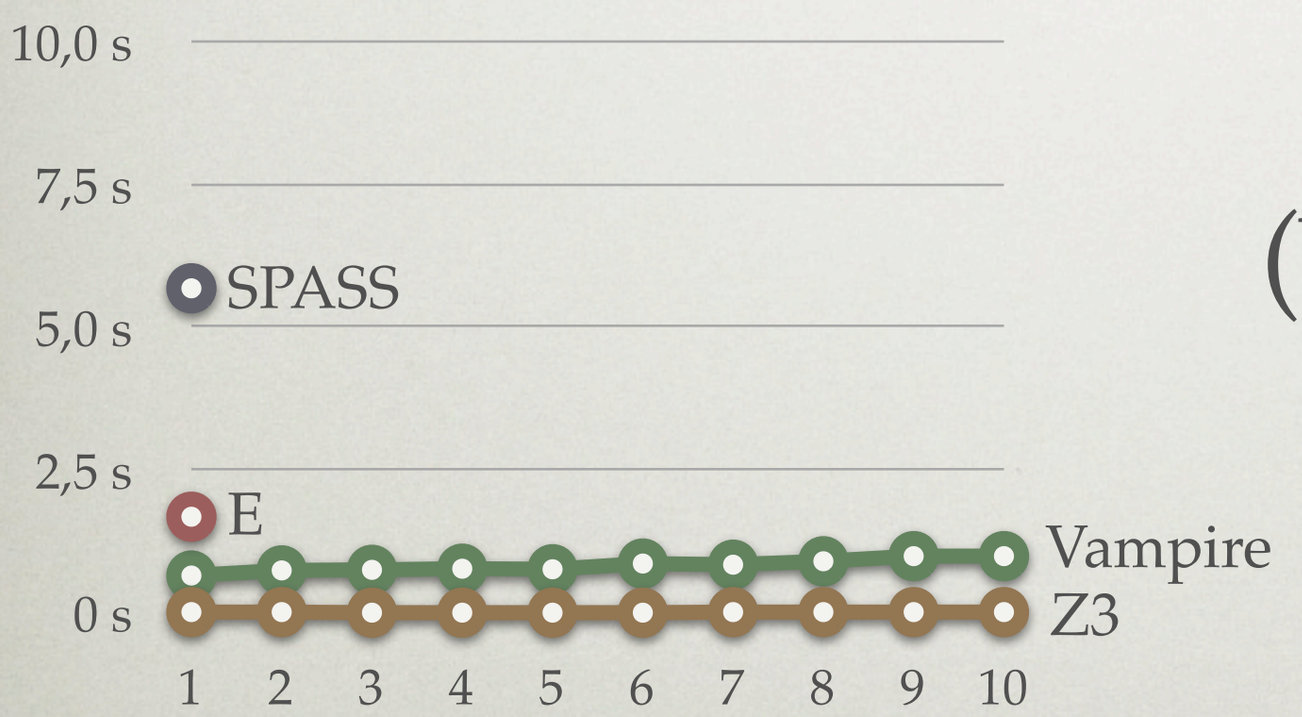
map f [x_1, ..., x_n] = [f x_1, ..., f x_n]



1 + … + 1 = (**n**::nat)



rev [x_1, …, x_n] = [x_n, …, x_1]



map f [x_1, …, x_n] = [f x_1, …, f x_n]

(but simp can solve
all of these
within 10 ms…)

Future Work

Isabelle

ATP / SMT

Future Work

Isabelle

ATP / SMT

Improve
Relevance Filter

Lighten
Translation

Provide
Extralogical Info.

Future Work

Isabelle

Improve
Relevance Filter

Lighten
Translation

Provide
Extralogical Info.

ATP / SMT

Handle Large
Axiom Bases

Support
Types

Exploit
Extralogical Info.

Future Work

Sledgehammer Hell
The Day after Judgment

Jasmin C. Blanchette
blanchette@in.tum.de