Human Competitiveness of Genetic Programming for Spectrum Based Fault Localisation

Shin Yoo[1], Xiaoyuan Xie[2], Fei-Ching Kuo[3], Tsong Yueh Chen[3], Mark Harman[4]

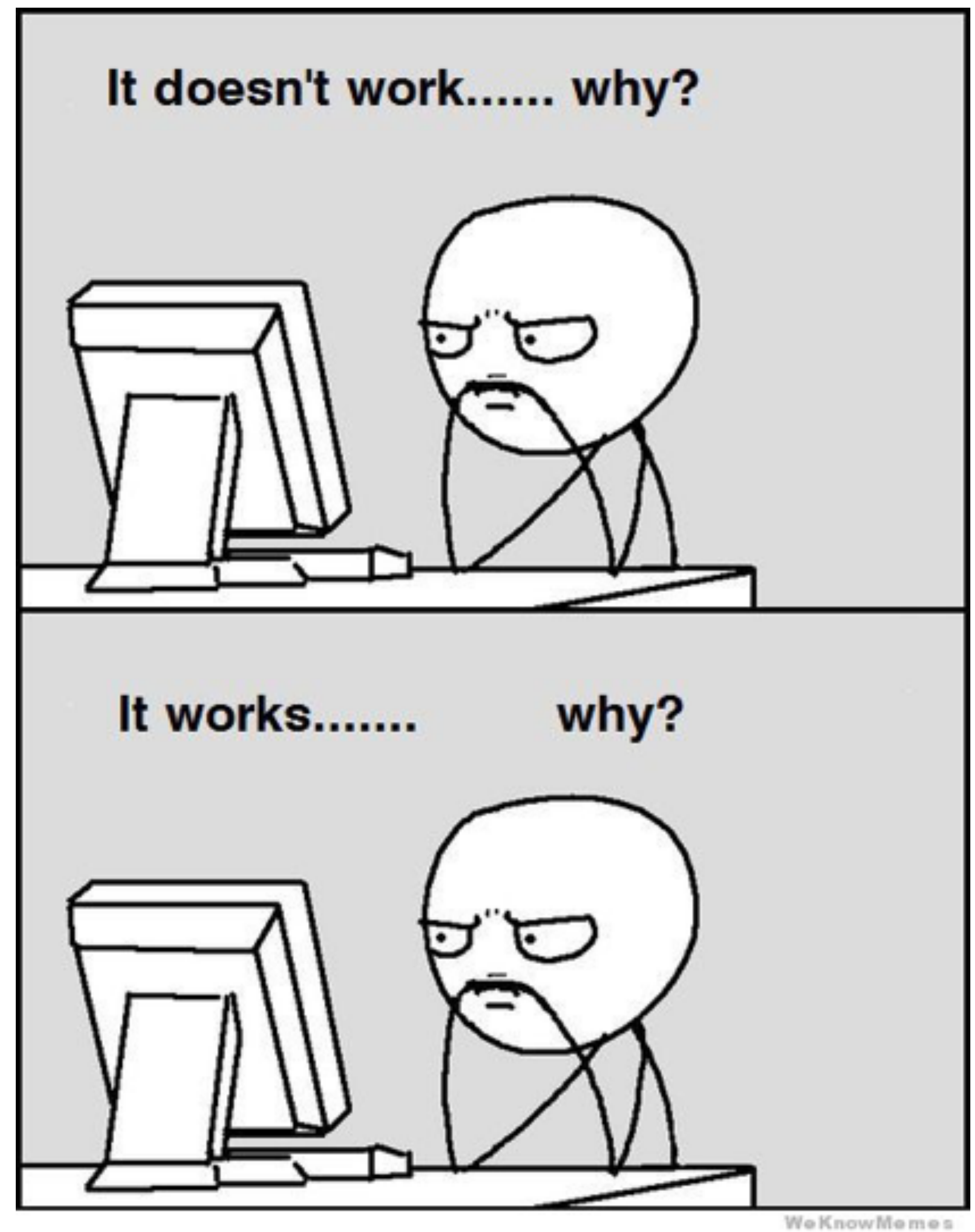
1: KAIST, Republic of Korea
2: Wuhan University, China
3: Swinburne University of Technology, Australia
4: University College London/Facebook, UK

HUMIES@GECCO 2017

Automated Debugging

- Debugging is **hard for humans:** we increasingly have to work on and with large code base written by others.

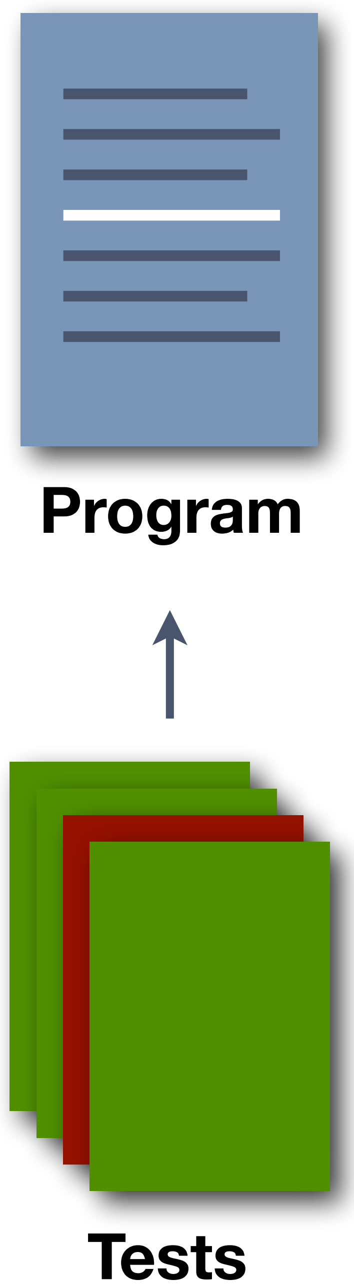
- Debugging is **hard for machines**: automated repair techniques rely heavily on automated fault localisation.



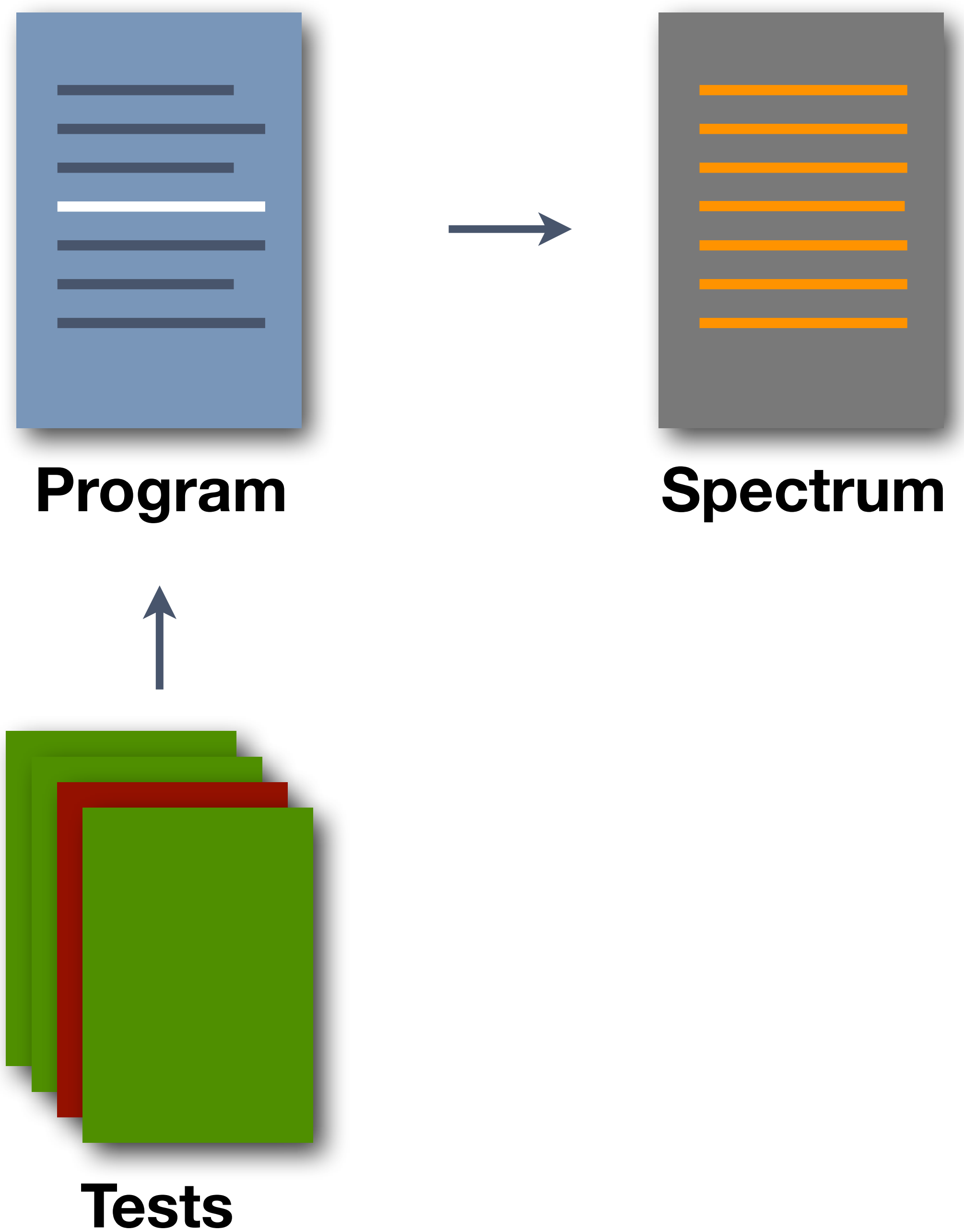
Spectrum Based Fault Localisation

Program

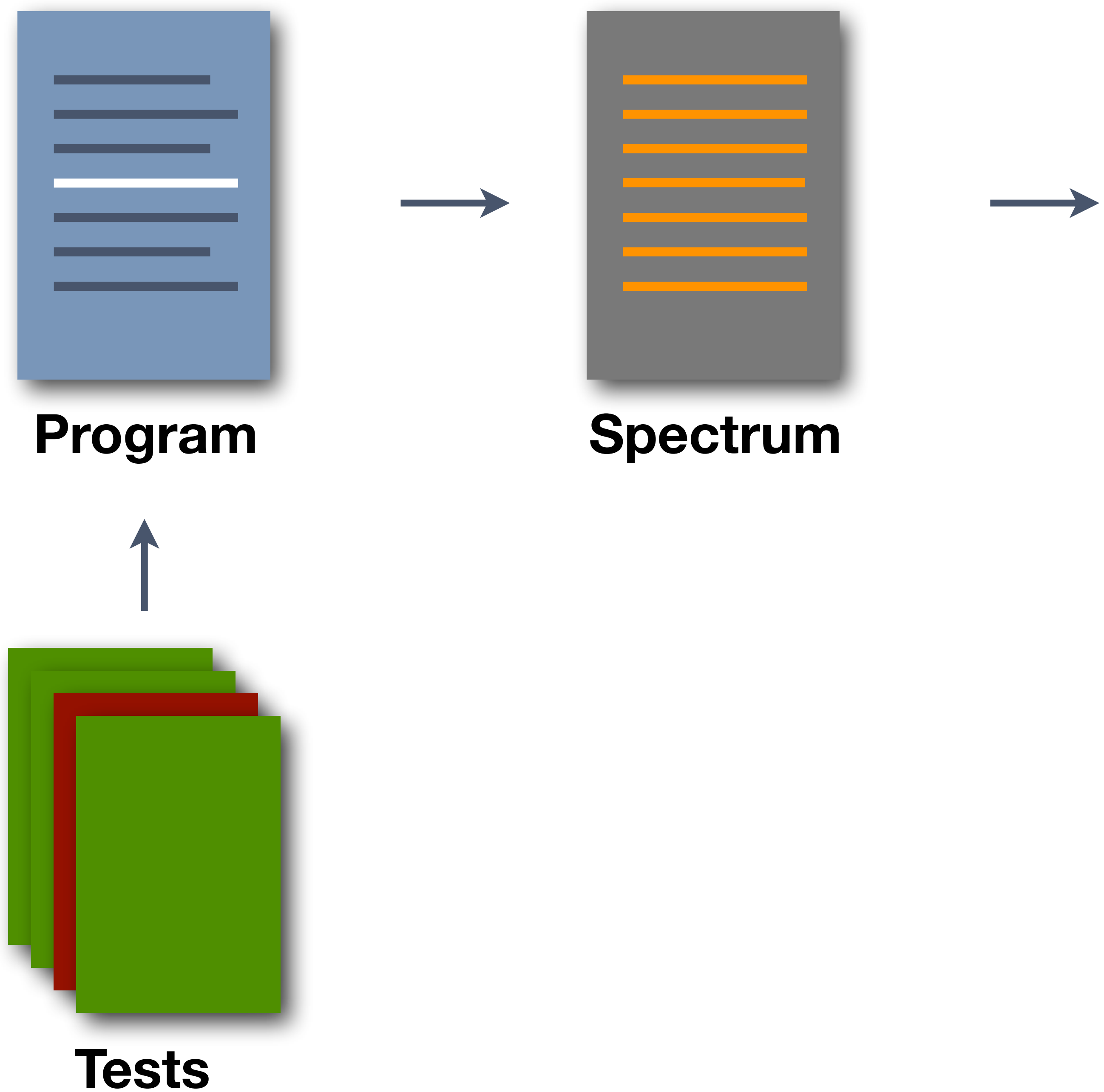
Spectrum Based Fault Localisation



Spectrum Based Fault Localisation



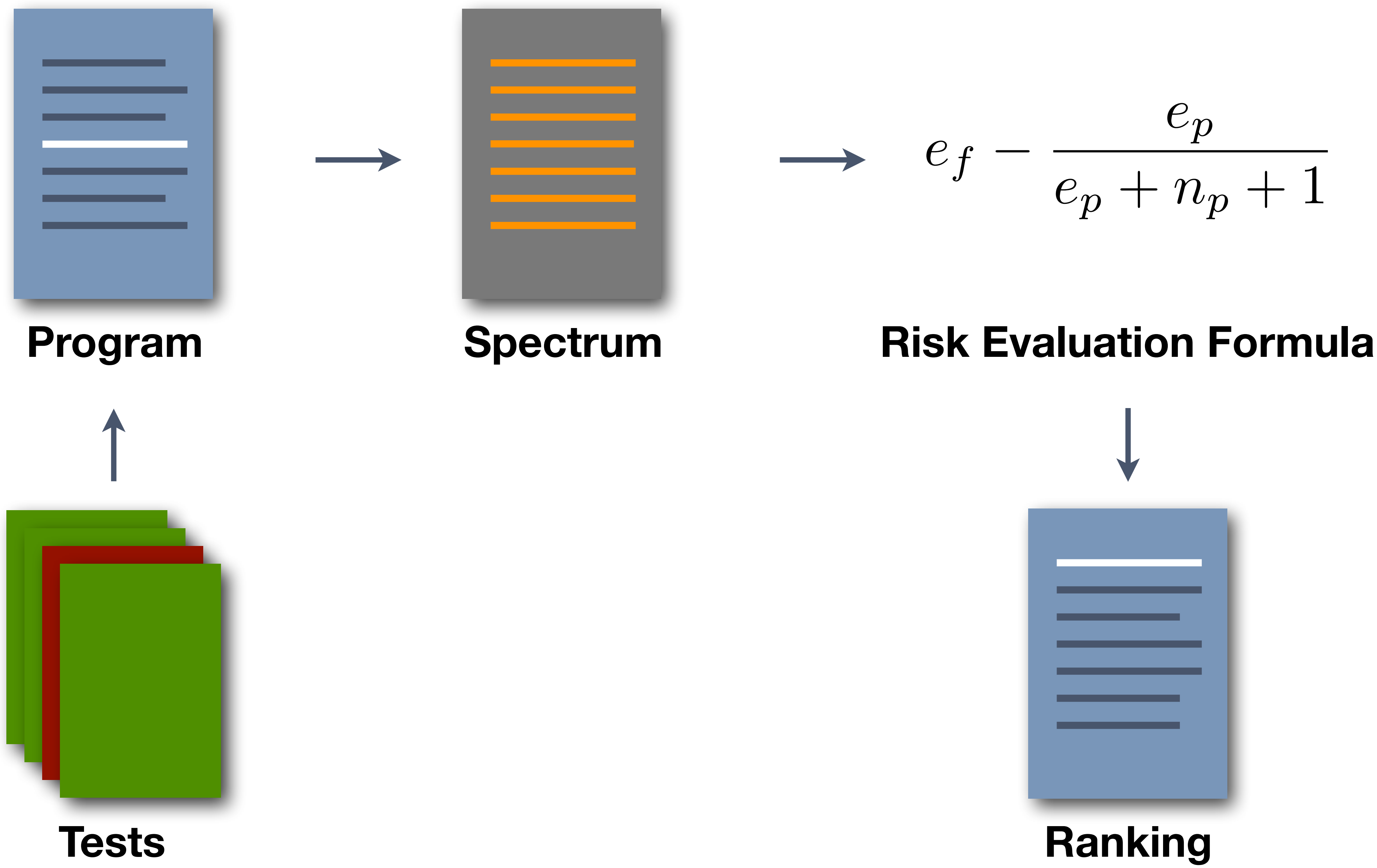
Spectrum Based Fault Localisation



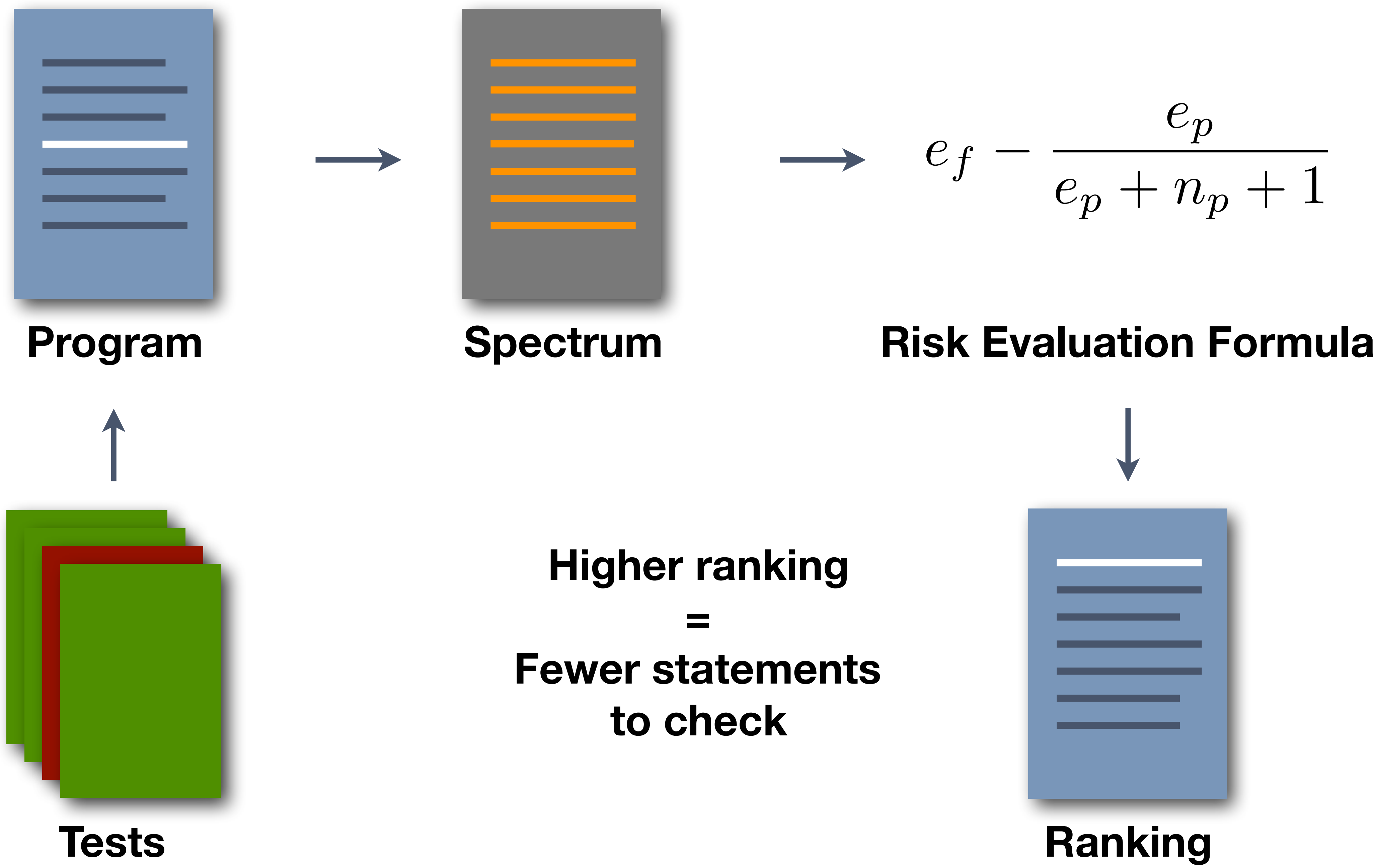
$$e_f - \frac{e_p}{e_p + n_p + 1}$$

Risk Evaluation Formula

Spectrum Based Fault Localisation



Spectrum Based Fault Localisation



Spectrum Based Fault Localisation

Structural Elements	Test t_1	Test t_2	Test t_3	Spectrum e_p	e_f	n_p	n_f	Tarantula	Rank
s_1	•			1	0	0	2	0.00	9
s_2	•			1	0	0	2	0.00	9
s_3	•			1	0	0	2	0.00	9
s_4	•			1	0	0	2	0.00	9
s_5	•			1	0	0	2	0.00	9
s_6	•		•	1	1	0	1	0.33	4
s_7 (faulty)		•	•	0	2	1	0	1.00	1
s_8	•	•		1	1	0	1	0.33	4
s_9	•	•	•	1	2	0	0	0.50	2
Result	P	F	F						

Spectrum Based Fault Localisation

Structural Elements	Test t_1	Test t_2	Test t_3	Spectrum e_p	e_f	n_p	n_f	Tarantula	Rank
s_1	•			1	0	0	2	0.00	9
s_2	•			1	0	0	2	0.00	9
s_3	•			1	0	0	2	0.00	9
s_4	•			1	0	0	2	0.00	9
s_5	•			1	0	0	2	0.00	9
s_6	•		•	1	1	0	1	0.33	4
s_7 (faulty)		•	•	0	2	1	0	1.00	1
s_8	•	•		1	1	0	1	0.33	4
s_9	•	•	•	1	2	0	0	0.50	2
Result	P	F	F						

Spectrum Based Fault Localisation

Structural Elements	Test t_1	Test t_2	Test t_3	Spectrum e_p	e_f	n_p	n_f	Tarantula	Rank
s_1	•			1	0	0	2	0.00	9
s_2	•			1	0	0	2	0.00	9
s_3	•			1	0	0	2	0.00	9
s_4	•			1	0	0	2	0.00	9
s_5	•			1	0	0	2	0.00	9
s_6	•		•	1	1	0	1	0.33	4
s_7 (faulty)		•	•	0	2	1	0	1.00	1
s_8	•	•		1	1	0	1	0.33	4
s_9	•	•	•	1	2	0	0	0.50	2
Result	P	F	F						

Spectrum Based Fault Localisation

Structural Elements	Test t_1	Test t_2	Test t_3	Spectrum e_p	e_f	n_p	n_f	Tarantula	Rank
s_1	•			1	0	0	2	0.00	9
s_2	•			1	0	0	2	0.00	9
s_3	•			1	0	0	2	0.00	9
s_4	•			1	0	0	2	0.00	9
s_5	•			1	0	0	2	0.00	9
s_6	•		•	1	1	0	1	0.33	4
s_7 (faulty)		•	•	0	2	1	0	1.00	1
s_8	•	•		1	1	0	1	0.33	4
s_9	•	•	•	1	2	0	0	0.50	2
Result	P	F	F						

$$\text{Tarantula} = \frac{\frac{e_f}{e_f+n_f}}{\frac{e_p}{e_p+n_p} + \frac{e_f}{e_f+n_f}}$$

Spectrum Based Fault Localisation

Structural Elements	Test t_1	Test t_2	Test t_3	Spectrum e_p	e_f	n_p	n_f	Tarantula	Rank
s_1	•			1	0	0	2	0.00	9
s_2	•			1	0	0	2	0.00	9
s_3	•			1	0	0	2	0.00	9
s_4	•			1	0	0	2	0.00	9
s_5	•			1	0	0	2	0.00	9
s_6	•		•	1	1	0	1	0.33	4
s_7 (faulty)		•	•	0	2	1	0	1.00	1
s_8	•	•		1	1	0	1	0.33	4
s_9	•	•	•	1	2	0	0	0.50	2
Result	P	F	F						

Spectrum Based Fault Localisation

Structural Elements	Test t_1	Test t_2	Test t_3	Spectrum e_p	e_f	n_p	n_f	Tarantula	Rank
s_1	•			1	0	0	2	0.00	9
s_2	•			1	0	0	2	0.00	9
s_3	•			1	0	0	2	0.00	9
s_4	•			1	0	0	2	0.00	9
s_5	•			1	0	0	2	0.00	9
s_6	•		•	1	1	0	1	0.33	4
s_7 (faulty)		•	•	0	2	1	0	1.00	1
s_8	•	•		1	1	0	1	0.33	4
s_9	•	•	•	1	2	0	0	0.50	2
Result	P	F	F						

Spectrum Based Fault Localisation

Structural Elements	Test t_1	Test t_2	Test t_3	Spectrum e_p	e_f	n_p	n_f	Tarantula	Rank
s_1	•			1	0	0	2	0.00	9
s_2	•			1	0	0	2	0.00	9
s_3	•			1	0	0	2	0.00	9
s_4	•			1	0	0	2	0.00	9
s_5	•			1	0	0	2	0.00	9
s_6	•		•	1	1	0	1	0.33	4
s_7 (faulty)		•	•	0	2	1	0	1.00	1
s_8	•	•		1	1	0	1	0.33	4
s_9	•	•	•	1	2	0	0	0.50	2
Result	P	F	F						

(Empirical) State of the Art (circa 2012)

Over 30 formulæ in the literature,
manually developed over a decade's time:
none guaranteed to perform best for all types of faults

(Empirical) State of the Art (circa 2012)

$$\frac{2e_f}{e_f + n_f + e_p}$$

$$\frac{e_f}{e_f + n_p + 2(e_p + n_f)}$$

$$\frac{e_f}{e_f + n_f + e_p}$$

$$\frac{2(e_f + n_p)}{2(e_f + n_p) + e_p + n_f}$$

$$\frac{e_f}{e_f + 2(n_f + e_p)}$$

$$\frac{e_f}{n_f + e_p}$$

Over 30 formulæ in the literature,
manually developed over a decade's time:
none guaranteed to perform best
for all types of faults

$$\frac{e_f + n_p}{n_f + e_p}$$

$$\frac{e_f}{e_f + n_f + e_p + n_p}$$

$$\frac{e_f + n_p}{e_f + n_f + e_p + n_p}$$

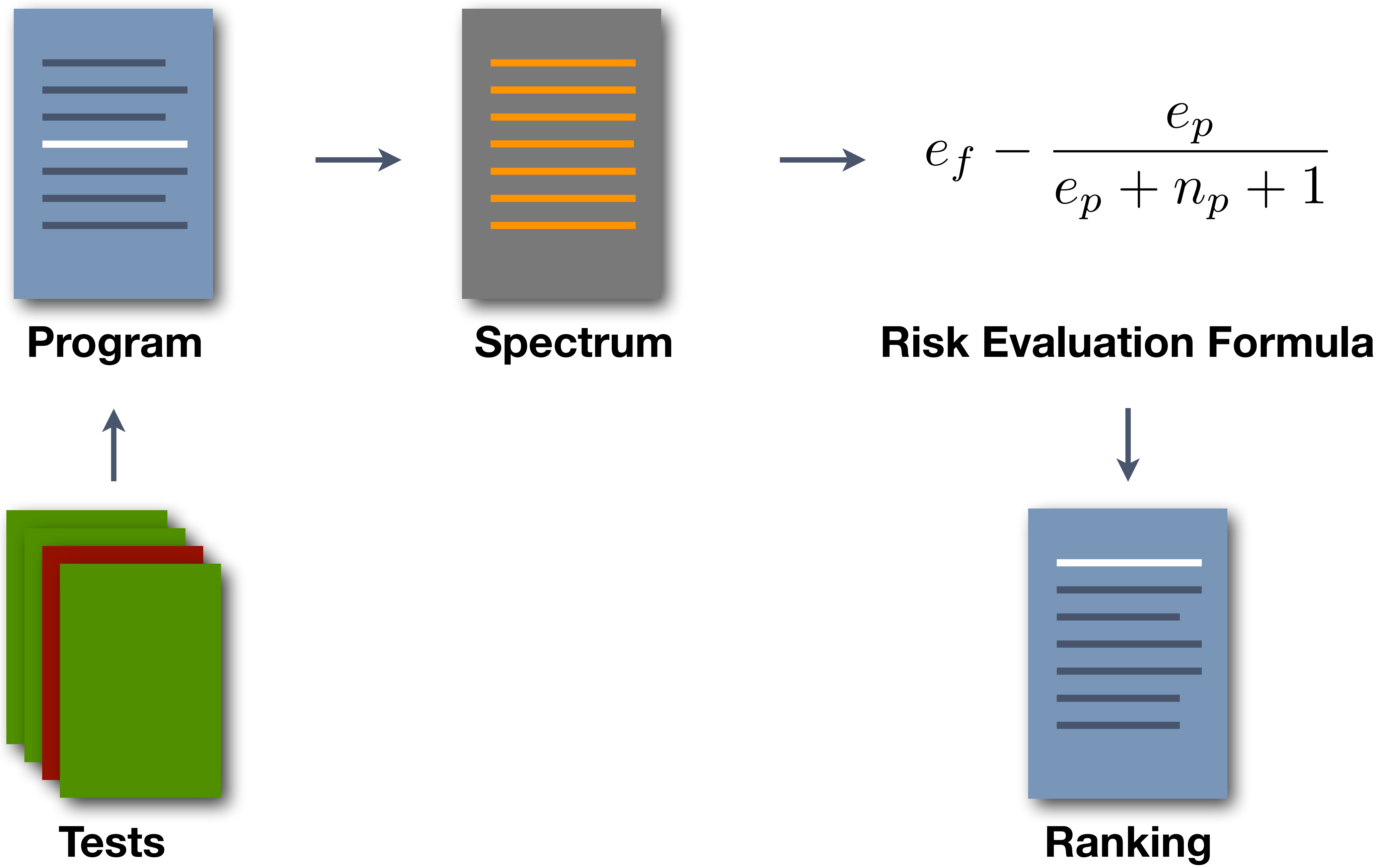
$$\frac{2e_f}{2e_f + n_f + e_p}$$

$$\frac{e_f + n_p - n_f - e_p}{e_f + n_f + e_p + n_p}$$

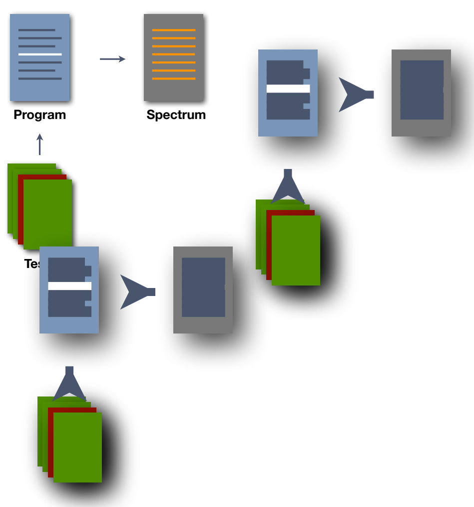
$$\frac{1}{2}\left(\frac{e_f}{e_f + n_f} + \frac{e_f}{e_f + e_p}\right)$$

$$\frac{\frac{e_f}{e_f+n_f}}{\frac{e_p}{e_p+n_p} + \frac{e_f}{e_f+n_f}}$$

Evolving Formulæ



Evolving Formulæ



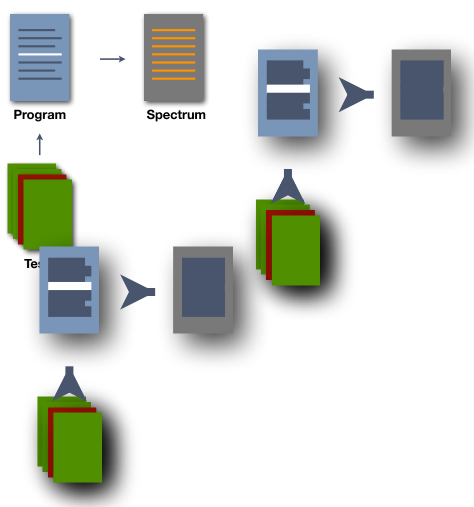
Training Data

$$e_f - \frac{e_p}{e_p + n_p + 1}$$

Risk Evaluation Formula

Ranking

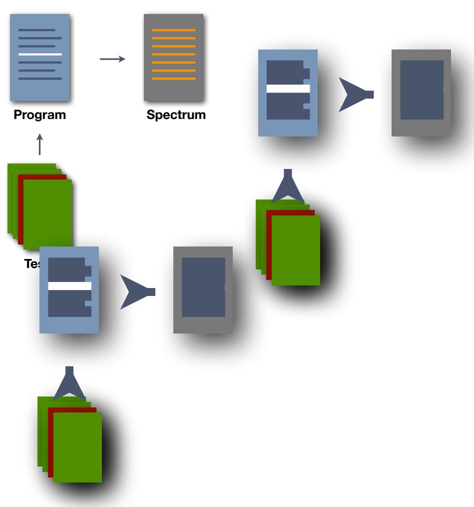
Evolving Formulæ



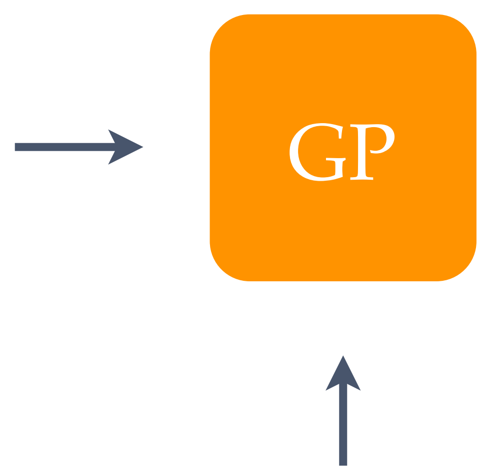
Training Data

Ranking

Evolving Formulæ

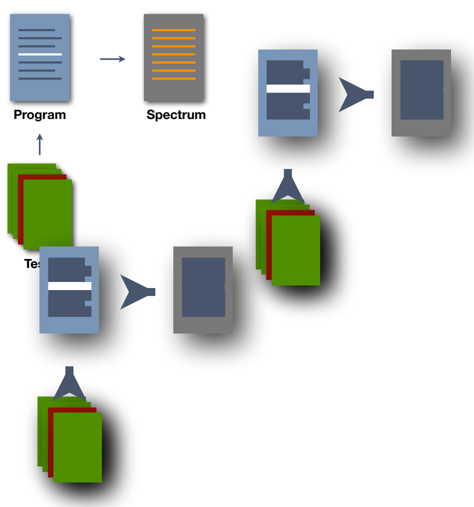


Training Data



Fitness
(minimise)

Evolving Formulæ



Training Data

GP

$$e_f^2(2e_p + 2e_f + 3n_p)$$

$$e_f^2(e_f^2 + \sqrt{n_p})$$

$$\ldots$$

Fitness (minimise)

Our Claims

Our Claims

GP evolved SBFL formulas are **provably better** than many human designs.

Our Claims

GP evolved SBFL formulas are **provably better** than many human designs.

We proved that **no human can surpass what GP evolved**, ever.

Our Claims

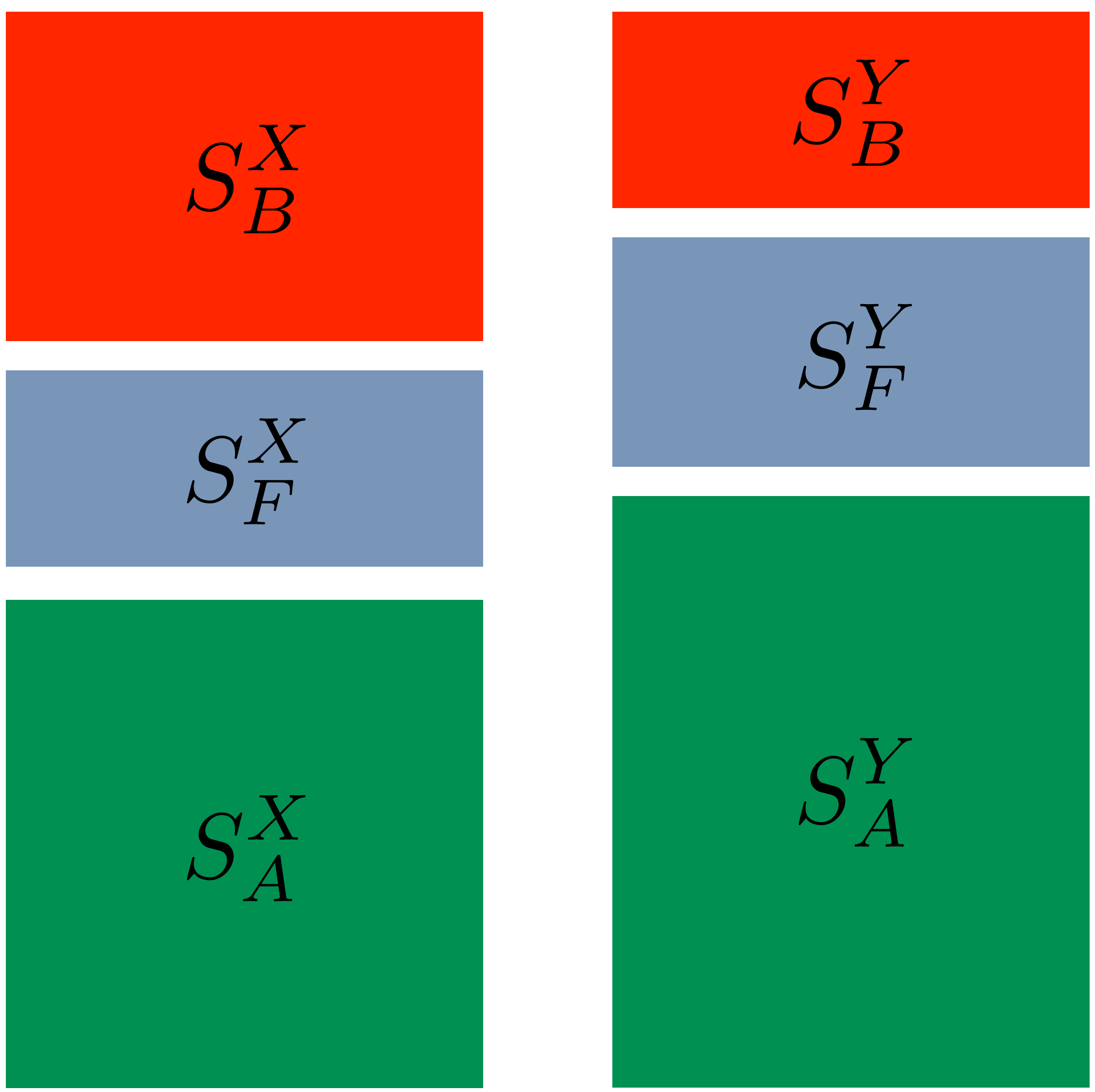
GP evolved SBFL formulas are **provably better** than many human designs.

We proved that **no human can surpass what GP evolved**, ever.

GP has **transformed the future research** on fault localisation.

Crash Course into Our Proof System

Statement Ranking



Formula X **Formula Y**

To show that Y dominates X, we show that:

$$S_B^Y \subseteq S_B^X \wedge S_A^X \subseteq S_A^Y$$

(assuming that we break ties in F sets consistently)

Equivalence is defined as:

$$X \leftrightarrow Y \iff X \rightarrow Y \wedge Y \rightarrow X$$

Crash Course into Our Proof System

- **Maximal** Groups: a set of formulas that are equivalent to each other, but are strictly better to some others
- Previous work theoretically proved the existence of maximal groups with respect to the space of known formulas:
  - ER1 (contains 2 manually designed formulas) and ER5 (contains 3 manually designed formulas)

GP's Human Competitiveness

- **GP expanded the known maximal groups:**
  - GP added one additional formula to ER1
  - GP founded three new maximal groups, each containing one GP-evolved formula

Name		Formula expression
ER1'	Naish1	$\begin{cases} -1 & \text{if } e_f < F \\ P - e_p & \text{if } e_f = F \end{cases}$
	Naish2	$e_f - \frac{e_p}{e_p + n_p + 1}$
	GP13	$e_f(1 + \frac{1}{2e_p + e_f})$
ER5	Wong1	e_f
	Russel & Rao	$\frac{e_f}{e_f + n_f + e_p + n_p}$
	Binary	$\begin{cases} 0 & \text{if } e_f < F \\ 1 & \text{if } e_f = F \end{cases}$
GP02		$2(e_f + \sqrt{n_p}) + \sqrt{e_p}$
GP03		$\sqrt{\lvert e_f^2 - \sqrt{e_p} \rvert}$
GP19		$e_f \sqrt{\lvert e_p - e_f + n_f - n_p \rvert}$

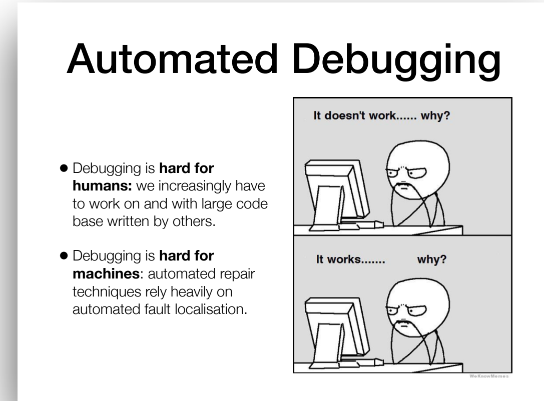
GP's Human Competitiveness

- We have **proved** that there is no greatest formula (i.e. one that outperforms all maximals):
  - GP evolved the best possible formula.
  - **No future human endeavour can surpass GP's results.**

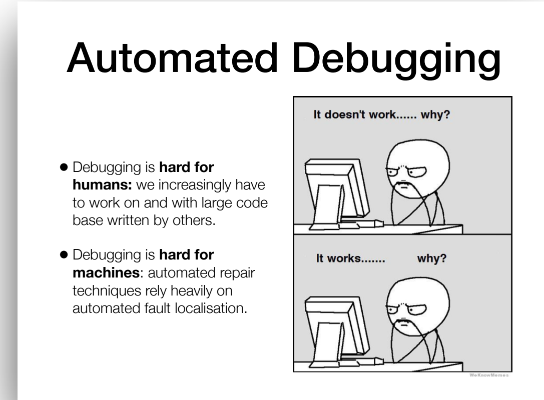
Q.E.D.

GP's Influence on Future Research

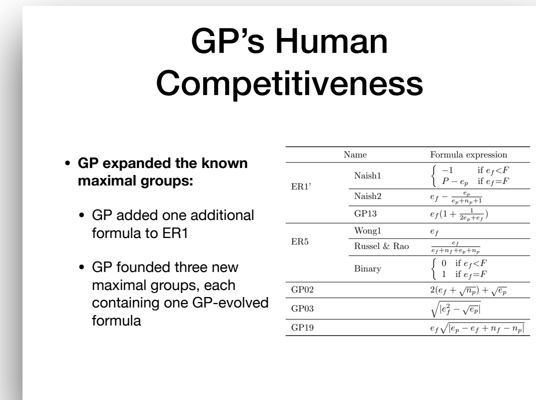
- Manually designing SBFL formulae is **no longer productive**.
- We need richer information than program spectrum: **GP can deal with increased complexity** better than human.
- **GP continues to produce state-of-the-art localisation results**, outperforming SVMs (ISSTA 2017).



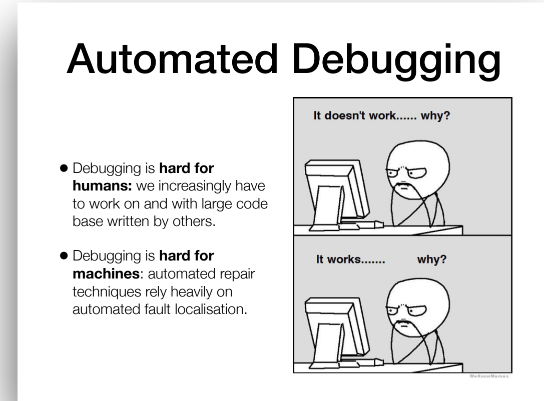
Debugging is hard.



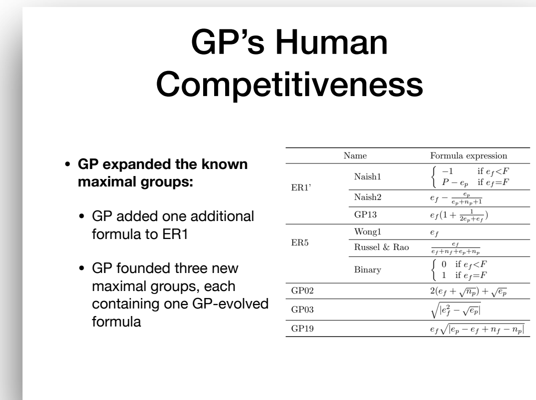
Debugging is hard.



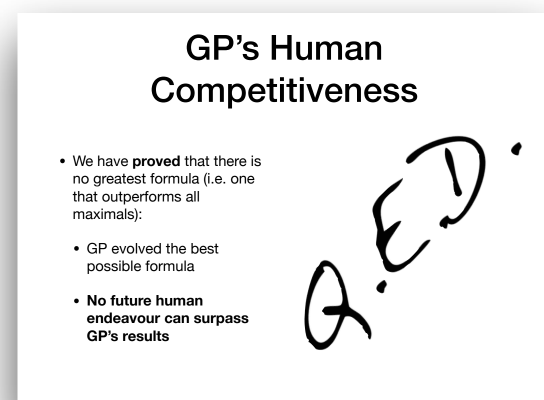
GP-evolved fault localisation techniques were provably better than over a decade's manual work.



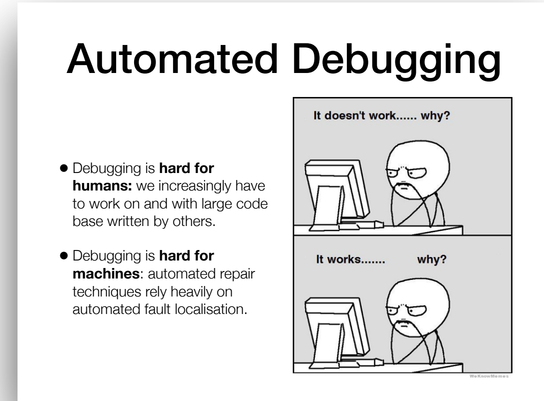
Debugging is hard.



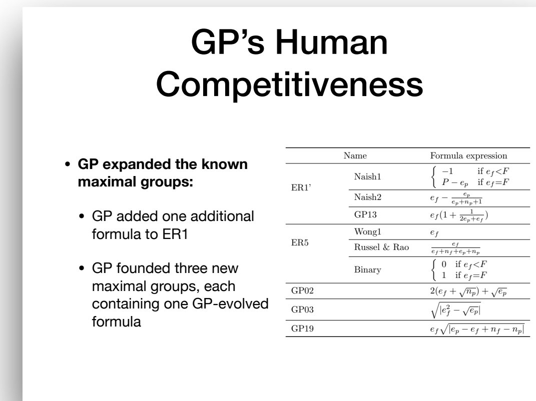
GP-evolved fault localisation techniques were provably better than over a decade's manual work.



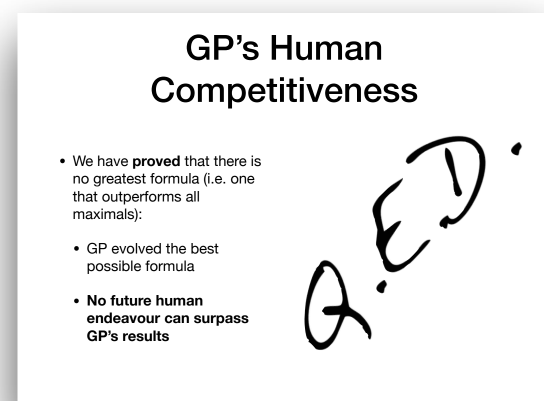
We proved that no human can surpass GP results, ever.



Debugging is hard.



GP-evolved fault localisation techniques were provably better than over a decade's manual work.



We proved that no human can surpass GP results, ever.

GP continues to have strong influence on future research.