Hinton & Nowlan's Computational Baldwin Effect Revisit:

Are we happy with it?

Akira Imada

Brest State Technical University

Belarus

Looking for "a-needle-in-a-haystack"

To break an unknown 4 digit decimal PIN

Only one out of 10,000 is correct (a-needle-in-a-haystack)

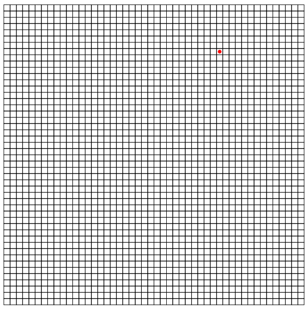
$\Downarrow$

How many random/exhaustive trials
will be necessary to break?

$\Downarrow$

An average of 5,000 trials.

A needle hidden in a huge grid world.



Can a navigation robot reach it?

$\Downarrow$

A 2-D version of a-needle-in-a-haystack.

What is a-needle-in-a-haystack

Assume we have a needle detector $F(x)$

"Find x (among N items) such that $F(x) = 1$"

when

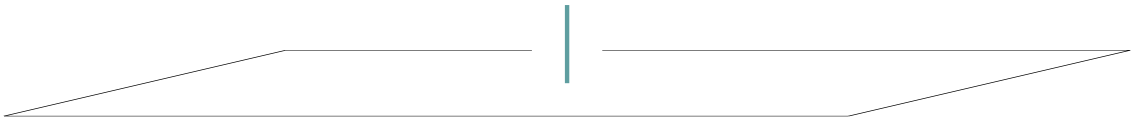
only x fulfills $F(x) = 1$ while $F(x) = 0$ for all others.

(We might call $F(x)$ a fitness function of this problem.)

$O(N)$ steps are necessary!

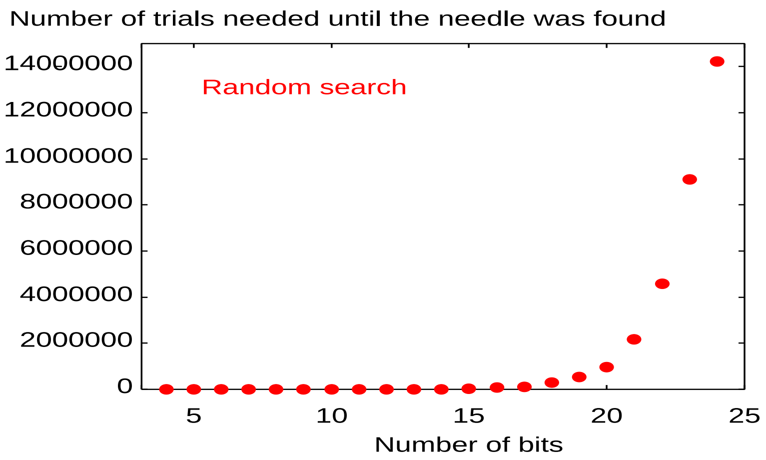
Can we look for a needle more efficiently?

Hinton & Nowlan's Needle (1987)



- A-needle ⇒ Just one configuration of 20-bit binary strings.
- Haystack ⇒ $2^{20} - 1$ search points.

Their choice of 20-bit was a good one!



Hence, their experiment was not a very difficult one.

How we evolve all-fitness-0 chromosomes?

(1011 1001011001001001)

(1000 1011011111011111)

$\xrightarrow{\text{crossover}}$ (1011 1011011111011111)

$\Big\downarrow$ mutation

(1011 1011011111 1 11111)

- What about in the case of "All-one-problem?"

$$(9 \times 14 \Rightarrow 16 \Rightarrow 17 \text{ vs. } 0 \times 0 \Rightarrow 0 \Rightarrow 0)$$

Hinton & Nowlan's brilliant trick

Evolution under Baldwin effect

— Lifetime learning of phenotype

A genotype:

(10901099011001001091)

Its phenotype:

(10101001011001001001)
(10001011011001001011)

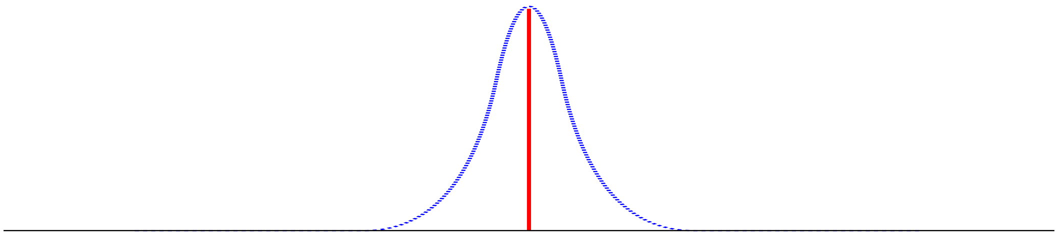
...

Their assumption

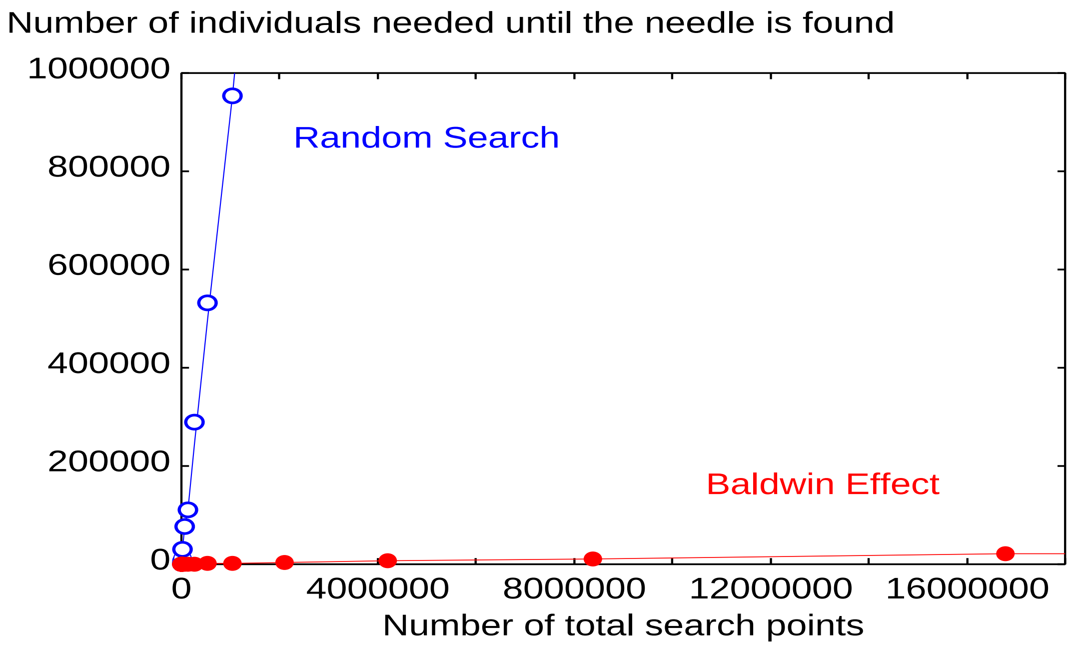
The closer the *genotype* to the needle,
the faster the learning of *phenotype.*

$\Downarrow$

This makes the sharp-fitness-peak a more smooth one.



Baldwin effect looks great!



Are we really happy with this?

Why should we continue evolution
when lifetime-learning already has found the needle?

$\Downarrow$

(Turney 1987)

"Not from an engineering but a biological interest."

Their GA implementation

1024 genotypes with 25% flexible genes,
each allowed 1000 lifetime learnings;

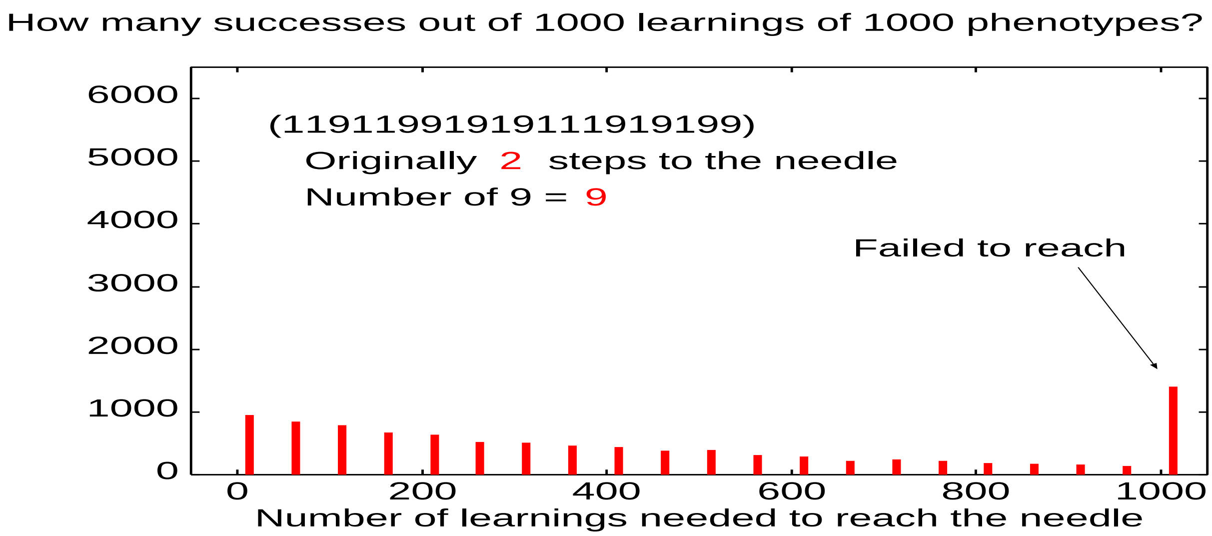
under

Roulette wheel selection;
One-point cross over;
Mutation;

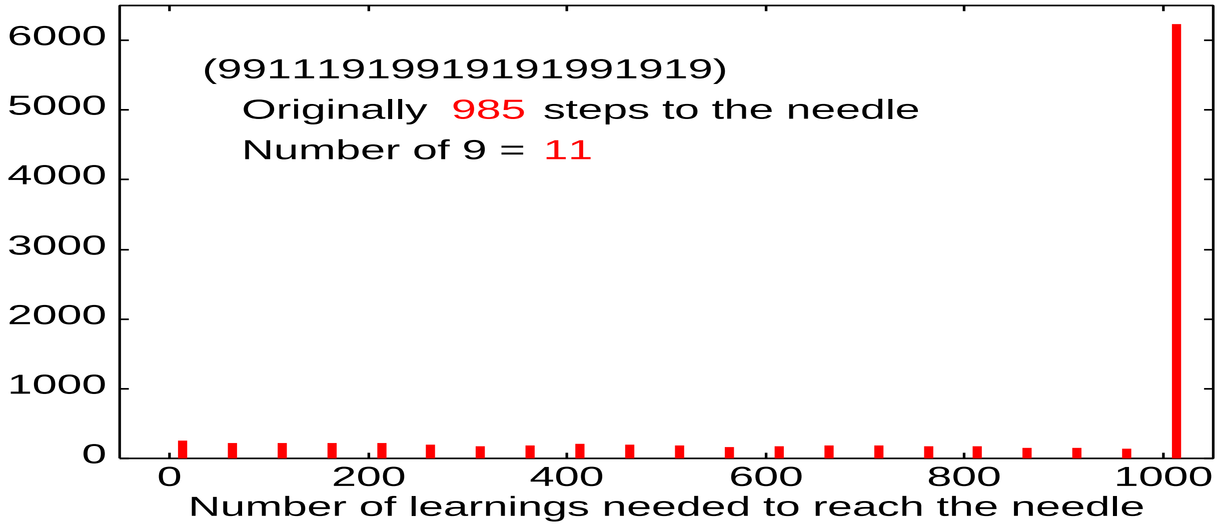
......

Is a high fitness gene a good gene?

Assume the needle is (11111111111111111111) without a loss of generality



Goodness depends on number of 9?



How many 9's are optimal?

or

Is the number of 9's decreasing?

Which will be a better one?

Assuming (11111111111111111111) is the needle

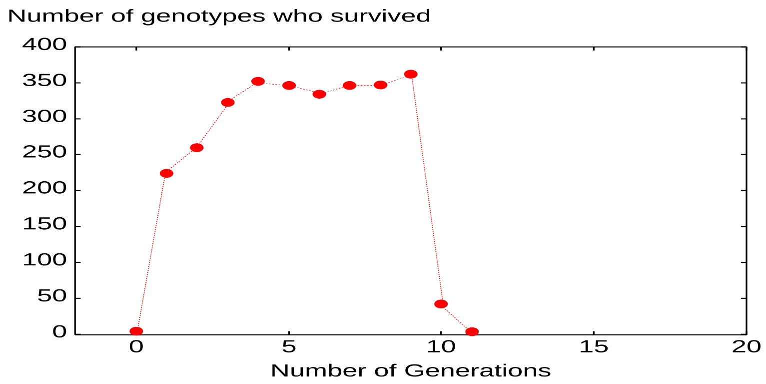
(9999999991999999999 9)

or

(1111111119111111111 1)

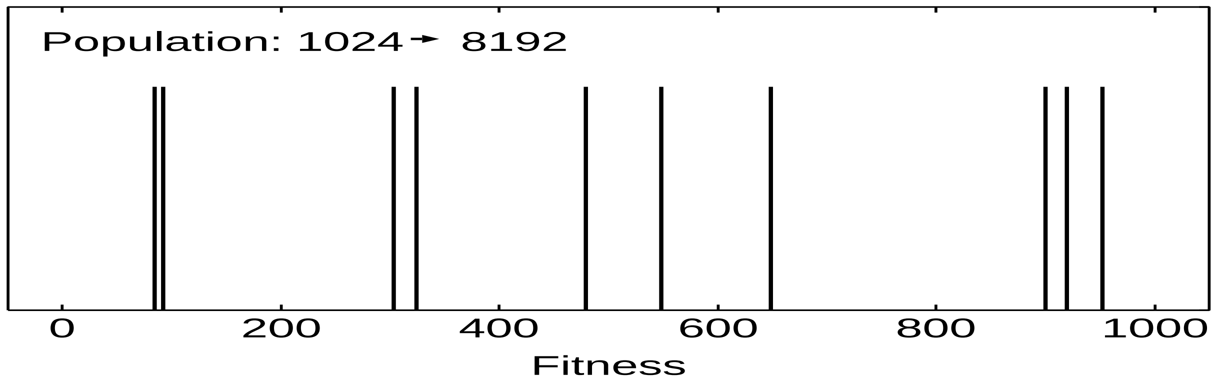
An extinction!

Only 4 successful genotypes are found in the 1st generation, at the luckiest case.



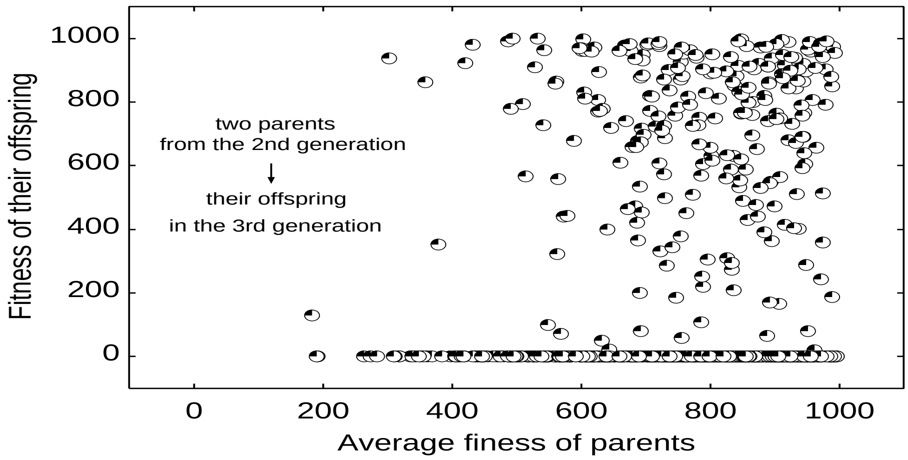
Genotype survived at the start

Let's increase the population.

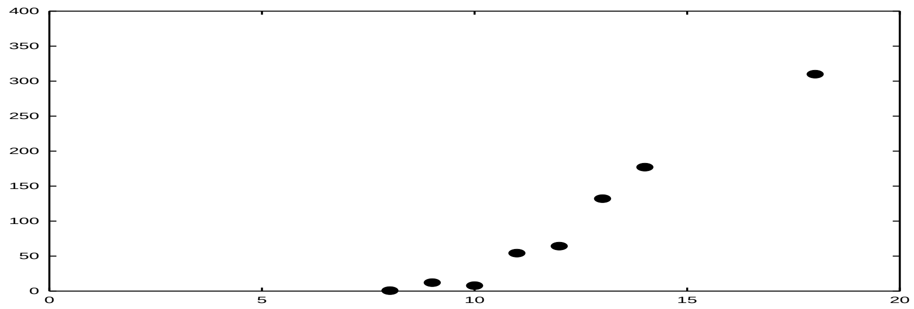


Why we failed?

No fitness correlation between parents & offspring



The more 9's the higher the fitness?



A dillemma

Assuming (1111111111111111111) is the needle,

e.g. (1191991099919911919 9) has no chance to reach the needle,

while e.g. (9999999199999999999) has 1000 chances.

$\Downarrow$

What if e.g. (9999999199999999999) dominates the population?

What about Lamarckian Inheritance?

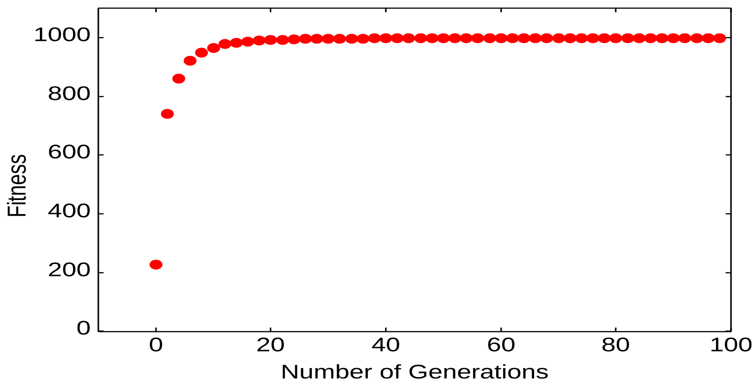
An inverse mapping from phenotype to genotype is necessary.

Turney (1996)

"We believe that computing this mapping is intractable,"

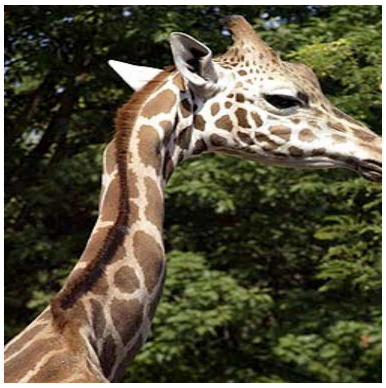
With Larmackian Inheritance

Let's re-map a few of successful 9's in phenotype to its genotype.



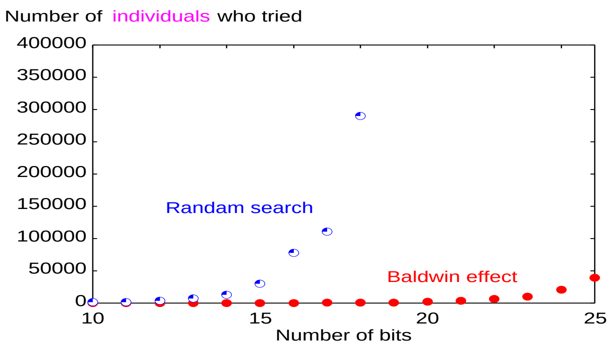
Eventually, all 9's disappeared to converge to the needle itself.

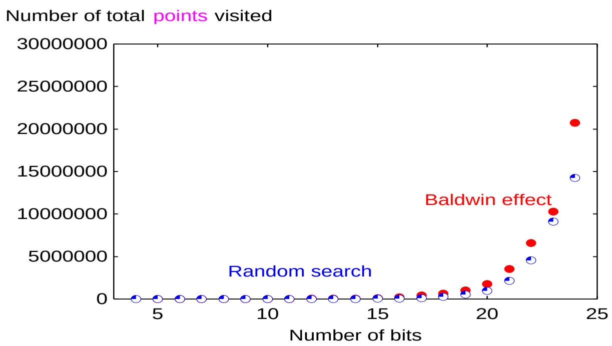
Why giraffe has a long neck?



Still we are not so happy.

The number of total points searched (not the number of individuals,) is almost similar as a random search.





“Alas! Where is the needle longer than 25-bit?”

Computational analogue of a-needle-in-a-haystack

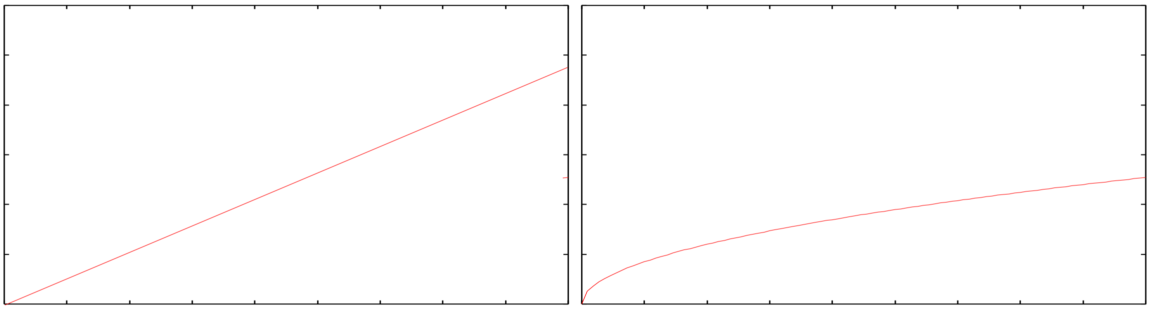
A specific rare cases in a huge database;
A real necessary information from world-wide-web;
Oil spills in the ocean from satellite image;
A cause of failures in diagnosing a huge code;
A collision of a hash function;
etc......

$\Downarrow$

Such searches are very important for us.

Grover's quantum search (1997)

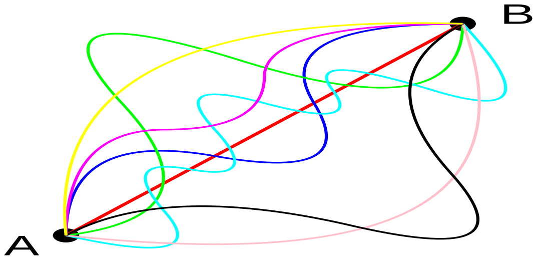
A speed up from $O(N)$ to $O(\sqrt{N})$.



A strange path of quantum computation

When a particle goes from A to B,

it takes all possible paths at the same time.



Concluding Remarks

- Searching for a needle in a hay is open & important issue.
- We need real efficient approaches.
- Not yet so far a real efficient one.
- We have to avoid an effect of

  like-to-hear-what-we-would-like-to-hear.

  Success at a small scale is not a royal road to a real success.