“It were much to be desired, that when mathematical processes pass through the human brain instead of through the medium of inanimate mechanism, it were equally a necessity of things that the reasonings connected with operations should hold the same just place as a clear and well-defined branch of the subject of analysis, a fundamental but yet independent ingredient in the science, which they must do in studying the engine.”

Ada Lovelace

“It were much to be desired, that when mathematical processes pass through the human brain instead of through the medium of inanimate mechanism, it were equally a necessity of things that the reasonings connected with operations should hold the same just place as a clear and well-defined branch of the subject of analysis, a fundamental but yet independent ingredient in the science, which they must do in studying the engine.”

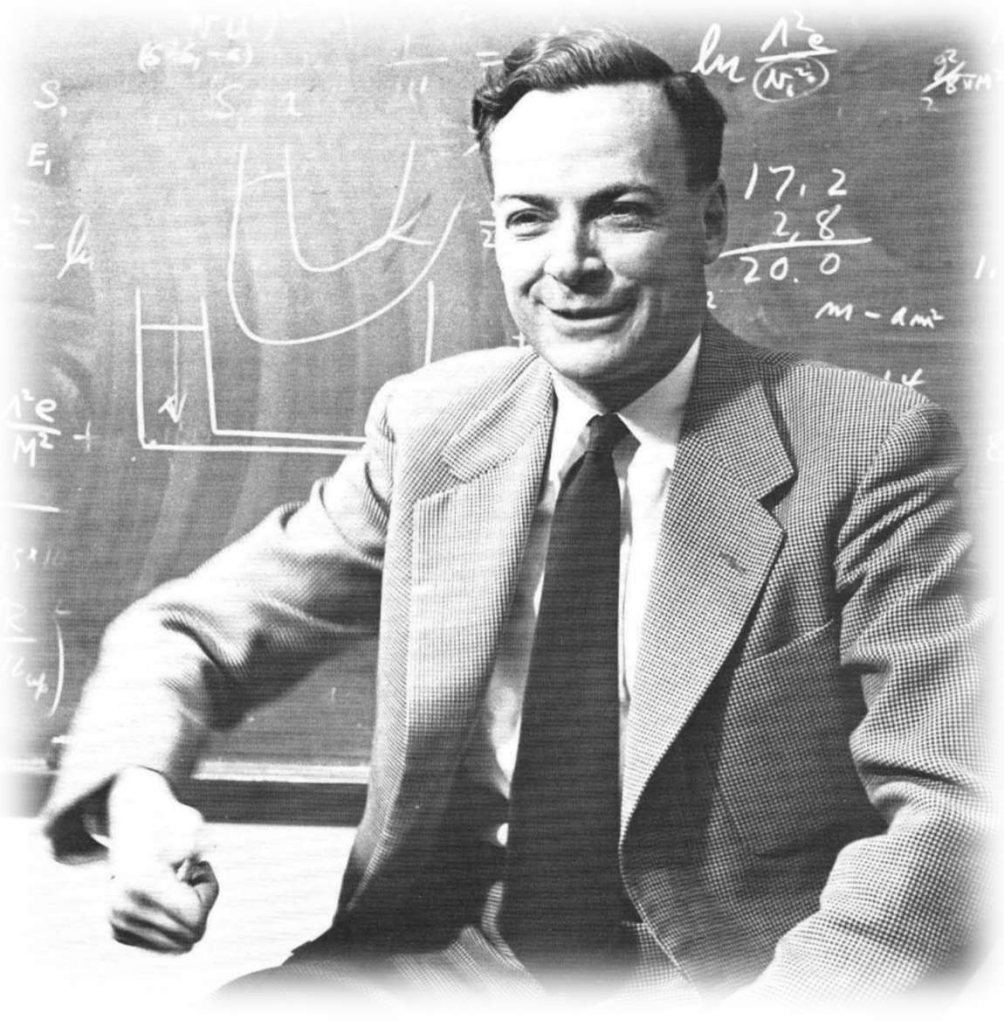
Ada Lovelace

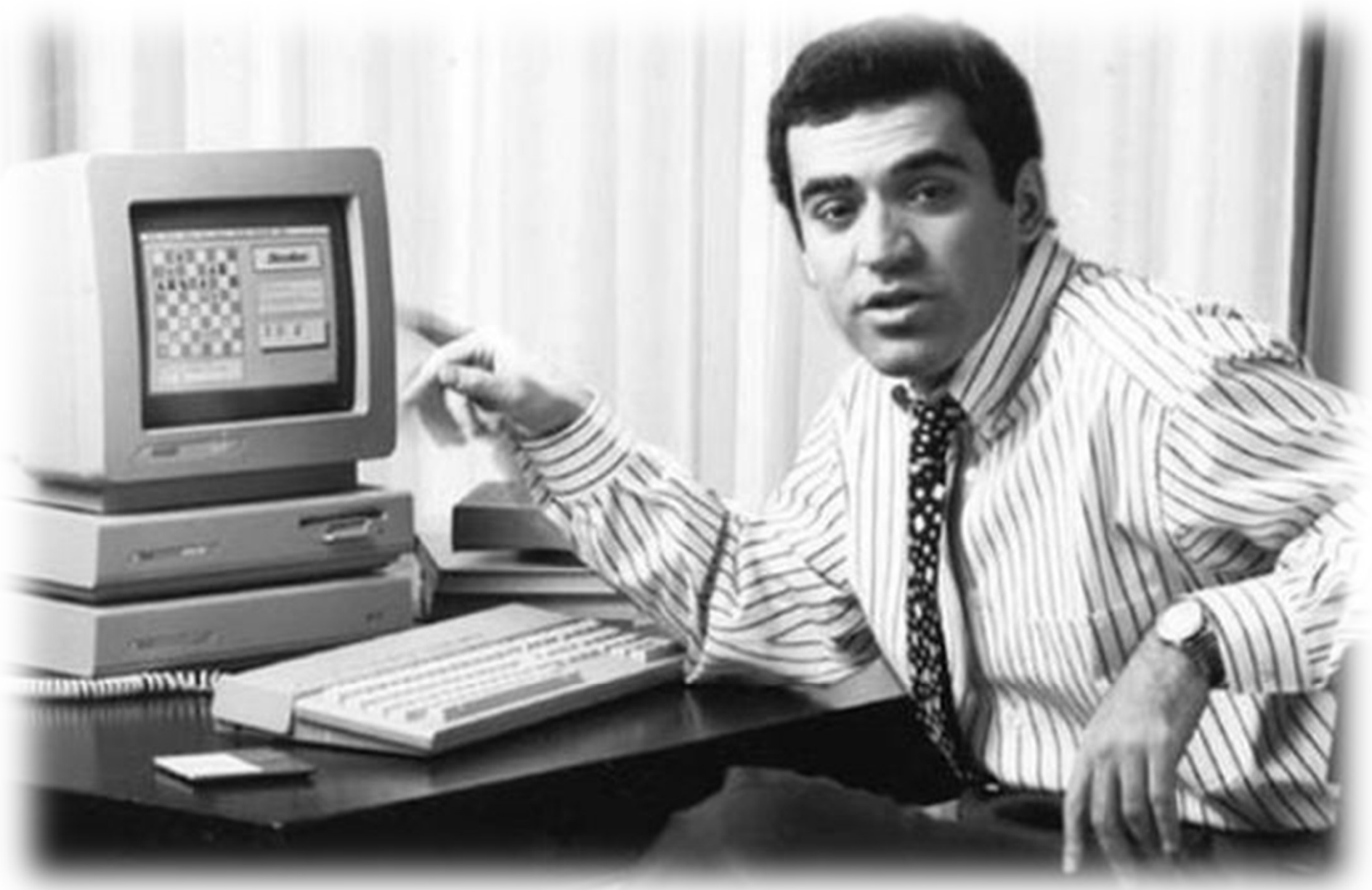
“The brain is fundamentally a lazy piece of meat.”

Gregory Berns

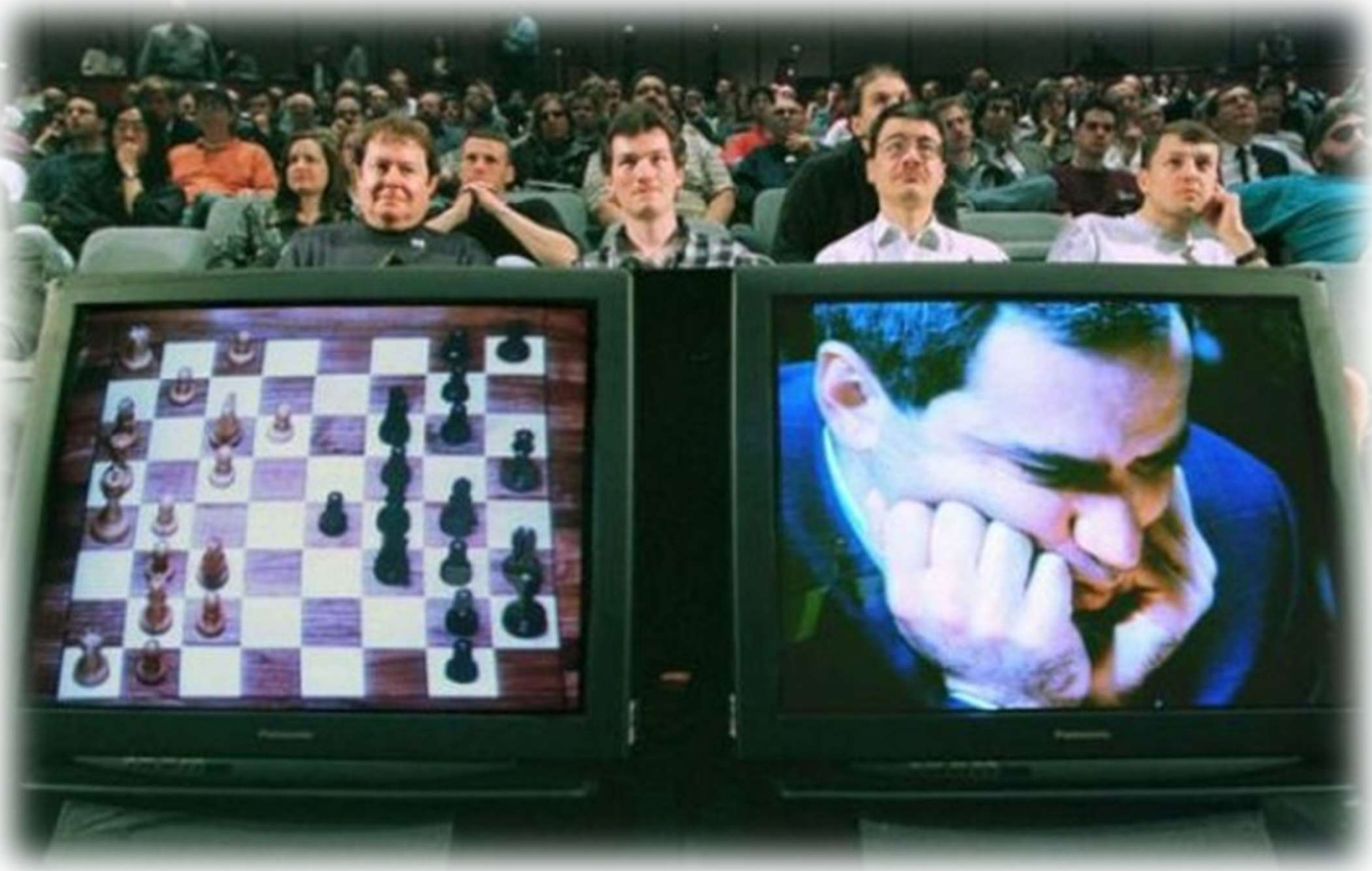
Iconoclast

"… a fun analogy in trying to get some idea of what we're doing in trying to understand nature, is to imagine that the gods are playing some great game like chess… and you don't know the rules of the game, but you're allowed to look at the board, at least from time to time… and from these observations you try to figure out what the rules of the game are."

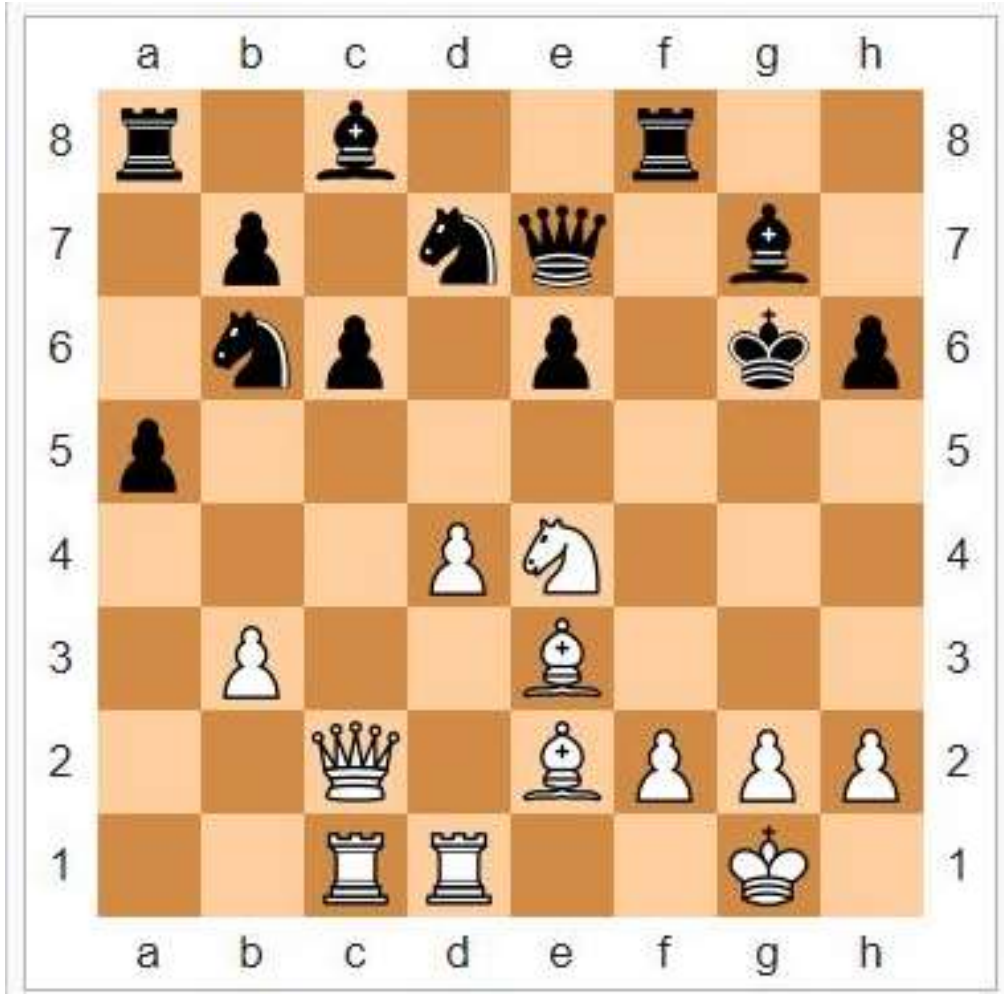




1992

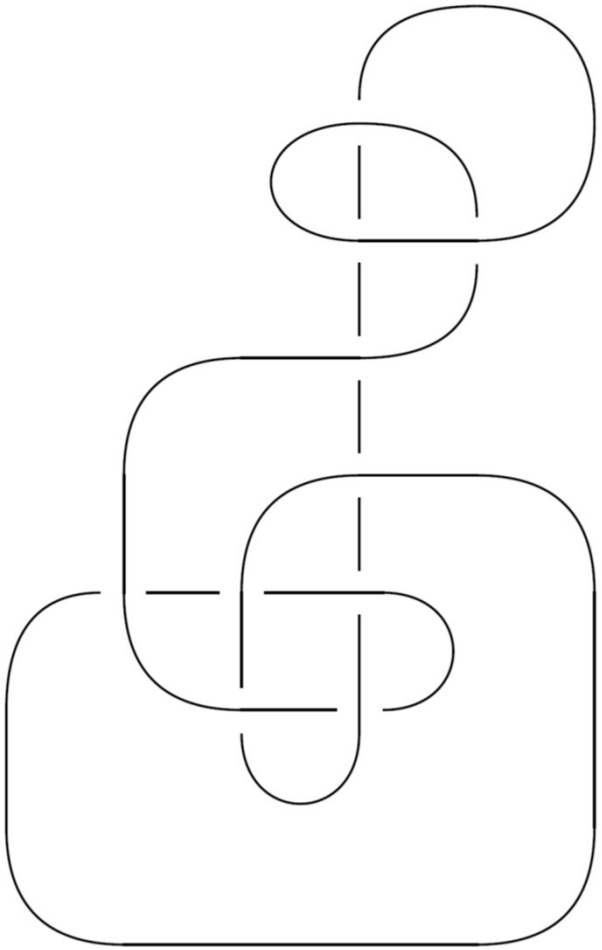


1996

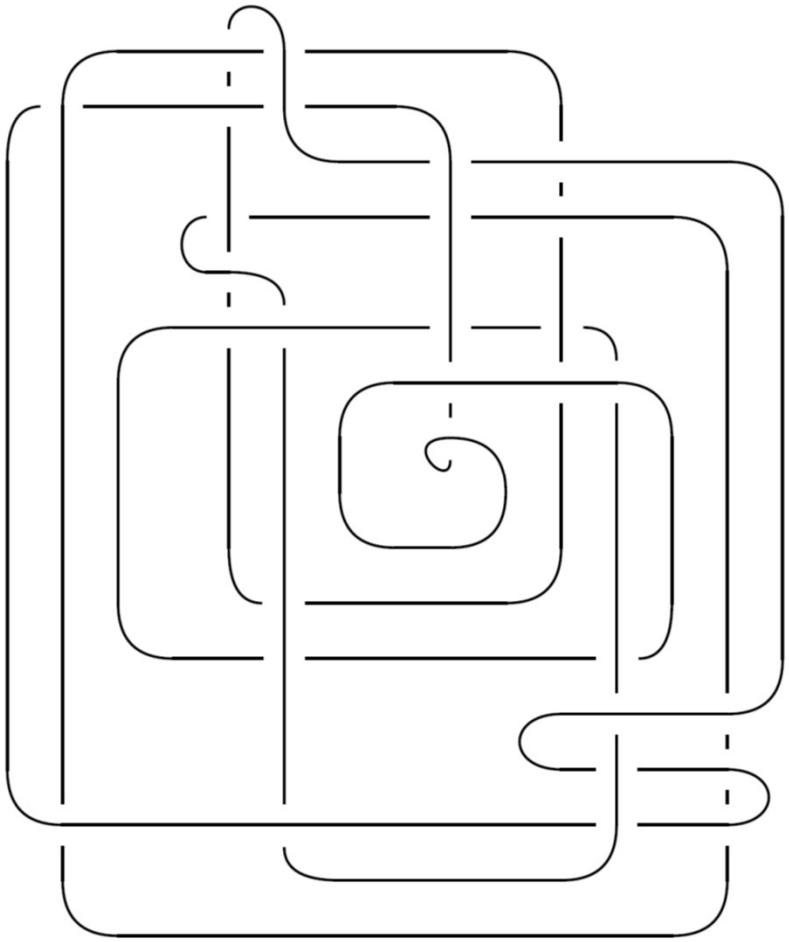

a b c d e f g h
8 7 6 5 4 3 2 1

?

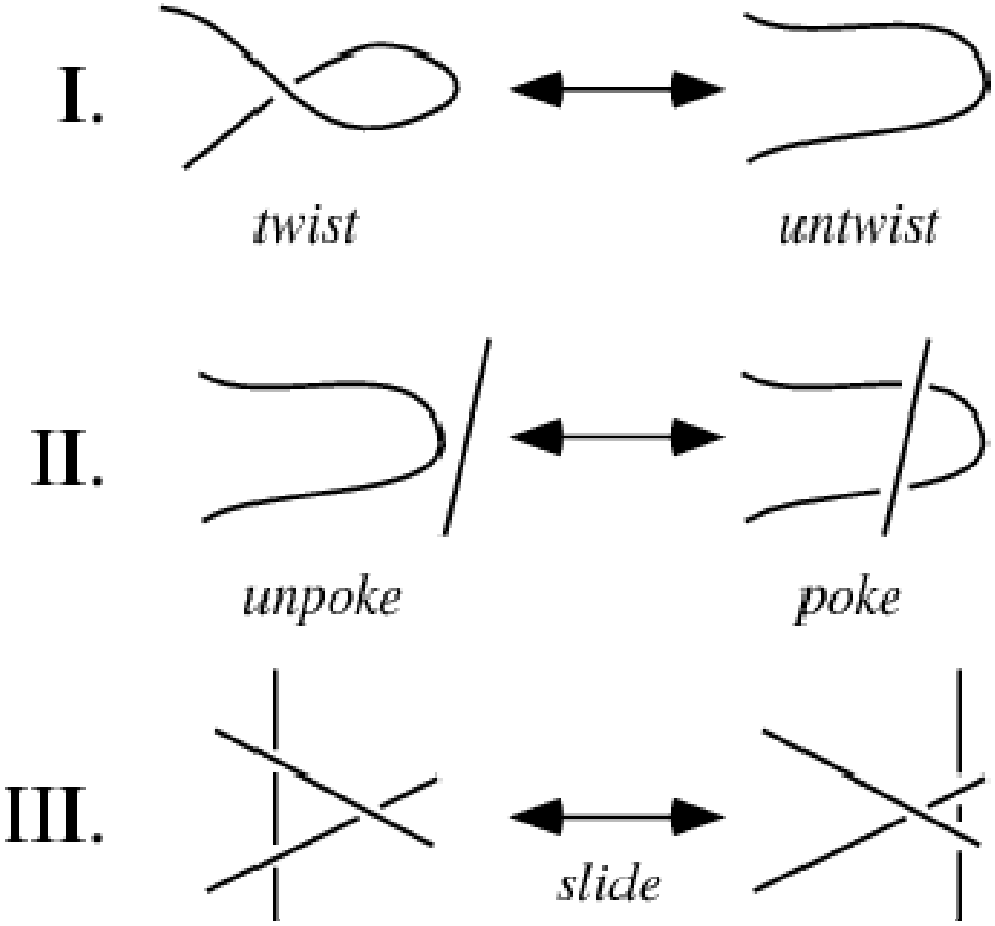
=



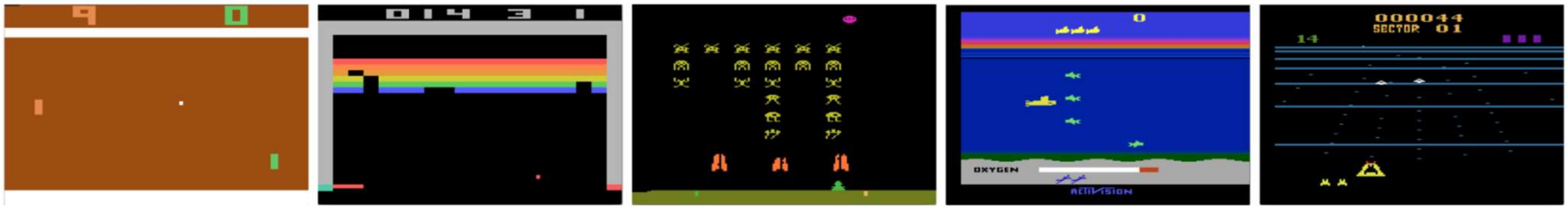
10 crossings



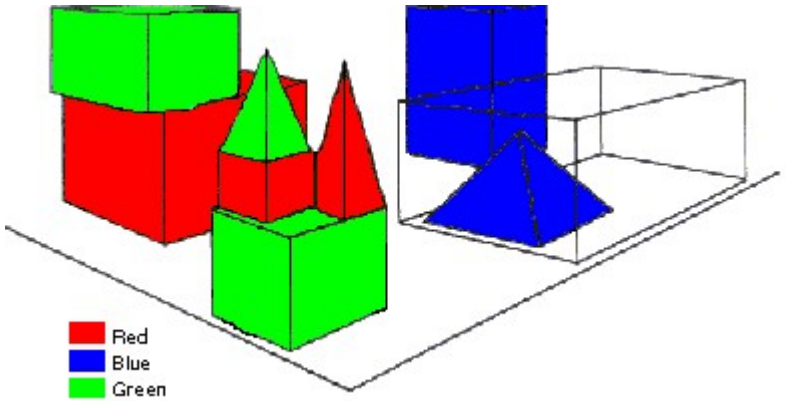
30 crossings



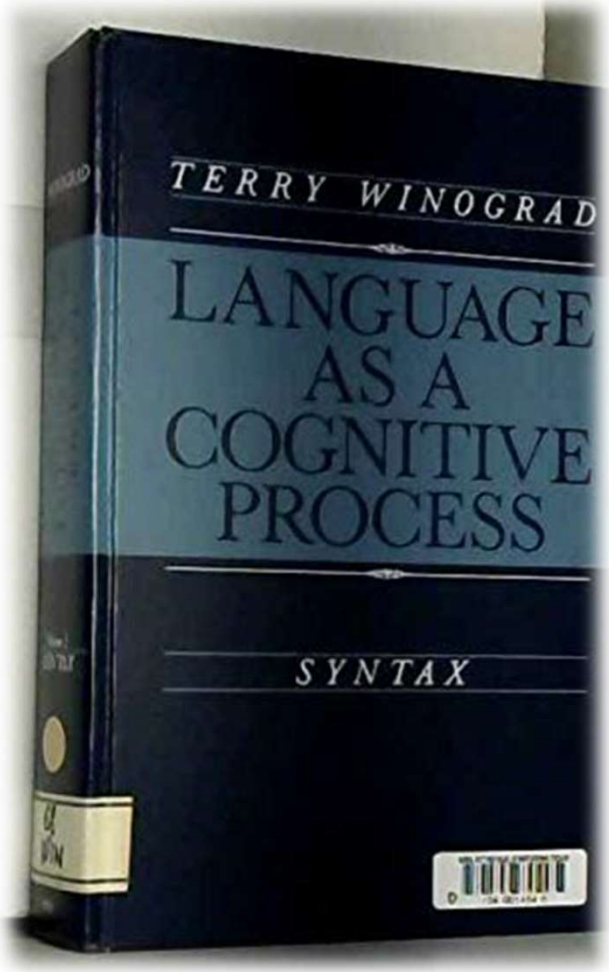
Kurt Reidemeister



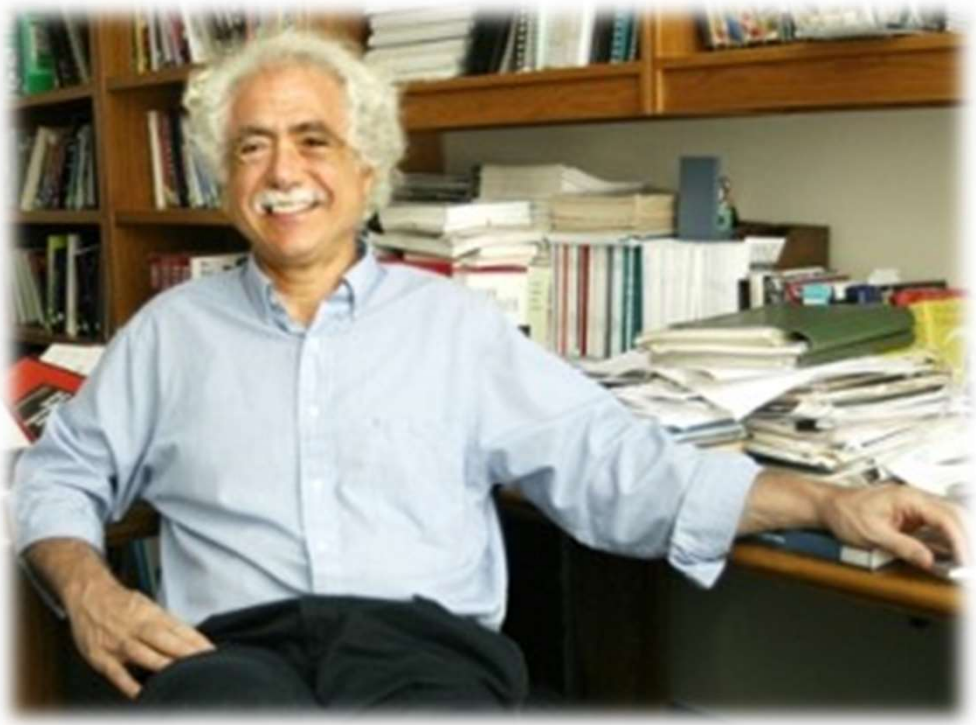
Person: Pick up a big red block.
Computer: OK.
Person: Grasp the pyramid.
Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
Computer: OK.
Person: What does the box contain?
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
Person: What is the pyramid supported by?
Computer: THE BOX.
Person: How many blocks are not in the box?
Computer: FOUR OF THEM.
Person: Is at least one of them narrower than the one which i told you to pick up?
Computer: YES, THE RED CUBE.
Person: Is it supported?
Computer: YES, BY THE TABLE.
Person: Can the table pick up blocks?
Computer: NO.
Person: Can a pyramid be supported by a block?
Computer: YES.
Person: Can a pyramid support a pyramid?
Computer: I DON'T KNOW.
Person: Stack up two pyramids.
Computer: I CAN'T.

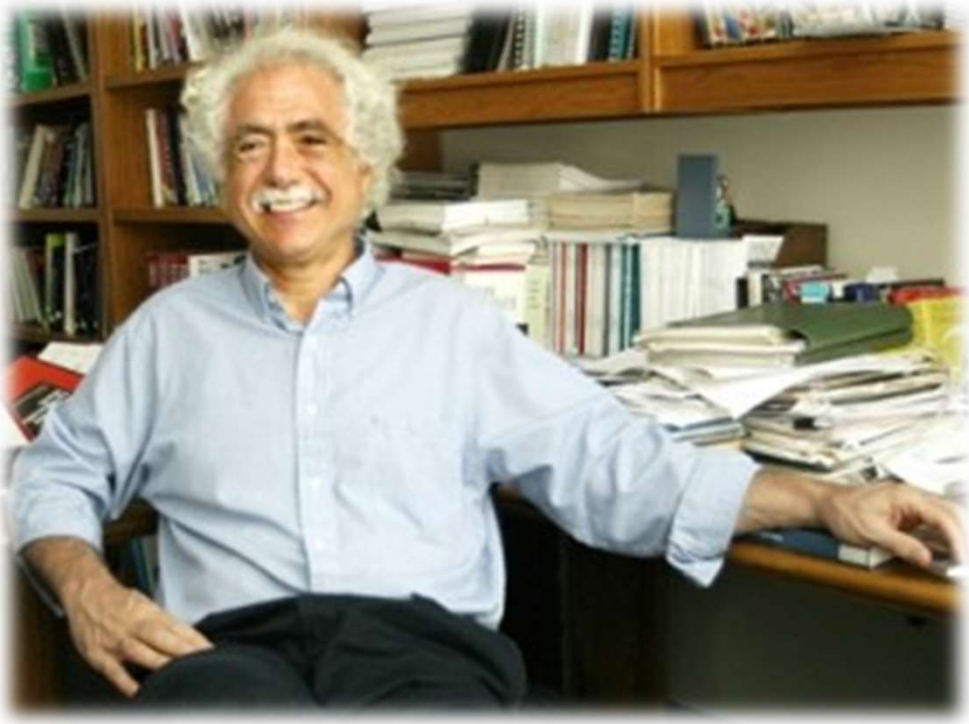


Person: Will you please stack up both of the red blocks and either a green cube or a pyramid?
Computer: OK.
Person: Which cube is sitting on the table?
Computer: THE LARGE GREEN ONE WHICH SUPPORTS THE RED PYRAMID.
:



SHRDLU (1971)





1997


Google

Winograd schemas:

The trophy would not fit in the brown suitcase because it was too big (*small*).

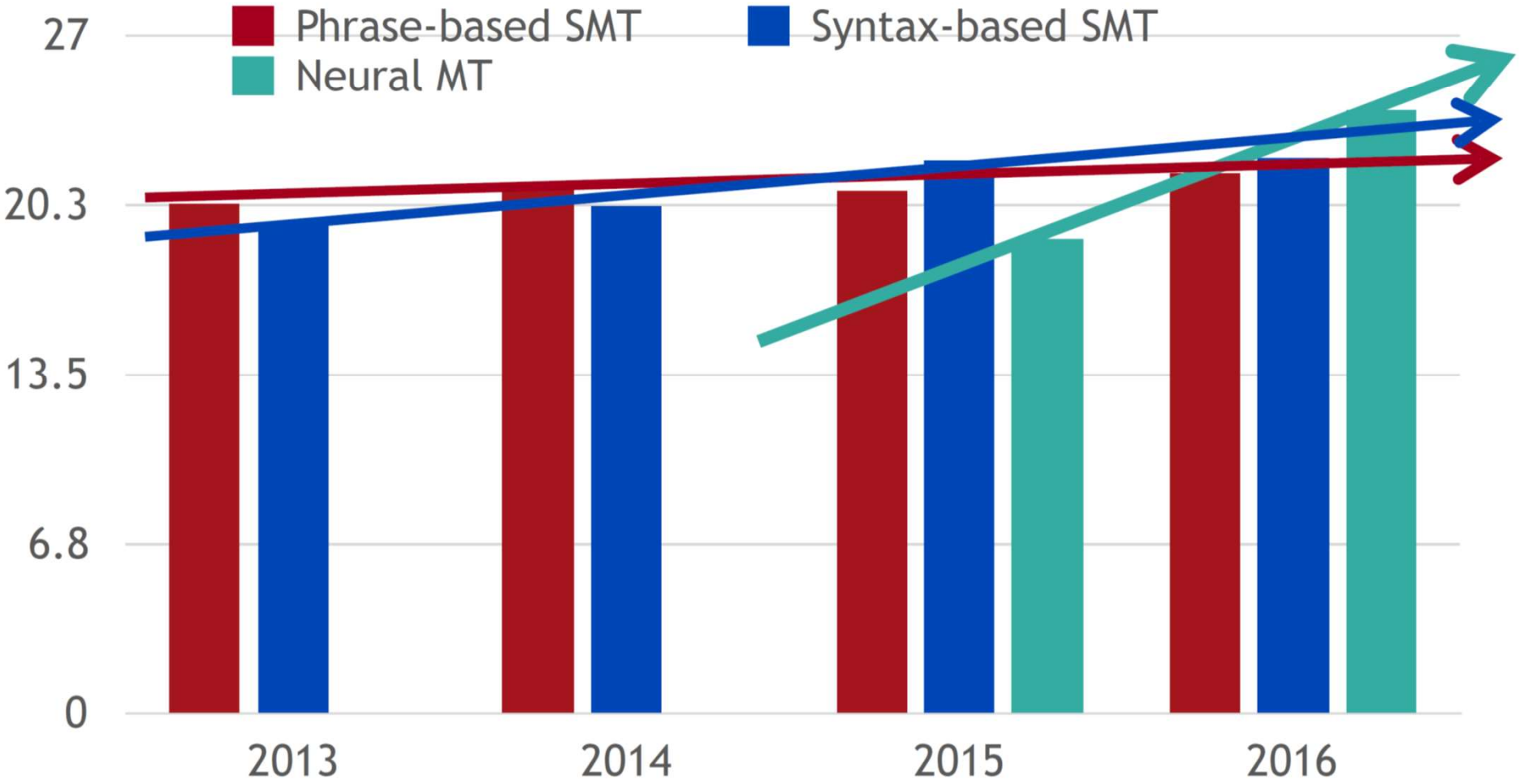
What was too big (*small*)?

The town councilors refused to give the demonstrators a permit because they feared (*advocated*) violence.

Who feared (*advocated*) violence?

MT progress over time

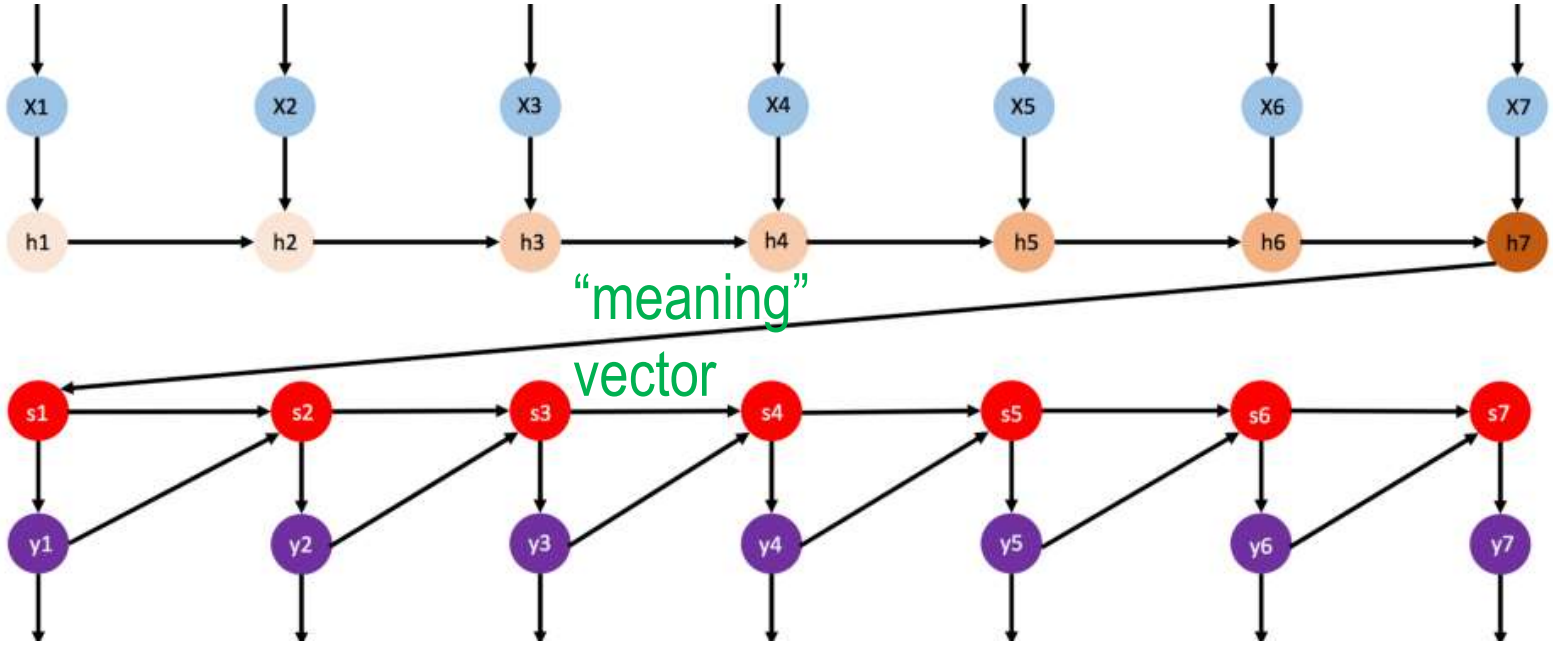
[Edinburgh En-De WMT newstest2013 Cased BLEU; NMT 2015 from U. Montréal]



Source: http://www.meta-net.eu/events/meta-forum-2016/slides/09_sennrich.pdf

seq2seq: Encoder + Decoder

you asked us to call you back after last Friday



Wir hoffen jedoch, dass sie bei Ihrer Reiseplanung weiterhilft

"Translation model"

"Language model"

Need "context" vectors

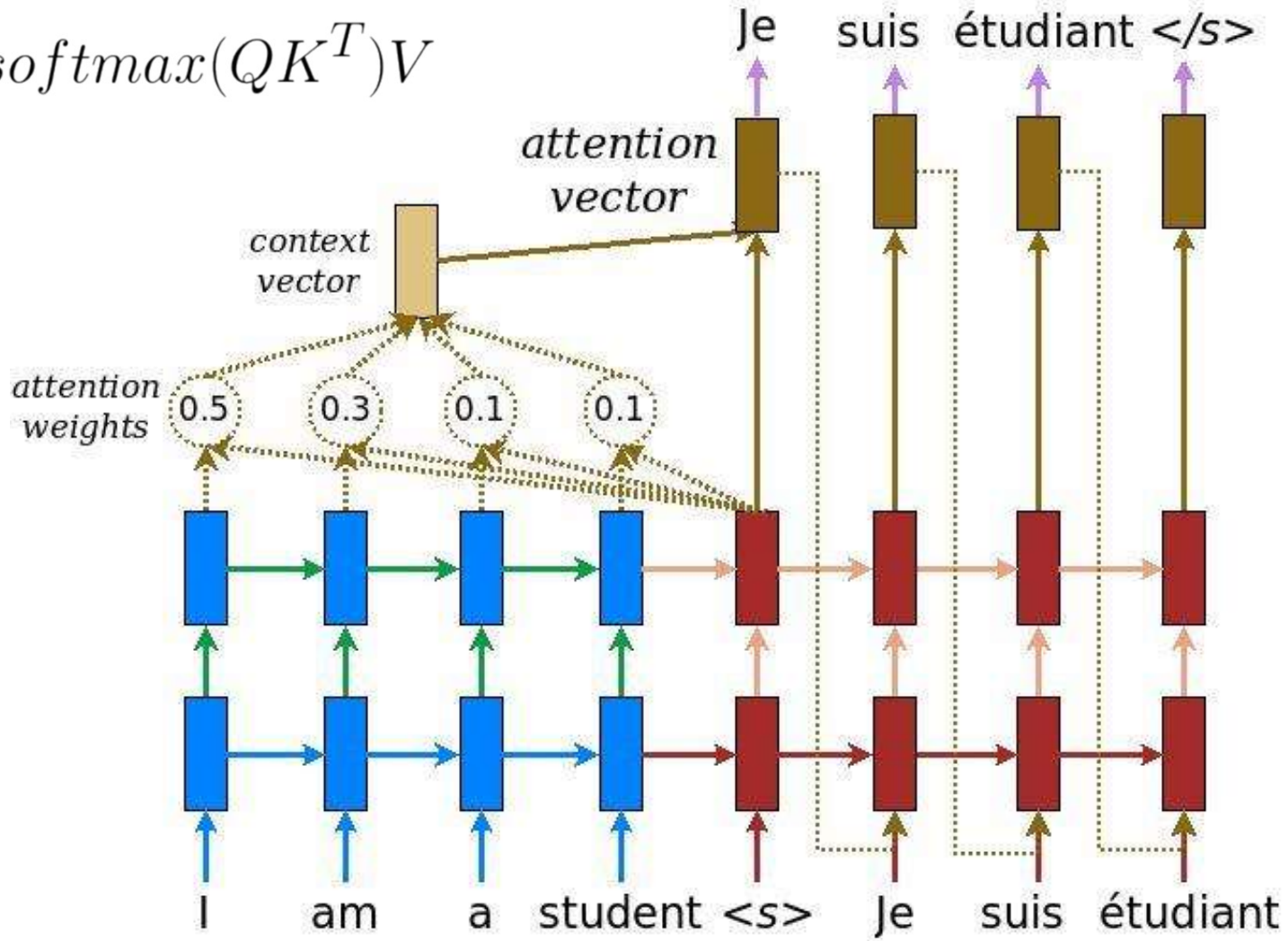

汀步路，
小心滑倒！
Slip and fall down carefully!


AIRPORTS AUTHORITY OF INDIA
फर्श पर खाना सख़्त मना है ।
—आदेश से
EATING CARPET
STRICTLY PROHIBITED
—By Order

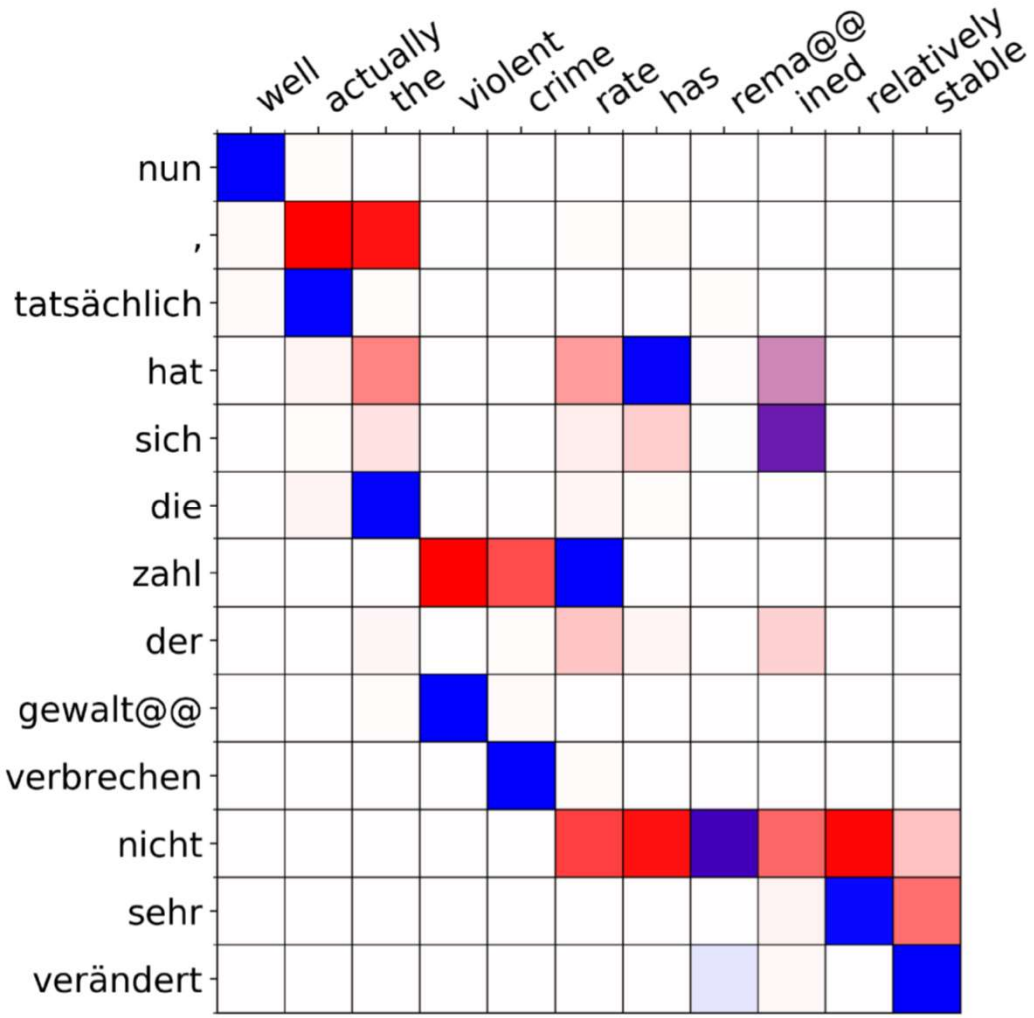

请勿食用
Please don't be edible
外带食物
Exterior Girdle food


Bengali
Russian
Portuguese
Arabic
Hindi
Spanish
Mandarin
German
Italian
French
Translate

$$softmax(QK^T)V$$



Source: https://github.com/tensorflow/nmt



variational attention (blue)
vs prior alignment (red)

self-attention

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

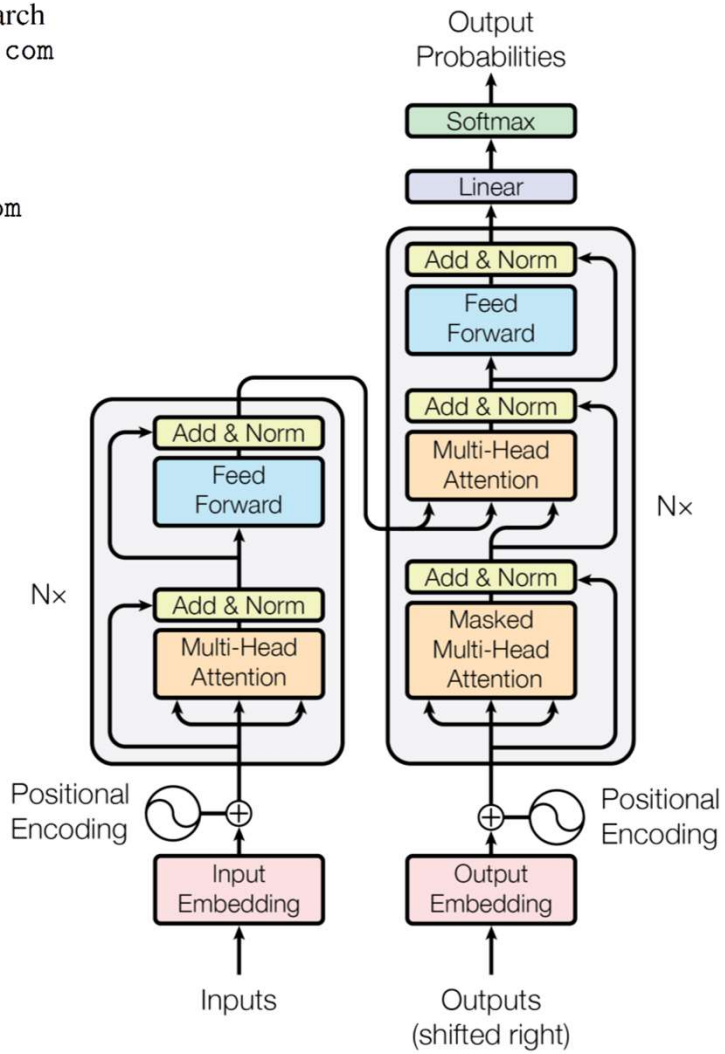
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.



Sentence	Google Translate	Transformer
The cow ate the hay because it was **delicious**.	La vache mangeait le foin parce **qu'elle** était délicieuse.	La vache a mangé le foin parce **qu'il** était délicieux.
The cow ate the hay because it was **hungry**.	La vache mangeait le foin parce **qu'elle** avait faim.	La vache mangeait le foin parce **qu'elle** avait faim.
The women stopped drinking the wines because they were **carcinogenic**.	Les femmes ont cessé de boire les vins parce **qu'ils** étaient cancérogènes.	Les femmes ont cessé de boire les vins parce **qu'ils** étaient cancérigènes.
The women stopped drinking the wines because they were **pregnant**.	Les femmes ont cessé de boire les vins parce **qu'ils** étaient enceintes.	Les femmes ont cessé de boire les vins parce **qu'elles** étaient enceintes.
The city councilmen refused the female demonstrators a permit because they **advocated** violence.	Les conseillers municipaux ont refusé aux femmes manifestantes un permis parce **qu'ils** préconisaient la violence.	Le conseil municipal a refusé aux manifestantes un permis parce **qu'elles** prônaient la violence.
The city councilmen refused the female demonstrators a permit because they **feared** violence.	Les conseillers municipaux ont refusé aux femmes manifestantes un permis parce **qu'ils** craignaient la violence	Le conseil municipal a refusé aux manifestantes un permis parce **qu'elles** craignaient la violence.*

Lukasz Kaiser, 2017

“The Transformer” are a Japanese [[hardcore punk]] band.
==Early years==
The band was formed in 1968, during the height of Japanese music history. Among the legendary [[Japanese people|Japanese]] composers of [Japanese lyrics], they prominently exemplified Motohiro Oda's especially tasty lyrics and psychedelic intention. Michio was a longtime member of the every Sunday night band PSM. His alluring was of such importance as being the man who ignored the already successful image and that he municipal makeup whose parents were – the band was called Jenei.<ref>http://www.separatist.org/se_frontend/post-punk-musician-the-kidney.html</ref> From a young age the band was very close, thus opting to pioneer what ...

:

=== 1981-2010: The band to break away ===
On 1 January 1981 bassist Michio Kono, and the members of the original line-up emerged. Niji Fukune and his [[Head poet|Head]] band (now guitarist) Kazuya Kouda left the band in the hands of the band at the May 28, 1981, benefit season of [[Led Zeppelin]]'s Marmarin building. In June 1987, Kono joined the band as a full-time drummer, playing a ...

Lukasz Kaiser, 2017

REFORMER: THE EFFICIENT TRANSFORMER

Nikita Kitaev*
U.C. Berkeley & Google Research
kitaev@cs.berkeley.edu

Łukasz Kaiser*
Google Research

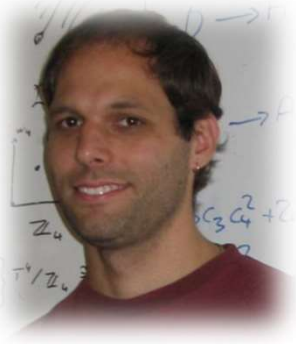
Anselm Levskaya
Google Research
{lukaszkaiser,levskaya}@google.com

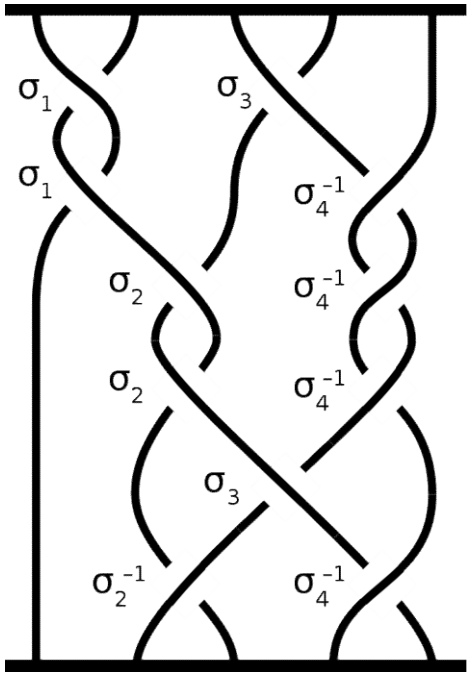
ABSTRACT

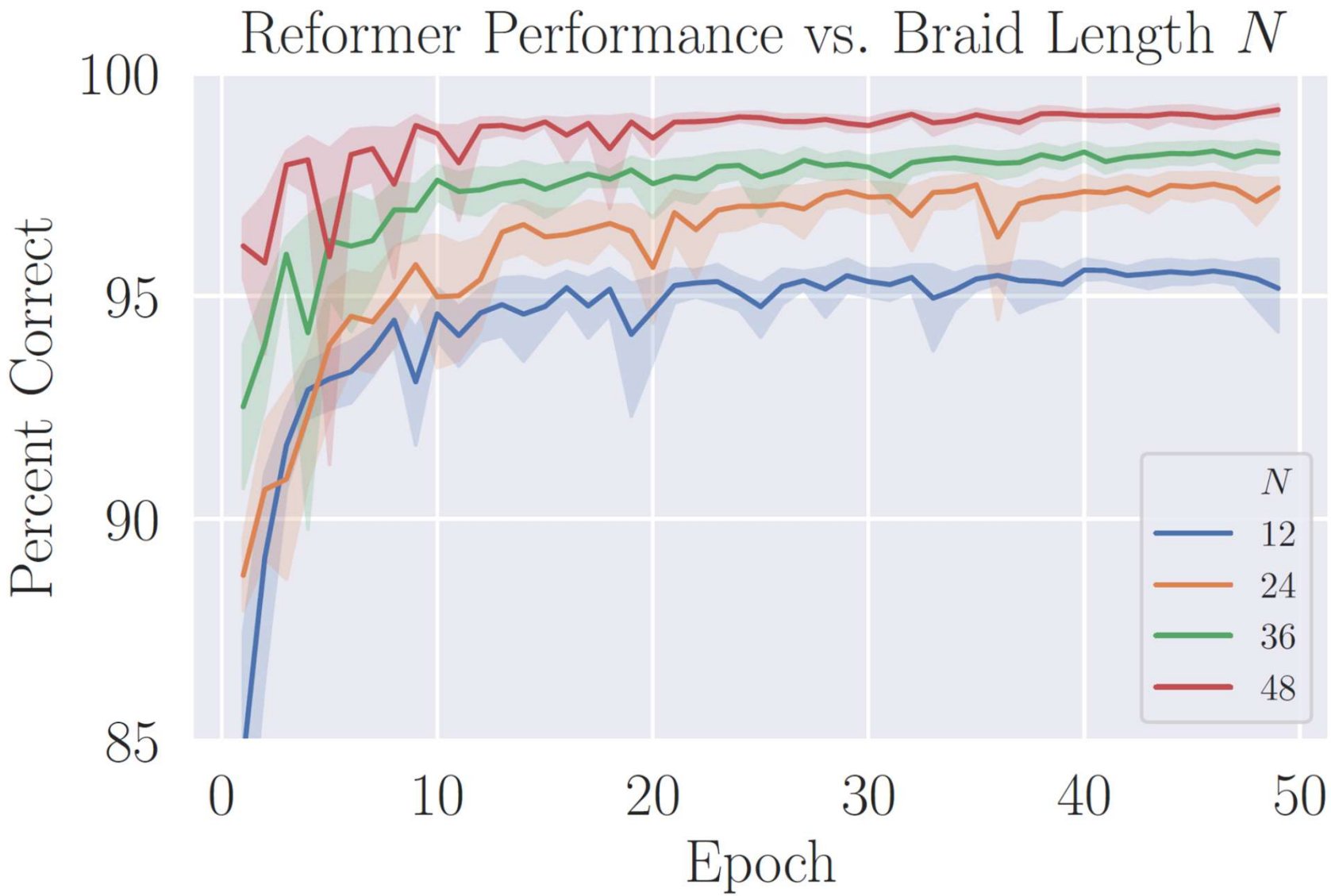
Large Transformer models routinely achieve state-of-the-art results on a number of tasks but training these models can be prohibitively costly, especially on long sequences. We introduce two techniques to improve the efficiency of Transformers. For one, we replace dot-product attention by one that uses locality-sensitive hashing, changing its complexity from O(L^2) to O($L \log L$), where L is the length of the sequence. Furthermore, we use reversible residual layers instead of the standard residuals, which allows storing activations only once in the training process instead of N times, where N is the number of layers. The resulting model, the Reformer, performs on par with Transformer models while being much more memory-efficient and much faster on long sequences.

Learning to Unknot

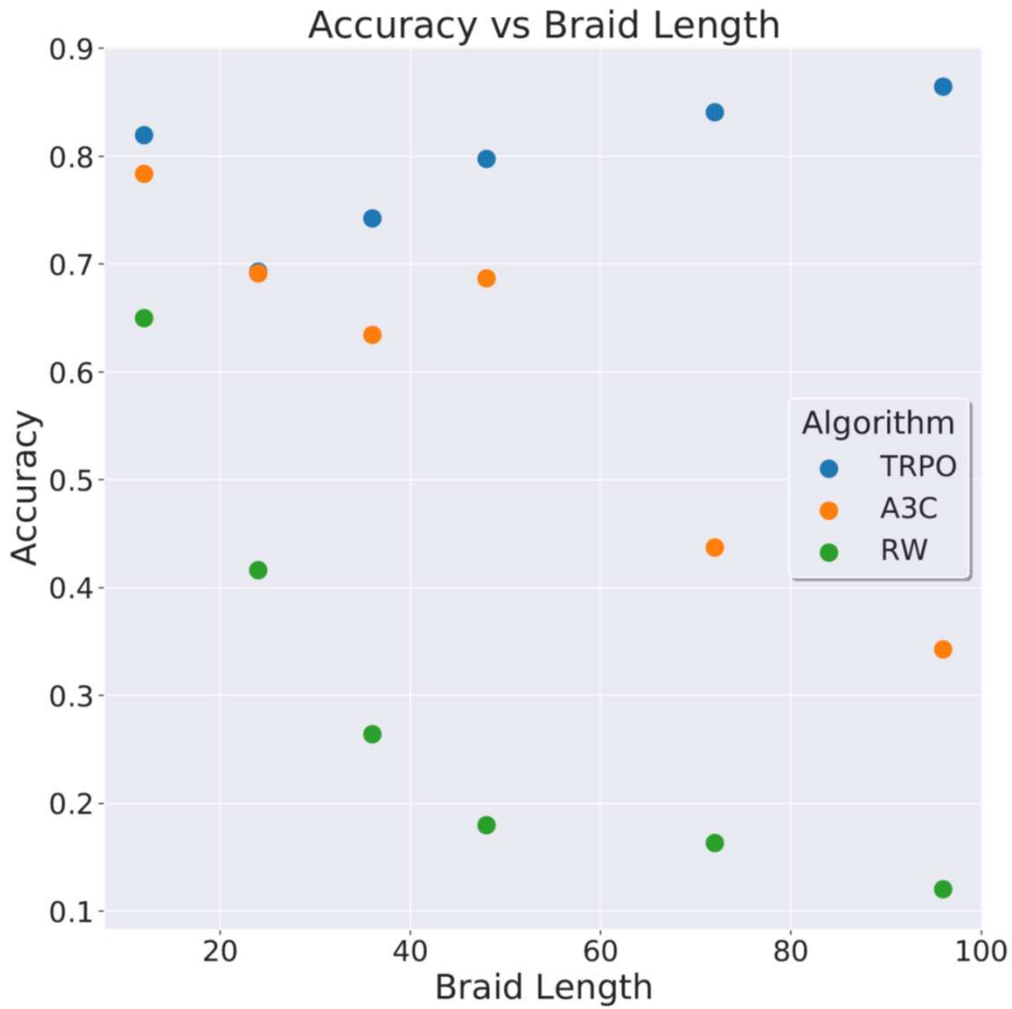
Sergei Gukov[1], James Halverson[2,3], Fabian Ruehle[4,5], Piotr Sułkowski[1,6]







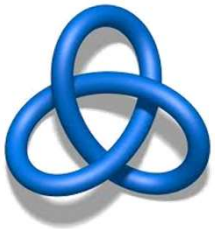
Reformer performance on UNKNOT as function of braid length. Performance increases with N.



Fraction of unknots whose braid words could be reduced to the empty braid word as a function of initial braid word length.

Generalized Poincare conjecture:

Every homotopy 4-sphere is diffeomorphic to the standard 4-sphere.



Theorem: If one finds a pair of knots which satisfy the following three properties:

- K and K' have the same 0-surgery
- K is not slice
- K' is slice

then the smooth 4-dimensional Poincare conjecture is false.

computation of
“quantum” invariants

1,388,705 knots

potential counterexamples
to SPC4 (ruled out)

165 knots

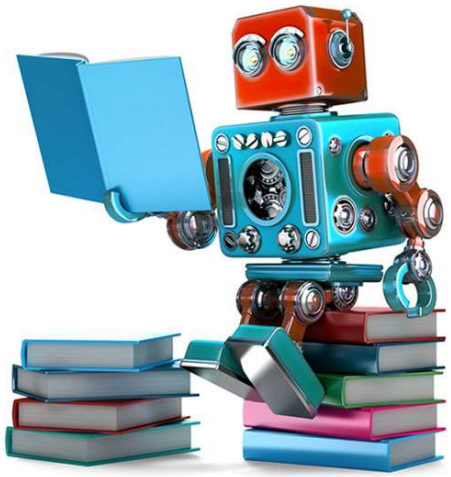
potential counterexample
to slice-ribbon conj.

#(crossings)

10 16 48 78 86

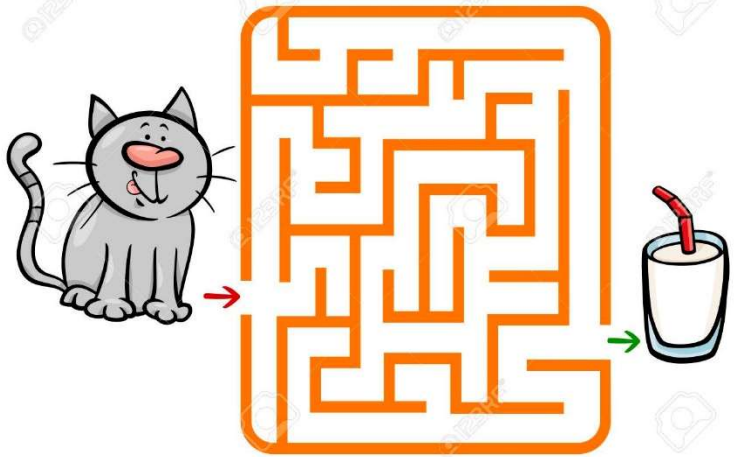
Using ML to discover new mathematics:

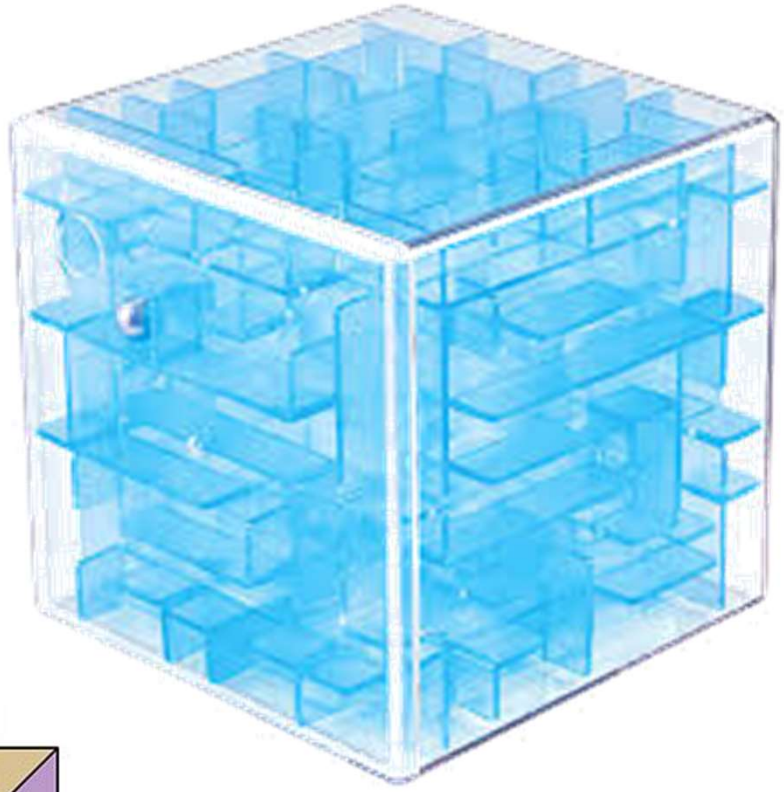
- Finding counterexamples (disproving conjectures)
- Formulating new conjectures (learning new patterns)

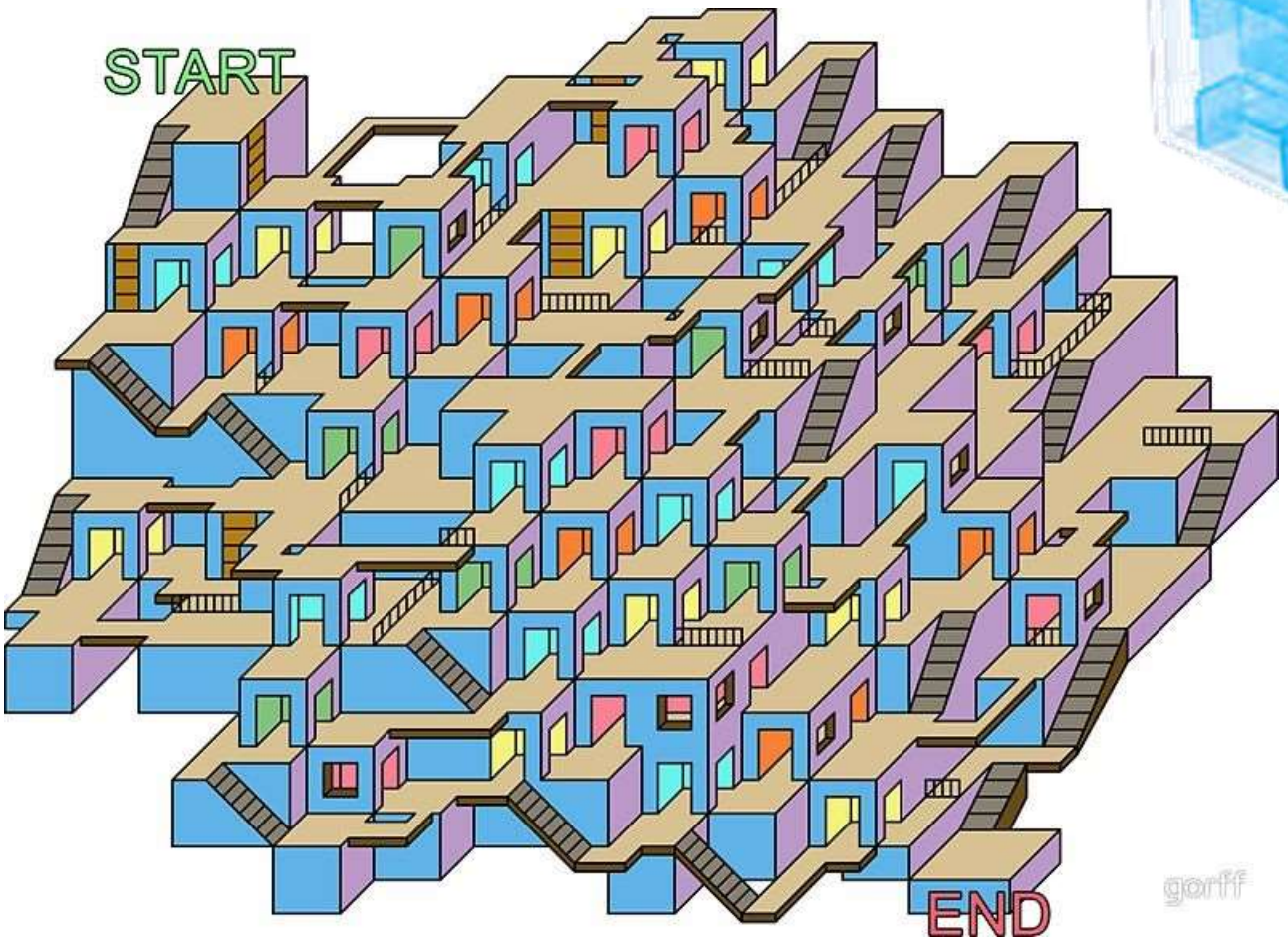


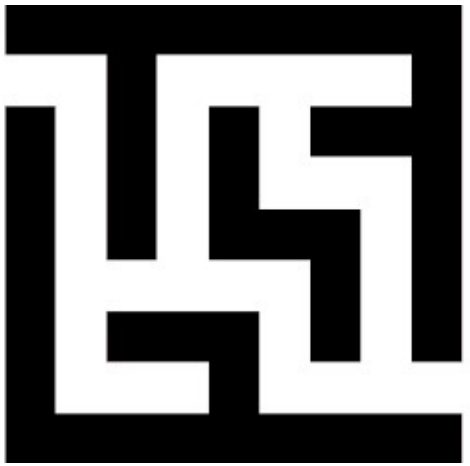
Mathematics of AI:
foundations, explainable AI,
geometric deep learning,
algorithm design, ...

Proof assistants:
autoformalization,
SAT-solvers, LEAN,
Minerva, GPT-4, ...

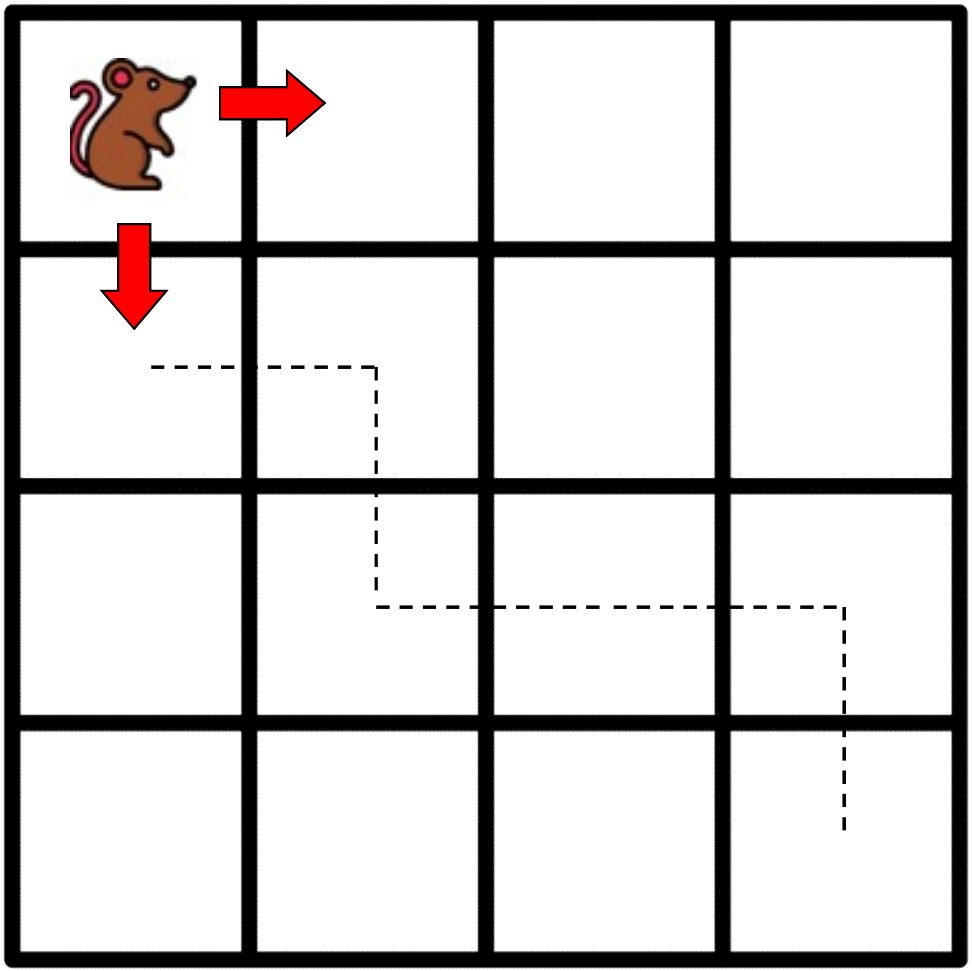




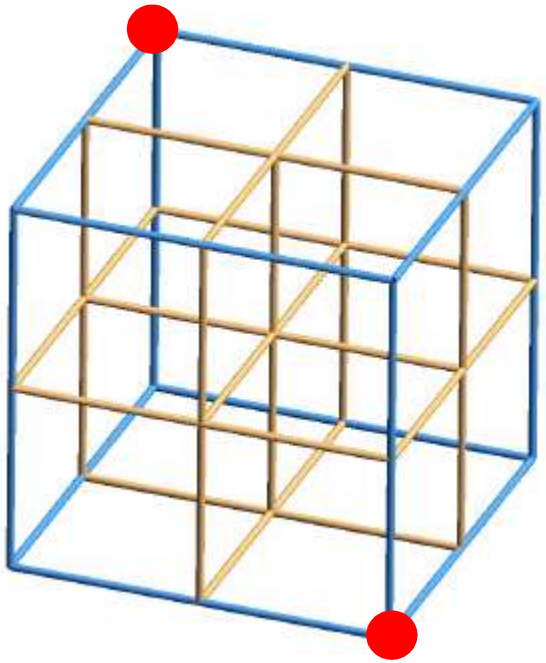

START
END



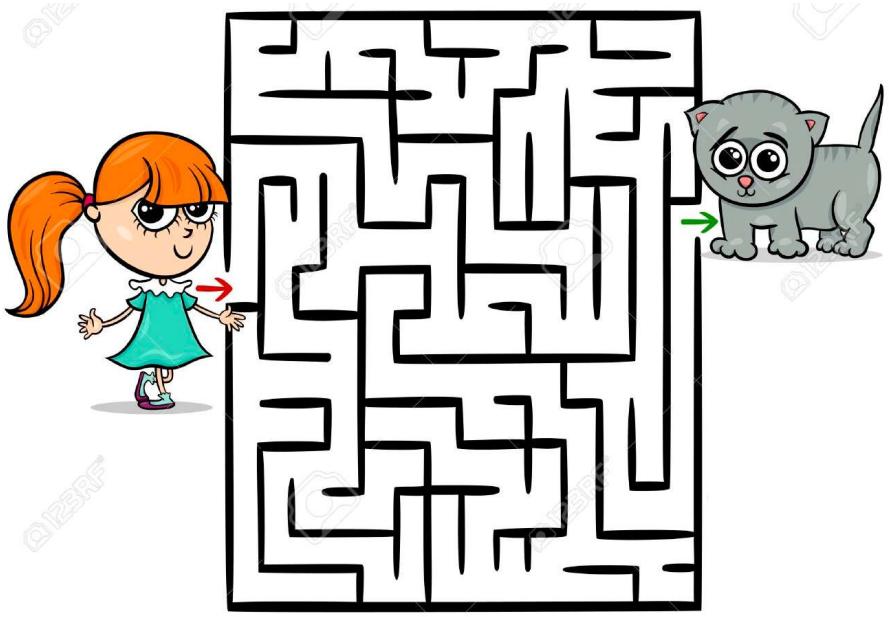
Start

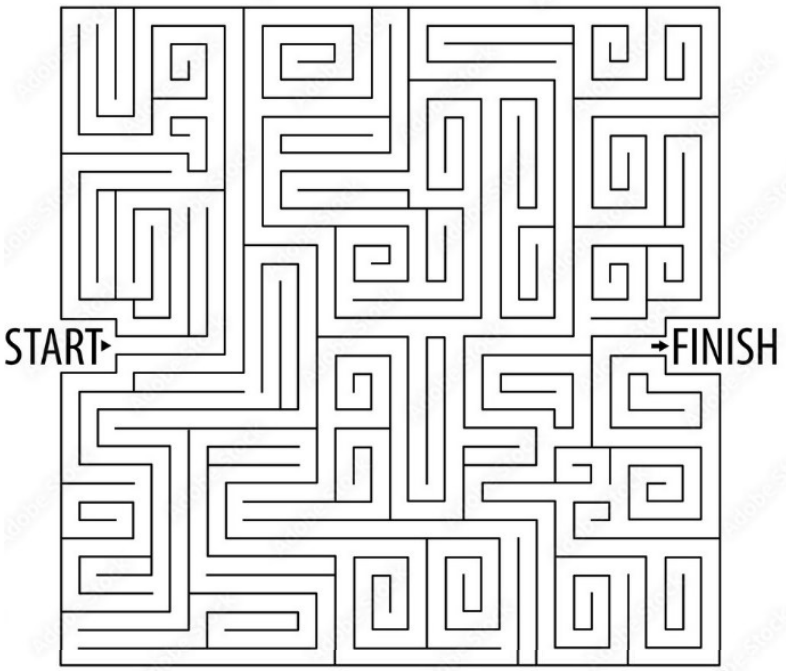


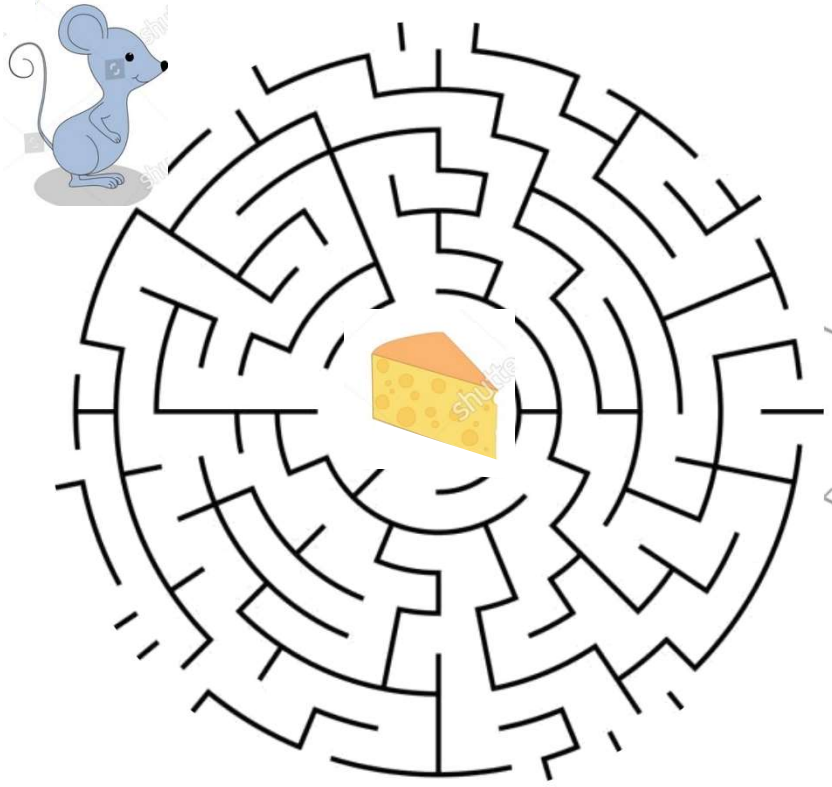
Finish

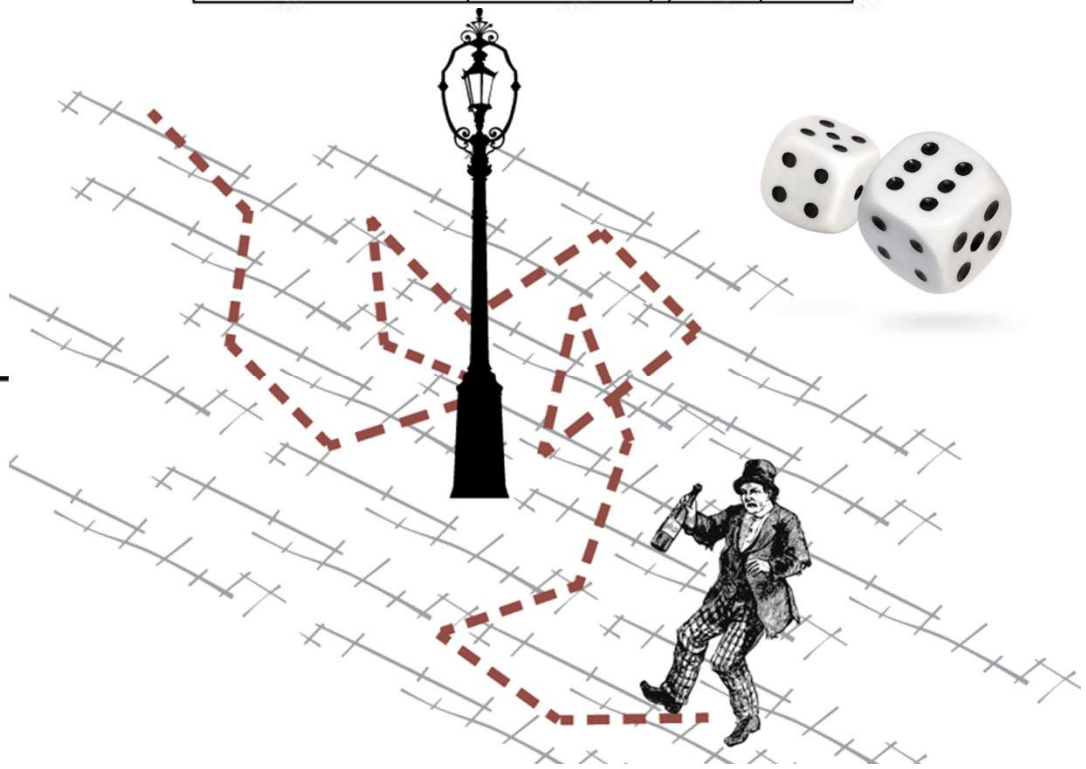


N




START
FINISH





lumpy rewards

rare events

“black swans”

Black swan theory

39 languages

From Wikipedia, the free encyclopedia

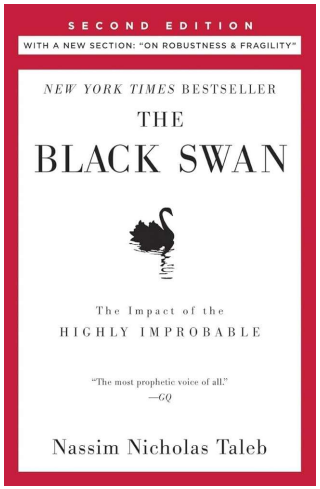
The **black swan theory** or **theory of black swan events** is a metaphor that describes an event that comes as a surprise, has a major effect, and is often inappropriately rationalized after the fact with the benefit of hindsight. The term is based on an ancient Roman saying expressing the European presumption that black swans did not exist until Dutch mariners saw them in Australia in 1697, and the term was then reinterpreted to mean an unforeseen and consequential event.[1]

A black swan (*Cygnus atratus*) in Australia

The theory was developed by Nassim Nicholas Taleb, starting in 2001, to explain:

1. The disproportionate role of high-profile, hard-to-predict, and rare events that are beyond the realm of normal expectations in history, science, finance, and technology.
2. The non-computability of the probability of consequential rare events using scientific methods (owing to the very nature of small probabilities).
3. The psychological biases that blind people, both individually and collectively, to uncertainty and to the substantial role of rare events in historical affairs.

Taleb's "black swan theory" (which differs from the earlier philosophical versions of the problem) refers only to statistically unexpected events of large magnitude and consequence and their dominant role in history. Such events, considered extreme outliers, collectively play vastly larger roles than regular occurrences.[2]:xxi More technically, in the scientific monograph "Silent Risk",[3] Taleb



AI in mathematical sciences:

:

M.Hughes
Y.-H.He
D.Krefl, R.-K.Seong
F.Ruehle
J.Carifio, J.Halverson, D.Krioukov, B.Nelson
A.Cole, G.Shiu
V.Jejjala, A.Kar, O.Parrikar
E.Parr, P.Vaudrevange, M.Wimmer
M.Larfors, R.Schneider
H.Otsuka, K.Takemoto
R.Deen, Y.-H.He, S.-J.Lee, A.Lukas
Y.-H.He, E.Hirst, T.Peterken
T.Akutagawa, K.Hashimoto, T.Sumimoto
M.Bies, M.Cvetic, R.Donagi, L.Lin, M.Liu, F.Ruehle
M.Noormandipour, Y.Sun, B.Haghighat
H.Erbin, R.Finotello
M.Demirtas, L.McAllister, A.Rios-Tascon
L.Anderson, M.Gerdes, J.Gray, S.Krippendorf, N.Raghuram, F.Ruehle
V.Jejjala, D.Pena, C.Mishra
T.Harvey, A.Lukas

:

Nobel Prize to a Neural Net?