Universal laws and architecture 4:
Layering, learning, and decentralized control

John Doyle
John G Braun Professor
Control and Dynamical Systems, EE, BE
Caltech

Compute

Turing

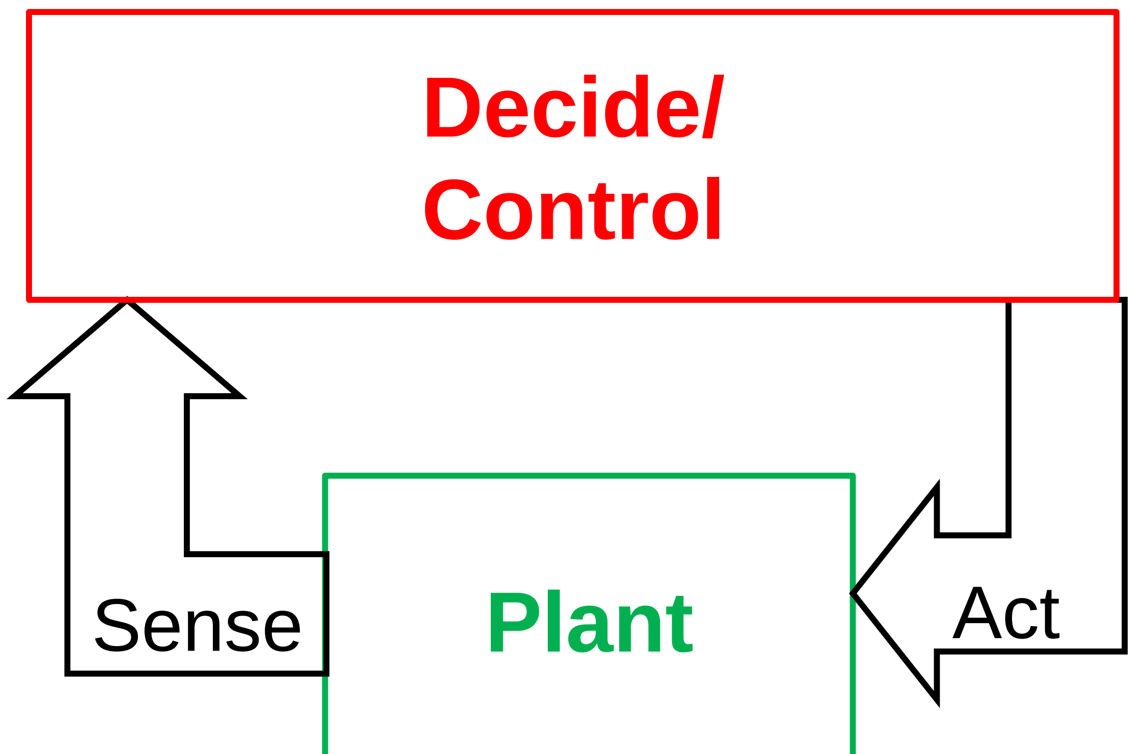
Closing the loop

- On the “plant”
- On the “story”

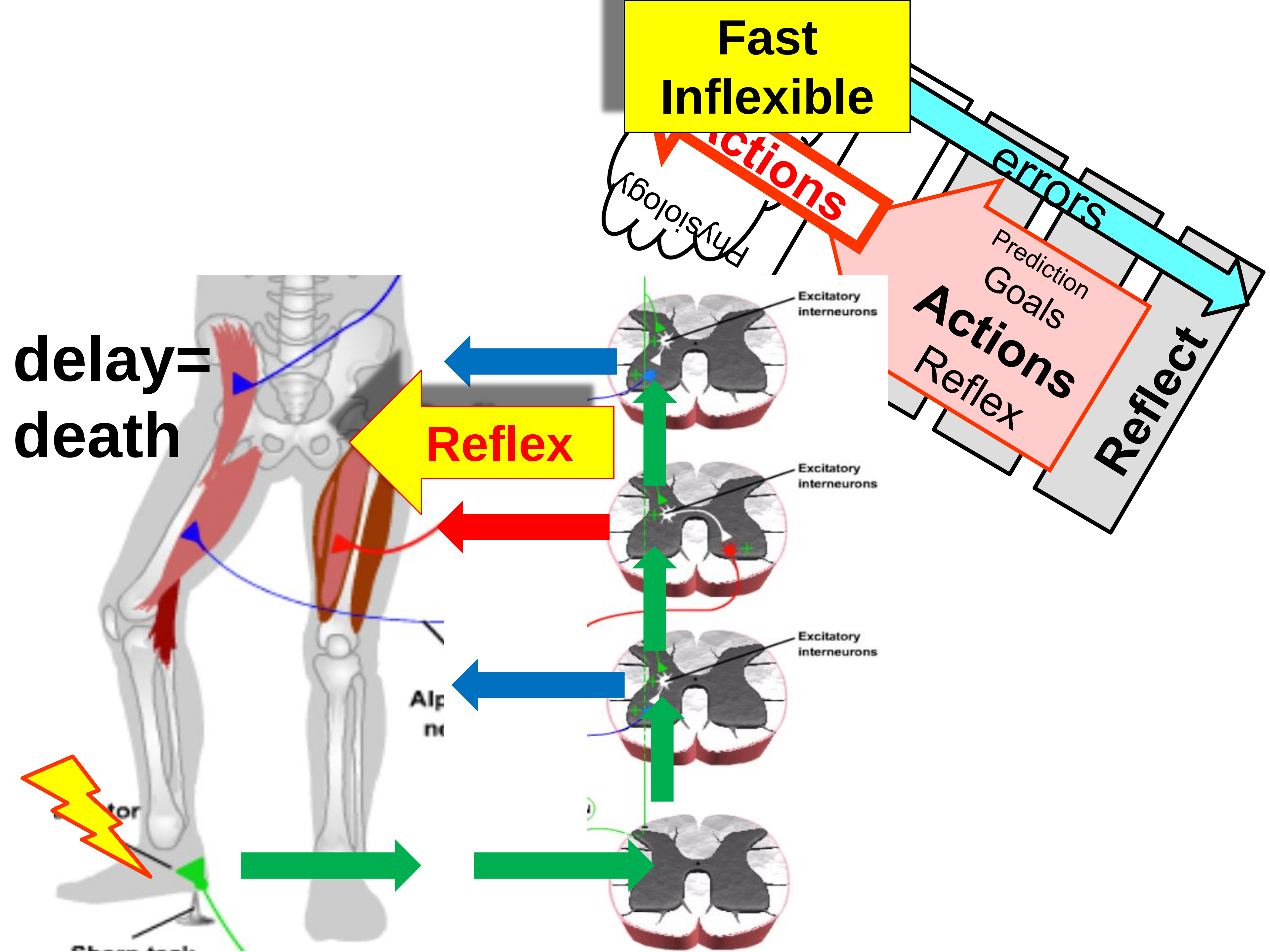
Delay is *most* important

Bode

Control



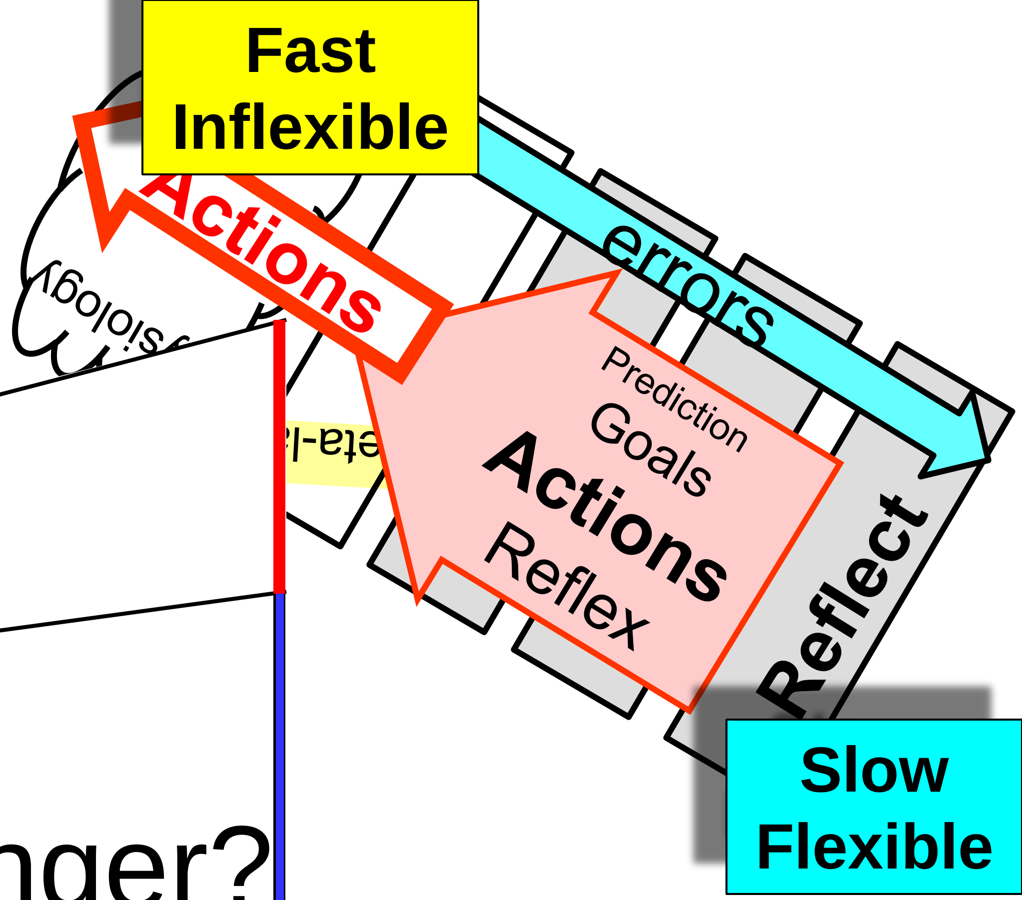
delay= death

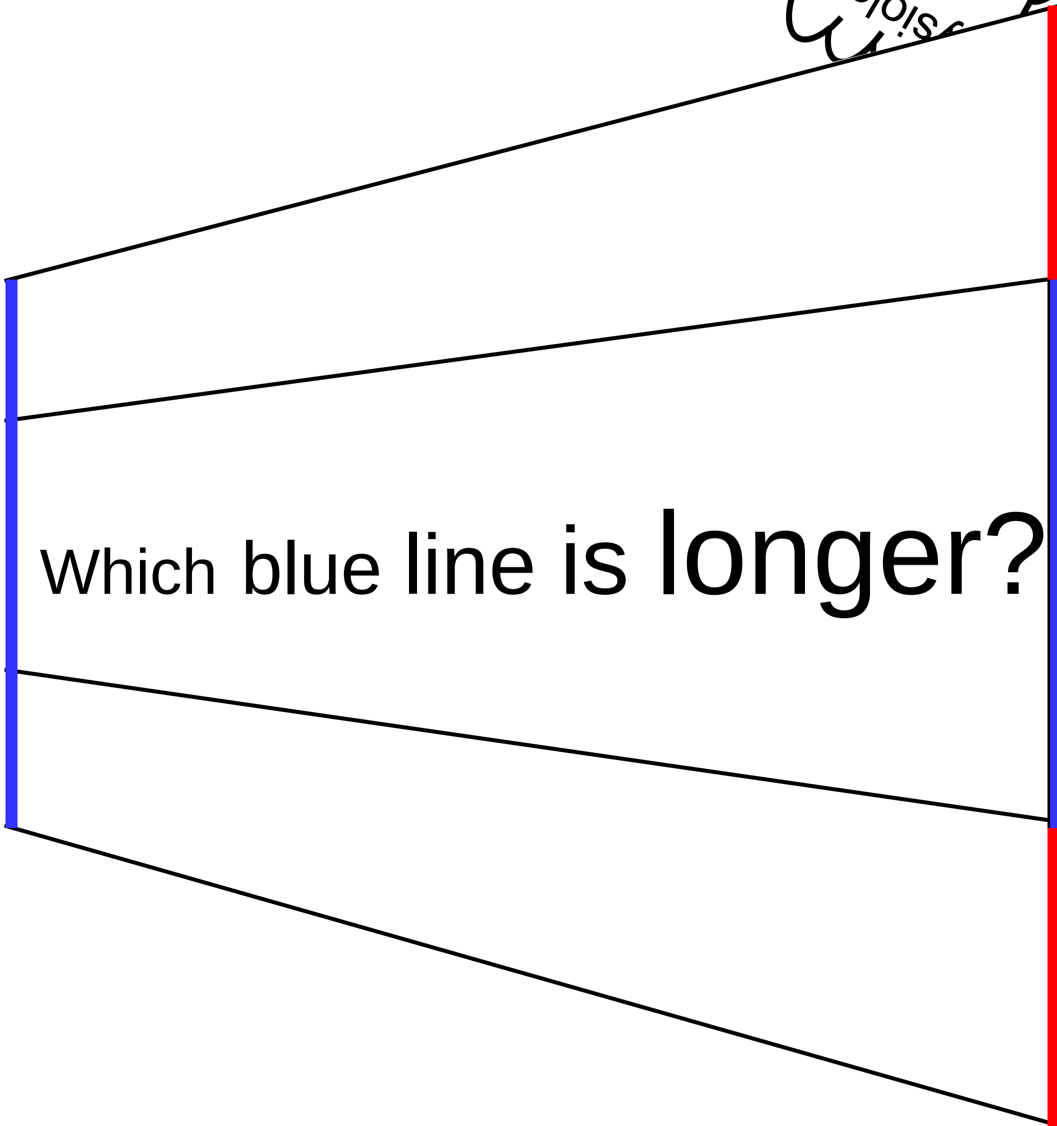


Fast
Inflexible
Physiology
ctions
errors
Prediction
Goals
Actions
Reflex
Reflect
Excitatory interneurons
Excitatory interneurons
Excitatory interneurons
Reflex

"Seeing is dreaming?"

"Seeing is believing?"



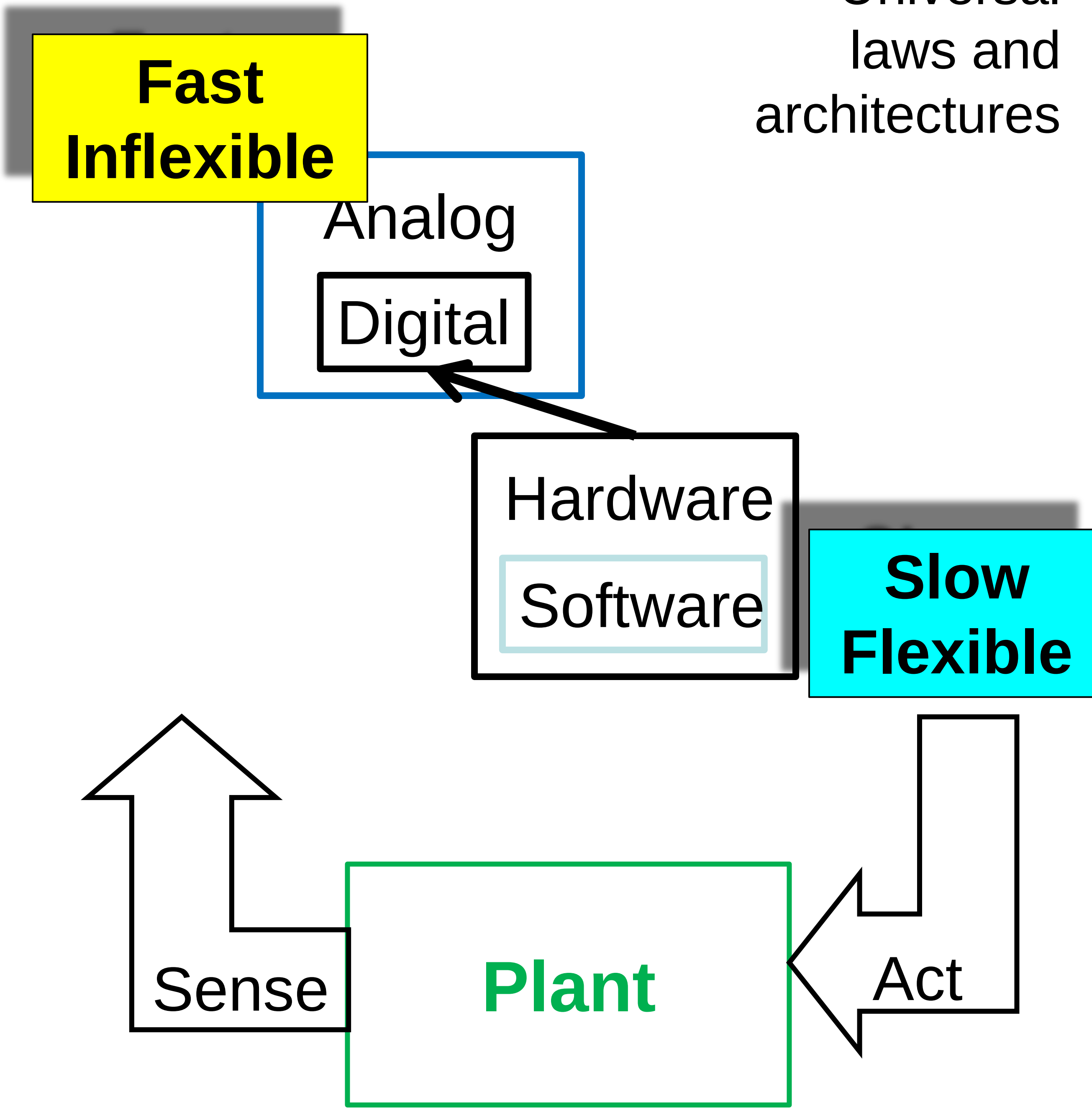


Universal laws and architectures

Compute

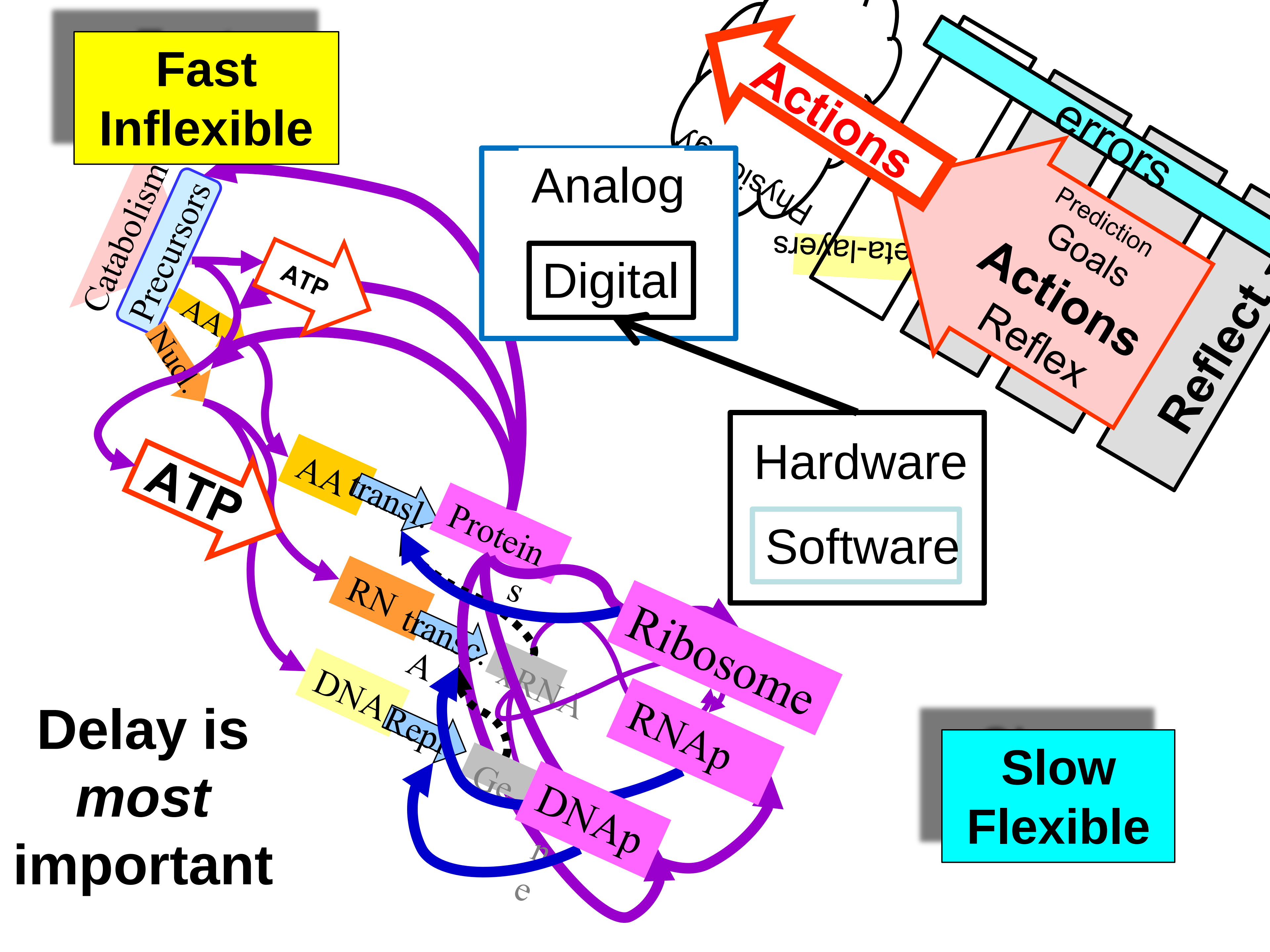
Turing

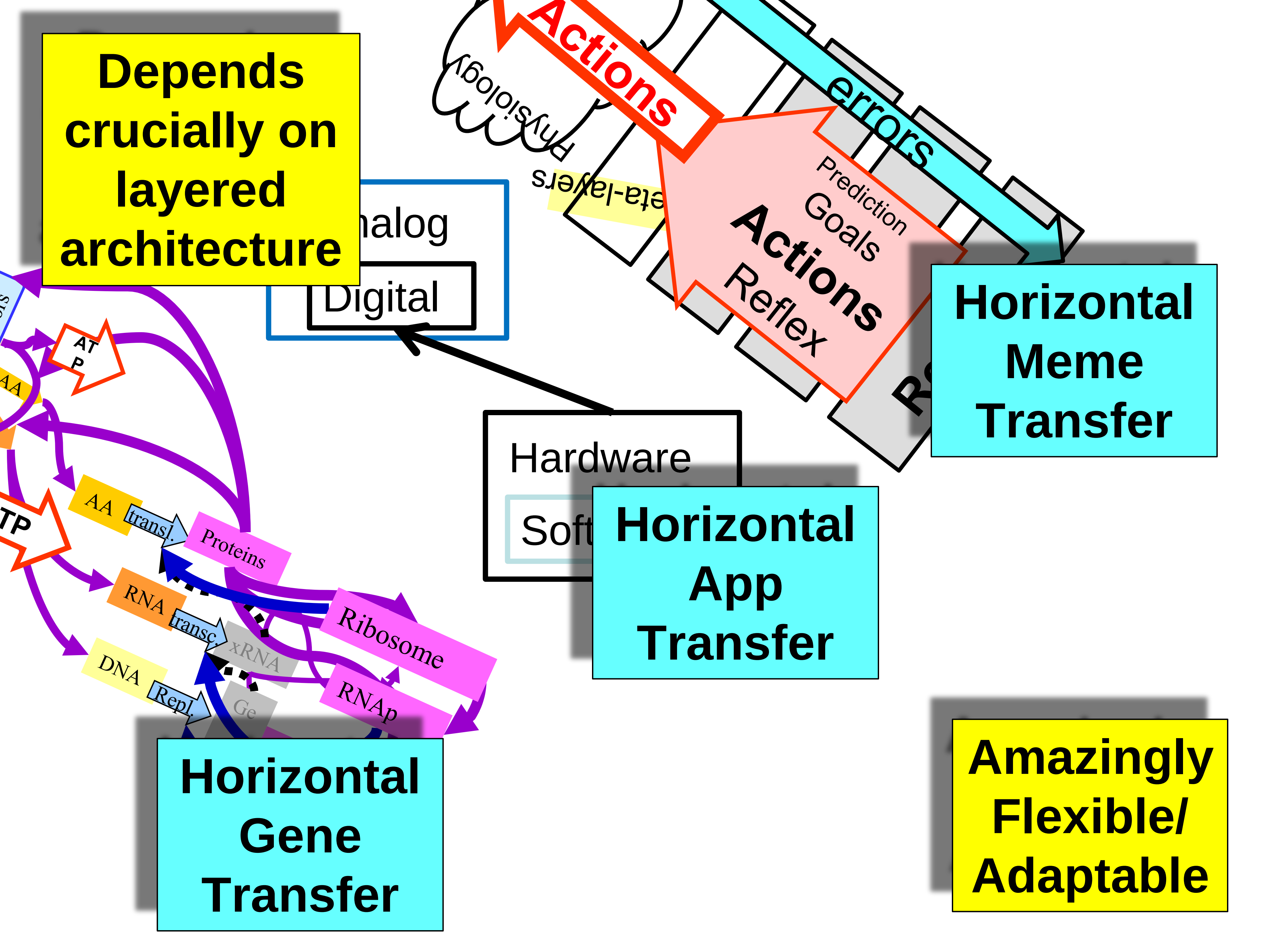
Delay is *most* important



Bode

Control


Fast
Inflexible
Catabolism
Precursors
ATP
AA
Nucl.
ATP
AA transl.
Protein
RN transc.
DNA Repl.
Ribosome
RNAp
DNAp
Analog
Digital
Hardware
Software
Actions
errors
Prediction
Goals
Actions
Reflex
Reflect
Slow
Flexible
Delay is
most
important


Depends crucially on layered architecture
Horizontal Meme Transfer
Horizontal App Transfer
Horizontal Gene Transfer
Amazingly Flexible/ Adaptable
Actions
errors
Physiology
Prediction
Goals
Actions
Reflex
Digital
Hardware
Soft
AA
transl.
Proteins
RNA
transc.
xRNA
Ribosome
DNA
Repl.
RNAp

Exploiting layered architecture

In case I forget…

Horizontal **Bad Meme** Transfer

Fragility?

Horizontal **Virus** Transfer

Parasites & Hijacking

Horizontal **Virus** Transfer

Depends crucially on layered architecture

Build on Turing to show what is *necessary* to make this work.

Horizontal Meme Transfer

- Acquire
- Translate/ integrate
- Automate

Horizontal App Transfer

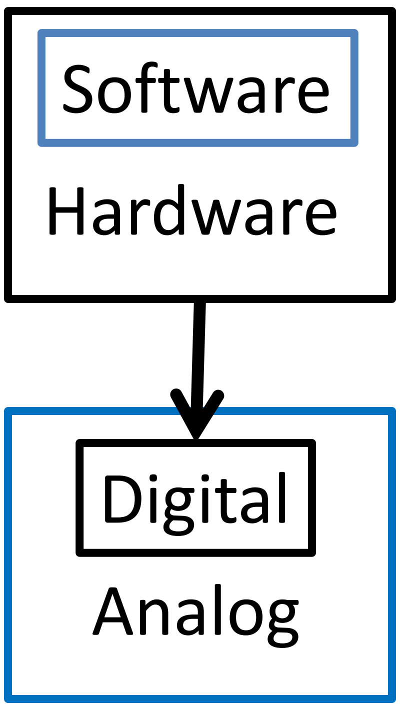
Horizontal Gene Transfer

Amazingly Flexible/ Adaptable

Turing as "new" starting point?

Essentials:

0. **Model**
1. **Universal laws**
2. **Universal architecture**
3. **Practical implementation**



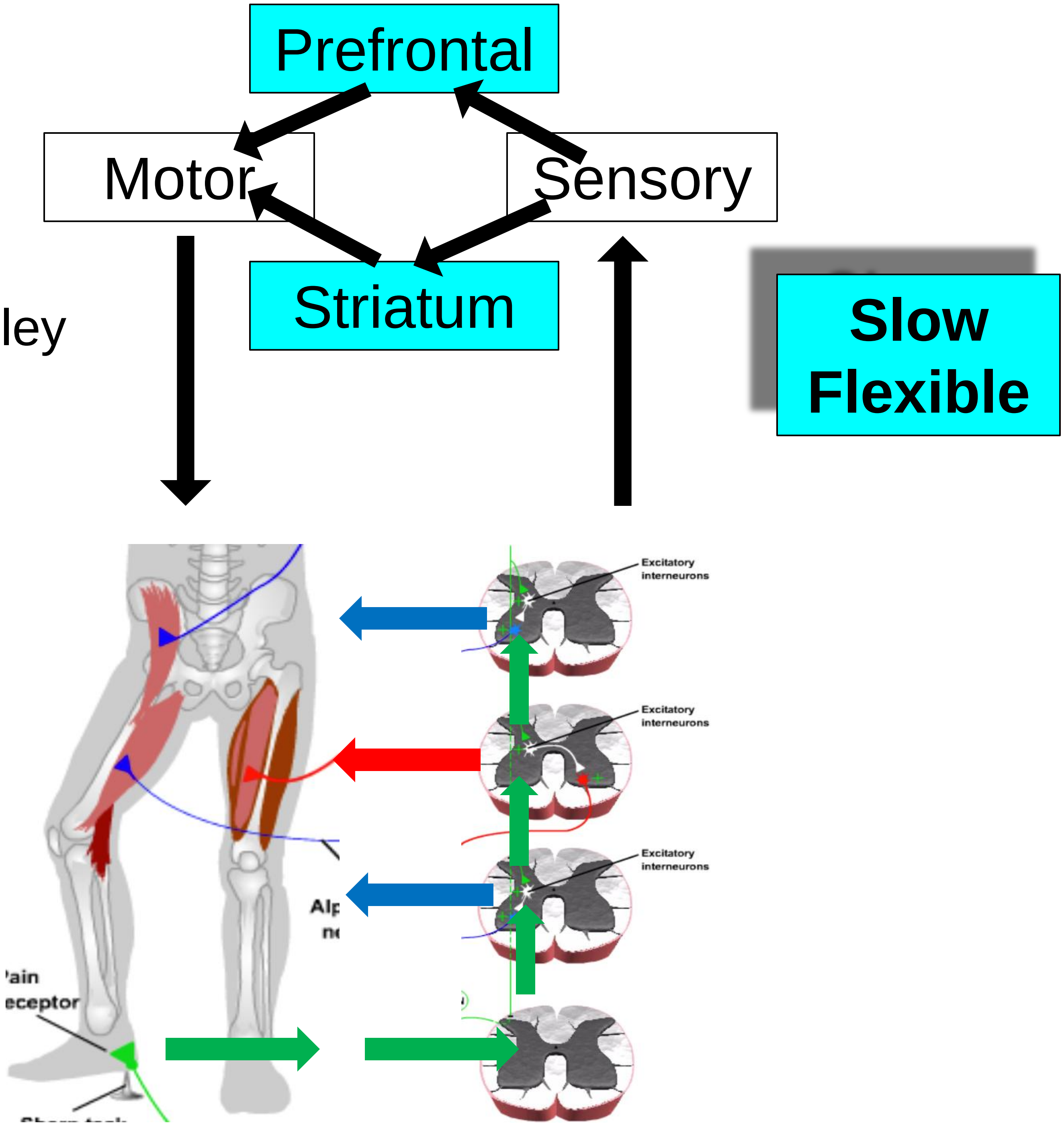
Turing's 3 step research:

0. Virtual (TM) machines
1. hard limits, (un)decidability using standard model (TM)
2. Universal architecture achieving hard limits (UTM)
3. Practical implementation in digital electronics (biology?)

Learning

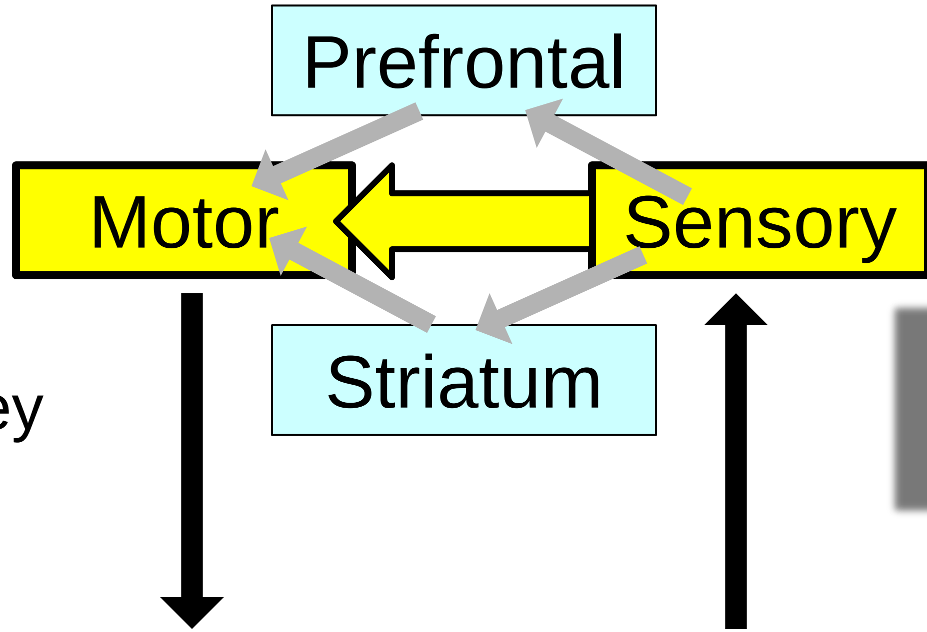
Ashby & Crossley

- **Acquire**
- Translate/ integrate
- Automate



Thanks to
Bassett & Grafton

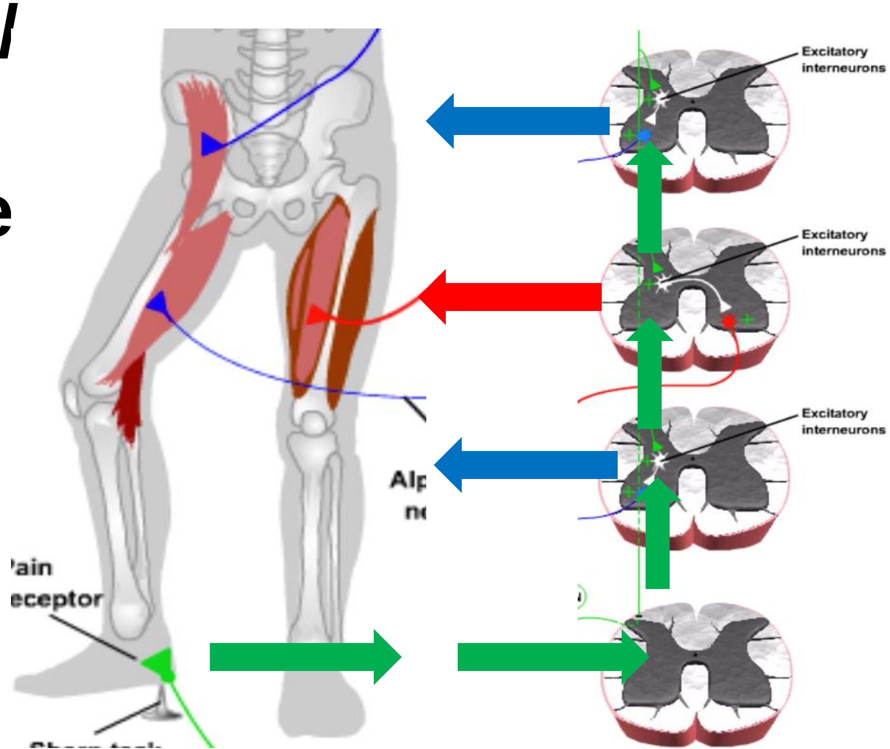
Fast
Inflexible

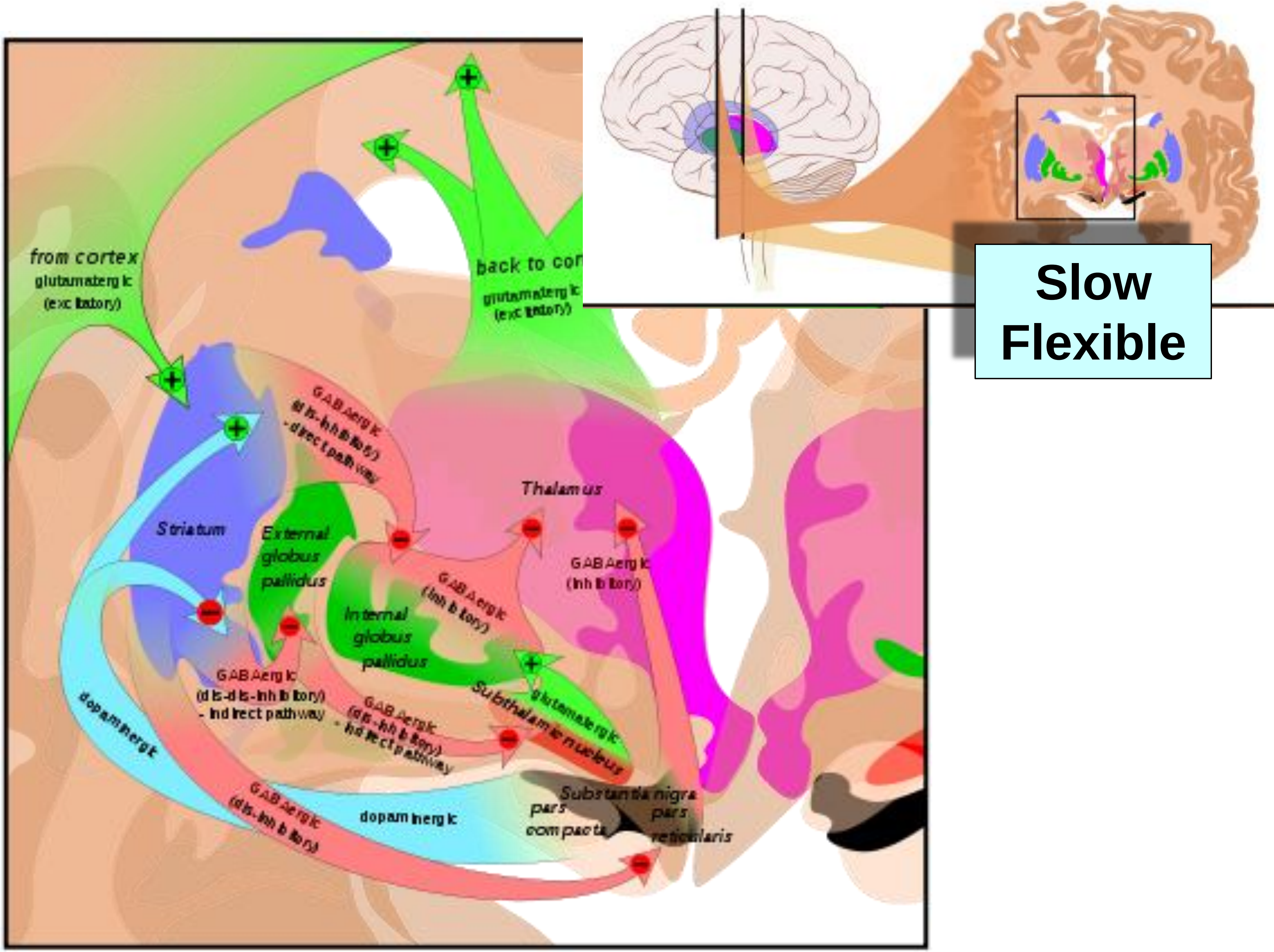


Slow
Flexible

Ashby & Crossley

- Acquire
- **Translate/ integrate**
- **Automate**





Slow
Flexible

Build on Turing to show what is *necessary* to make this work.

Horizontal Meme Transfer

Fast Inflexible

Prefrontal

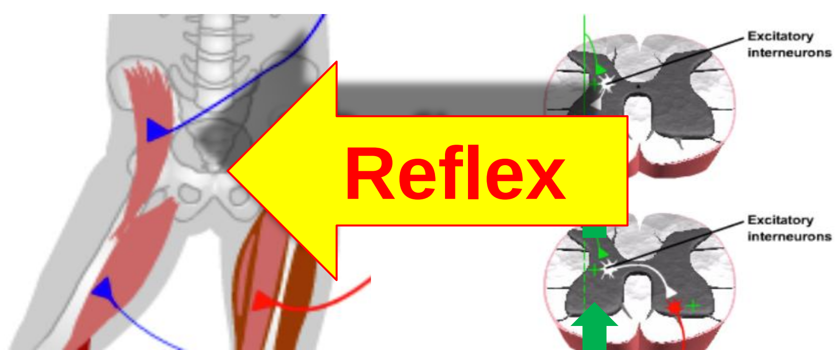
Motor

Sensory

Striatum

Slow Flexible

- **Acquire**
- **Translate/ integrate**
- **Automate**

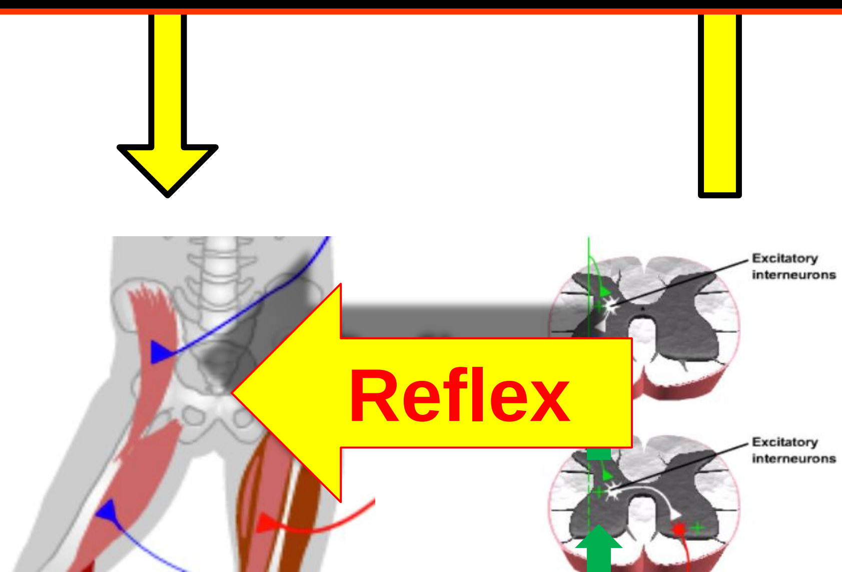


Wolpert, Grafton, etc

robust

Brain as ~~optimal~~ controller

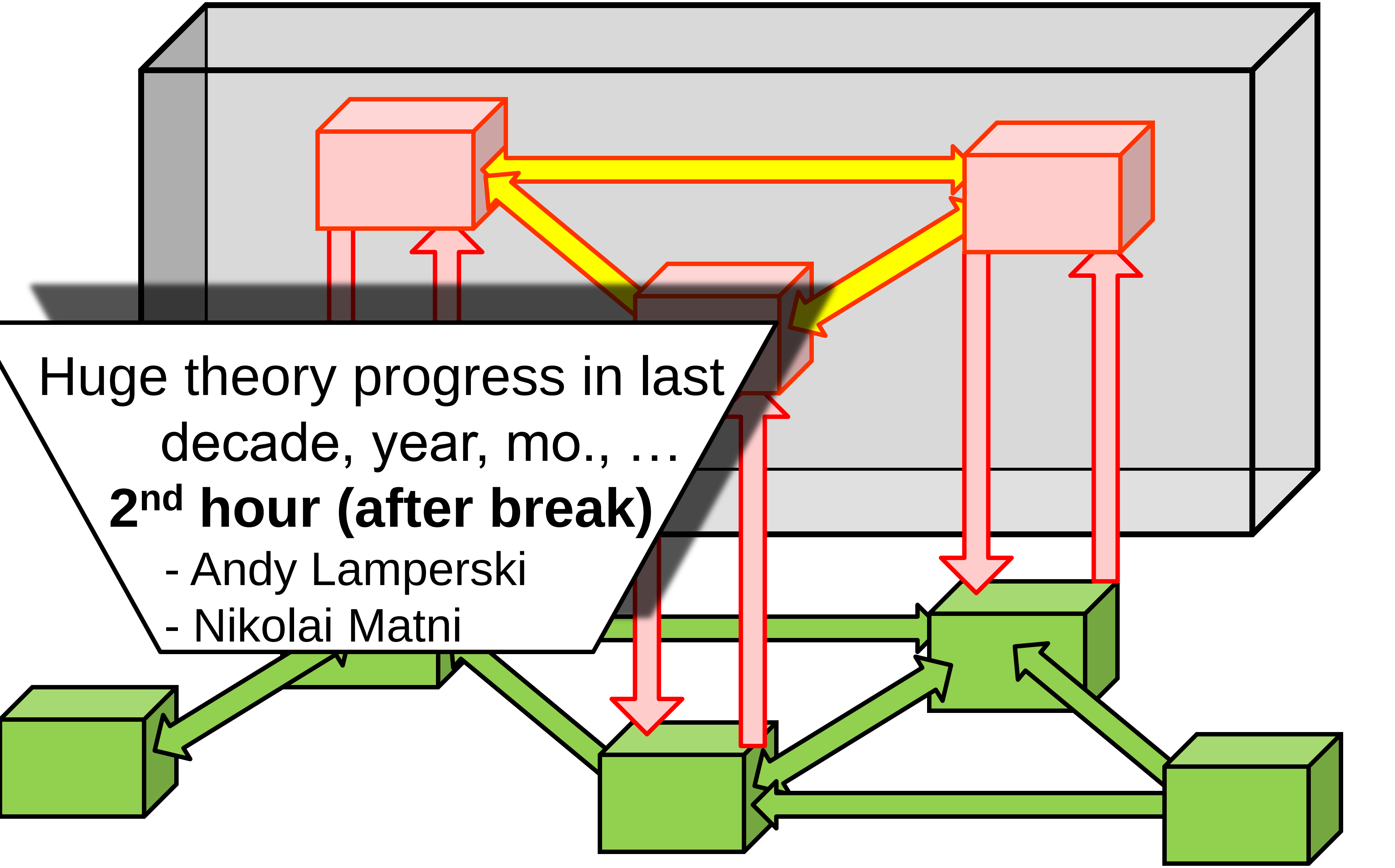
- Acquire
- Translate/ integrate
- **Automate**



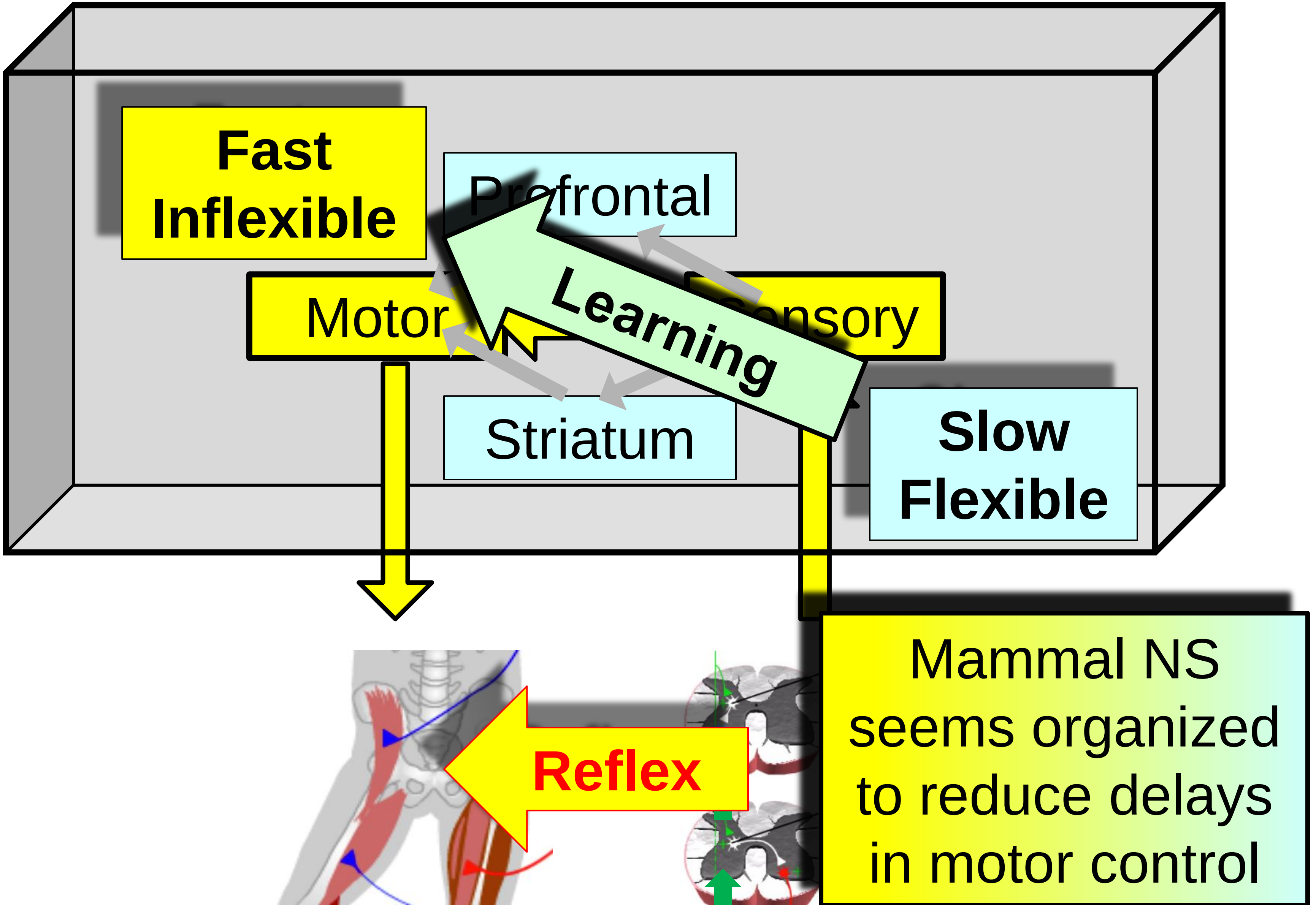
What I’m not going to talk about

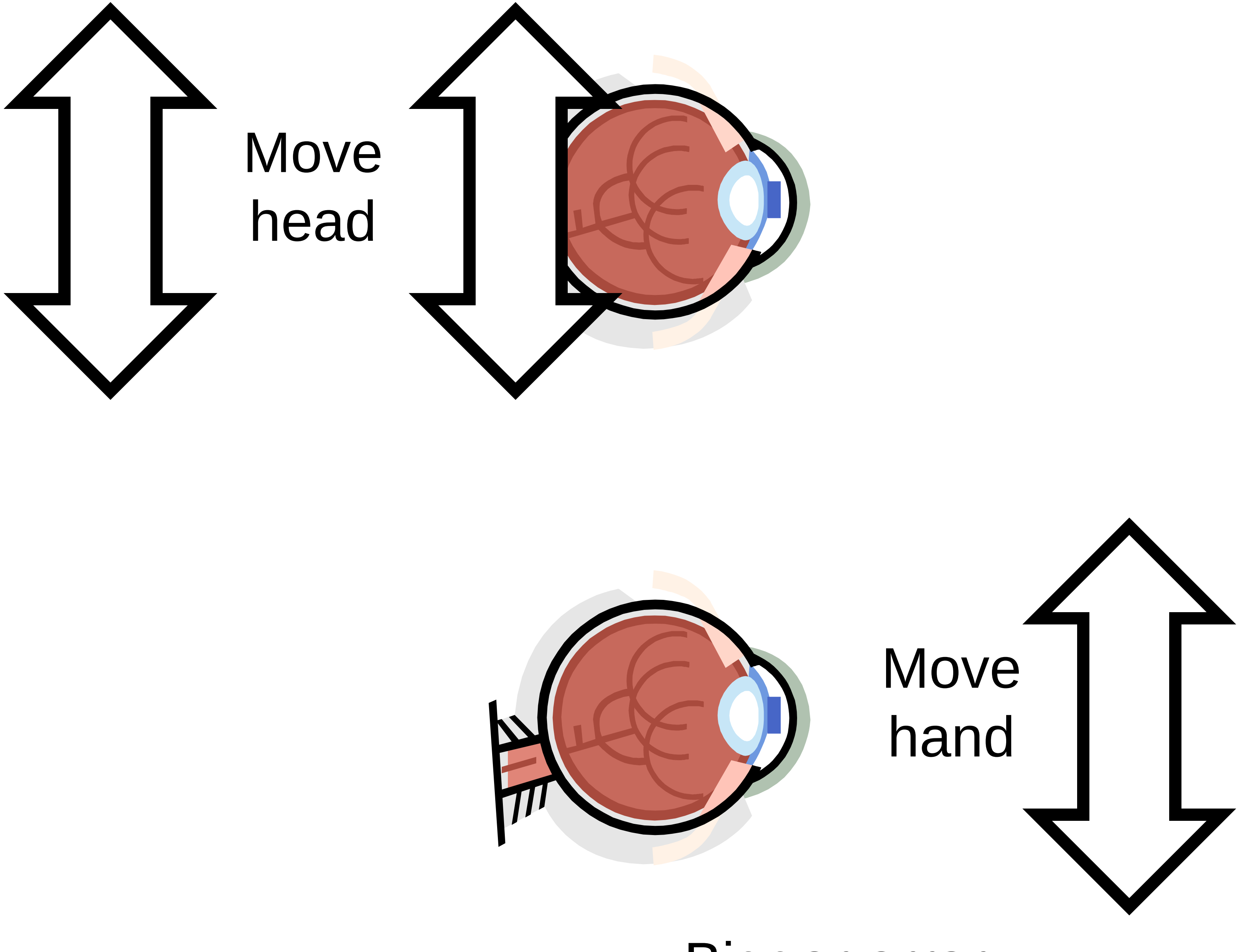
- Connections between robustness and risk sensitivity
- Asymmetry between false positives and negatives
- Risk aversion and risk seeking
- Uncertainty is more in models than in probabilities
- Life is not like a casino

Going beyond black box: control is decentralized with internal delays.

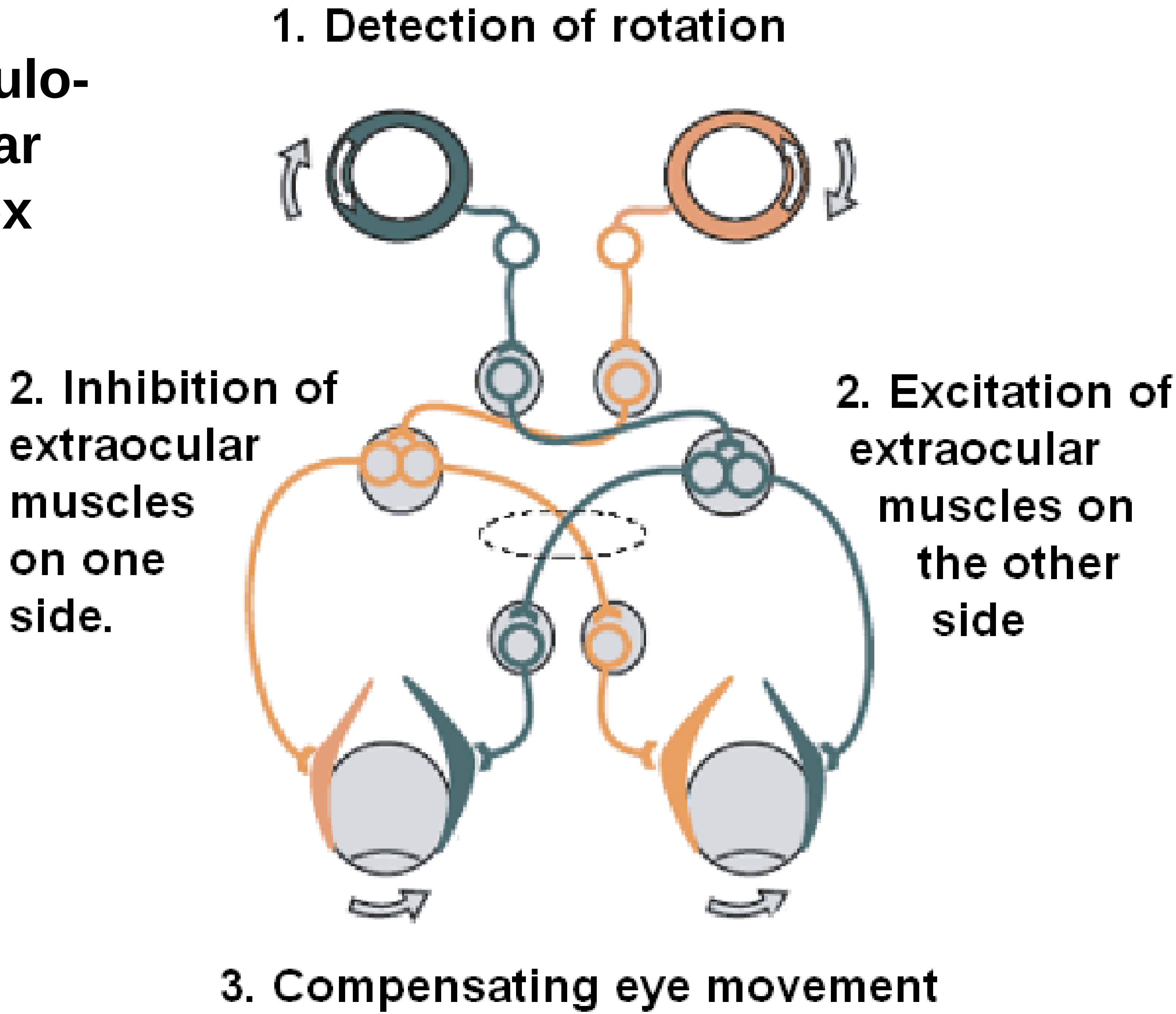


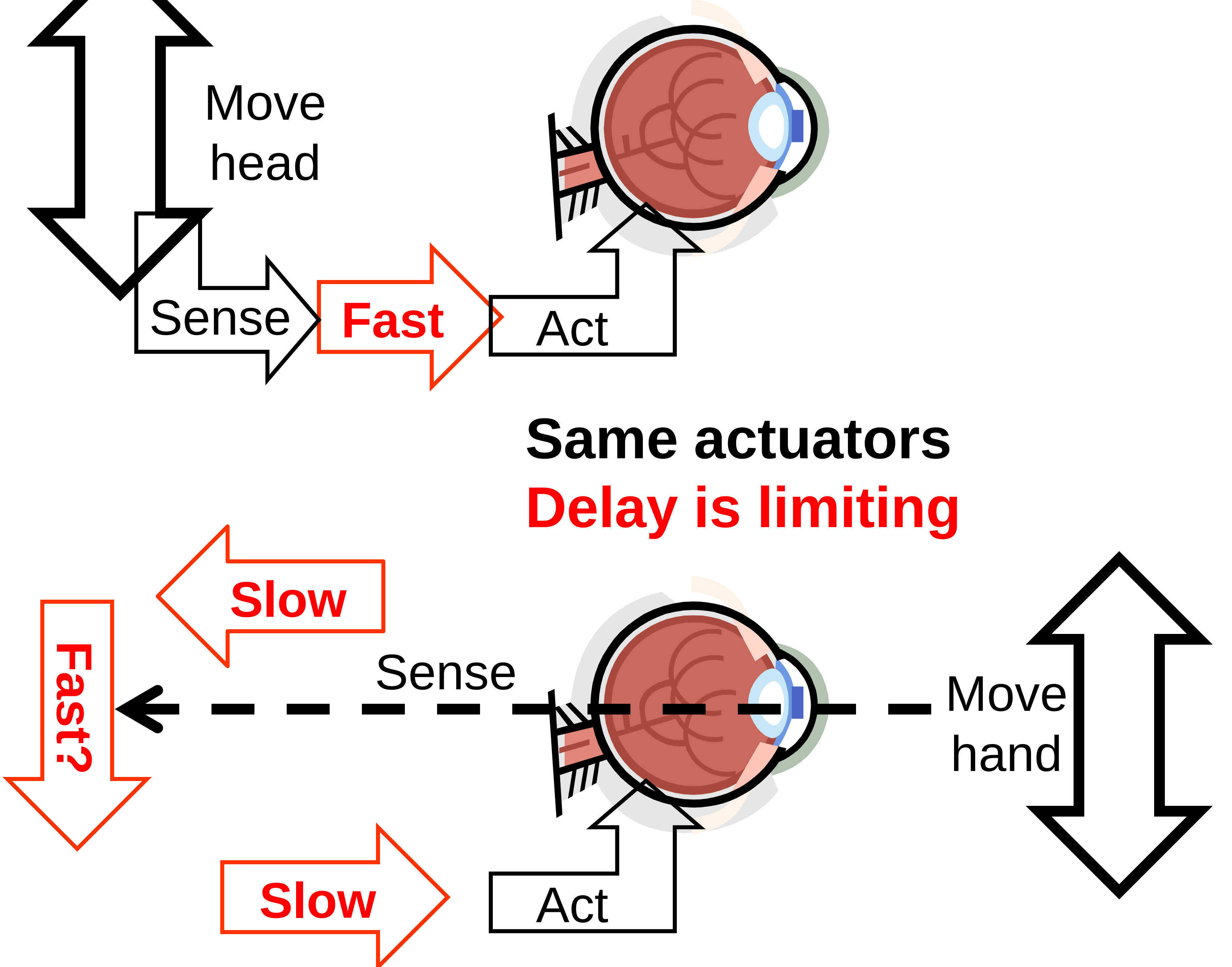
Going beyond black box: control is decentralized with internal delays.




Move
head
Move
hand
Bigger error

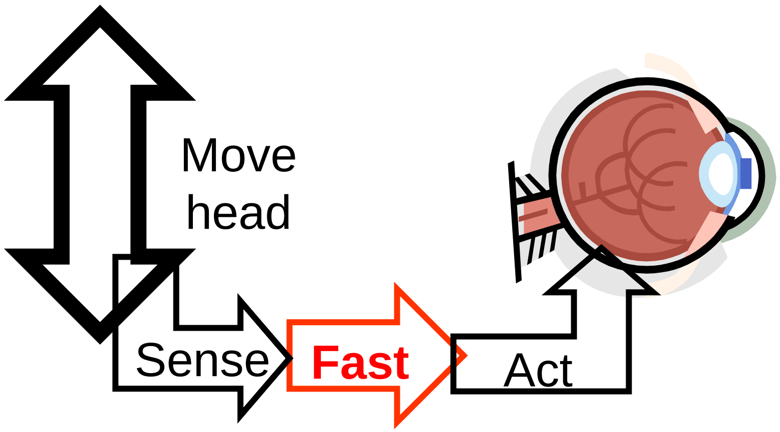
Vestibulo-ocular reflex




Move
head
Sense
Fast
Act
Same actuators
Delay is limiting
Slow
Fast?
Sense
Move
hand
Slow
Act

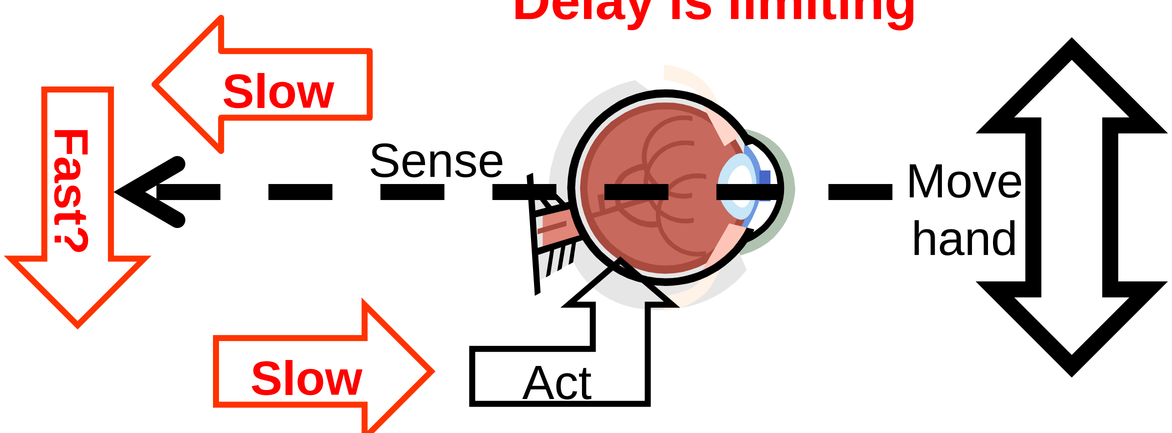
Versus standing on one leg

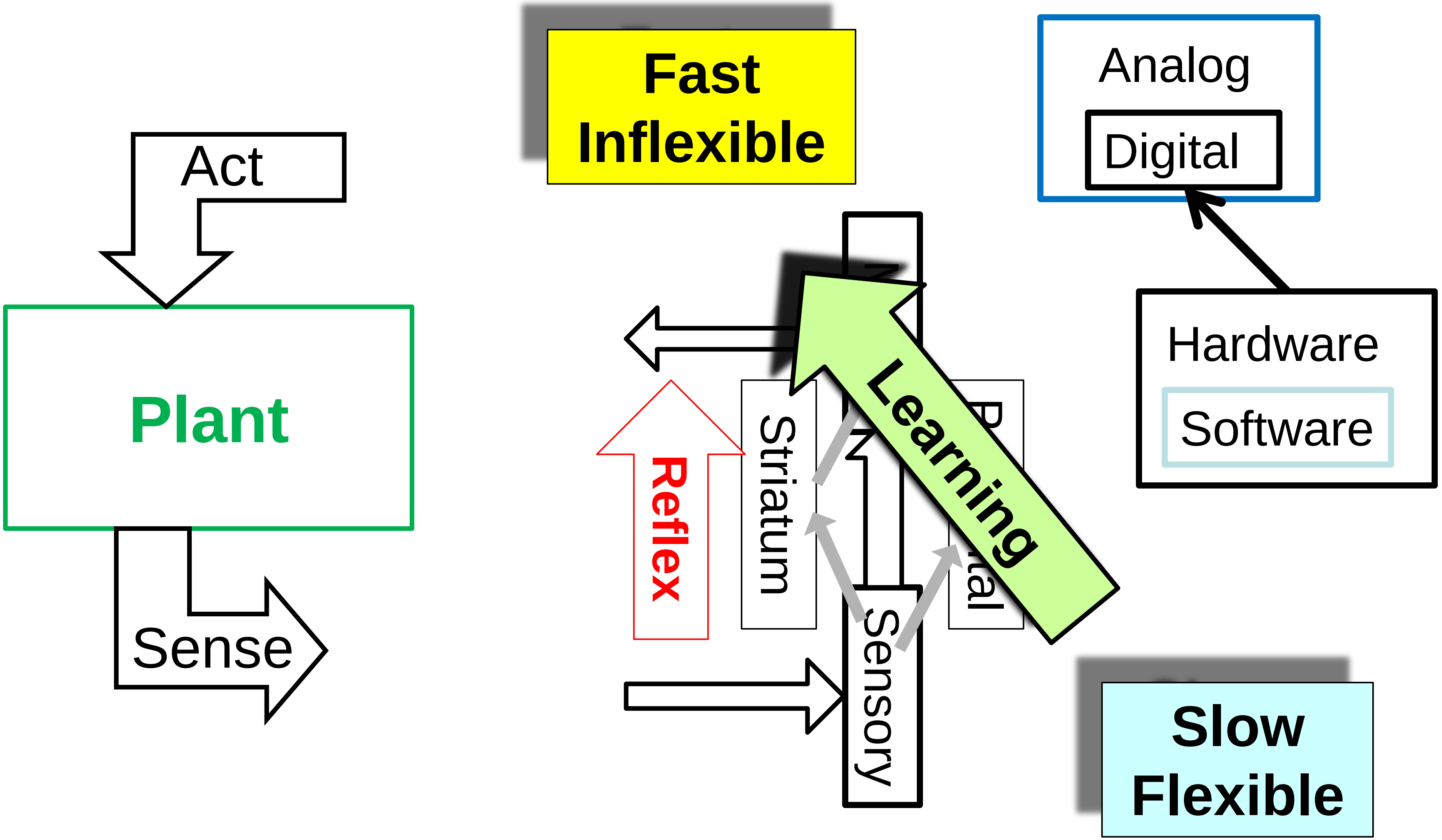
- Eyes open vs closed
- Contrast
  - young surfers
  - old football players



Same actuators

Delay is limiting





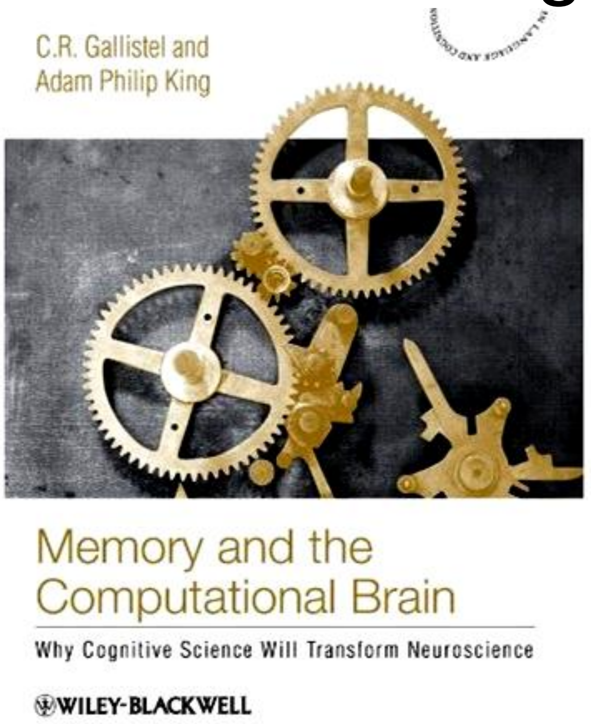
Universal tradeoffs*:*

Speed and ***flexibility*** are crucial
to implementing robust controllers.

Issues for neuroscience

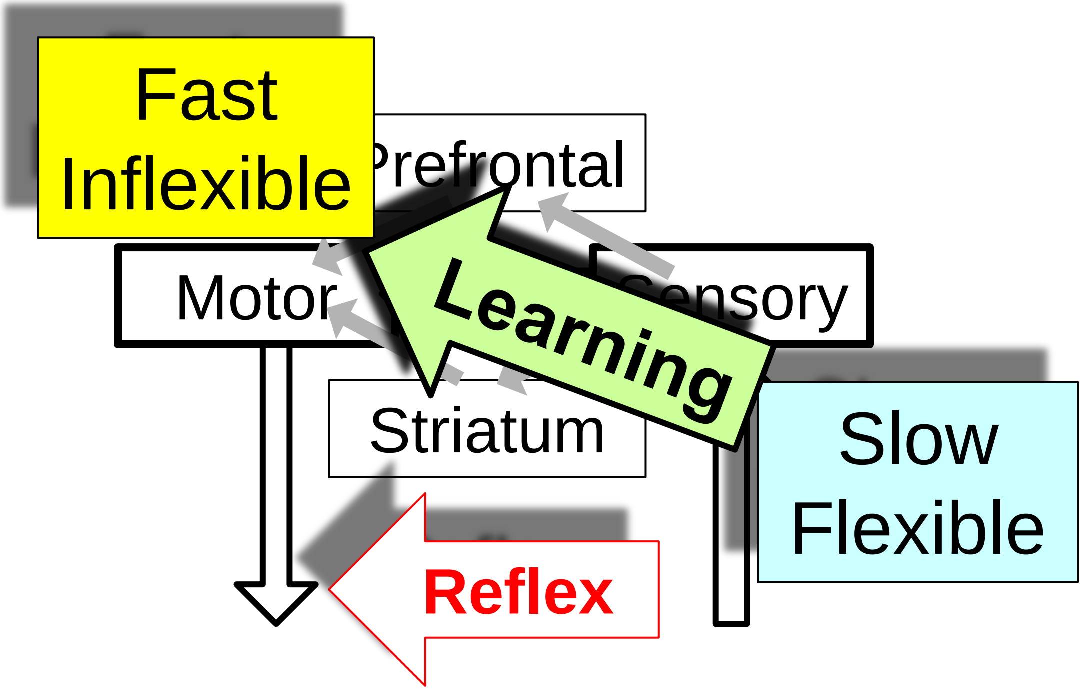
- Brains and UTMs?
  - Time is most critical resource?
  - Space (memory) almost free?
- Read/write random access memory hierarchies?
- Brain >> UTM?

Gallistel and King

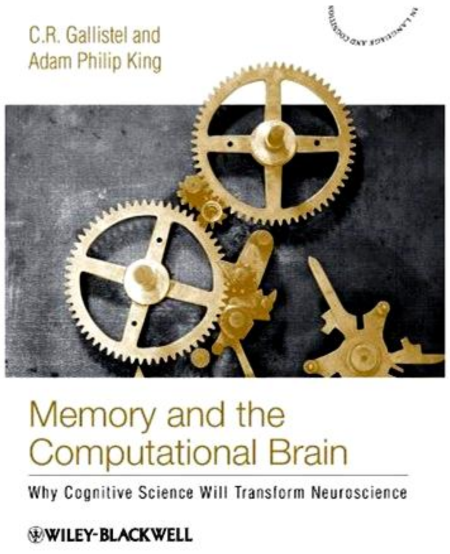


Conjecture

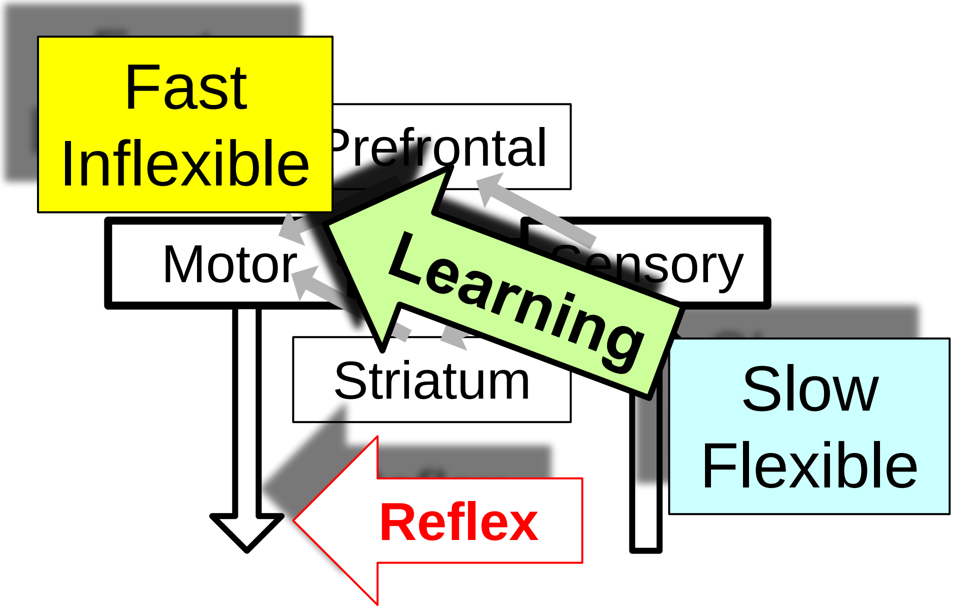
- Memory potential ≈ ∞
- Examples
  - Insects
  - Scrub jays
  - Autistic Savants



Gallistel and King

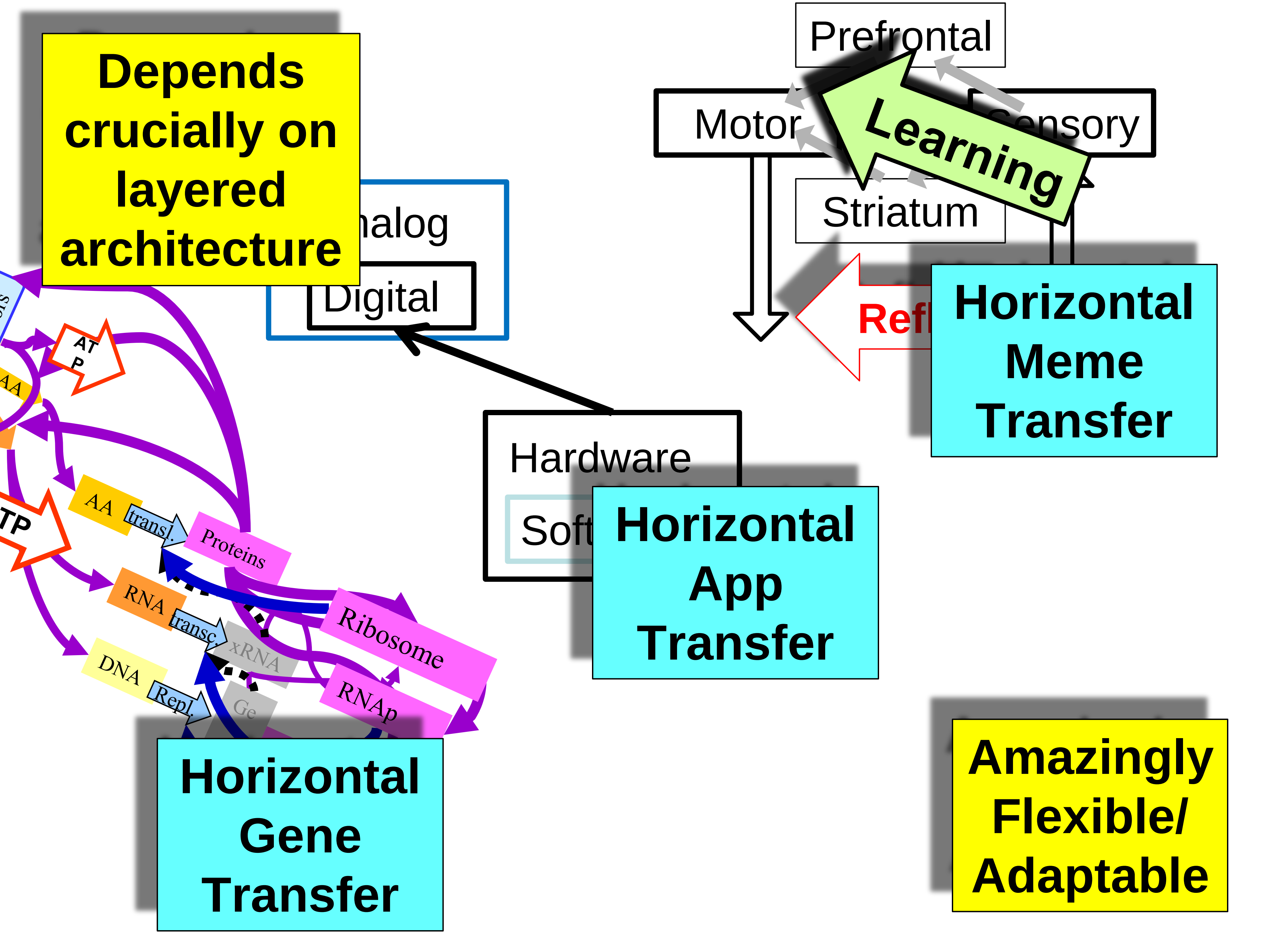


- Sensori-motor memory potential ≈ ∞ (Ashby)
- Limits are on ***speed*** of
  - nerve propagation delays
  - learning
- But control is ***never*** centralized
- Is there a random access read/write memory?



The EvoPsycho question: why?

- Sensori-motor memory potential ≈ ∞ (Ashby)
- Limits are on ***speed*** of
  - nerve propagation delays (fish parts?)
  - learning ???
- I'm probably confused
- What about robust learning


Depends crucially on layered architecture
Prefrontal
Motor
Sensory
Learning
Striatum
nalog
Digital
Refl
Horizontal Meme Transfer
Hardware
Soft
Horizontal App Transfer
AT P
AA
TP
AA
transl.
Proteins
RNA
transc.
xRNA
Ribosome
DNA
Repl.
Ge
RNAp
Horizontal Gene Transfer
Amazingly Flexible/ Adaptable

What I'm not going to talk about

- It's true that most "really smart scientists" think almost everything in this talk is nonsense
- Why they think this
- Why they are wrong

- Time (not space) is our problem, as usual
- Don't have enough time for what is true, so have to limit discussion of what isn't
- No one ever changes a made up mind (almost)

What I'm not going to talk about

Compute

Turing

Delay is *most* important

Bode

Control, OR

Communicate

Shannon

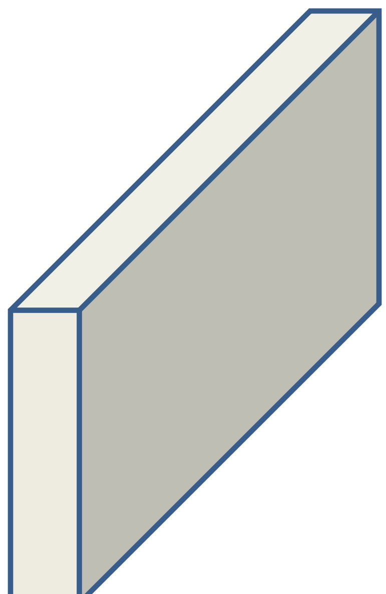
Delay is ~~*least*~~ important

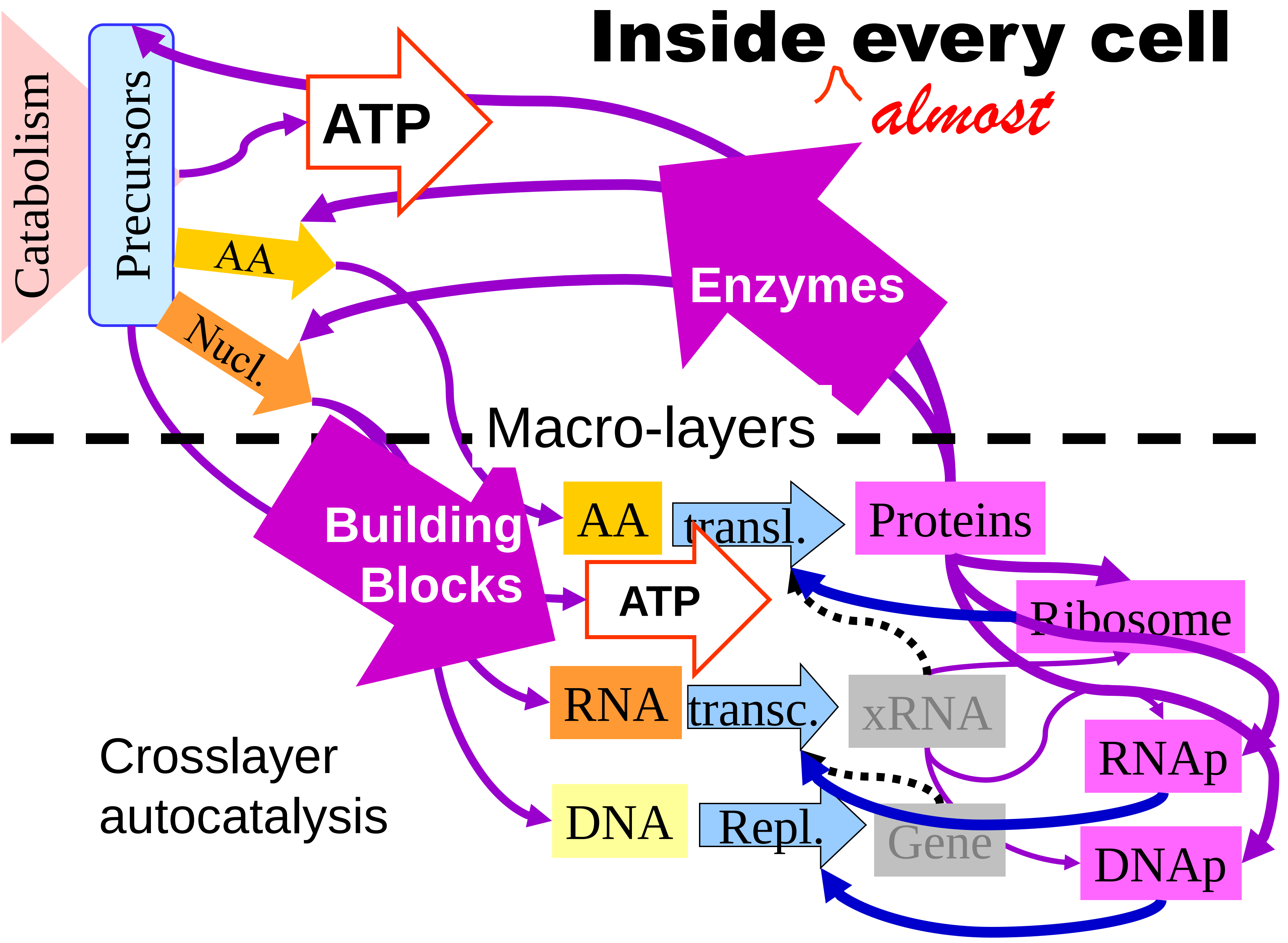
Carnot

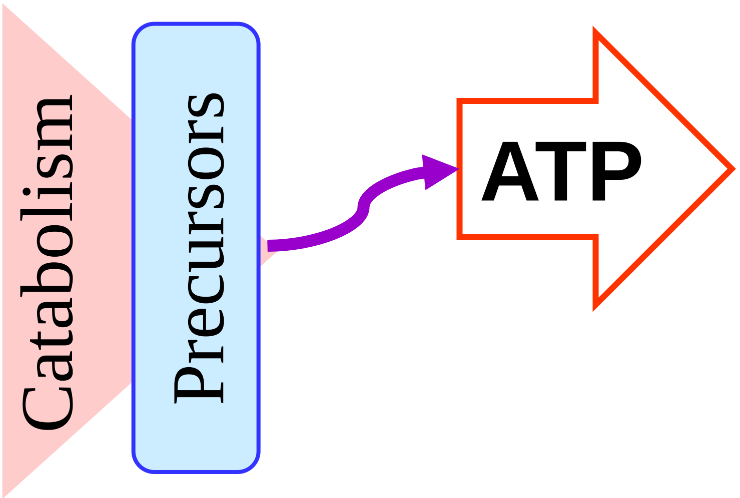
Boltzmann

Heisenberg

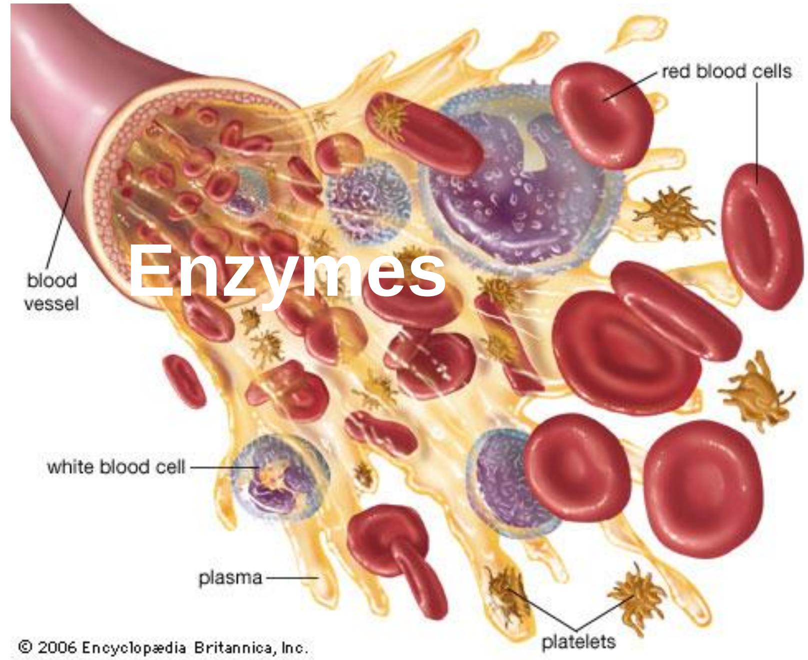
Physics




Inside every cell
almost
Catabolism
Precursors
ATP
AA
Nucl.
Enzymes
Macro-layers
Building Blocks
AA
transl.
Proteins
ATP
Ribosome
RNA
transc.
xRNA
RNAp
DNA
Repl.
Gene
DNAp
Crosslayer autocatalysis



Inside every cell

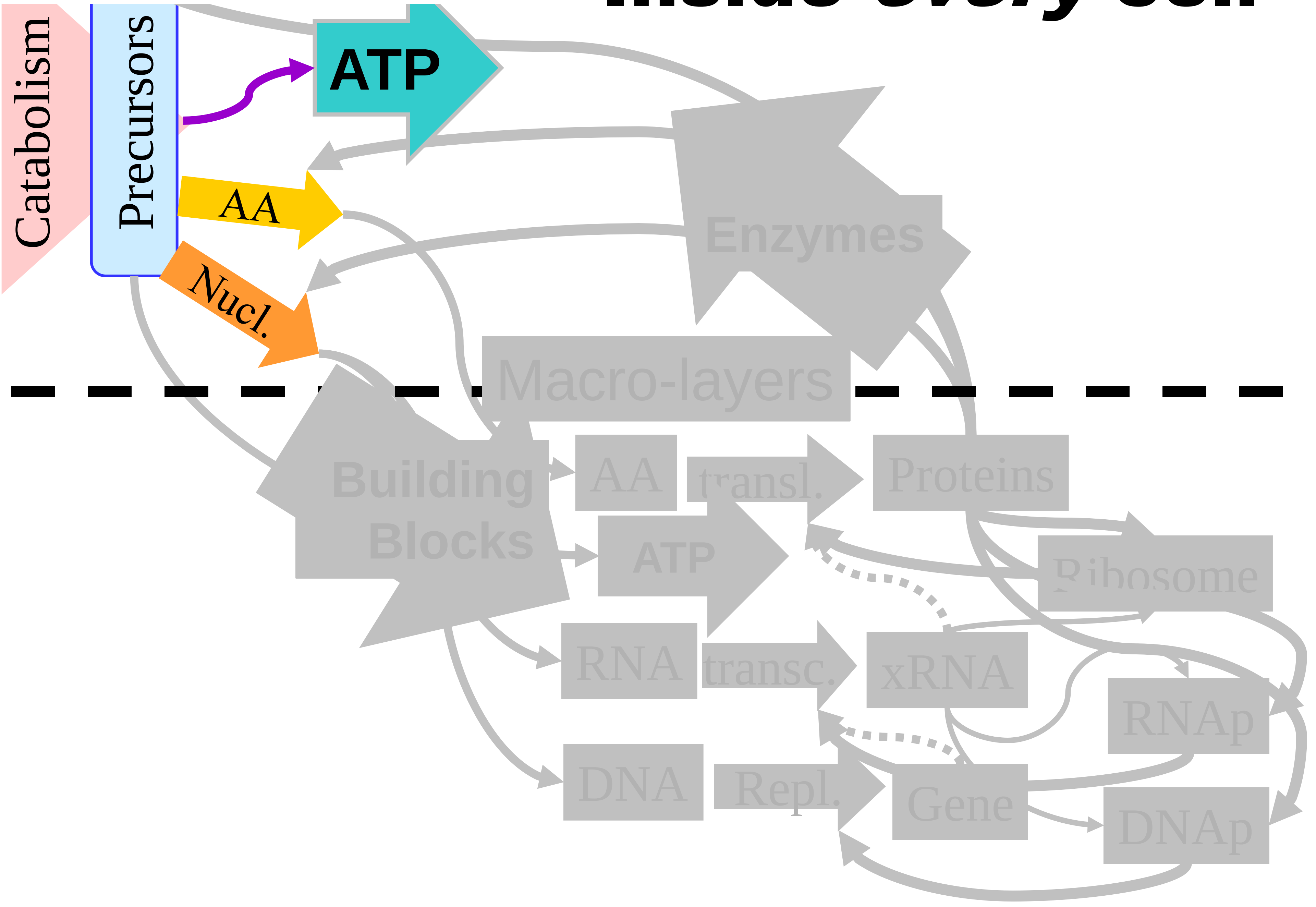


Mature red blood cells live 120 days

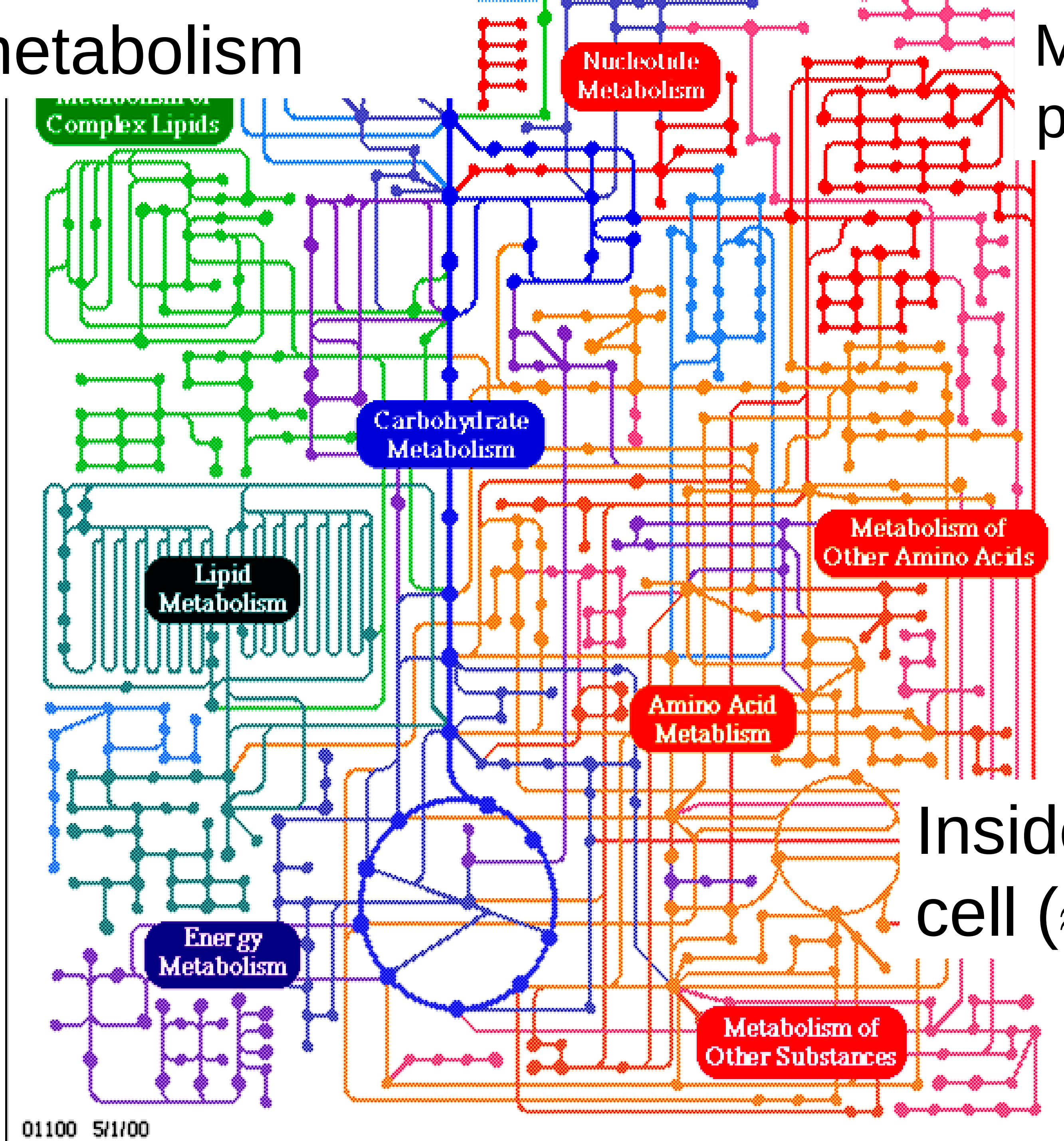
or "metabolism first" origins of life?

Core metabolism

Inside *every* cell



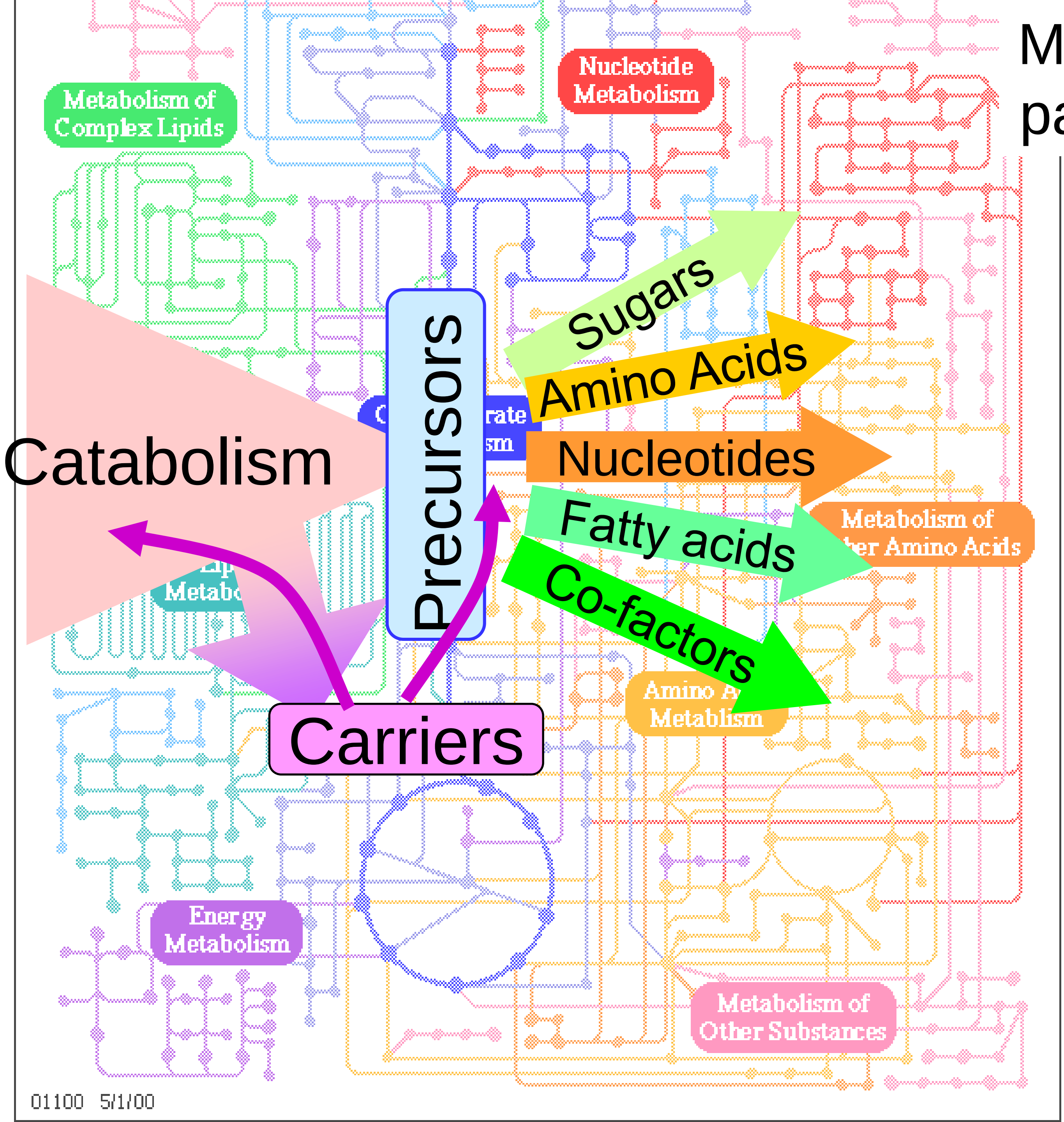
Core metabolism

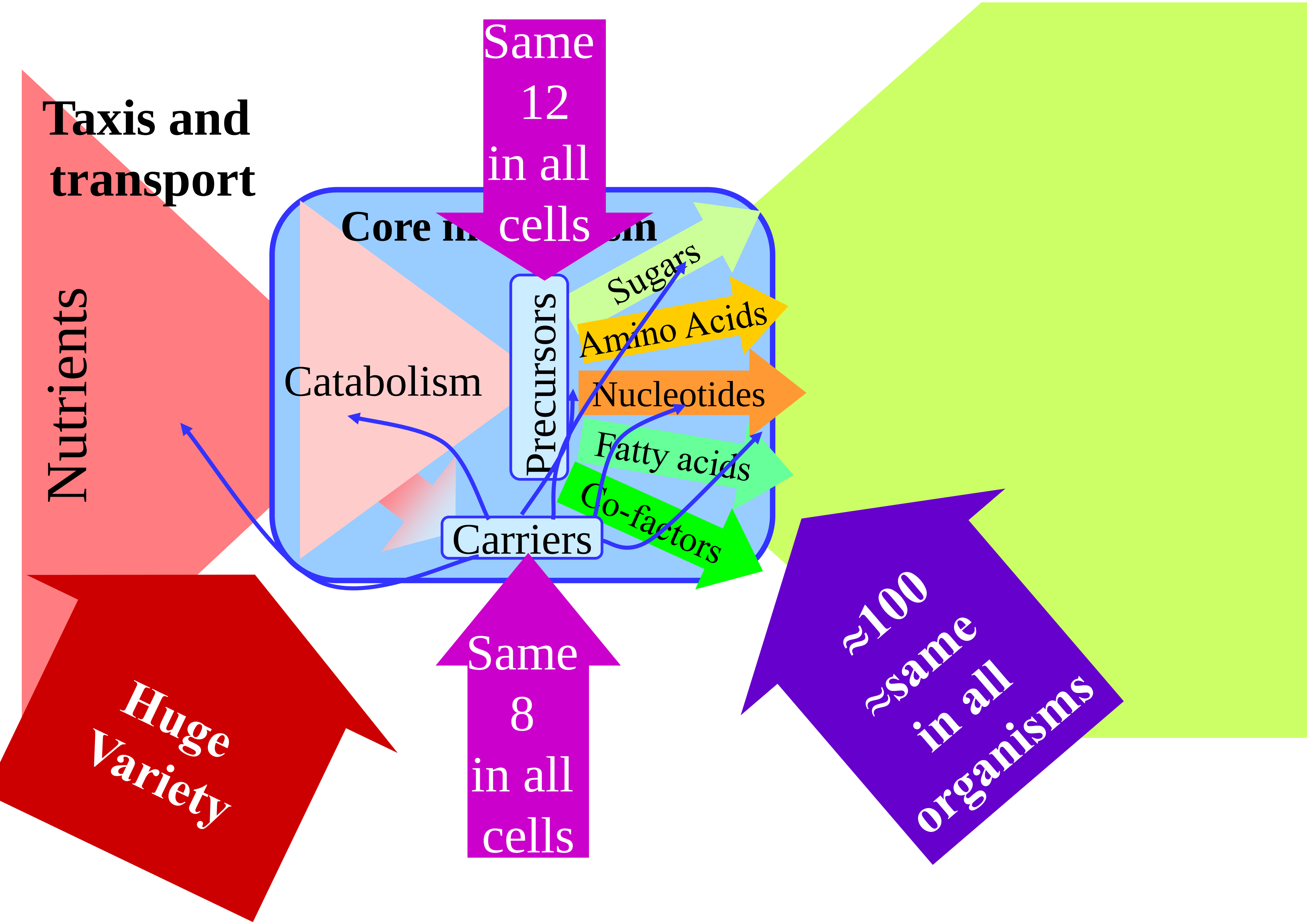


Metabolic pathways

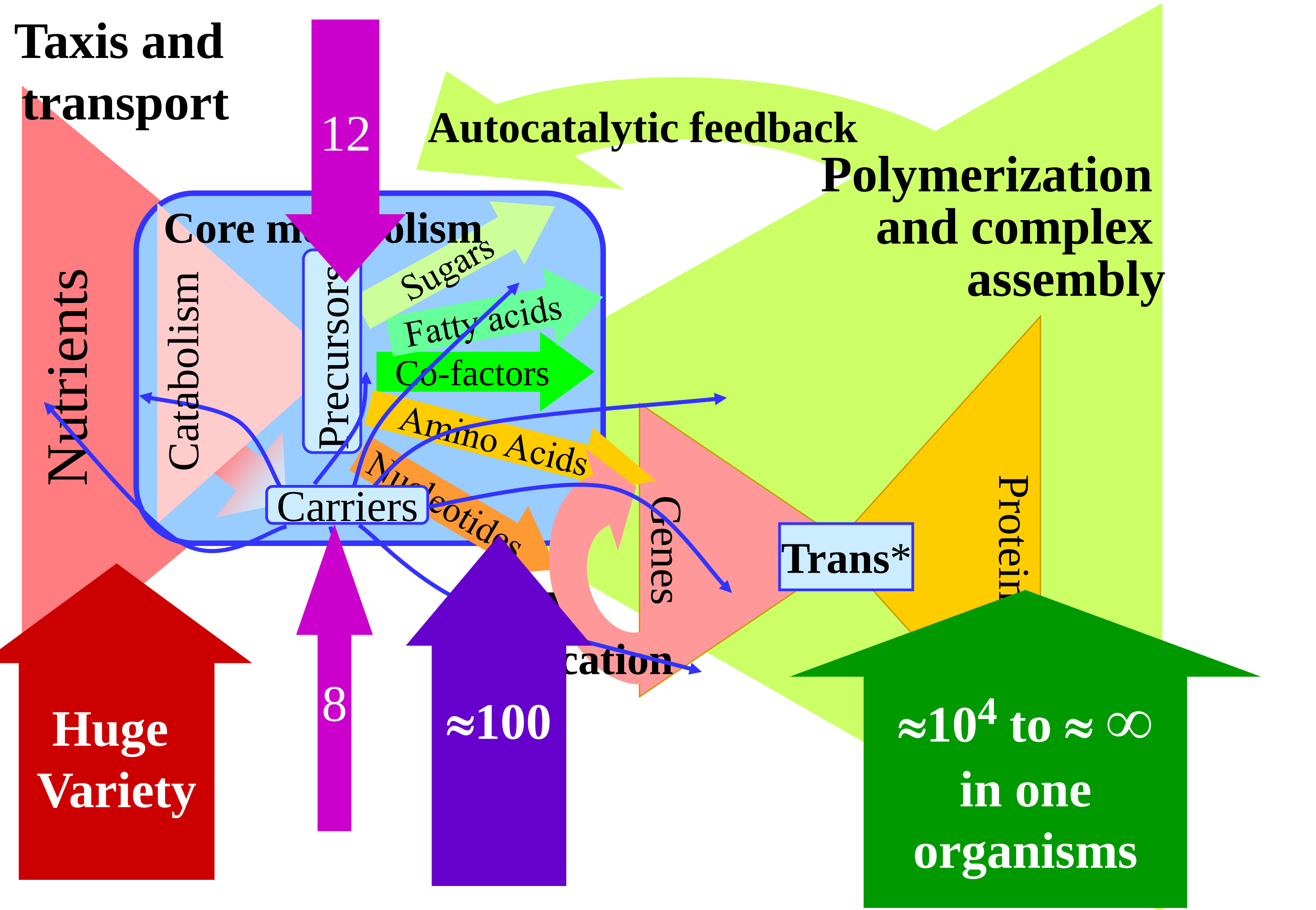
Inside every cell (≈10^{30})

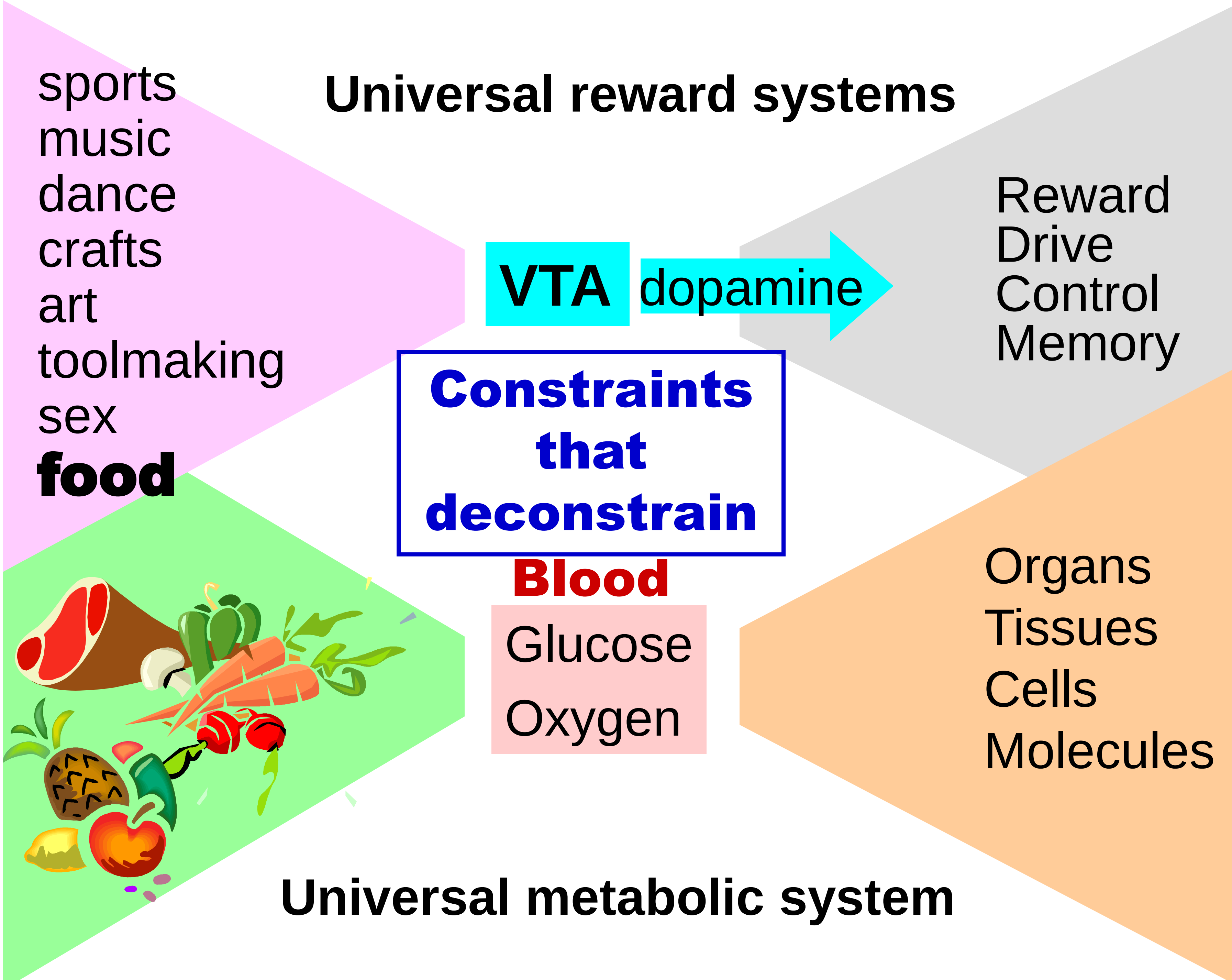
Metabolic pathways





Taxis and transport
Nutrients
Same 12 in all cells
Core metabolism
Catabolism
Precursors
Sugars
Amino Acids
Nucleotides
Fatty acids
Co-factors
Carriers
Same 8 in all cells
Huge Variety
≈100 ≈same in all organisms


Taxis and transport
12
Autocatalytic feedback
Polymerization and complex assembly
Core metabolism
Nutrients
Catabolism
Precursors
Sugars
Fatty acids
Co-factors
Amino Acids
Nucleotides
Carriers
Genes
Trans*
Proteins
cation
Huge Variety
8
≈100
≈10^4 to ≈ ∞ in one organisms


Universal reward systems
sports
music
dance
crafts
art
toolmaking
sex
food
VTA
dopamine
Reward
Drive
Control
Memory
Constraints that deconstrain
Blood
Glucose
Oxygen
Organs
Tissues
Cells
Molecules
Universal metabolic system

Modularity 2.0

Constraints

dopamine

Blood

Glucose

Oxygen

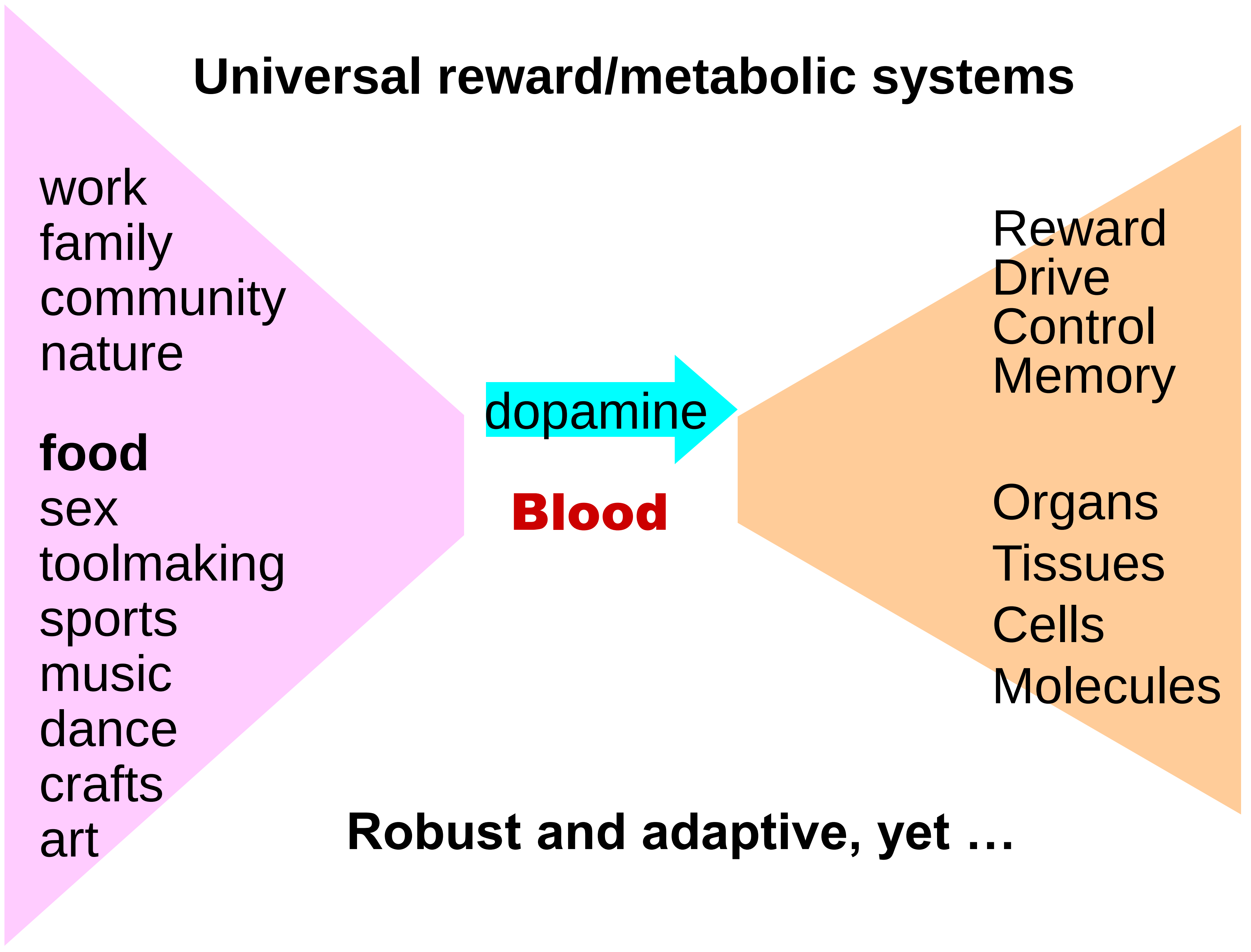
Modularity 2.0

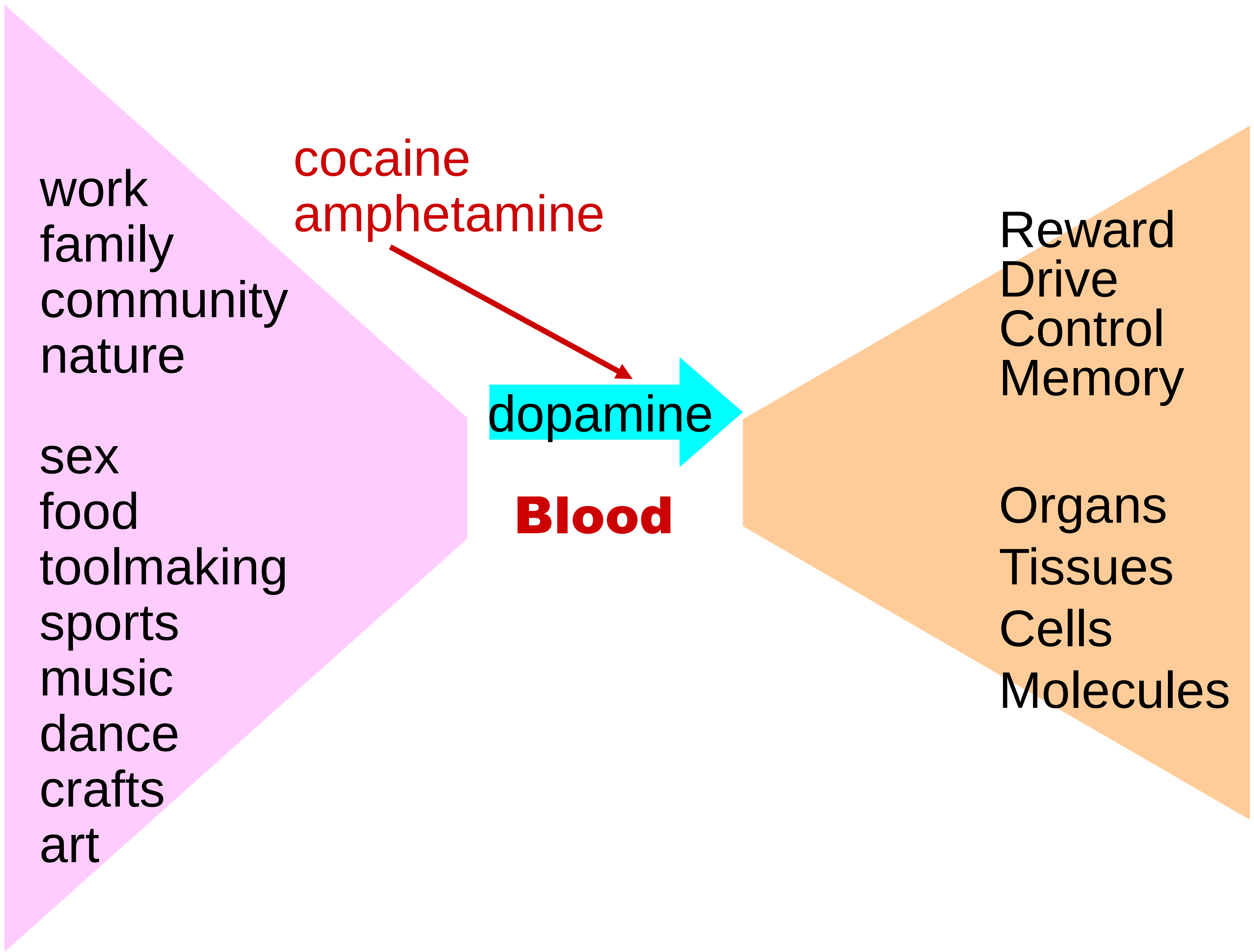
sports
music
dance
crafts
art
toolmaking
sex
food

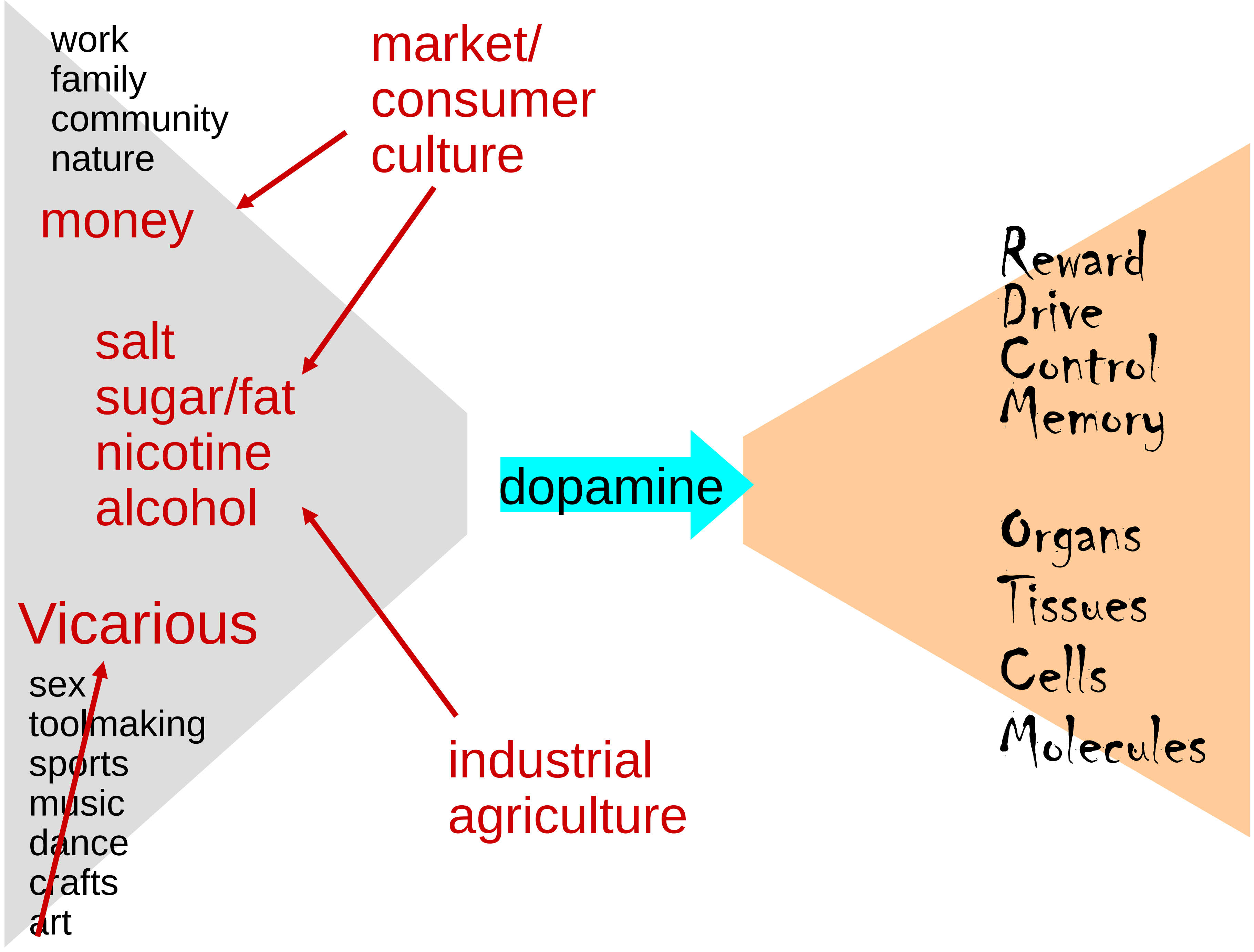
that deconstrain

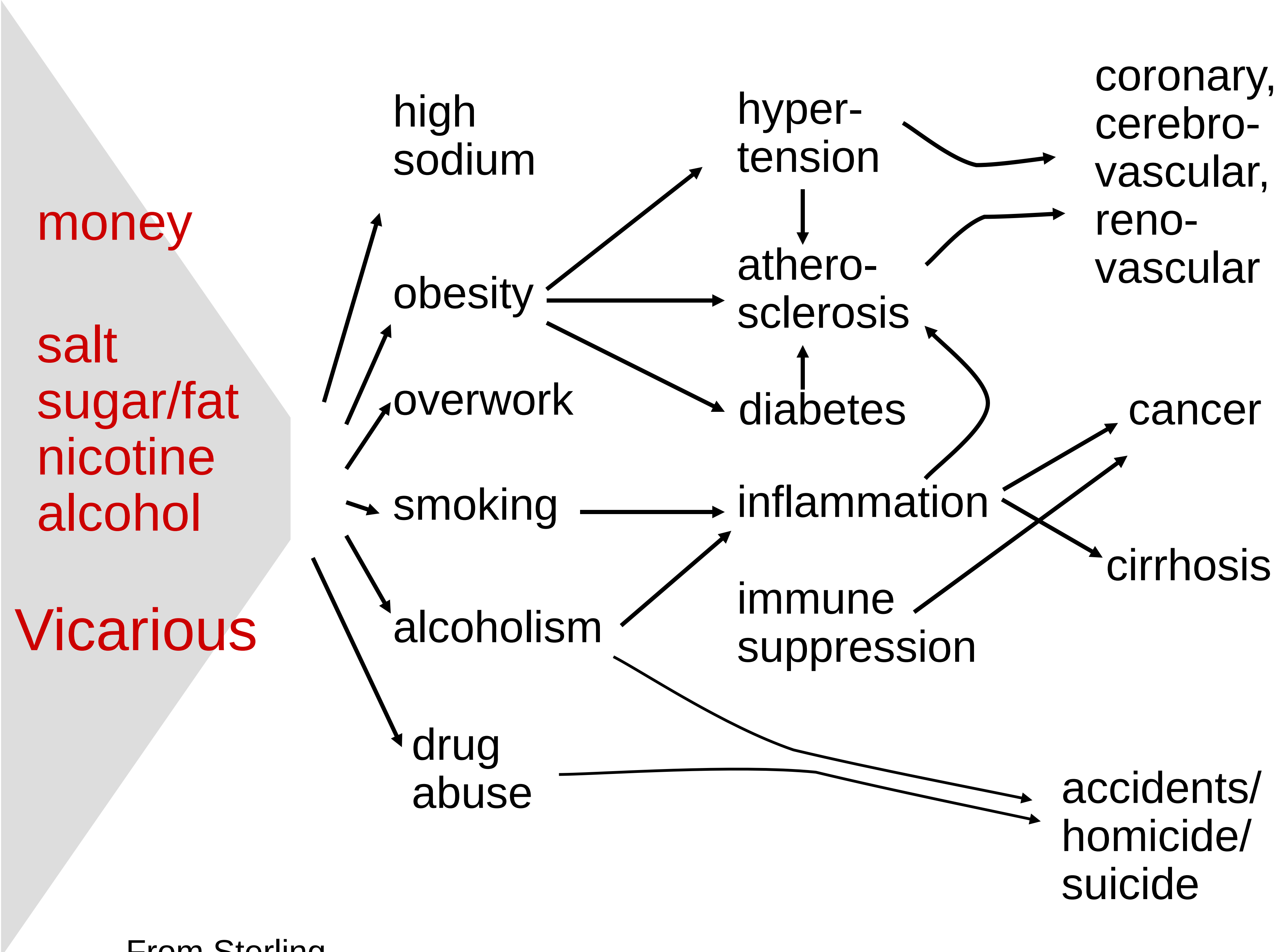
Reward
Drive
Control
Memory

Organs
Tissues
Cells
Molecules

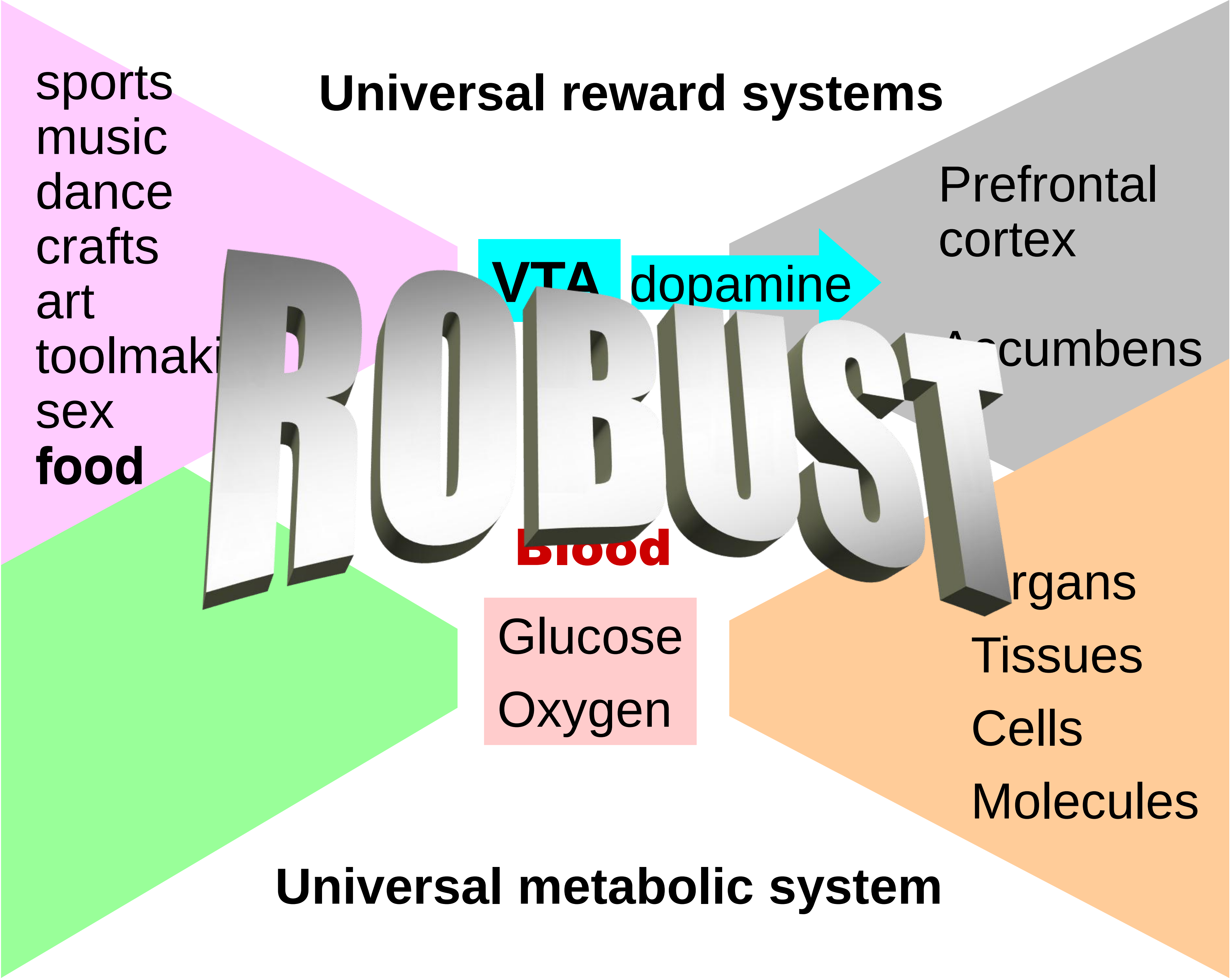

Universal reward/metabolic systems
work
family
community
nature
food
sex
toolmaking
sports
music
dance
crafts
art
dopamine
Blood
Reward
Drive
Control
Memory
Organs
Tissues
Cells
Molecules
Robust and adaptive, yet …

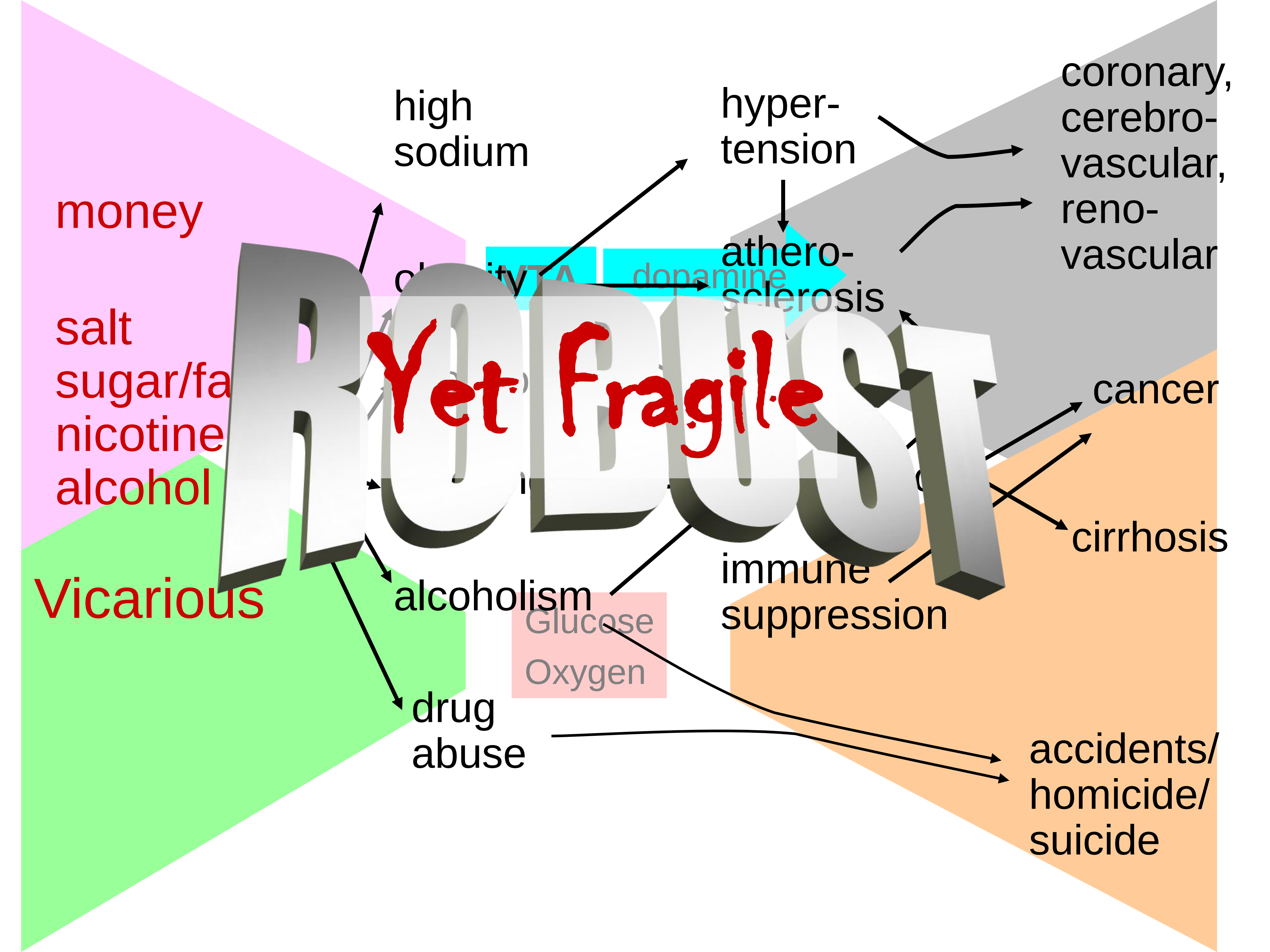

work
family
community
nature
sex
food
toolmaking
sports
music
dance
crafts
art
cocaine
amphetamine
dopamine
Blood
Reward
Drive
Control
Memory
Organs
Tissues
Cells
Molecules


work
family
community
nature
market/
consumer
culture
money
salt
sugar/fat
nicotine
alcohol
dopamine
Reward
Drive
Control
Memory
Organs
Tissues
Cells
Molecules
Vicarious
sex
toolmaking
sports
music
dance
crafts
art
industrial
agriculture



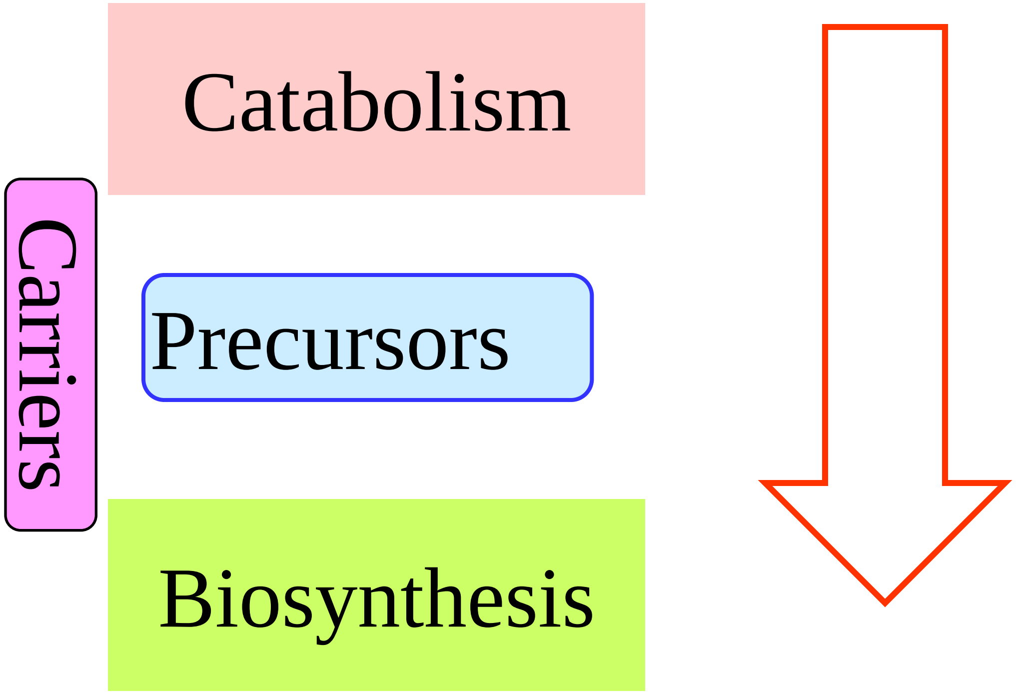
From Sterling


Universal reward systems
sports
music
dance
crafts
art
toolmaki
sex
food
VTA
dopamine
Prefrontal
cortex
ccumbens
ROBUST
Blood
Glucose
Oxygen
rgans
Tissues
Cells
Molecules
Universal metabolic system



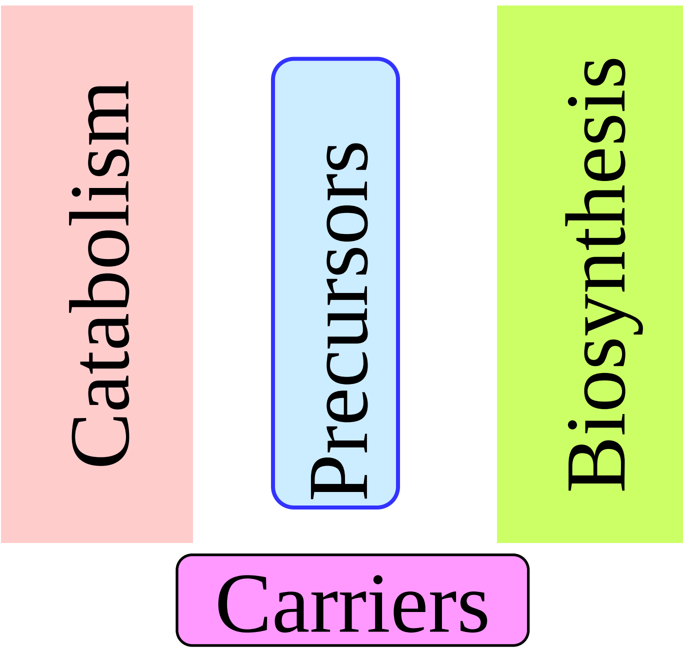
money
salt
sugar/fa
nicotine
alcohol
Vicarious
high
sodium
hyper-
tension
athero-
sclerosis
dopamine
coronary,
cerebro-
vascular,
reno-
vascular
cancer
cirrhosis
immune
suppression
alcoholism
Glucose
Oxygen
drug
abuse
accidents/
homicide/
suicide
ROBUST
Yet Fragile

Inside every cell

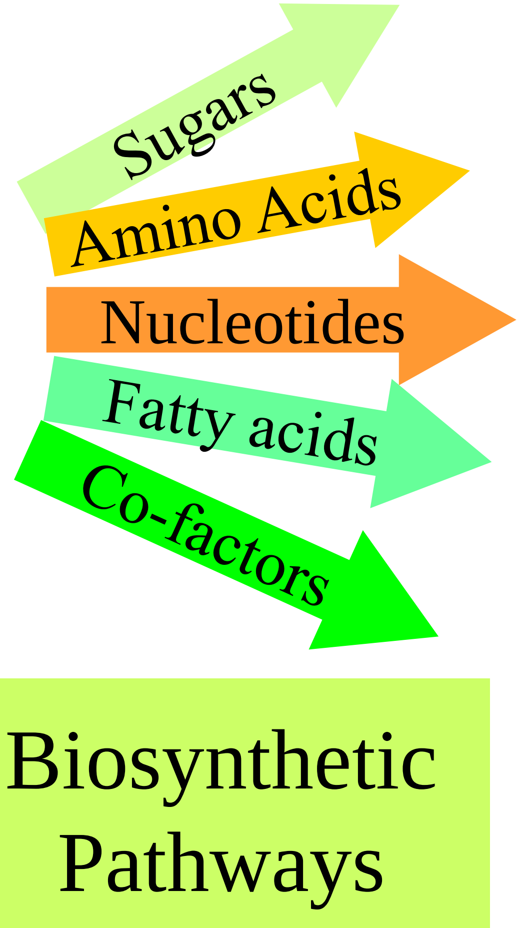


Layered architecture

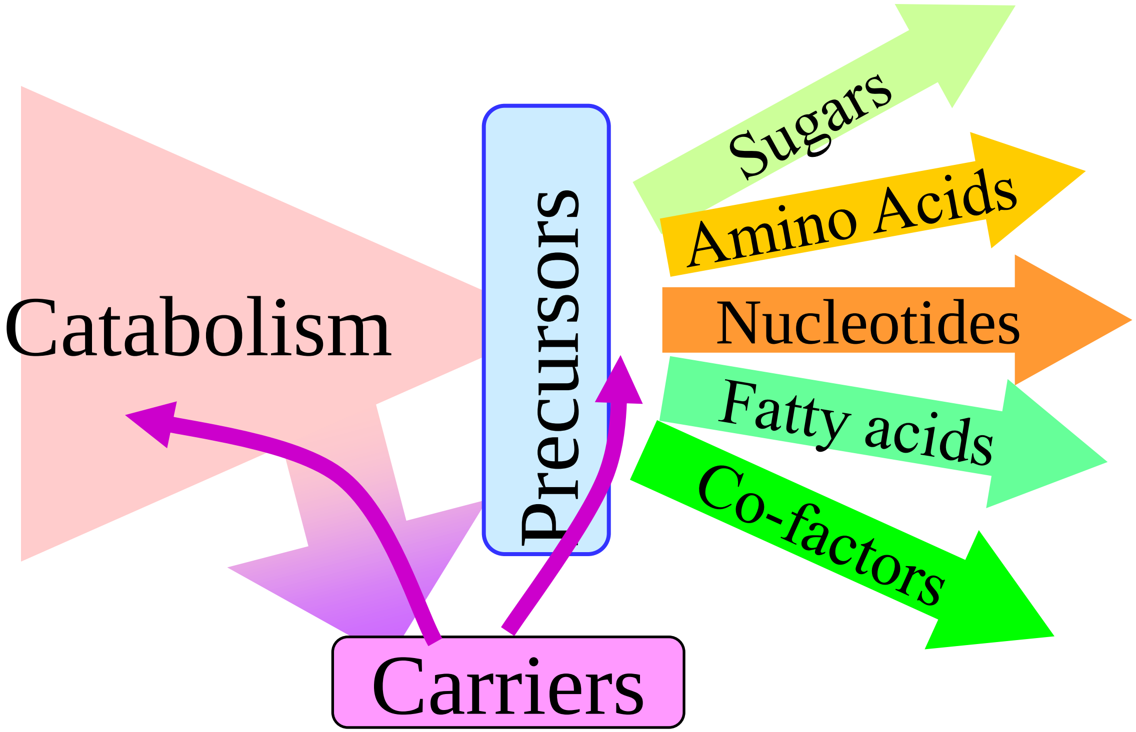
Inside every cell



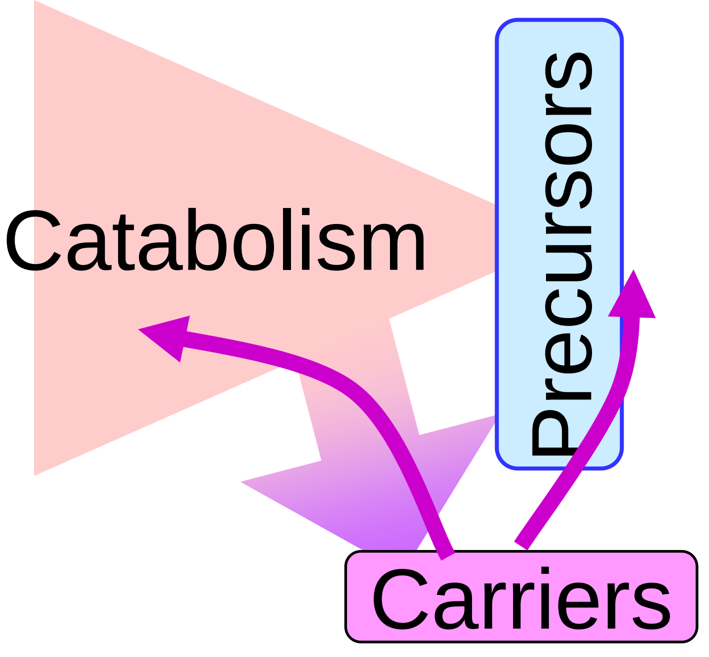
Layered architecture

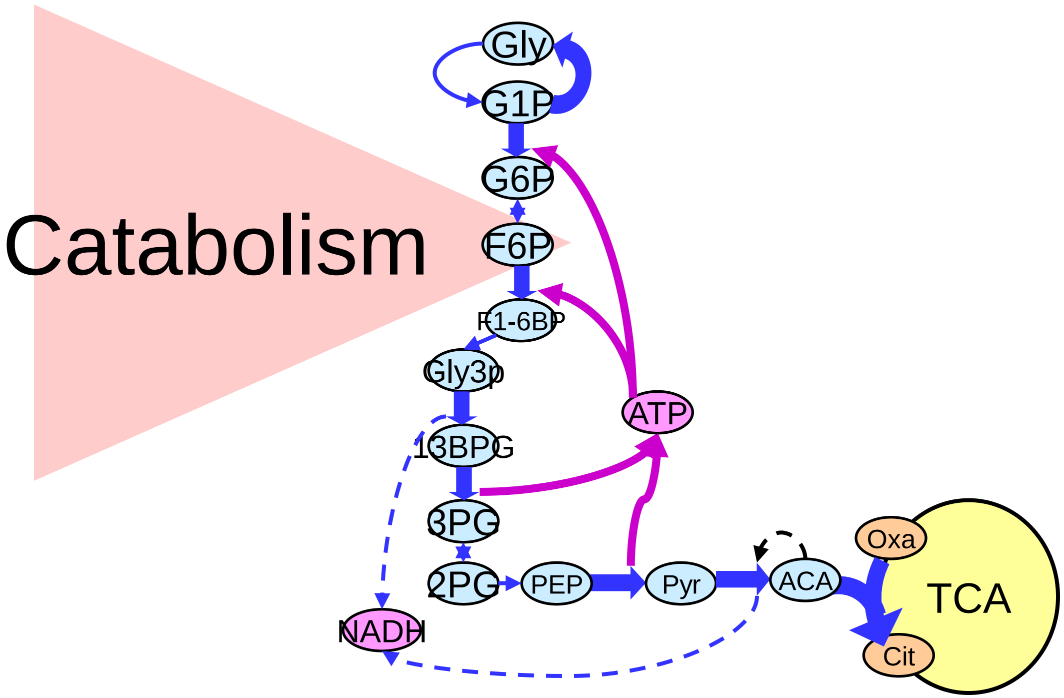


Inside every cell

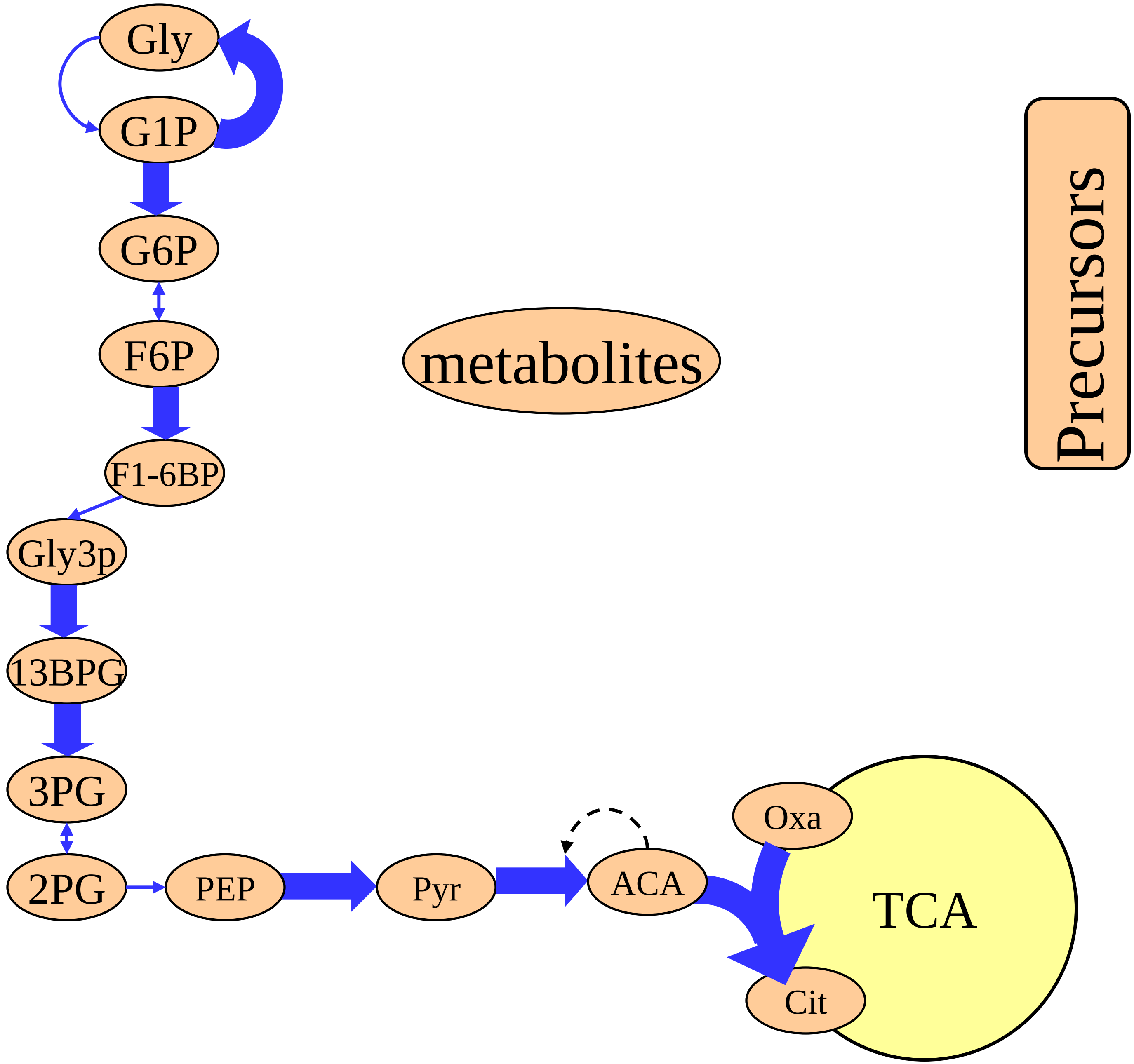


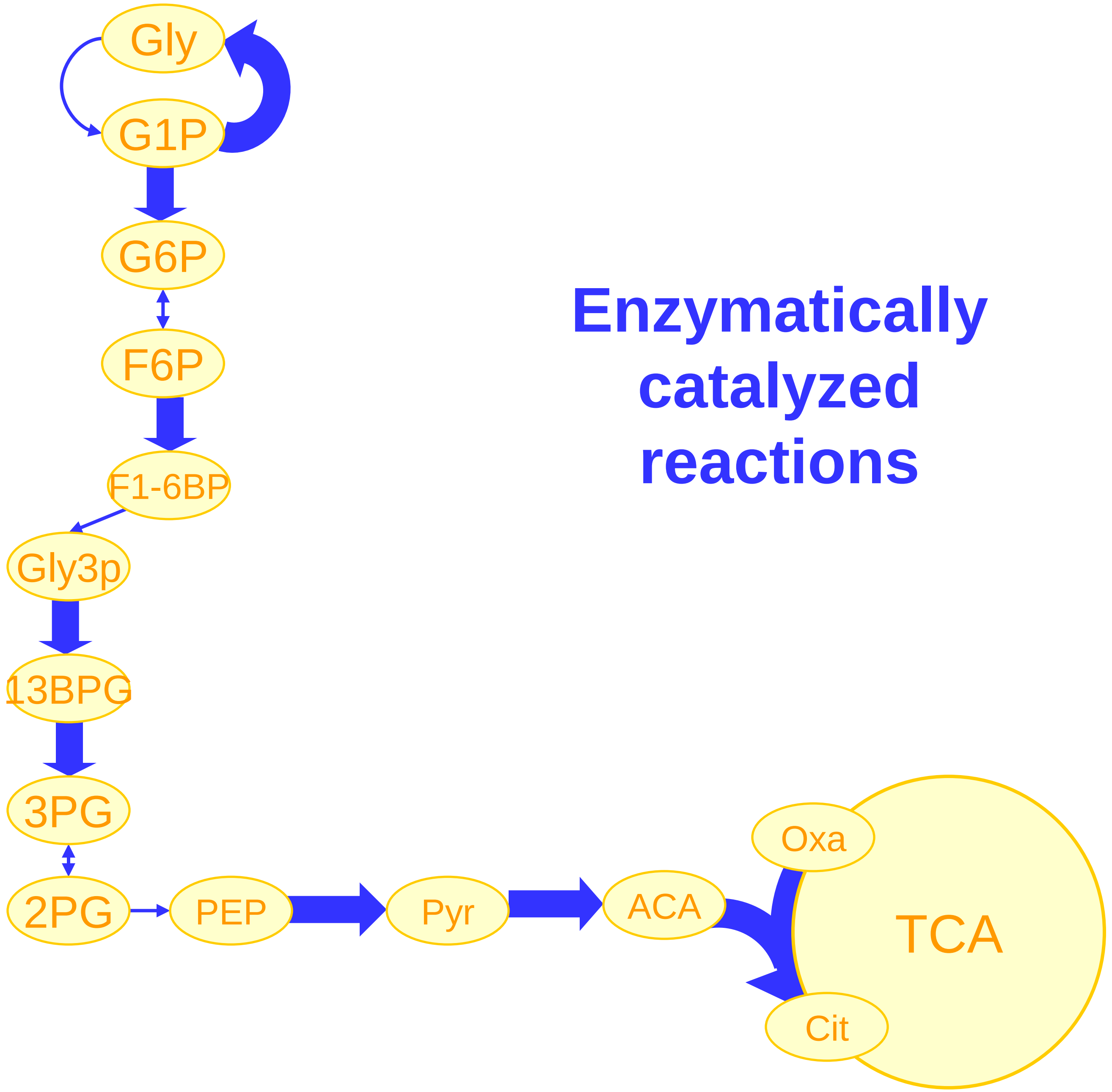
Core metabolic bowtie
Layered architecture


Catabolism
Precursors
Carriers



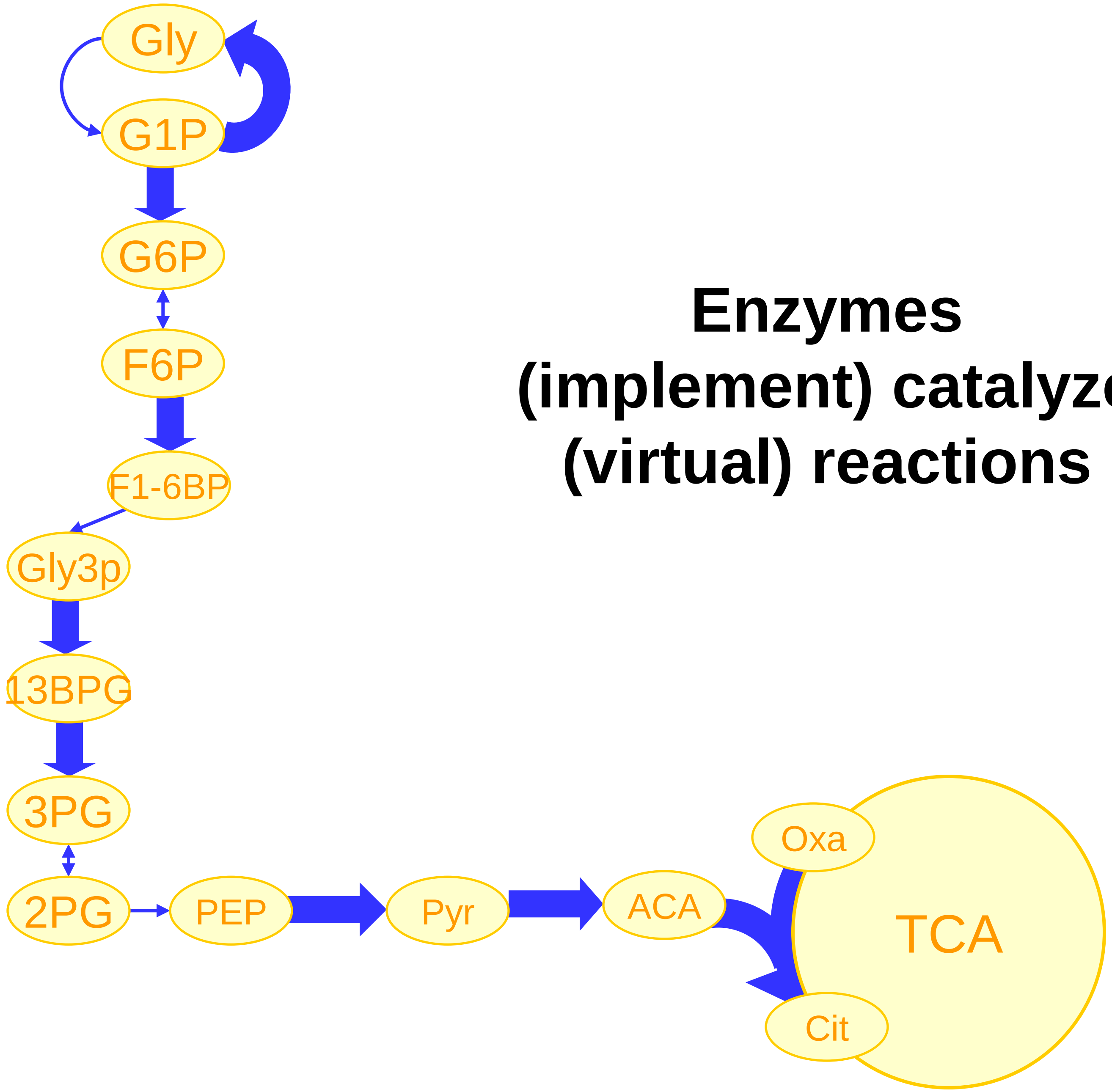
Catabolism
Gly
G1P
G6P
F6P
F1-6BP
Gly3p
13BPG
3PG
2PG
PEP
Pyr
ACA
ATP
NADH
Oxa
Cit
TCA

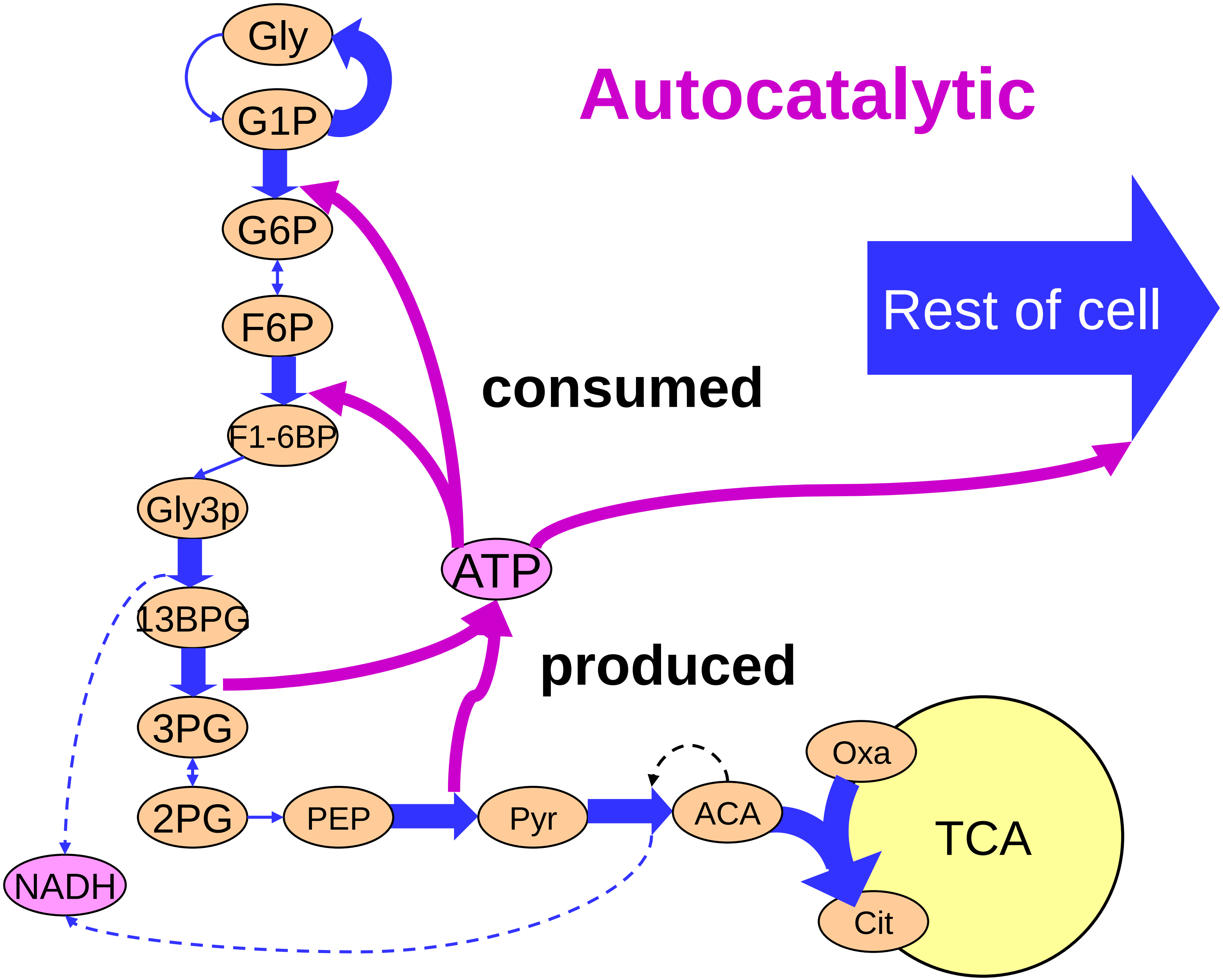

Gly
G1P
G6P
F6P
F1-6BP
Gly3p
13BPG
3PG
2PG
PEP
Pyr
ACA
Oxa
TCA
Cit
metabolites
Precursors



Enzymatically catalyzed reactions

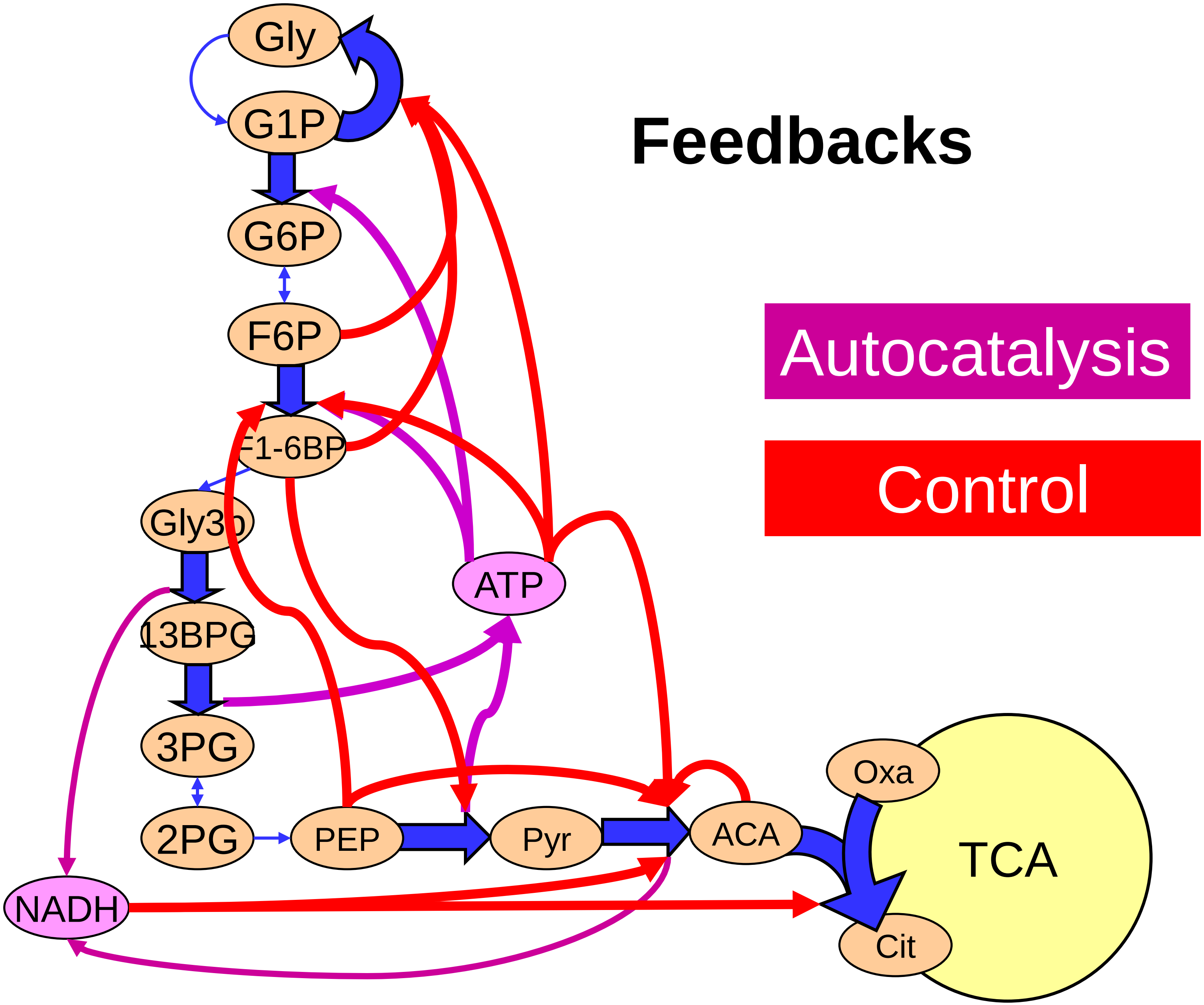
Enzymes (implement) catalyze (virtual) reactions



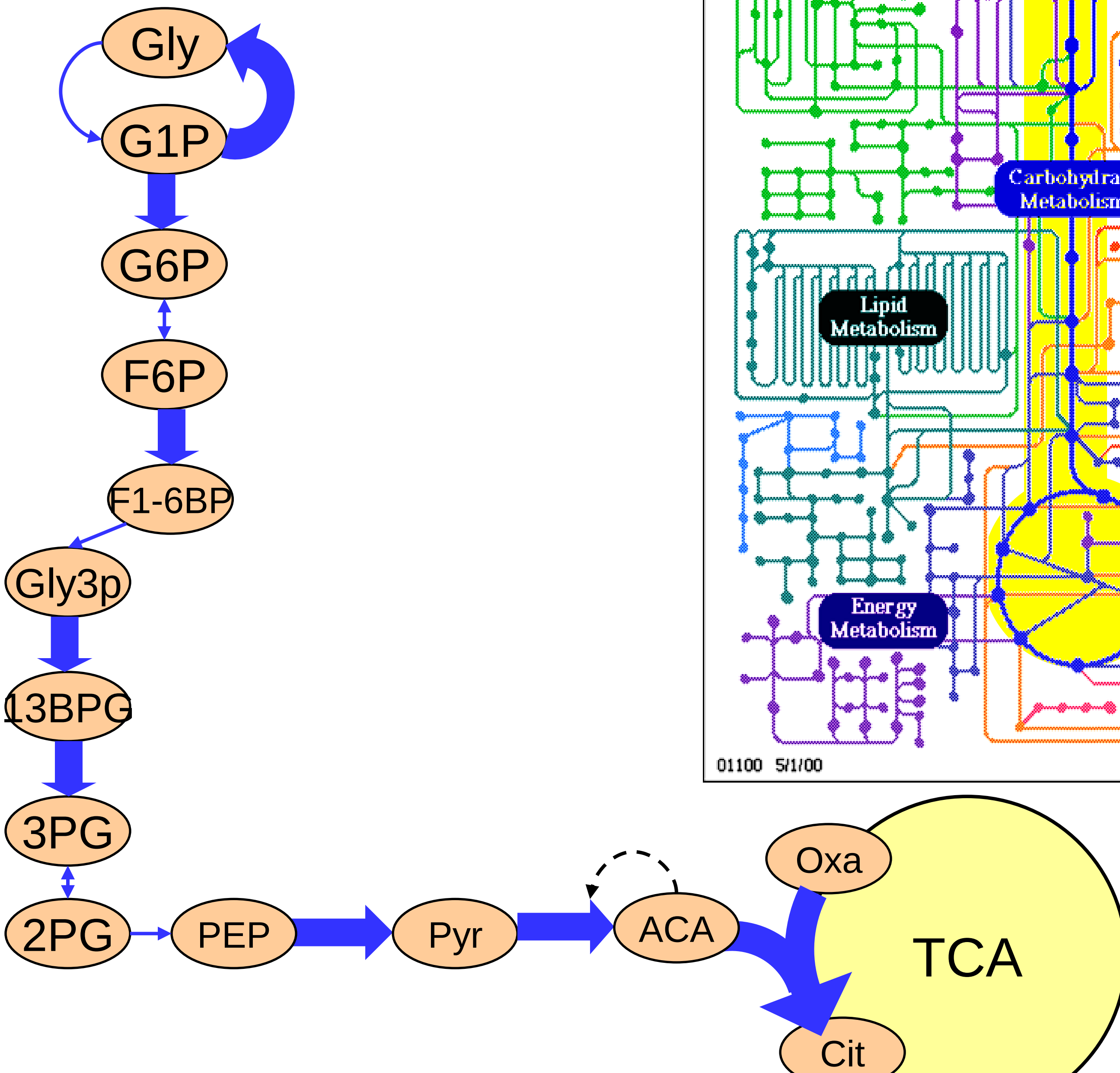


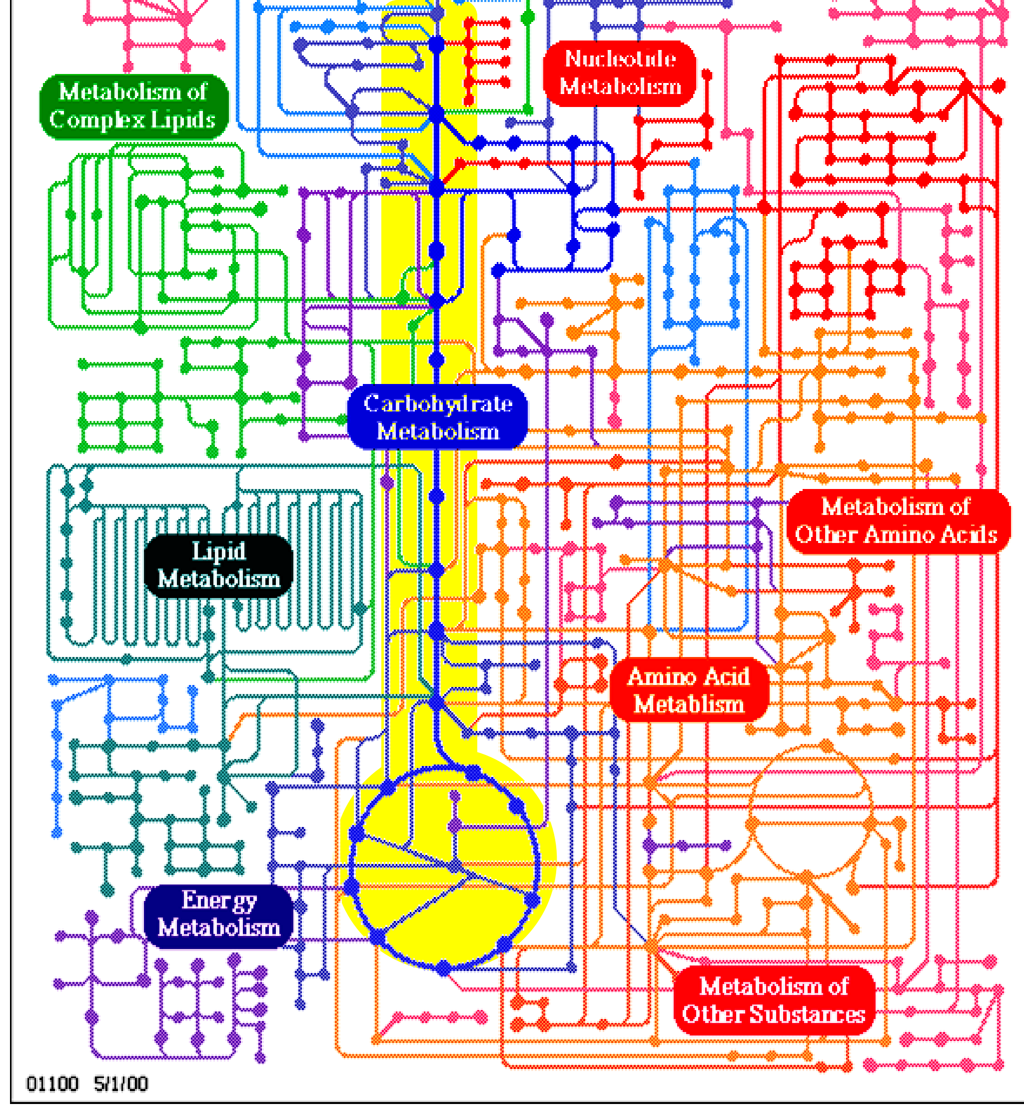
Autocatalytic
Gly
G1P
G6P
F6P
F1-6BP
Gly3p
13BPG
3PG
2PG
PEP
Pyr
ACA
Oxa
Cit
TCA
NADH
ATP
consumed
produced
Rest of cell

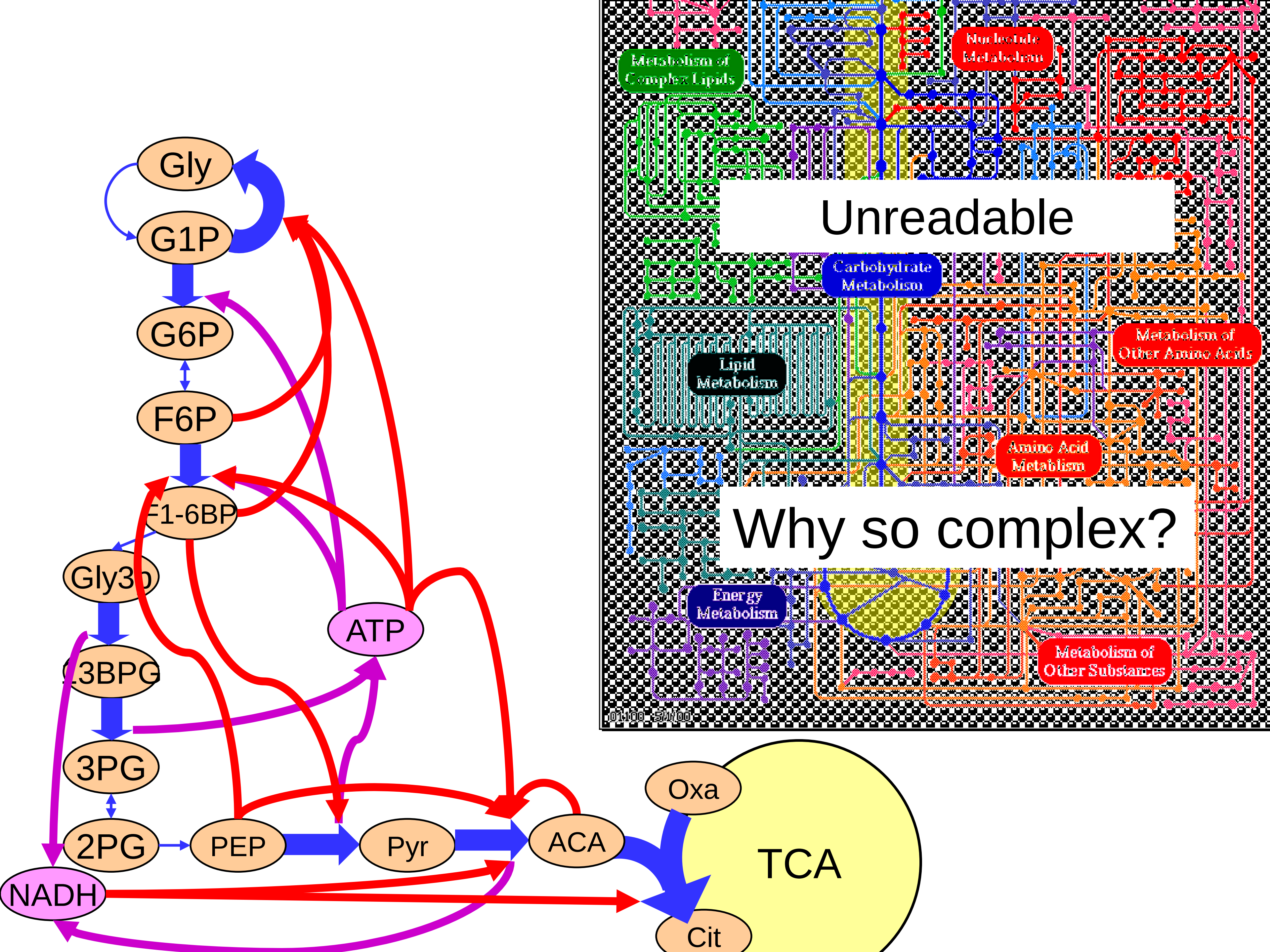
Feedbacks



Gly
G1P
G6P
F6P
F1-6BP
Gly3p
13BPG
3PG
2PG
PEP
Pyr
ACA
ATP
NADH
Oxa
Cit
TCA
Autocatalysis
Control

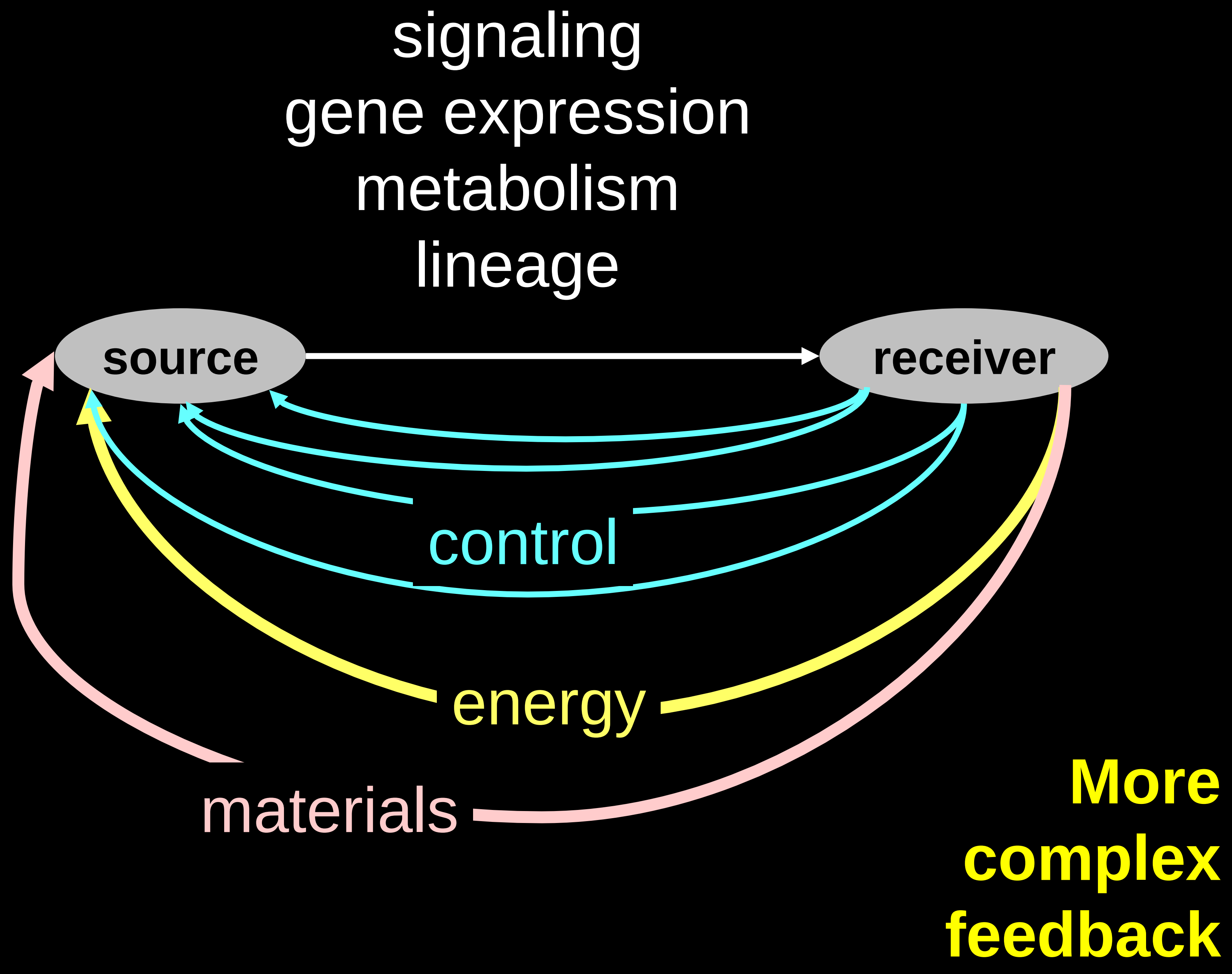

Gly
G1P
G6P
F6P
F1-6BP
Gly3p
13BPG
3PG
2PG
PEP
Pyr
ACA
Oxa
TCA
Cit


Metabolism of Complex Lipids
Nucleotide Metabolism
Carbohydrate Metabolism
Lipid Metabolism
Metabolism of Other Amino Acids
Amino Acid Metablism
Energy Metabolism
Metabolism of Other Substances
01100 5/1/00

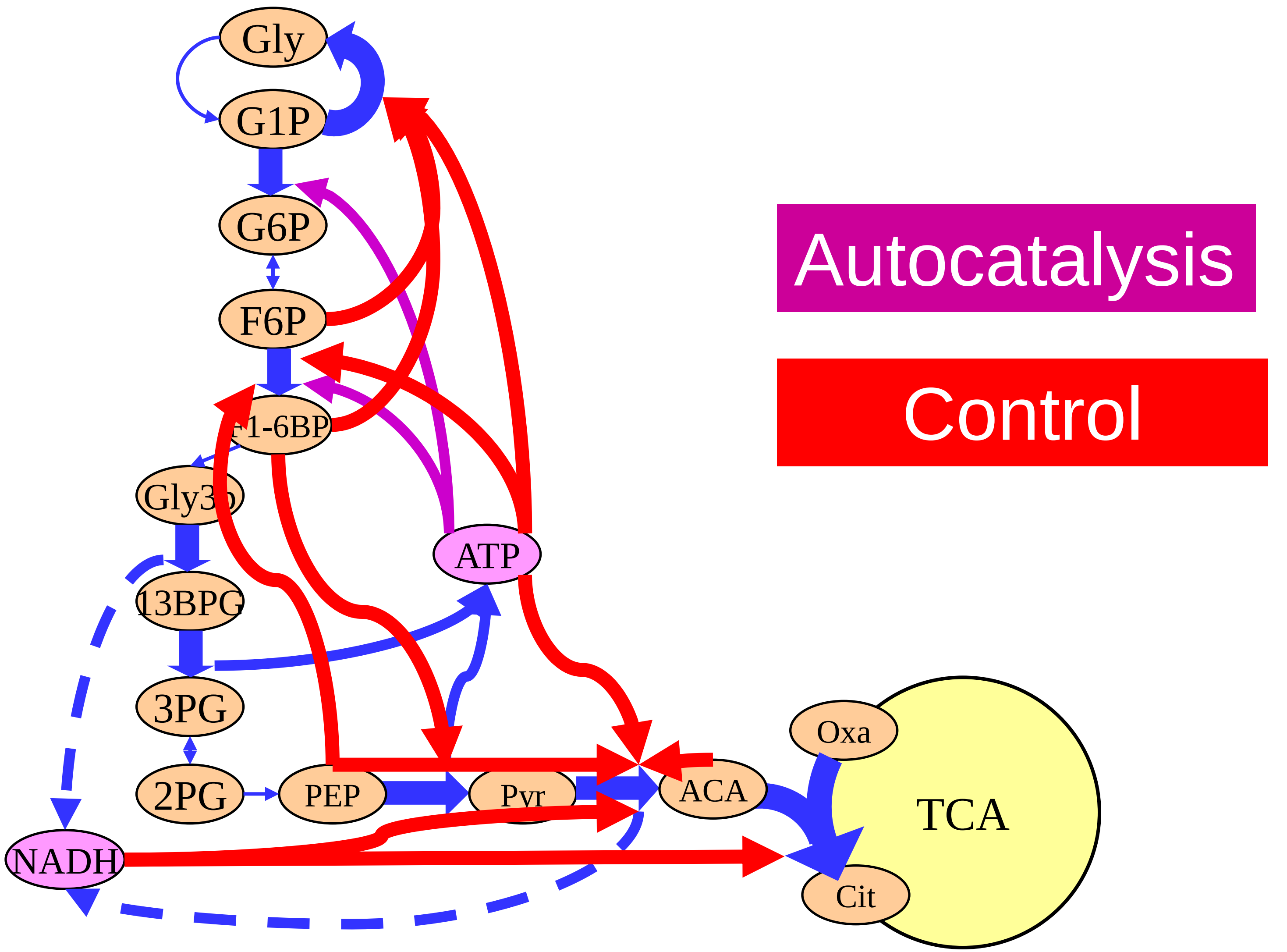

Unreadable
Why so complex?
Metabolism of Complex Lipids
Nucleotide Metabolism
Carbohydrate Metabolism
Metabolism of Other Amino Acids
Lipid Metabolism
Amino Acid Metablism
Energy Metabolism
Metabolism of Other Substances
Gly
G1P
G6P
F6P
F1-6BP
Gly3p
13BPG
3PG
2PG
PEP
Pyr
ACA
ATP
NADH
Oxa
TCA
Cit

signaling
gene expression
metabolism
lineage

Biological pathways

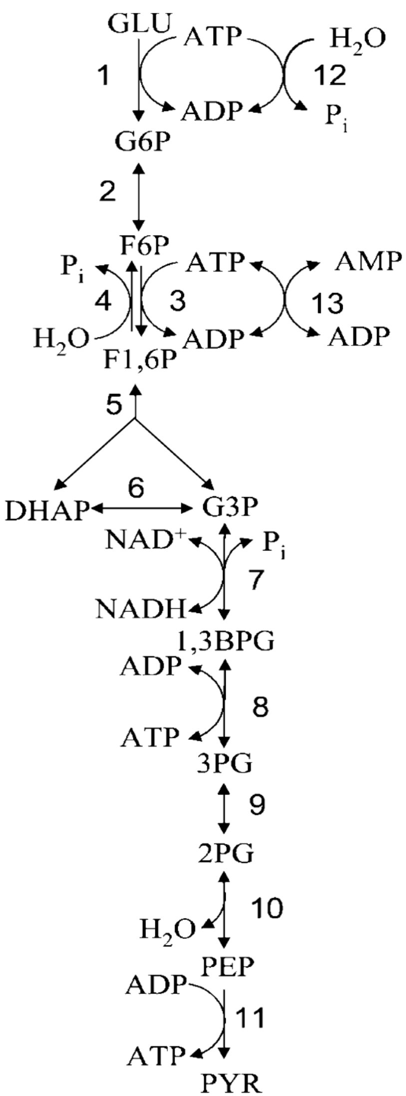

signaling
gene expression
metabolism
lineage
source
receiver
control
energy
materials
More
complex
feedback

Feedbacks



Gly
G1P
G6P
F6P
F1-6BP
Gly3p
13BPG
3PG
2PG
PEP
Pyr
ATP
NADH
ACA
Oxa
TCA
Cit
Autocatalysis
Control

Nikolaev, …, Elucidation and Structural Analysis of **Conserved Pools** for Genome-Scale Metabolic Reconstructions, Biophysical Journal, Volume 88, Issue 1, January 2005, Pages 37-49



(P1) Total NAD moiety: $[NAD^+] + [NADH]$

(P2) Total adenylate moiety: $[ATP] + [ADP] + [AMP]$

(P3) Total carbon moiety: $2[GLU] + 2[G6P] + 2[F6P] + 2[F1,6P] + [DHAP] +$

(P4) Total phosphate moiet

(P5) Total oxygen moiety:

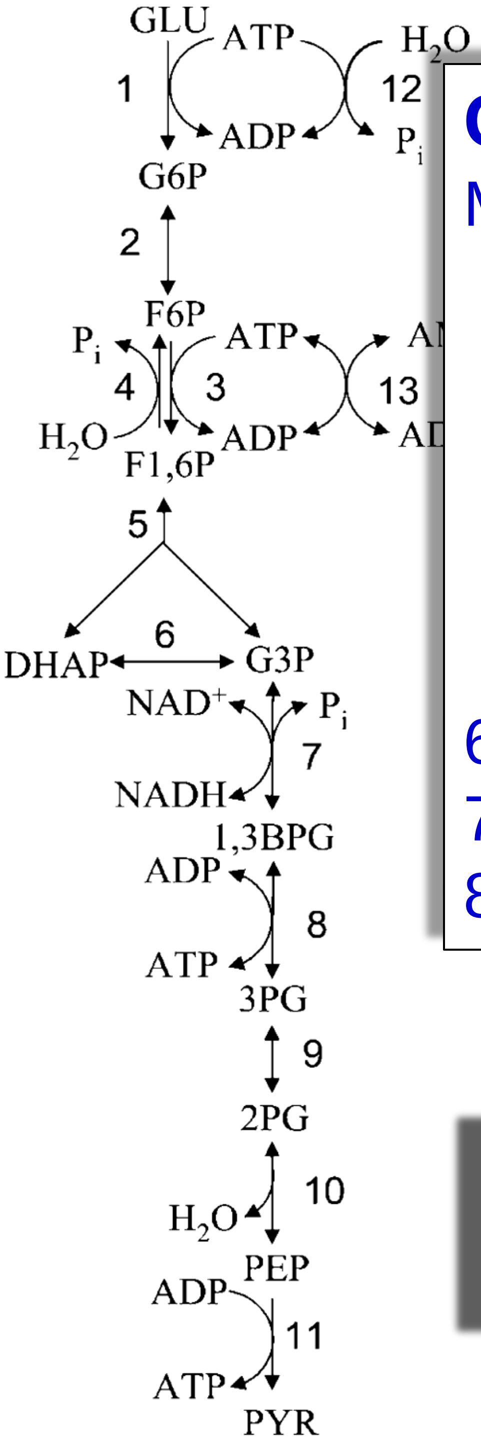
(P6) Oxidized state of meta

(P7) Reduced state of metabolites: $2[GLU] + 2[G6P] + 2[F6P] + 2[F1,6P] +$ $+ [DHAP] + [G3P] + [NADH]$

(P8) High energy potential release: $2[GLU] + [G6P] + [F6P] + [3PG] + [2PG] +$ $+ [PYR] + [ADP] + 2[AMP] + [H_2O]$

Constrained (conserved):

1. Total NAD moiety
2. Total Adenylate moiety
3. Total Carbon moiety
4. Total phosphate moiety
5. Total oxygen moiety
6. Oxidized state of metabolites
7. Reduced state of metabolites
8. High energy potential release



(P1) Total NAD moiety: $[NAD^+] + [NADH]$

Constrained ("conserved"):

Moieties

1. NAD
2. Adenylate
3. Carbon
4. phosphate
5. oxygen

6. Oxidized state of metabolites
7. Reduced state of metabolites
8. High energy potential release

2[F1,6P] + [DHAP] +

[2PG] + [PEP] + [PYR]

HAP] + [G3P] +

+ [PEP] +

DP] + $[P_i]$

P_i] + $[H_2O]$

] +

+ [PEP] + [PYR] + $[NAD^+]$

(P7) Reduced state of metabolites: 2[GLU] + 2[G6P] + 2[F6P] + 2[F1,6P] +

$$\frac{1}{\pi}\int_0^\infty \ln|S(j\omega)|\left(\frac{z}{z^2+\omega^2}\right)d\omega \geq \ln\left|\frac{z+p}{z-p}\right|$$

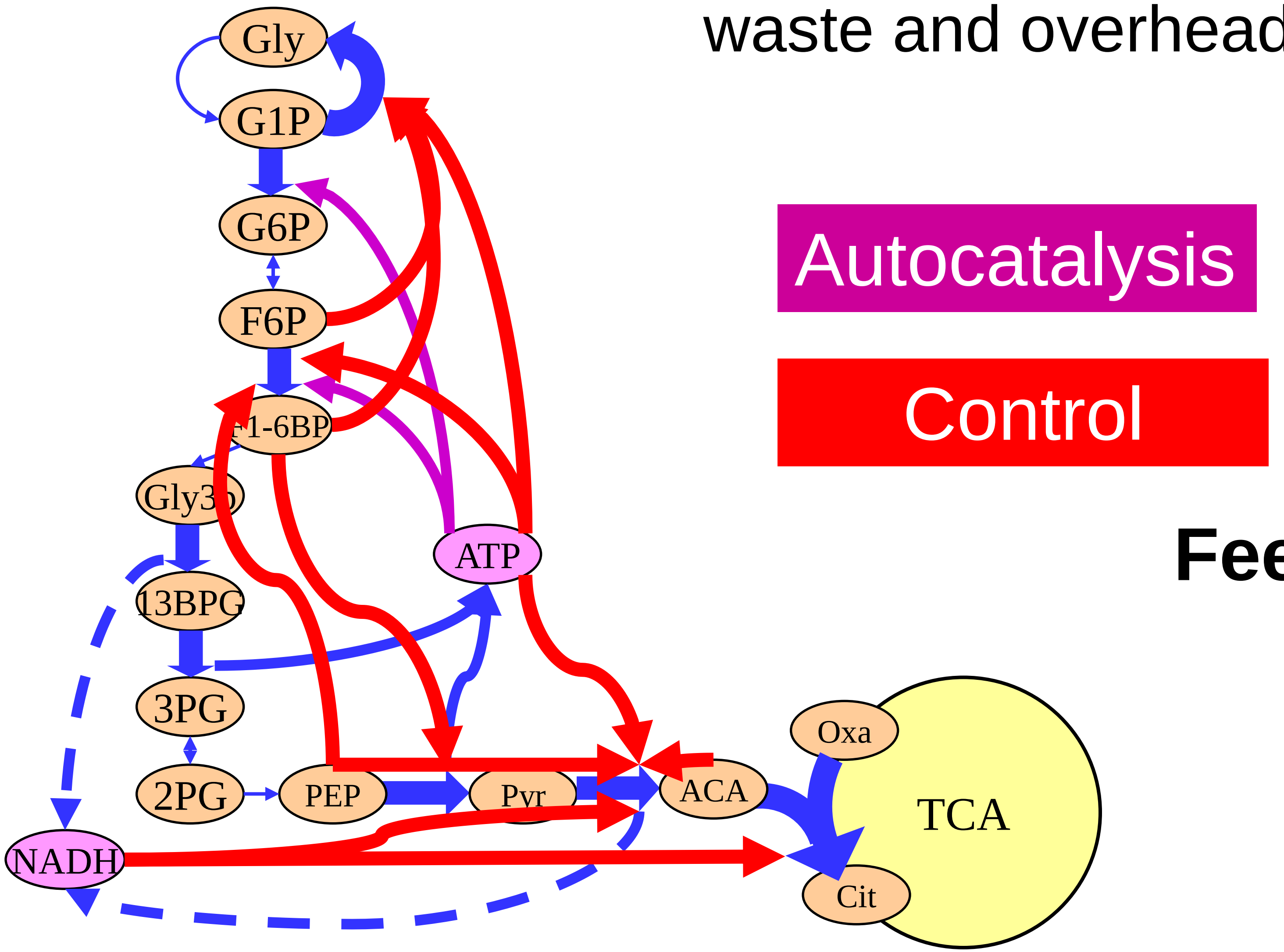
ADH]

[2PG] +

+ [PYR] + [ADP] + 2[AMP] + $[H_2O]$

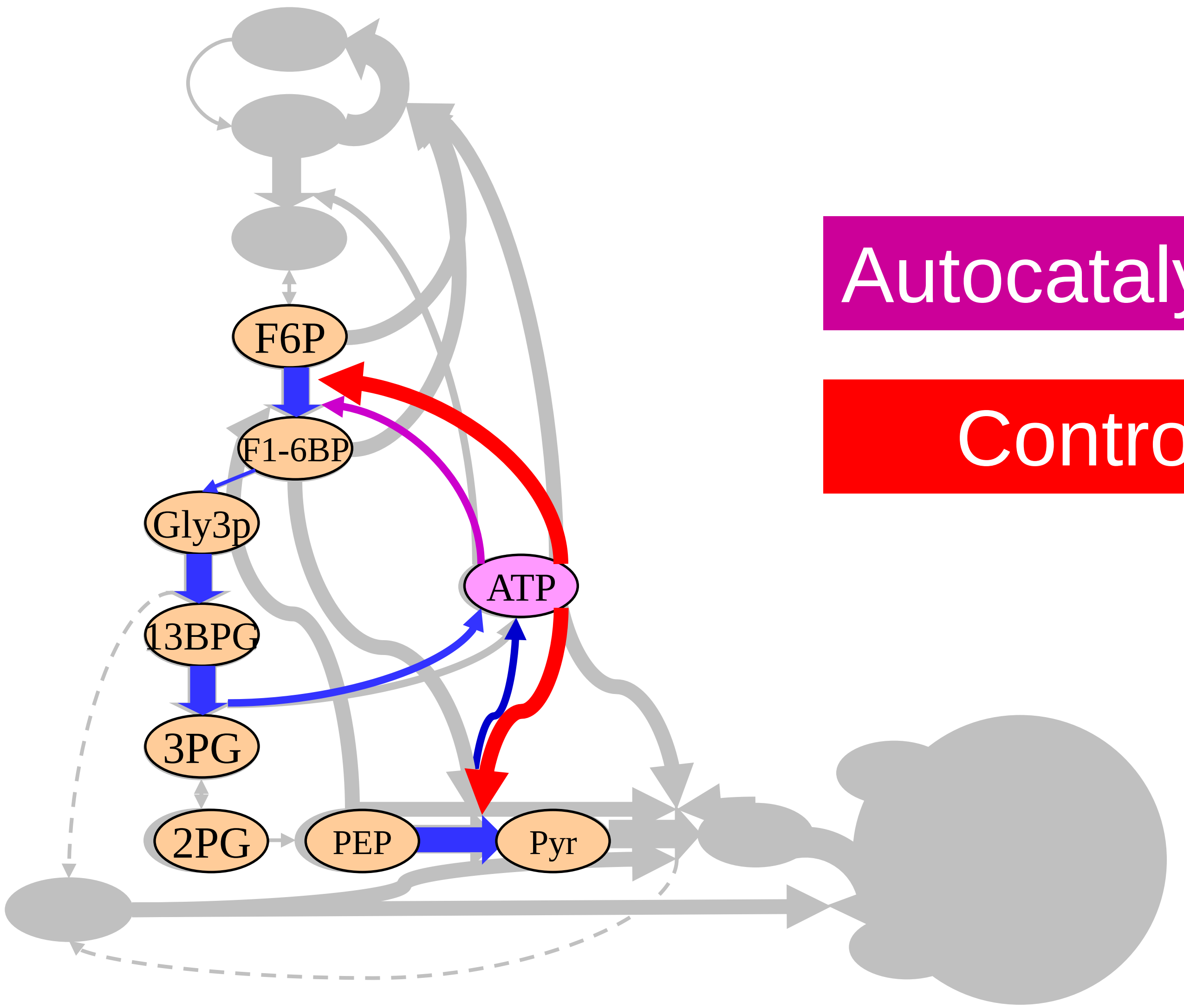
Robust=maintain energy charge w/fluctuating cell demand

Efficient=minimize metabolic waste and overhead



Feedbacks

Minimal model?

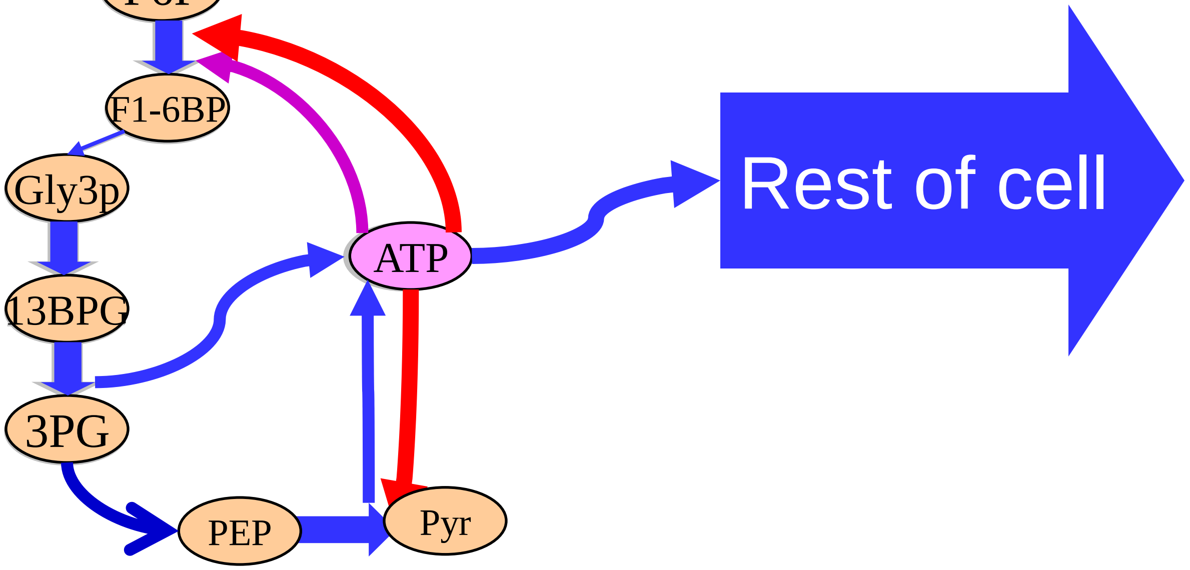


Autocatalysis

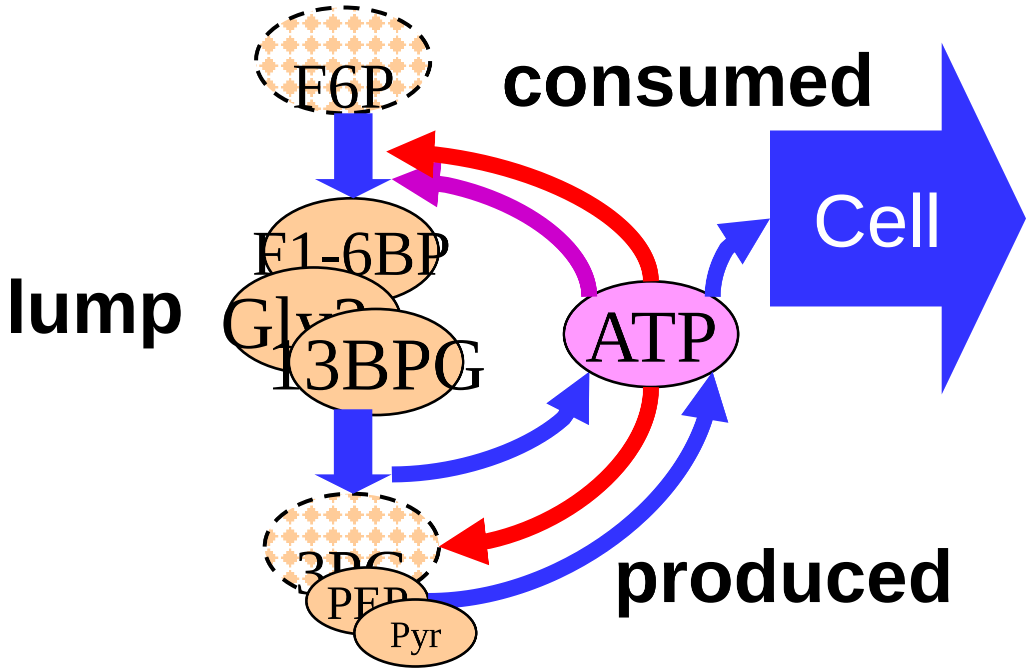
Control

Minimal model
~1 equilibrium
2 metabolites
3 “reactions”

Control
Plus
Autocatalytic
Feedback



Minimal model
~1 equilibrium
2 metabolites
3 “reactions”



UG biochem, math, control theory

Glycolytic Oscillations and Limits on Robust Efficiency

Fiona A. Chandra,[1]* Gentian Buzi,[2] John C. Doyle[2]

Both engineering and evolution are constrained by trade-offs between efficiency and robustness, but theory that formalizes this fact is limited. For a simple two-state model of glycolysis, we explicitly derive analytic equations for hard trade-offs between robustness and efficiency with oscillations as an inevitable side effect. The model describes how the trade-offs arise from individual parameters, including the interplay of feedback control with autocatalysis of network products necessary to power and catalyze intermediate reactions. We then use control theory to prove that the essential features of these hard trade-off "laws" are universal and fundamental, in that they depend minimally on the details of this system and generalize to the robust efficiency of any autocatalytic network. The theory also suggests worst-case conditions that are consistent with initial experiments.

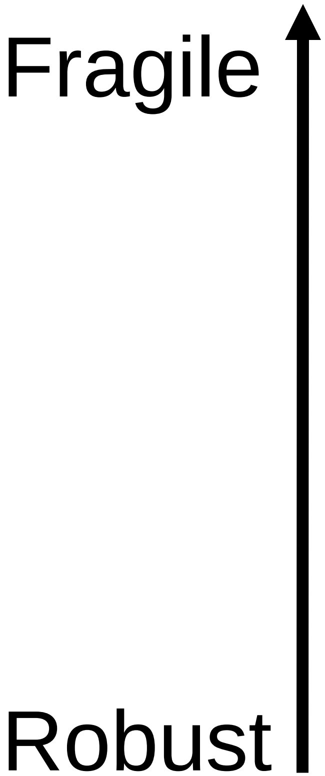
the cell's use of ATP. In glycolysis, two ATP molecules are consumed upstream and four are produced downstream, which normalizes to $q = 1$ (each y molecule produces two downstream) with kinetic exponent $a = 1$. To highlight essential trade-offs with the simplest possible analysis, we normalize the concentration such that the unperturbed ($\delta = 0$) steady states are $\bar{y} = 1$ and $\bar{x} = 1/k$ [the system can have one additional steady state, which is unstable when $(1, 1/k)$ is stable]. [See the supporting online material (SOM) part I]. The basal rate of the PFK reaction and the consumption rate have been normalized to 1 (the 2 in the numerator and feedback coefficients of the reactions come from these normalizations). Our results hold for more general systems as discussed below and in SOM, but the analysis

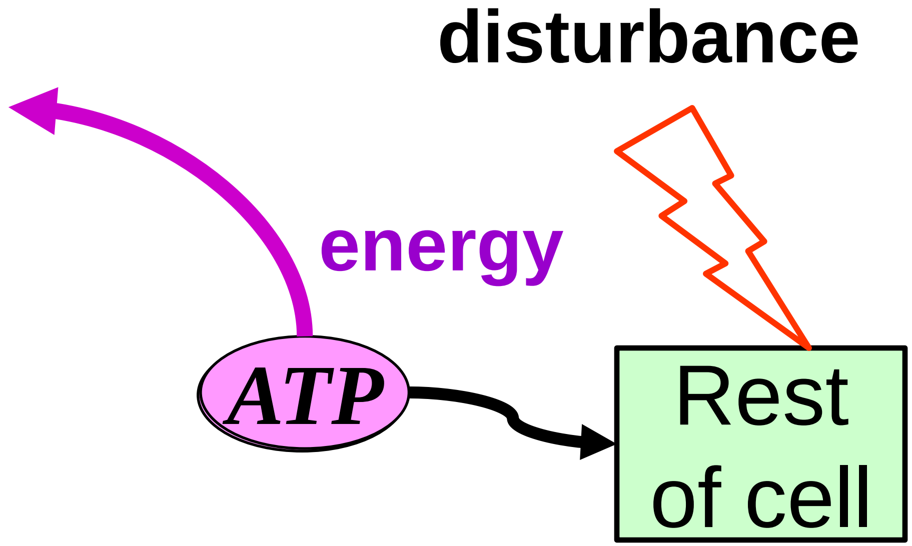
Chandra, Buzi, and Doyle

Most important paper so far.

www.sciencemag.org SCIENCE VOL 333 8 JULY 2011

(May 21): Hard tradeoff in glycolysis



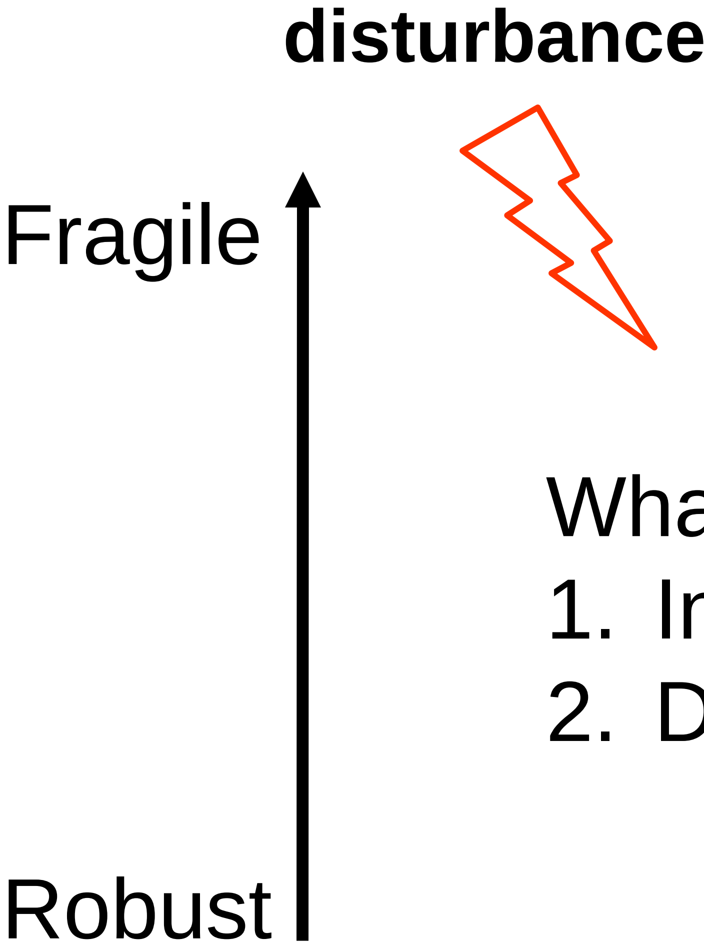


Robust =

Maintain energy

(ATP concentration)

despite demand fluctuation



Accurate vs sloppy

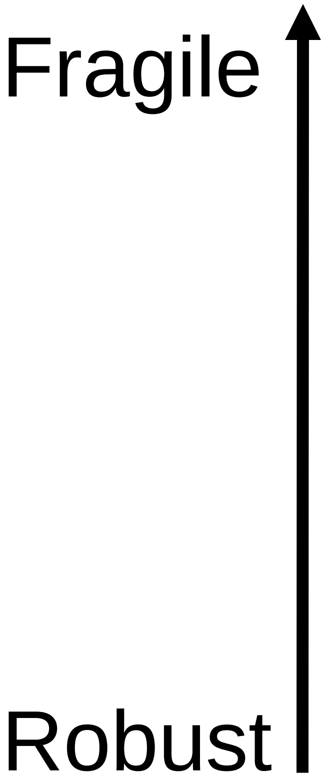
What makes this hard?

1. Instability (autocatalysis)
2. Delay (enzyme amount)

Robust

≈Disturbance rejection

≈ Accurate



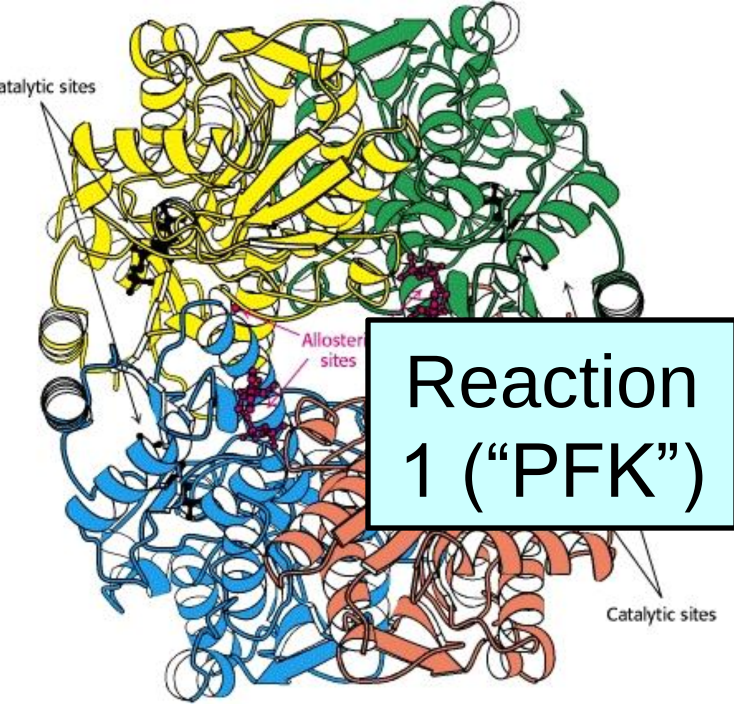
What makes this hard?

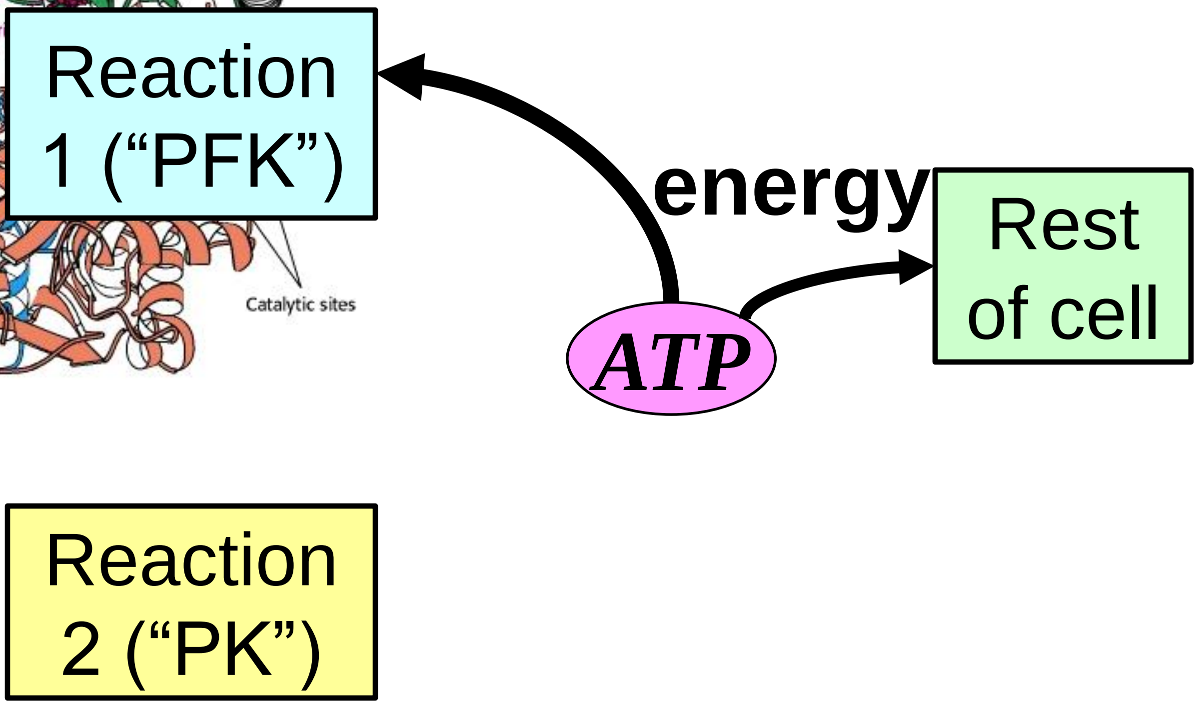
1. Instability
2. Delay

The CNS must cope with both

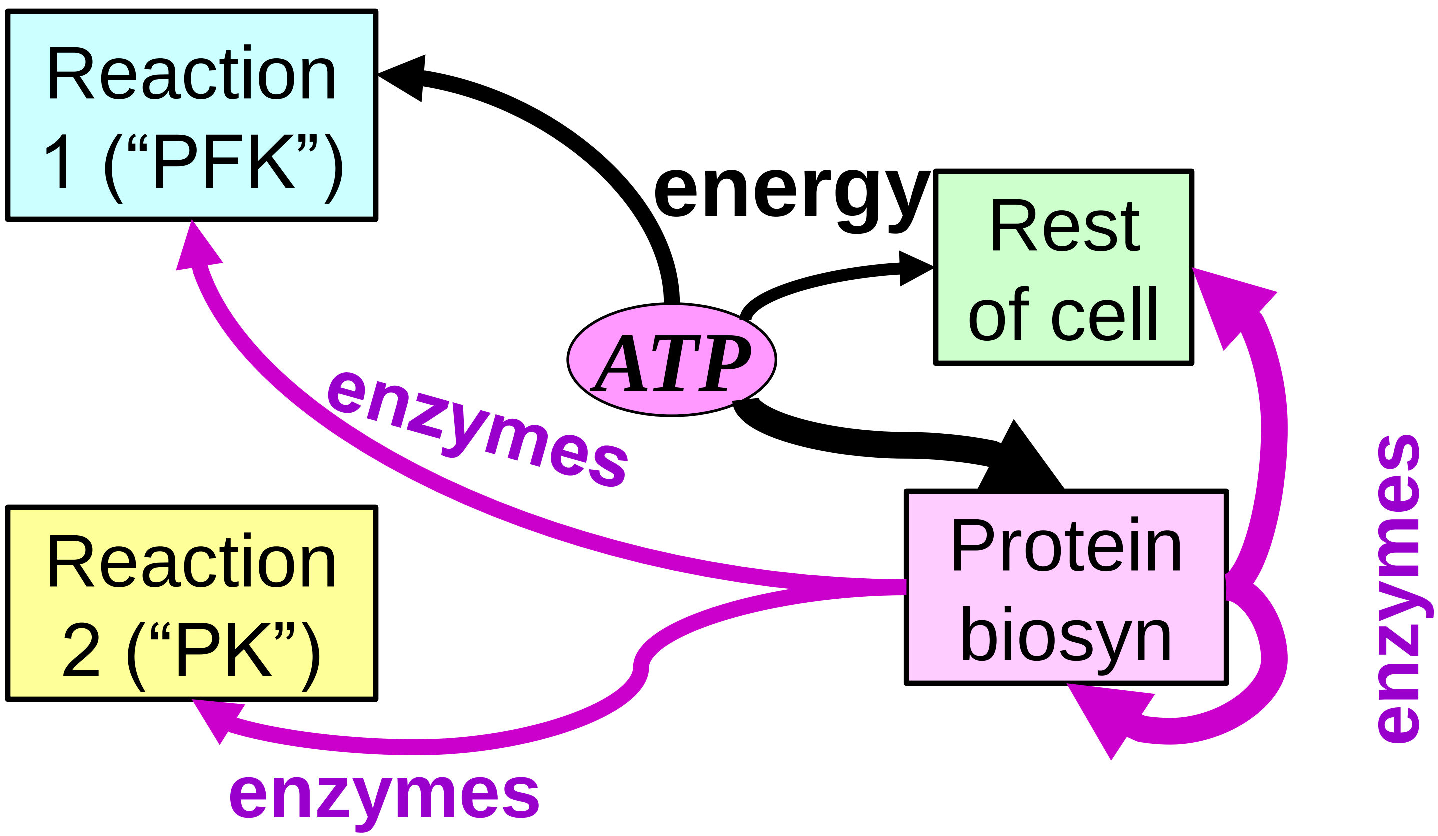
Today’s important point

enzymes catalyze reactions





enzymes catalyze reactions, another source of autocatalysis



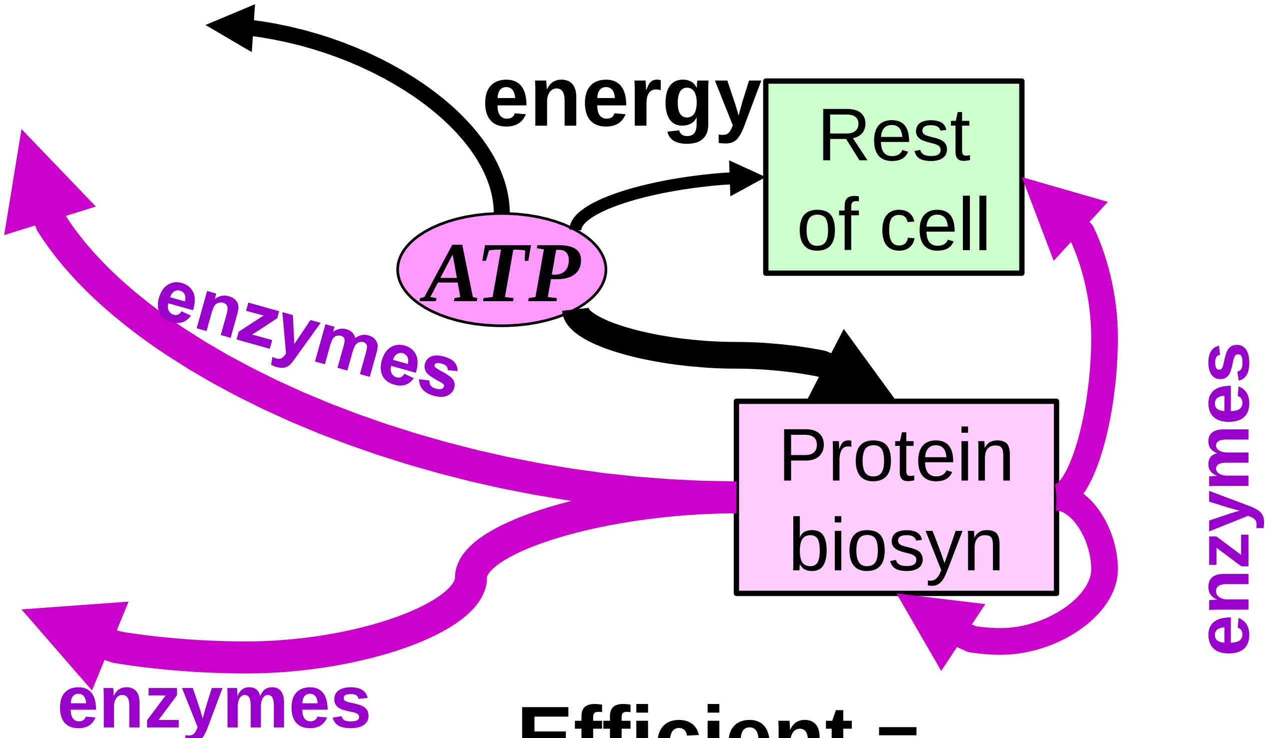
Efficient =

low metabolic overhead

≈ low enzyme amount

enzymes catalyze reactions, another source of autocatalysis

reaction rates ∝ enzyme amount



Can't make too many enzymes here, need to supply rest of the cell.

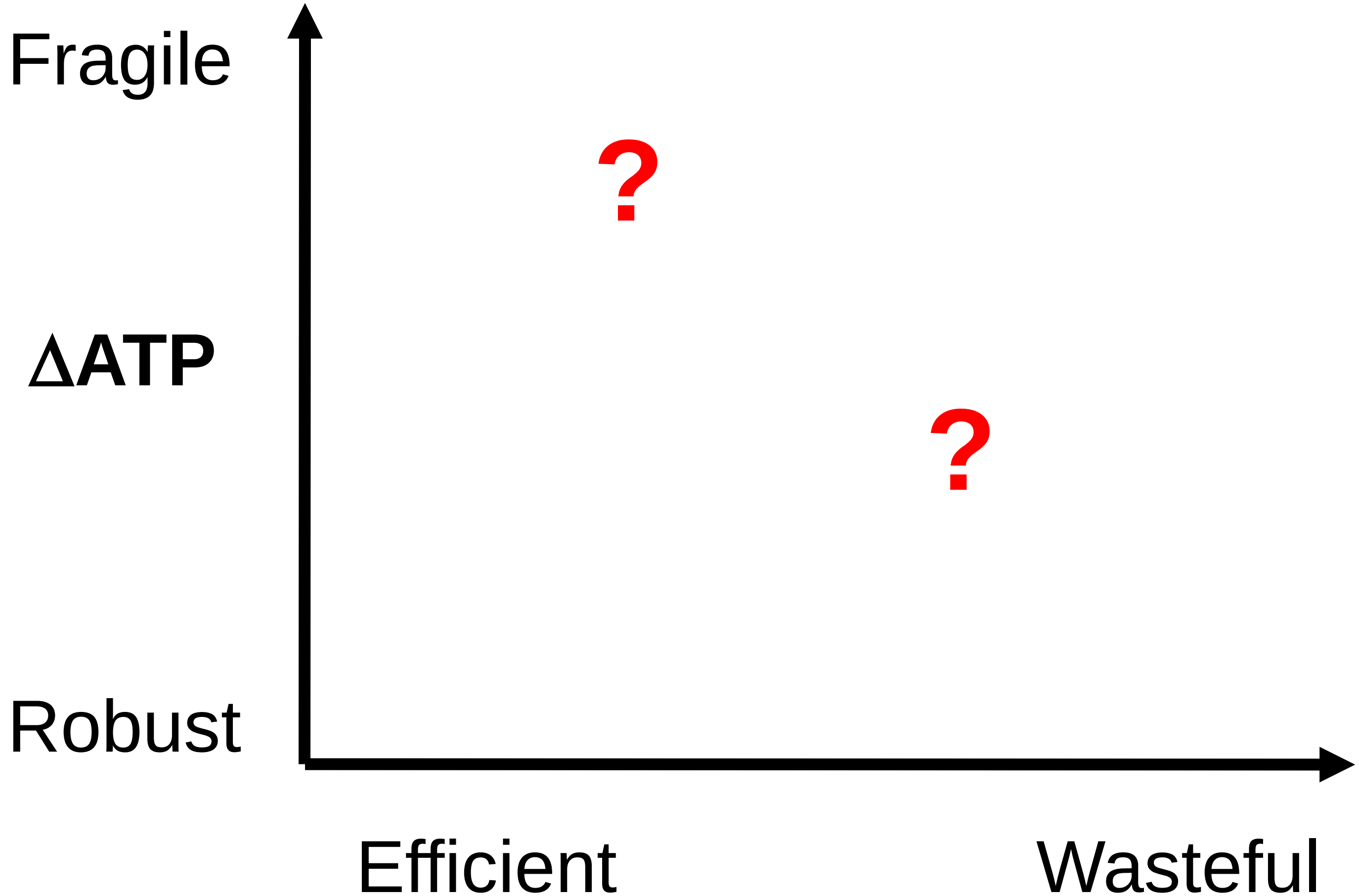
Efficient =

low metabolic overhead

≈ low enzyme amount

(⇒ slow reactions)

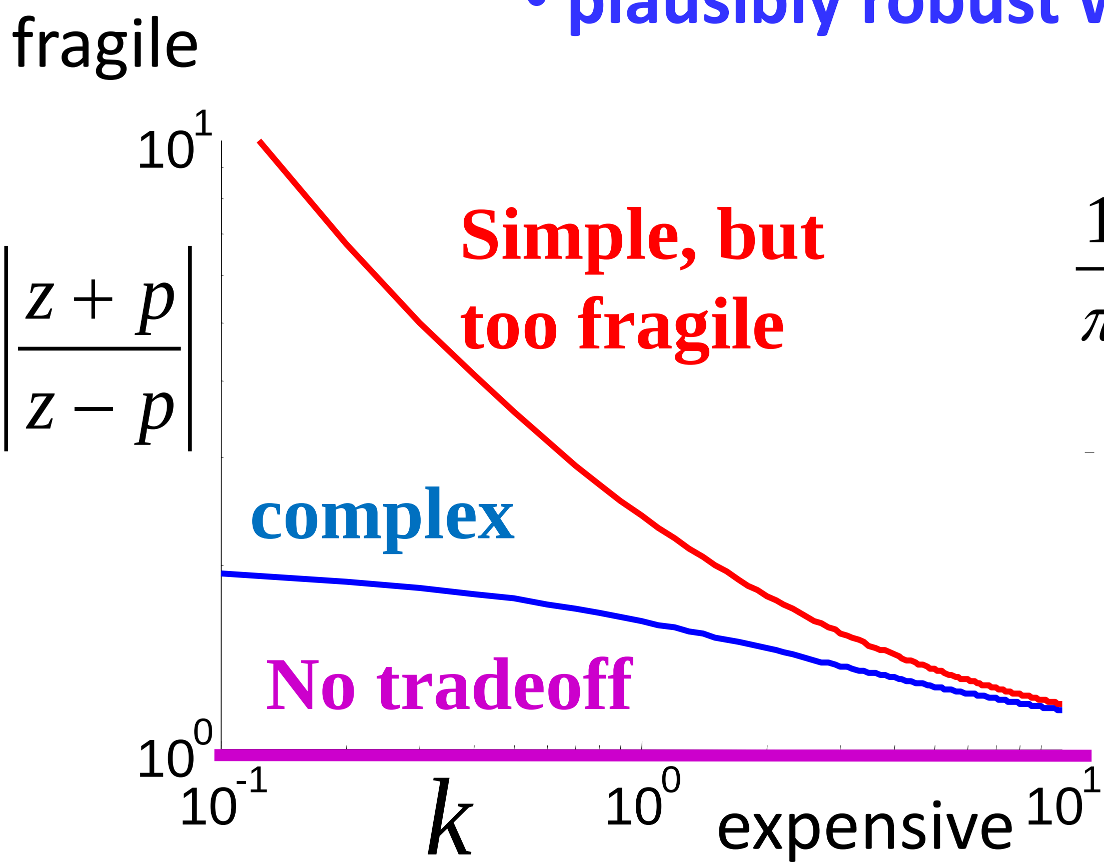
Robust = Maintain ATP



Efficient = low enzyme amount ($\Rightarrow$ slow reactions)

(May 21): Hard tradeoff in glycolysis is

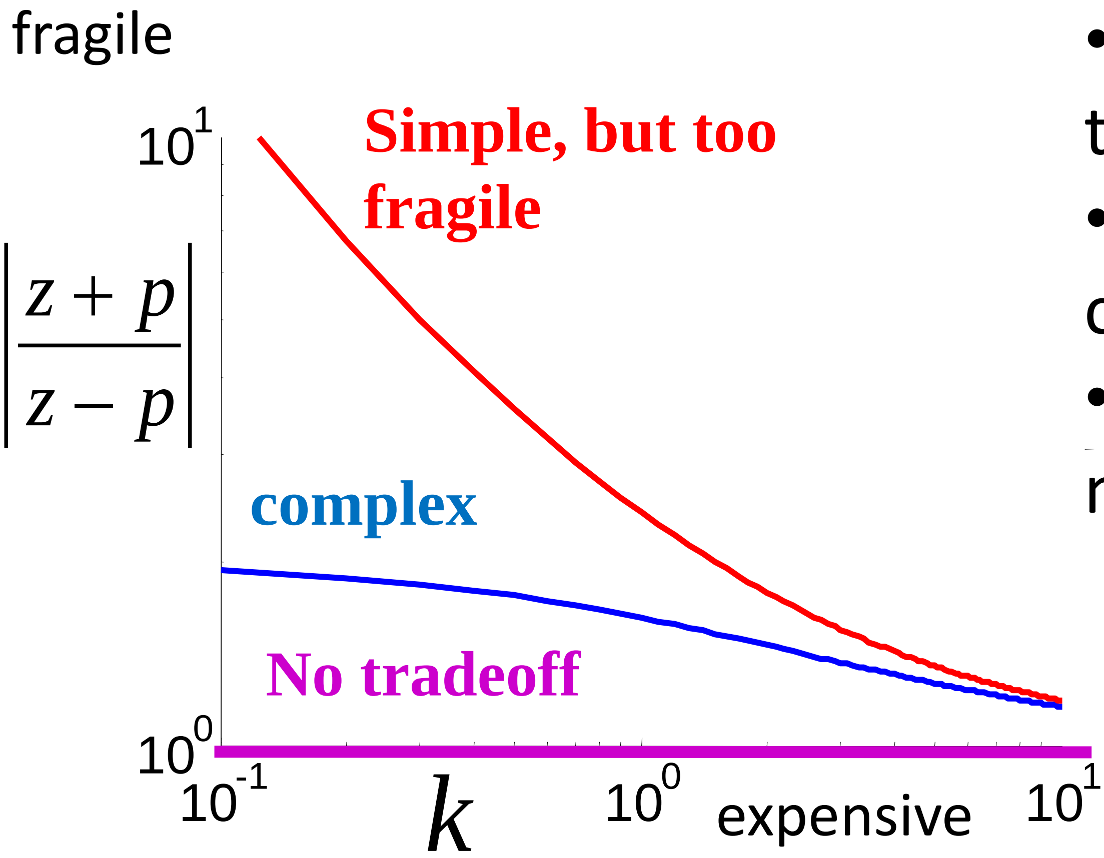
- **robustness vs efficiency**
- **absent without autocatalysis**
- **too fragile with simple control**
- **plausibly robust with complex control**



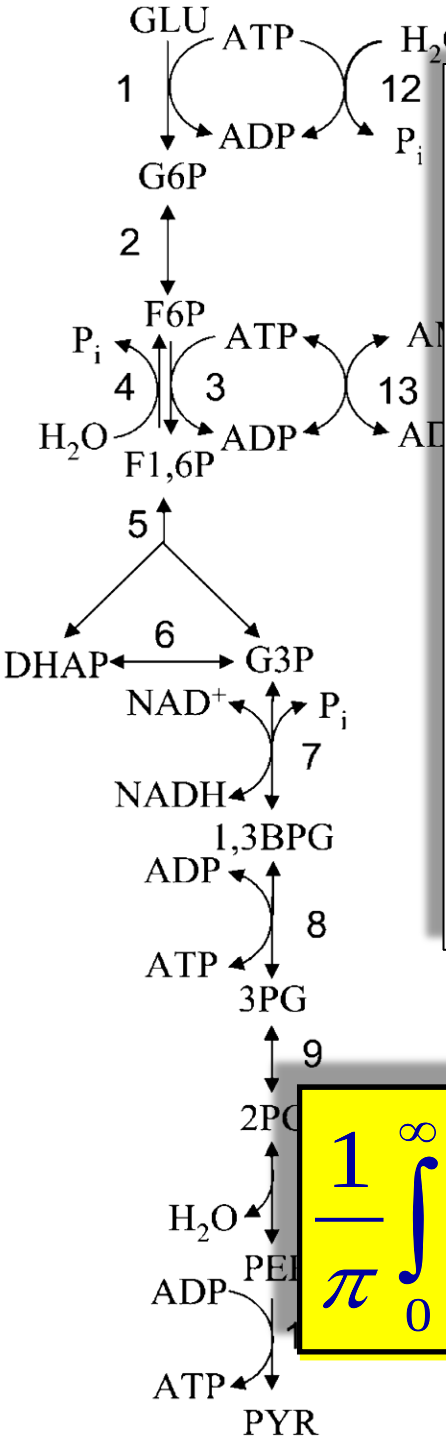
$$\frac{1}{\pi}\int_0^\infty \ln\left|S(j\omega)\right|\left(\frac{z}{z^2+\omega^2}\right)d\omega \geq \ln\left|\frac{z+p}{z-p}\right|$$

(May 21): Hard tradeoff in glycolysis is

- robustness vs efficiency
- absent without autocatalysis
- too fragile with simple control
- plausibly robust with complex control



- Evolution can
- increase complexity
- to improve robustness tradeoffs.
- But this complexity creates new fragilities
- so there is always more to this story.



(P1) Total NAD moiety: $[NAD^+] + [NADH]$

Constrained ("conserved"):

Moieties

1. NAD
2. Adenylate
3. Carbon
4. phosphate
5. oxygen

6. Oxidized state of metabolites
7. Reduced state of metabolites
8. High energy potential release

[F1,6P] + [DHAP] +

[2PG] + [PEP] + [PYR]

HAP] + [G3P] +

+ [PEP] +

DP] + $[P_i]$

$P_i] + [H_2O]$

] +

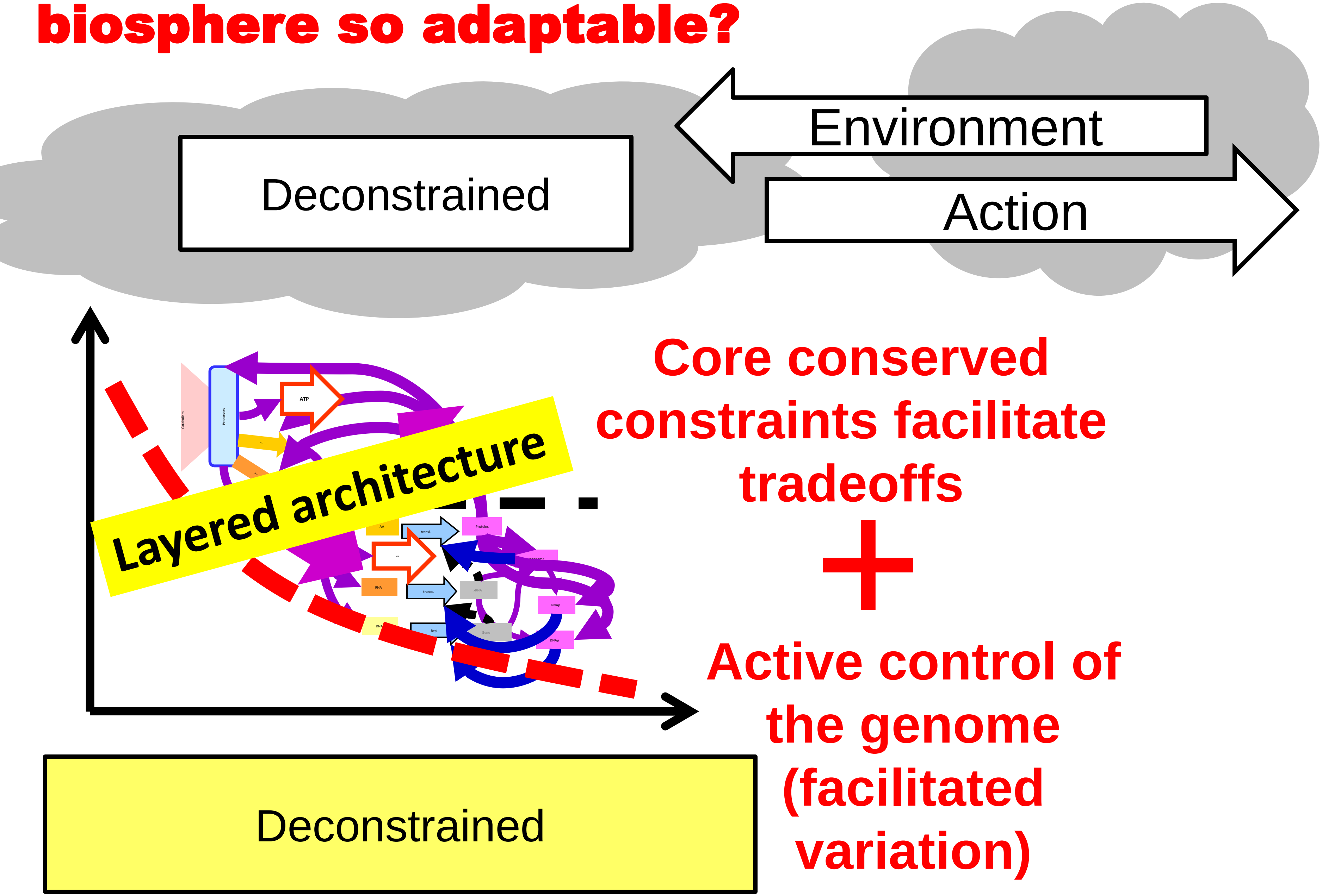
+ [PEP] + [PYR] + $[NAD^+]$

$$\frac{1}{\pi}\int_0^\infty \ln\left|S(j\omega)\right|\left(\frac{z}{z^2+\omega^2}\right)d\omega \geq \ln\left|\frac{z+p}{z-p}\right|$$

] +

+ [PYR] + [ADP] + 2[AMP] + $[H_2O]$

What makes the bacterial biosphere so adaptable?

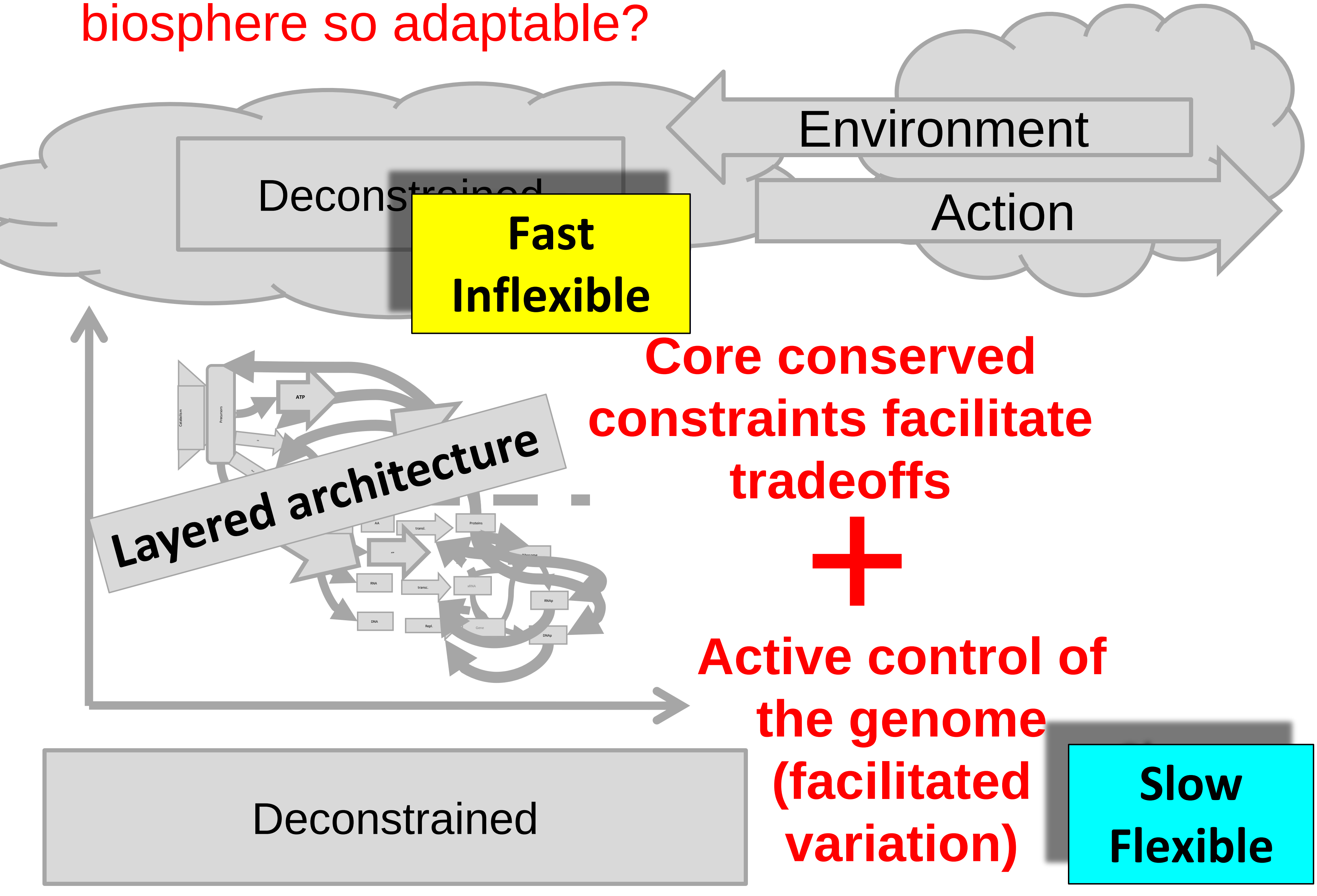


Core conserved constraints facilitate tradeoffs

+

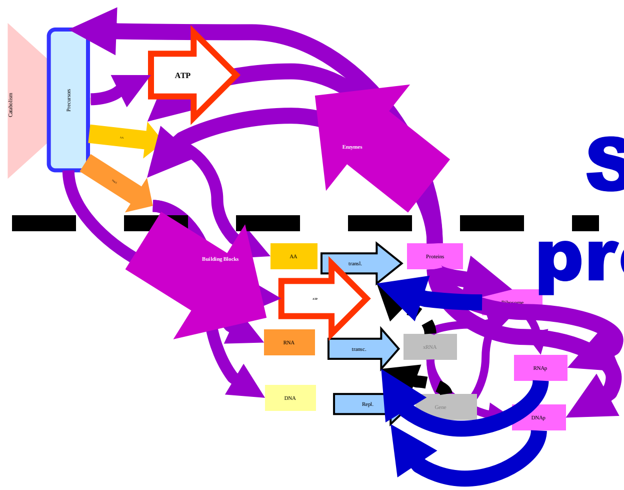
Active control of the genome (facilitated variation)

What makes the bacterial biosphere so adaptable?



Bacterial biosphere

Deconstrained Environments



Shared protocols

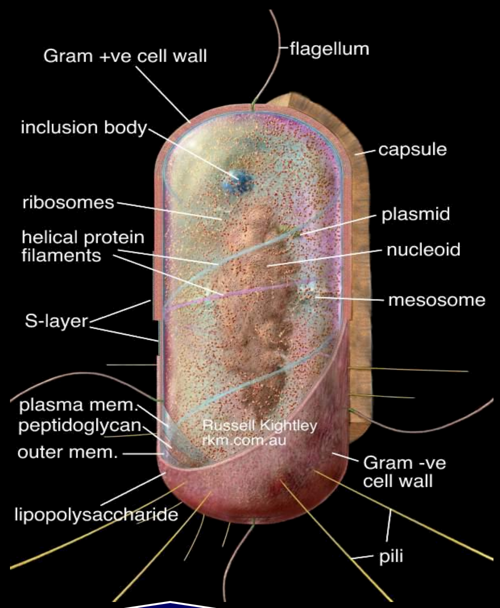
Architecture
=
Constraints
that
Deconstrain

Deconstrained Genomes

System

“Emergent”: “Nontrivial” consequences of other constraints

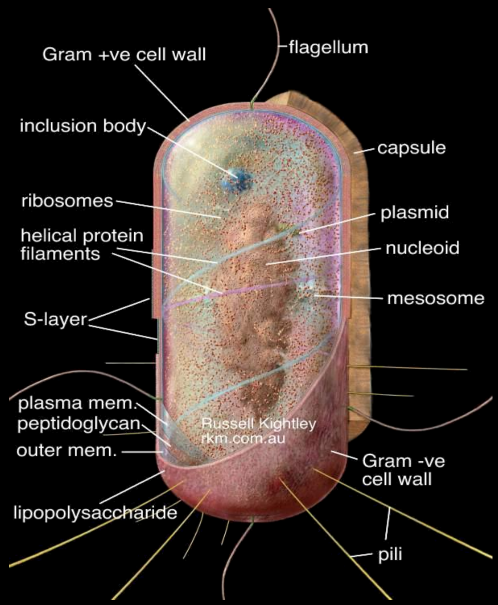
Architecture =Constraints



Protocols

Components

Systems requirements: Survive in hostile environments



Constraints

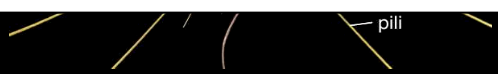
Components and materials: “Chemistry”

Constrained (“conserved”):

Moieties

1. NAD
2. Adenylate
3. Carbon
4. phosphate
5. oxygen

6. Oxidized state of metabolites
7. Reduced state of metabolites
8. High energy potential release



Constraints

Components and materials:
“Chemistry”

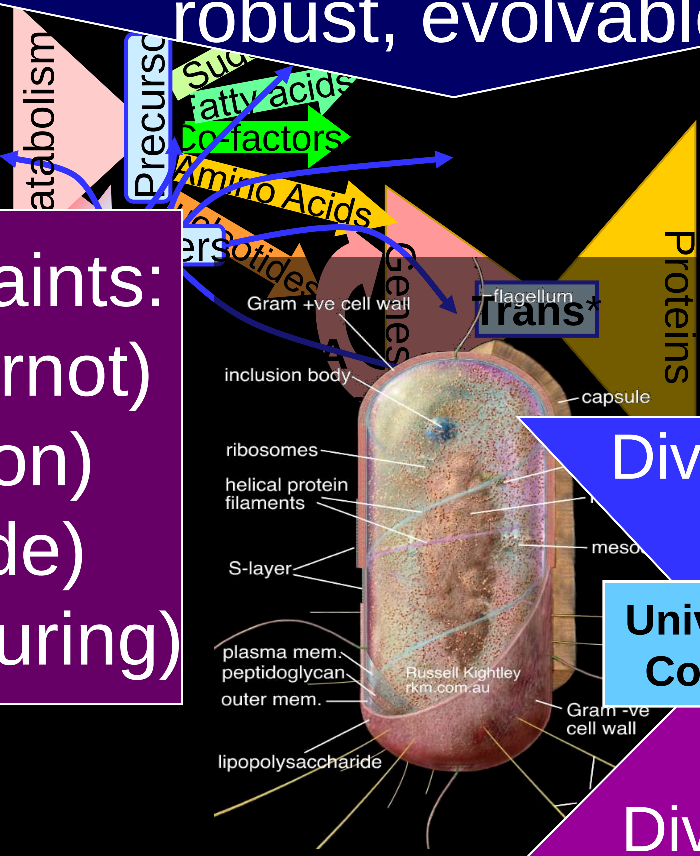
Bacterial biosphere

- carriers: ATP, NADH, etc
- Precursors, …
- Enzymes
- Translation
- Transcription
- Replication
- …

Protocols

Architecture = protocols
= "constraints that deconstrain"

Systems requirements: functional, efficient, robust, evolvable



Hard constraints:
Thermo (Carnot)
Info (Shannon)
Control (Bode)
Compute (Turing)

Constraints

Diverse

Universal Control

Diverse

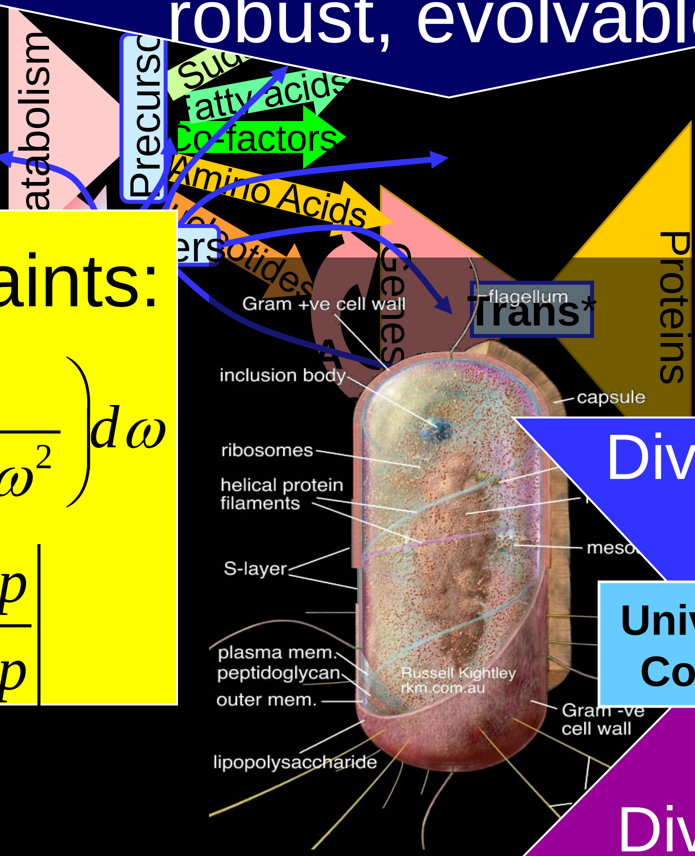
Protocols

Components and materials: Energy, moieties

Systems requirements: functional, efficient, robust, evolvable

Hard constraints:

$$\frac{1}{\pi}\int_0^\infty \ln\left|S(j\omega)\right|\left(\frac{z}{z^2+\omega^2}\right)d\omega \ge \ln\left|\frac{z+p}{z-p}\right|$$



Diverse

Universal Control

Diverse

Protocols

Constraints

Components and materials: Energy, moieties

OPEN ACCESS Freely available online PLoS BIOLOGY

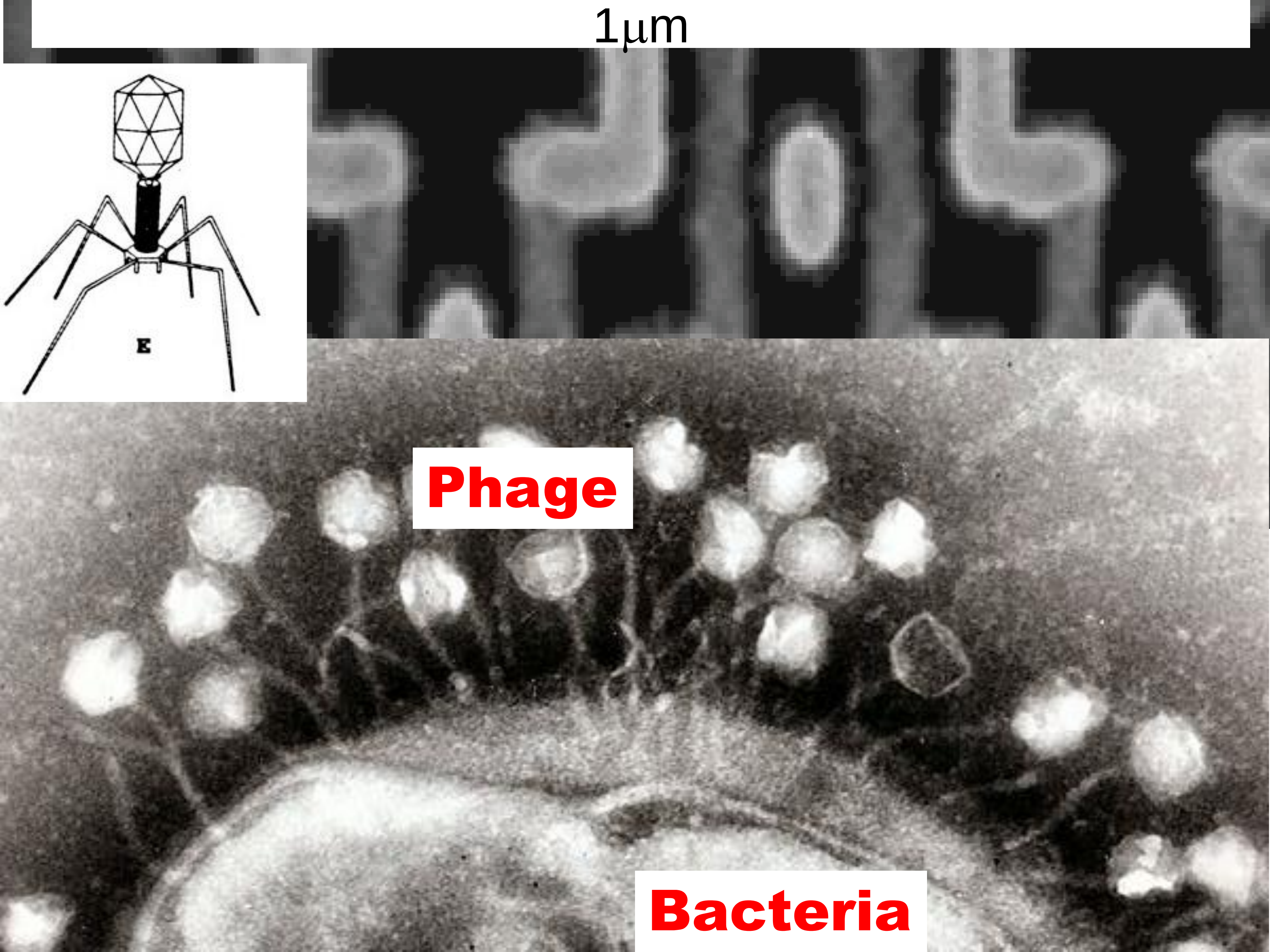
Viruses' Life History: Towards a Mechanistic Basis of a Trade-Off between Survival and Reproduction among Phages

Marianne De Paepe, François Taddei*

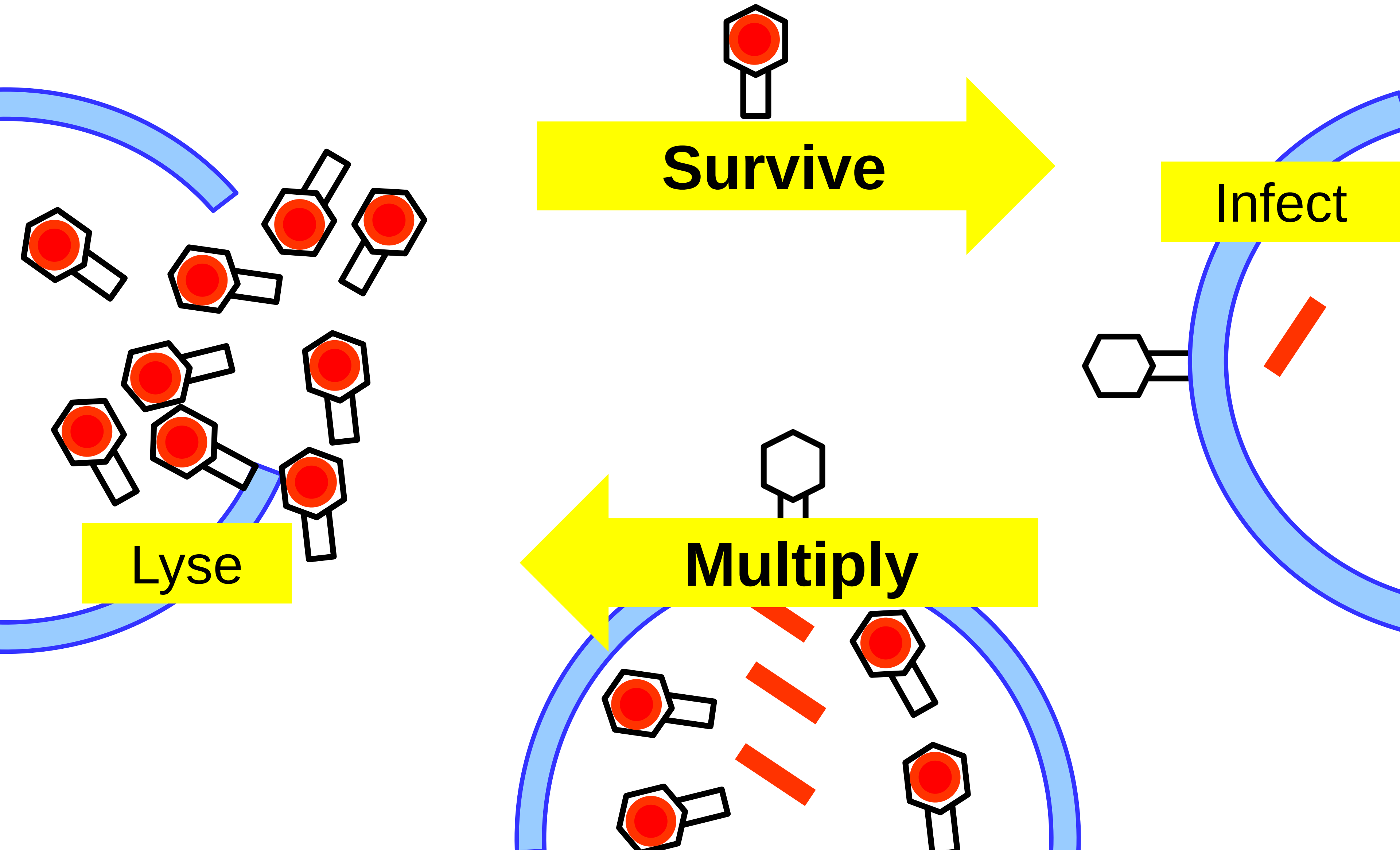
Laboratoire de Genetique Moleculaire, Evolutive et Medicale, University of Paris 5, INSERM, Paris, France

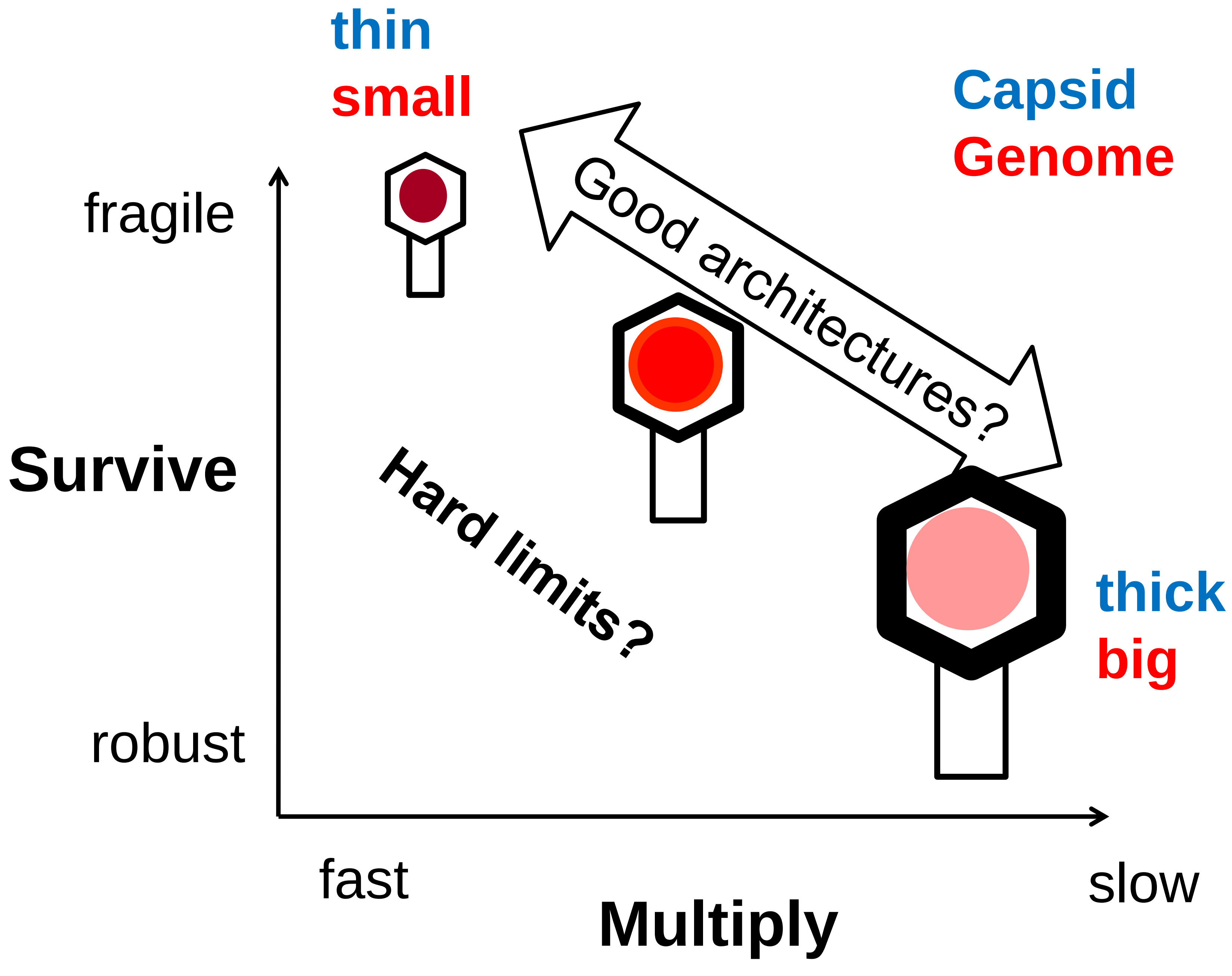
July 2006 | Volume 4 | Issue 7 | e193

I recently found this paper, a rare example of exploring an explicit tradeoff between robustness and efficiency. This seems like an important paper but it is rarely cited.


1μm
E
Phage
Bacteria

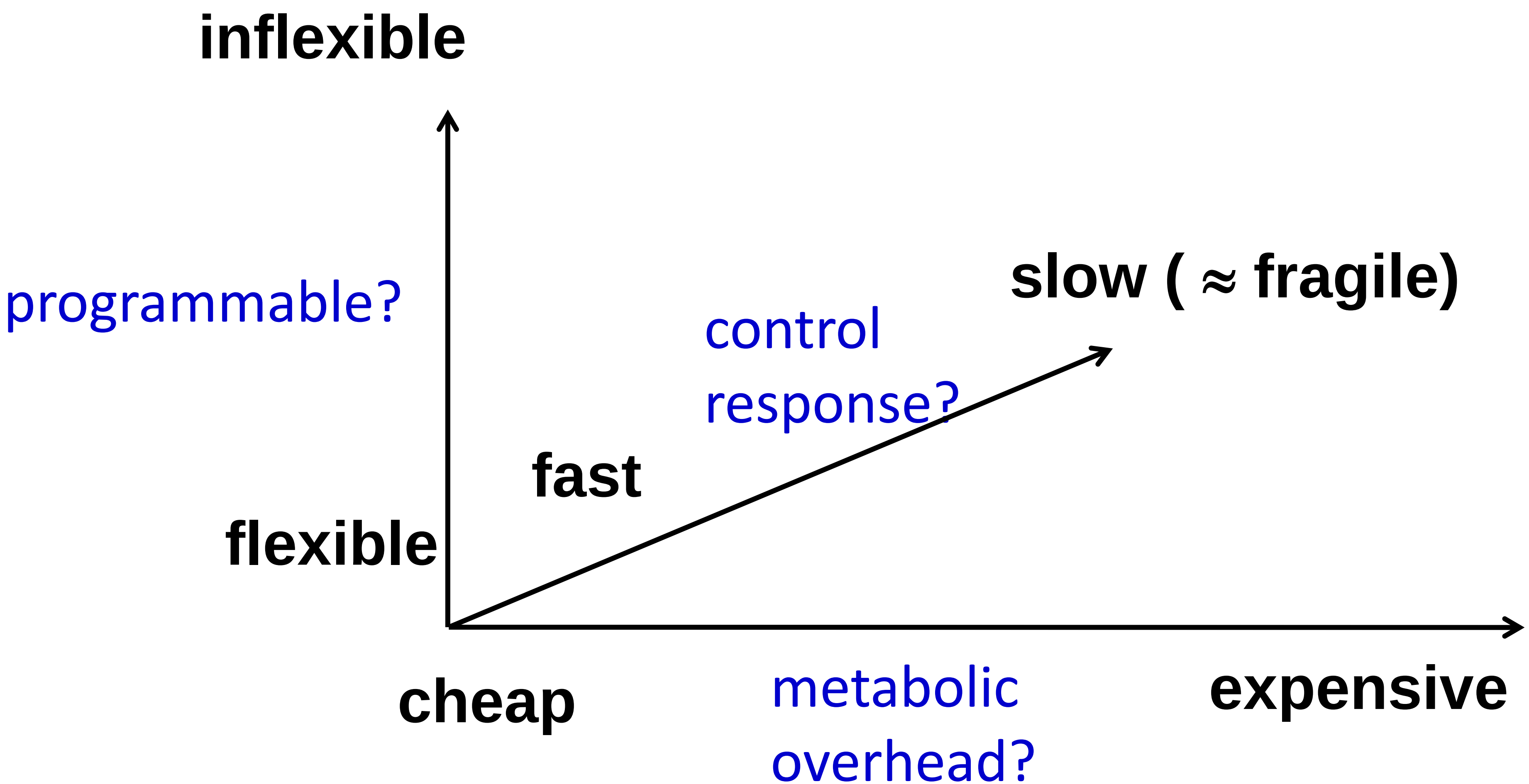
Phage lifecycle




thin
small
Capsid
Genome
fragile
Good architectures?
Survive
Hard limits?
thick
big
robust
fast
slow
Multiply

Tradeoffs?

Accurate vs sloppy is now an implicit dimension of robust/fragile



Conjecture: human brain tradeoffs dominated by fast vs flexible more than robust vs cheap

1. For hunter/gatherer metabolism is far above basal, and dominated by active muscle
2. Brain homeostasis is a much greater challenge than basal metabolic demands

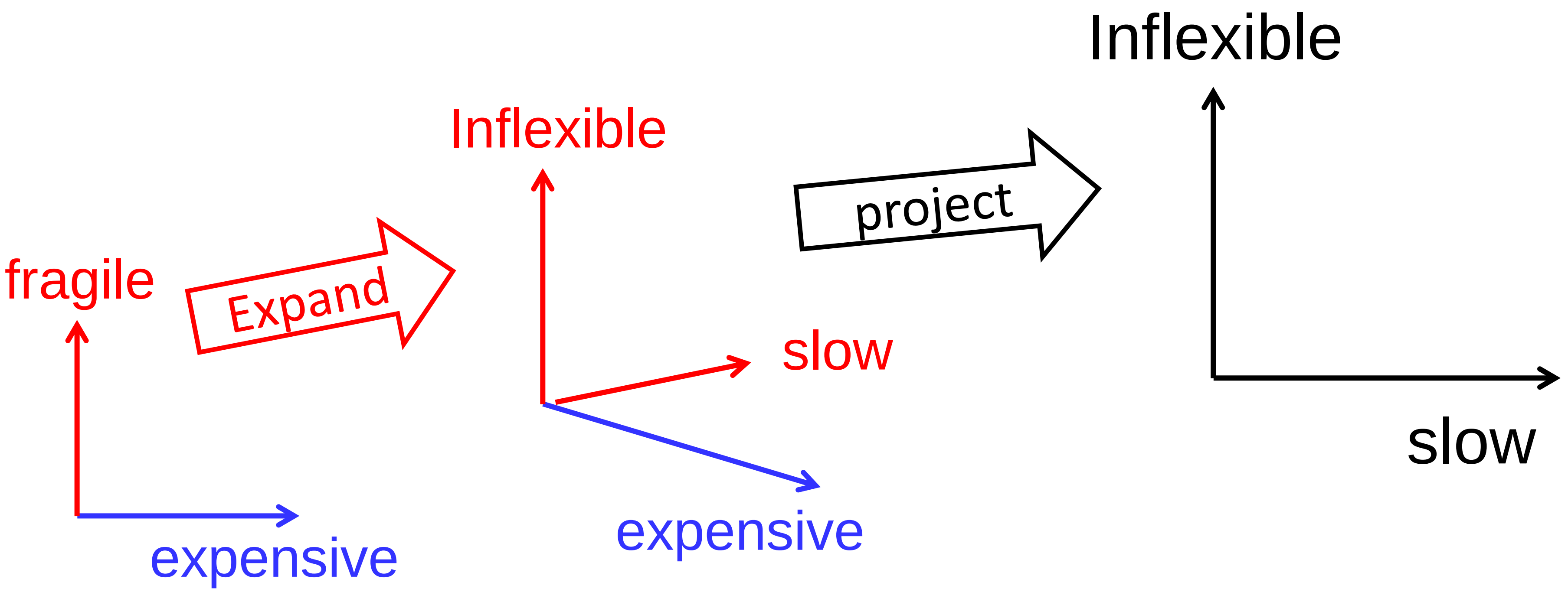
Creates new fragilities in modern lifestyle

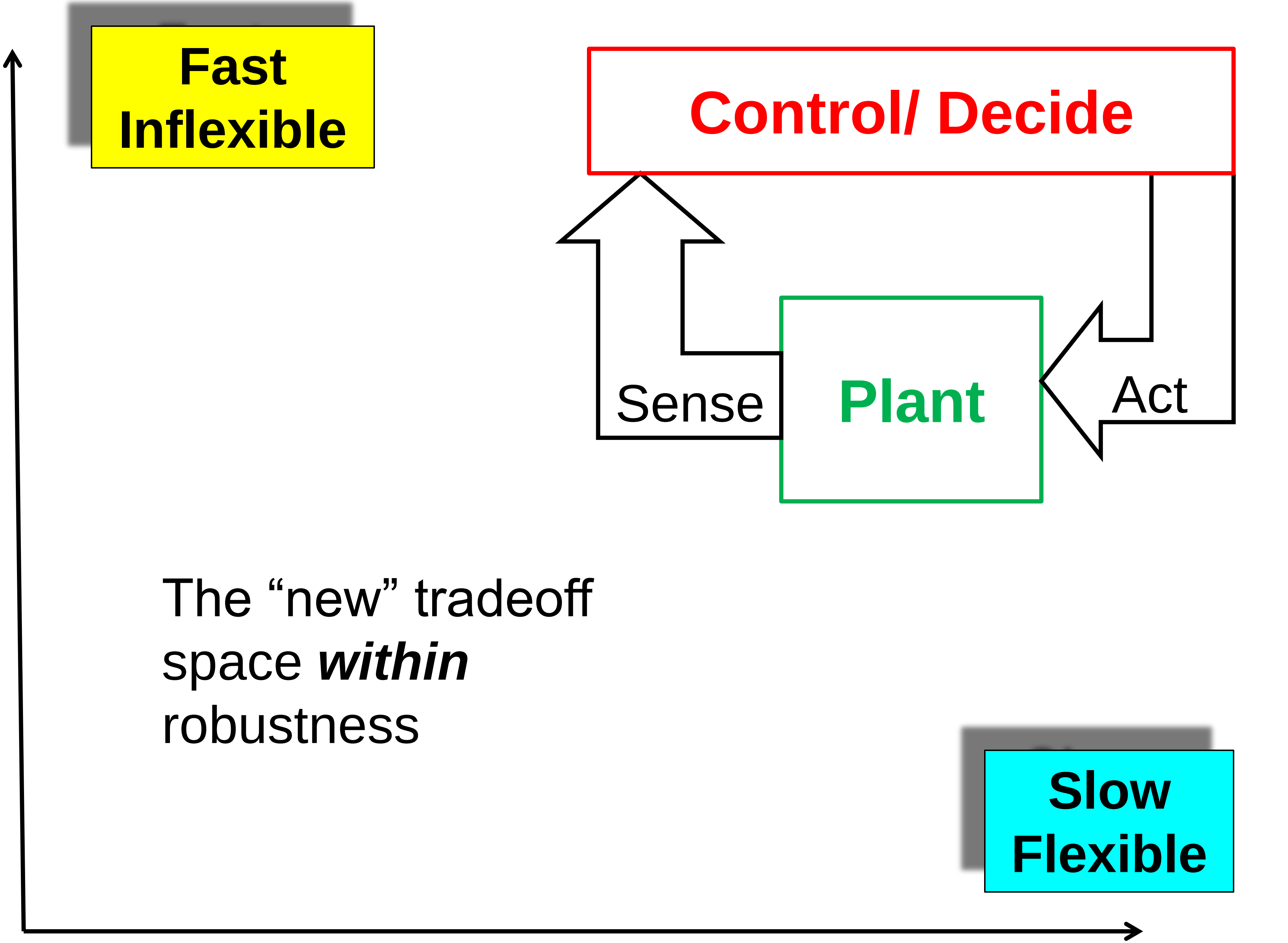
Not true for sedentary organisms with limited nutrient diets (e.g. Koala, Panda, …)

Conjecture: human brain tradeoffs dominated by fast vs flexible more than robust vs cheap

Fragility dimensions with most important tradeoffs:

1. latency/delay/speed of control vs.
2. flexibility/adaptability

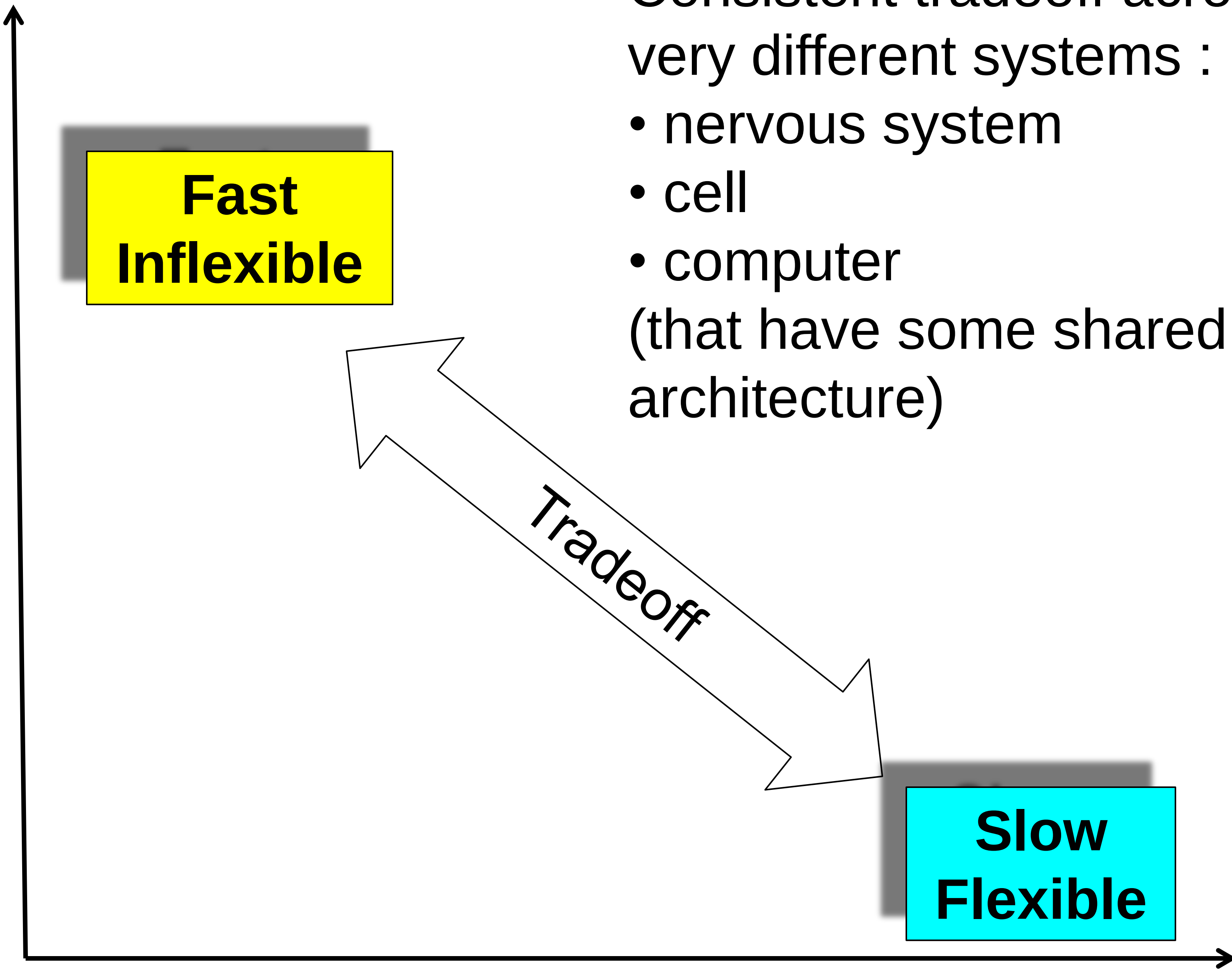



Fast
Inflexible
Control/ Decide
Sense
Plant
Act
The “new” tradeoff space within robustness
Slow
Flexible

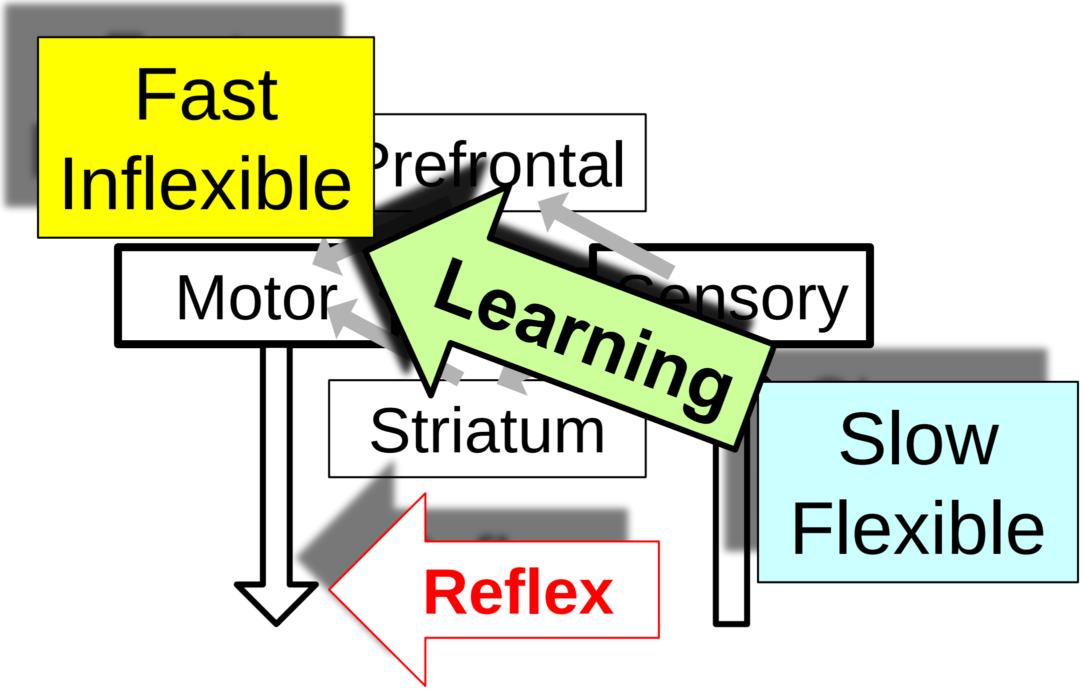
Consistent tradeoff across very different systems :

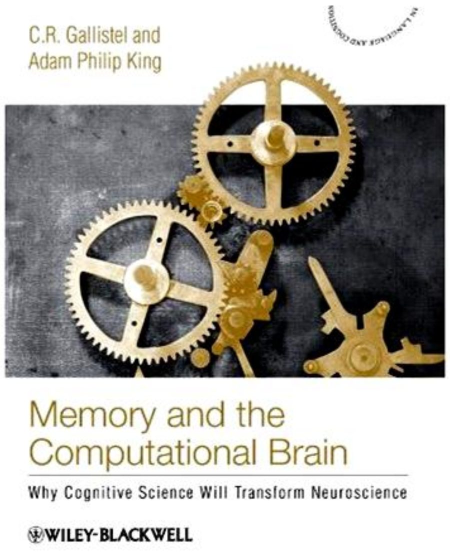
- nervous system
- cell
- computer

(that have some shared architecture)

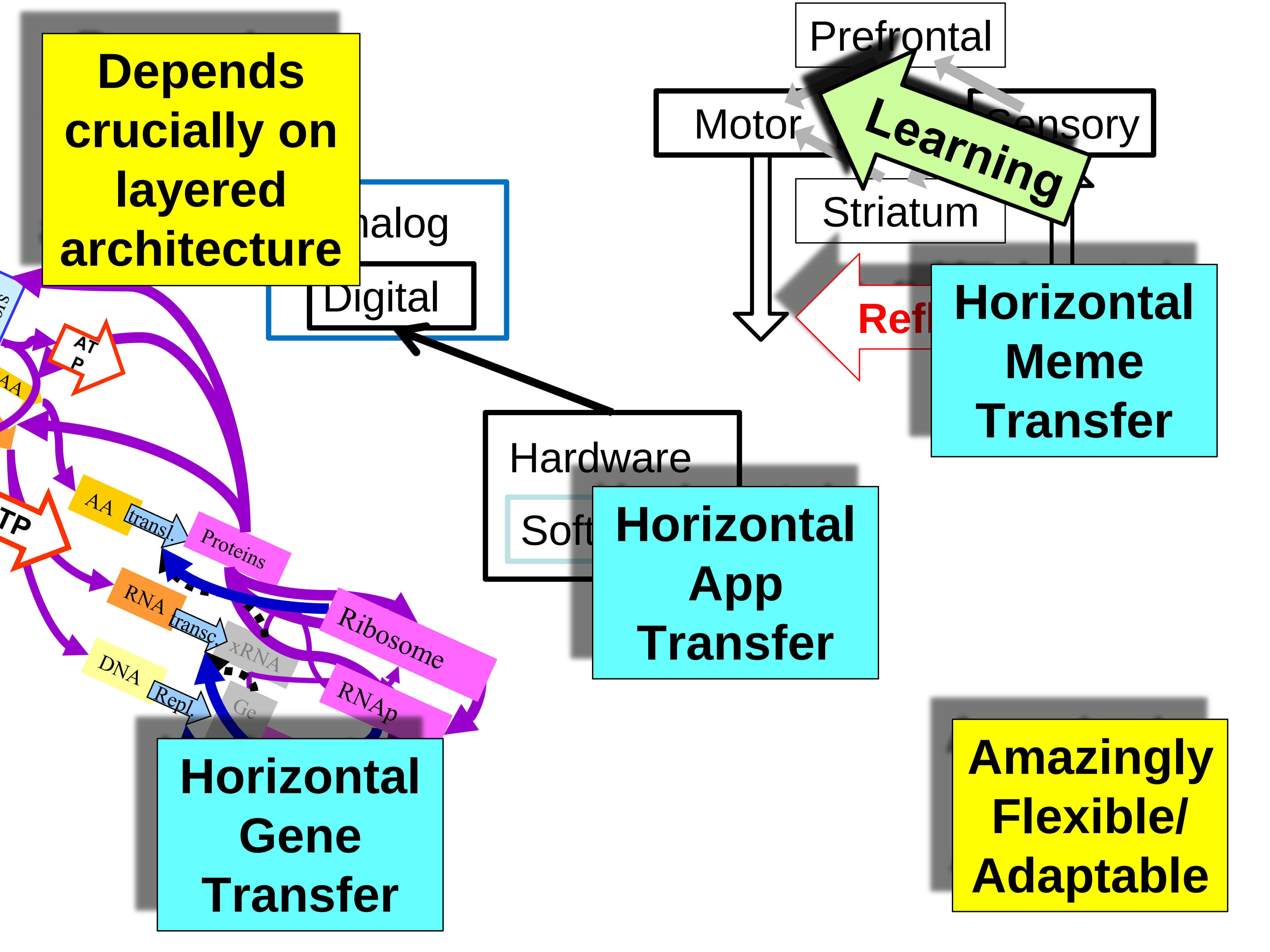


Gallistel and King

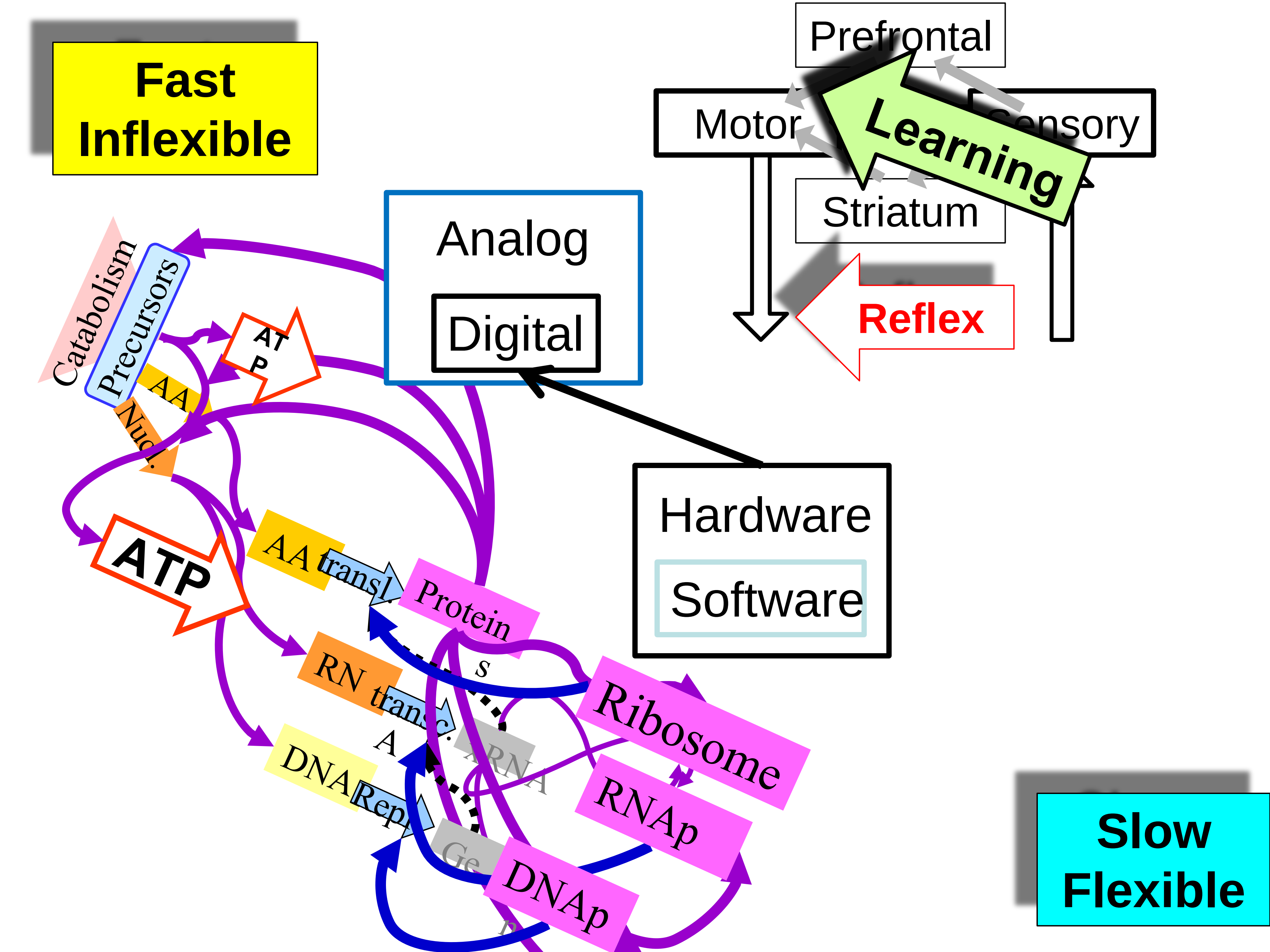


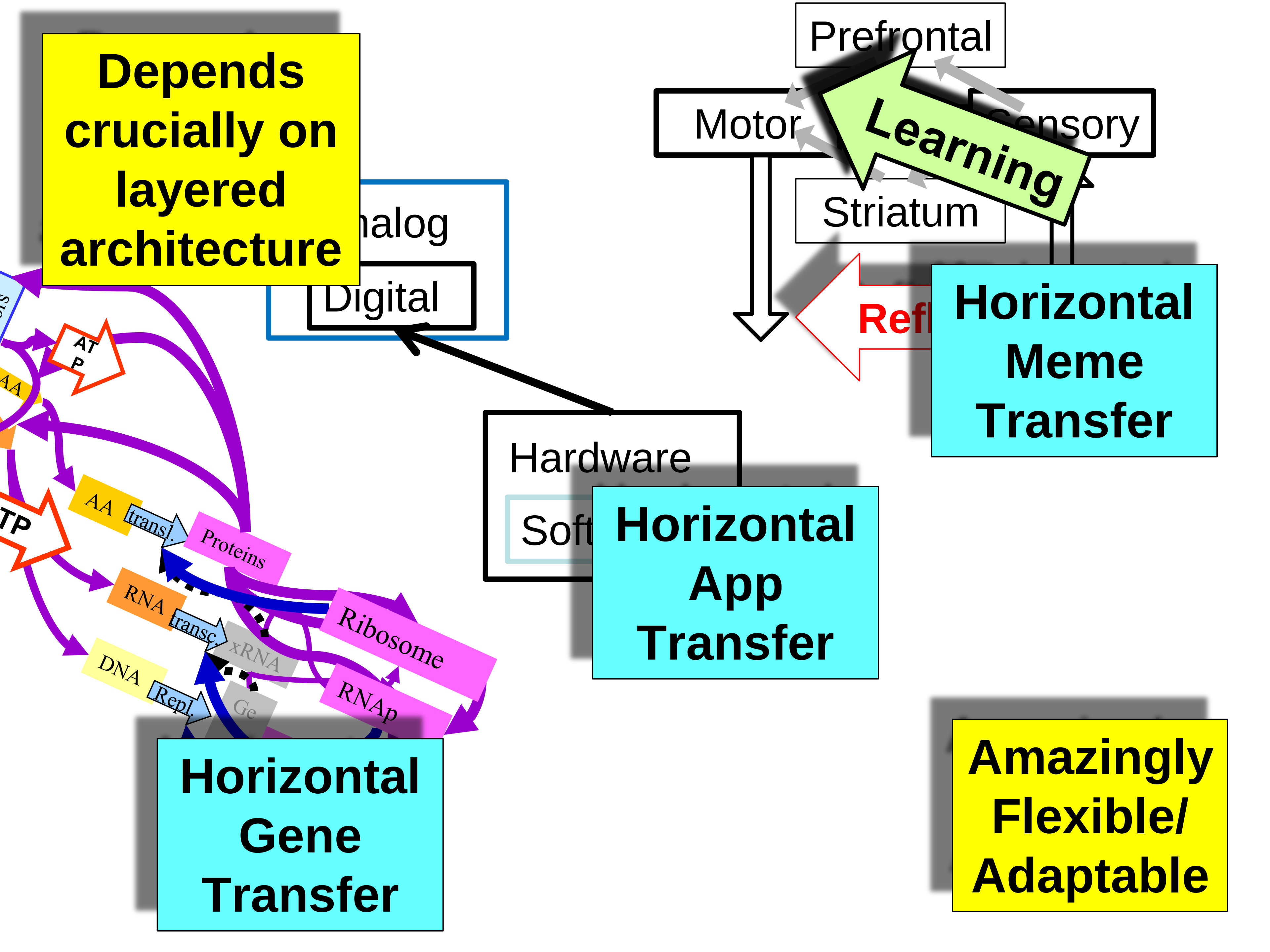


- Sensori-motor memory potential ≈ ∞ (Ashby)
- Limits are on ***speed*** of
  - nerve propagation delays
  - learning
- But control is ***never*** centralized
- Is there a random access read/write memory?

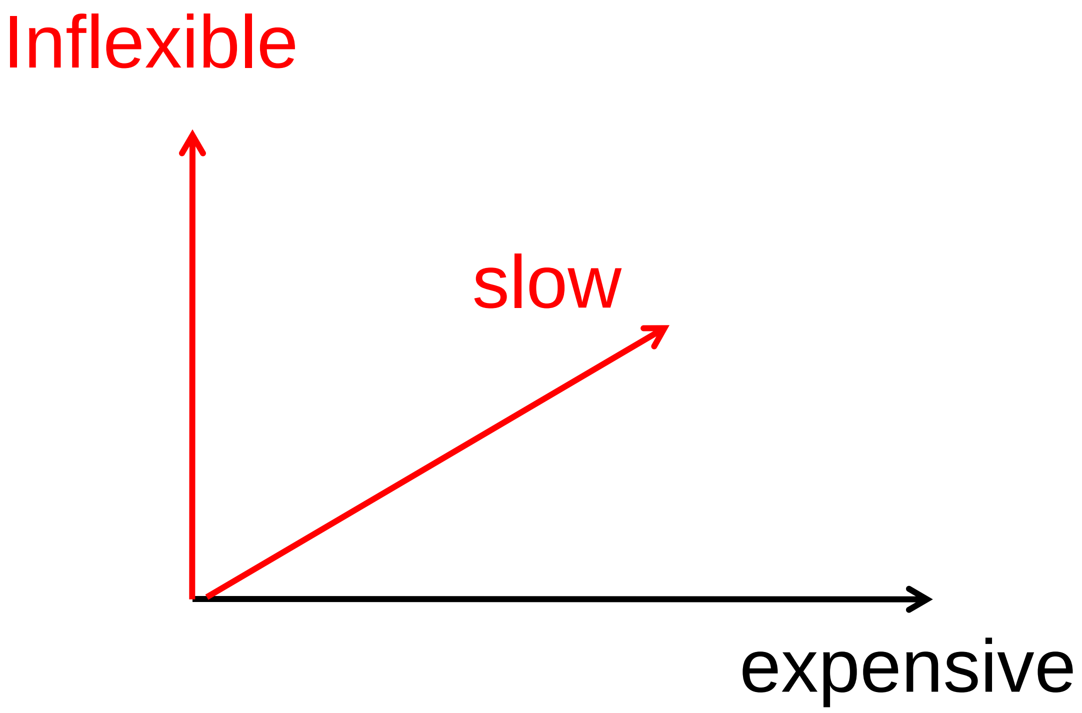


Depends crucially on layered architecture
Prefrontal
Motor
Sensory
Learning
Striatum
Digital
Hardware
Horizontal Meme Transfer
Horizontal App Transfer
AA
transl.
Proteins
RNA
transc.
xRNA
DNA
Repl.
Ribosome
RNAp
Horizontal Gene Transfer
Amazingly Flexible/ Adaptable

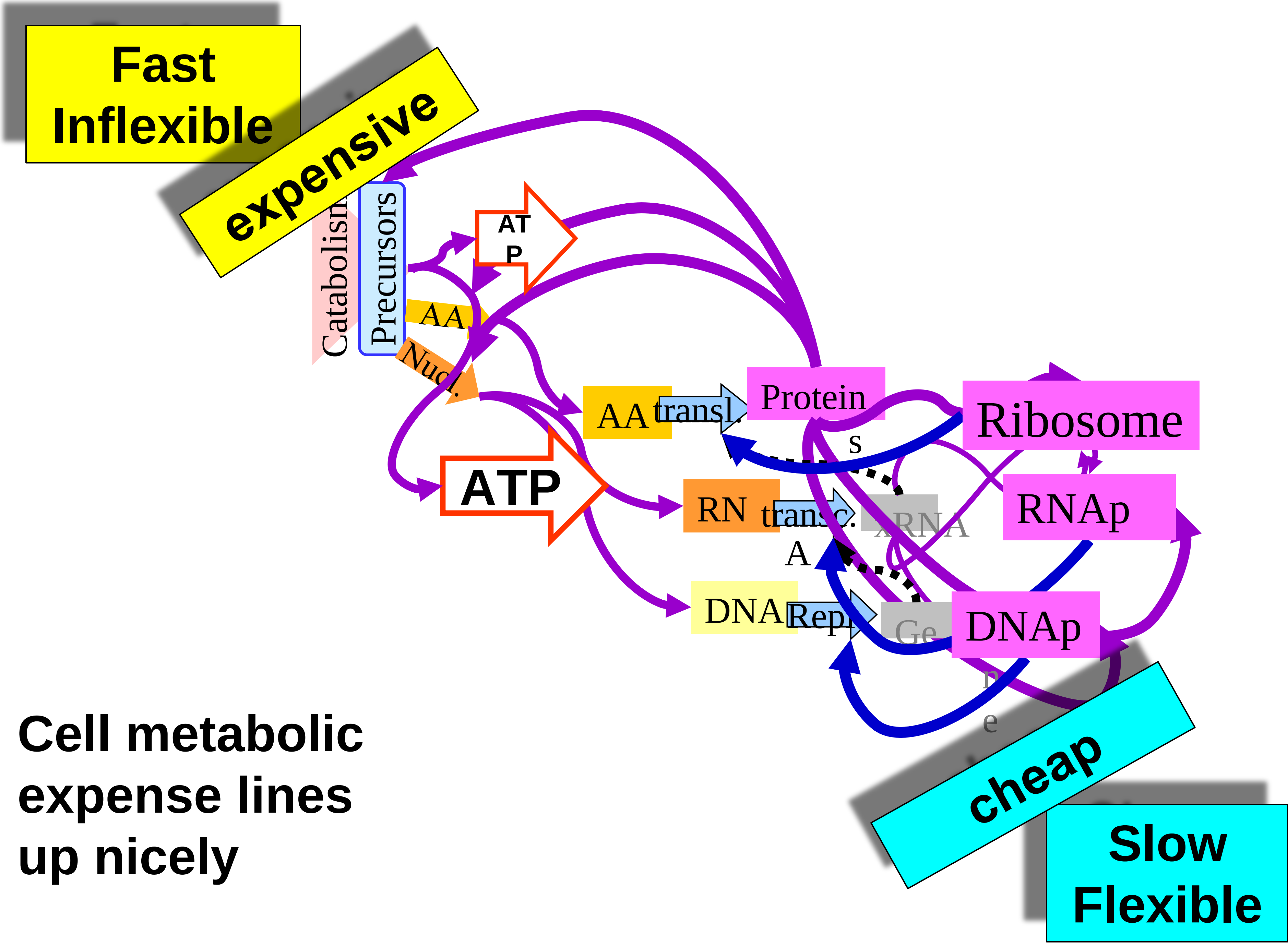

Fast
Inflexible
Prefrontal
Motor
Sensory
Learning
Striatum
Reflex
Analog
Digital
Hardware
Software
Catabolism
Precursors
ATP
AA
Nucl.
ATP
AA transl.
Protein
RN transc.
DNA Repl.
RNA
Ribosome
RNAp
DNAp
Slow
Flexible



Depends crucially on layered architecture
nalog
Digital
Prefrontal
Motor
Sensory
Striatum
Learning
Refl
Horizontal Meme Transfer
Hardware
Soft
Horizontal App Transfer
AT P
AA
TP
AA
transl.
Proteins
RNA
transc.
xRNA
DNA
Repl.
Ribosome
RNAp
Horizontal Gene Transfer
Amazingly Flexible/ Adaptable

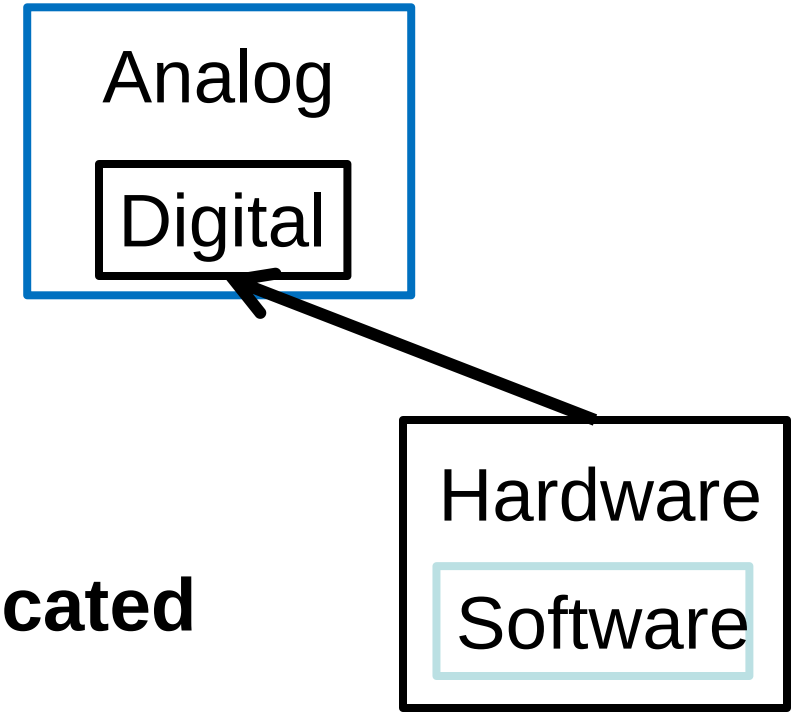


But efficiency tradeoffs are different.



Cell metabolic expense lines up nicely

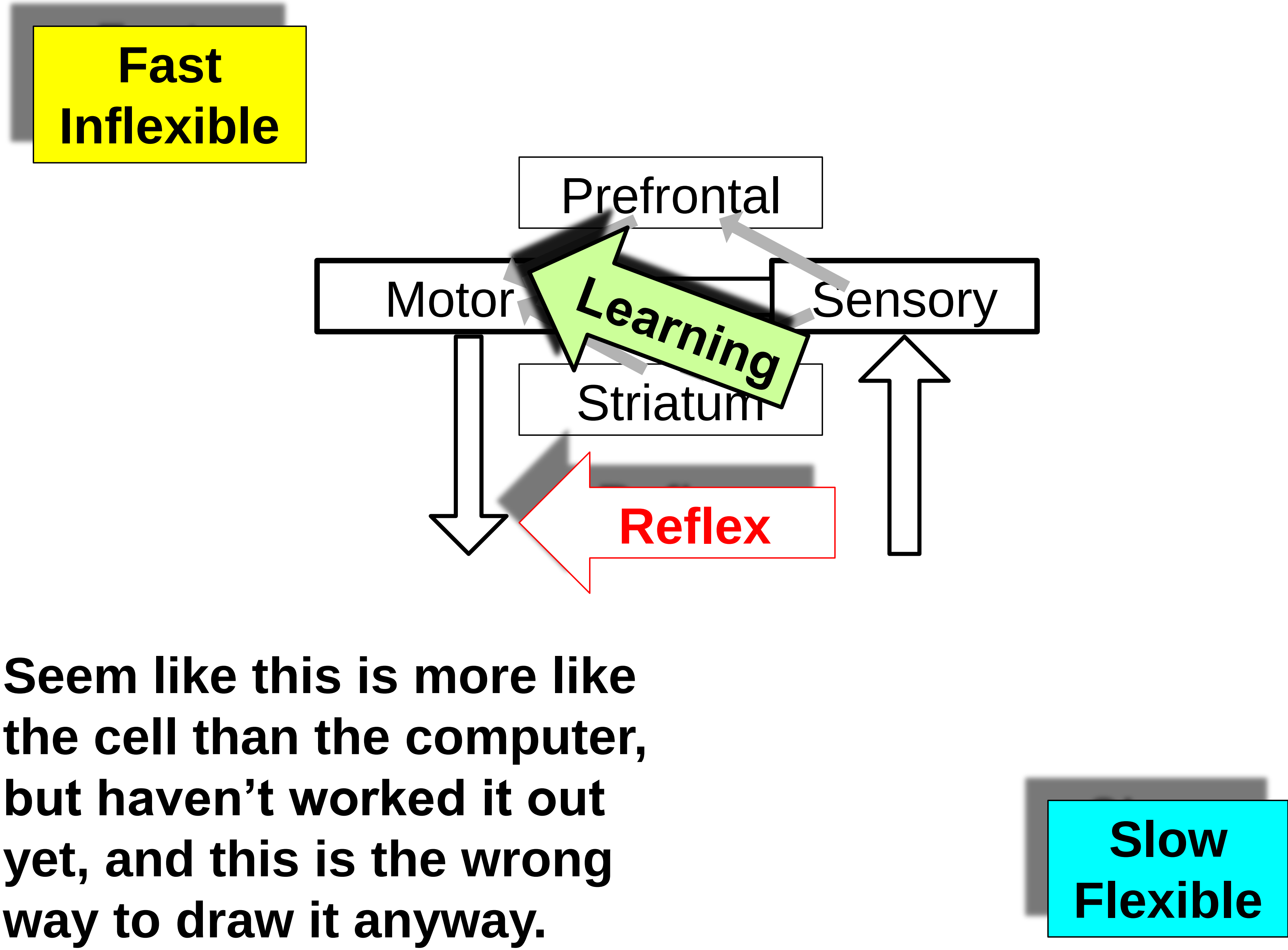
Fast
Inflexible



Expense is complicated tradeoff between

- **design effort**
- **fabrication cost**
- **energy use**
- **etc etc**

Slow
Flexible



Seem like this is more like the cell than the computer, but haven't worked it out yet, and this is the wrong way to draw it anyway.

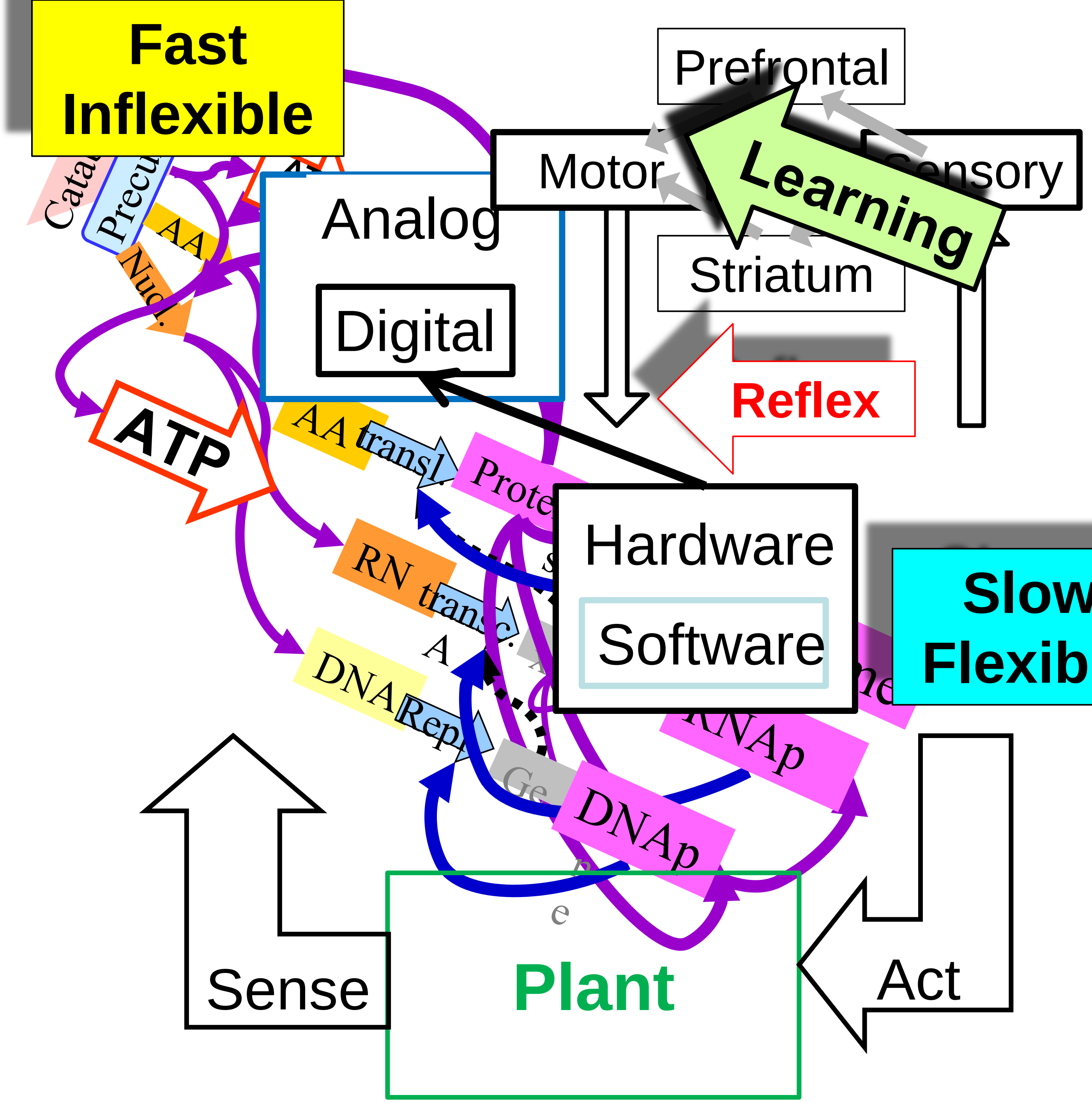
Compute

Turing

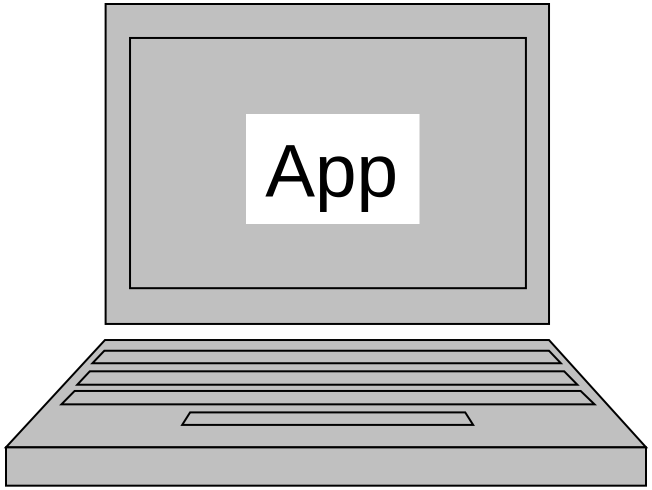
Delay is *most* important

Bode

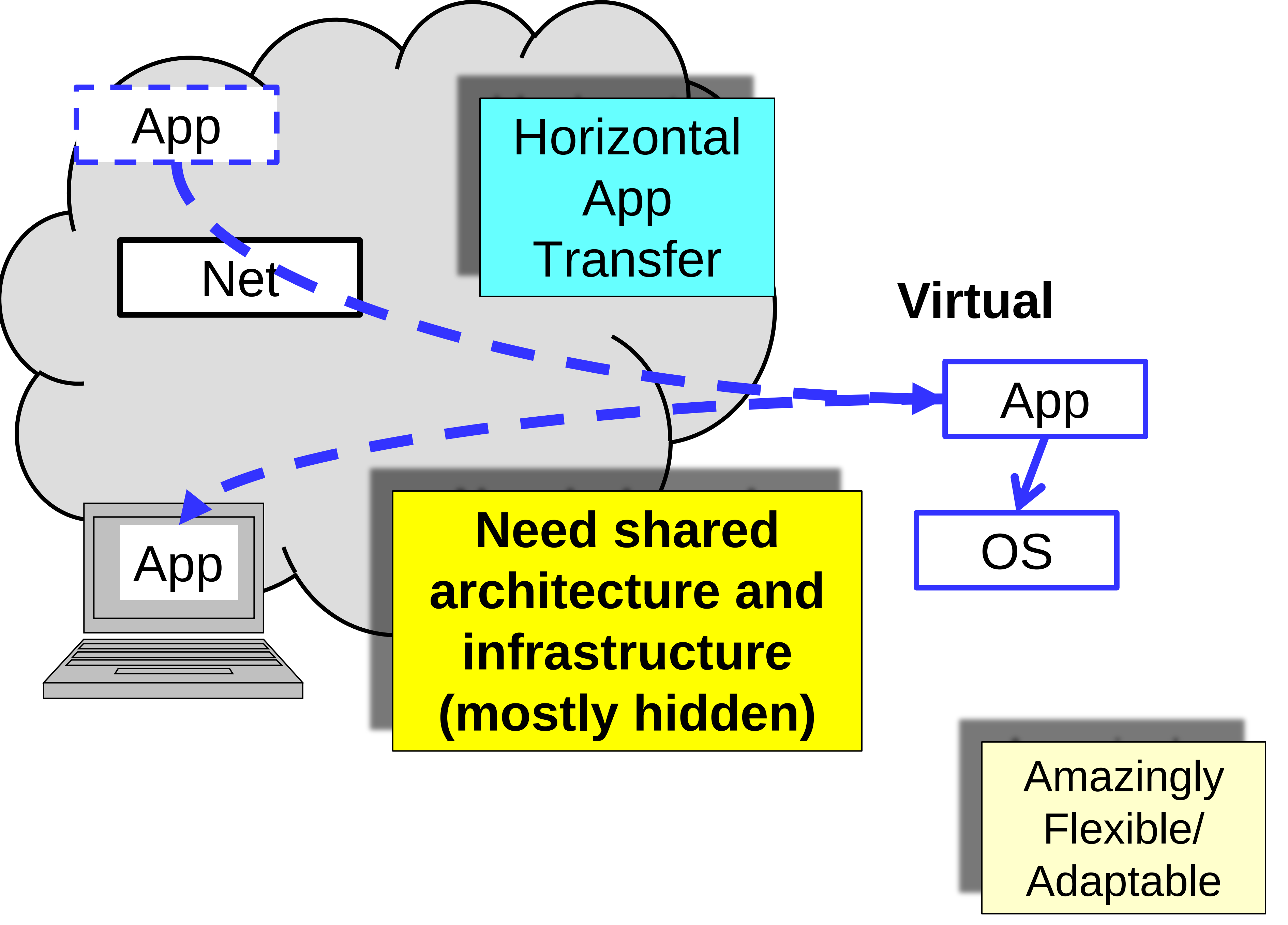
Control

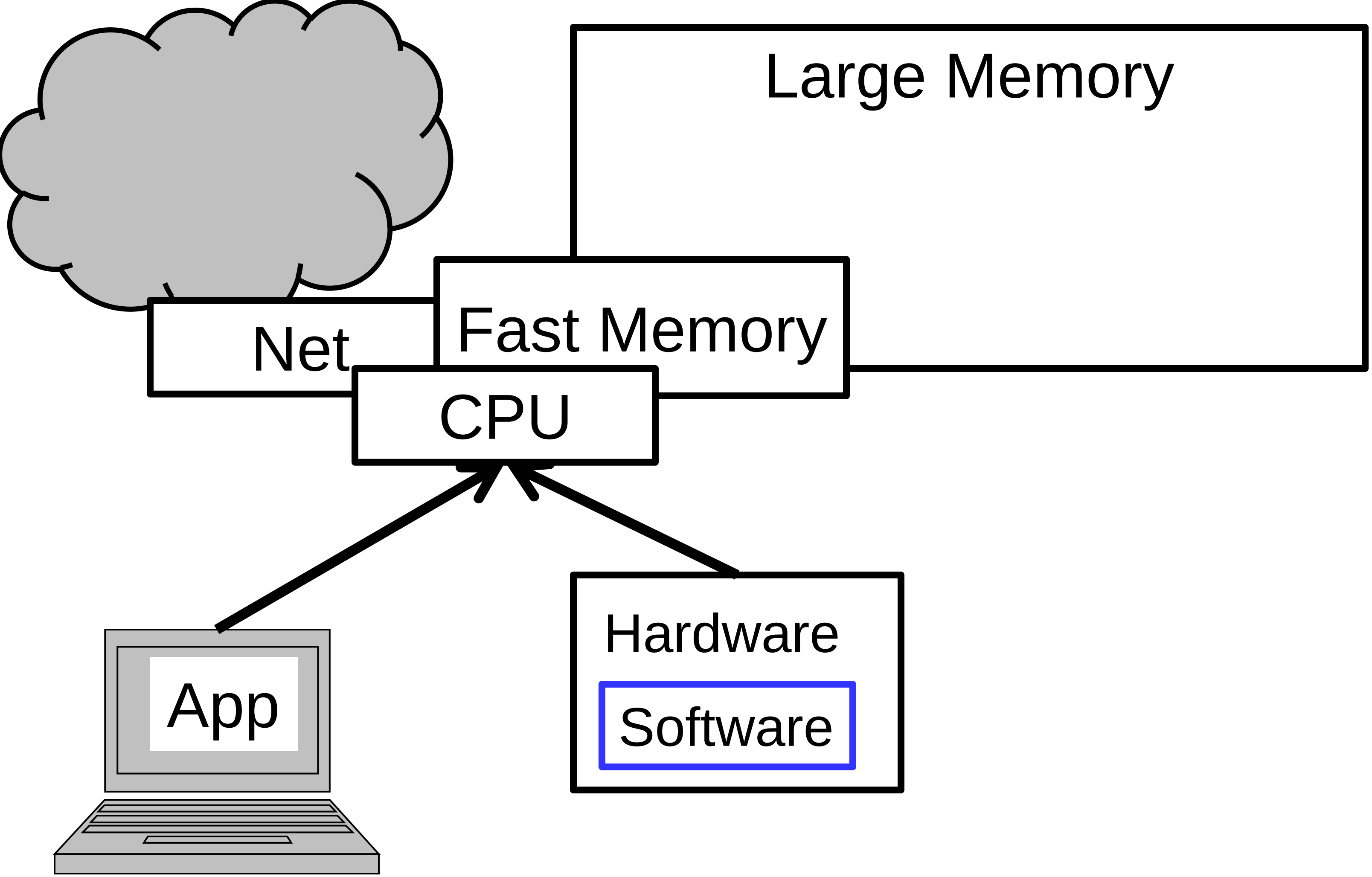


What you see: The hardware interface and the application function

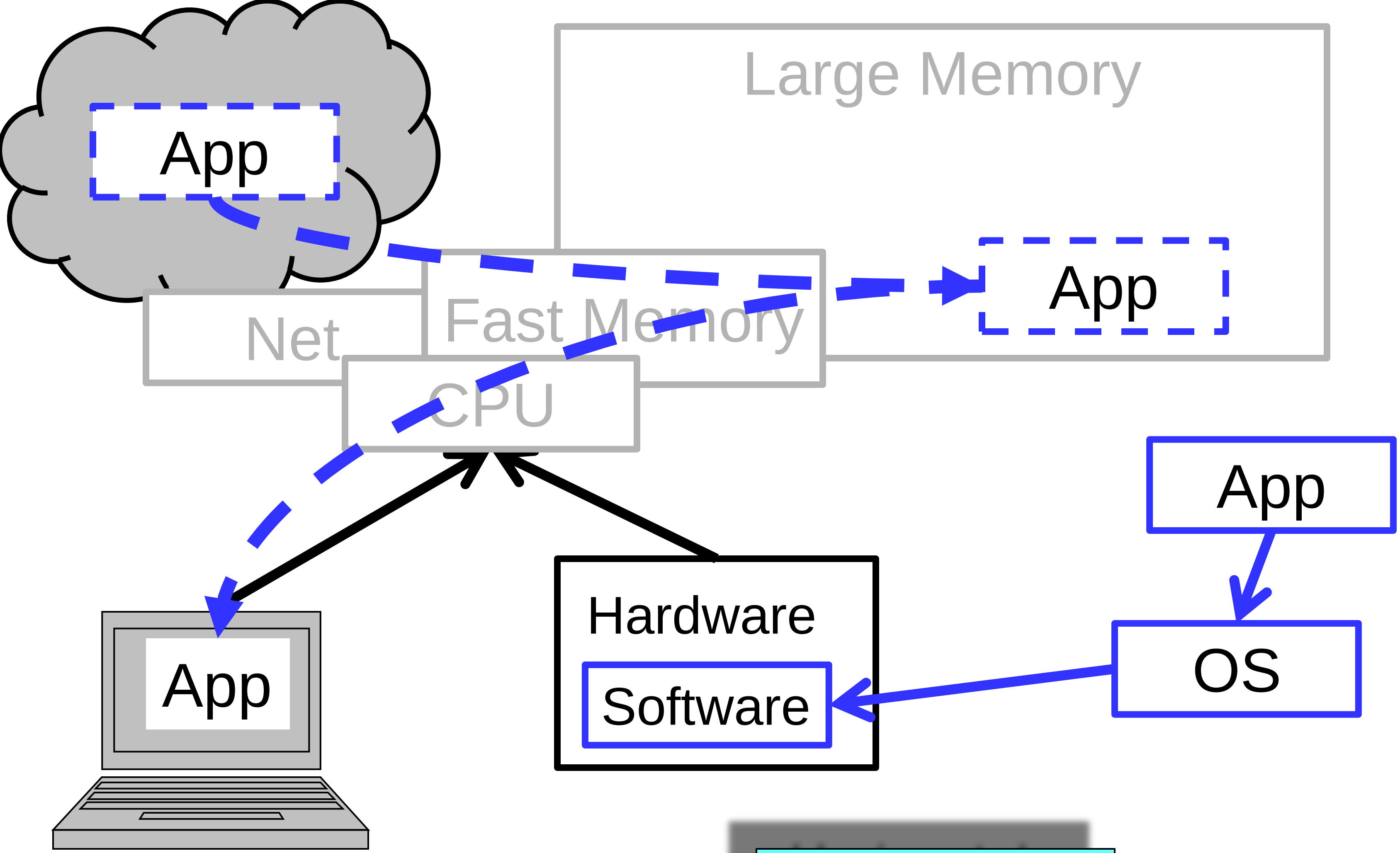


Need shared architecture and infrastructure (mostly hidden)


App
Net
Horizontal App Transfer
Virtual
App
OS
App
Need shared architecture and infrastructure (mostly hidden)
Amazingly Flexible/ Adaptable



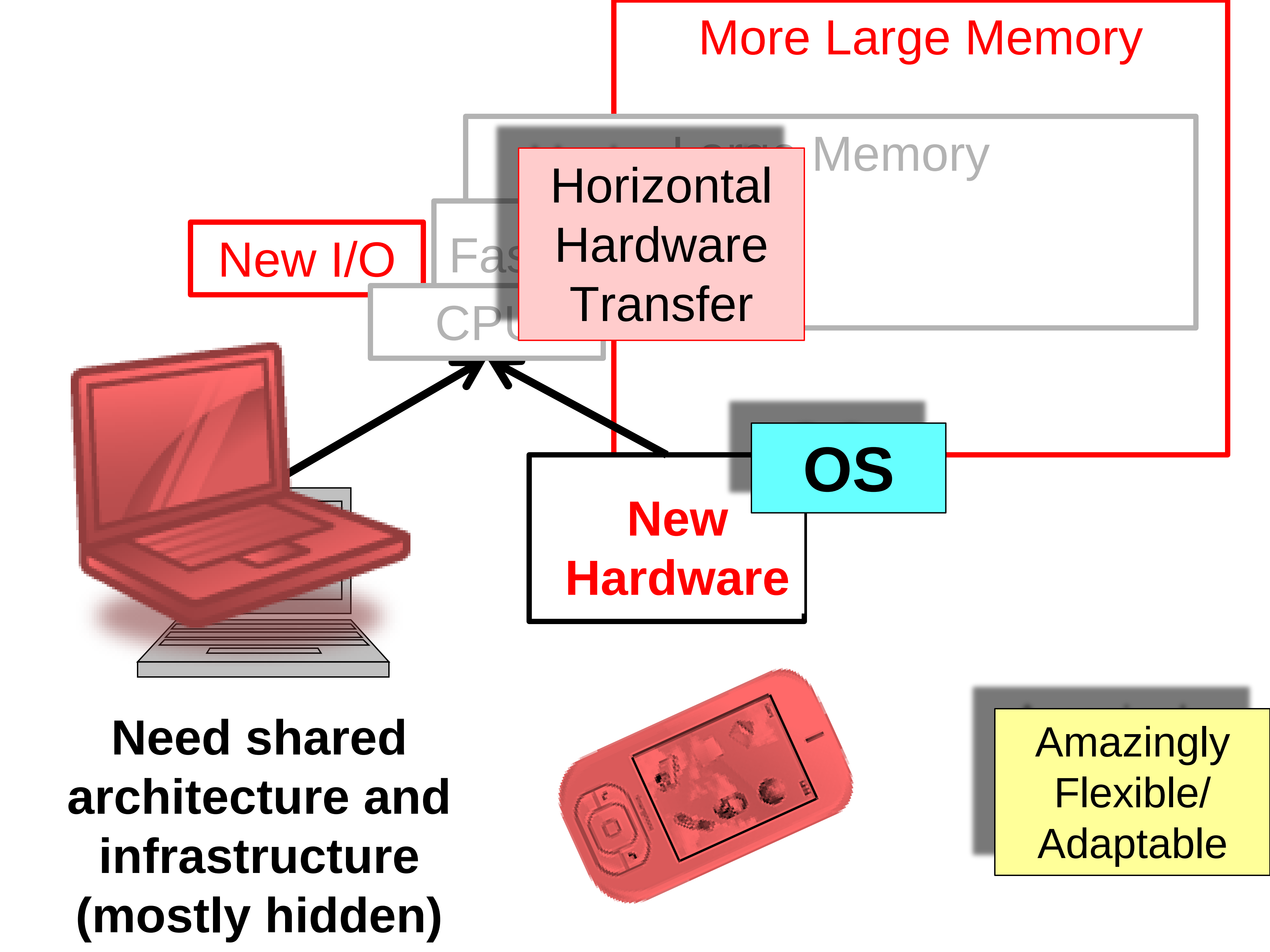
Need shared architecture and infrastructure (mostly hidden)



Need shared architecture and infrastructure (mostly hidden)

Horizontal App Transfer

Amazingly Flexible/ Adaptable



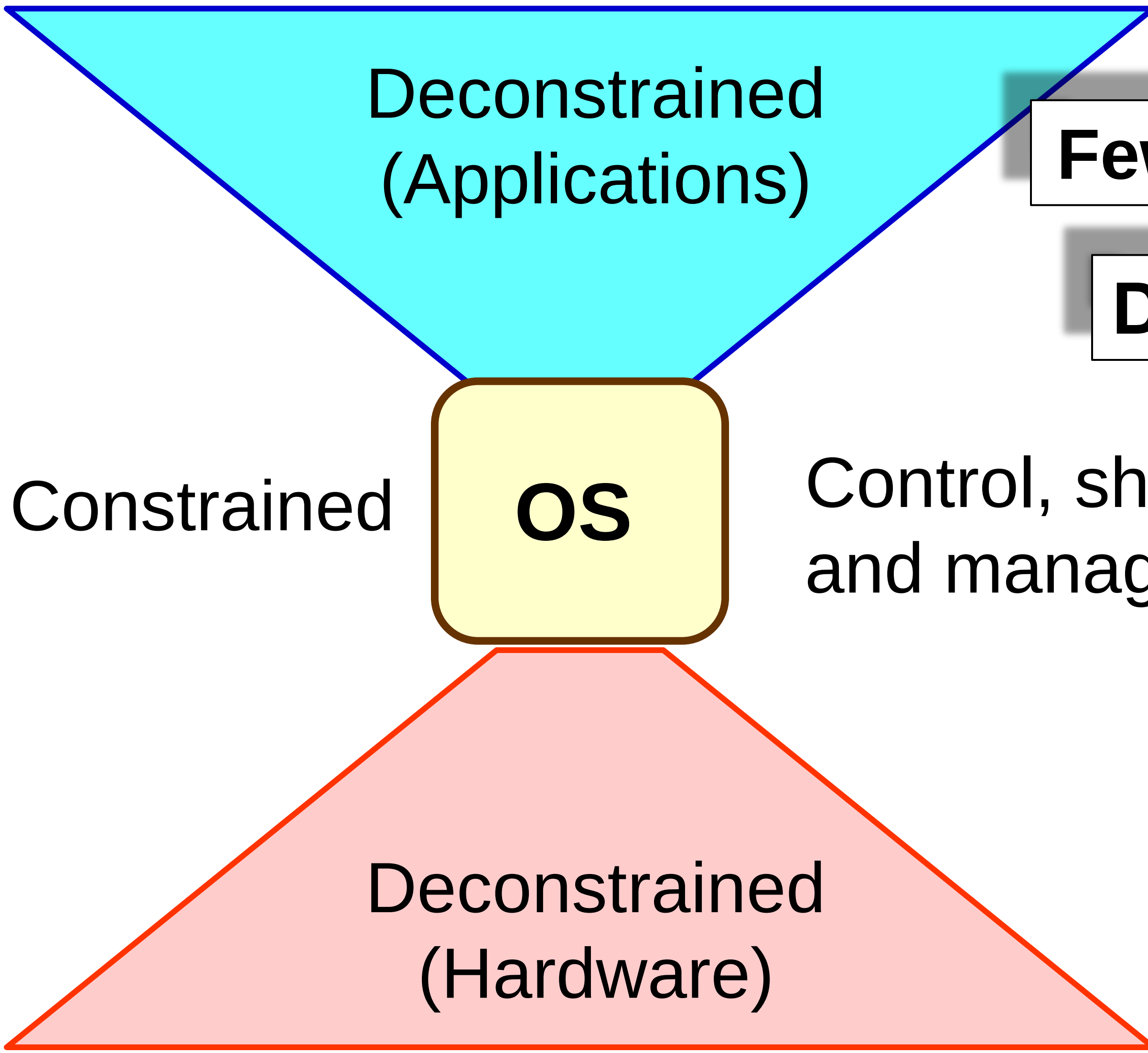
More Large Memory
Memory
New I/O
Horizontal Hardware Transfer
OS
New Hardware
Need shared architecture and infrastructure (mostly hidden)
Amazingly Flexible/ Adaptable

Layered architectures

Essentials

Few global variables

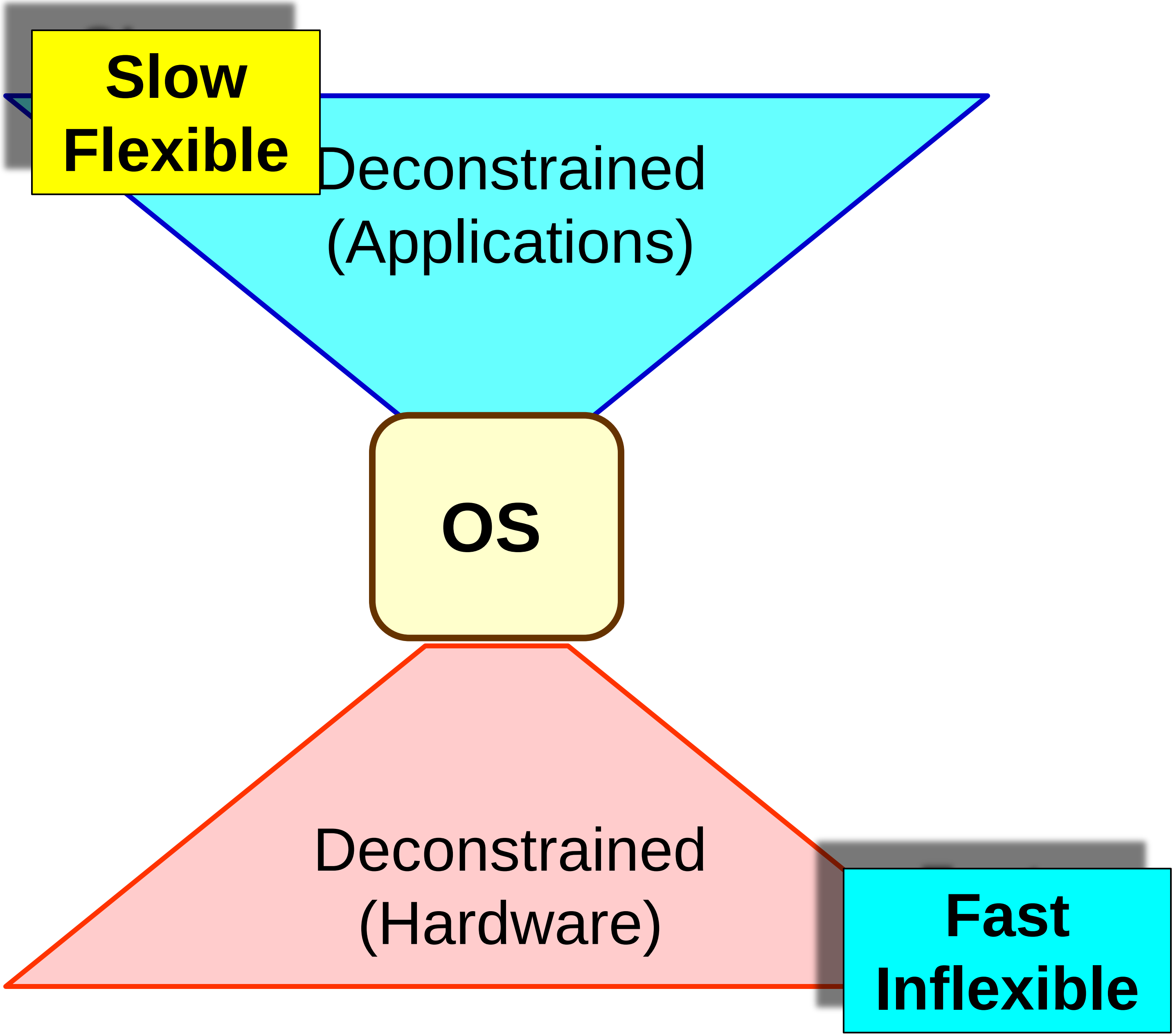
Don’t cross layers

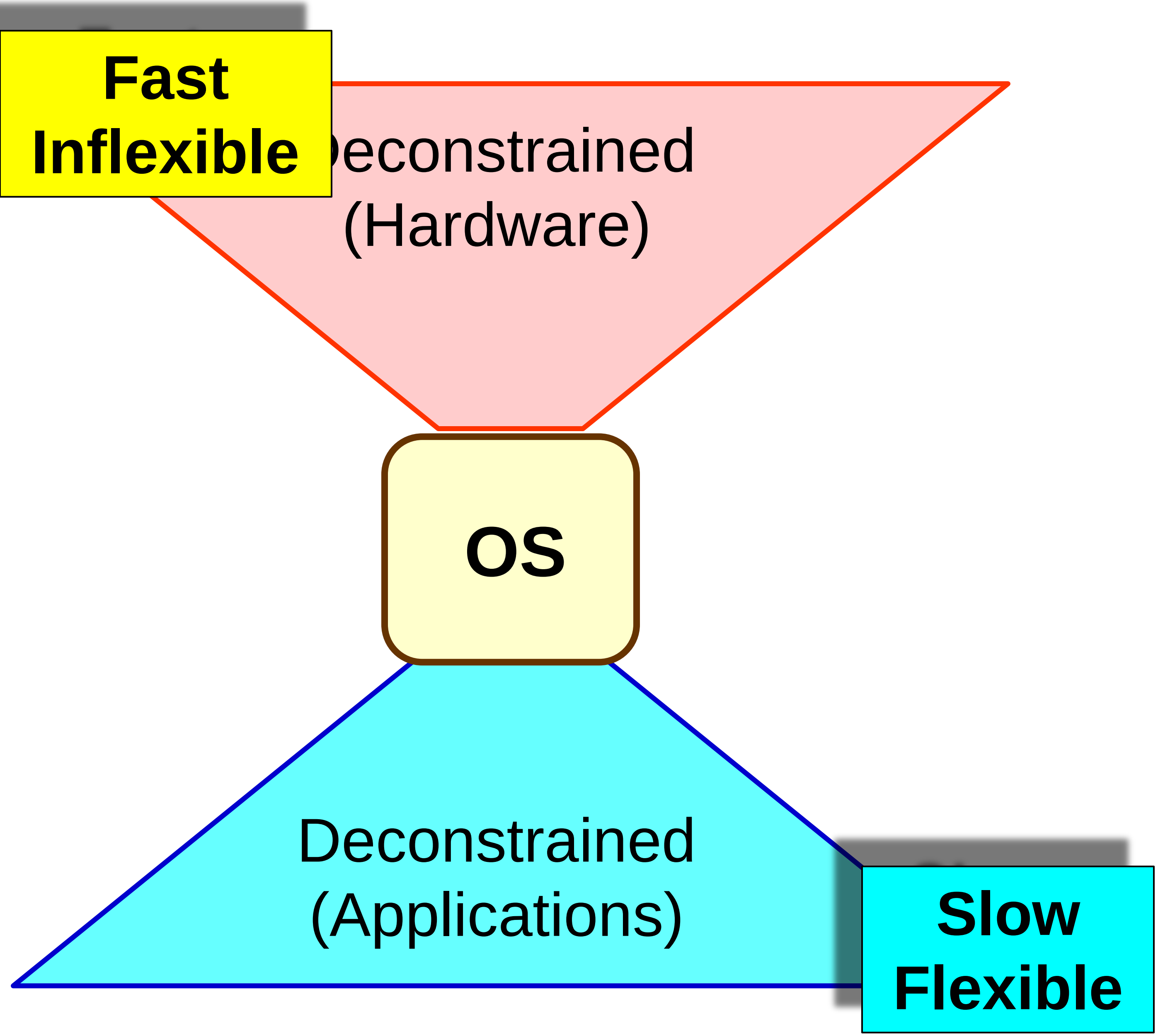


Constrained

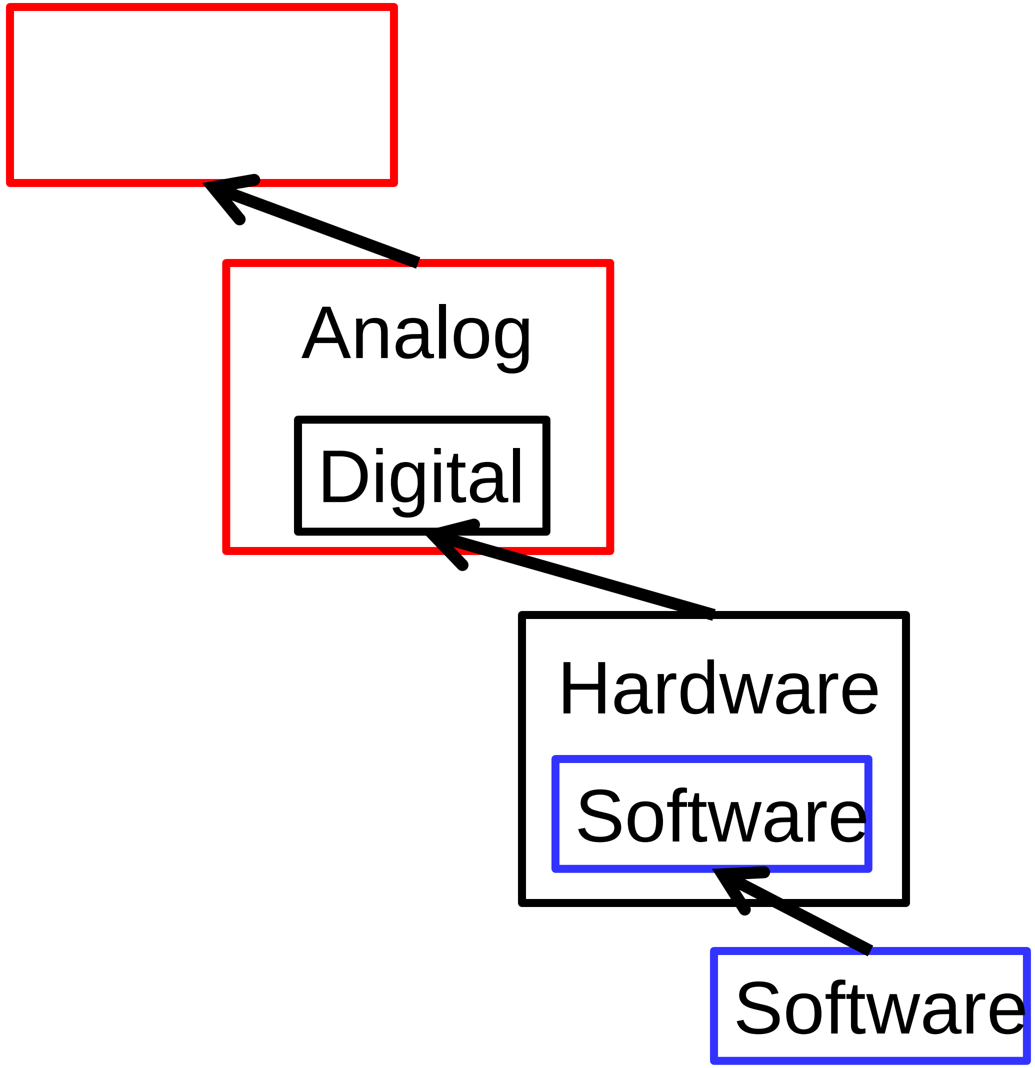
Control, share, virtualize, and manage resources

Processing
Memory
I/O


Slow
Flexible
Deconstrained
(Applications)
OS
Deconstrained
(Hardware)
Fast
Inflexible


Fast
Inflexible
Deconstrained
(Hardware)
OS
Deconstrained
(Applications)
Slow
Flexible

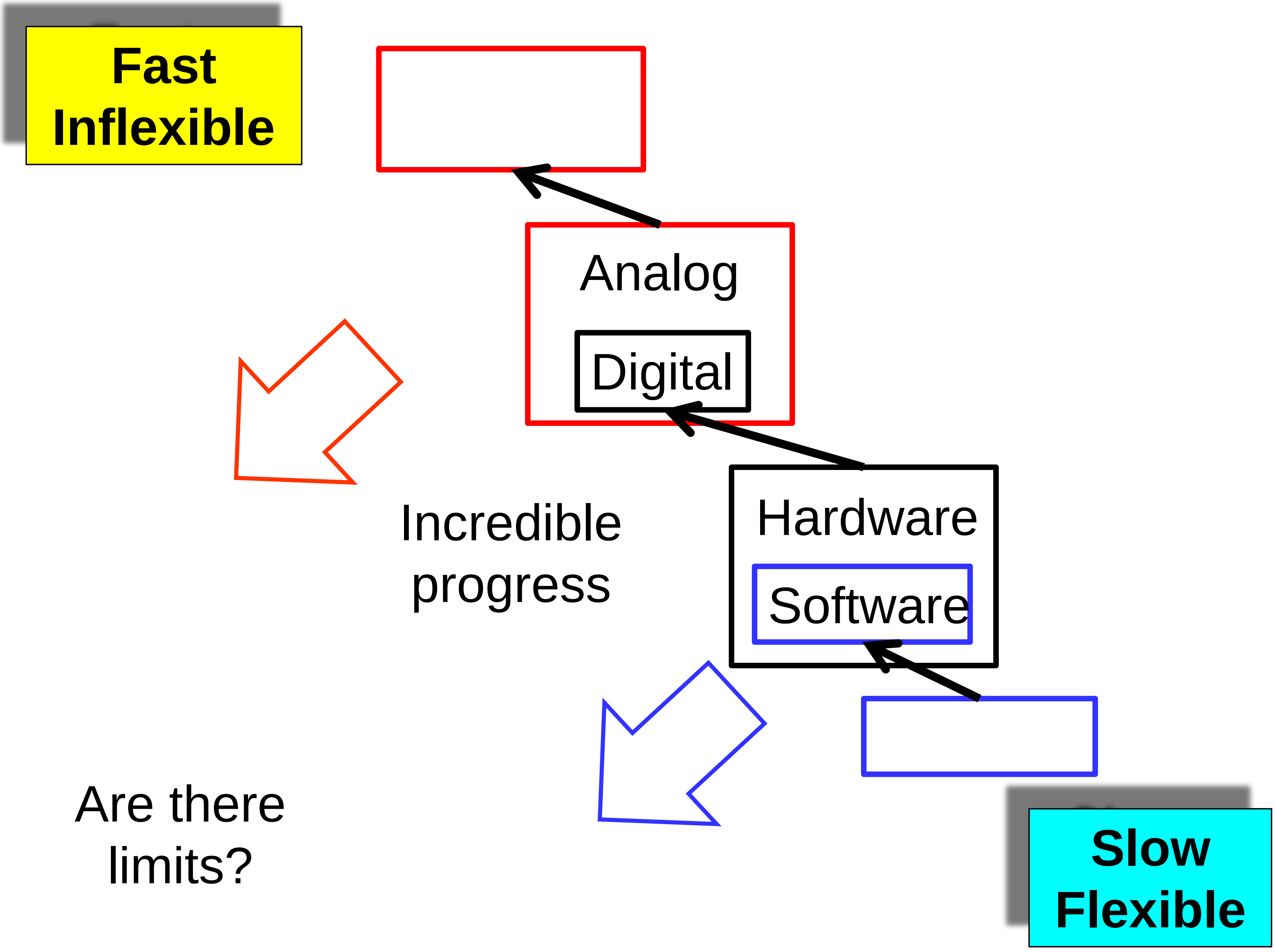
Fast
Inflexible

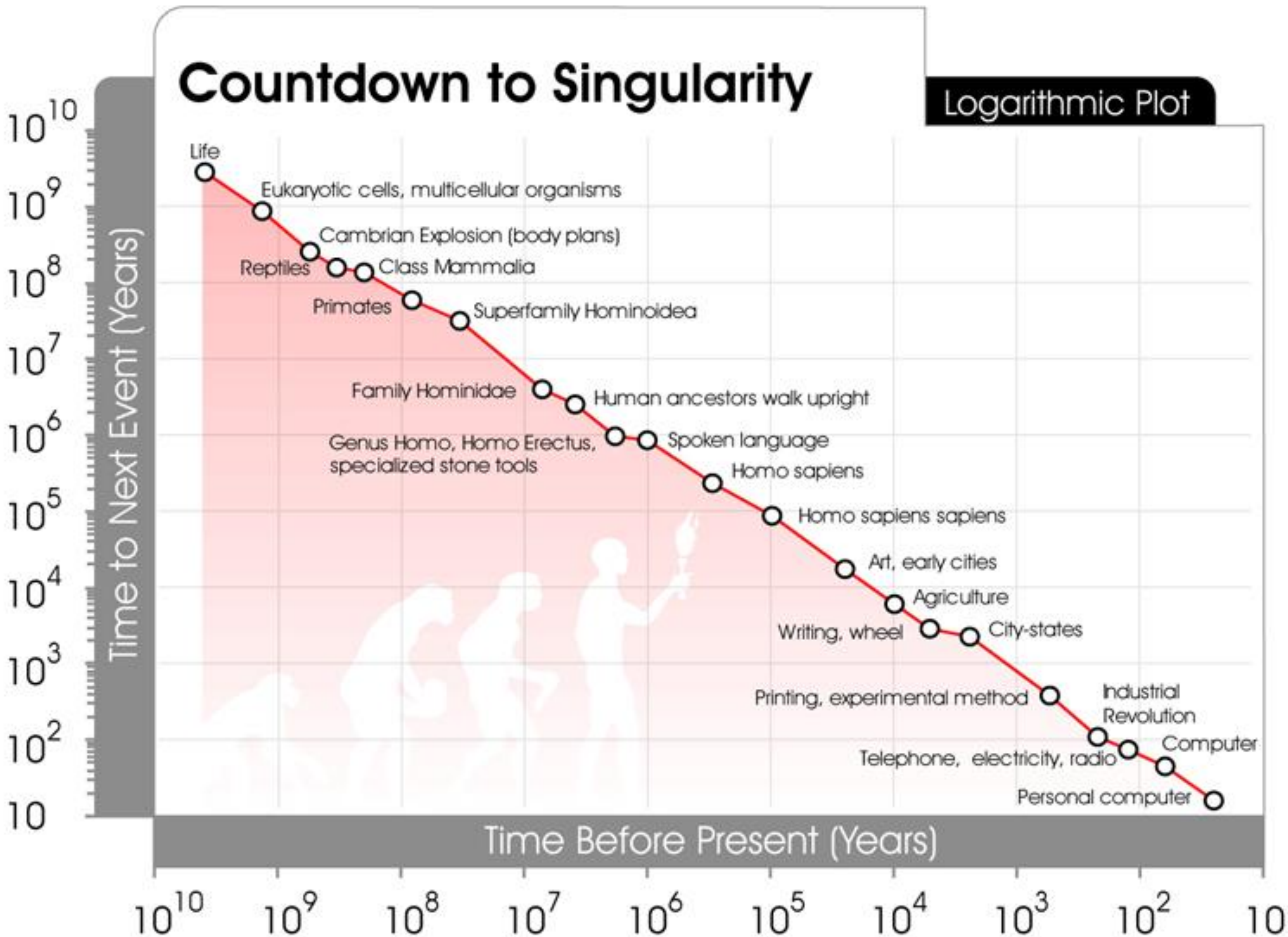


Tradeoff across multiple layers

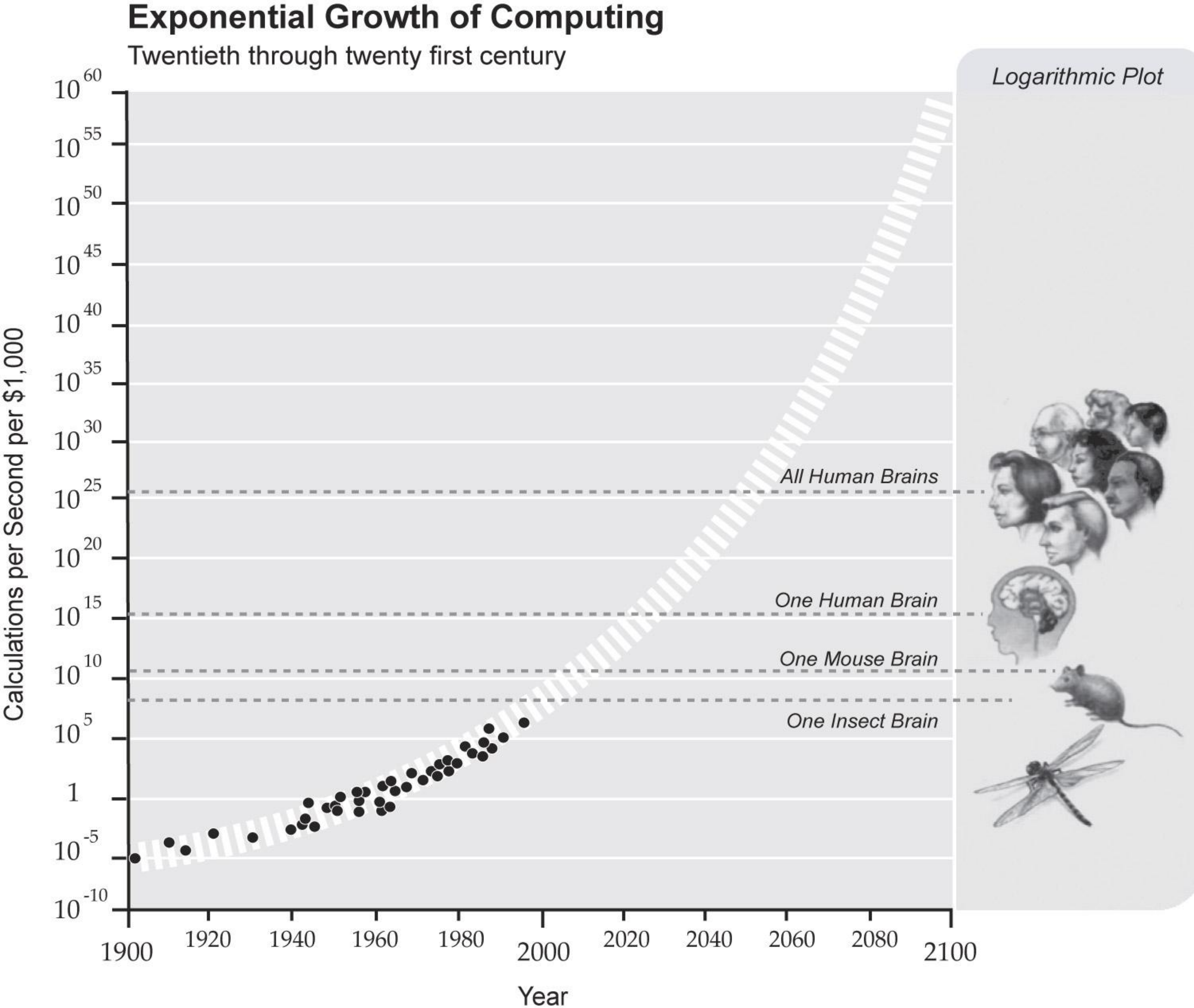
- Distributed
- Analog
- ASICs
- FPGA
- …
- Compiled
- Interpreted

Slow
Flexible


Fast
Inflexible
Analog
Digital
Incredible
progress
Hardware
Software
Are there
limits?
Slow
Flexible



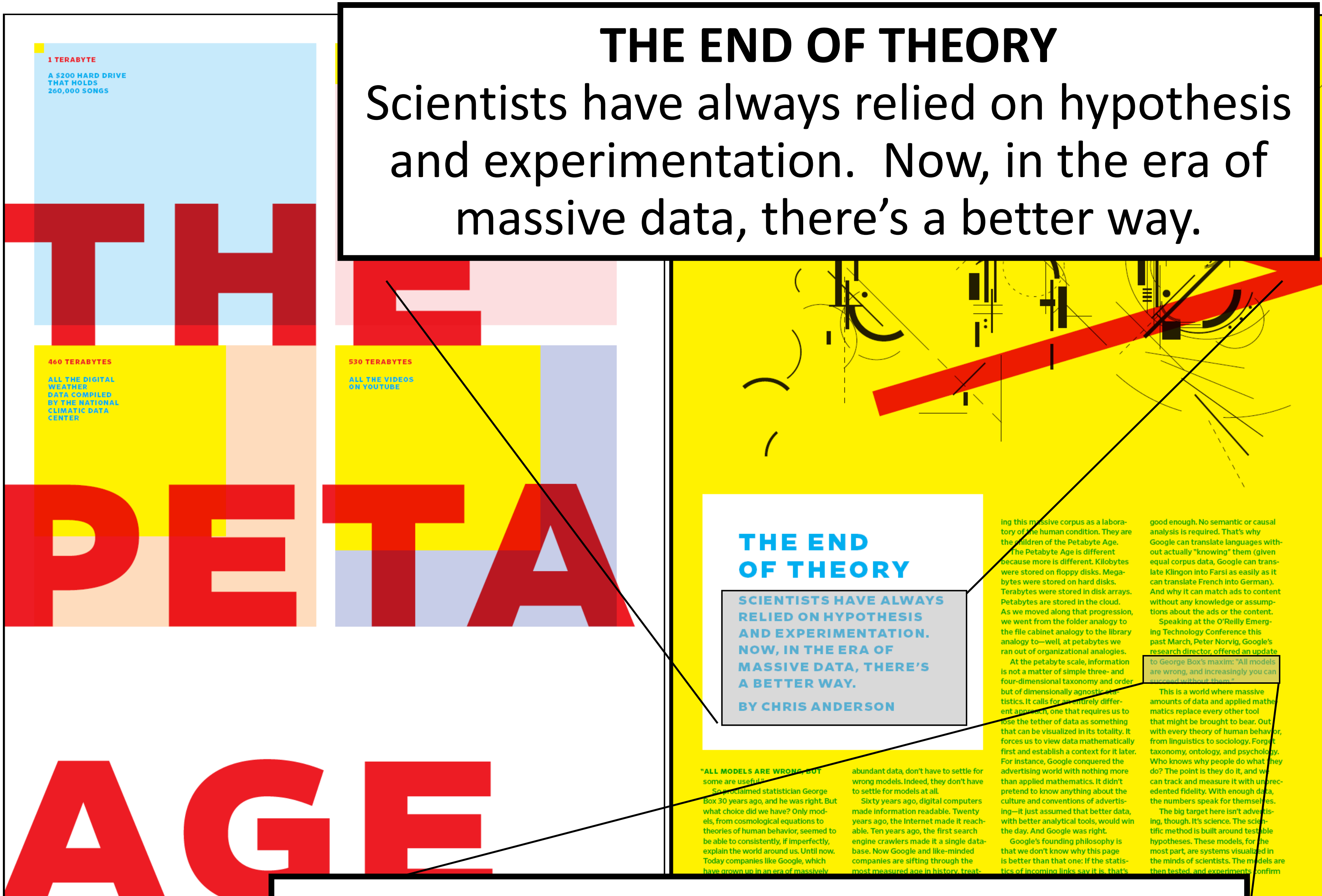
Countdown to Singularity
Logarithmic Plot
Time to Next Event (Years)
Time Before Present (Years)
Life
Eukaryotic cells, multicellular organisms
Cambrian Explosion (body plans)
Reptiles
Class Mammalia
Primates
Superfamily Hominoidea
Family Hominidae
Human ancestors walk upright
Genus Homo, Homo Erectus, specialized stone tools
Spoken language
Homo sapiens
Homo sapiens sapiens
Art, early cities
Agriculture
Writing, wheel
City-states
Printing, experimental method
Industrial Revolution
Telephone, electricity, radio
Computer
Personal computer
10^{10} 10^9 10^8 10^7 10^6 10^5 10^4 10^3 10^2 10



Exponential Growth of Computing
Twentieth through twenty first century
Logarithmic Plot
Calculations per Second per $1,000
10^60
10^55
10^50
10^45
10^40
10^35
10^30
10^25
10^20
10^15
10^10
10^5
1
10^-5
10^-10
All Human Brains
One Human Brain
One Mouse Brain
One Insect Brain
1900
1920
1940
1960
1980
2000
2020
2040
2060
2080
2100
Year

THE END OF THEORY

Scientists have always relied on hypothesis and experimentation. Now, in the era of massive data, there's a better way.



1 TERABYTE

A $200 HARD DRIVE THAT HOLDS 260,000 SONGS

460 TERABYTES

ALL THE DIGITAL WEATHER DATA COMPILED BY THE NATIONAL CLIMATIC DATA CENTER

530 TERABYTES

ALL THE VIDEOS ON YOUTUBE

THE PETA AGE

THE END OF THEORY

SCIENTISTS HAVE ALWAYS RELIED ON HYPOTHESIS AND EXPERIMENTATION. NOW, IN THE ERA OF MASSIVE DATA, THERE'S A BETTER WAY.

BY CHRIS ANDERSON

"ALL MODELS ARE WRONG, BUT some are useful."

So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? Only models, from cosmological equations to theories of human behavior, seemed to be able to consistently, if imperfectly, explain the world around us. Until now. Today companies like Google, which have grown up in an era of massively abundant data, don't have to settle for wrong models. Indeed, they don't have to settle for models at all.

Sixty years ago, digital computers made information readable. Twenty years ago, the Internet made it reachable. Ten years ago, the first search engine crawlers made it a single database. Now Google and like-minded companies are sifting through the most measured age in history, treating this massive corpus as a laboratory of the human condition. They are the children of the Petabyte Age.

The Petabyte Age is different because more is different. Kilobytes were stored on floppy disks. Megabytes were stored on hard disks. Terabytes were stored in disk arrays. Petabytes are stored in the cloud. As we moved along that progression, we went from the folder analogy to the file cabinet analogy to the library analogy to—well, at petabytes we ran out of organizational analogies.

At the petabyte scale, information is not a matter of simple three- and four-dimensional taxonomy and order but of dimensionally agnostic statistics. It calls for an entirely different approach, one that requires us to lose the tether of data as something that can be visualized in its totality. It forces us to view data mathematically first and establish a context for it later. For instance, Google conquered the advertising world with nothing more than applied mathematics. It didn't pretend to know anything about the culture and conventions of advertising—it just assumed that better data, with better analytical tools, would win the day. And Google was right.

Google's founding philosophy is that we don't know why this page is better than that one: If the statistics of incoming links say it is, that's good enough. No semantic or causal analysis is required. That's why Google can translate languages without actually "knowing" them (given equal corpus data, Google can translate Klingon into Farsi as easily as it can translate French into German). And why it can match ads to content without any knowledge or assumptions about the ads or the content.

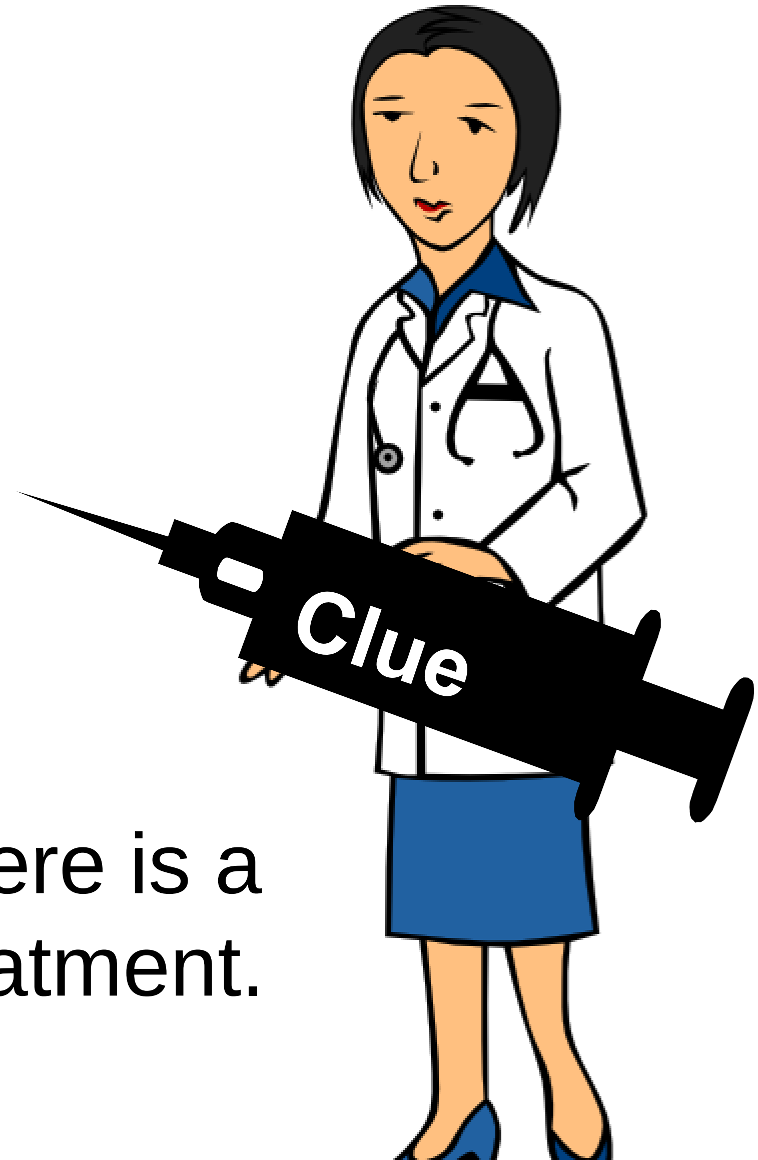
Speaking at the O'Reilly Emerging Technology Conference this past March, Peter Norvig, Google's research director, offered an update to George Box's maxim: "All models are wrong, and increasingly you can succeed without them."

This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.

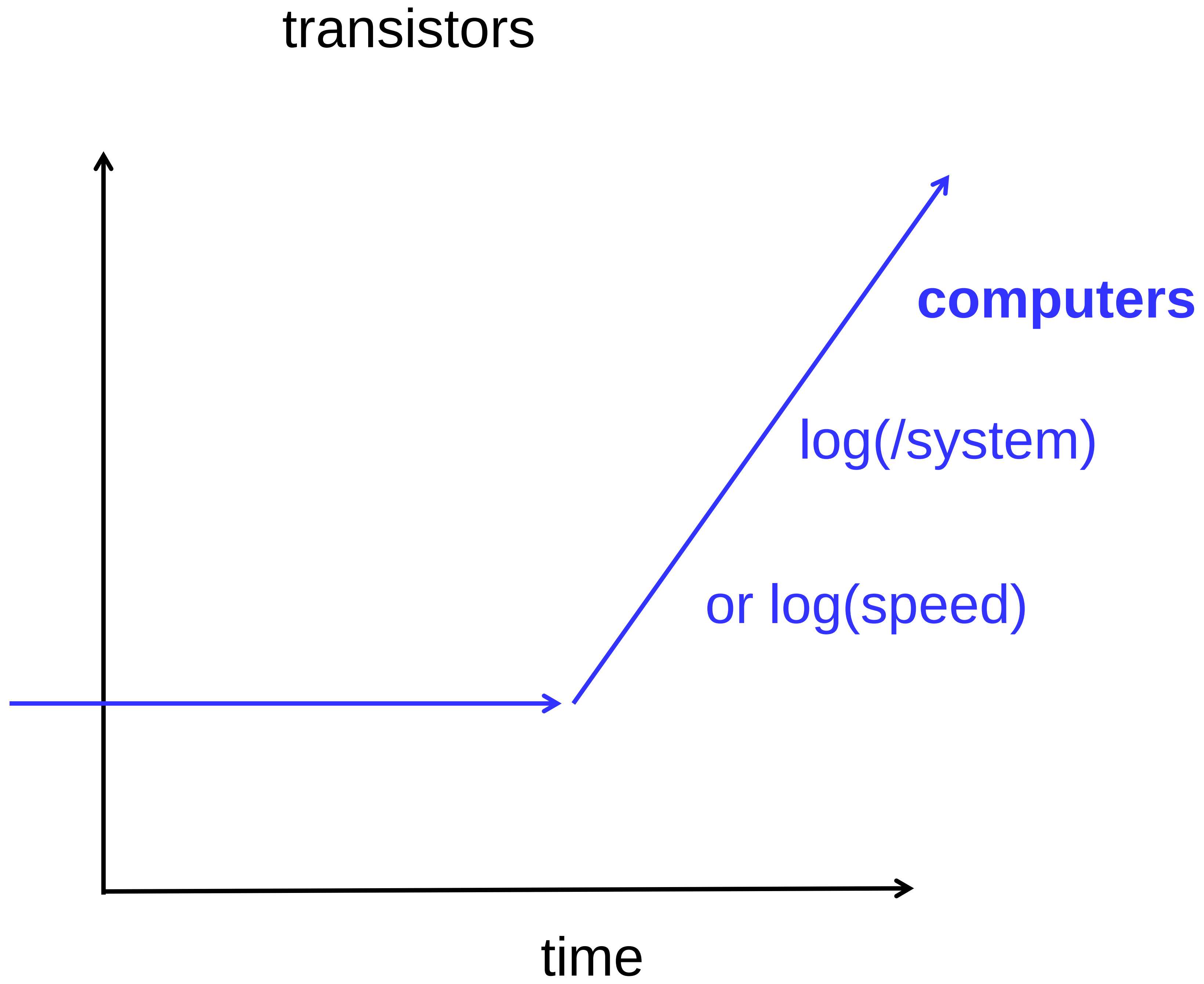
The big target here isn't advertising, though. It's science. The scientific method is built around testable hypotheses. These models, for the most part, are systems visualized in the minds of scientists. The models are then tested, and experiments confirm

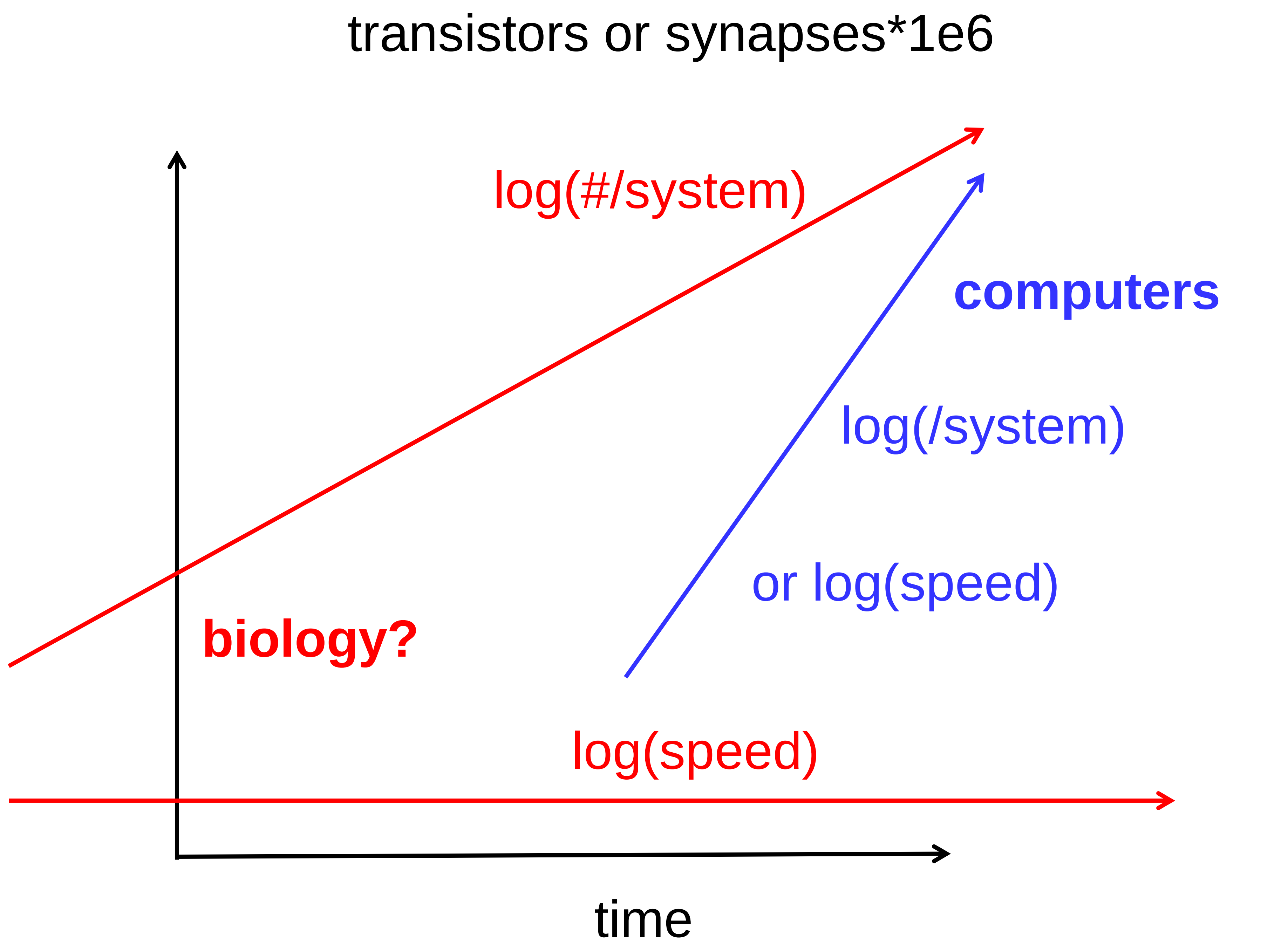
"All models are wrong, and increasingly you can succeed without them."

Save our children, stop

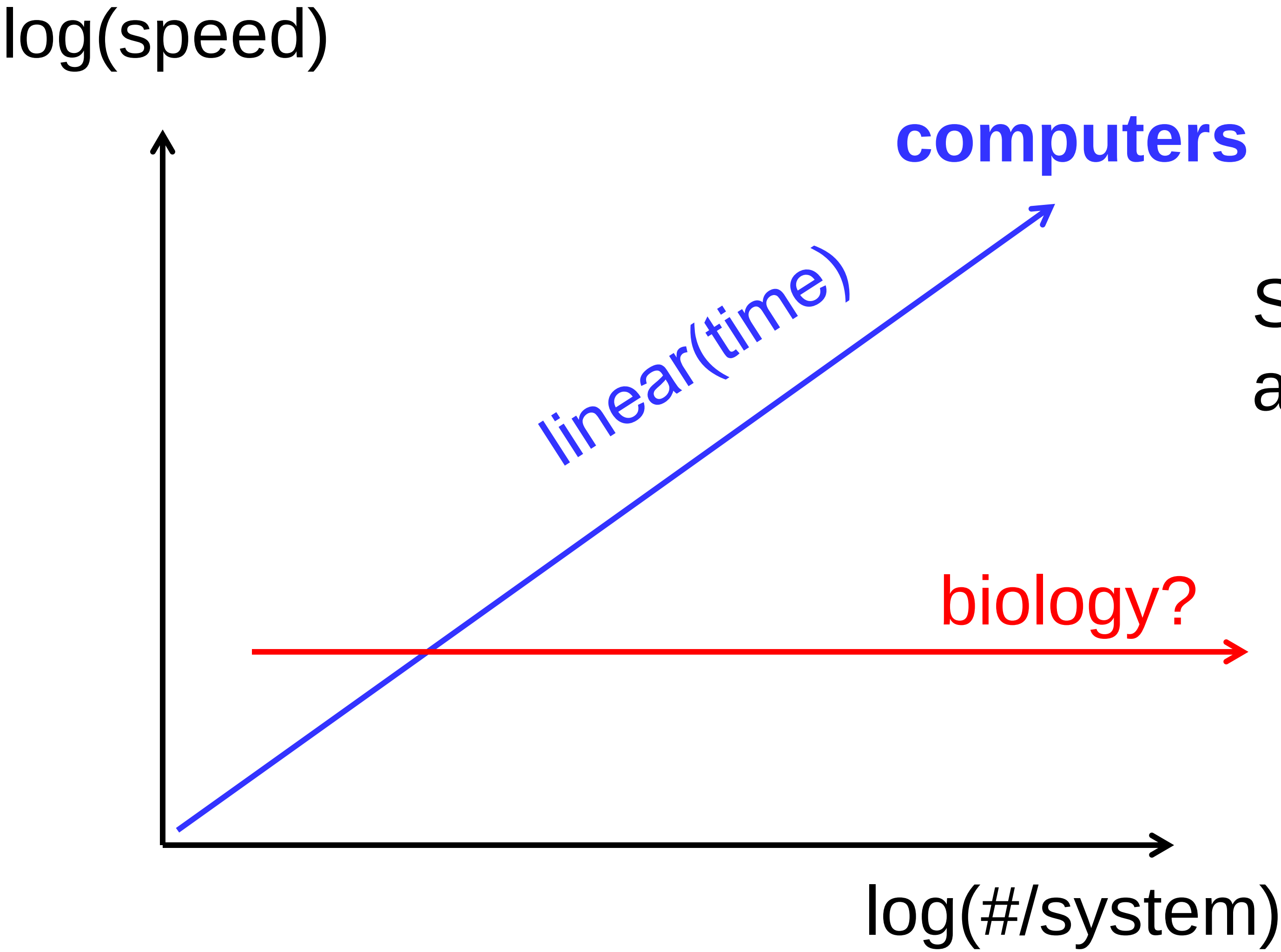


There is a treatment.


transistors
computers
log(/system)
or log(speed)
time

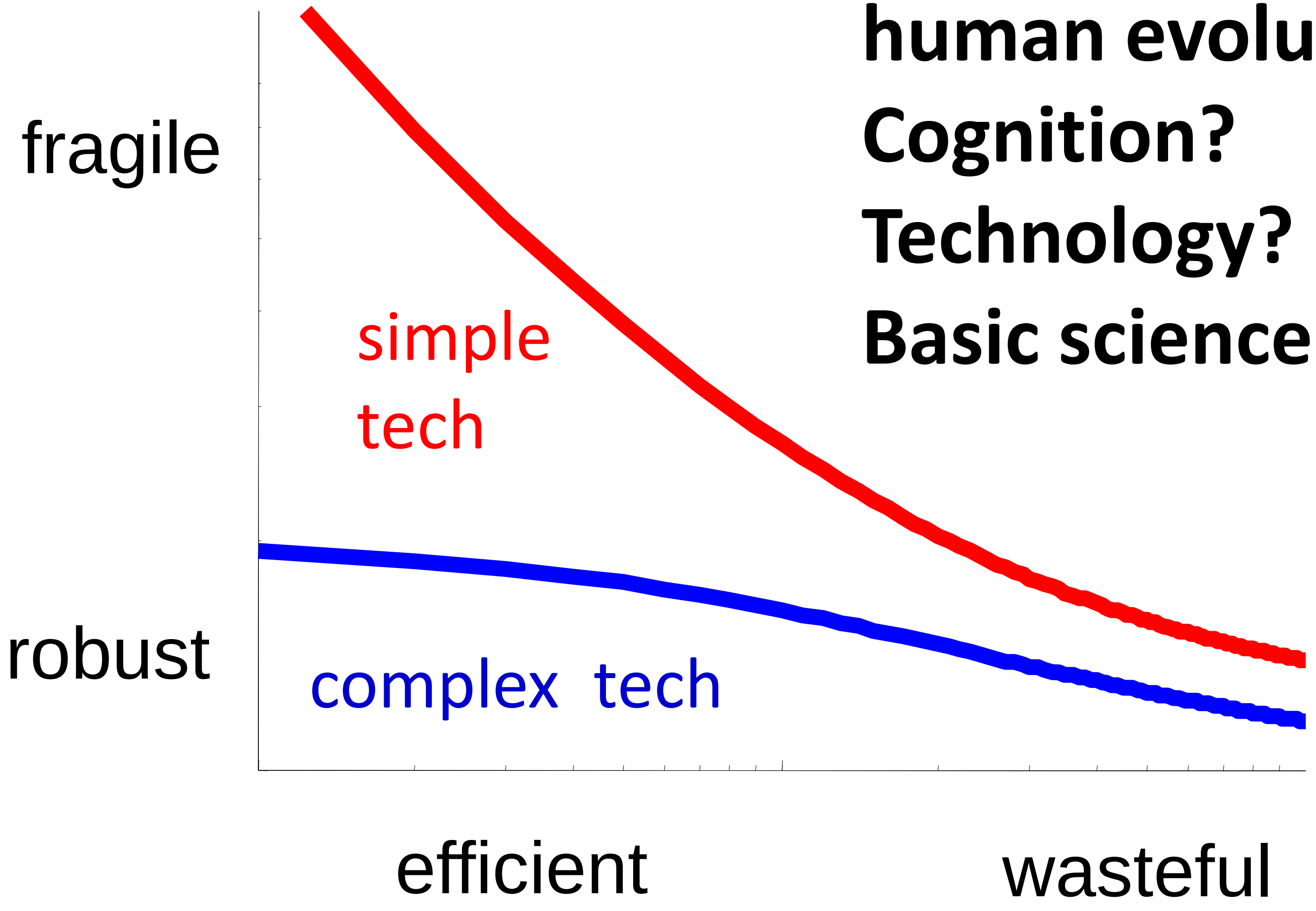

transistors or synapses*1e6
log(#/system)
computers
log(/system)
or log(speed)
biology?
log(speed)
time

transistors or synapses



So different architectures?

How general is this picture?



Implications for human evolution?
Cognition?
Technology?
Basic sciences?