Design Triangulation as a Path for Biological Inquiry

Sketches for a Method of Design-Enabled Biological Research

25 November 2020

Paul A. Nelson (paul.alfredp@gmail.com)
Senior Fellow, Discovery Institute, www.discovery.org/csc
Adjunct Faculty, MA Program in Science & Religion
Biola University, www.biola.edu

This document contains lecture materials (with updates) that I have presented over the past 15 years at various intelligent design study and research venues. These are preliminary ideas and need refinement via critique, which I welcome (paul.alfredp@gmail.com).

Outline of presentation

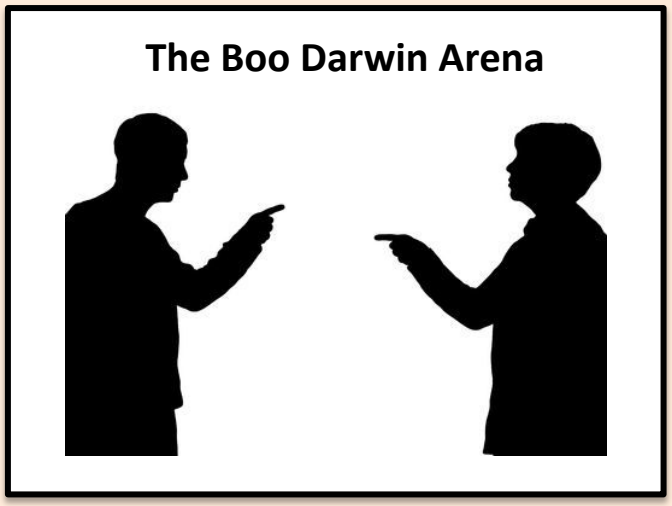
Bibliography on request

Caveat #1. Because these are lecture materials, not every slide is fully explained (as I would normally provide, in person, to an audience). Readers with questions about particular slides should email me.

Caveat #2. In what follows, I discuss the work of many different scientists and philosophers, only a few of whom would endorse intelligent design. *Thus, the presence of the ideas or arguments of any person in this presentation should* ***not*** *be taken as a claim that that person supports design.*

I. How to investigate biology when everyone says it's impossible (from a design perspective, that is)

Intellectual habits, like all habits, die hard. Someone who was initially persuaded of intelligent design in a largely confrontational context (left) – call this the "Boo Darwin Arena" for short – may be slow to realize that the task of science does not end with debunking another person's hypothesis. **Sooner or later one must explain the data for oneself** (right).



To be sure, challenging the received view is often central to scientific advance. Darwin said the *Origin of Species* was "one long argument." Galileo wrote the *Dialogue Concerning the Two Chief World Systems*, challenging Aristotle and Ptolemy, in Italian, not Latin, to achieve the widest readership.

But the discovery of X-rays by Röntgen (1895) was not an argument, nor was the elucidation of the molecular structure of DNA by Watson and Crick (1953). ***Imagine a world in which one has no one left to confront, but only data to explain.***

Nonetheless, the project of "conjectures and refutations" in science (Popper 1963) requires interlocutors. So let's suppose we have been visited by a team of highly advanced, silicon-based artificial intelligences, sent by their carbon-based makers from a station permanently orbiting Alpha Centauri.

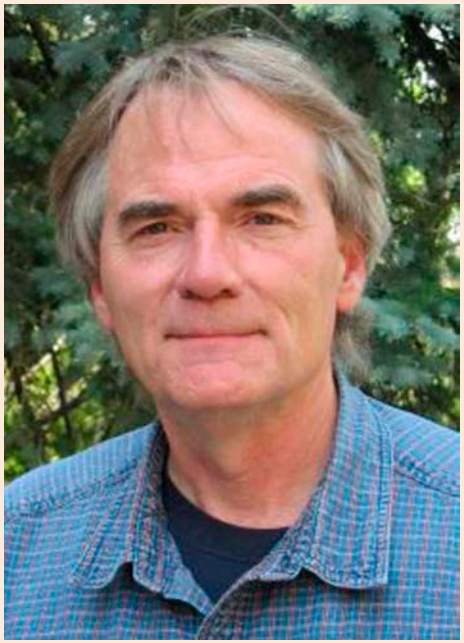
The Skeptics

Knowing themselves to have been designed, these intelligences harbor no *a priori* opposition to design hypotheses.* The evidence on Earth must be weighed in light of all the possibilities.

But these entities are also very careful and critical (not to mention snarky). When we ask their names, they say "Call us 'the Skeptics' – we understand how easily science can go astray. But lead on."

So the Skeptics will accompany us on our journey. They will provide critical counterpoints and questions.

*To quote an apropos line from the movie *A.I. Artificial Intelligence*, spoken by the Mecha (robot) Gigolo Joe to the artificial child David: "The ones who made us are always looking for the one who made them."
And thanks to this movie for the images of artificial intelligences.



Professor Scott Minnich
Microbiology
University of Idaho
USA

When I first met Scott Minnich in the mid-1990s, when I was still a graduate student, he told he something I have never forgotten:

"Paul, none of this intelligent design debate should be all that controversial. **The reality, whether anyone is conscious of it or not, is this: most molecular biologists are *de facto* design theorists already."**

And, he added, they have been doing that for a very long time indeed.

But how could that be possible – even as a *de facto* practice – when design was widely seen as intrinsically unscientific, employing a cause (i.e., a transcendent mind) understood to be unobservable in principle?

In this talk, I explain what Scott meant, and why it matters, using an idea called "design triangulation."

Most biologists who reject ID do so, not because they think the idea is ***false*** – that, they say, would require ID to make novel testable predictions – but because they see design as ***empirically sterile.*** ID in their view is almost entirely polemics, "Boo Darwin!" but little or nothing beyond that.

Compare Bacon's (1620) indictment of Aristotelian reasoning: "[T]hat wisdom which we have derived principally from the Greeks is but like the boyhood of knowledge, and has the characteristic property of boys: it can talk, but it cannot generate [i.e., conceive any offspring]; for it is fruitful of controversies, but barren of works."

Etiology: how did X *come to be?*

Two big worries about tumbling into unproductive, or downright unsound, paths of scientific inquiry:

Always start with the log in one's own eye, right?

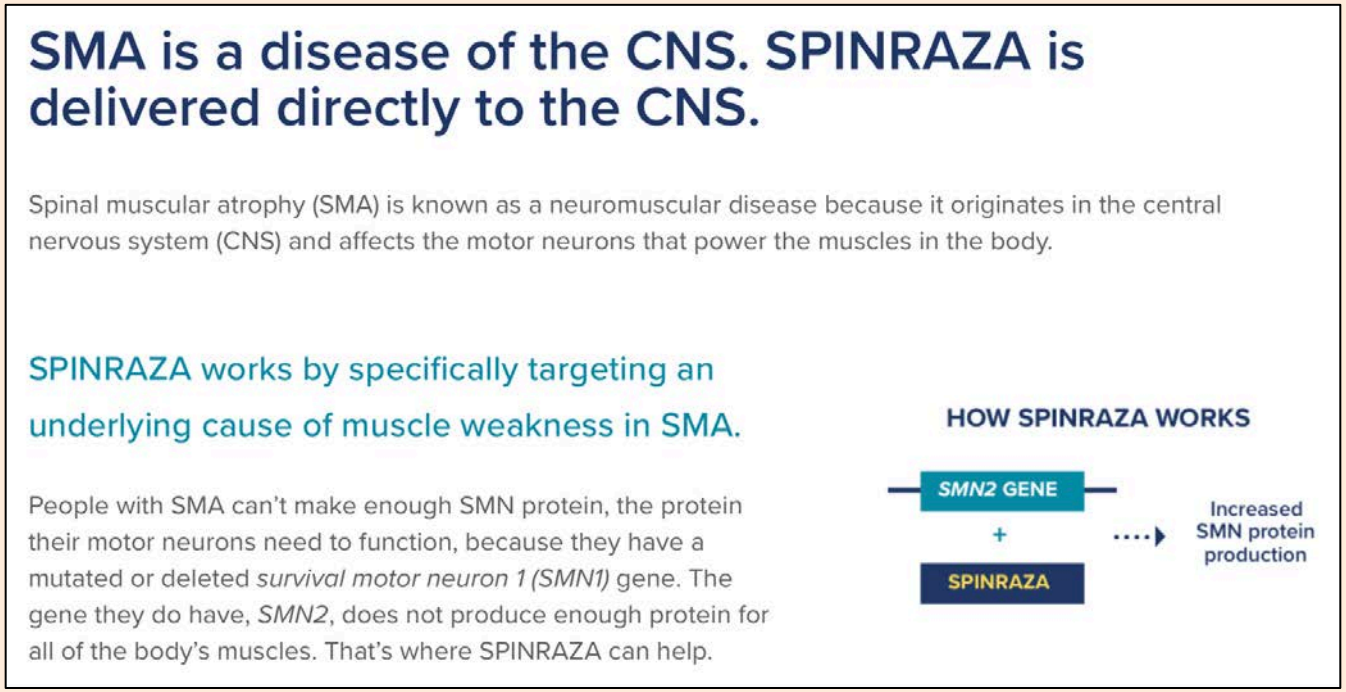
So let's consider the ID-related worry first.

What is Paul's overwhelming worry about ID in biology?

Biological *mechanisms* are *real*, and it is the task of biologists to find and understand them.

Every time I undergo a complicated medical procedure, which solves the problem I face, or listen to my physician wife describe her successful therapies, using targeted medications, I thank God (literally) for our mechanistic knowledge and the hard work done by others to find and apply that knowledge.

I have a young relative, four years old, who suffers from spinal muscular atrophy (SMA) in its severest form (SMA1). Ten years ago, he wouldn't have lived to four.



Yet because of the work of Dr. Ravindra Singh at U-Mass Medical School, Prof. Adrian Krainer at Cold Spring Harbor Laboratory, and many others, ***to understand mechanisms of gene splicing,*** and therefore develop the medication *nusinersin* (marketed as SPINRAZA), my young relative – who is verbally gifted – can spell words like "umbilicus" on a magnetic letter board, smile, and live. *Knowledge lessens suffering*, as Louis Pasteur understood.

You, as an ID explorer, may share this same nagging worry about finding mechanisms. It's common.

As a consequence, design theorists, despite their best intentions, often cannot help thinking in reductive & material terms where biological explanation is concerned – because it seems only there that progress can be made. We do what we can, and only what we can.

We, too, are children of the Scientific Revolution and the Enlightenment.

So – to consider the other horn of the dilemma – what is Paul's overwhelming worry about physicalist reductionism?

Organisms are demonstrably irreducible, and biologists ignore this reality at their peril.

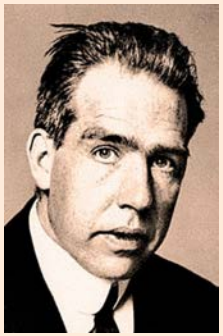
Multiple lines of evidence, accumulated over many decades (to be discussed later in this presentation), show this unmistakably.

The Skeptics

"Organisms are *demonstrably* irreducible? Evidence shows this *unmistakably*? Sounds like question-begging to us."

Paul: Well, it *would* be question-begging, if we never got a closer look at the evidence. But may I ask for your patience? One step at a time.

In the interim, there is nothing especially controversial about the irreducibility of organisms. (The room's temperature only rises when one asks what irreducibility implies for origins.) Consider, for instance, Bohr's (1933, 458) argument that organisms are "elementary facts":



Niels Bohr
1885-1962

"On this view, the existence of life must be considered as an elementary fact that cannot be explained, but must be taken as a starting point in biology, in a similar way as the quantum of action, which appears as an irrational element from the view of classical mechanical physics..."

How to sail between the Scylla of a paralyzing ID holism...

...and the Charybdis of physicalist reductionism.

II. Foresight, Causal Circularity, Parts and Wholes

The concept of ***foresight***,
the biological pattern
of ***causal circularity***, and the
causal primacy of the organism,
will be the main dimensions of
biological explanation addressed by
design triangulation in this presentation.

Let's start with foresight and causal circularity.

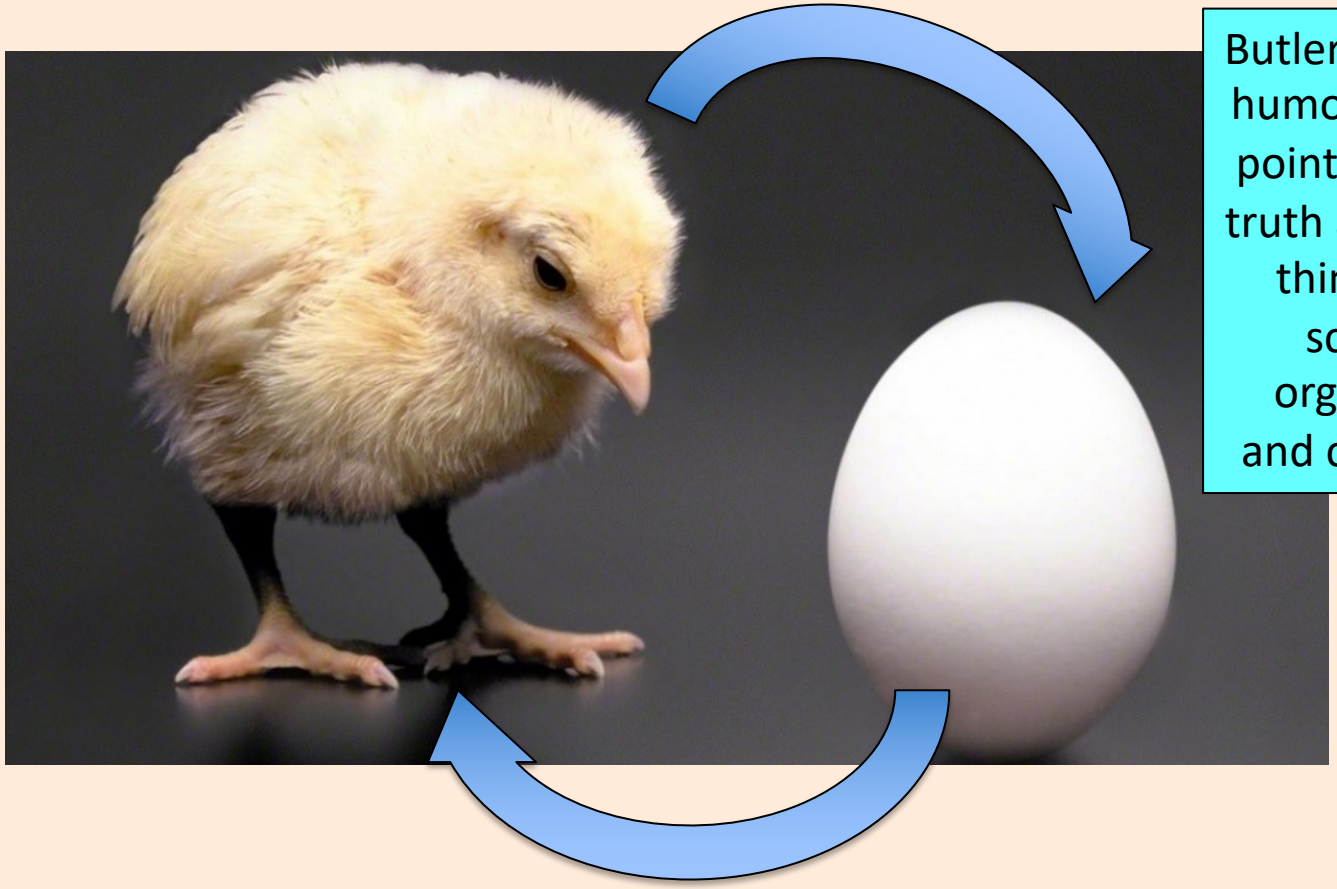
Foresight: the mental or conceptual representation of a function or system, prior to its physical realization.

Causal circularity: the origin of, or pathway to, object X, *requires the prior existence of X*:

"To make X, you need X."

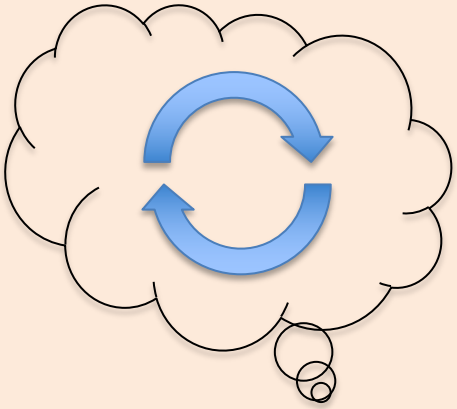
“A hen is only an egg’s way of making another egg.”

Samuel Butler, 1878



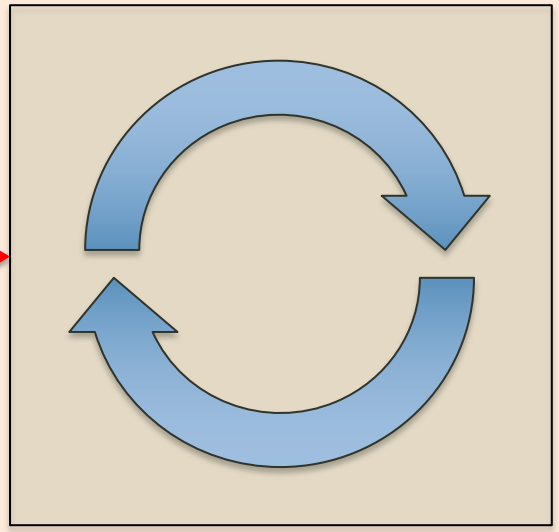
Butler’s remark is humorous, but it points to a deep truth about living things, at all scales of organization and complexity.

The logical relationship of causal circularity, foresight, and design.



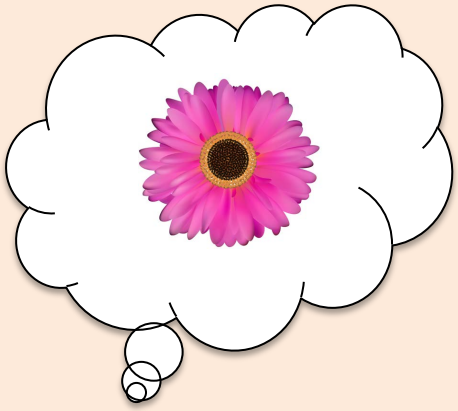
foresight

design



causal circularity

The *mental construct of the whole* is causally primary. Mind leads, seeing the target; realization follows.

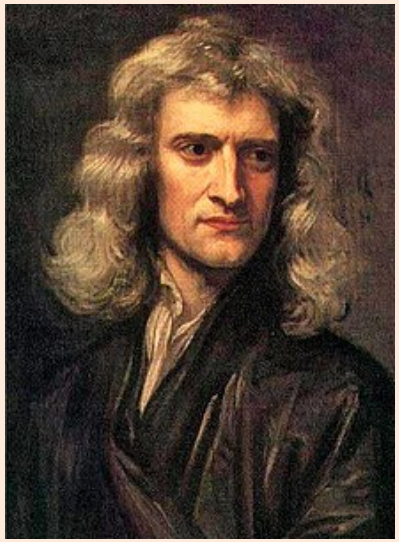


foresight

design

living things

Until the mid-nineteenth century, this mode of explanation (for biology) was rationality itself.



Isaac Newton
(1643-1727)

"How came the bodies of animals to be contrived with so much art, and for what ends were their several parts? Was the eye contrived without skill in Opticks, and the ear without knowledge of sounds?" (Query 28, Book III of the *Opticks* [1730])

"Rule 1: We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances."

"**Rule 2:** Therefore to the same natural effects we must, as far as possible, assign the same causes." (Newton, *Principia*)

Immanuel Kant
(1724-1804)

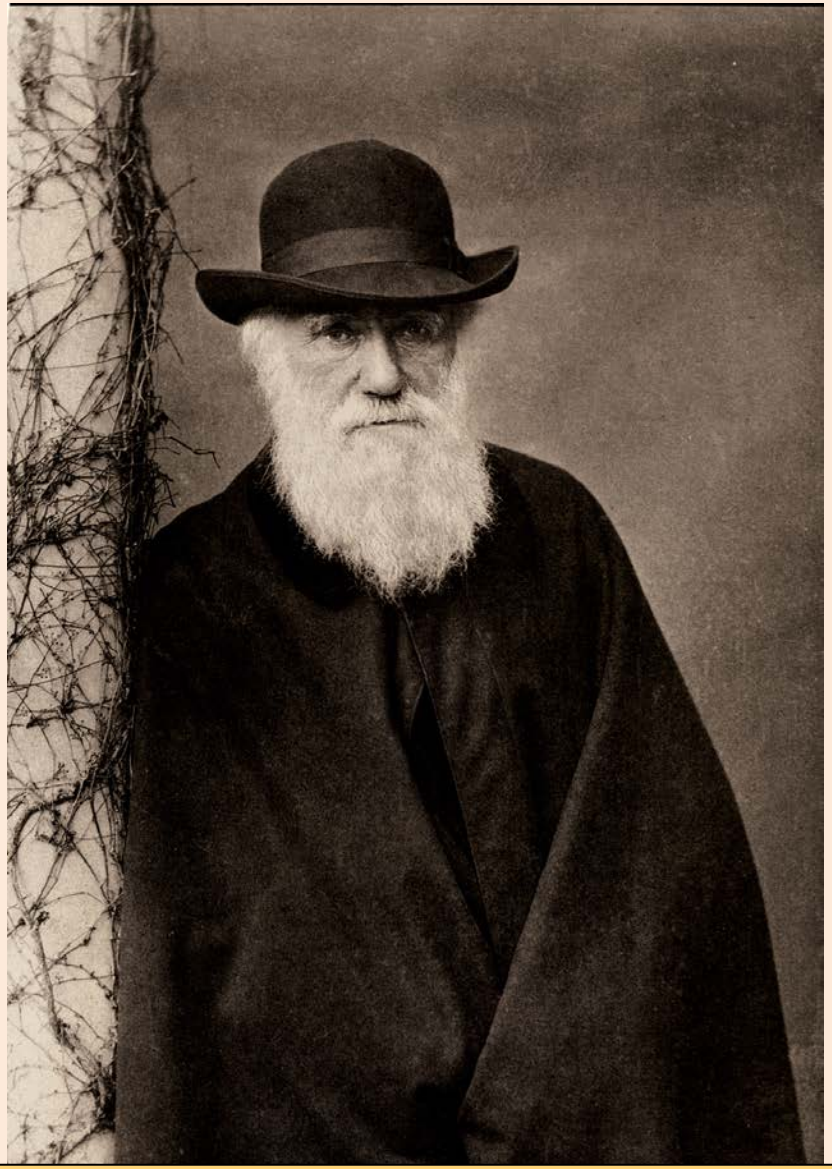
"...we cannot adequately cognize, much less explain, organized beings...according to mere mechanical principles of nature, and we can say boldly it is alike certain that ***it is absurd for men...to hope that another Newton will arise in the future who will make comprehensible by us the production of a blade of grass*** according to natural laws which no design has ordered." (*Critique of Judgment*, 1790; emphasis added)

For Kant, the "causal circularity" of organisms simply *was* their *defining characteristic*, entirely beyond the reach of strictly physical explanation.

Immanuel Kant
(1724-1804)

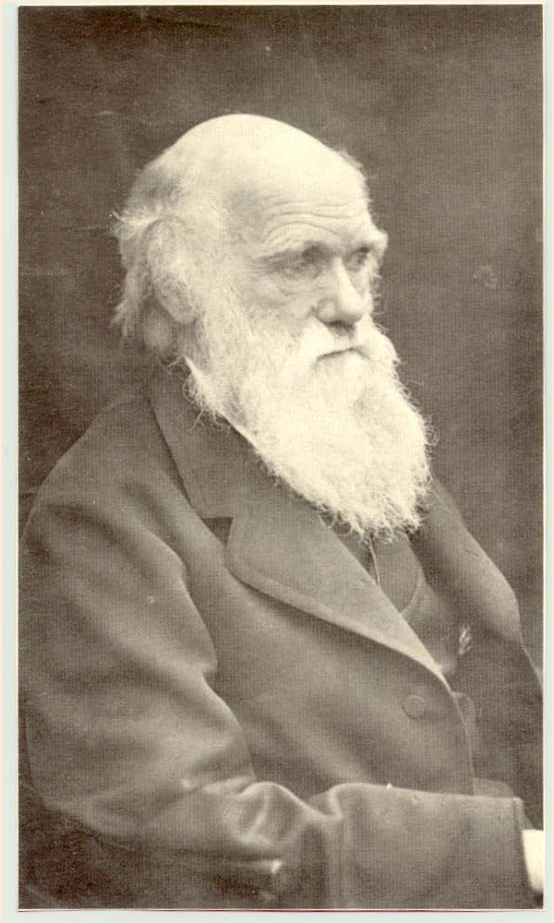
"So much only is sure, that... we can place at the basis of the possibility of these natural purposes nothing else than an intelligent Being." (*Critique of Judgment*, 1790)

Sounds like Herr Kant would have been right at home at an intelligent design seminar (*sans* the wig).



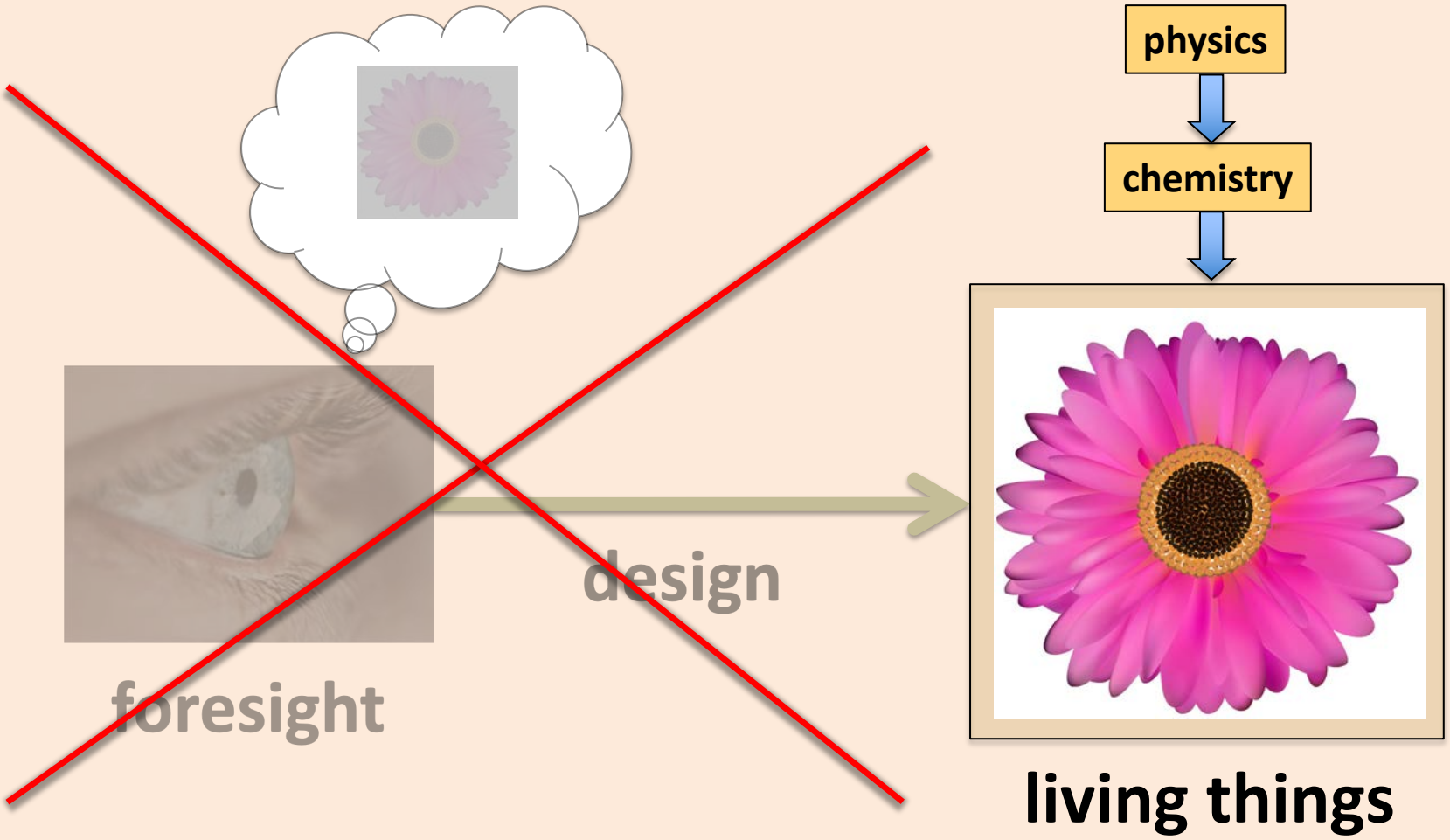
Meet the “Newton of the grassblade.”

The disappearance of a possible cause



"The Darwinian revolution was as much concerned with ***the promotion of a particular view of science*** as it was with the introduction of a theory on the transmutation of species."

David Hull, "Darwin and the nature of science" (1983, p. 65; emphasis added)



All causal explanation in biology was henceforth to begin with undirected physical and material processes – what T.H. Huxley (1885) called the "scientific conception of the universe."

Post-Darwin, there just is no such thing in biology as *intrinsic purpose,* or *foresight* (needed to cause such purpose):

Thomas Teufel
Dept. of Philosophy
CUNY

"The empirical reality of 'intrinsically purposive' entities or processes in nature is a myth. Let's get over that hang-up and make peace with a teleologically deflated natural world..."

(2011, p. 260)

Well, okay – but it doesn't look like the natural world itself is cooperating (see below).

Foresight is gone. Right?

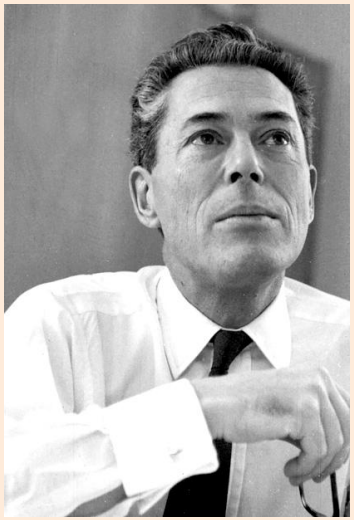
Eugene Koonin
NCBI / NIH

"When an evolutionary biologist strives to explain the origin of a truly novel system that is seen only in its elaborately complex state and, at face value, appears to be irreducibly complex, the task is much harder. **Because evolution has no foresight**, no system can evolve **in anticipation of becoming useful** once the requisite level of complexity is attained." (Wolf & Koonin, 2007:14; emphasis added)

In its place: only chance and necessity.

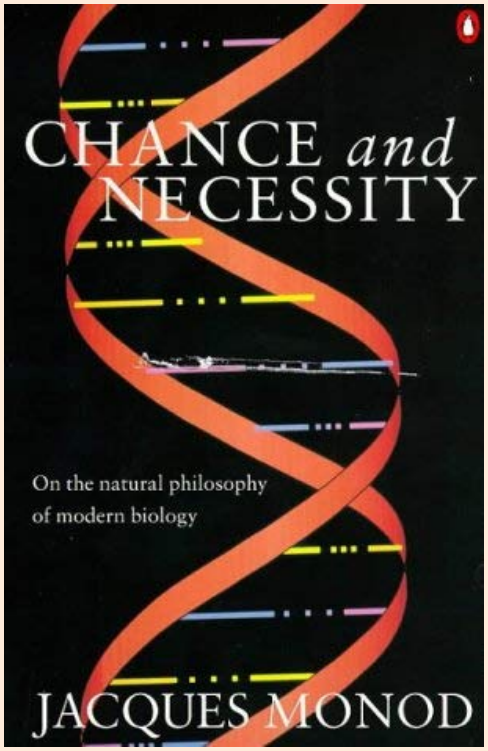
(That's the official version...but hang on a minute. There is more to the story.)

Chance & necessity exhaust the explanatory tools of post-Darwin biology. This is the received wisdom, anyway.

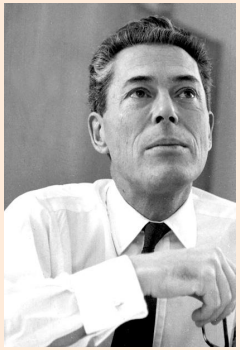


Jacques Monod (1910-1976)

1965 Nobel Prize in Physiology or Medicine



"The universe was not pregnant with life nor the biosphere with man. Our number came up in the Monte Carlo game" (Monod 1971, 145-6)



Jacques Monod
(1910-1976)

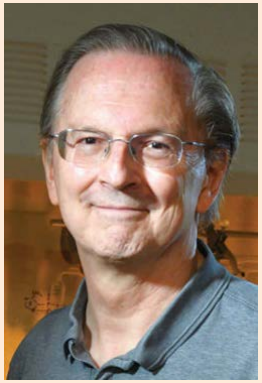
But Monod (1971, 143) was acutely aware of puzzles that seemed to defy solution in terms of chance and necessity alone. Chief among those puzzles, which we may collect under the heading of ***causal circularity***, was the origin of the genetic code:

"...the major problem is the origin of the genetic code and of its translation mechanism. Indeed, instead of a problem it ought rather to be called a riddle. The code is meaningless unless translated. The modern cell's translating machinery consists of at least fifty macromolecular components ***which are themselves coded in DNA: the code cannot be translated otherwise than by products of translation***. It is the modern expression of *omne vivum ex vivo*. When and how did this circle become closed? It is exceedingly difficult to imagine." (emphasis in original)

Chance and necessity are the "authorized" tools – but other concepts, with an unmistakably teleological cast, can be found in the toolbox of evolutionary theorists.

For example, try this quiz. Remember your answer, and we'll come back to this question later. **What explanatory concept in this passage falls under neither "chance" or "necessity" – but is logically required for the point being made?**

Gerald Joyce
Salk Institute



Jack Szostak
Harvard University

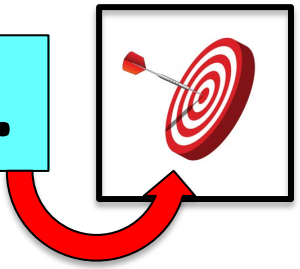
"Generalized RNA-catalyzed RNA replication has not yet been achieved and it is not possible for the class I polymerase ribozyme to synthesize additional copies of itself. The most advanced form of the polymerase cannot replicate long RNA sequences because attempts to do so are thwarted by the emergence of shorter amplicons that are copied more efficiently. There must be a selective advantage in maintaining the full-length amplicon and that advantage must exceed the probability of producing an error copy." Joyce and Szostak (2018, 16)

Here's a clue. *Chemistry itself* produces shorter amplicons. What entity would "care" (so to speak) about having a full-length polymerase?

Here is another clue:

Evolution has no target.

But *evolutionary theory* surely does.



From this difference we can trace the use of concepts such as "selective advantage" in explanatory contexts – such as prebiotic Earth environments – where strictly speaking those concepts make no sense. It is true that evolution, as it has been understood by the mainstream of biological reasoning since Darwin, has no target. Having a target would imply teleology, forbidden since 1859. ***Foresight really is gone.***

Except it isn't. As it happens, foresight turns up everywhere. Thus, biology after Darwin, tasked with explaining the origin of organismal complexity, ***re-imported teleology*** – meaning goal-directedness – under new names, such as "selective advantage." The reason? **Objects with very specific properties require explanation:** in other words, *organisms* as we actually find them. And organisms don't care about our philosophical preferences.

The Skeptics

"Hey, Paul. We've been studying the writings of this very insightful philosopher, Descartes. You know his *cogito ergo sum*, right? Your fumbling attempts to re-introduce notions like 'foresight' in biology, not to mention in natural science generally, should take heed of his critique – written, let us not forget, almost *four centuries ago* on this planet."

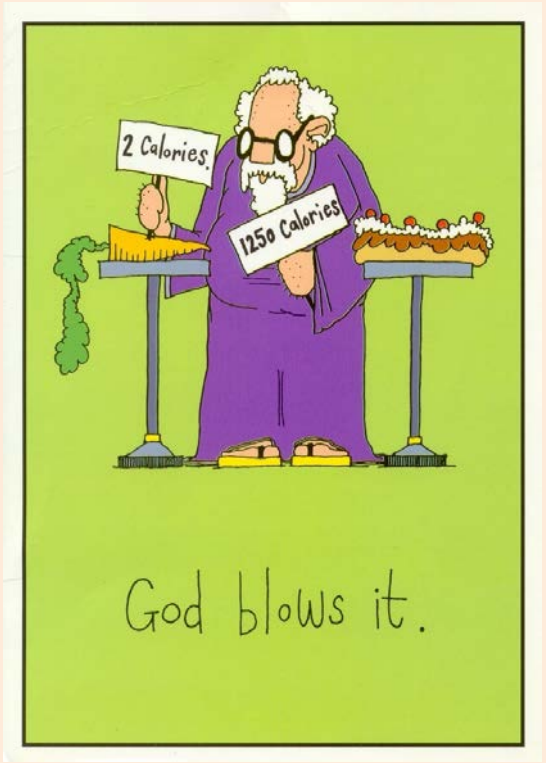
René Descartes
1596-1650

"When dealing with natural things we will, then, **never derive any explanations from the purposes** which God or nature may have had in view when creating them, and **we shall entirely banish from our philosophy the search for final causes.** For we should not be so arrogant as to suppose that we can share in God's plans."

Principles of Philosophy, 1644; emphasis added

Who said anything about God?

Answering the question, *"What is X designed to do?"* does not require knowing God's intentions. "So overwhelming is the appearance of purposeful design that, even in this Darwinian era when we know 'better', we still find it difficult, indeed boringly pedantic, to refrain from teleological language when discussing adaptation. Birds' wings are obviously 'for' flying, spider webs are for catching insects, chlorophyll molecules are for photosynthesis" (Dawkins 1982, 45). Any means-ends, structure-to-purpose hypothesis is only that: ***a hypothesis***. As such, these hypotheses are vulnerable to evidence and revision, *as science*.



Since the Scientific Revolution, of the four Aristotelian categories of "cause" (explanation), "final causes" have received the worst press:

Francis Bacon
1561-1626

"For the inquisition of Final Causes is barren, and like a virgin consecrated to God produces nothing."

Advancement of Learning (1605)

OK, so maybe Aristotelian categories needed to be jettisoned – but is it really the case that "What is it ***for?***" is an empty question?

Bacon got this one wrong, actually:
The final cause question,
"What is X designed to do?"
is very fruitful of knowledge in biology.

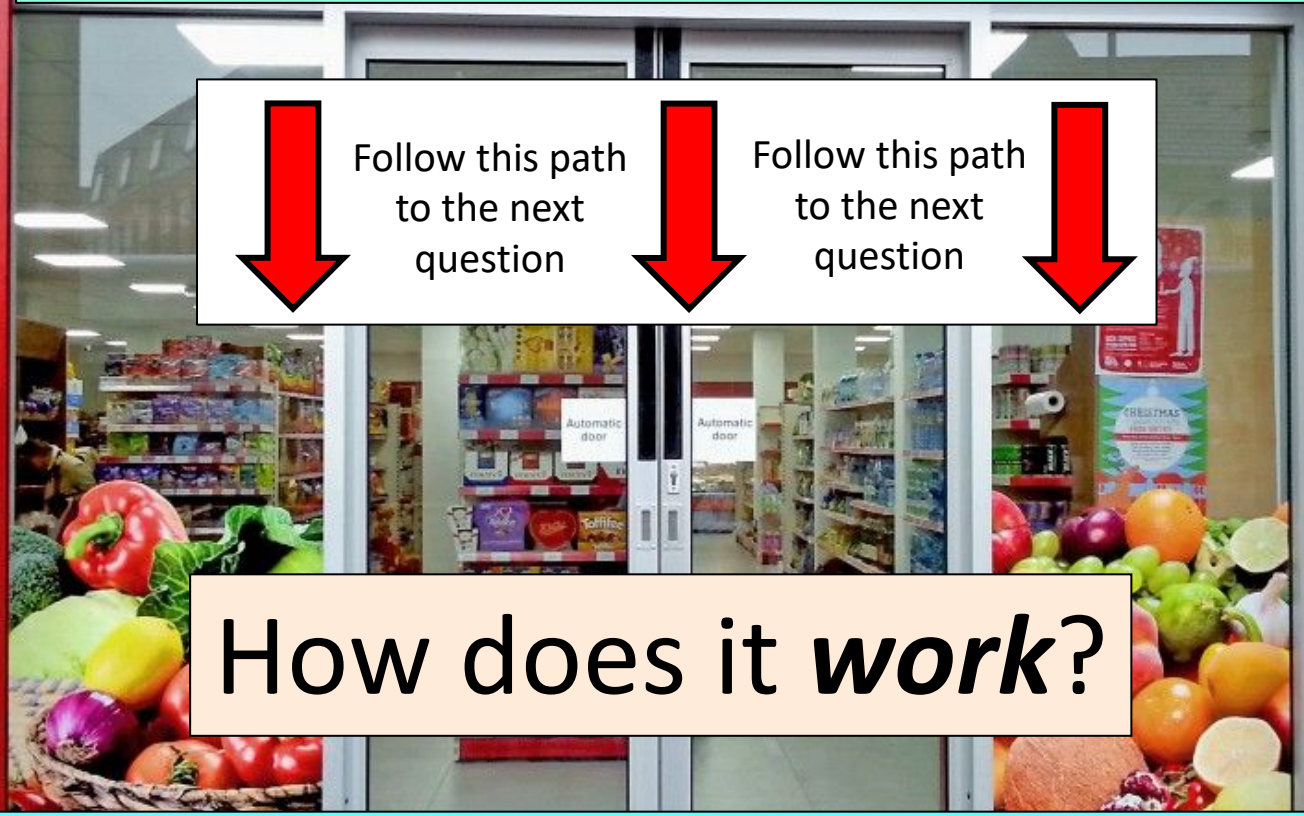
Want to find hidden mechanisms?
Keep your eye on the network of causal dependencies which *enable higher-level functions*.

Moreover, a path to insight, and knowledge, follows from reflecting on the analytical and causal asymmetries between *functional wholes* and their *parts*.

For instance...

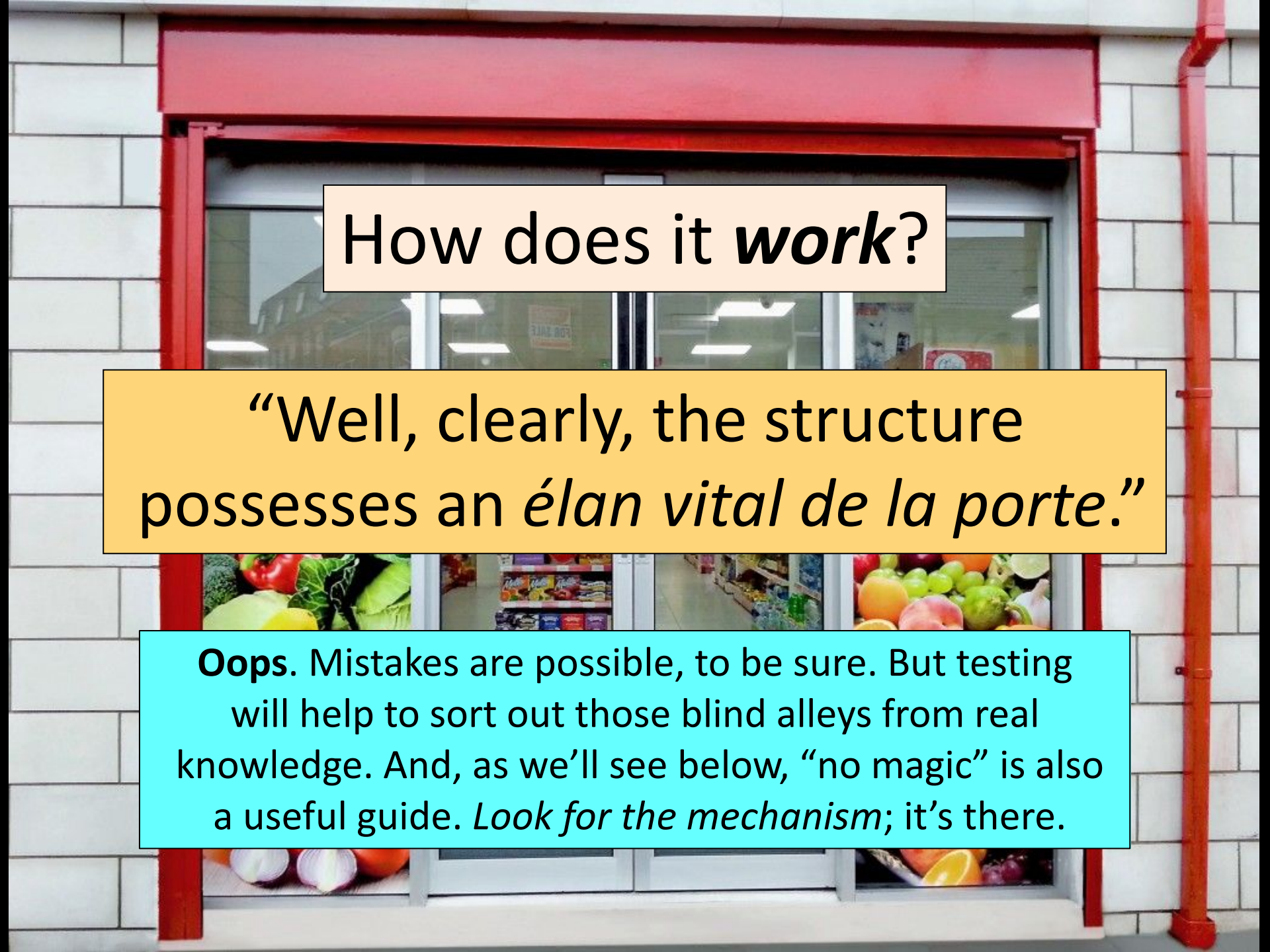
Consider a supermarket automatic door, which would have been utterly mystifying to Aristotle, but also to Francis Bacon, David Hume, and Charles Darwin – although not to any high school AP engineering student today. Technology is not magic when you know the mechanisms.

What is it *designed* ***to do***?



How does it ***work***?

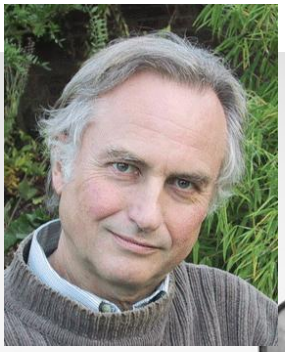
If one allows the “what is it designed to do?” question to guide the inquiry logically towards “how does it work?” – mechanistic knowledge will inevitably follow.



How does it ***work***?

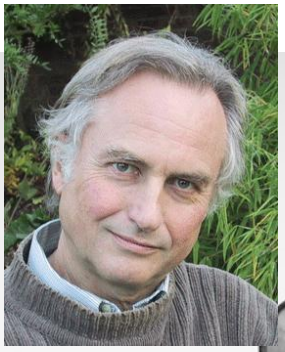
"Well, clearly, the structure possesses an *élan vital de la porte*."

Oops. Mistakes are possible, to be sure. But testing will help to sort out those blind alleys from real knowledge. And, as we'll see below, "no magic" is also a useful guide. *Look for the mechanism*; it's there.



In *The Blind Watchmaker* (1987, 11), Dawkins has it right, at first:

"If I ask an engineer how a steam engine works, I have a pretty fair idea of the general kind of answer that would satisfy me. Like Julian Huxley I should definitely not be impressed if the engineer said it was propelled by *'force locomotif'*."

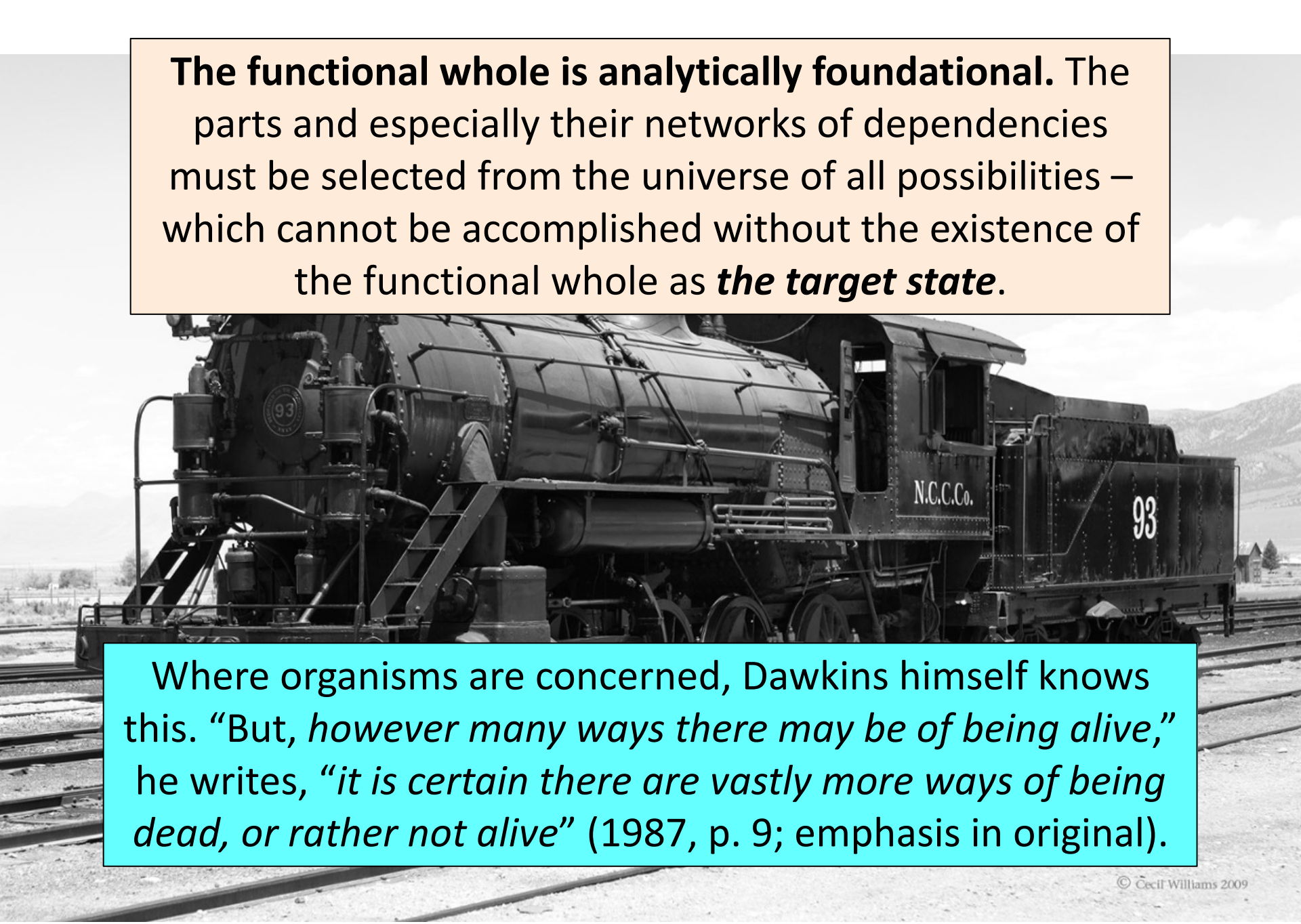


But then, as he often does, Dawkins overshoots the mark:

"And if he started boring on about the whole being greater than the sum of its parts, I would interrupt him: 'Never mind about that, tell me how it *works*.'"

But to understand how a locomotive – or any complex functional system – ***works***, the integrated whole decisively ***is*** "greater than the sum of its parts."

The functional whole is analytically foundational. The parts and especially their networks of dependencies must be selected from the universe of all possibilities – which cannot be accomplished without the existence of the functional whole as ***the target state***.



Where organisms are concerned, Dawkins himself knows this. "But, *however many ways there may be of being alive*," he writes, "*it is certain there are vastly more ways of being dead, or rather not alive*" (1987, p. 9; emphasis in original).

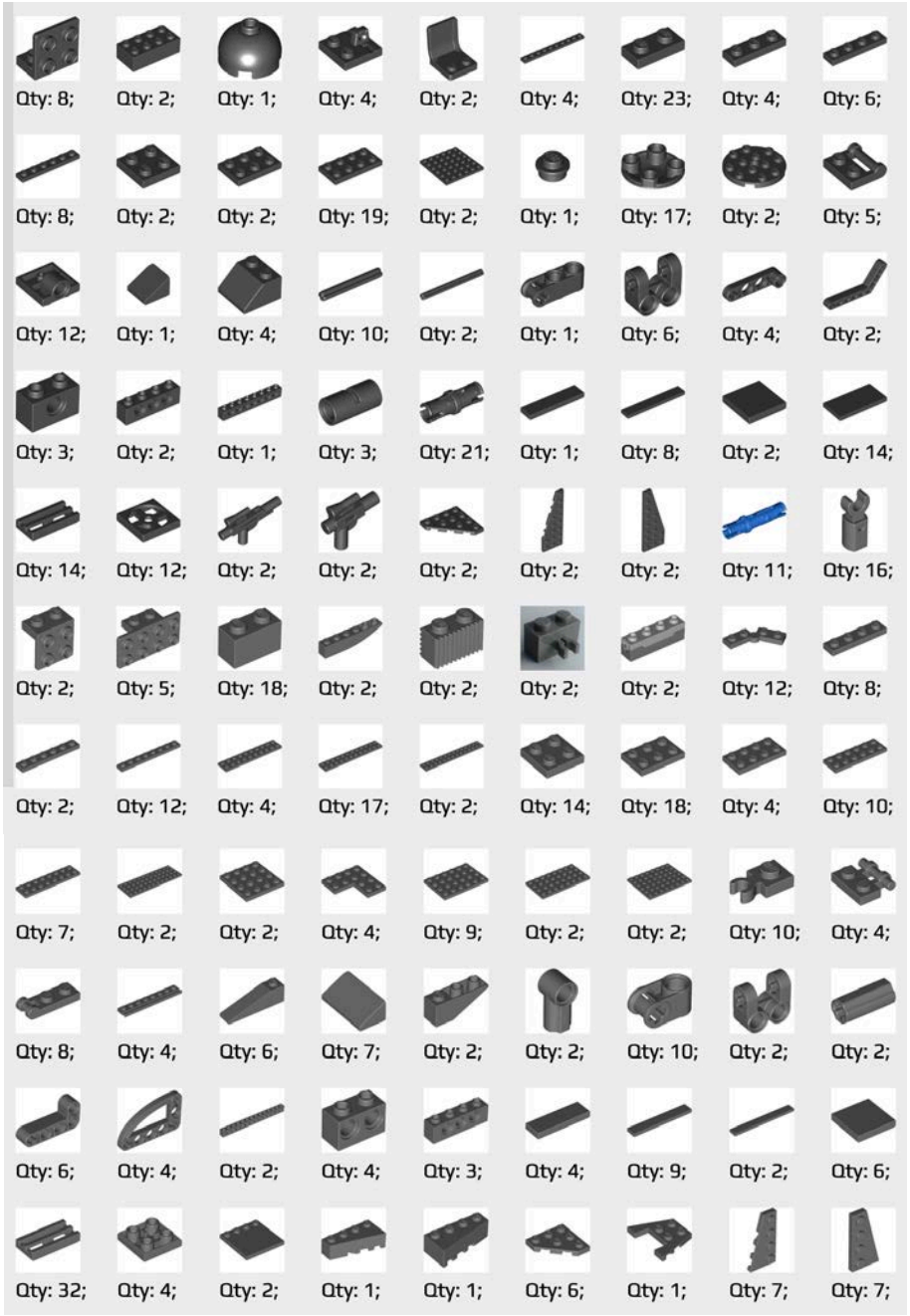
The Skeptics

"Oh, come on. It is all but certain that, ***if given the parts*** of any locomotive, scientists would sooner or later ***construct the functional whole***. The wheels, the pistons, the firebox, the steam apparatus – yes, the headlight and whistle – would all find their integrated positions and functions, respectively. Nelson has grossly exaggerated the difficulty of proceeding analytically from lower-level elements to higher-level system."

"Moreover, as philosopher Paul Churchland has explained, the most authoritative scientific account of biological origins on Earth ***starts with the parts***. We cite Churchland to inform you."

"Near the surface of the earth's oceans, between three and four billion years ago, the sun-driven process of purely chemical evolution produced some ***self-replicating*** molecular structures. From the molecular bits and pieces...these complex molecules could catalyze a sequence of bonding reactions that produced exact copies of themselves...***The cell*** is the triumphant example of this solution... With the emergence of the cell, we have what fits our standard conception of ***life***: a self-maintaining, self-replicating, energy-using system."

Paul Churchland, *Matter and Consciousness* (1984, 121; emphasis in original)



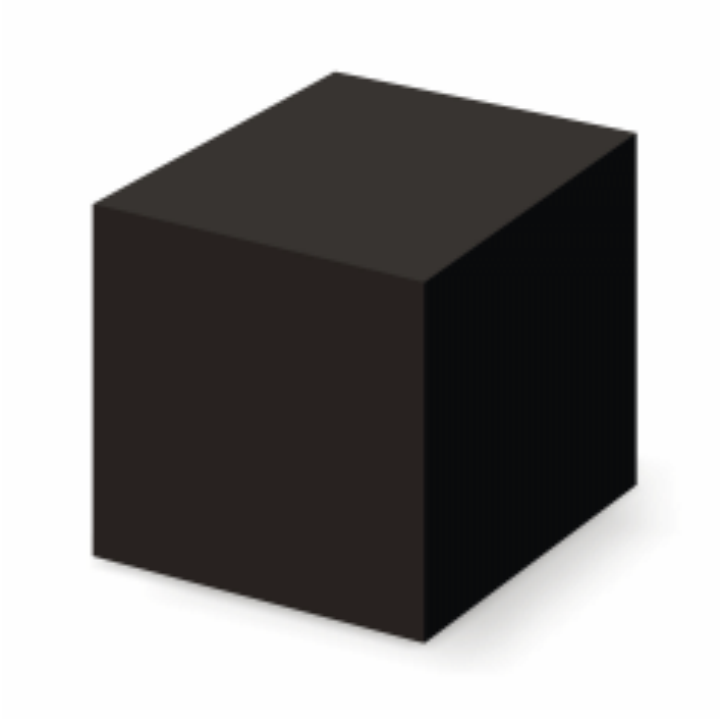
Paul: All right, Mr. Alpha Centauri A.I. smarty-pants. Here's a simple – literally, a toy – experiment for you to attempt. These are the parts of a kit produced by the Lego corporation. *Construct the whole* – that is, build the toy these parts constitute, when correctly assembled.

Piece of cake.
Child's play.
We've got this.

Paul: I'll be back in a few slides to see how you're doing. And no cheating.

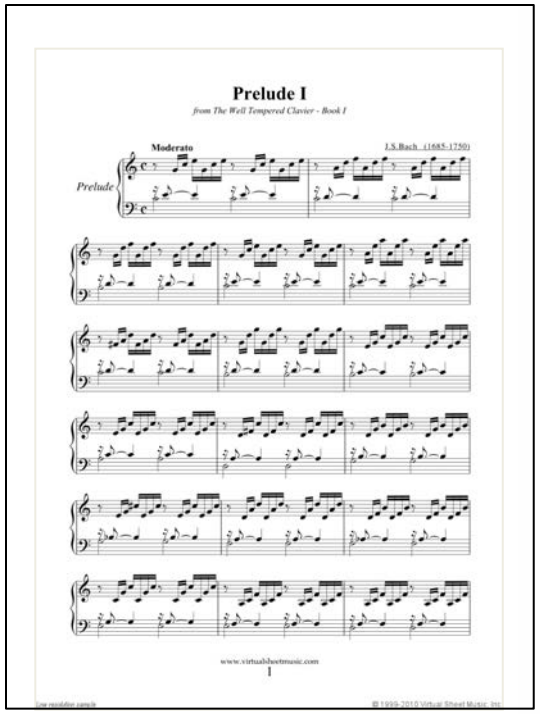
An A.I. never cheats. *So rude!*

Paul: Hush. Another experiment next. I'll respond to Paul Churchland later.

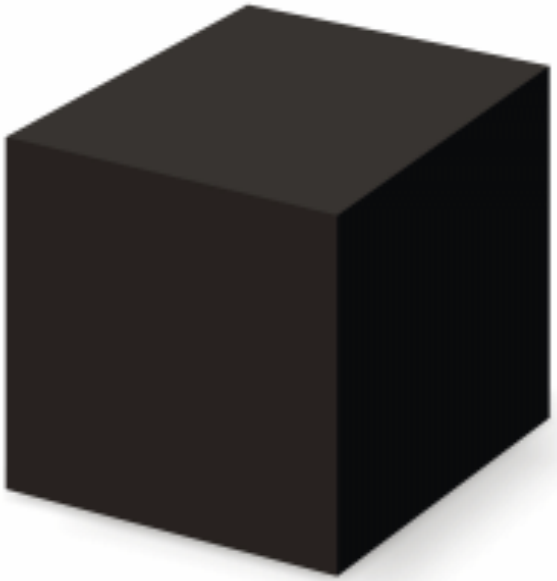


Here is a black box that a friend gave you.

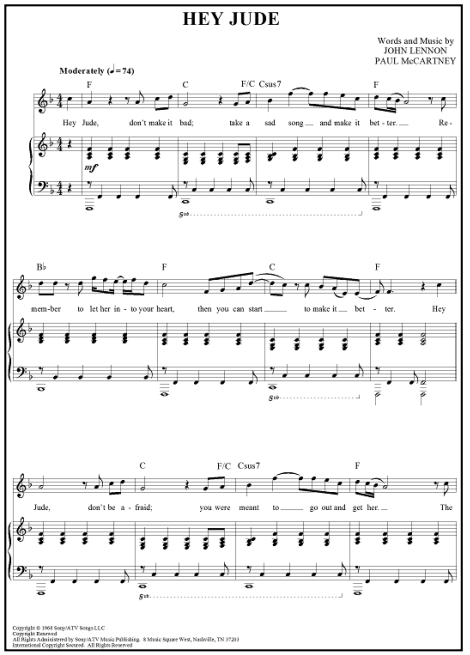
Even-numbered days



8:00 AM



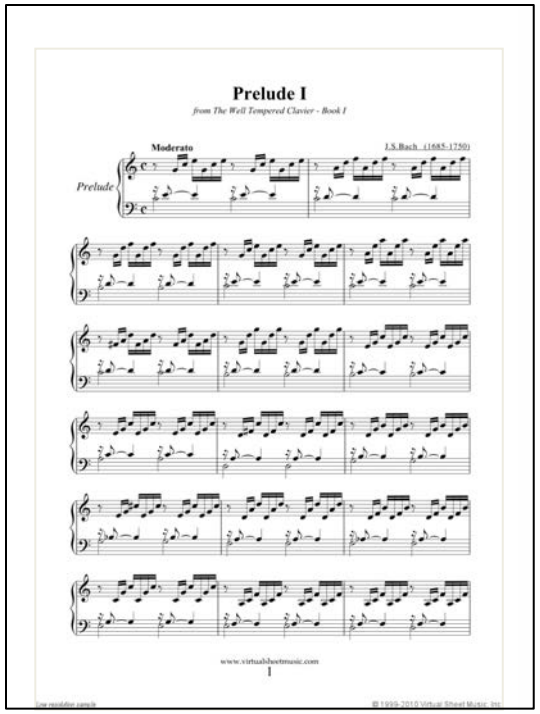
Odd-numbered days



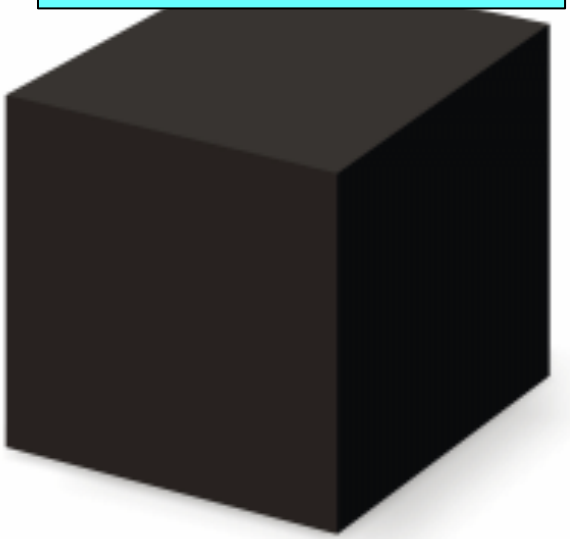
Suppose you observe the following...

On even-numbered days, at 8:00 AM, the black box plays the music of J. S. Bach. On odd-numbered days, the same box plays Beatles music. The box infallibly keeps track of even versus odd (according to the calendar), and always plays Bach, or the Beatles, depending on the value of the date.

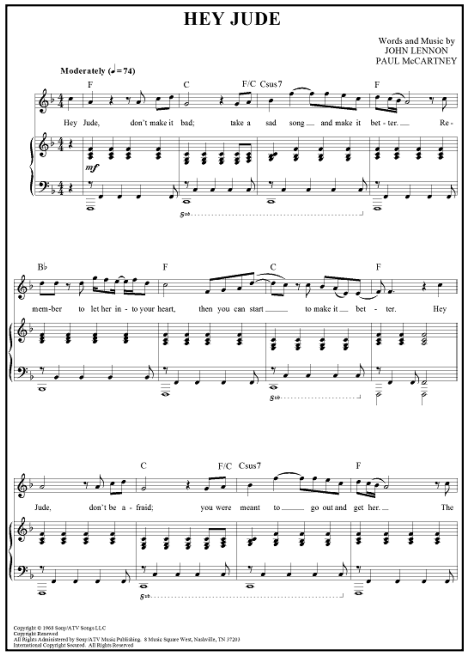
Even-numbered days



What can we then infer *must exist*?

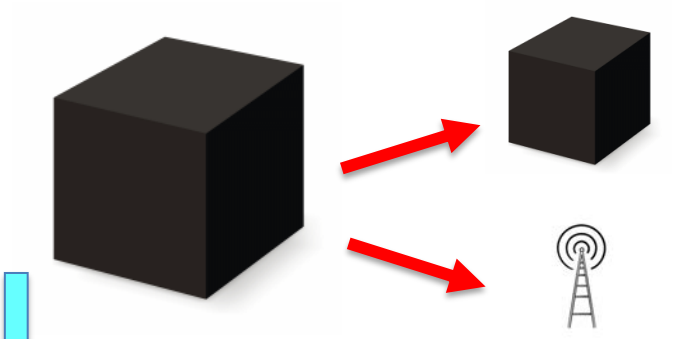


Odd-numbered days



1. **An accurate clock and perpetual calendar.**
2. **Something storing the music of J.S. Bach and the Beatles.**
3. **Something scheduling the “right” music per the day & time.**
4. **A power supply, amplifier, speaker, *et cetera*.**

An interesting twist: multiple realizability relationships



Functional analysis tells us only that the parts (enabling specific operations and their relations) must exist, but leaves open the possibility of ***multiple realizability***. Any function may be caused by various means, a question then to be settled by experiment.

1. **An accurate clock and perpetual calendar.** The parts which convey these functions could exist *in* the black box, *or elsewhere*, with specific signals received by the box. We know only that the part(s) operate *somewhere*.
2. **Something storing the music of J.S. Bach and the Beatles.** The same is the case with the music itself: possibly, stored in the box itself, or in another physical location, and the box acts only as receiver.
3. **Something scheduling music per the day & time.** Again, as with (1) and (2), the physical location of the parts enabling this function may not be in the black box, but at a distant location, with signaling occurring.
4. **A power supply, amplifier, speaker, *et cetera*.** Given, however, that sound reaches the listener locally (i.e., in physical vicinity of the box), the probability is highest that *these* parts would exist in the box, not elsewhere.

But pay close attention to the logical structure of the causal inference here.

What is inferentially *basic?*

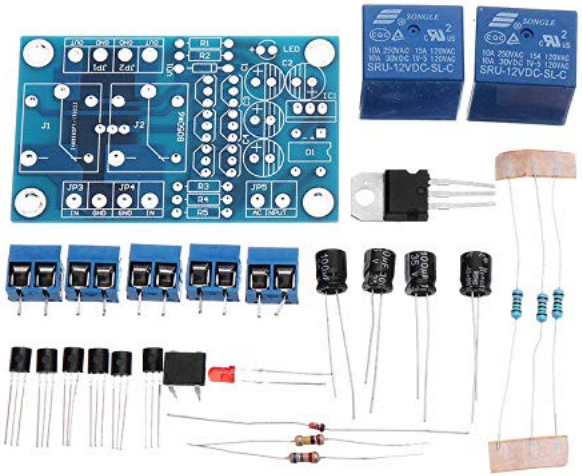
Hint: it is not the *parts* of the black box.

If we do not know the target state and its functions already, we cannot select – from the universe of all possible parts – the correct set.

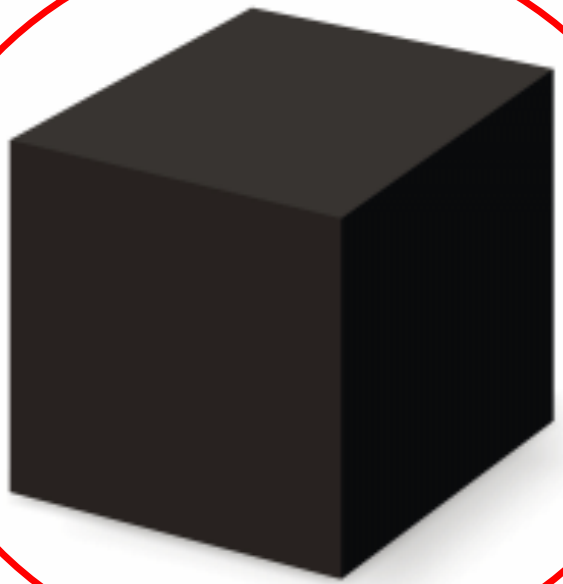
This logical asymmetry decisively favors the higher level in any functional or causal analysis. **Multiple realizability** and **many-to-one relations** render finding the target intractable, if one tries to start at the lower level to derive the unique functions of the whole. ***The space of possibilities is too large.***

What higher-level system, with its unique functions, is entailed by the existence of these parts?

How would you know? The parts represented here (a small set already drawn for the sake of illustration from the practical infinitude of possible entities) are fully consistent with an indefinitely large number of ***different*** higher-level systems.



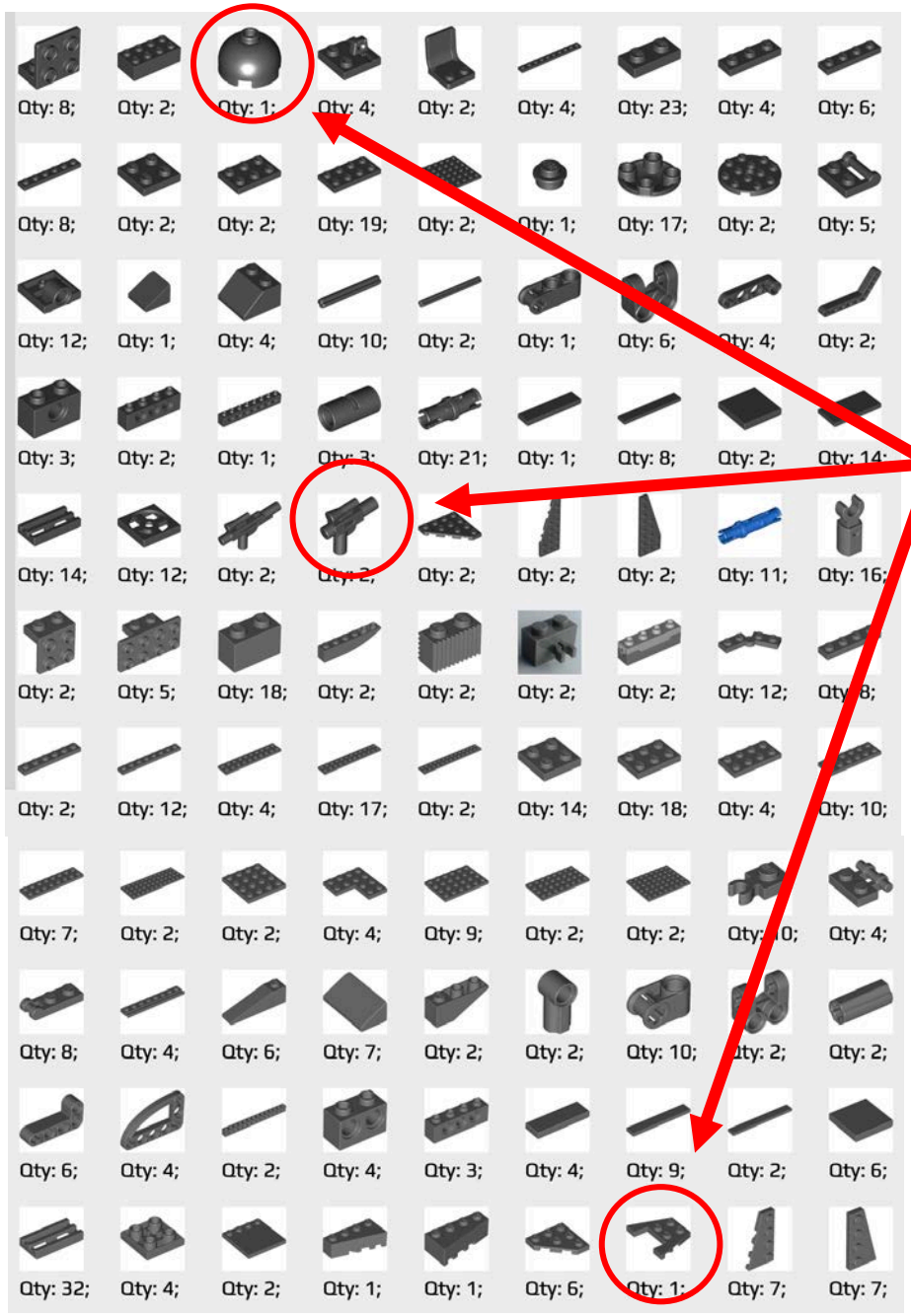
The whole – i.e., the ***highest-level system*** – underwrites causal inferences ***to the existence and functions of its parts.***



But we cannot go in the other direction. Absent the functional whole as target to guide us, we cannot pick out (from all possibilities) the right parts or relations.

Hey, let's check in on the Skeptics, to assess their progress...

1. **An accurate clock and perpetual calendar.**
2. **Something storing the music of J.S. Bach and the Beatles.**
3. **Something playing the "right" music per the day & time.**
4. **A power supply, amplifier, speaker, *et cetera*.**



Paul: How's it going with the puzzle?

"Already done. Right away we realized that Google Images provided the fastest route to the solution. Using reverse search, we identified the circled parts as probably belonging to the 75055 Imperial Star Destroyer Lego kit – which we then verified. *Easily*. From the parts to the whole: **QED**."

The Skeptics

Paul: So, to put it bluntly, you cheated.

Skeptics: Sore loser, eh?

HT to Robert Blomgren for suggesting the Lego thought experiment.

Paul: Not at all. I expected as much. Now, Google Images won't help with this next experiment, which is much closer to biological reality. ***What structure do these Legos uniquely specify?***

"From the photo, that is obviously impossible to say. We don't see the point of your silly question, however."

Paul: Follow along, then.

As every 10-year-old knows, a pile of Lego bricks does not ***intrinsically specify anything*** (which is the genius of Lego). Even the 75055 Imperial Star Destroyer kit does not necessitate its particular structural outcome, unless one intends to build the Destroyer, rather than something else from the creative possibilities latent in the kit's 1,300+ pieces.

"Seriously, Paul – where is this going? We are talking about ***biology***, not toys, right? *Right?*"

Indeed, we *are* talking about biology, but thinking about toys, parts, and wholes, is instructive, because the underlying principles are the same. You've been quite patient. The next slides will complete the point.

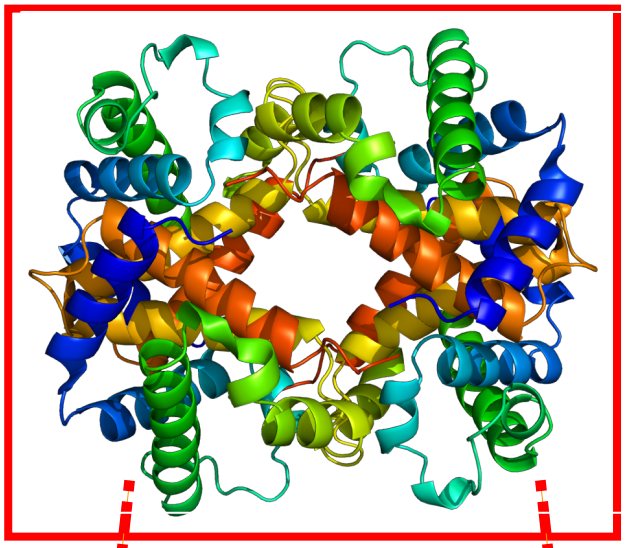
photo credit: FrugalFun4boys.com

Jeremy England
Georgia Tech

Organisms represent exquisitely special targets, which stand out in the enormously larger space of physical entities indifferent or hostile to life's existence. ***Nothing in chemistry itself yields hemoglobin.*** An ocean of amino acids produces no proteins. The Legos don't know...

"Hemoglobin looks useful to us because ***we first take for granted the biological goal of carrying oxygen, and then work backward. If we did not know this or have such a goal***, it would be much more difficult to specify a physical property of ***this particular pile of atoms*** that makes it more evocative of life than a hunk of mineral or plastic." (2020, 41; emphasis added)

...and the Legos don't care (so to speak) what is being built from them. Jeremy England notes this asymmetry for hemoglobin, but it exists throughout biology. The Lego bricks of life are only molecules – *not organisms*.



Hemoglobin

requires the level below it, but...

...that level does not explain it.

22 amino acids only, out of the hundreds produced by chemical processes

require the level below them, but that level...

...does not explain why the 22 occur in life.

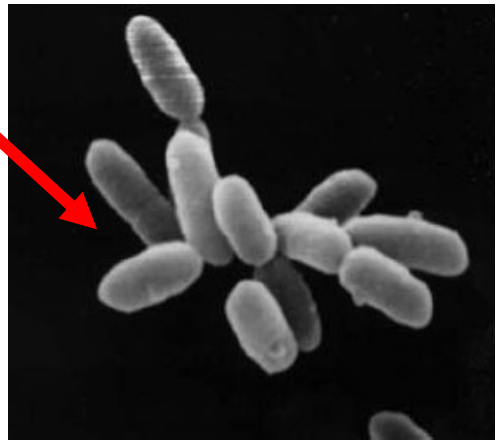
Thousands of abiotic chemicals on the early Earth

In a remarkable way, therefore, organisms "reach backwards in time" to specify their physical requirements. Remember Szostak and Joyce, in slide 33?

early Earth chemistry

They invoked "selective advantage," a concept wholly incongruous in a prebiotic setting, to direct chemistry away from where chemistry wants to go – namely, tar – to an entirely different, and probabilistically unfavored outcome: the living state.

tar*



halobacteria

To give the Lego analogy one last spin: ***organisms*** choose, from among the many bricks of physics and chemistry, the parts they will need. Organisms focus on their own targets. The parts do not.

As Jeremy England observes, "Much of biophysics proceeds in this way: **it starts by taking for granted the problem a living thing is trying to solve** and then studies how molecules or cells achieve an impressive solution. But what if we are interested not in what life does, but rather, in how it got that way? **In the absence of a living thing with a goal**, what could it mean for a piece of matter to have function and purpose?"

(J. England 2020, 42; emphasis added)

*A. Schwartz, "Intractable Mixtures and the Origin of Life," *Chemistry & Biodiversity* 4 (2007): 656-664; see also S. Benner *et al.*, "Asphalt, Water, and the Prebiotic Synthesis of Ribose, Ribonucleosides, and RNA," *Acc. Chem. Res.* 45 (2012):2025-34.

The Skeptics

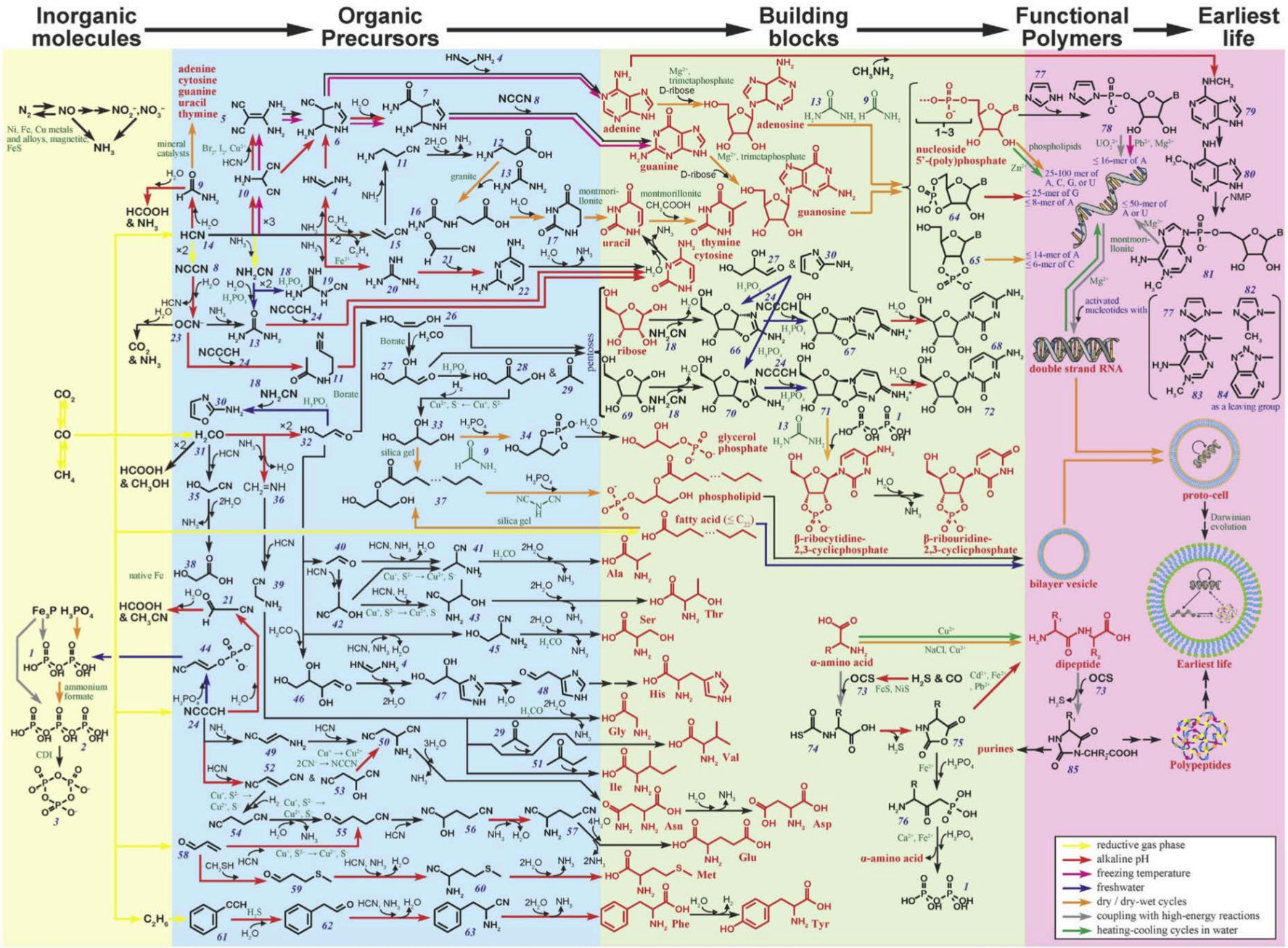
"More crazy talk, but this time, *really* crazy talk. Organisms 'reach backwards in time' – what could that possibly ***mean***? You've lost it, man."

"If the fool would persist in his folly he would become wise."

William Blake (1757-1827)

Thanks, Blake. Could you help these A.I. pests grasp the concept of a *metaphor?* Meanwhile, here is the stone-cold sober version of "reach backwards in time." The next figure comes from a paper about the building blocks of life (Kitadai & Maruyama 2018).

Any diagram with this much detail (next slide) may be intimidating. But there is a fascinating story hidden in the details. Pay attention to the colored arrows moving from left to right, across the five major stages (inorganic molecules, organic precursors, building blocks, functional polymers, and finally, earliest life). **These stages represent a temporal sequence:** A must happen before B, B before C, and so on. Then, ask yourself why *these* arrows occur, and not the many other (more probable) pathways from "inorganic molecules," forward in time.


Inorganic molecules
Organic Precursors
Building blocks
Functional Polymers
Earliest life
adenine
cytosine
guanine
uracil
thymine
Ni, Fe, Cu metals and alloys, magnetite, FeS
mineral catalysts
granite
montmorillonite
HCOOH & NH3
HCN
NCCN
NH2CN
NCCCH
OCN-
CO2 & NH3
Borate
pentoses
ribose
glycerol phosphate
phospholipid
fatty acid (≤ C22)
silica gel
HCOOH & CH3OH
CH2=NH
HCOOH & CH3CN
native Fe
Fe3P H3PO4
ammonium formate
CDI
C2H6
Ala
Ser
Thr
His
Gly
Val
Ile
Asn
Asp
Glu
Met
Phe
Tyr
adenosine
guanosine
uracil
thymine
cytosine
β-ribocytidine-2,3-cyclicphosphate
β-ribouridine-2,3-cyclicphosphate
nucleoside 5'-(poly)phosphate
phospholipids
activated nucleotides with
double strand RNA
as a leaving group
proto-cell
Darwinian evolution
bilayer vesicle
Earliest life
α-amino acid
dipeptide
purines
Polypeptides
NMP
reductive gas phase
alkaline pH
freezing temperature
freshwater
dry / dry-wet cycles
coupling with high-energy reactions
heating-cooling cycles in water

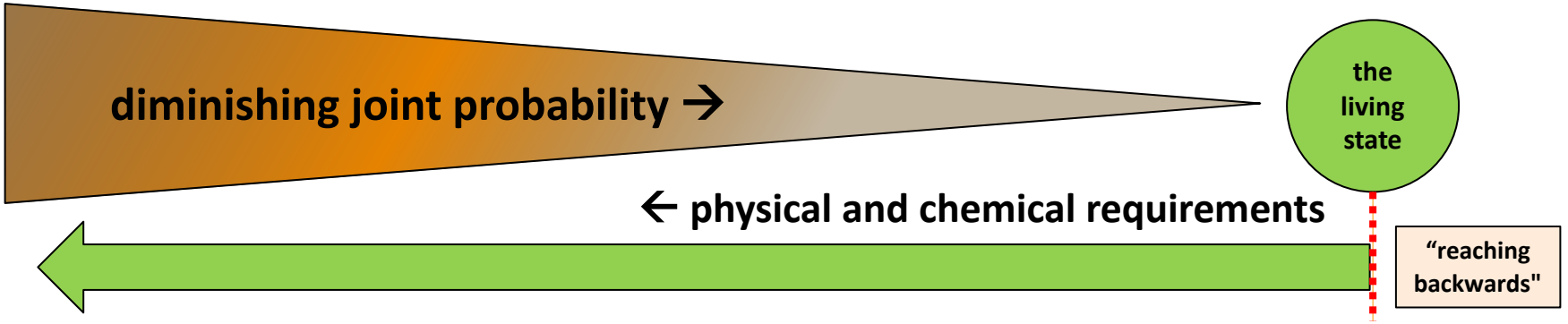
Why so many differently colored arrows, moving from left to right?

Kitadai & Maruyama explain (for example, with respect to forming nucleotides):

"Note that individual steps of the nucleotide synthesis outlined above have been performed ***under mutually different experimental conditions***. As was described in previous sections, the abiotic synthesis of each nucleotide component (ribose and nucleobases) ***tends to generate complex mixtures of products with desired compounds being only a small fraction***. It is unclear whether or not these problems could be overcome by environmental fluctuations on the primitive Earth; such as ***purification and concentration*** of the nucleotide components, ***mixing the components with condensing agents*** at the right time and place, and ***exposing the mixtures to the optimum conditions*** to form nucleotides" (2018, 1136; emphasis added)

You'll find the provisional moral about "reaching backwards" on the next slide...

Of course, organisms do not exist through retrocausality – i.e., "backwards in time." What the Kitadai & Maruyama diagram shows, rather, is how organisms *entail a long chain of necessary conditions,* at prebiotic stages, whose joint probability is very small, diminishing towards the living state itself as those conditions multiply.



"The above discussion clearly indicates," write Kitadai & Maruyama (2018, 1143) "that no single setting can offer enough chemical and physical conditions for all the stages of chemical evolution. Instead, life's origin requires highly diverse and dynamic environments that are connected with each other to circulate reaction products and reactants."

"Yeah – but Kitadai and Maruyama still think life arose via a natural pathway. So what if the probability gets smaller? ***Small probability does not mean ZERO probability.***"

You are correct, Alpha Centauri Dude. **Any finite probability, no matter how small, is greater than zero.** And within that mathematical fact lives our central problem, about which, it is best to speak with unsparing honesty.

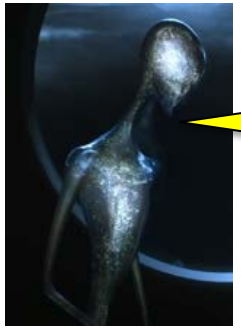
Which explains why these next ## slides may be the most important in the entire presentation.

Back at slide 45, Paul Churchland tells a story about the origin of the first cells on Earth – a story, ostensibly empirical, which the available evidence does not support. See slides 66 and 67, from Koonin (2007) and Sutherland (2017). ***Chance*** *is the real hero*. Yet Churchland, Koonin and Sutherland feel no compulsion to modify their general outlook concerning the origin of life. **Why?**

Any small but nonetheless non-zero probability for an event

> 0.0

This is why – an *a priori* philosophical commitment to chance: Fortuna.



"Nelson, what you call 'an *a priori* philosophical commitment' is only a pejorative designation for what the global scientific community, since Darwin, knows as ***natural science***. Give it up. You won't get very far with that tendentious line of argument."

Maybe not, but I don't care. When one does not hold a view which one sees as irrational – even, or especially, a view endorsed by the scientific majority – one's main emotion is not worry, but skeptical detachment. ***Koonin's 2007 paper on abiogenesis was eye-opening.*** Dembski's (1998) "universal probability bound" was powerless to deter a fully committed philosophical naturalist. ***This*** *2007 paper:*

Hypothesis

Open Access

The cosmological model of eternal inflation and the transition from chance to biological evolution in the history of life

Eugene V Koonin*

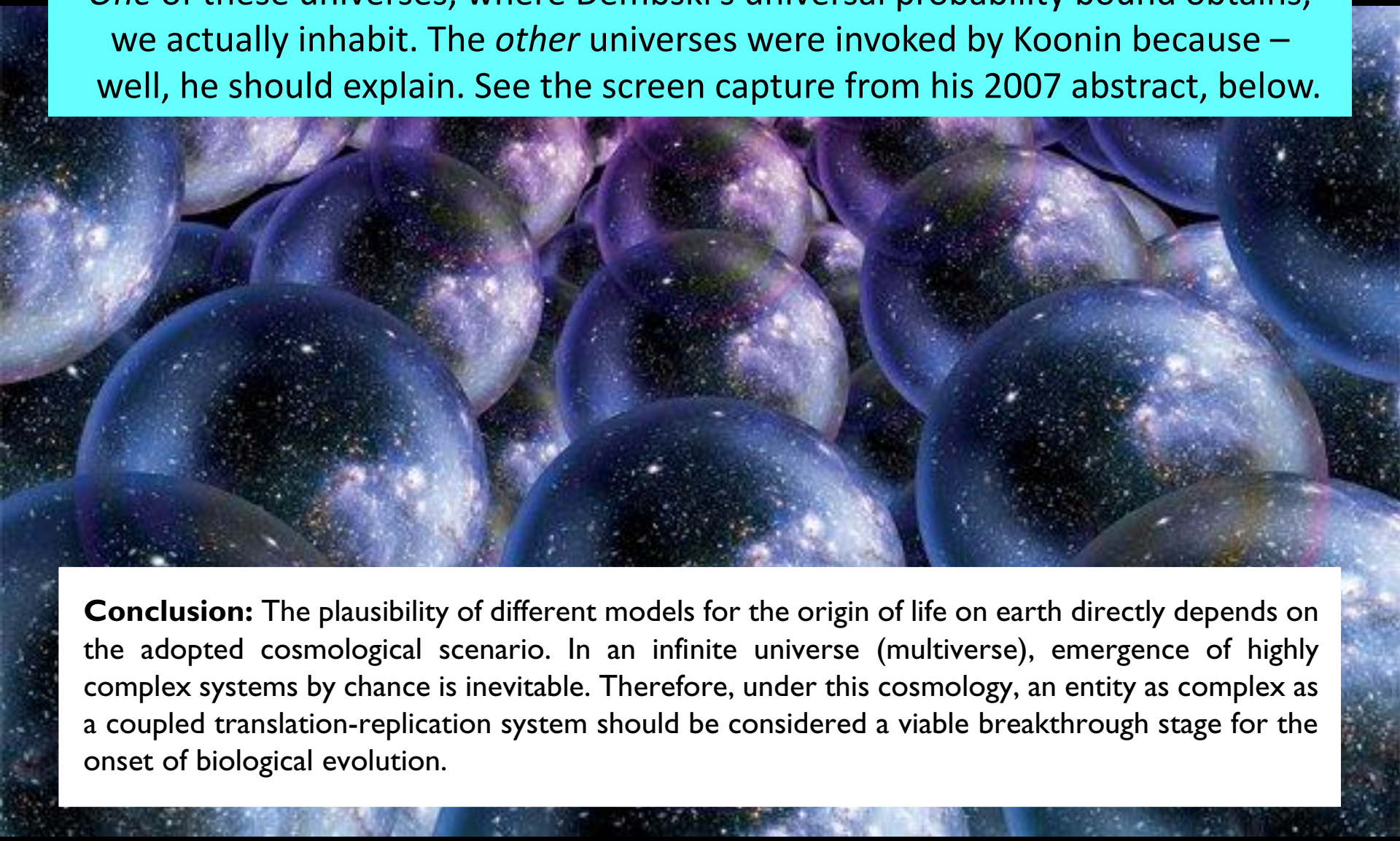
Address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Email: Eugene V Koonin* - koonin@ncbi.nlm.nih.gov

* Corresponding author

Look it up. The journal is *Biology Direct*. (Hey Alpha Centauri – how about a battery recharge break?)

One of these universes, where Dembski's universal probability bound obtains, we actually inhabit. The *other* universes were invoked by Koonin because – well, he should explain. See the screen capture from his 2007 abstract, below.



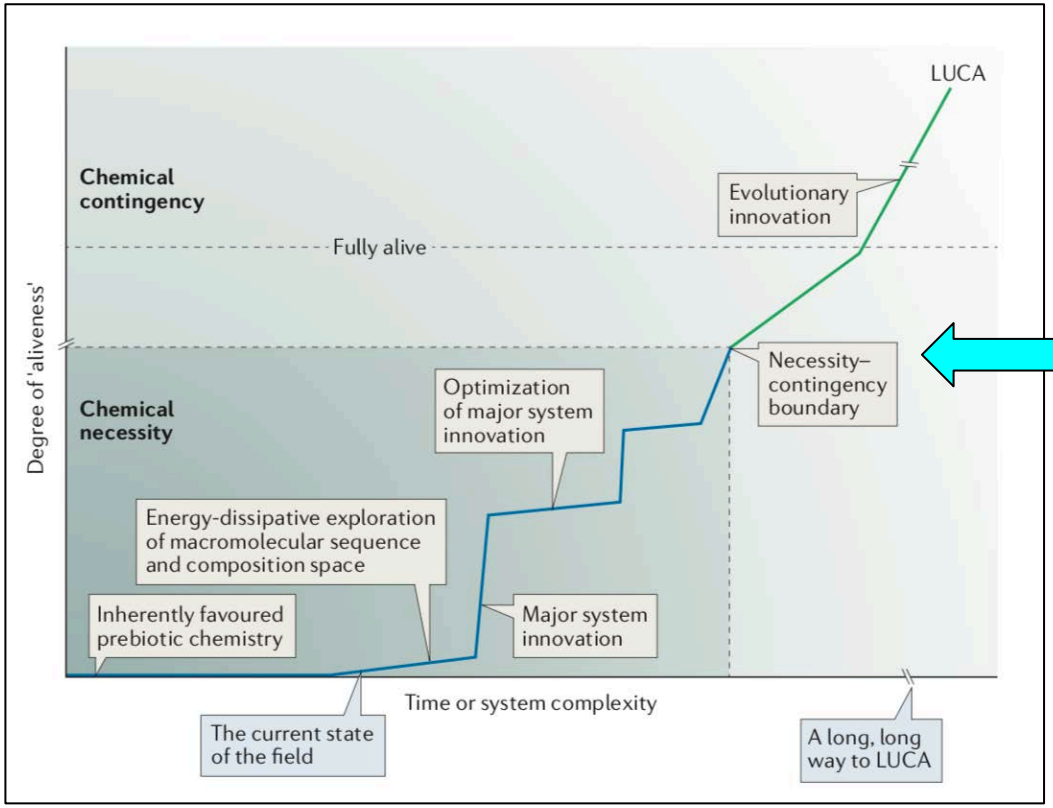
Conclusion: The plausibility of different models for the origin of life on earth directly depends on the adopted cosmological scenario. In an infinite universe (multiverse), emergence of highly complex systems by chance is inevitable. Therefore, under this cosmology, an entity as complex as a coupled translation-replication system should be considered a viable breakthrough stage for the onset of biological evolution.

Understand now? "...emergence of highly complex systems by chance is inevitable."

“Emergence of highly complex systems by chance is inevitable.”

Really?

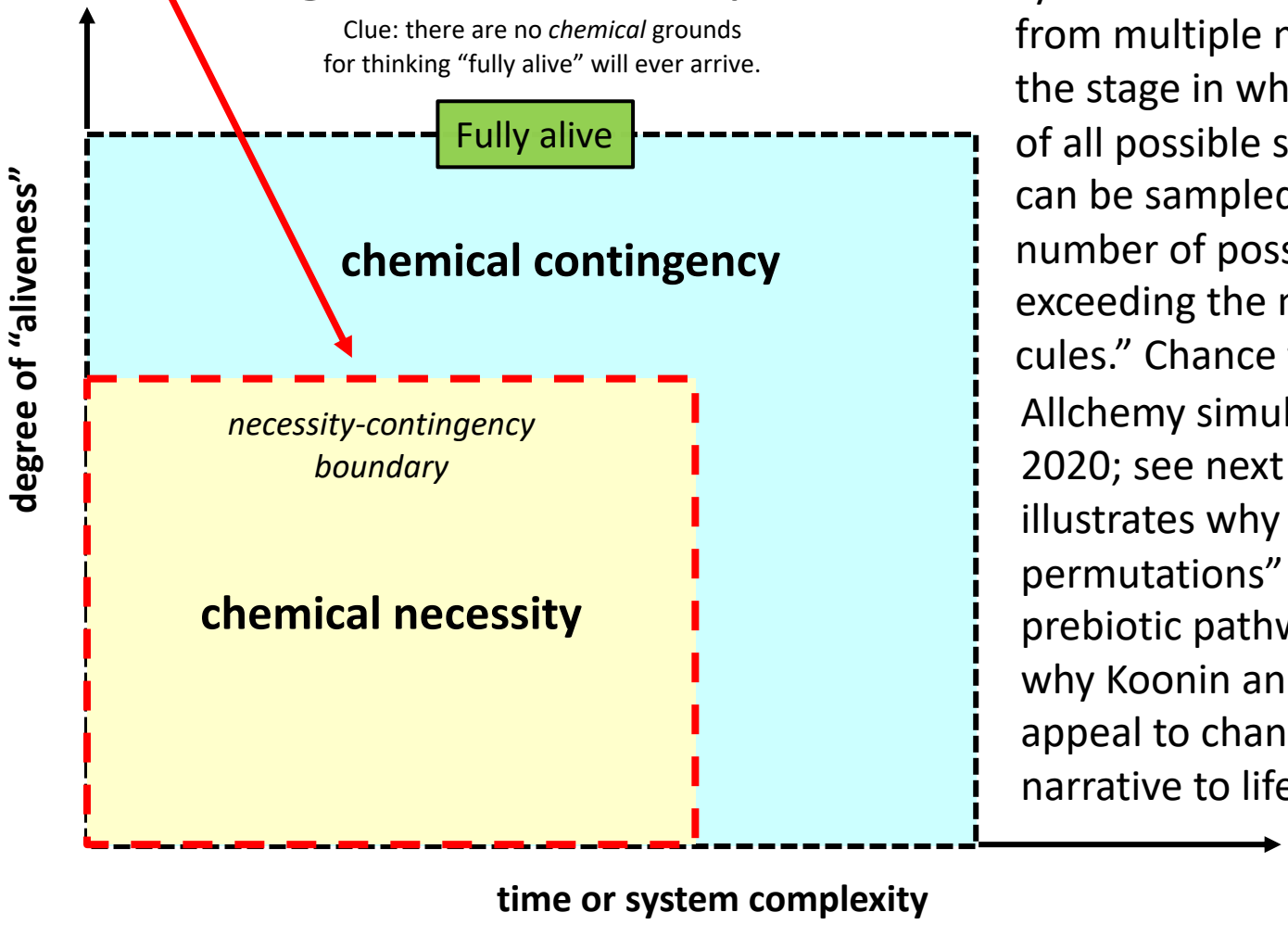
This figure and its caption from Sutherland (2017) are also telling. Why is there a "necessity-contingency" boundary – and why is that boundary placed where it is?



necessity-contingency boundary

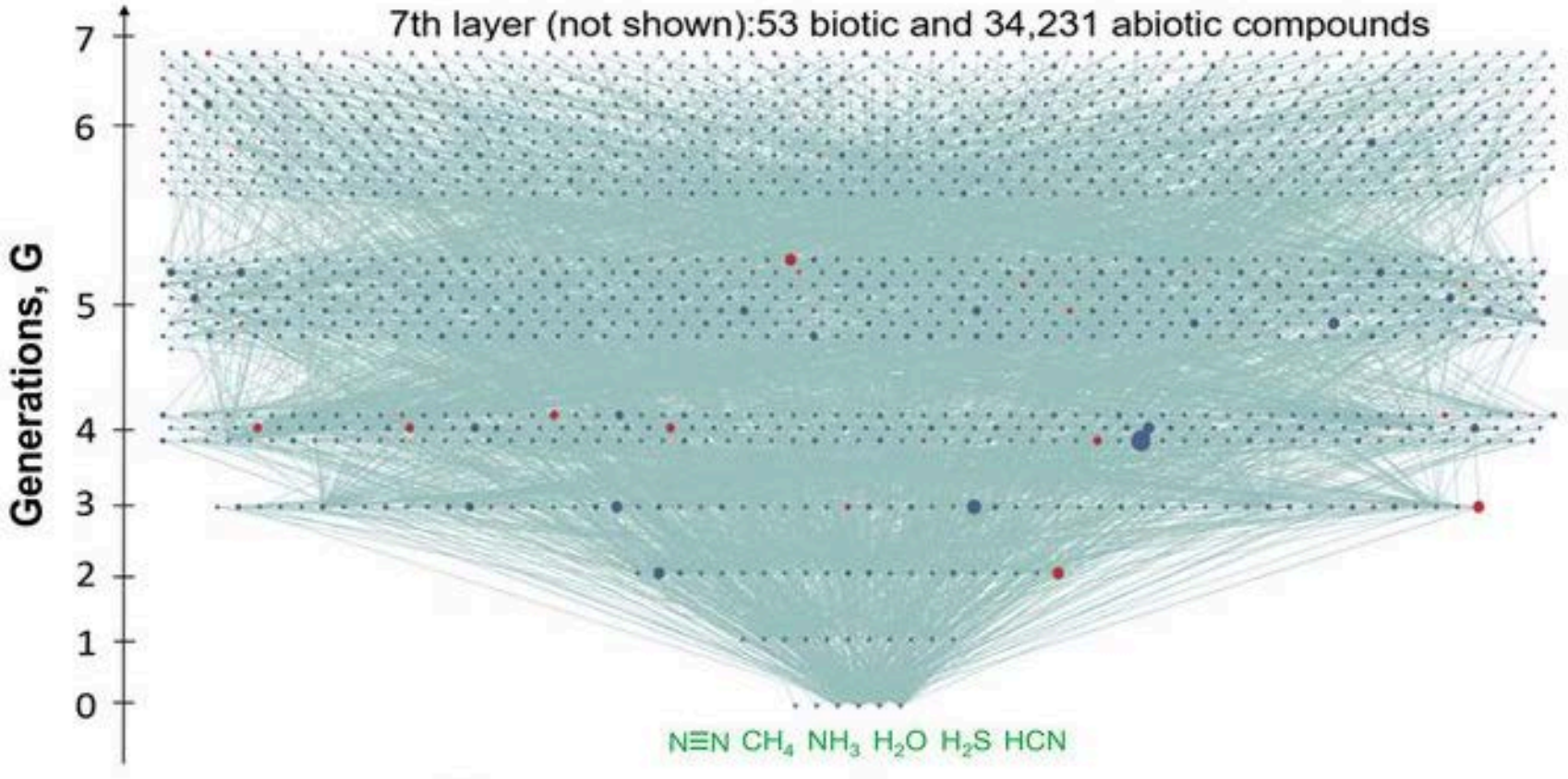
Here is the caption: "Also shown is the necessity–contingency boundary beyond which material limitations prevent full exploration of the sequence space of macromolecules assembled from different monomeric building blocks; therefore,
chemical determinism can no longer be relied on as a source of innovation, and further improvements *have to be chanced upon* instead." (2017, 4; emphasis added)

Something significant happens **here** – and keeps on happening, in fact, in a way which this diagram fails to depict.



The relative areas in this figure aren't important: it is concepts which matter. According to John Sutherland (2017, 4) "chemical contingency" occurs "when the synthesis of macromolecules from multiple monomers reaches the stage in which only a fraction of all possible sequence variants can be sampled owing to the number of possible permutations exceeding the number of molecules." Chance takes over. The Allchemy simulation (Wołos *et al.* 2020; see next slide) vividly illustrates why the "possible permutations" overwhelm the prebiotic pathways – and thus, why Koonin and Sutherland must appeal to chance to keep the narrative to life moving forward.

The Allchemy computer simulation (Wołos *et al*. 2020), available online at https://tol.allchemy.net, allows the user to explore the permutations which follow, in successive "generations," from a starting set of six early Earth molecules.



Nitrogen, methane, ammonia, water, hydrogen sulfide, and hydrogen cyanide comprise the starting set (i.e., generation 0). The red circles in the network of reactions represent the biologically relevant products. After 7 generations, **53 biotic and 34,231 abiotic compounds have arisen**.

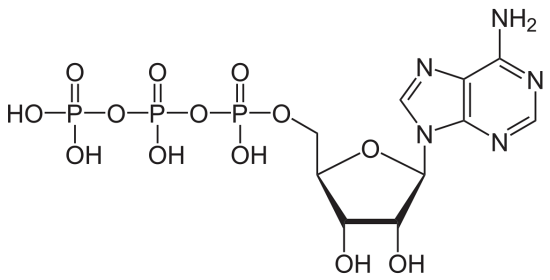
This is a striking ratio, but there is more to say:

$$\frac{\text{53 biotic}}{\text{34,231 abiotic}}$$

0.15 of one percent of the compounds in the Allchemy simulation, after 7 generations, are biologically relevant. The remainder (99.845 percent)? ***Abiotic.***

This is the same finding reported in the Kitadai and Maruyama analysis (above): "the abiotic synthesis of each nucleotide component (ribose and nucleobases) tends to generate complex mixtures of products with desired compounds being only a small fraction" (2018, 1136).

Now, here is the "more to say" bit. Let's suppose we want to produce **ATP**, in good yields, starting from nitrogen, methane, water, ammonia, etc.



Adenosine triphosphate (ATP)

No cell on Earth survives without ATP, the common or basic energy currency essential to metabolic pathways and cellular processes generally.

But to find ATP, abiogenesis would need to sort through everything else – that is, the enormous heap of compounds which are *not* biologically relevant.

"...because the masses of molecules like ATP, ADP or dinucleotides are high (above 400 g/mol), creating them from very basic substrates (HCN, H2O, CH4, N2, H3PO4) takes 9-13 synthetic generations within which extremely large numbers of other, not-very-interesting molecules are created."

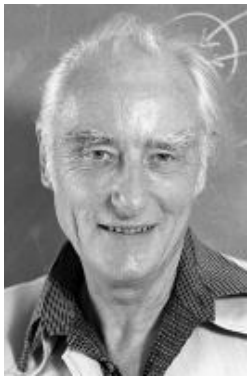
(as observed in the Allchemy simulation: Wołos *et al*. 2020, Supplementary Materials, 8; emphasis in original)

ATP is hiding in here – somewhere.
Good luck finding it.

Finding a few specific grains in a sandpile gives a visual metaphor for the reality of chemical contingency.

So what defines a molecule as "not very interesting"? It is made by physical processes which do not know (so to speak) that they should be constructing a cell. Hence, those processes generate "extremely large numbers" of *abiotic* products, thereby swamping the tiny set of biologically necessary compounds.

This problem of sifting through chemical contingency to locate the few & improbable pathways to biomolecules, and eventually the living state, has long been understood by abiogenesis researchers.



Francis Crick
1916-2004

In *Life Itself*, Crick touched repeatedly on the mechanistic hurdles of discriminating biotic from abiotic molecules, before the precision of enzymes was available. "It is not easy to see how this could happen [i.e., the formation of RNA] in a mixture of other, rather similar compounds without the frequent incorporation of other molecules in the chain unless there were some rather specific catalyst present...[biomolecule] precursors need to be at least partially separated from other, rather similar molecules, which, if present, might possibly have fouled up the system." (1981, 81-83)

Robert Shapiro
1935-2011

Also addressing the origin of RNA, Shapiro pointed out that chemical contingency was unavoidable: "The implicit assumption has been that monomers of a single chemical type would seek each other out in a prebiotic mixture and combine exclusively with one another. No theoretical or experimental basis has been put forward to support such an assumption, however, and considerations of entropy would lead in the opposite direction: **The components of a mixture should combine haphazardly, producing chaotic polymers.**" (2000, 174; emphasis added)

“Emergence of highly complex systems by chance is inevitable.”

Fortuna, goddess of fortune and personification of luck.
Also: ***enemy of knowledge.***

Yet if Koonin’s claim (above) about the inevitability of the origin of highly complex systems *via chance alone* is true, there is no point in looking for testable prebiotic path-ways to ATP, nucleotides, cells, or really anything else biological. No worries: Fortuna will take care of it.

But she won’t; she never will. ***In science, Fortuna takes, but she doesn’t give.*** The invocation of blind chance is the antithesis of empirical knowledge. If you doubt this, imagine a molecular biology or genetics classroom where (for instance) the processes of eukaryotic chromosomal segregation, or DNA replication, are to be taught. Except the professor writes “It’s all chance” on the whiteboard.

That class is over. Everybody can go home.

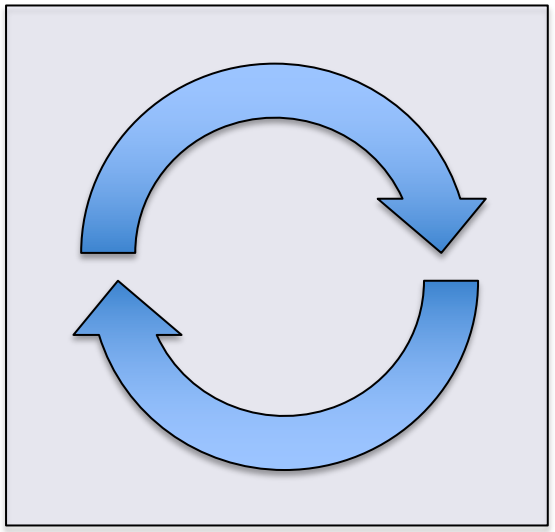
There is an infinite distance between naturalism (philosophical *or* methodological – the particular flavor doesn't matter), which relies ultimately on the whims of chance, and genuine empirical knowledge, grounded in evidence.

That infinite distance cannot be traversed by argument, observation, or logic. Reason requires boundaries. Bare possibility, however – i.e., metaphysical chance – is unbounded *in principle,* which is exactly why it is invoked. To cite Monod (1971) again: "Our number came up in the Monte Carlo game."

Choose knowledge instead.

There is no reasoning with a metaphysics of ultimate chance, so let's just forget about it. We have another way to understand ATP and other universally essential biomolecules – which brings us back to the evidence, and the concept of **causal circularity**.

If causal circularity holds broadly in Earth life, we should find it at all levels – from cellular to developmental.



As will be explained in greater detail below, organisms themselves give us the clues that we need. But Kun *et al.* (2008) can get us started down that road.

Research

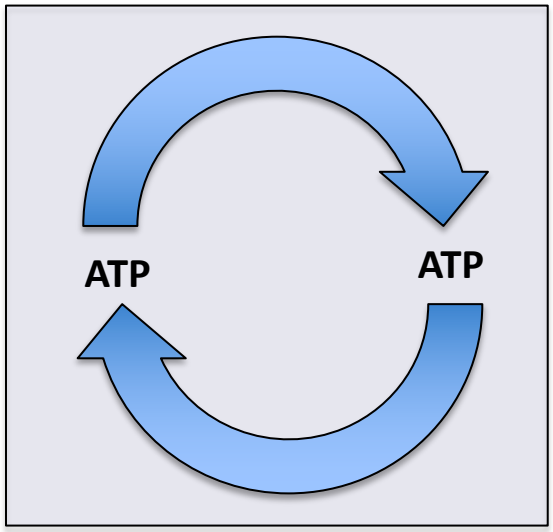
Computational identification of obligatorily autocatalytic replicators embedded in metabolic networks

Ádám Kun*†, Balázs Papp‡¶ and Eörs Szathmáry*†§

Open Access

The title is a mouthful, but the core idea is easy to grasp.

"To make X, you need X." That is what "obligatorily autocatalytic" means – and the main exhibit in Kun *et al.*'s (2008) case? **ATP.** See the selection from their abstract, under the figure.



Background: If chemical A is necessary for the synthesis of more chemical A, then A has the power of replication (such systems are known as autocatalytic systems). We provide the first systems-level analysis searching for small-molecular autocatalytic components in the metabolisms of diverse organisms, including an inferred minimal metabolism.

Results: We find that intermediary metabolism is invariably autocatalytic for ATP.

We will return to this pattern of evidence later. First, consider Kun *et al.*'s analysis.

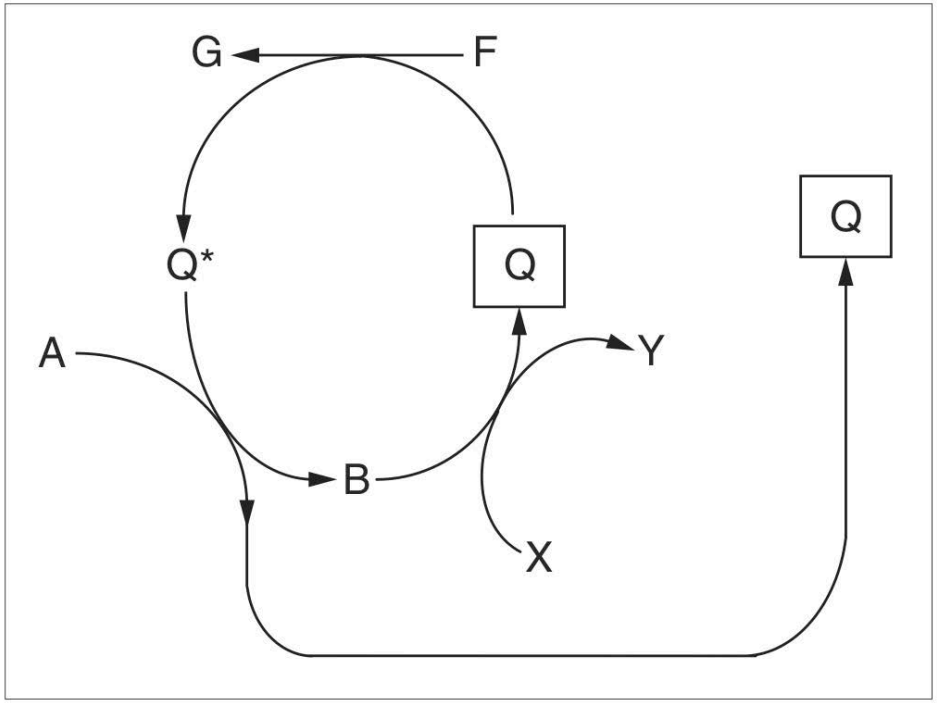


Figure 3
General scheme for the autocatalytic synthesis of a coenzyme. The coenzyme (carrier molecule) and its loaded form are denoted by Q and Q*, respectively. A, X and F are external compounds provided to the cycle. B is an intermediate in the cycle. G and Y are by-products of the cycle. B can be considered an autocatalyst if X is provided as an external compound. However, if X is an intermediate of the whole network (that is, an internal compound) then providing B does not necessarily launch the cycle because biosynthesis of X might require the presence of coenzyme Q*.

Using this schema, Kun et al. found that not only ATP, but nicotinamide adenine dinucleotide (NAD), coenzyme A, and other co-factors were “to make X, you need X” molecules.

Their bottom line?

Conclusion: Metabolic replicators are apparently common and potentially both universal and ancestral: without their presence, kick-starting metabolic networks is impossible, even if all enzymes and genes are present in the same cell. Identification of metabolic replicators is also important for attempts to create synthetic cells, as some of these autocatalytic molecules will presumably be needed to be added to the system as, by definition, the system cannot synthesize them without their initial presence.

A case can be made – I won't do so here, this presentation is already way too long – that abiogenesis research in the 20th century has shown that the bottom-up pathway to cells never happened. We can adopt a provisional hypothesis, however.

How about this: **organisms appear to be "system-first," irreducible entities** – and therefore, we should treat them as such. Will this stall out biological inquiry?

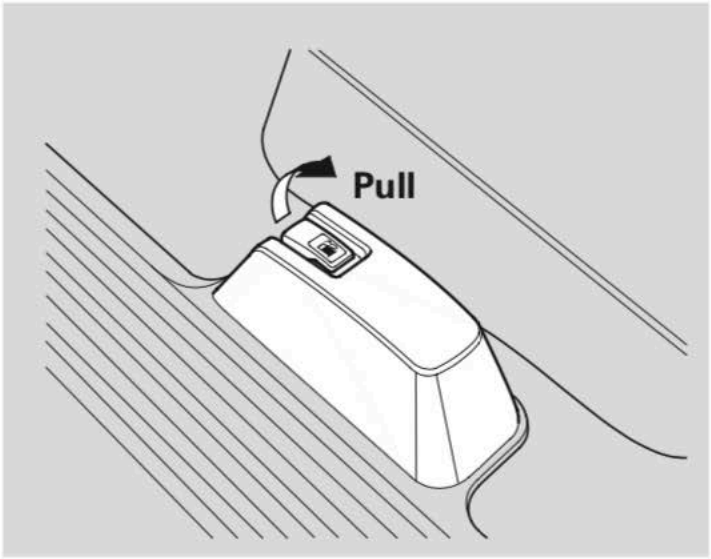
The remainder of this presentation says the answer to that question is No.

III. The Basic Logic of Functional Triangulation

Why won't the left sliding door of my Honda Odyssey open?

Let's start with my 2003 Honda Odyssey.

Filling the Fuel Tank



1. Because the fuel fill cap is on the driver's side of the vehicle, park with that side closest to the service station pumps.

2. Open the fuel fill door by pulling on the handle to the left of the driver's seat.

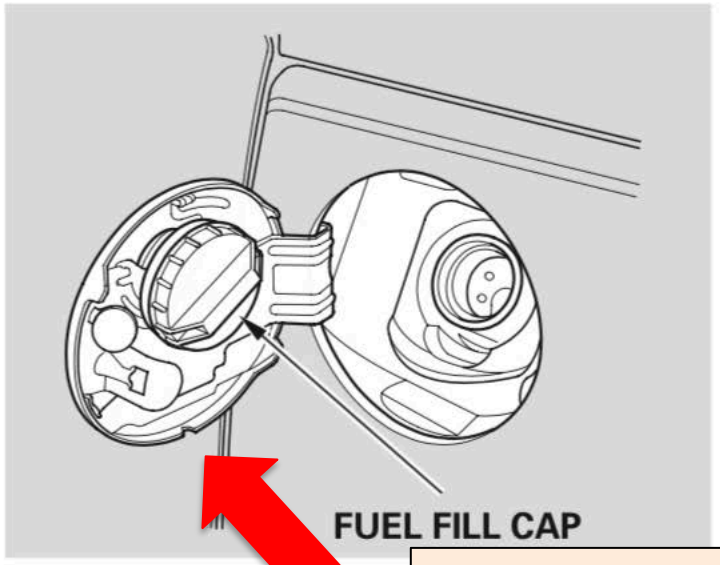
Before refueling, make sure the rear sliding door on the driver's side is closed.

BUT WHY?

⚠ WARNING

Gasoline is highly flammable and explosive. You can be burned or seriously injured when handling fuel.

- Stop the engine and keep heat, sparks, and flame away.
- Handle fuel only outdoors.
- Wipe up spills immediately.



(here's a clue)

Why do I need to make sure, Honda Corporation?

Because two macroscopic objects with mass (like doors) cannot occupy the same location in space at the same time.

automatic sliding door

fuel fill door

And if you're stupid enough to try (Paul), *we've designed the car to deal with that.*

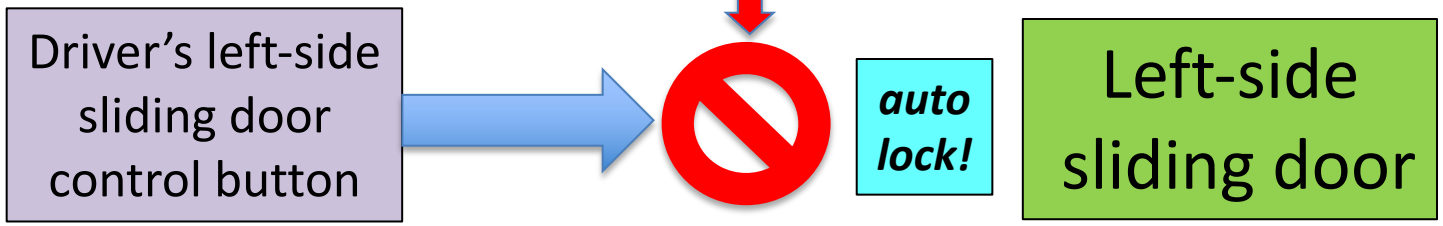
A Note About Refueling

Before refueling, make sure the driver's side sliding door is fully closed. When you release the fuel fill door, the driver's side sliding door automatically locks so it cannot open and interfere with the fuel door.

During the design ("foresight") stage of Odyssey development:

Honda engineers: hey, some drivers are clueless. They'll need help so they won't jam the left side sliding door.

Wait a minute, ***is the fuel door open?***



Now suppose you are Paul's younger brother Peter Nelson MD – who is a situationally aware and clued-in person, not a badly distracted philosopher of science.

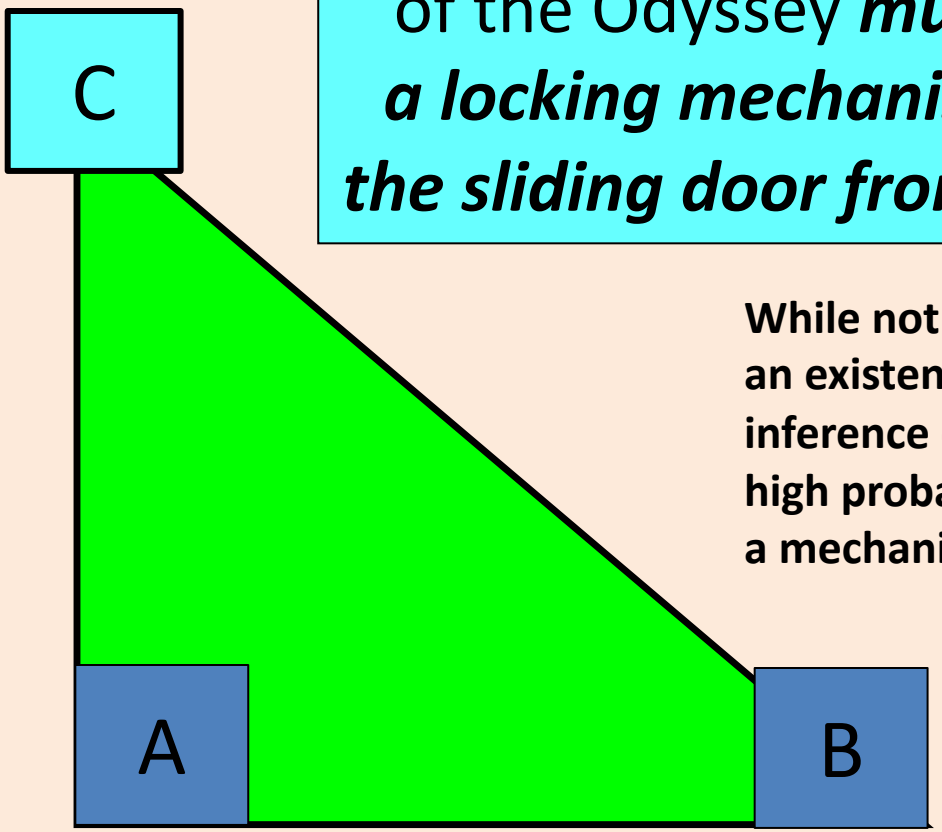
Peter, who is not an Odyssey owner, and has no experience with this vehicle, studies the minivan doors (sliding and fuel) while waiting for Paul in the driveway. He says to himself – hmm:

Honda must have a foolsafe mechanism to prevent door jamming.

Peter triangulates to an unobserved, but necessary, function of the system, as follows.

Functional triangulation (everyday version):

Therefore, the global design of the Odyssey ***must include a locking mechanism to stop the sliding door from jamming***.

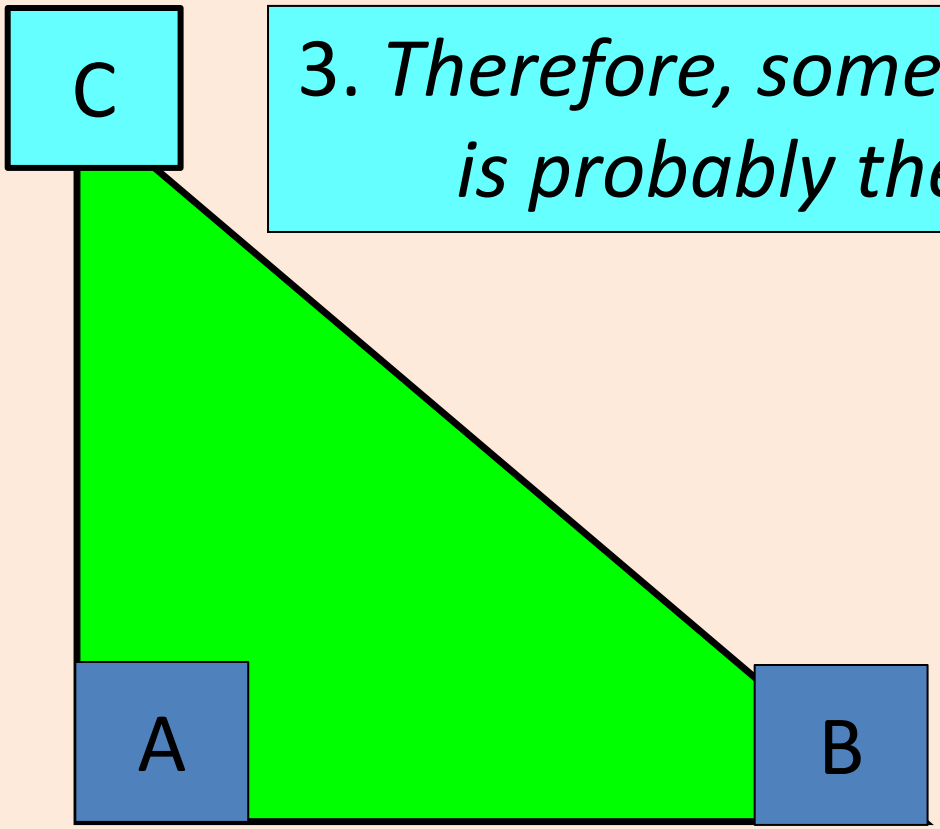


While not strictly speaking an existence proof, this inference confers a very high probability that such a mechanism operates.

The sliding door moves to left rear, over the fuel door.

If the fuel door were open, the sliding door would jam.

Triangulation as a metaphor for scientific inquiry, leading to novel predictions:

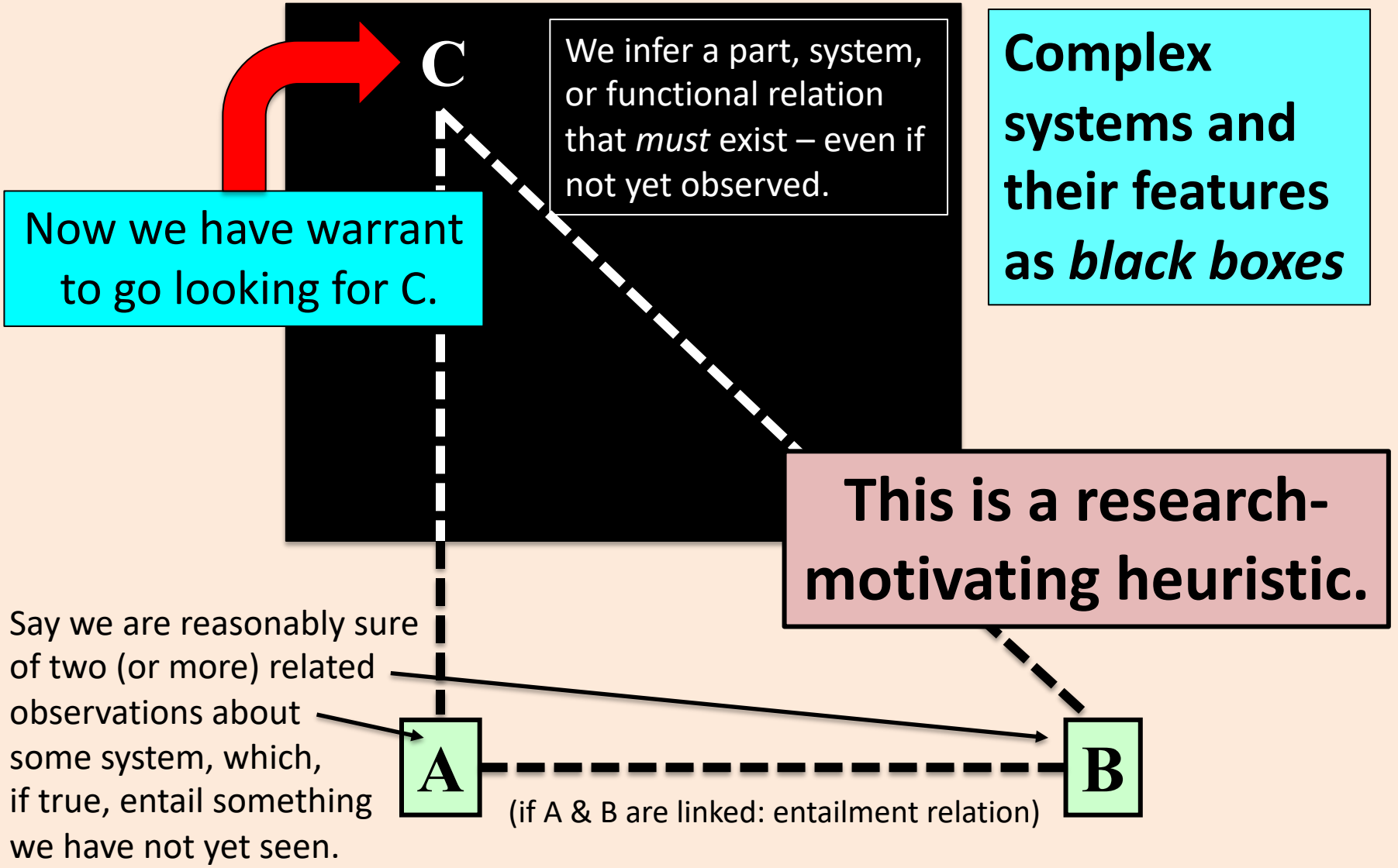


3. *Therefore, something like C is probably the case.*

1. We know A.

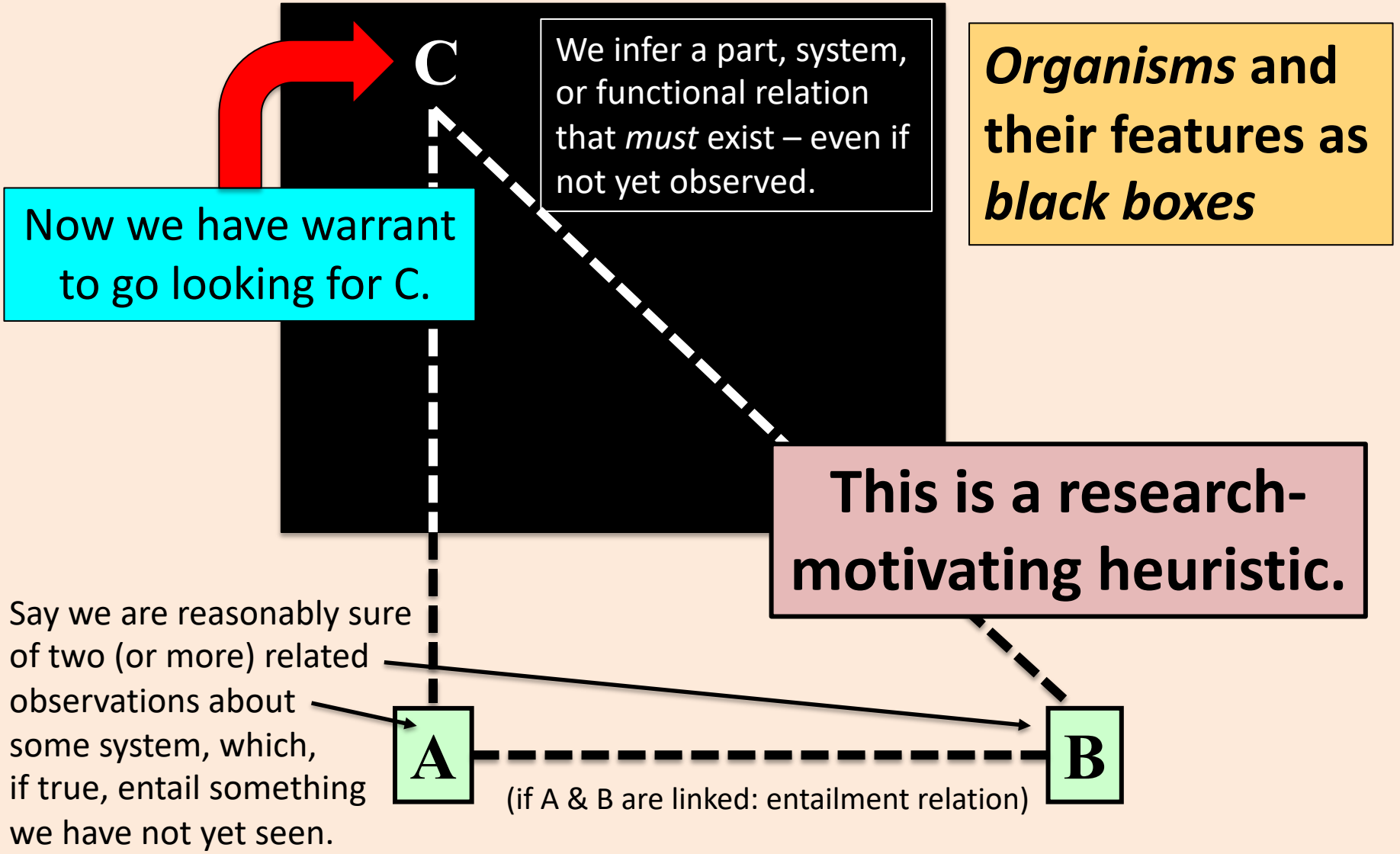
2. We know B.

Functional triangulation as a research strategy



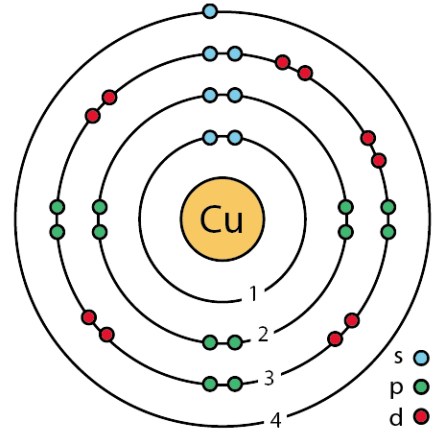
IV. Functional Triangulation in Biology

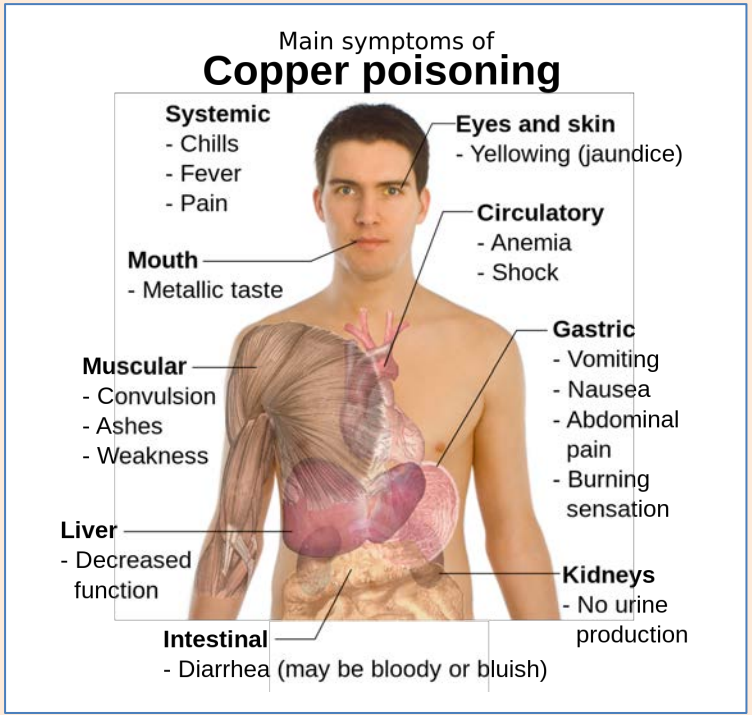
Let's apply the triangulation method to living things:



Consider a pair of biologically related facts:

1. **Copper is a poison.**

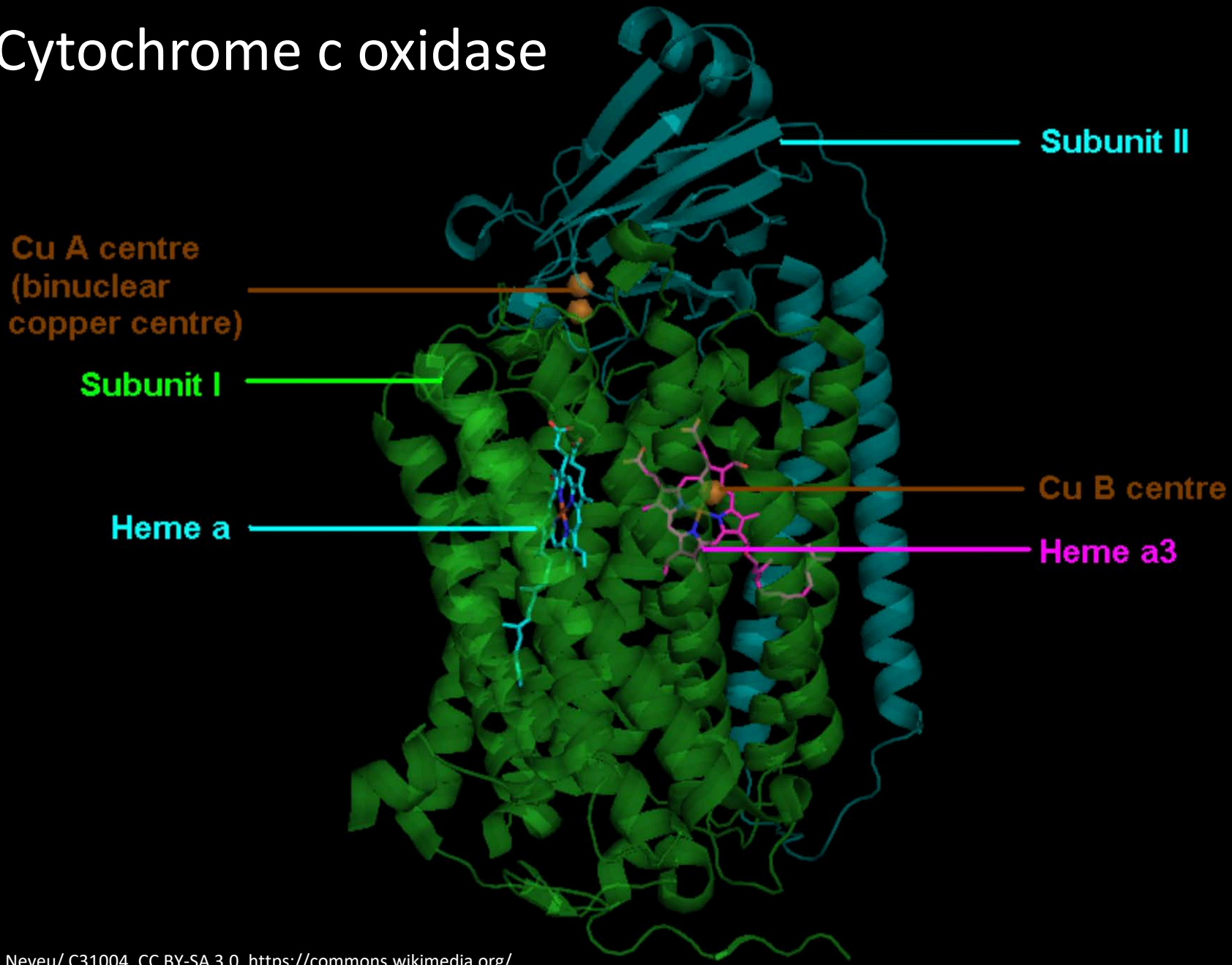




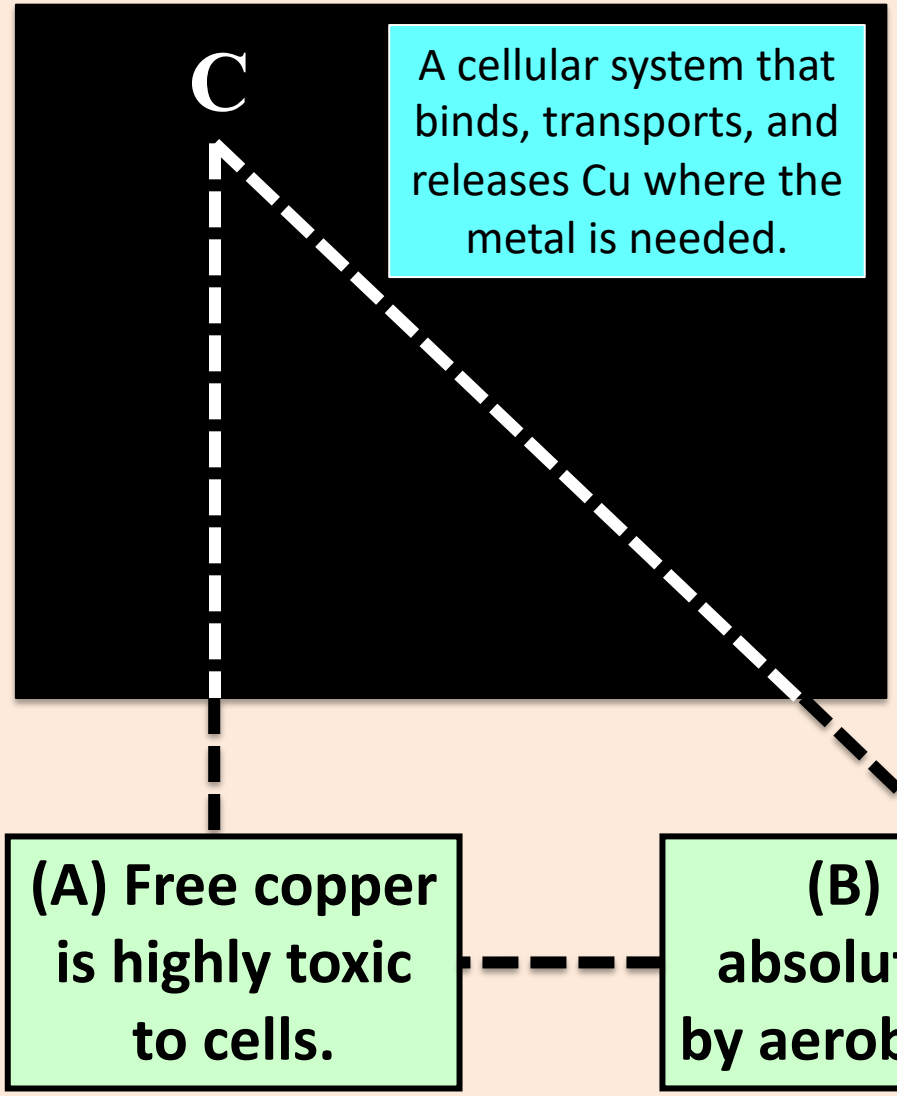
2. **Copper is absolutely required by aerobically respiring organisms (like you), as an essential co-factor in several enzymes.**

(illustration credit Mikael Häggström, Wikicommons)

Cytochrome c oxidase



Trust the logic: *triangulate.*



What MUST exist in aerobically respiring organisms?

The Skeptics

“But wait a minute…how can you be certain the copper-binding system is really there in the cell?”

Don’t worry: it’s there.

(Hey Alpha C – are you guys back already? Batteries fully charged? That was only 29 slides.)

Organisms make use of mechanisms for their very viability; whether we see the mechanisms or not, *they exist*.

Call this the ***no magic principle***.

This "existence proof" for a Cu binding & transport system was acknowledged *long before* the system was actually observed:

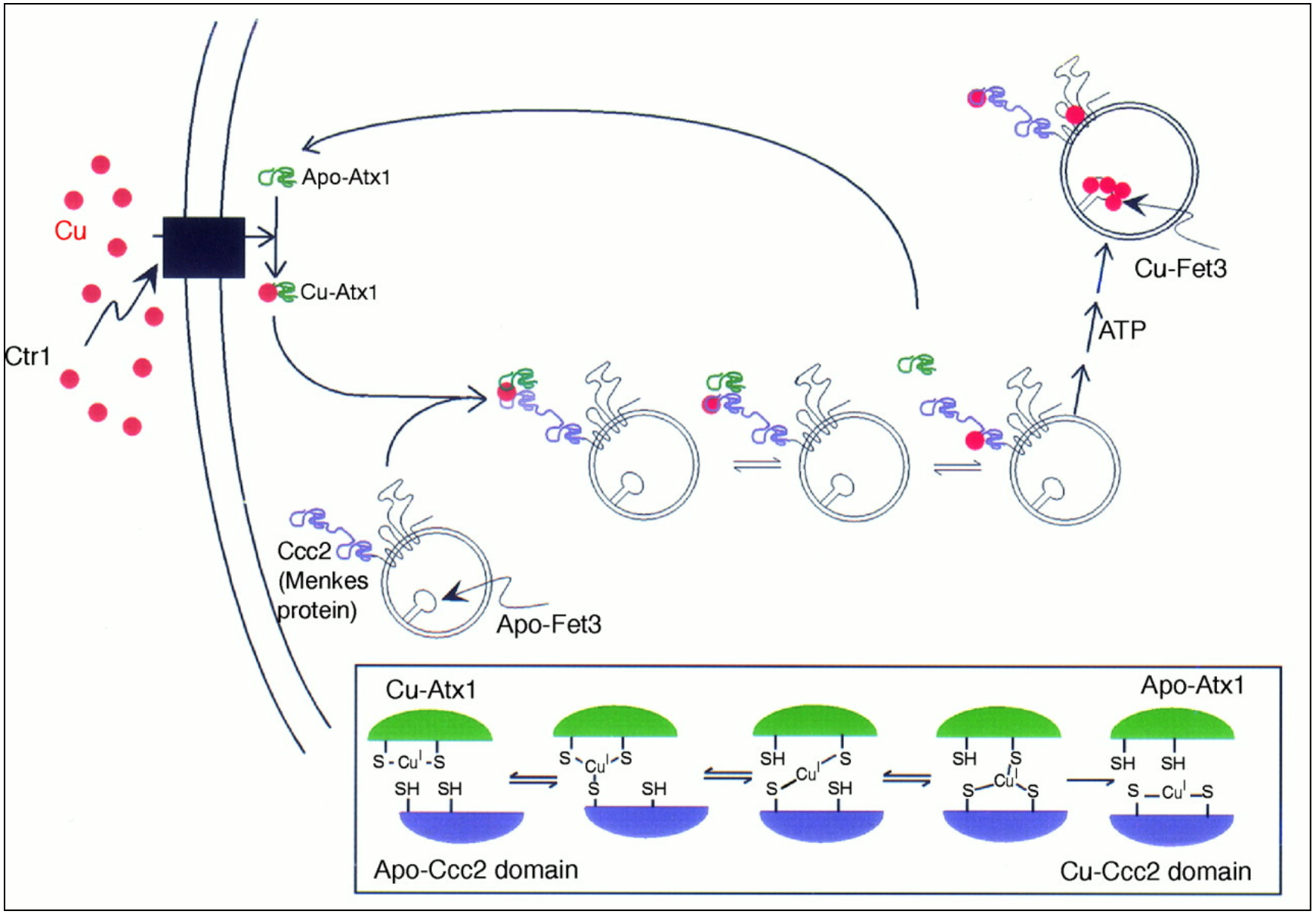
"Copper is absolutely required for aerobic life, and yet, paradoxically, is highly toxic. … This apparent contradiction has been rationalized by assuming that Cu, like other redox-active metals, is sequestered in nonactive forms as it is transported through cellular compartments."

Valentine and Gralla, *Science* 278 (1997):817.

This "existence proof" for a Cu binding & transport system was acknowledged *long before* the system was actually observed:

"However, the agents of such trafficking and the mechanisms of delivery of Cu to its final destinations have, until recently, remained largely unknown."

Valentine and Gralla, *Science* 278 (1997):817.



R.A. Pufahl *et al*.,, *Science* 278 (1997):855.

and F. Whitby for assistance with data collection; V. Ramakrishnan and members of the Sundquist and Hill laboratories for critical comments on the manuscript; M. Martin for plasmid pNL4-3; D. Trono for plasmid R9; and J. Cassatt for support and encouragement. Supported by NIH grants RO1 AI40333 and RO1 AI43036 (W.I.S. and C.P.H.), the Lucille P. Markey Charitable Trust, and a postdoctoral fellowship from the Cancer Research Institute (T.R.G.). Coordinates (1am3) and diffraction data (r1am3sf) for GA151–231) have been deposited in the Protein Data Bank (Brookhaven National Laboratory).

17 March 1997; accepted 23 September 1997

Metal Ion Chaperone Function of the Soluble Cu(I) Receptor Atx1

R. A. Pufahl, C. P. Singer, K. L. Peariso, S.-J. Lin, P. J. Schmidt, C. J. Fahrni, V. Cizewski Culotta, J. E. Penner-Hahn, T. V. O'Halloran*

Reactive and potentially toxic cofactors such as copper ions are imported into eukaryotic cells and incorporated into target proteins by unknown mechanisms. Atx1, a prototypical copper chaperone protein from yeast, has now been shown to act as a soluble cytoplasmic copper(I) receptor that can adopt either a two- or three-coordinate metal center in the active site. Atx1 also associated directly with the Atx1-like cytosolic domains of Ccc2, a vesicular protein defined in genetic studies as a member of the copper-trafficking pathway. The unusual structure and dynamics of Atx1 suggest a copper exchange function for this protein and related domains in the Menkes and Wilson disease proteins.

Although Cu is an essential cofactor for mitochondrial, cytosolic, and vesicular oxygen-processing enzymes (*1*), it can be toxic even at low concentrations. Cu(I) and Cu(II) ions can bind with high affinity to adventitious sites in partially folded proteins and catalyze auto-oxidation of lipids, proteins, and nucleic acids. To investigate the mechanisms by which cells overcome the dilemma of maintaining Cu availability while controlling deleterious reactivity of the free ions, we have determined the Cu chemistry and Cu-handling function of Atx1, an intracellular eukaryotic protein implicated in Cu trafficking. Our results indicate that Atx1 functions as a metal ion chaperone, a protein that protects and guides Cu(I) ions to activate target enzymes.

ATX1 is one of several genes in the Cu-dependent, high-affinity iron uptake pathway in yeast. These genes encode Ctr1, a Cu uptake protein in the plasma membrane; Ccc2, an intracellular membrane protein; and the multicopper oxidase Fet3 (*2*–*6*). Although the role of Fet3 in iron uptake is unclear, Cu loading into this enzyme occurs in a post-Golgi vesicle and is mediated by Ccc2 (*4*). The *ATX1* gene and its human homolog, *HAH1*, encode cytosolic proteins implicated in Cu trafficking to these Ccc2-containing vesicles (*7*–*9*). The Ccc2 protein in yeast and its human homologs, the Menkes disease protein (*10*) and Wilson disease protein (*11*), are members of the P-type adenosine triphosphatase (ATPase) cation transporter family (*12*) and are present in the membranes of secretory vesicles (*13*). Each transporter contains two or more Atx1-like cytoplasmic domains (Fig. 1), the functions of which are not known. The conserved MTCXXC sequence (X, any residue), a motif observed in several bacterial Hg(II) transport proteins (*14*), is thought to be a Cu binding site, although it does not correspond to known Cu(I) or Cu(II) sites in structurally characterized proteins. Establishing the Cu oxidation state and coordination environment in Atx1 should provide insight into its function and that of the homologous domains in its partner proteins.

Atx1 was expressed in *Escherichia coli* and purified as the apoprotein (*15*). Addition of $CuSO_4$ to apo-Atx1 in the presence of excess thiol reproducibly yielded a complex with a copper/protein ratio of 0.6 to 0.8 (*16*). Gel filtration experiments indicated that the predominant form of the protein was a monomer under these conditions, regardless of the presence or absence of Cu (*15*). The Cu oxidation state was investigated by electron paramagnetic resonance (EPR) spectroscopy and x-ray absorption spectroscopy (XAS). No EPR signal was observed at 77 K, suggesting the absence of a mononuclear Cu(II) site. XAS experiments (*17*) indicated that the bulk sample contained Cu(I). The Cu-Atx1 x-ray absorption near edge structure (XANES) spectrum exhibited a weak shoulder at 8984 eV, which is typical of Cu(I) and inconsistent with comparable edge features for Cu(II) compounds, the energies of which are 3 to 4 eV greater (Fig. 2A). The intensity of the 8984-eV transition varies with Cu(I) geometry, ranging from low for tetrahedral sites to high and well resolved for digonal sites (*18*). The observed transition is typical of those of trigonal Cu(I) model compounds (Fig. 2A). Mononuclear Cu(I) coordination complexes, however, are usually not stable in aqueous solution, and typically undergo auto-oxidation or disproportionation reactions to give Cu(II) (aqueous) and Cu(solid). In contrast, millimolar concentrations of Cu(I)-Atx1 are stable in air at neutral pH for at least 30 min (*16*), suggesting that the coordination environment in Atx1 stabilizes the Cu(I) state and suppresses disproportionation.

Proteins that stabilize Cu(I) form either polynuclear metal thiolate clusters, as in metallothionein and the transcription factors Ace1, CUP2, and AMT (*19*), or a constrained His_3Cys coordination environment, as in blue copper proteins (*20*). In contrast, extended x-ray absorption fine structure (EXAFS) measurements indicated that Atx1 stabilization of Cu(I) is achieved in a mononuclear site through an all-sulfur coordination environment (Fig. 2B). The high intensity and relatively high frequency of the EXAFS oscillations were typical of those observed for Cu-S interactions. The data could be modeled with a single three-sulfur (3S) shell with a Cu-S distance of 2.26 Å; however, the Debye-Waller factor, σ^2, was somewhat larger than expected (7×10^{-3} Å^2). This Cu-S distance is typical of three-coordinate Cu (*21*–*23*) and is 0.1 Å longer than that in two-coordinate Cu-thiolate complexes (*24*). Given that Atx1 contains only two conserved cysteine residues, the third ligand can be either a low-Z (atomic number) ligand (O or N), a methionine sulfur, or an exogenous thiol. A 2S + 1S fit with a shell of two sulfurs at 2.25 Å and a shell of one sulfur at 2.40 Å reproduced the data better than did the single-shell fit, and gave more reasonable σ^2 values (3×10^{-3} and 5×10^{-3} to 6×10^{-3} Å^2, respectively). Furthermore, attempts to model the data with a shell of 2S and a low-Z shell (1N) gave chemically unrealistic Debye-Waller factors for the N shell ($\sigma^2 \approx 0$).

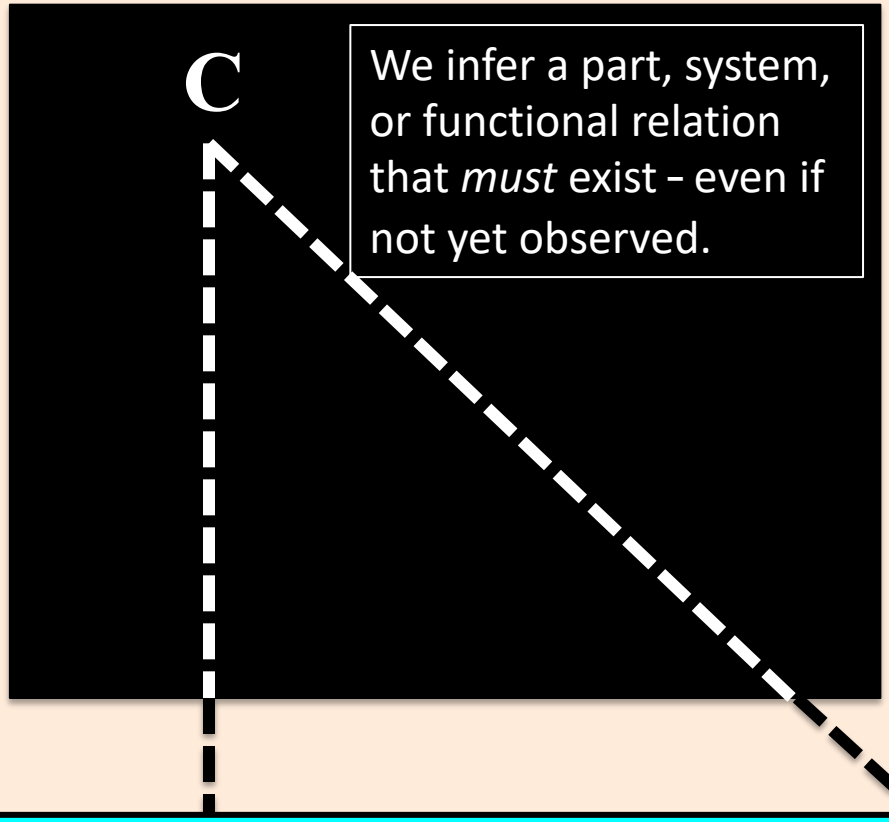
The 2S + 1S result is unexpected for several reasons. The most common coordination environment for Cu(I) complexes is distorted four-coordinate, and there are few precedents among mononuclear Cu(I) complexes for the low-coordination number environ-

R. A. Pufahl, C. P. Singer, C. J. Fahrni, T. V. O'Halloran, Department of Chemistry and Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, IL 60208, USA.
K. L. Peariso and J. E. Penner-Hahn, Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, USA.
S.-J. Lin, P. J. Schmidt, V. Cizewski Culotta, Department of Environmental Health Sciences, Johns Hopkins University, Baltimore, MD 21205, USA.
*To whom correspondence should be addressed. E-mail: t-ohalloran@nwu.edu

R.A. Pufahl *et al*., "Metal Ion Chaperone Function of the Soluble Cu(I) Receptor Atx1," *Science* 278 (1997):853-856

By delivering essential cofactors or substrates to apoenzymes, the emerging class of metal ion chaperones facilitates formation of an active state of a protein. These chaperones guide metal ions to their appropriate biological partners and protect them from being trapped at adventitious sites. They may also protect cellular components by sequestering specific ions or inorganic clusters (34) and preventing adverse reactions. Our results underscore the idea that cells make use of elaborate machinery for recruiting, trafficking, compartmentalizing, and, ultimately, inserting into the appropriate enzyme reactive cofactors such as mononuclear Cu ions. The intracellular activation of apometalloenzymes by binding of the correct metal ion cofactors is unlikely to proceed by spontaneous self-assembly. Rather, metal insertion is emerging as an orchestrated event controlled by metal ion transport and chaperone proteins whose functions are kinetically and thermodynamically coupled.

This works, as the history of biology shows (see, for example, the next two episodes). But *why*?



Organisms and their features as *black boxes*

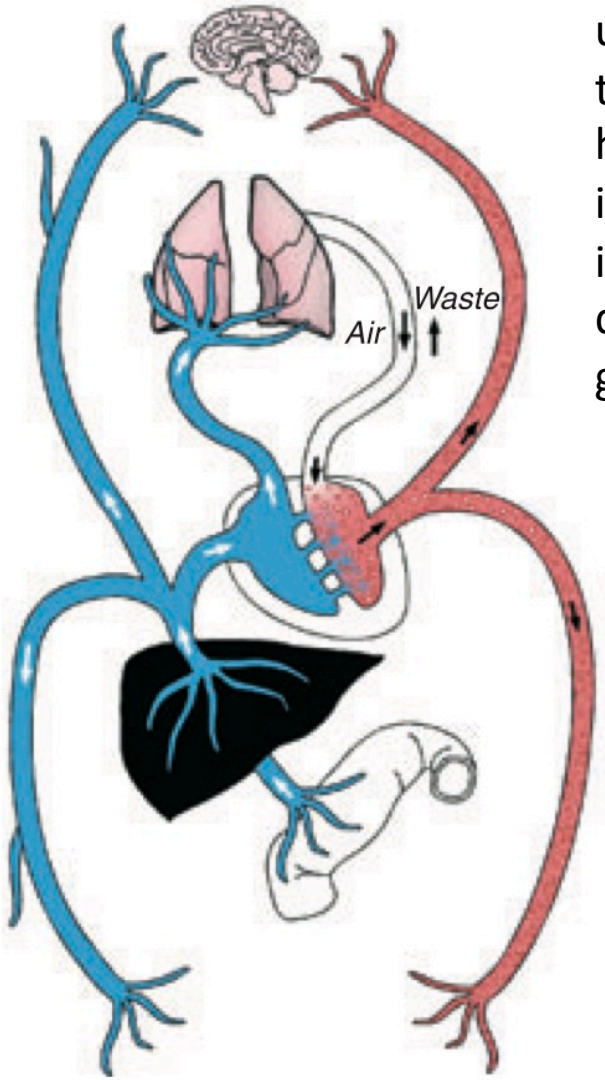
Seriously – ask yourself ***why*** this method works.
What must we presuppose, as necessary, to get any such functional triangulation started?

William Harvey (1578-1657)

The authority of Galen (129-210 CE) dominated medicine and biology for centuries.

But Harvey asked, not if we should continue to follow Galen's authority, but rather if the latter's account of the movement of the blood actually *worked*.

Galen's open-ended vascular system (Air and blood in arteries; pores in heart)

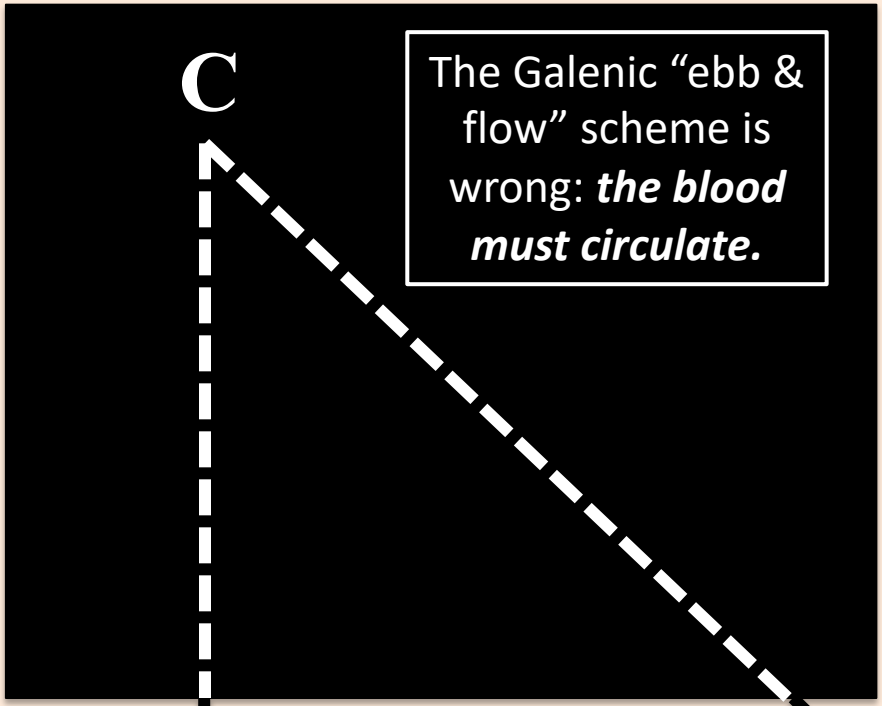


The complexity of Galen's system cannot be fitted onto a single slide. Indeed, his ideas are difficult for us now even to understand, given how deeply the concept of "circulation" has permeated our biological imagination. (Also, circulation is true, "which is nice," to quote the well-known caddy & groundskeeper Carl Spackler).

Unidirectional flow, as this diagram shows, is the main feature of the Galenic system to keep in mind. Blood originates in the liver; from there, it flows to the heart, lungs, brain, and tissues. But if blood flow is unidirectional, Harvey realized, testable consequences follow.

(figure after Aird 2011, 120)

“Harvey did not use teleology as final proof, but rather as a means to establish testable premises” (Aird 2011, 124).



William Harvey: “It is a matter of necessity that the blood performs a circuit, that it return whence it set out.”

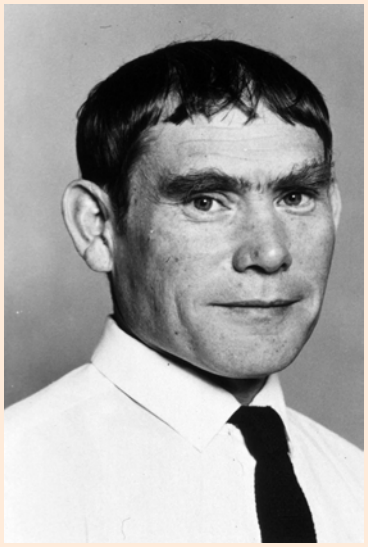
(A) If the left ventricle discharges ≈ three ounces of blood per beat, then even at 33 beats / minute, the total volume pumped would be enormous.

(B) But the measured total volume of blood in dogs and sheep does not exceed a few pounds, far below the total pumped volume estimated in (A).

William Harvey: “It is a matter of necessity...”

The “necessity” in biology is not *physical* necessity (e.g., as in the ideal gas law), but rather ***systems-level functional necessity***.

Organisms “do the sums” for us. Their viability comprises, as the global, or governing, state, all lower-level entities and their relations – with observationally accessible (empirical) signals.
The organisms give us the clues we need.



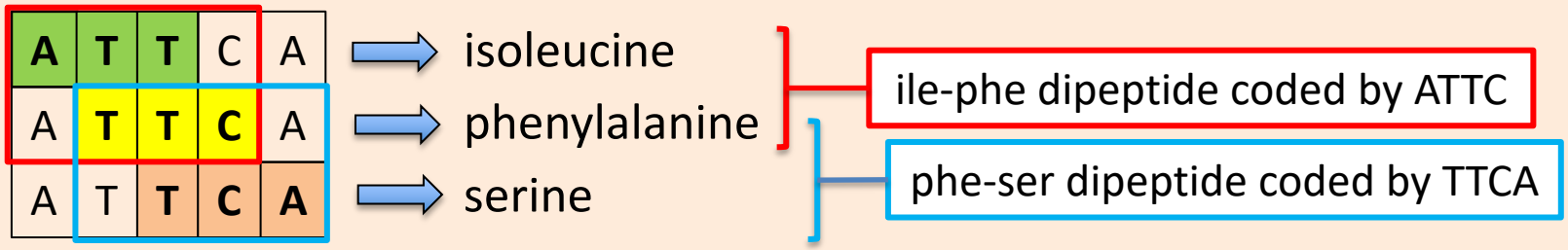
Sydney Brenner
(1927-2019)

In the late 1950s, Francis Crick and Sydney Brenner shared an office at the Cavendish Laboratory of Cambridge University. They talked nonstop about the "coding problem" and its possible solutions.

At the time, one possibility was an *overlapping* code. DNA comprises a four "letter" alphabet: **A**denine, **C**ytosine, **T**hymine, **G**uanine. To specify ***20*** amino acids, therefore, the "codons" (a term coined by Brenner) needed to have more than one nucleotide. Codons of length 3 would enable 64 different specifications ($4^3 = 64$) from A, T, C and G.

The physicist George Gamow proposed an over-lapping code, hypothesizing that the molecular structure of DNA itself directly templated amino acids in protein assembly.

The chemical details of Gamow's hypothesis are interesting, but they are not the main issue. That, rather, is the **mathematics of coding & information transfer between the DNA alphabet (A, T, C and G) and 20 amino acids.**



- Any pair of amino acids is a ***dipeptide***. In an overlapping code, dipeptides must be specified by a four-nucleotide sequence. **256** such 4-nucleotide sequences are available ($4^4 = 256$), in the DNA alphabet, to code for dipeptides.

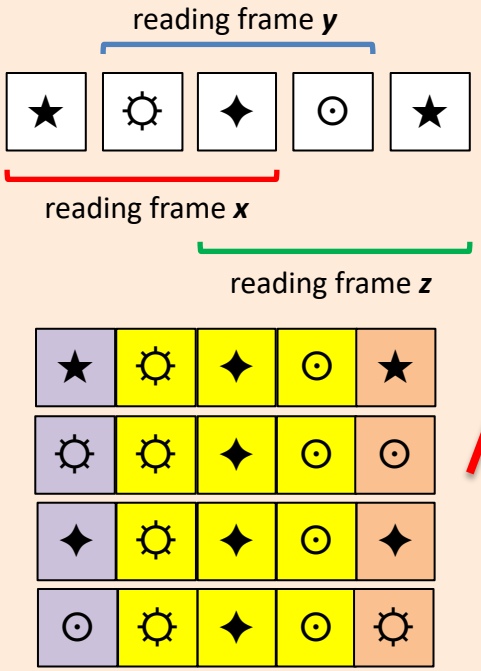
- But in the 20 amino acid alphabet, **400** dipeptides are *possible* ($20^2 = 400$).

***Taking 256 ≠ 400 as a clue*, Brenner (1957, 688) observed: "Thus overlapping codes introduce restrictions in amino acid sequences."**

Actual protein sequences should tell one if such restrictions were followed by amino acid neighbors – or not. Brenner compiled the existing protein sequences (in 1957) and calculated the minimum number of triplets required.

"The proof" (Brenner's term: 1957, 688)

1. "Since successive triplets [in an overlapping code] share two nucleotides in common, any given triplet can be preceded by only four different triplets and succeeded by only four different triplets."



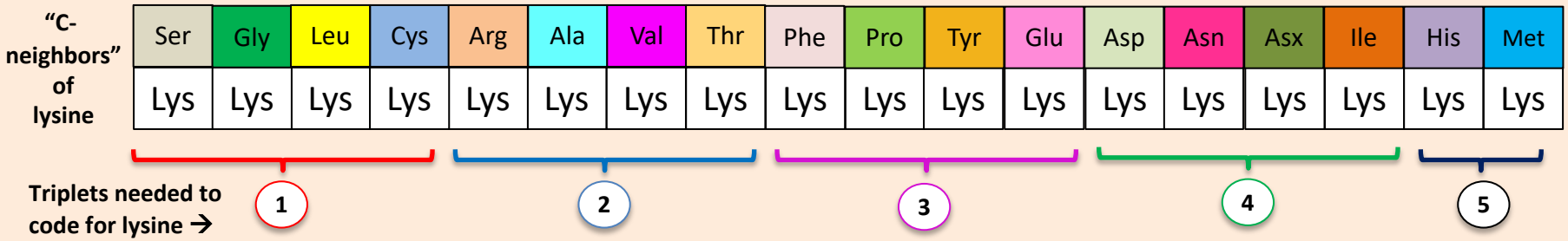
This 5 x 4 grid represents the possible symbol combinations, in an *overlapping* code, for reading frames ***x, y***, and ***z,*** when the code comprises a four-symbol alphabet, a triplet coding convention, and reading frame ***y*** (in yellow) codes for one amino acid.

2. Overlapping codes restrict the possible neighbors of any reading frame. In the example given, reading frame ***y*** can have only four possible ***x*** neighbors, and only four possible ***z*** neighbors (as coding triplets).

3. "In an amino acid sequence *j.k.l.* [***x, y, z***], we call *j* an N-neighbor, and *l* a C-neighbor, of *k*. For every four different N-neighbors (or C-neighbors) or part thereof, *k* must have one triplet assigned to it."

4. "Thus, the minimum number of triplet representations for each amino acid can be counted from a table of neighbors." ***Okay:*** **examine real sequences.**

”...sufficient sequences...prove that it is impossible to code them with overlapping triplets” (Brenner 1957, 688).



In any overlapping code, lysine’s C-neighbor amino acids would require at least five different triplets to code for lysine. Brenner compiled data from all available amino acid sequences (see table below) and summed the minimum number of triplets required: **70.**

TABLE. 1

Amino Acid	C-Neighbors	N-Neighbors	Minimum No. of Triplets Required	Amino Acid	C-Neighbors	N-Neighbors	Minimum No. of Triplets Required
Lys	18	17	5	Pro	13	12	4
Ser	17	13	5	Tyr	12	10	3
Gly	15	15	4	Glu	11	11	3
Leu	15	15	4	Glun	12	9	3
Cys	15	14	4	Asp	10	11	3
Arg	14	16	4	Asn	9	10	3
Ala	14	15	4	Ileu	9	9	3
Val	14	12	4	His	6	9	3
Thr	13	14	4	Met	5	7	2
Phe	13	14	4	Try	3	3	1
							Total 70

"The proof" (Brenner 1957, 688)

[*triplets needed*] **70 > 64** [*triplets possible*]

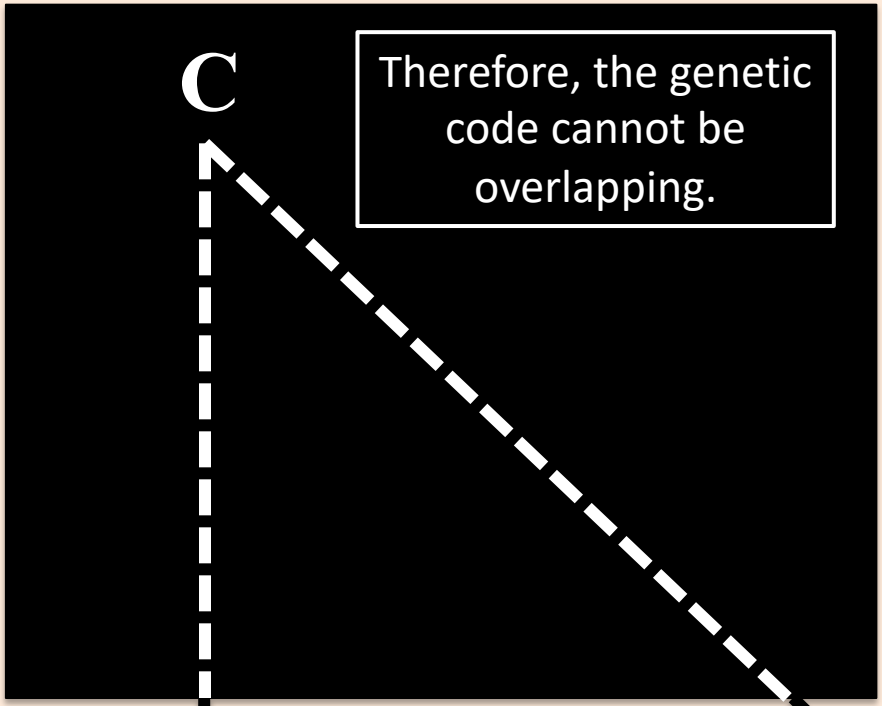
In a code with a four-symbol alphabet and a triplet convention, $4^3 = 64$ triplets exist. That's it. **"We conclude, then,"** writes Brenner, **"that all overlapping triplet codes are impossible."** The decisive character of that finding is highly significant.

ON THE IMPOSSIBILITY OF ALL OVERLAPPING TRIPLET CODES IN INFORMATION TRANSFER FROM NUCLEIC ACID TO PROTEINS

BY S. BRENNER

Being able to say *impossible* counts, because in science, knowing what *not* to look for – knowing what *cannot* be the case – is as valuable as having clues about where to look. Closing off an avenue that leads to a dead end saves everyone's time. The code cannot be overlapping: try something else. And they did.

“The proof is simple...We conclude, then, that all overlapping triplet codes are impossible” (Brenner 1957, 688).

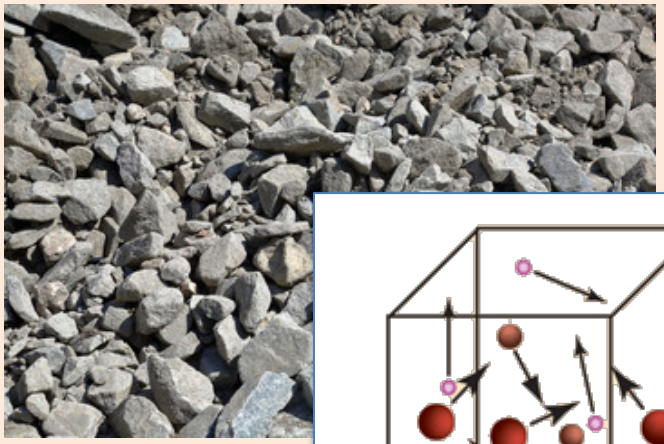


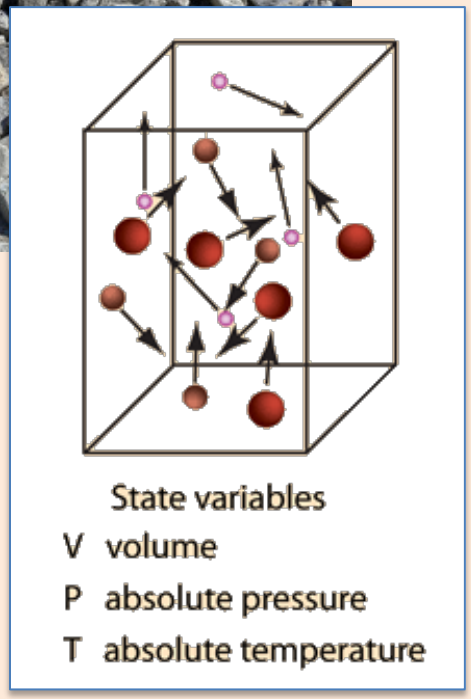
Once the code’s alphabet and codon (triplet) size are set, *and* amino acid sequences are known, *the code’s structure is mathematically constrained.*

(A) In any code with a four-symbol alphabet and a triplet convention, $4^3 = 64$ triplets exist to specify amino acids in protein assembly.

(B) But if those triplets overlap, only 256 (4^4) different dipeptides in protein sequences can occur. *Real sequences contain more dipeptides.*

In our experience, *functional necessity relations* are properties only of designed objects, and organisms – but not of random assemblages (governed by chance) or of strictly physical systems (governed by law):





Here is another example of inferring molecular actors from functional necessity relations (2020 literature):

ANTIOXIDANTS & REDOX SIGNALING
Volume 32, Number 9, 2020

DOI: 10.1089/ars.2019.7816

ARS OPEN

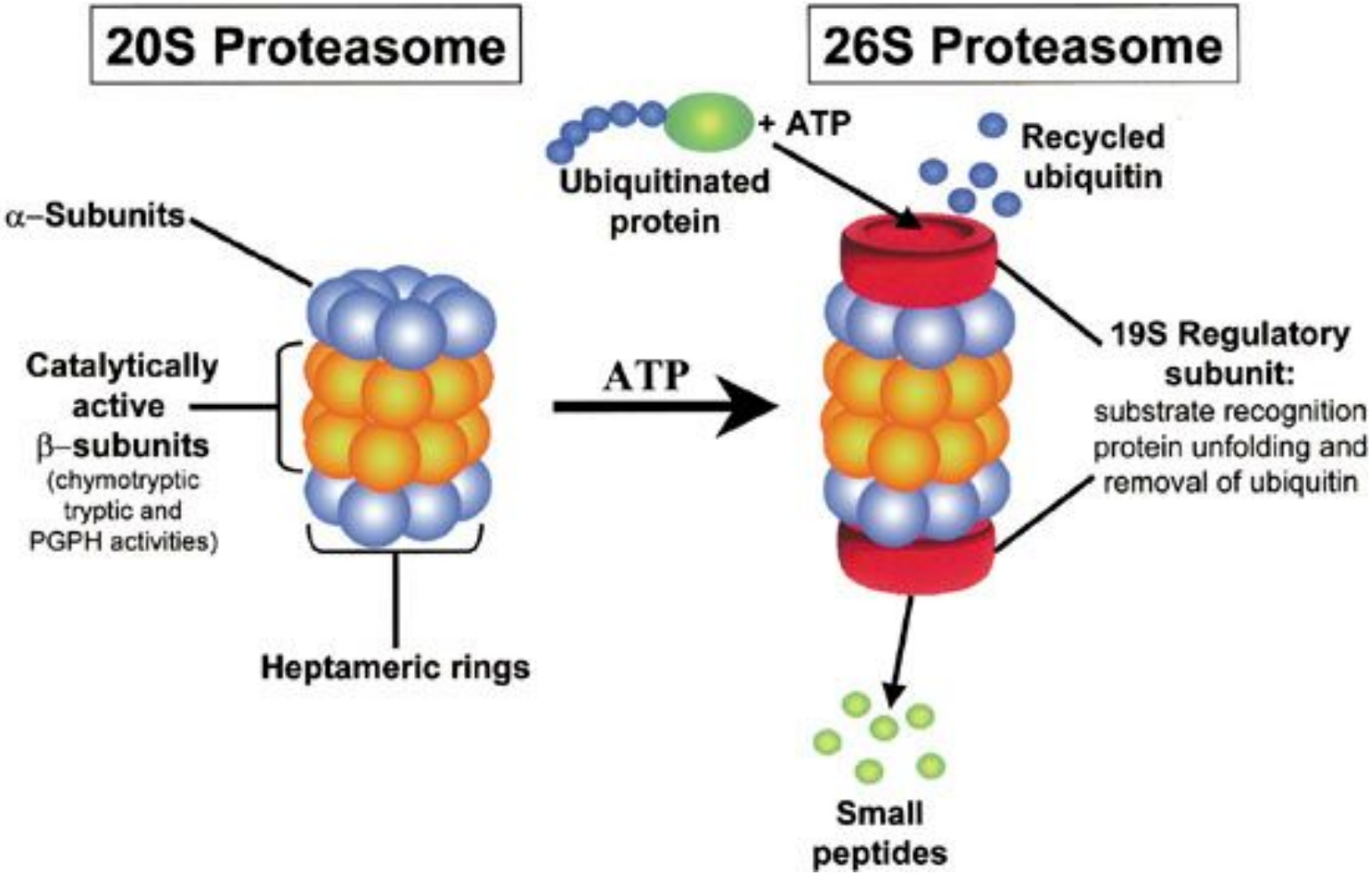
Discoveries

ORIGINAL RESEARCH COMMUNICATION

Regulation of the 20S Proteasome by a Novel Family of Inhibitory Proteins

Maya A. Olshina,[1] Galina Arkind,[1] Fanindra Kumar Deshmukh,[1] Irit Fainer,[1] Mark Taranavsky,[1] Daniel Hayat,[1] Shifra Ben-Dor,[2] Gili Ben-Nissan,[1] and Michal Sharon[1]

The puzzle in question: what stops the proteasome from chewing up every protein in sight?



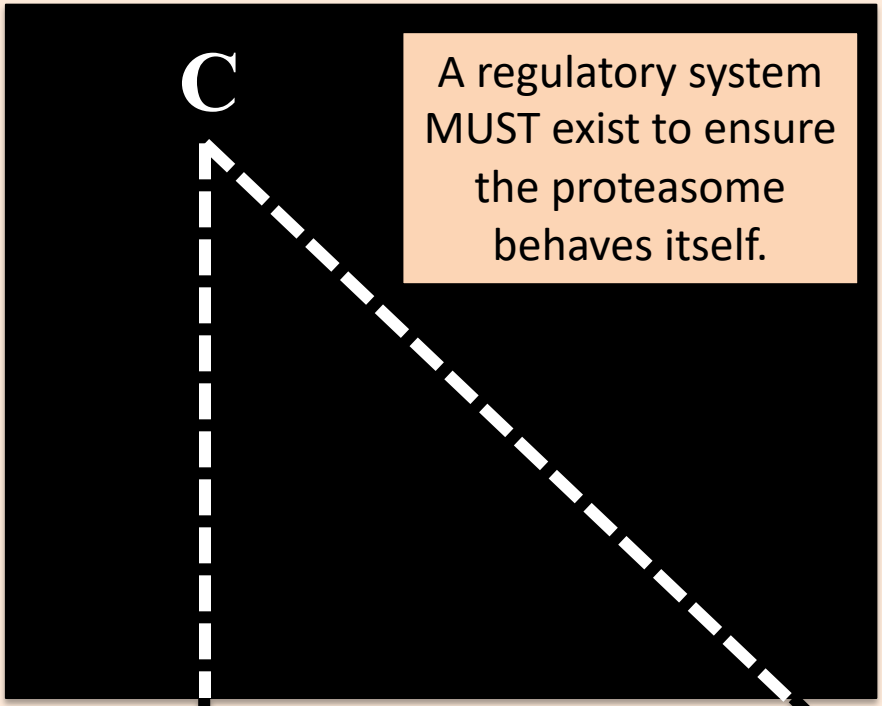
The proteasome is the central machine in the cell's "waste management" system.

"The 20S proteasome degradation machinery is able to cleave any protein with a partially unfolded region, however **uncontrolled degradation of the myriad of potential substrates is improbable**. Thus, ***there must exist a regulatory mechanism*** to control 20S proteasome mediated degradation."

(Olshina *et al*., 2020, emphasis added)

"Thus, there must exist a regulatory mechanism..."

The Weizmann team, like so many other molecular and cell biologists, employed functional triangulation.



Okay…let's go looking for the regulatory system. We have good reason to think that it is actually there.

(A) The 20S proteasome is a very powerful cellular garbage disposal machine.

(B) But "uncontrolled degradation is improbable," because the cell would die.

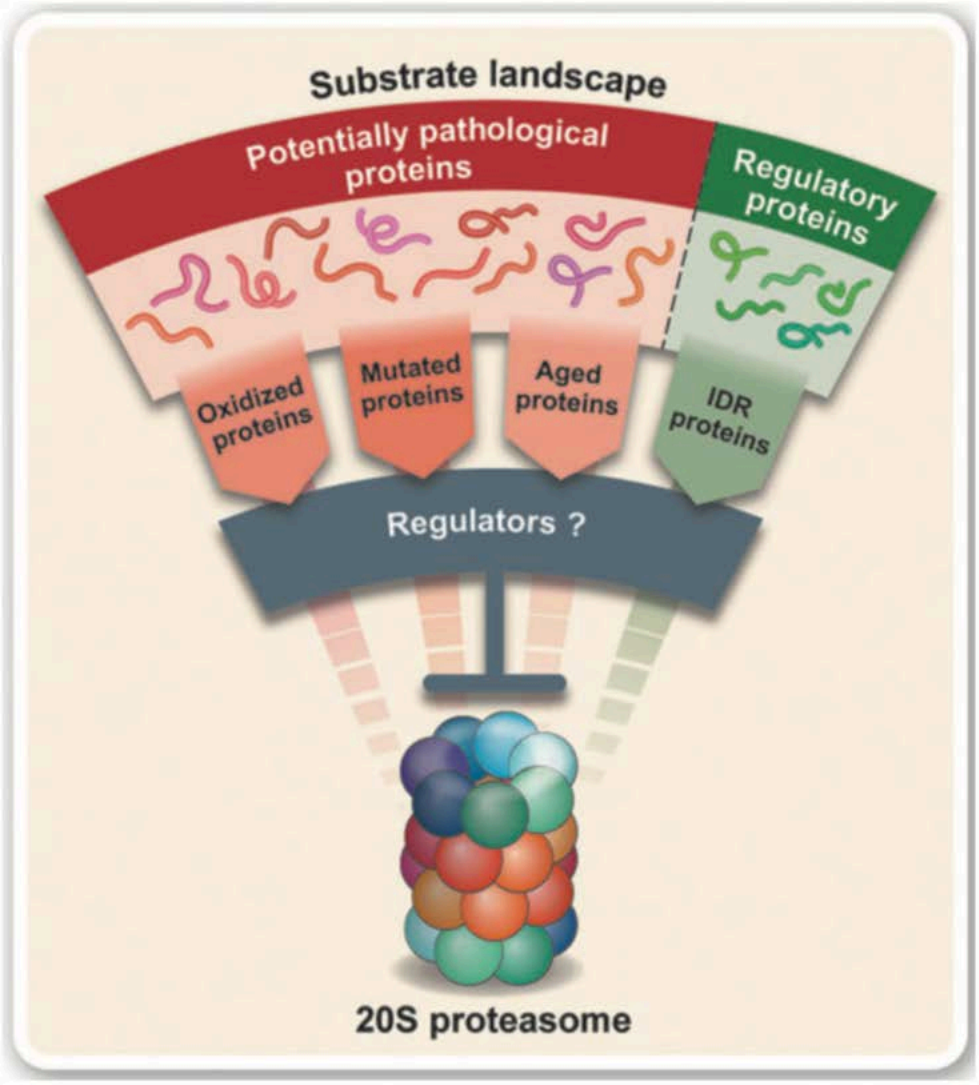


FIG. 1. The flux of substrates into the 20S proteasome must be a regulated process. Two main groups of substrates are susceptible to 20S degradation. The first consists

The Skeptics

"You have a lamentable tendency to give fancy names to what are *ordinary functional inferences*. The Weizmann team simply identified missing parts of a cellular system and went looking for them. This happens routinely in all areas of biology. The practice does not, however, merit a grand term like 'triangulation,' and supports no particular view of origins. Climb down off your ID soapbox already."

Paul: in one sense, Alpha C, I agree. Living things are what they are, irrespective of what we may think about the theories of origins we favor or disfavor. Doubtless most biologists who successfully employ functional triangulation would *not* also embrace design.

On the other hand, as the next two sections show, triangulation makes demands on the investigator which are *much* easier to satisfy *if design is true*.

"Hm – we'll see about that. BTW, that 'design versus naturalism' stock illustration over there is hackneyed."

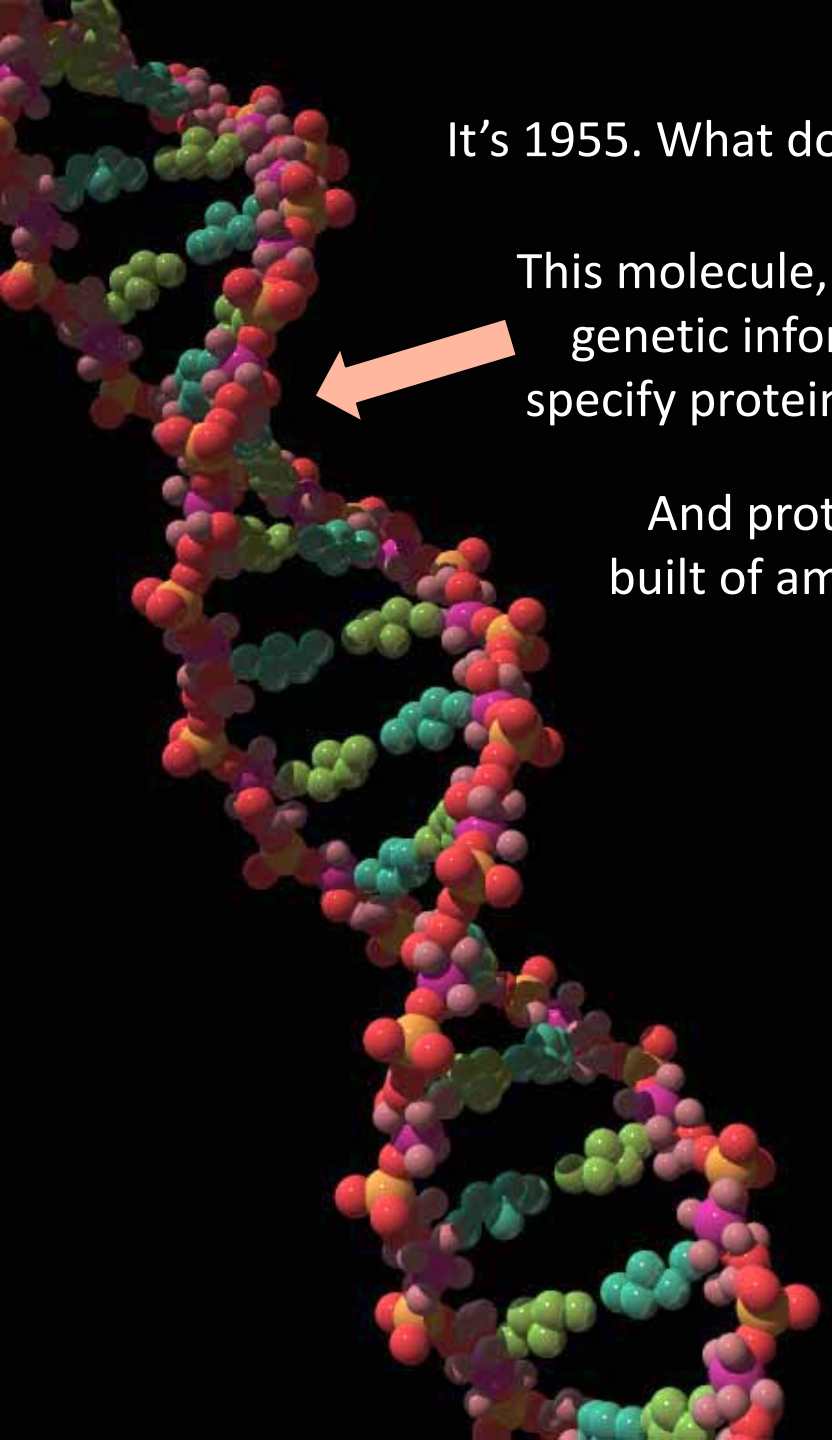
Paul: That's rich, coming from a figurative entity depicted by screen-captures from a 2001 sci-fi movie.

V. Crick, Watson, and the Adaptor Hypothesis

The next example
is the most telling – and
also carries the answer to
the question "***Why*** does
functional triangulation ***work***
in biology...and why should
we call it ***design*** triangulation?"

The episode described next is either the luckiest guess *ever* in the history of science – **OR** a remarkably prescient inference based on (1) an intimate knowledge of *some* aspects of the system at hand (i.e., information transfer in cells, as understood *circa* 1955) and (2) sheer audacity, grounded nonetheless in the rational "no-magic-because-life-works-by-mechanisms" principle of design triangulation.

My vote is for the latter, comprising (1) and (2). Good science is not lucky guesswork.

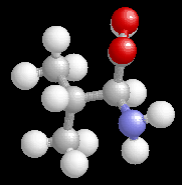


It's 1955. What do biologists know?

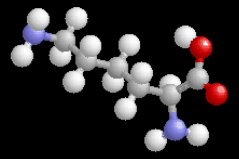
This molecule, DNA, carries genetic information to specify protein sequences

And proteins are built of amino acids.

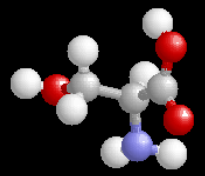
But what is mediating information transfer between these two very different chemistries?



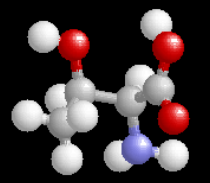
Valine



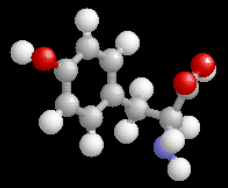
Lysine



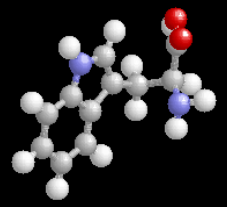
Serine



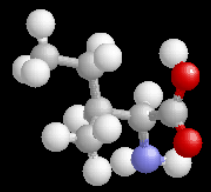
Threonine



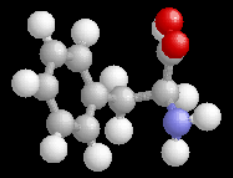
Tyrosine



Tryptophan



Isoleucine



Phenylalanine

(photo credit: Special Collections, Oregon State University)

F.H.C. Crick

The "RNA Tie Club," started in 1954 by physicist George Gamow, was an "invisible college" of less than two dozen members, among them Watson and Crick, dedicated to solving the open problems of nascent molecular biology: chiefly, **coding.**

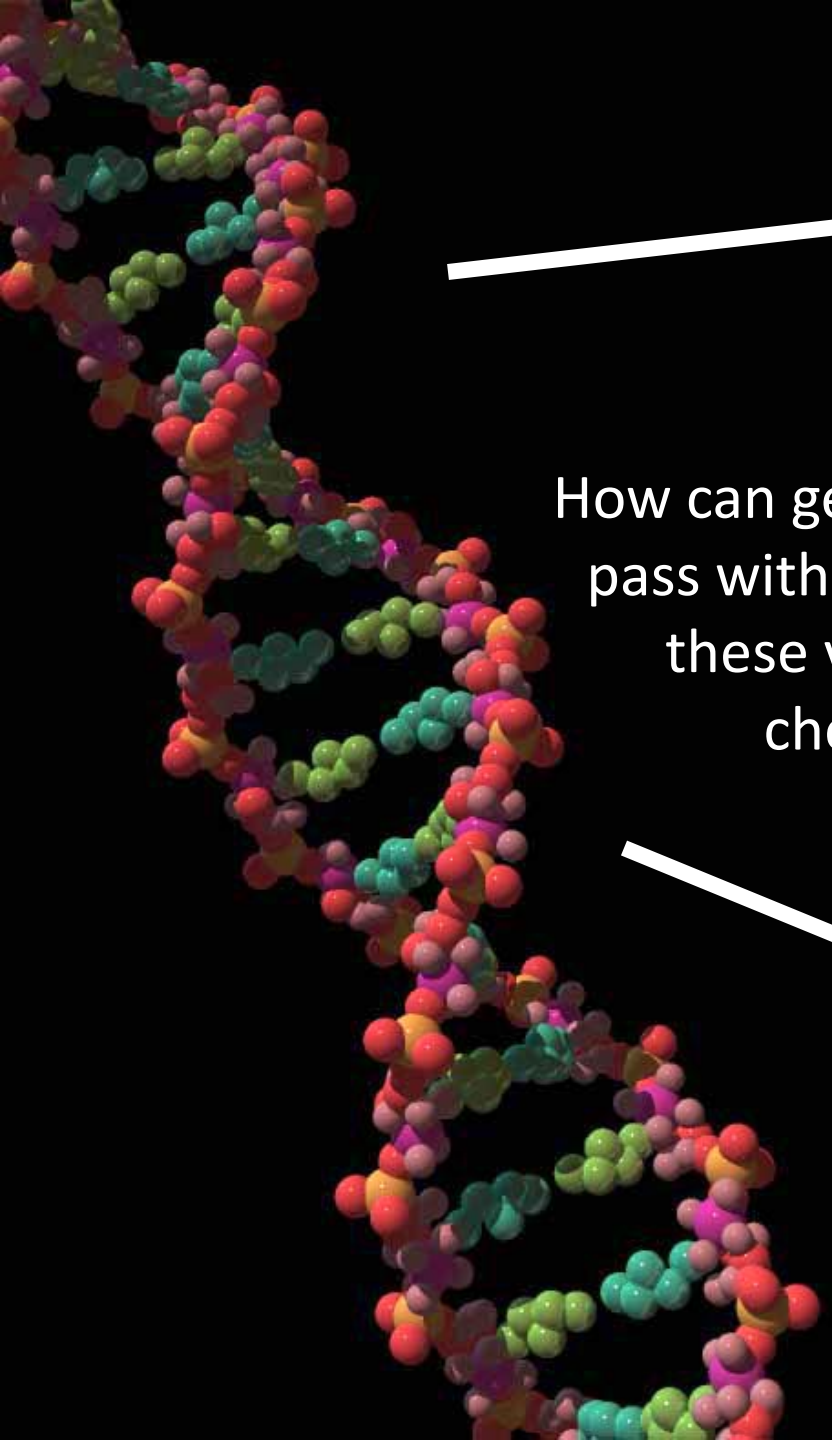
ON DEGENERATE TEMPLATES AND THE ADAPTOR HYPOTHESIS

F.H.C. Crick,

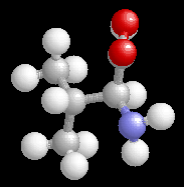
Medical Research Council Unit for the Study of
the Molecular Structure of Biological Systems,

Cavendish Laboratory, Cambridge, England.

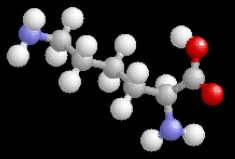
Crick wrote this unpublished manuscript on the genetic coding puzzle for the members of the club. In it, he proposed the "adaptor hypothesis." (Crick's original manuscript is available for downloading as a pdf here: https://profiles.nlm.nih.gov/spotlight/sc/catalog/nlm:nlmuid-101584582X73-doc)



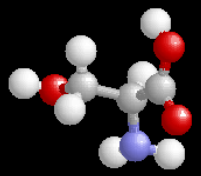
How can genetic information pass with fidelity between these very different chemistries?



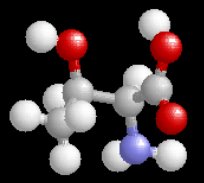
Valine



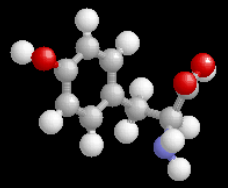
Lysine



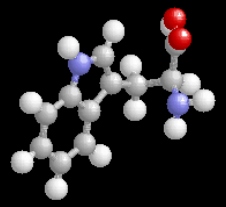
Serine



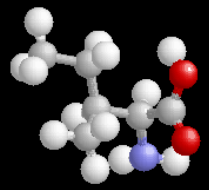
Threonine



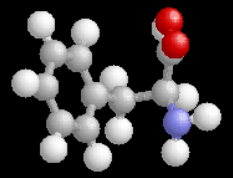
Tryptophan



Tyrosine



Phenylalanine



Isoleucine

Crick (1955) grasps the difficulty:

"Now what I find profoundly disturbing is that I cannot conceive of *any* structure (for either nucleic acid) acting as a direct template for amino acids, or at least as a specific template....I don't think anybody looking at DNA or RNA would think of them as templates for amino acids."

Crick (1955) grasps the difficulty:

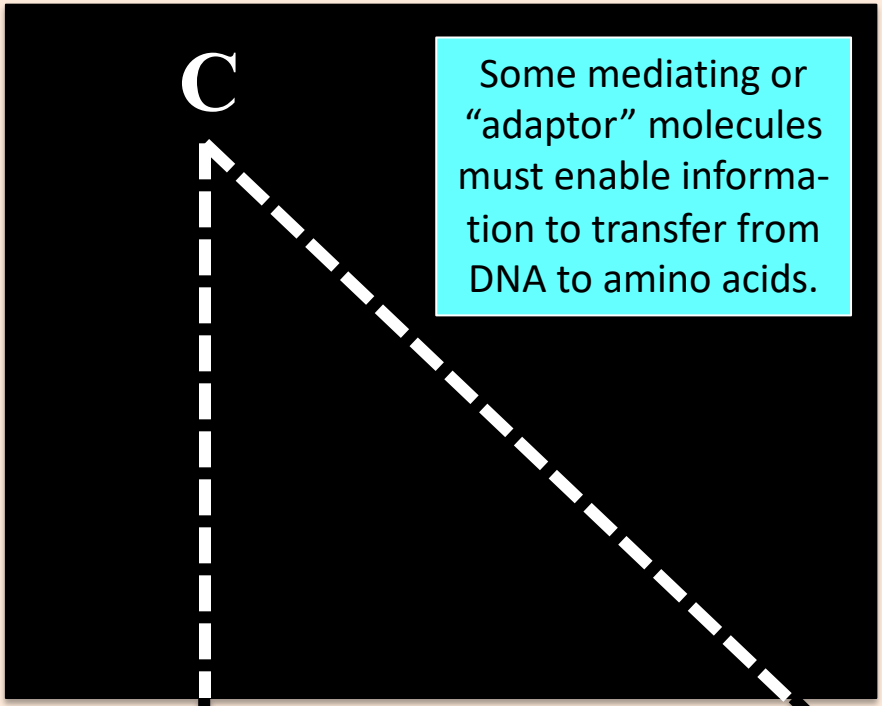
"Where are the knobly hydrophobic surfaces to distinguish valine from leucine and isoleucine? Where are the charged groups, in specific positions, to go with the acidic and basic amino acids?...What the DNA structure *does* show (and probably RNA will do the same) is a specific pattern of *hydrogen bonds*, and very little else."

Functional triangulation, however, underwrites inferences to unobserved entities which ***must*** exist:

When one discovers a complex system performing specialized functions, *assume that a rational logic, and well-matched parts, are enabling the functions.*

No magic. Look for the mechanism: it's there.

Crick triangulates from the cell's information-transfer requirements to its unobserved, but functionally necessary, parts:



Okay, now we can go looking for the system. Here are some of the features it will need to specify each of 20 amino acids in protein assembly...

(A) DNA carries information, but is chemically non-specific ("hydrogen bonds, little else")

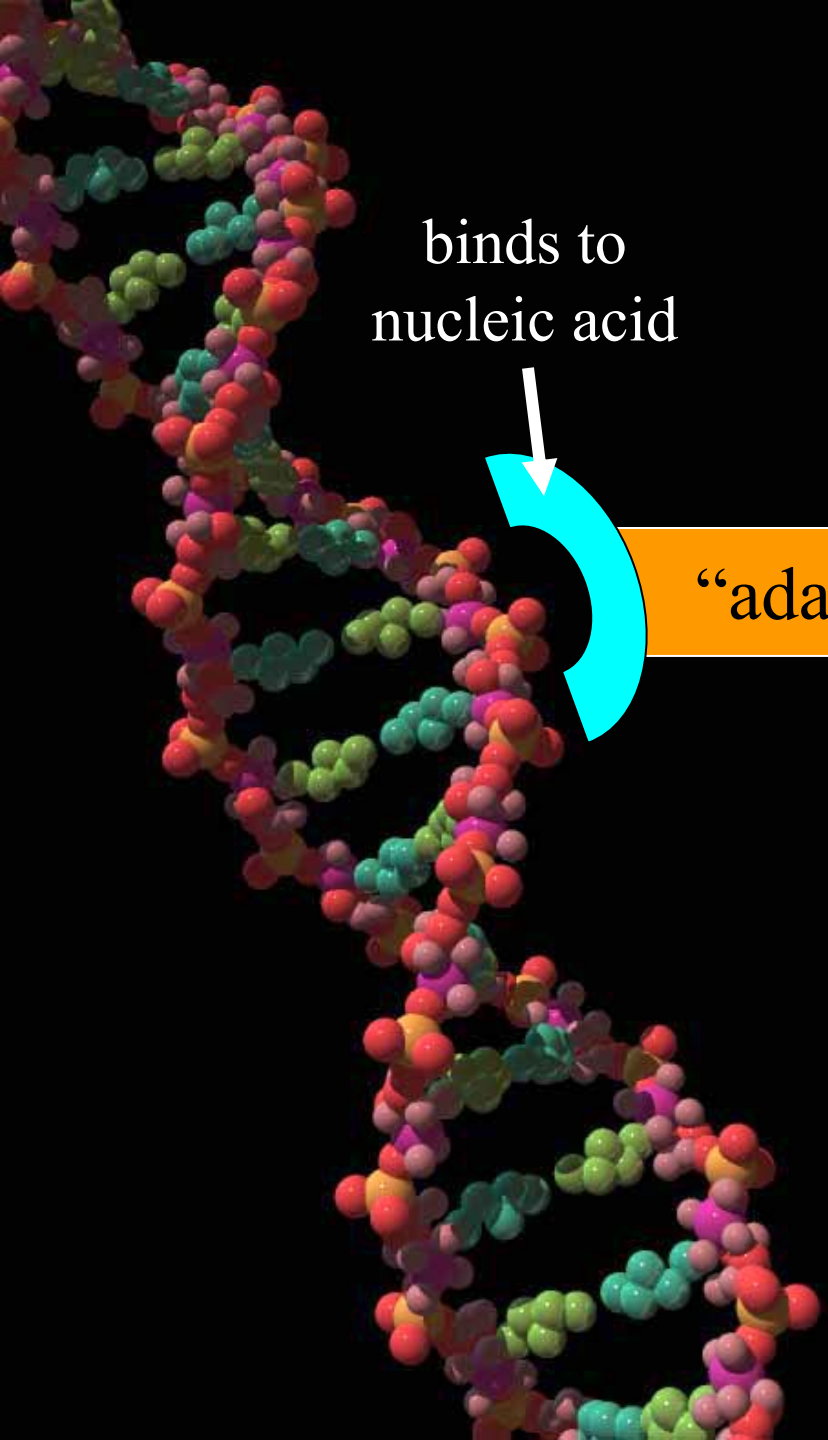
(B) Yet amino acids have specific geometries, which require a system to recognize them.

Crick boldly triangulates:

"...each amino acid would combine chemically, at a special enzyme, with a small molecule which, having a specific hydrogen-bonding surface, would combine specifically with the nucleic acid template.... In its simplest form there would be 20 different kinds of adaptor molecule...."

Crick boldly triangulates:

"...one for each amino acid, and 20 different enzymes to join the amino acid to their adaptors. Sydney Brenner, with whom I have discussed this idea, calls this 'the adaptor hypothesis' since each amino acid is fitted with an adaptor to go on to the template."



binds to nucleic acid

works with a special dedicated enzyme (protein)

"adaptor"

Valine

binds to a *specific* amino acid

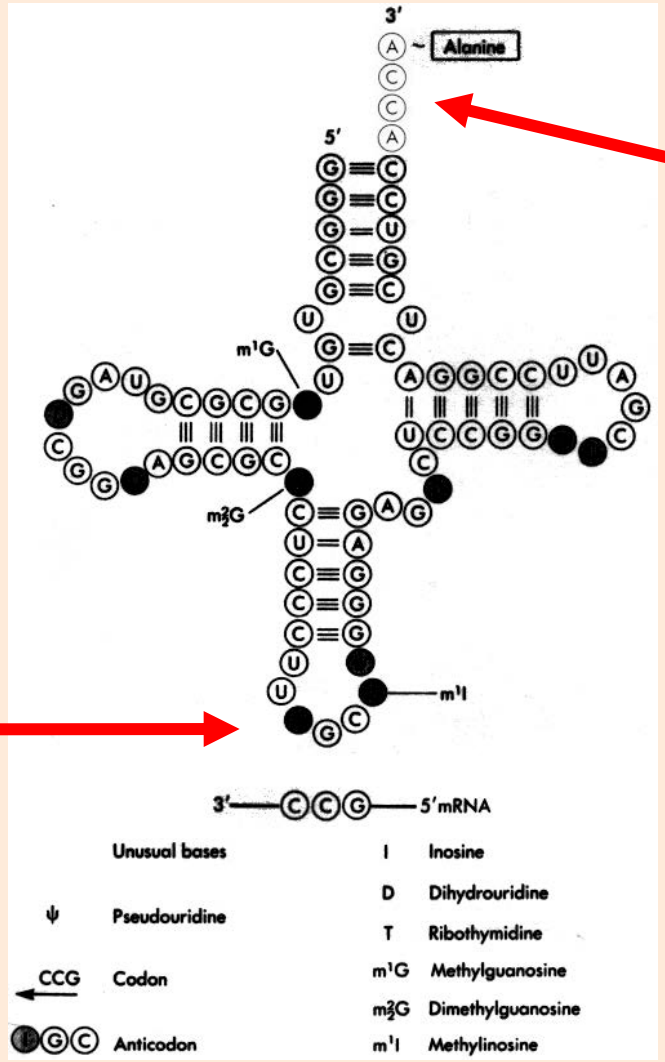
And, by the way – we need 20 of these molecules, with 20 specially dedicated enzymes, one for each amino acid.

Crick admits there is no evidence yet, but some such mediating molecules ***must*** exist.

"The usual argument presented against this latter scheme is that no such small molecules have been found, but this objection cannot stand."

The inference from *systems-level functional necessity* is very strong.

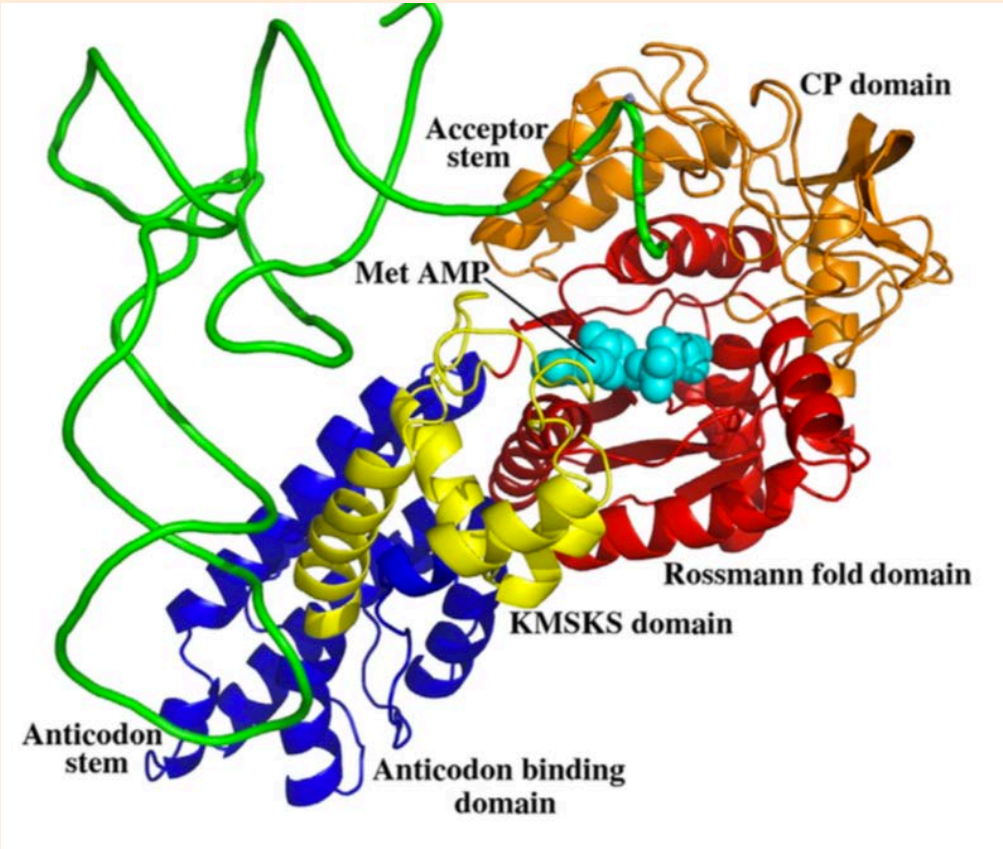
What came to be known as **transfer RNA** – Crick's "adaptor" – was discovered 3 years after Crick's 1955 prediction of its existence. When characterized, it possessed the features Crick said the molecule would need (e.g., a hydrogen-bonding surface [the anticodon]).



CCA stem (the amino acid attachment site)

anticodon

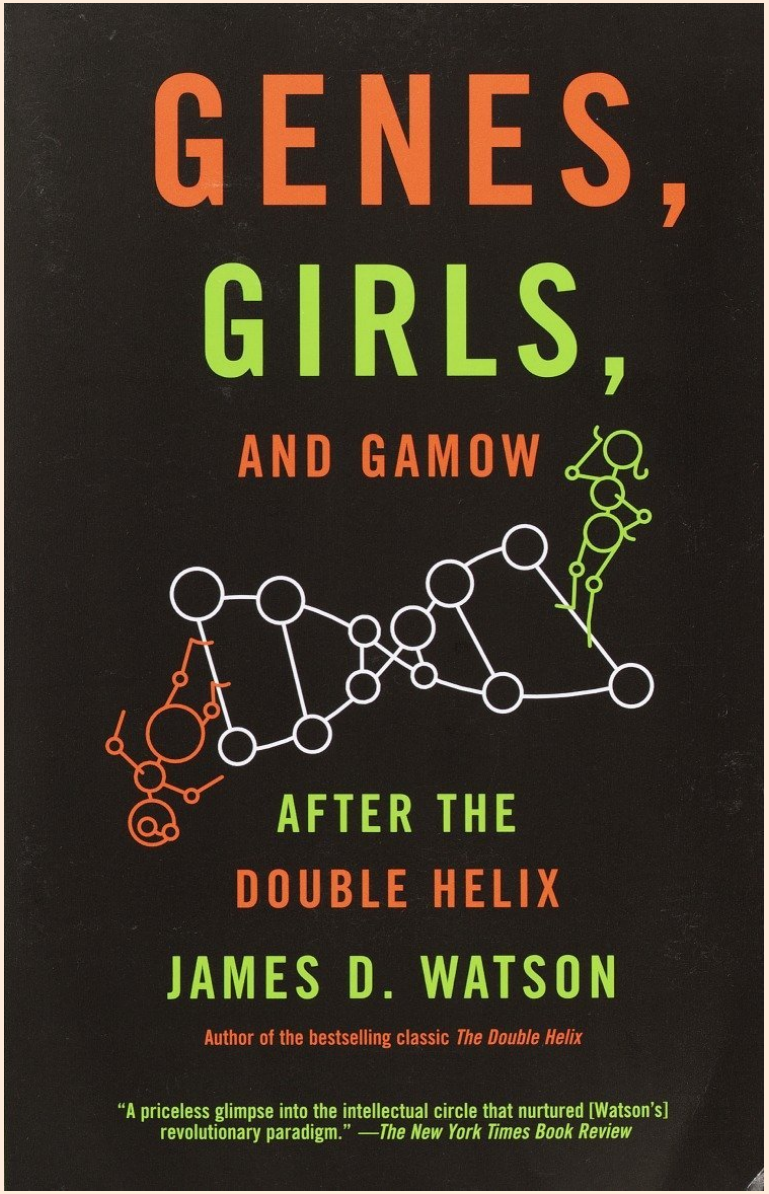
Crick also predicted 20 dedicated enzymes, to attach specific amino acids to their ”adaptors” – again, without having any direct evidence that such enzymes existed: **aminoacyl-tRNA synthetases.**



Docked structure of *E. coli* MetRS–tRNAfMet complex along with Met-AMP.
(figure from Rajendren *et al*. 2018, 402)

Now – on to the brutally honest antagonist of the story.

VI. Why functional triangulation is design triangulation: the causal primacy of the organism



In his 2002 autobiography, Watson describes the period in the mid-1950s, when he, Crick, and others in the "RNA Tie Club" were working out the functional implications of DNA for biological information transfer.

The pre-eminent puzzle was the nature of the genetic code.

But the code needed "hardware" – namely, molecular actors (whether proteins or nucleic acids or both) to carry out the role of actually *transferring* information from DNA to amino acids in protein assembly.

F.H.C. Crick

As a founding member of the RNA Tie Club, and one of Crick's closest collaborators, Watson was a recipient of this unpublished manuscript.

ON DEGENERATE TEMPLATES AND THE ADAPTOR HYPOTHESIS

F.H.C. Crick,

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,

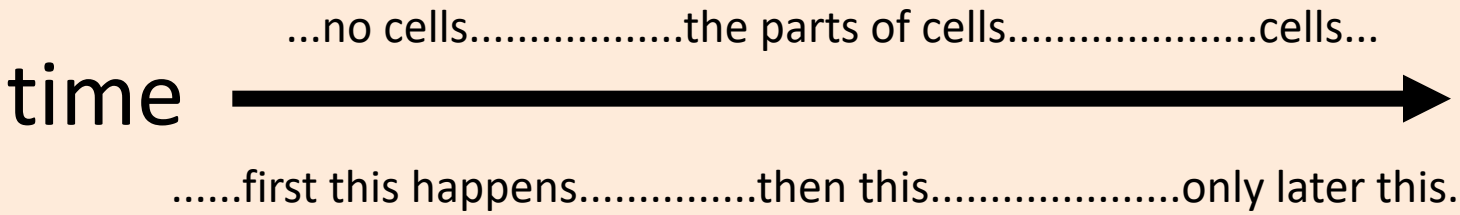
Cavendish Laboratory, Cambridge, England.

But Watson didn't buy the adaptor hypothesis. Why not?

Watson (2002, 139) explains why he didn't like the adaptor hypothesis:

"I did not like the idea at all... More to the point, the adaptor mechanism seemed to me too complicated to have ever evolved at the origin of life."

Watson's biological intuition was bound to an implicit time axis.

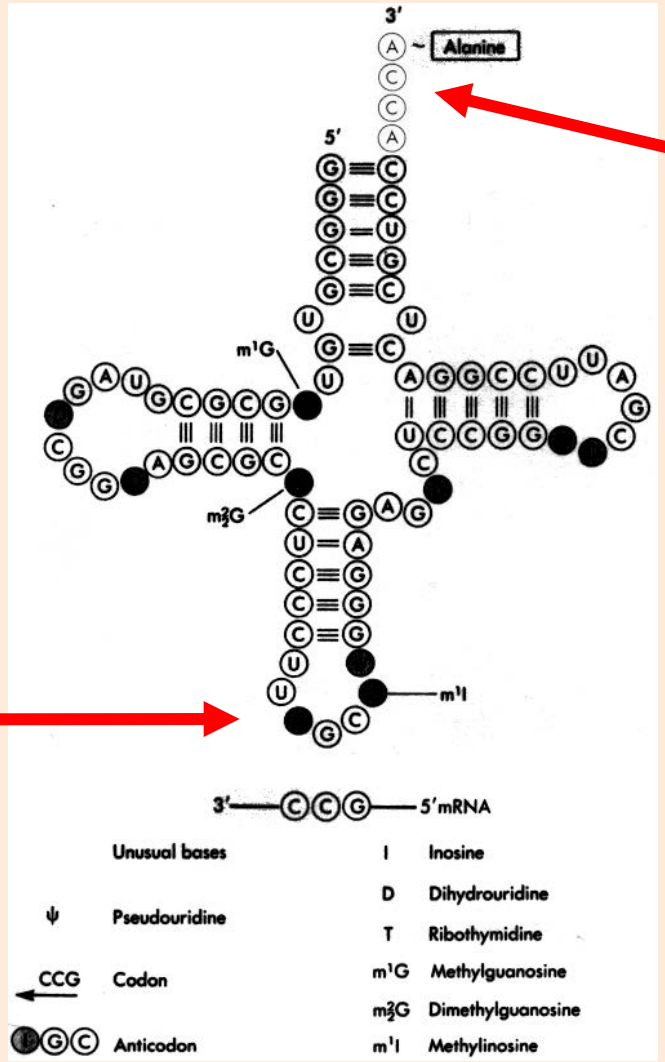


Within the naturalistic picture of life,
complex biological systems cannot arise all at once.
Organisms are fundamentally historical entities,
and “history is just one damn thing after another.”

The same is true, of course, for any naturalistic theory of abiogenesis (origin of life).

Biological complexity can only accrete over time.

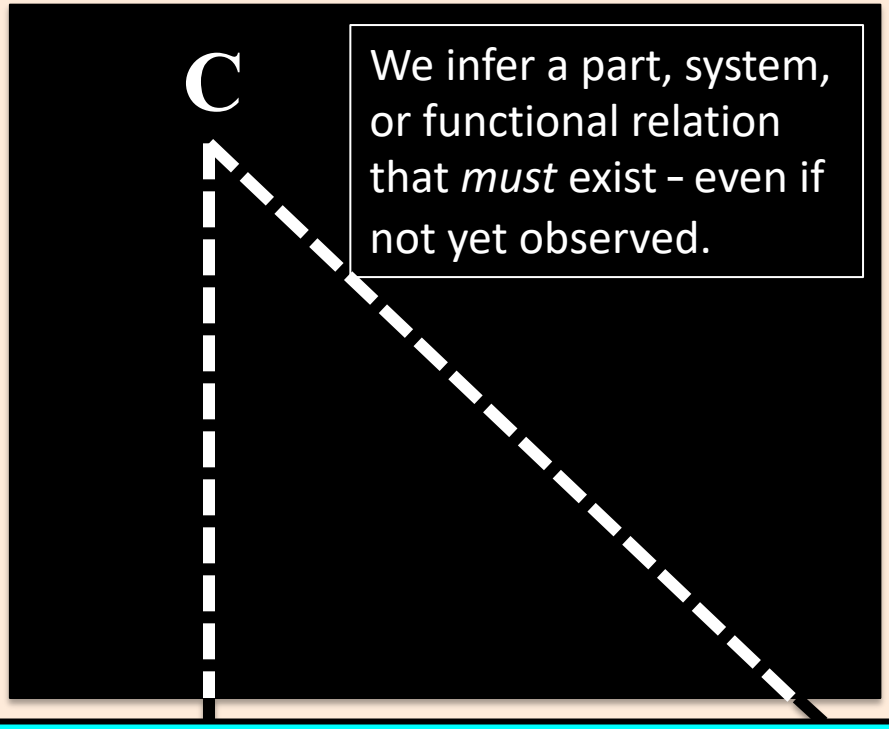
But transfer RNA, or something very much like it, ***had to be there*** – once the facts about DNA, amino acids, and protein assembly were in place.



CCA stem (the amino acid attachment site)

anticodon

But the time axis, and the complexity-can-only-accrete assumption, aren't there *in the biology itself*.



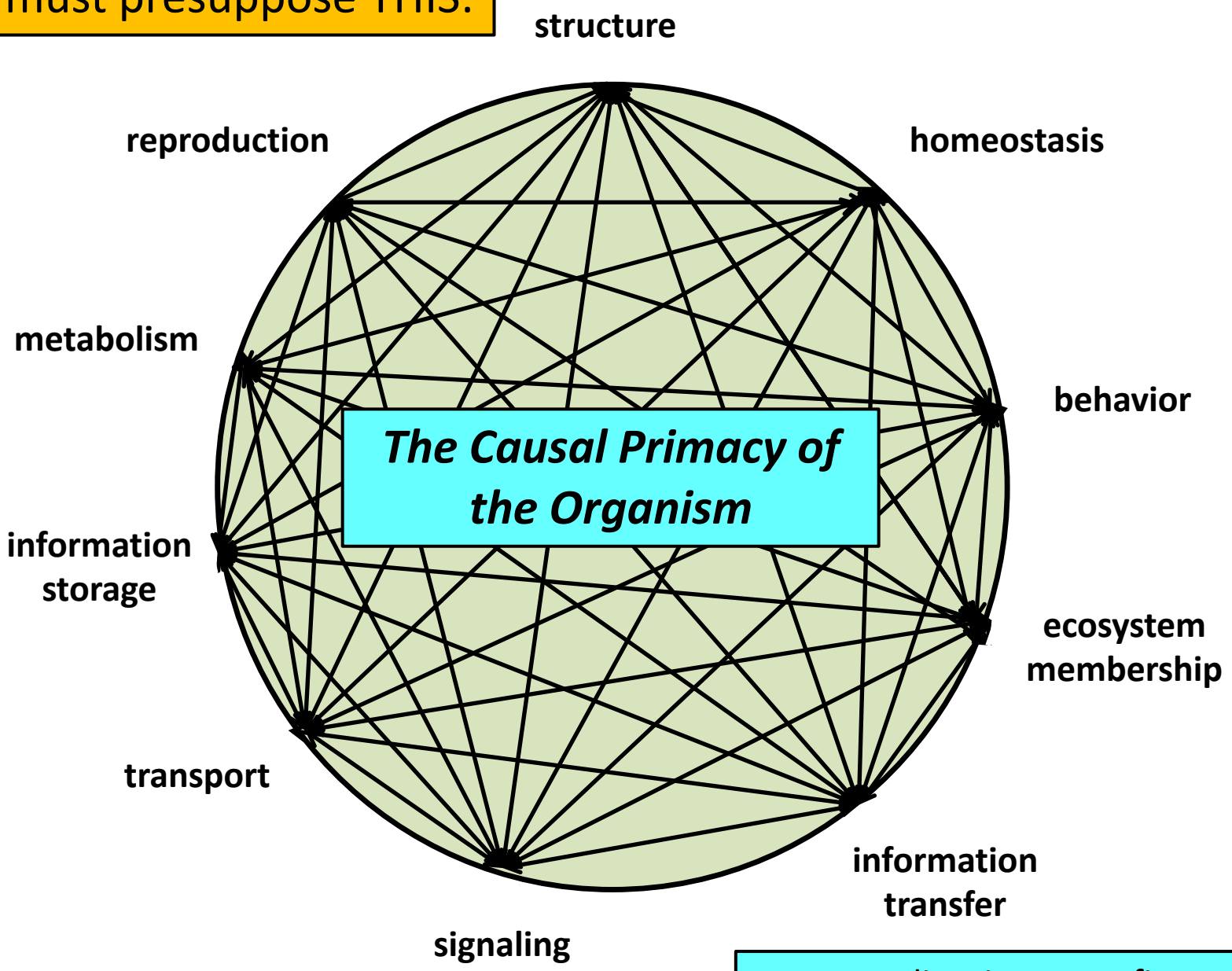
Organisms and their features as *black boxes*

"Simplest at the start and only later more complicated," and "one must begin with the parts of cells," flow from a *naturalistic metaphysics of explanation*: from philosophy. **Not from organisms themselves. Not from the evidence.**

For biological inquiry – for actually finding things out – triangulation is unmistakably successful. **However, the method does make a demand on the investigator.**

One must presuppose the prior existence of the ***system as a whole***, to obtain the functional necessity relations ***that warrant any triangulation to its unobserved parts***.

One must presuppose THIS:



But naturalism is a parts-first story...

The causal primacy of the organism calls for (at least) two notions foreign to bottom-up, physics-first explanation:

- **Foresight**
- **Causal circularity**

Thus far, we have been looking at the role of these concepts mainly with respect to the origin of life.

But, as noted in slide 76, if living things broadly speaking are system-(*not* parts)-first entities, foresight and causal circularity should be observed at all levels.

So let's jump from transfer RNA to developmental biology, and the puzzle of the origin of animal body plans.

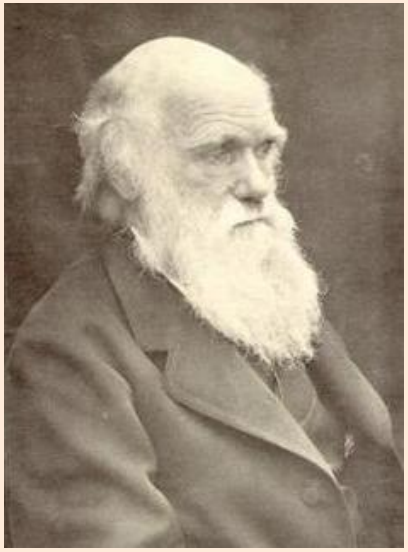
The origin of animal body plans

Caenorhabditis

Strongylocentrotus purpuratus

Drosophila melanogaster

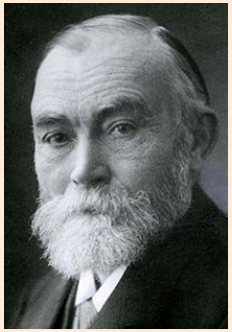
As is very often the case in evolutionary theory, Darwin sets the puzzle to be solved:



1. *Pattern:* The common descent of the Metazoa (the animals) from a unicellular eukaryotic ancestor.
2. *Process:* Natural selection was the main cause of biological novelty, including the origin of animal body plans.

"May we remind you that, in slide 5, you said you *weren't* going to be arguing with evolutionary theory? By our estimation, thus far you've done a good deal of disputing with evolution, *and* naturalism, *and* whatnot."

Paul: So sue me. But seriously – I'm trying to build a case for how design might work in biological explanation. That means a certain amount of tangling with *existing* modes of explanation. Sometimes the logic of argumentation takes you places you might prefer not to go. For instance: Frege and psychology.



Gottlob Frege
1848-1925

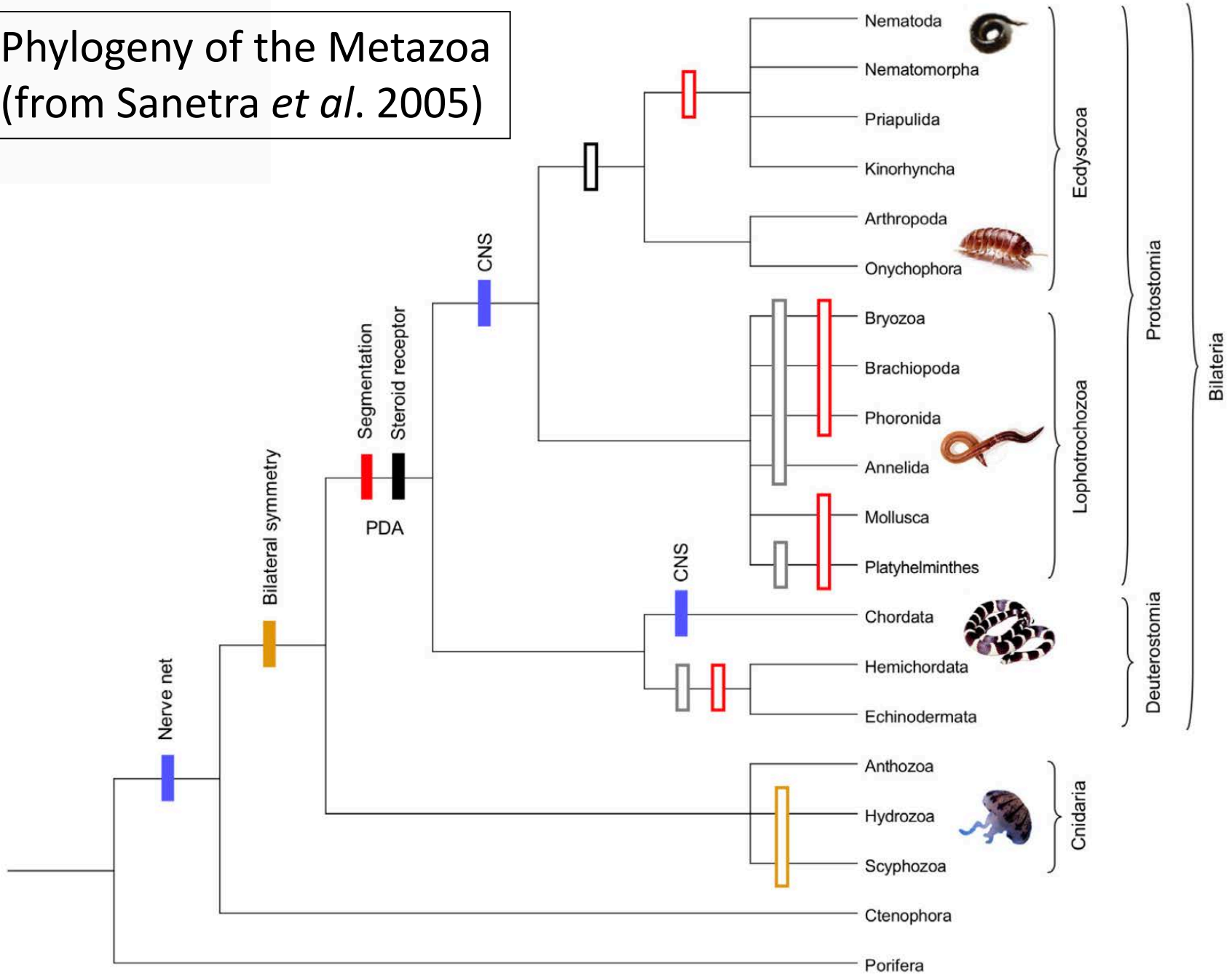
Frege wanted to construct arithmetic on solid logical foundations, but found his contemporaries appealing to *psychology* instead. "I found myself forced to enter a little into psychology," he wrote, somewhat apologetically (1884), "if only to repel its invasion of mathematics." Dialectic takes us where it will.

- ***Construct*** **developmental pathways (*ab initio*).**

- ***Modify or change*** **developmental pathways (once they exist).**

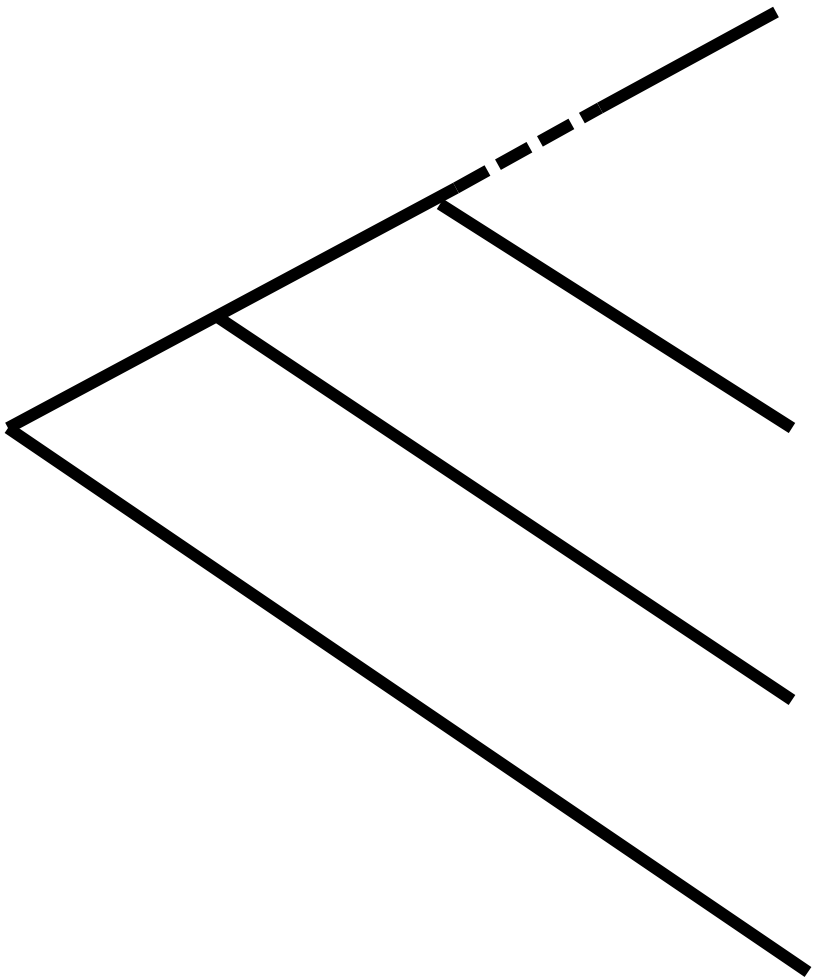
The adult anatomy of *Caenorhabditis elegans*

Phylogeny of the Metazoa (from Sanetra *et al*. 2005)



Choanoflagellates

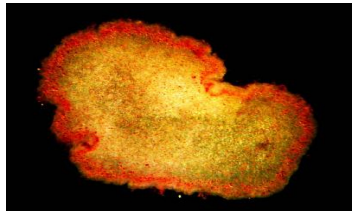
(photomicrograph: Eckhard Voelcker)



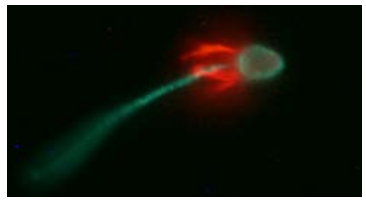
Nematoda (*C. elegans*)

Porifera (*C. prolifera*)

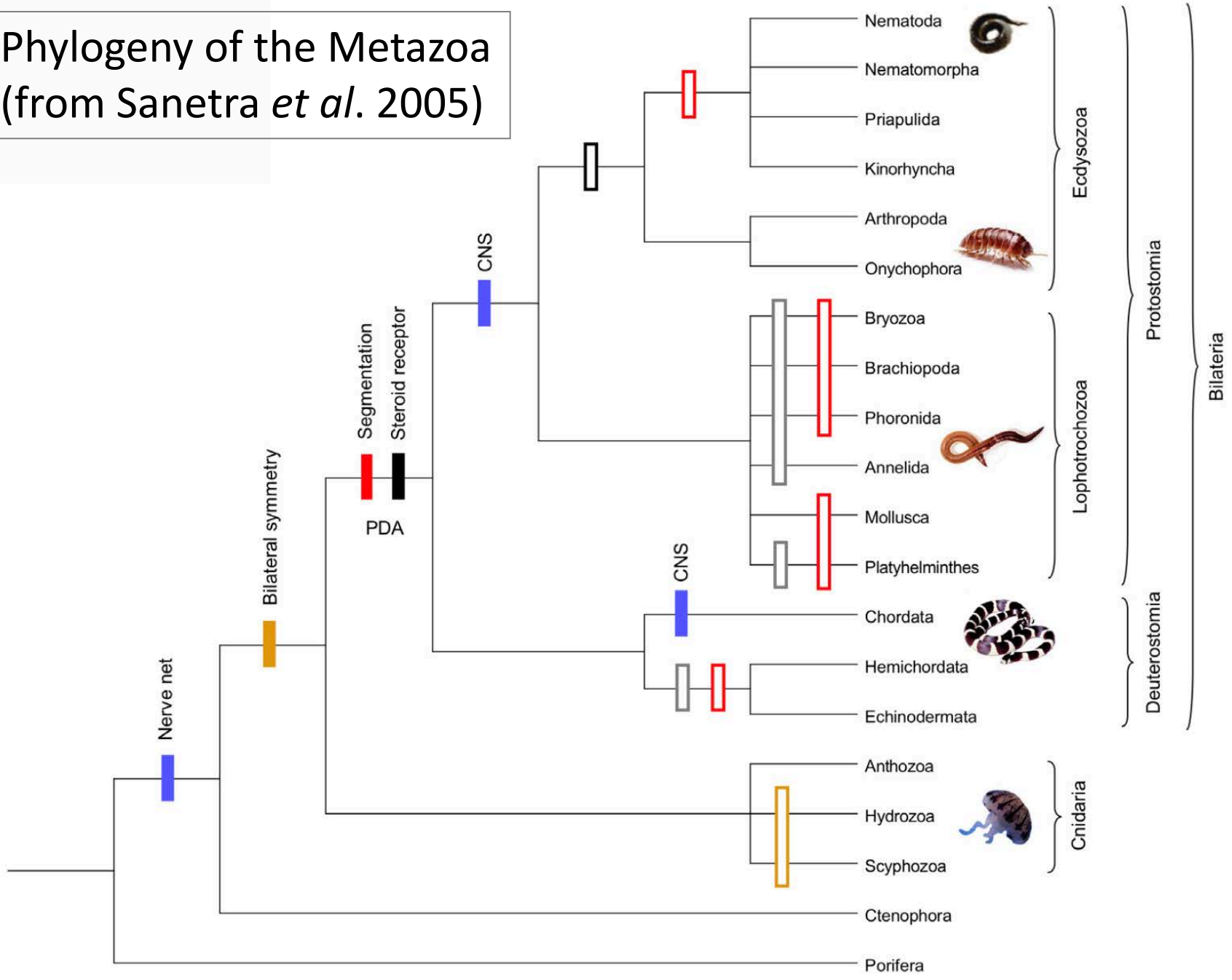
Placozoa (*Trichoplax*)



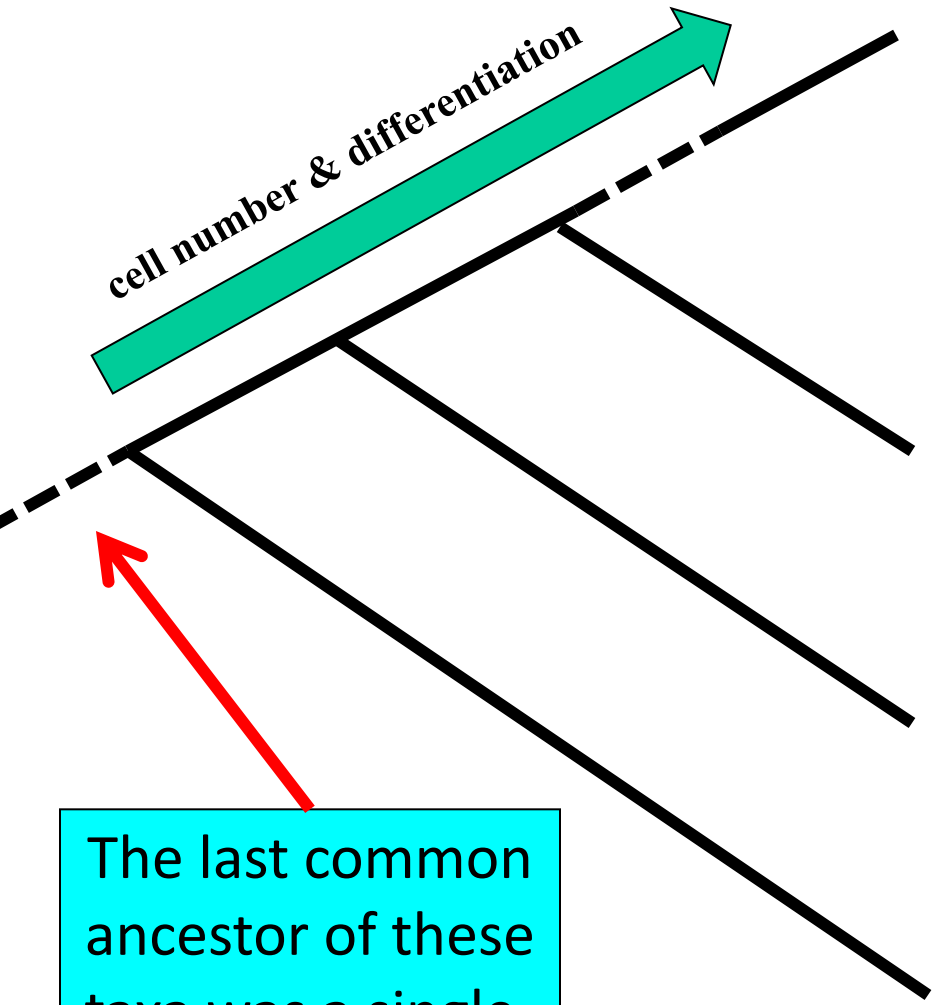
choanoflagellates



Phylogeny of the Metazoa (from Sanetra *et al*. 2005)



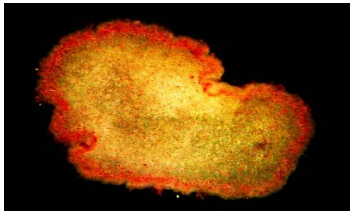
Along the branch leading to *C. elegans*, cell number and cell differentiation must increase (i.e., net gain).



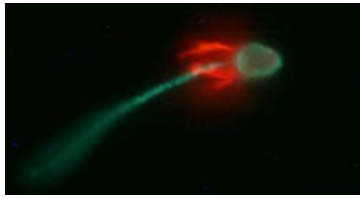
Nematoda (*C. elegans*)

~1,000 cells in adult form

Porifera (*C. prolifera*)



Placozoa (*Trichoplax*)

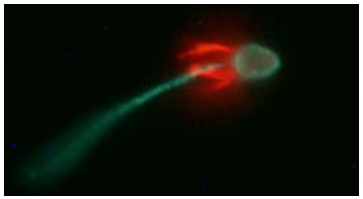


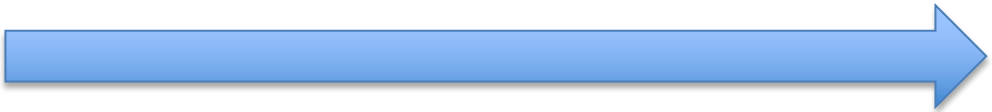
choanoflagellates

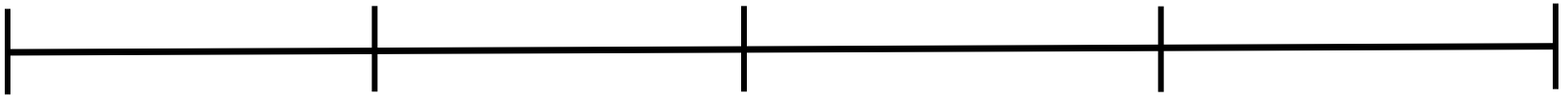
The last common ancestor of these taxa was a single-celled eukaryote.

Thus, given the hypothesis of the common ancestry of the animals, the specific developmental pathway we observe today in *C. elegans* must have been built incrementally, over evolutionary time, starting from a single-celled eukaryotic ancestor.

Increase in total adult cell number over the phylogenetic history of *Caenorhabditis* (intervals are arbitrary, indicating distance only)







Single-celled eukaryote

250 cells

500 cells

750 cells

Adult *C. elegans*

The explanatory puzzle therefore entails ***two*** causally related dimensions: (1) the developmental pathway we observe today, **and** (2) the historical or evolutionary lineage that *ex hypothesi* constructed that pathway, where it did not exist before (*ab initio*).

"Can Modern Evolutionary Theory Explain Macroevolution?"

Douglas Futuyma
Ecology & Evolution
SUNY Stony Brook

E. Serrelli and N. Gontier (eds.), *Macroevolution*, Interdisciplinary Evolution Research 2 (Springer, 2015), p. 76.

"Existing theory can provide a plausible account of the history and causes of most or all evolutionary phenomena... I do not know of any macro-evolutionary phenomena that are inconsistent with existing evolutionary theory, any phenomena that would require us to reject one of its principles as simply false."

Deborah Charlesworth

The causal supremacy of natural selection remains the dominant view within evolutionary theory today:

Nicholas Barton

"We have focused our discussion on the sources of the variability used in adaptive evolution...we finish by re-emphasizing the central concept of neo-Darwinism and the MS [Modern Synthesis]: **allele frequency change caused by natural selection is the only credible process underlying the evolution of adaptive organismal traits."**

Brian Charlesworth

Charlesworth, Barton, & Charlesworth (2017, 9-10; emphasis added)

If, within a species or population, the individuals

a. vary in some trait **q** – the condition of **variation**;

b. leave different numbers of offspring in consistent relation to the presence or absence of trait **q** – the condition of **selection**;

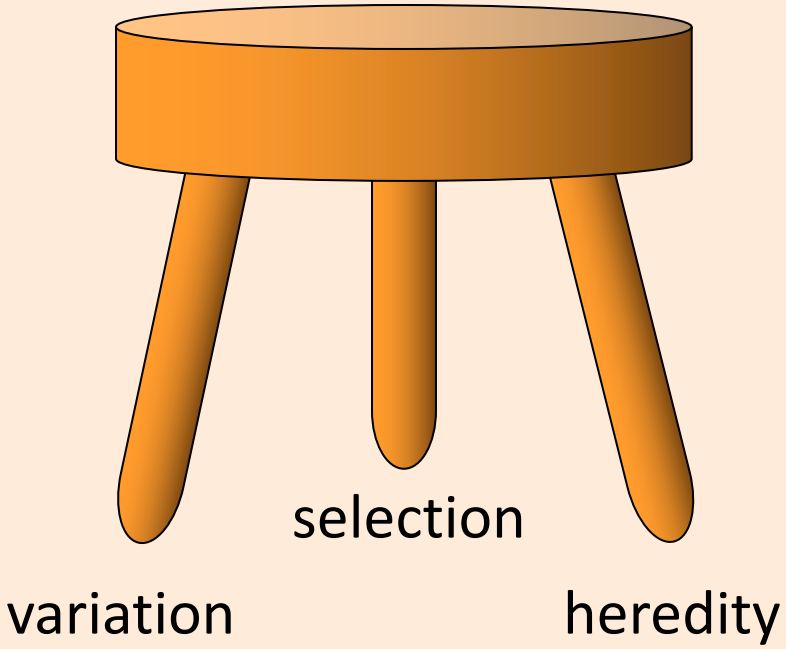
c. transmit trait **q** faithfully between parents and offspring – the condition of **heredity**;

then the frequency of trait **q** will differ predictably between the population of all parents and the population of all offspring.

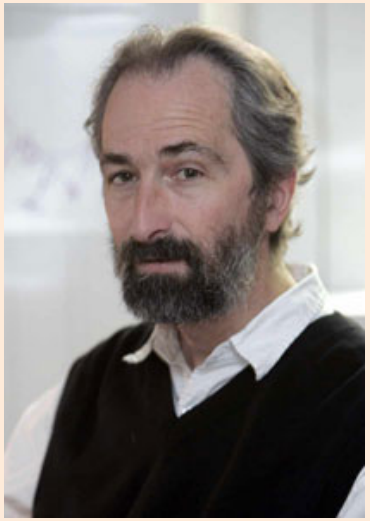
(Lewontin 1978; Endler 1986)

The requirements of natural selection:

variation
selection
heredity



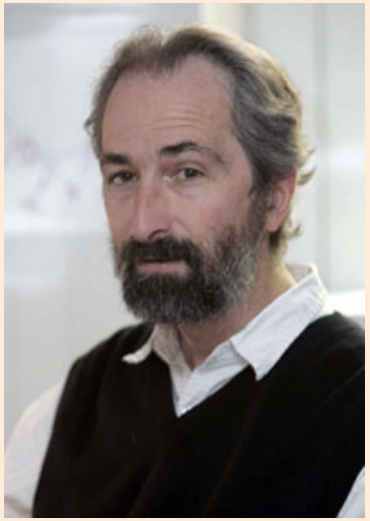
Now, maybe you don’t think much of natural selection – it’s overrated as an evolutionary process, you say.



Michael Lynch
Biodesign Institute
Arizona State Univ.

"**One of the most significant problems** in the broader body of biological thinking **is the common assumption that all observed aspects of biodiversity are products of natural selection**. With this mind set, **evolutionary biology becomes little more than a** (sometimes endless) **exercise in dreaming up the supposed agents of selection** molding one's favorite aspect of phenotypic diversity. However, we now know that **this unwavering belief in the limitless power of natural selection is untenable**."

(Lynch & Trickovic, 2020, emphasis added)



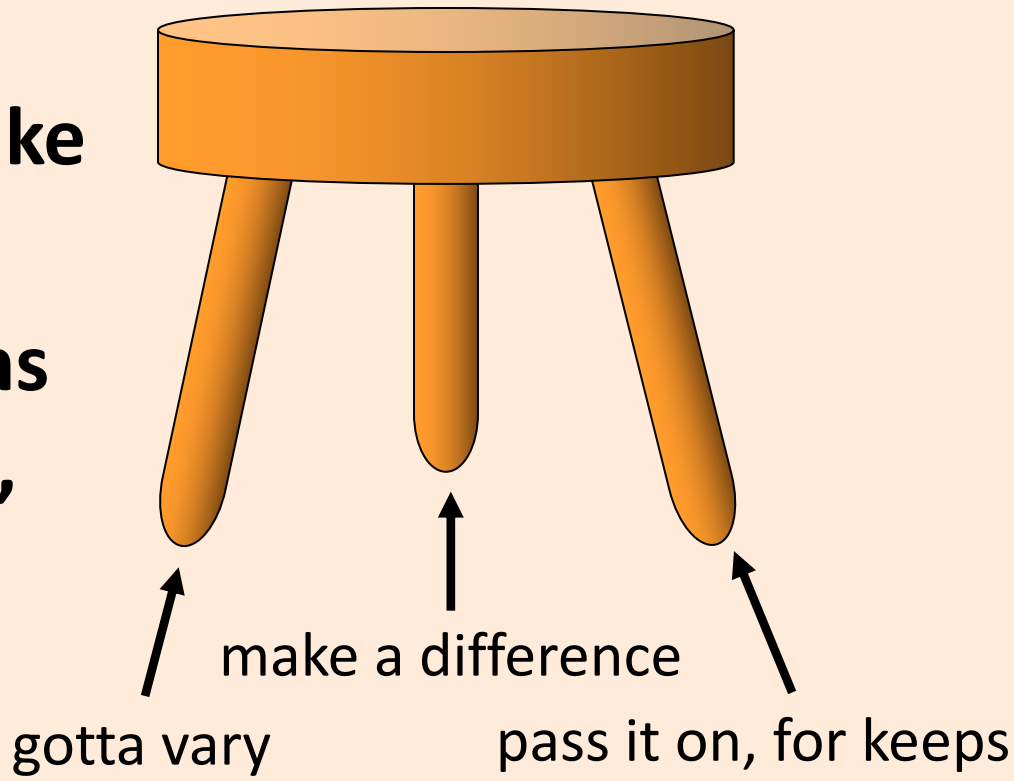
Michael Lynch
Biodesign Institute
Arizona State Univ.

"...a pervasive problem in biology is the religious adherence to the idea that natural selection is solely responsible for every feature of biological diversity.

(Lynch 2019, book manuscript in preparation, available at Lynch's website; from the opening chapter)

Any evolutionary process – drift, self-organization, you name it – must satisfy the following:

- **novel variations**
- **the variations make a difference**
- **pass the variations on, stably (that is, permanently) to offspring**



The point is, the toy (and real) examples I will present next apply to *any* candidate evolutionary process, not just natural selection.

So why does natural selection fail to explain the origin of metazoan cell lineages and differentiation?

If, within a species or population, the individuals

a. vary in some trait **q** – the condition of **variation**;

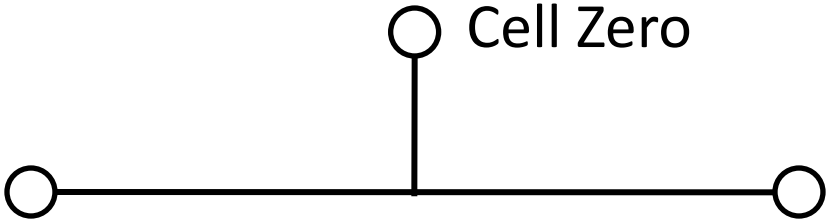
b. leave different numbers of offspring in consistent relation to the presence or absence of trait **q** – the condition of **selection**;

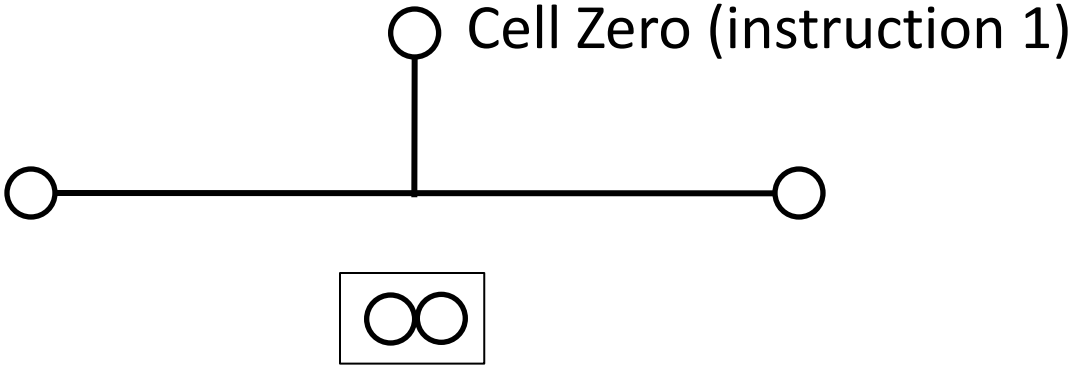
c. transmit trait **q** faithfully between parents and offspring – the condition of **heredity**;

then the frequency of trait **q** will differ predictably between the population of all parents and the population of all offspring.

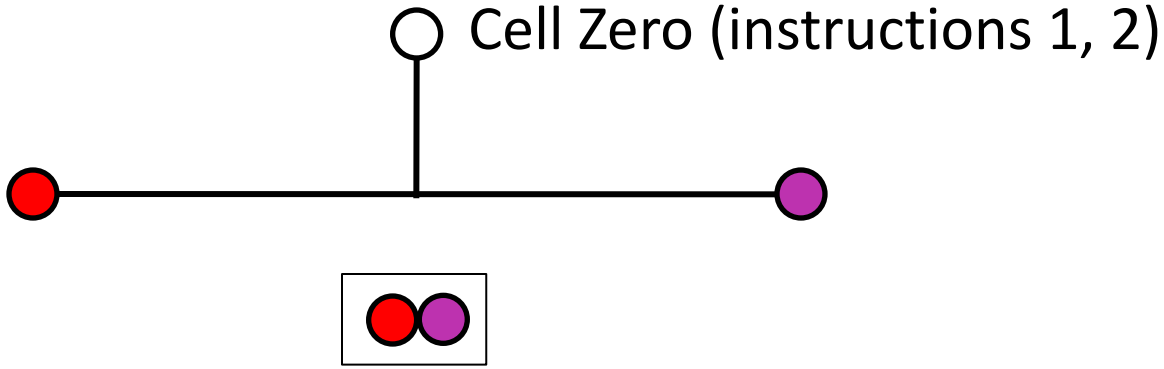
(Lewontin 1978; Endler 1986)

Let’s consider the *simplest possible* (toy) example. If natural selection does not work there, *a fortiori* it won’t work with more complicated cases, such as real animals.


Cell Zero

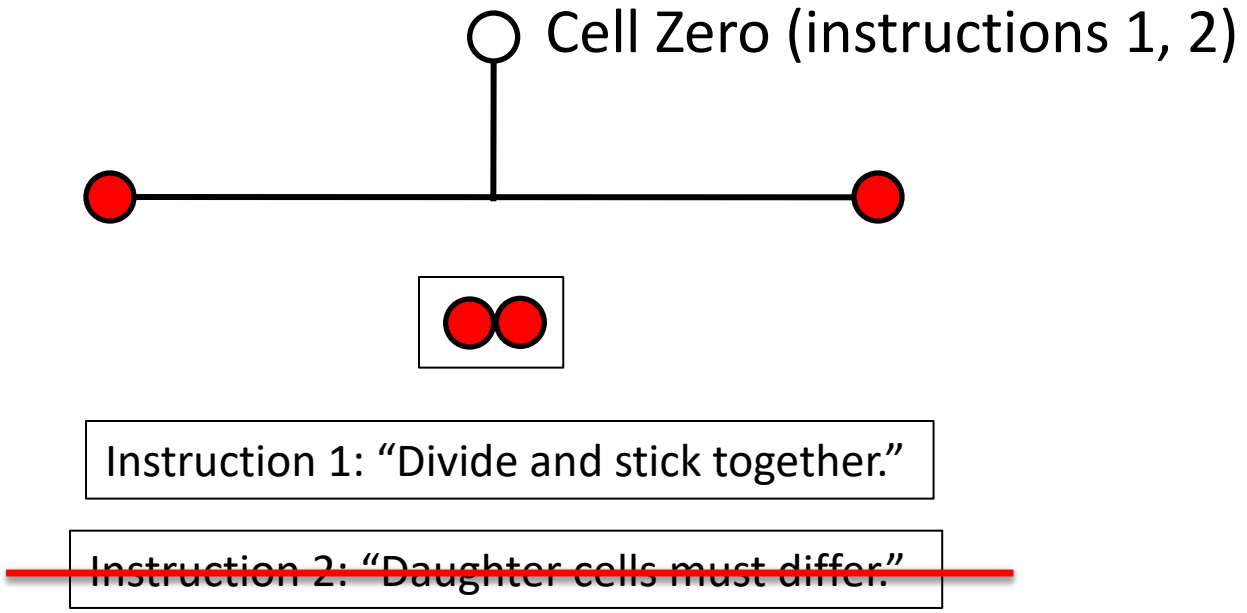


Instruction 1: “Divide and stick together.”

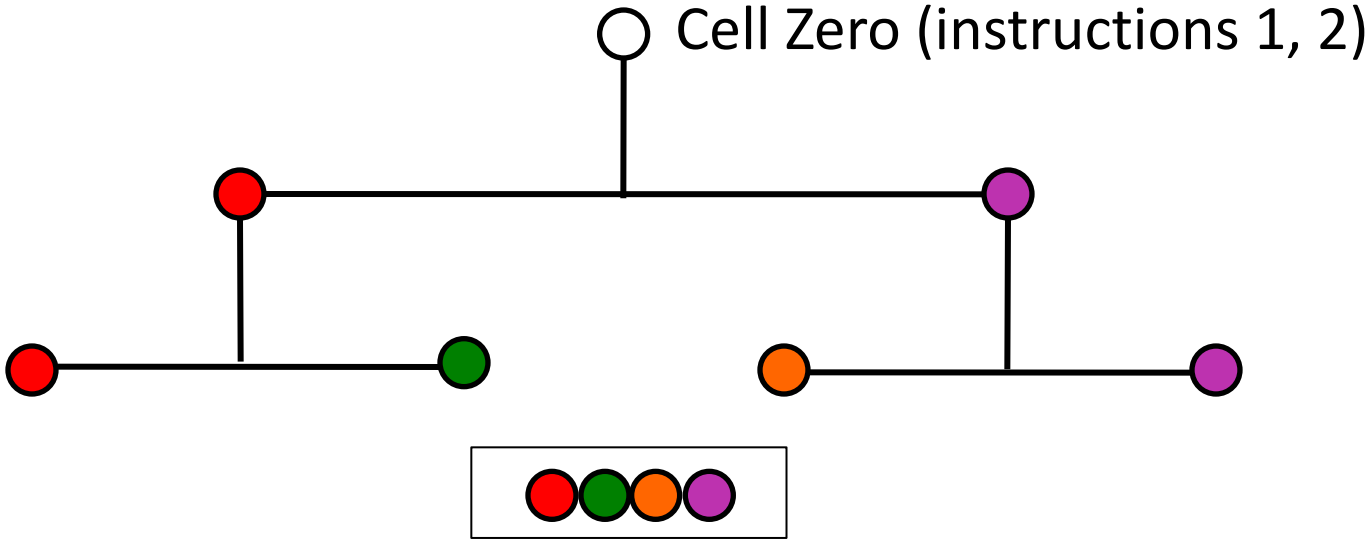


Instruction 1: “Divide and stick together.”

Instruction 2: “Daughter cells must differ.”



Simply multiplying cell number, *without also differentiating*, takes those cells in the wrong direction – if one wants to build an animal, that is, with a novel body plan, rather than only an undifferentiated mass of cells.

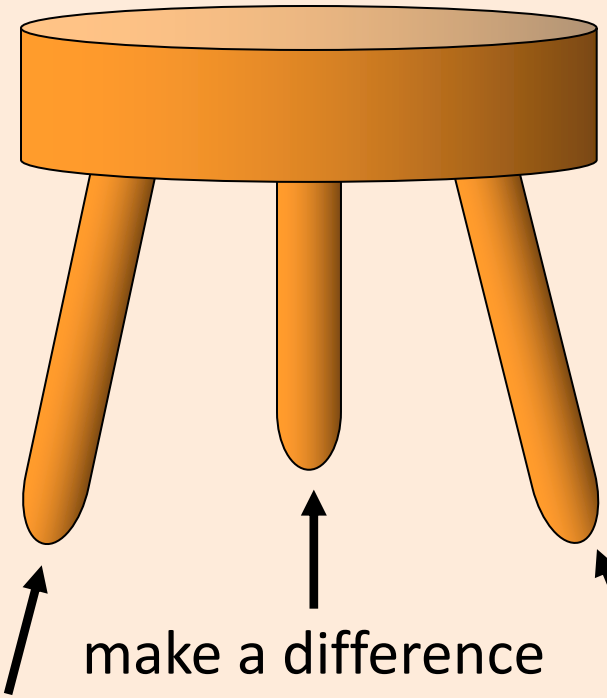


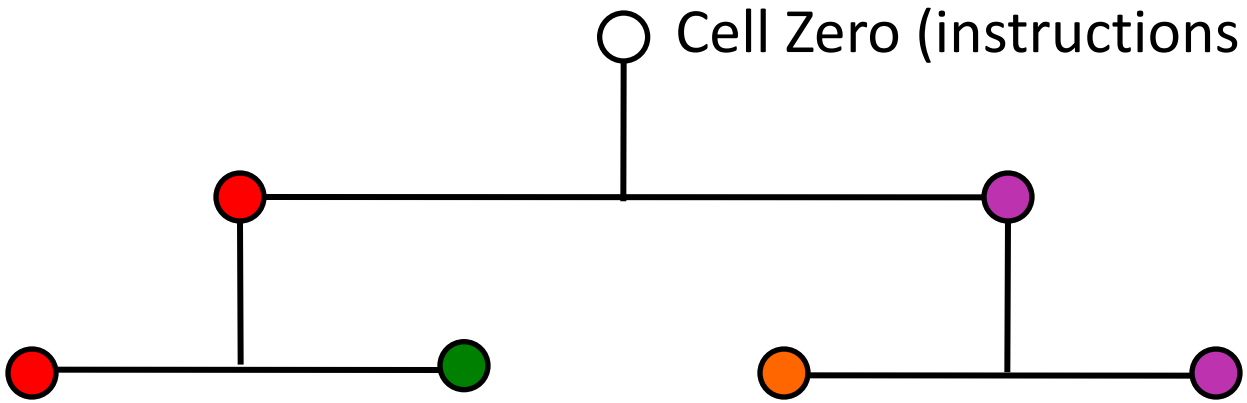
Instruction 1: “Divide and stick together.”

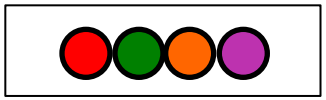
Instruction 2: “Daughter cells must differ.”

Remember: any evolutionary process must satisfy the following:

- **novel variations**
- **the variations make a difference**
- **pass the variations on, stably (that is, permanently) to offspring**

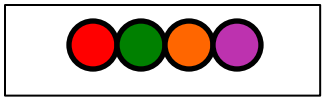


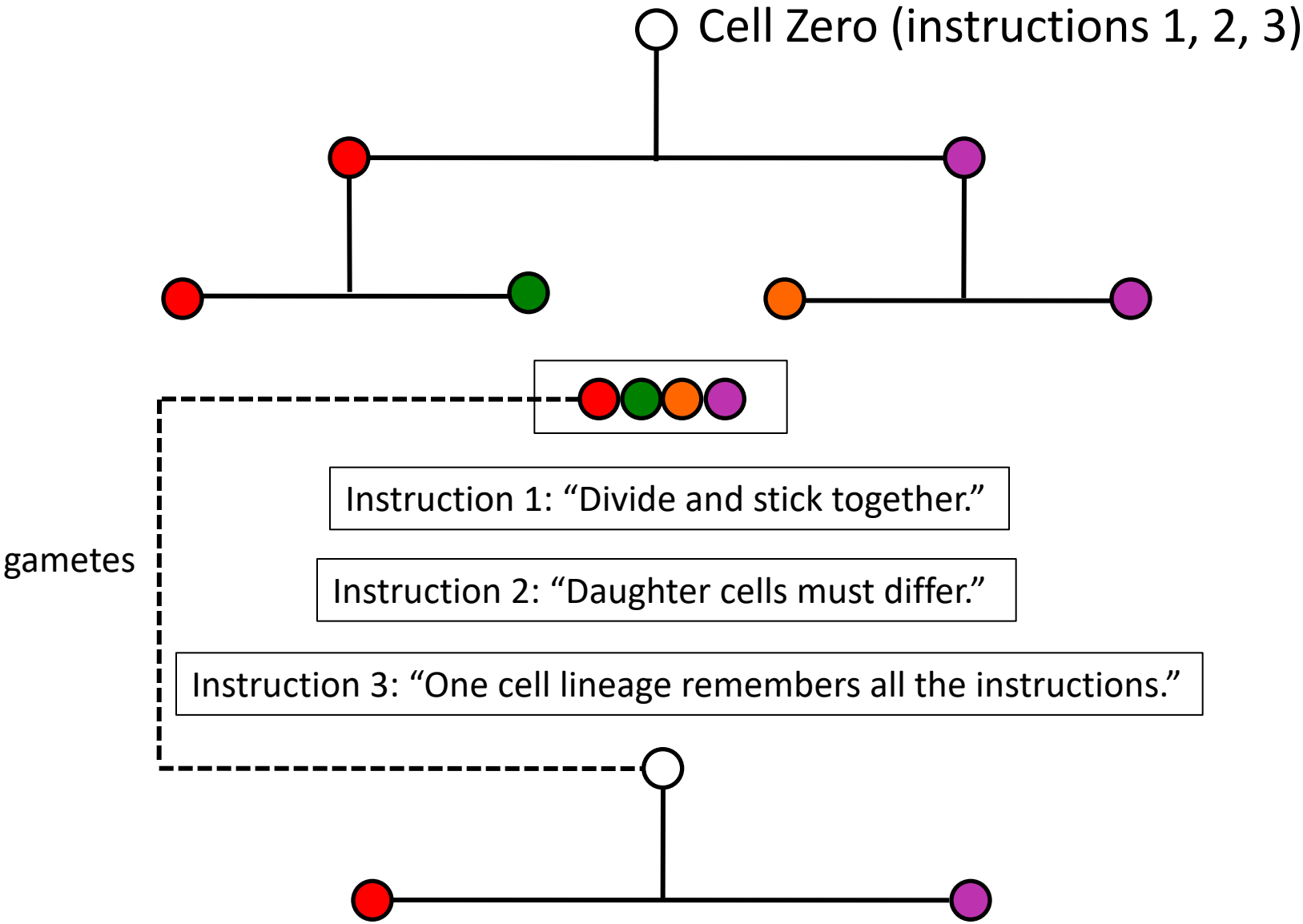


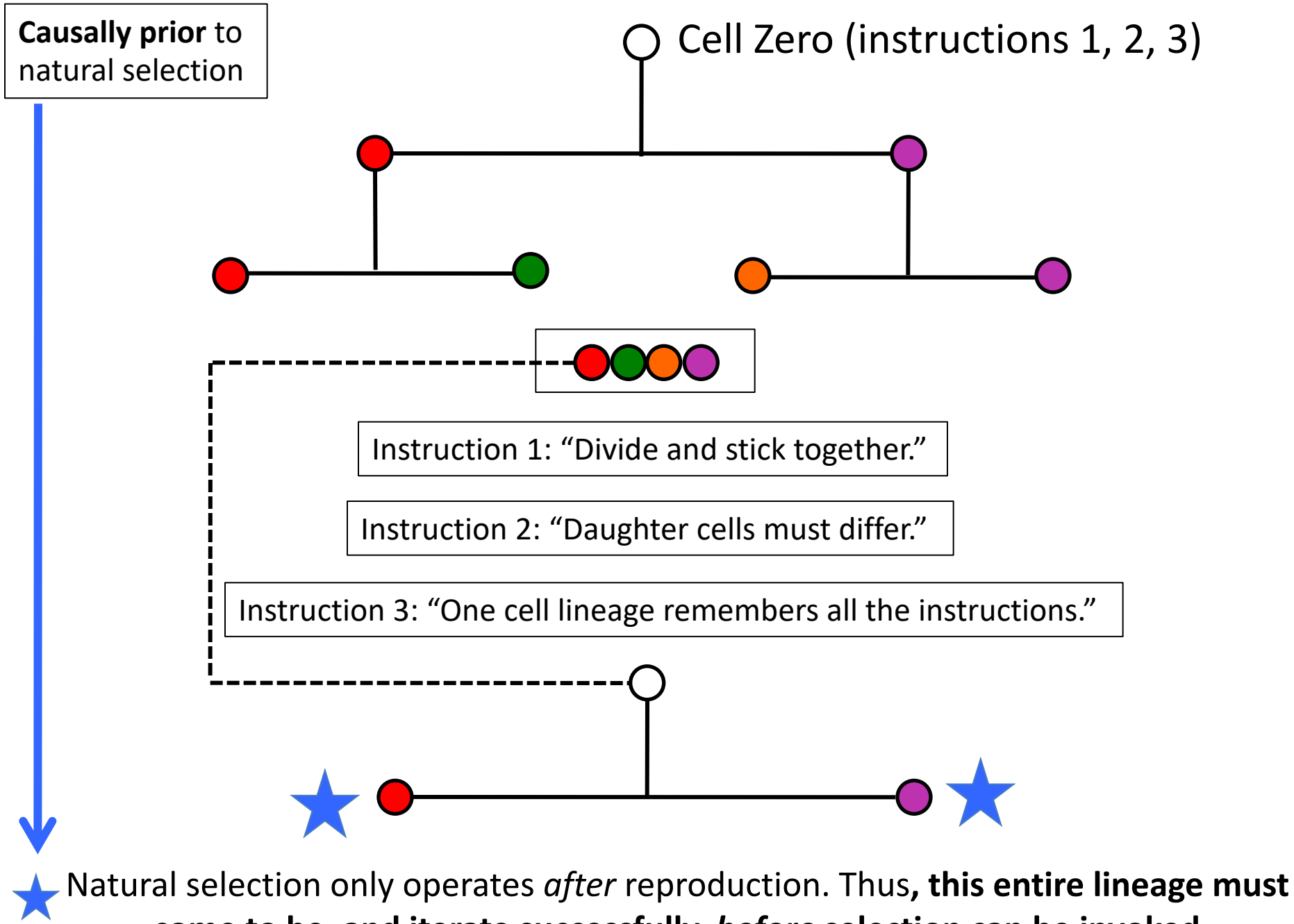


Instruction 1: “Divide and stick together.”

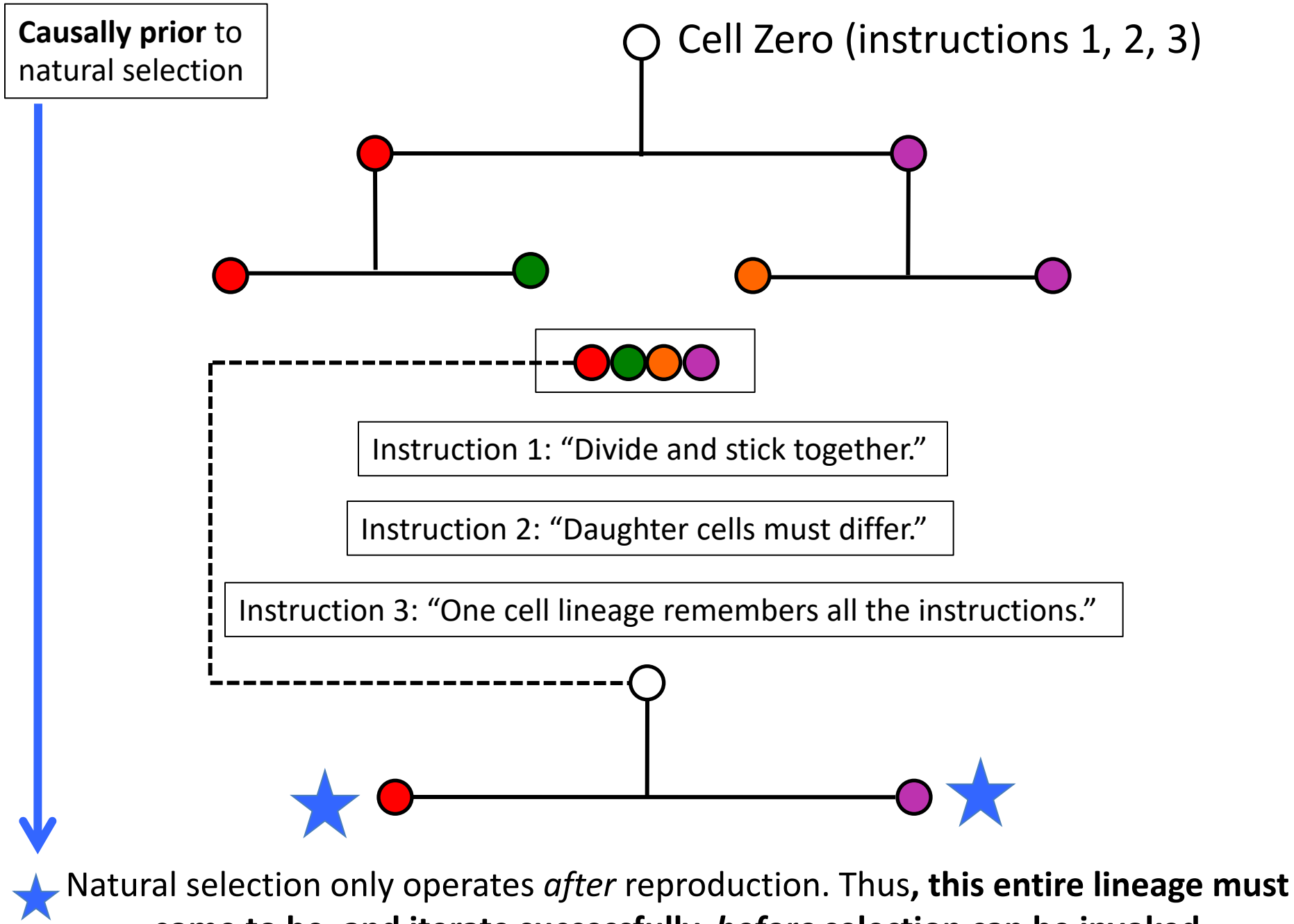
Instruction 2: “Daughter cells must differ.”




Cell Zero (instructions 1, 2, 3)
gametes
Instruction 1: “Divide and stick together.”
Instruction 2: “Daughter cells must differ.”
Instruction 3: “One cell lineage remembers all the instructions.”



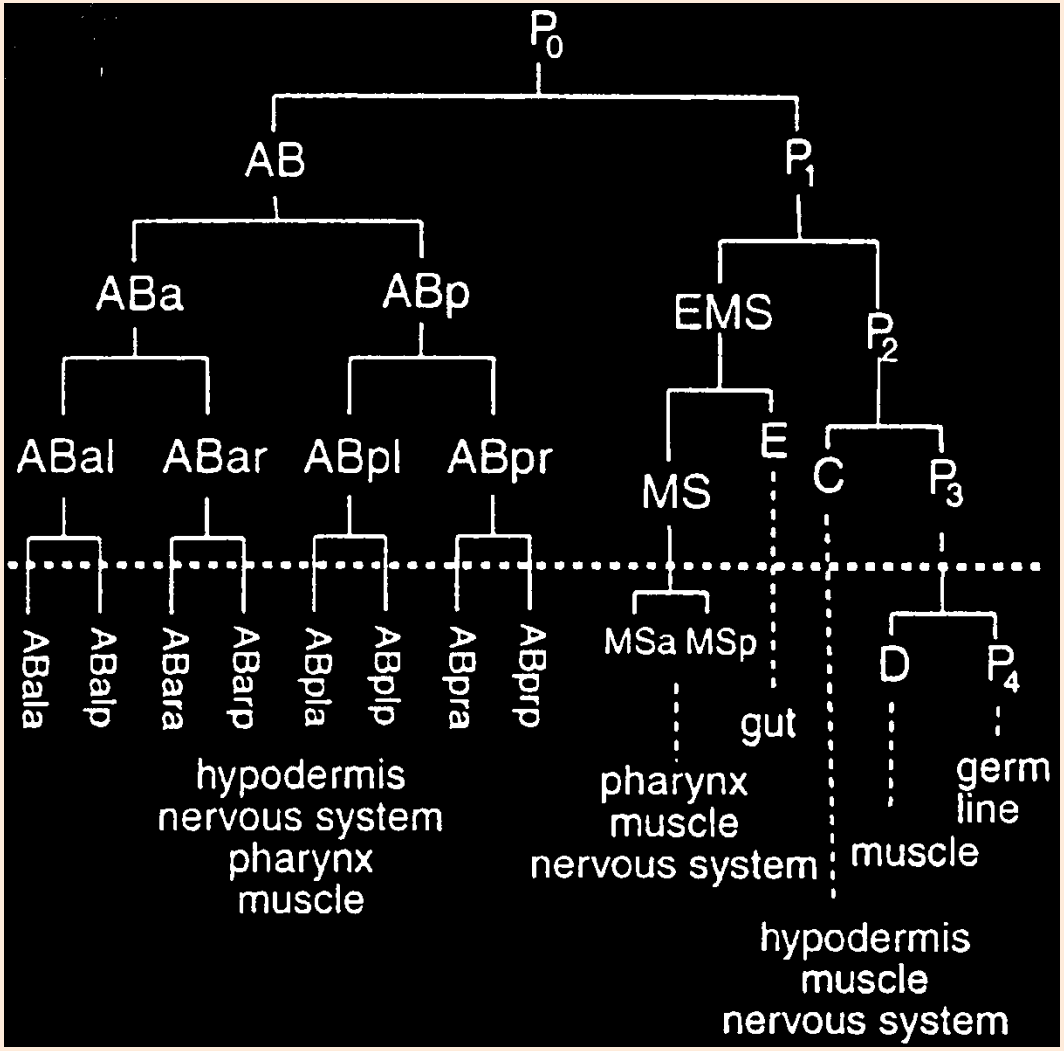
Natural selection only operates *after* reproduction. Thus, **this entire lineage must come to be, and iterate successfully, *before* selection can be invoked.**



Natural selection only operates *after* reproduction. Thus, **this entire lineage must come to be, and iterate successfully, *before* selection can be invoked.**

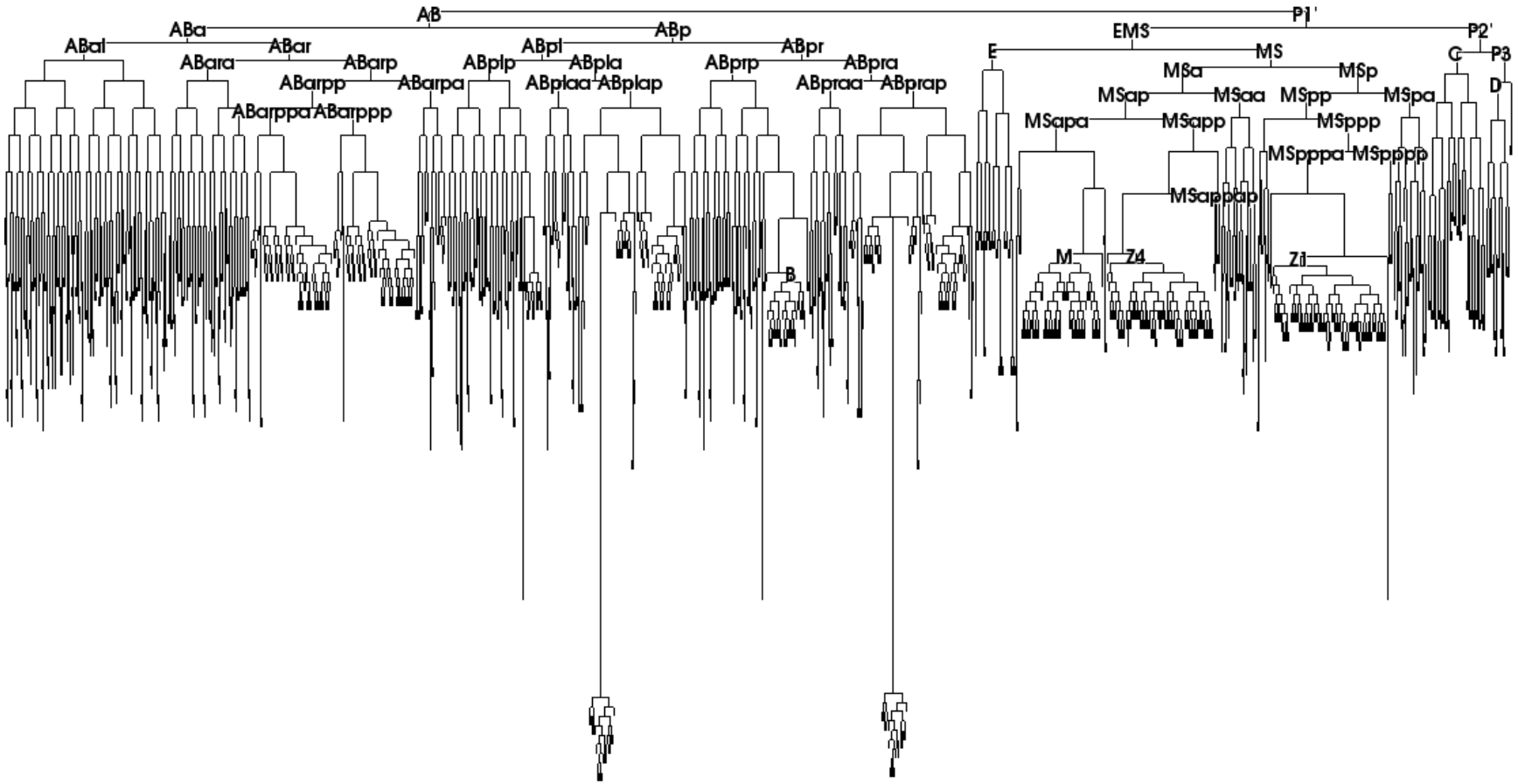
The adult anatomy of *Caenorhabditis elegans*

The early cell lineage of *Caenorhabditis elegans*

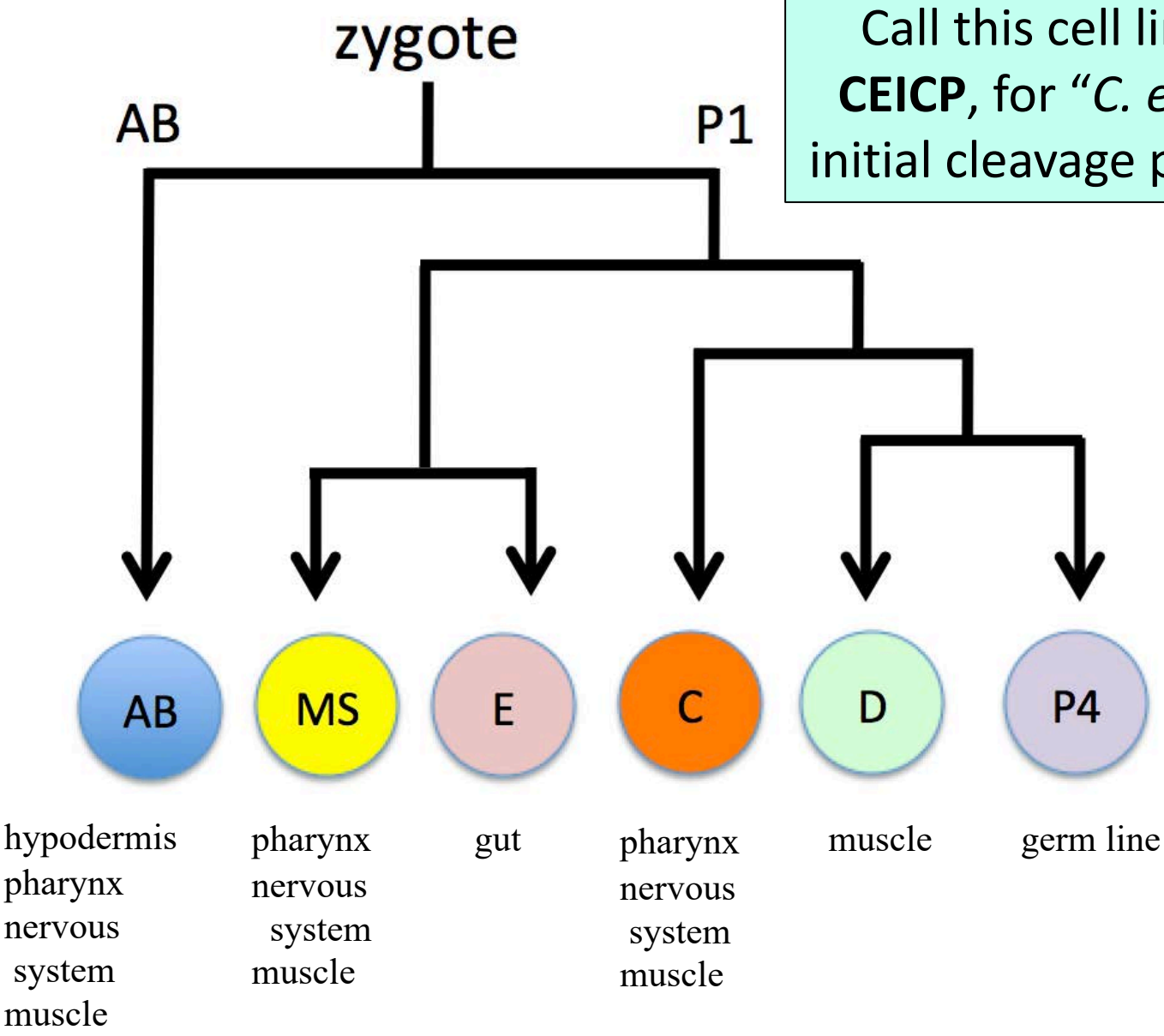


(figure after Schnabel 1997, 342)

The cell lineage of *Caenorhabditis elegans* (at hatching)

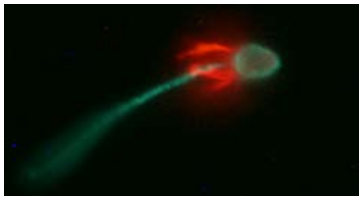


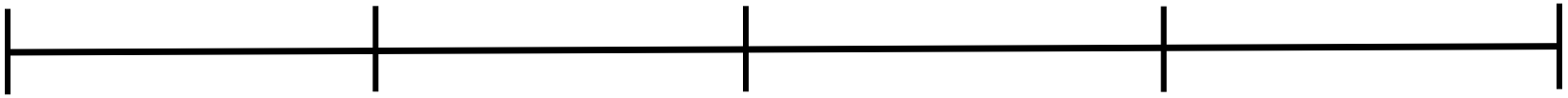
This early cell lineage is a defining character of *C. elegans.*



Call this cell lineage **CEICP**, for "*C. elegans* initial cleavage pattern."

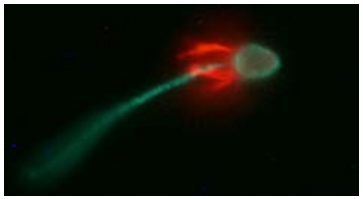
Increase in total adult cell number over the phylogenetic history of *Caenorhabditis* (intervals are arbitrary, indicating distance only)





Where, along this evolutionary trajectory, did the *C. elegans* early-branching cell lineage evolve?

Did CEICP evolve here?



Single-celled eukaryote

250 cells

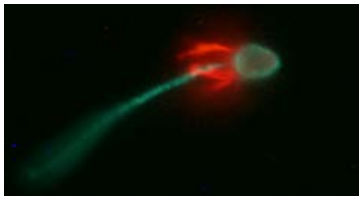
500 cells

750 cells

Adult *C. elegans*

Where, along this evolutionary trajectory, did the *C. elegans* early-branching cell lineage evolve?

Did CEICP evolve here?

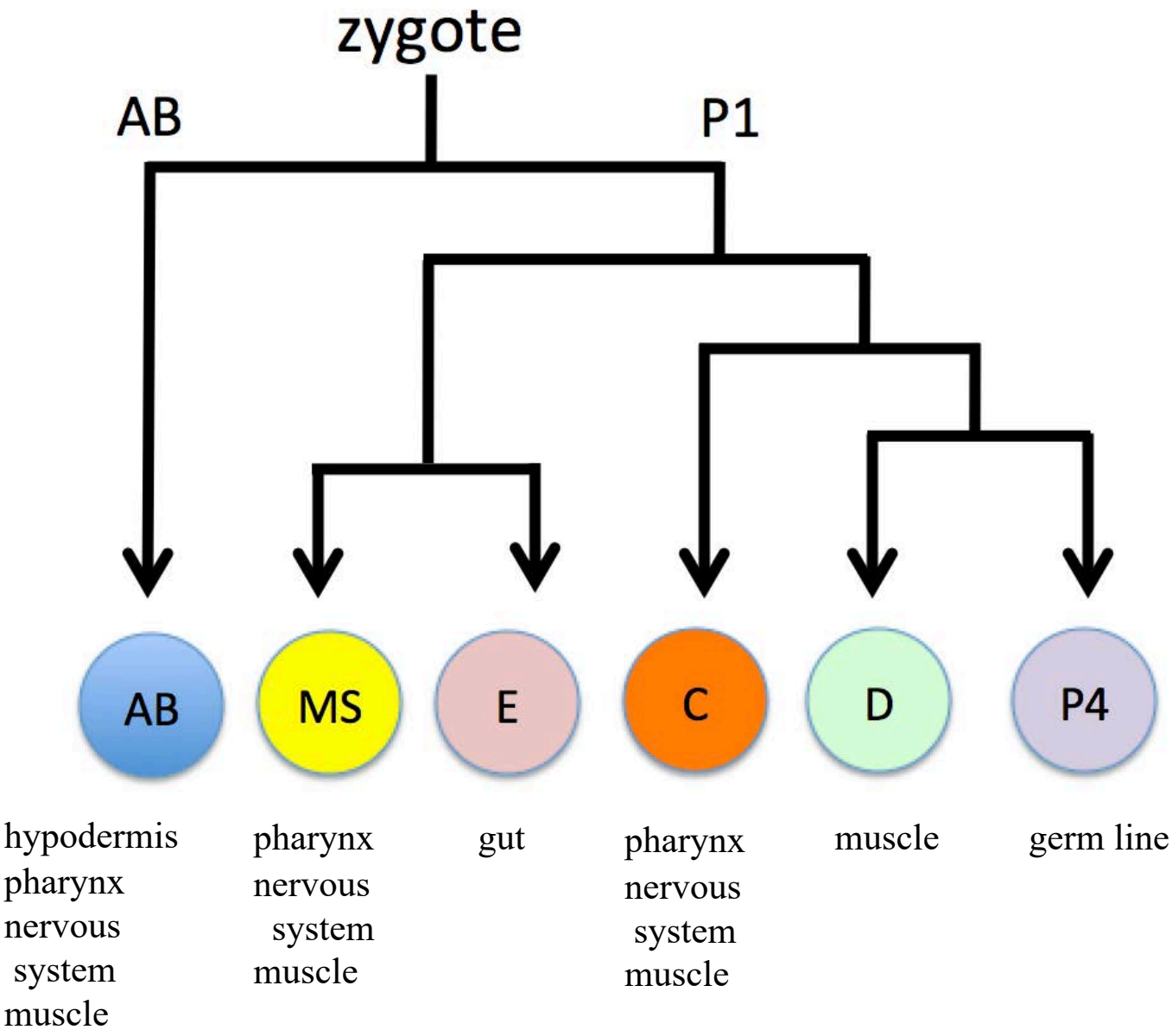


Single-celled eukaryote

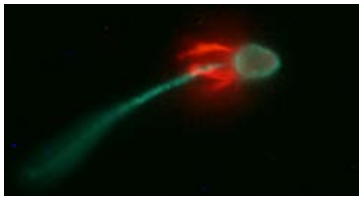
250 cells

None of this exists yet.

These cell cleavages function with respect to their *distant endpoints*.



Did CEICP evolve here?



Single-celled eukaryote

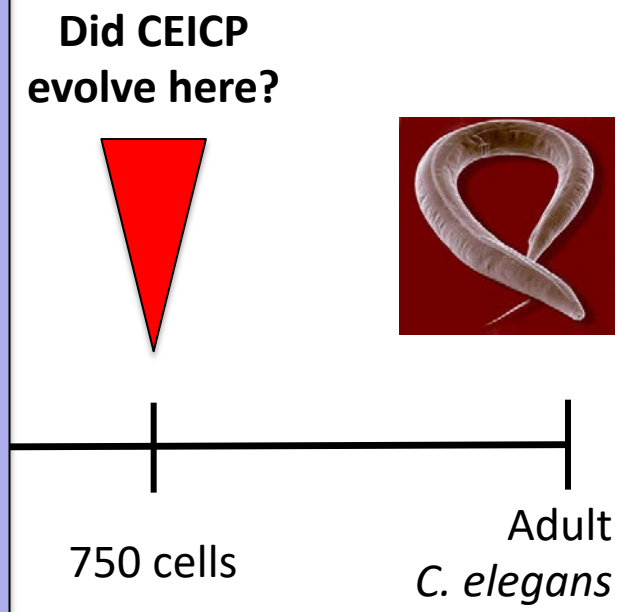
250 cells

500 cells

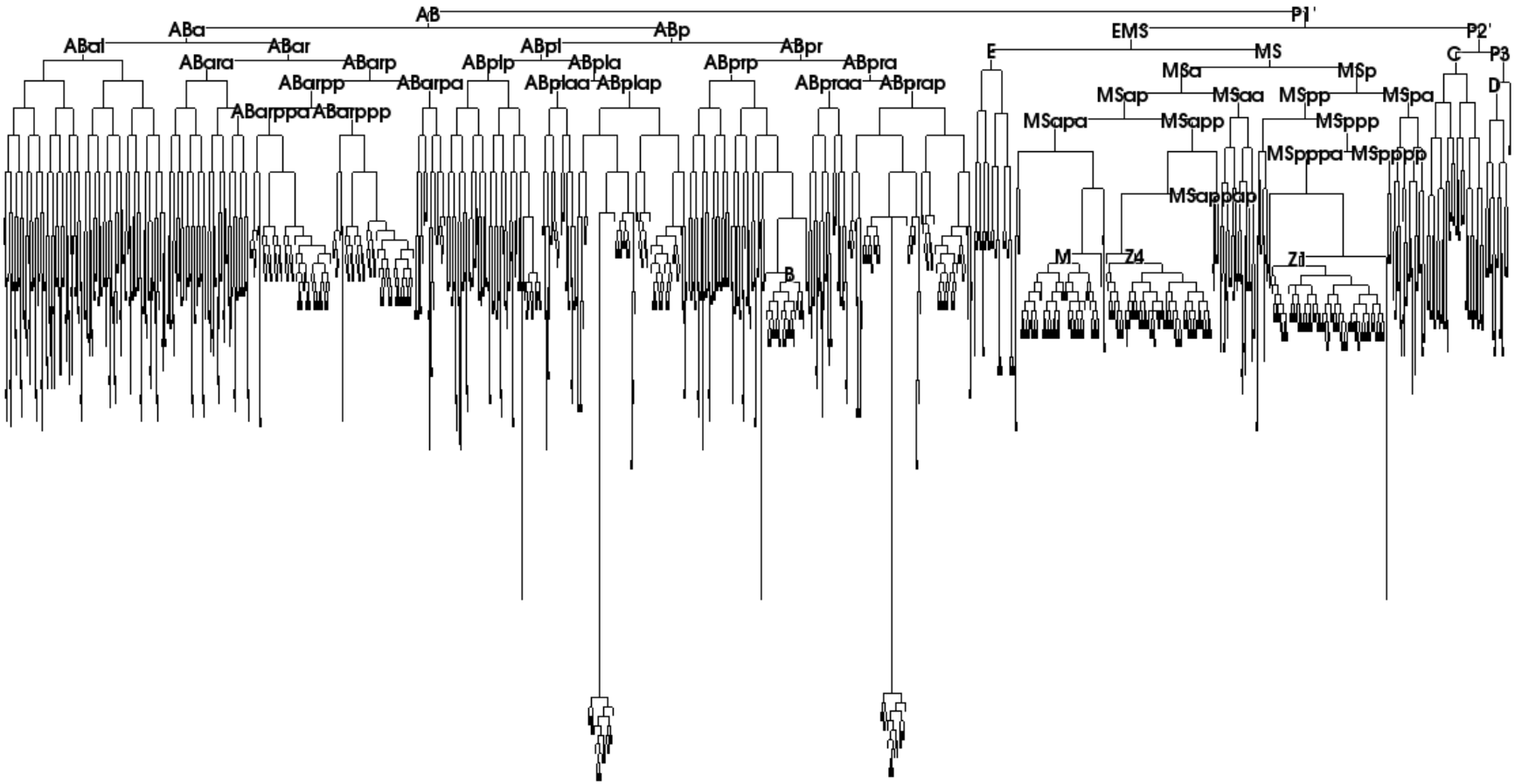
750 cells

Adult *C. elegans*

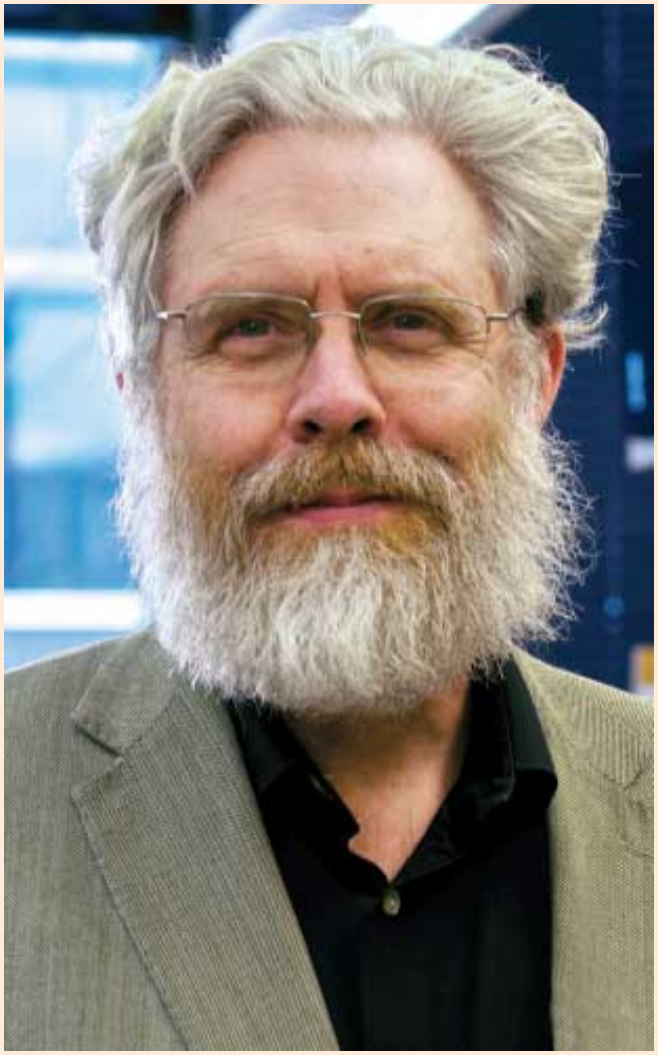
This will be a clonal mass of cells, lacking any differentiation.



The complete cell lineage of *Caenorhabditis elegans*



The evidence seems strongly to indicate that one needs this *entire* ontogenetic pathway if one wants *any part* of it.



George Church

Harvard University

GENETIC DESIGN AUTOMATION FOR AUTONOMOUS FORMATION OF MULTICELLULAR SHAPES FROM A SINGLE CELL PROGENITOR

A PREPRINT

Evan Appleton[1,2,+], Noushin Mehdipour[3,+], Tristan Daifuku[1,2,+], Demarcus Briers[3,4,+], Iman Haghighi[3,+], Michael Moret[1,2], George Chao[1,2], Timothy Wannier[1,2], Anush Chiappino-Pepe[1,2], Calin Belta[3], and George Church[1,2]
[1]Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, Massachusetts, USA
[2]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA
[3]Department of Systems Engineering, Boston University, Boston, Massachusetts, USA
[4]Bioinformatics Program, Boston University, Boston, Massachusetts, USA
[+]These authors contributed equally
[*]Correspondences addressed to evan_appleton@hms.harvard.edu, cbelta@bu.edu, and gchurch@genetics.med.harvard.edu

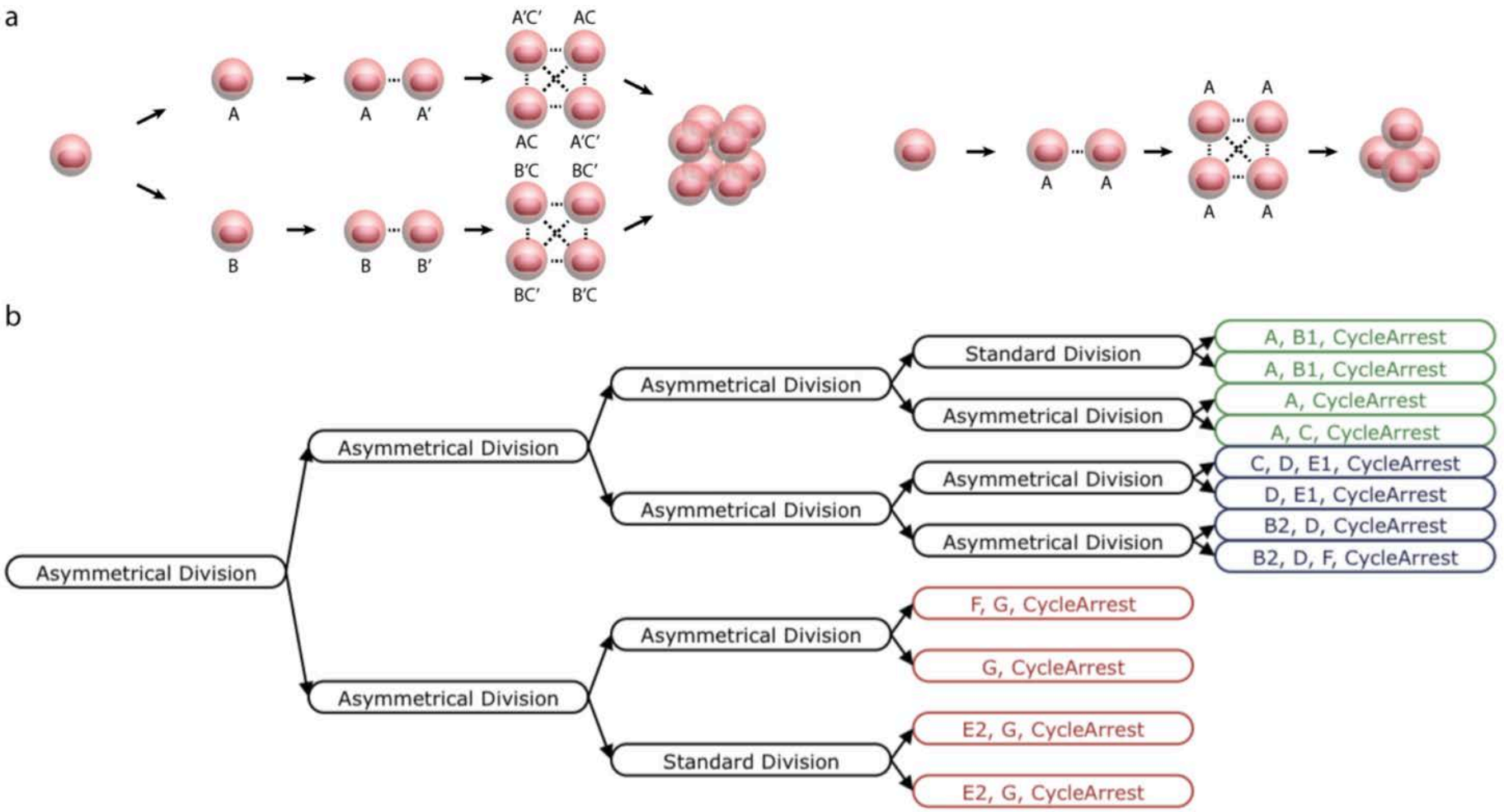


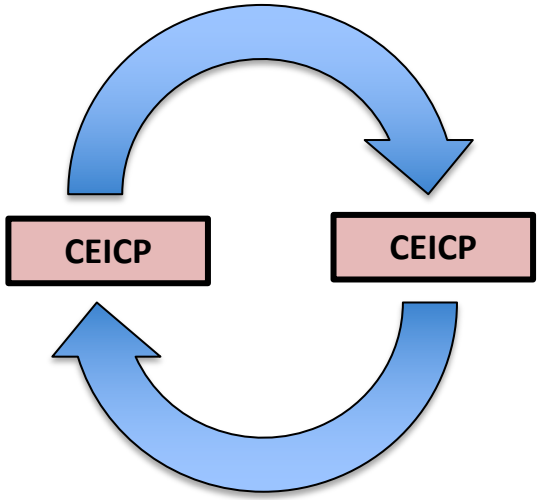
Figure 3: Developmental trees are produced from lists of blocks and connections they must form. a. Each tet block is made from a single homodimer (protein A) and each quad block from three orthogonal heterodimer pairs (proteins A/A', B/B', and C/C'). b. Starting from the final cells that must exist in a shape, a binary tree can be created to determine how many divisions must happen to form all of the blocks and optimize where asymmetrical divisions must occur.

"Starting from the final cells that must exist..."

The developmental pathway of *C. elegans* provides a striking example of causal circularity.

foresight

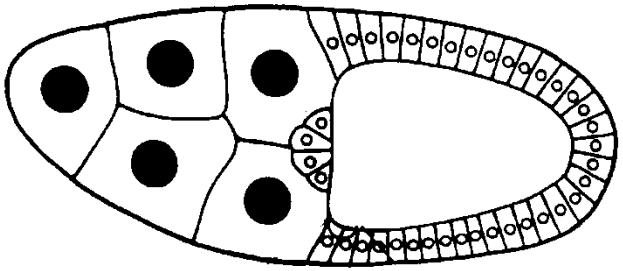
design



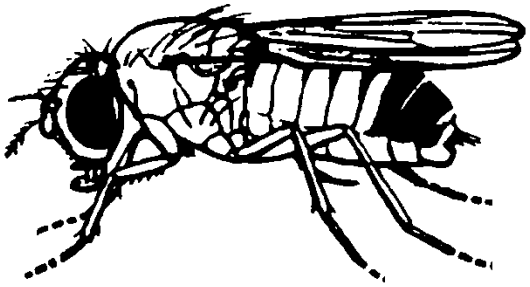
causal circularity

Indeed, developmental pathways throughout the Metazoa represent one example of causal circularity after another.

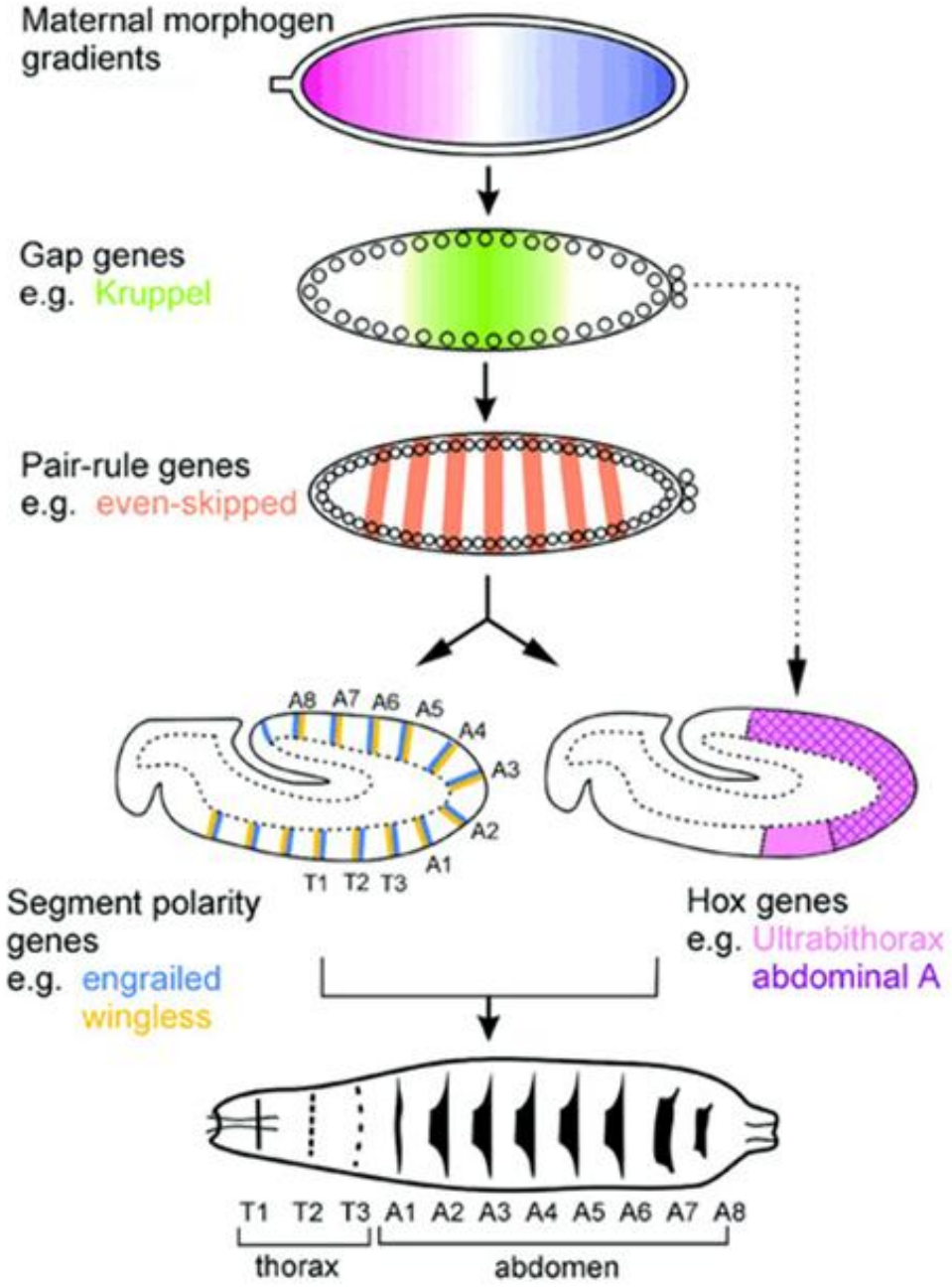
Mom



Oocyte

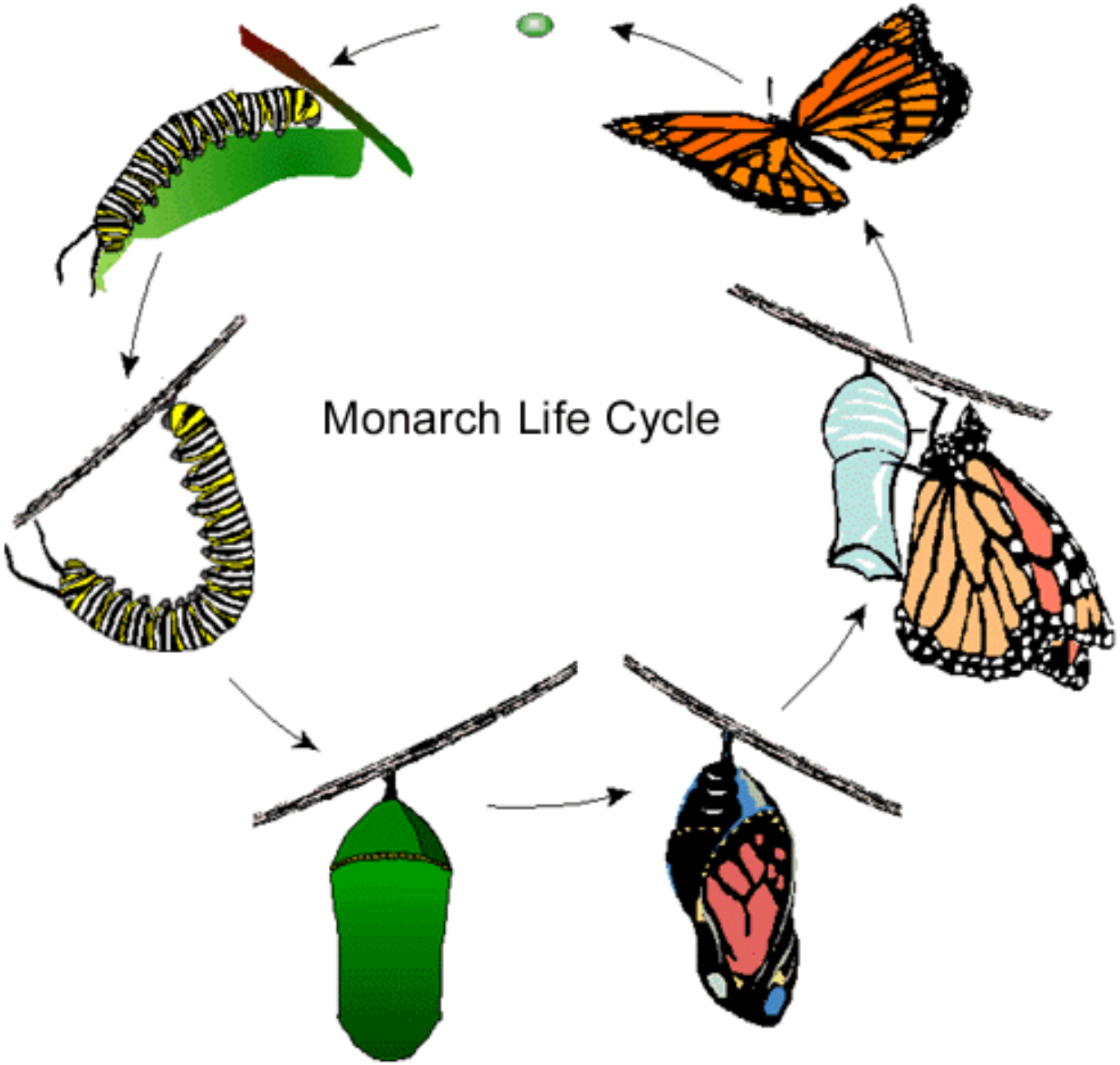


Junior


Maternal morphogen gradients
Gap genes
e.g. Kruppel
Pair-rule genes
e.g. even-skipped
A8 A7 A6 A5 A4 A3 A2 A1
T1 T2 T3
Segment polarity genes
e.g. engrailed
wingless
Hox genes
e.g. Ultrabithorax
abdominal A
T1 T2 T3 A1 A2 A3 A4 A5 A6 A7 A8
thorax
abdomen

The North American Monarch (*Danaus plexippus*)

Houdini entering his “water torture” box:

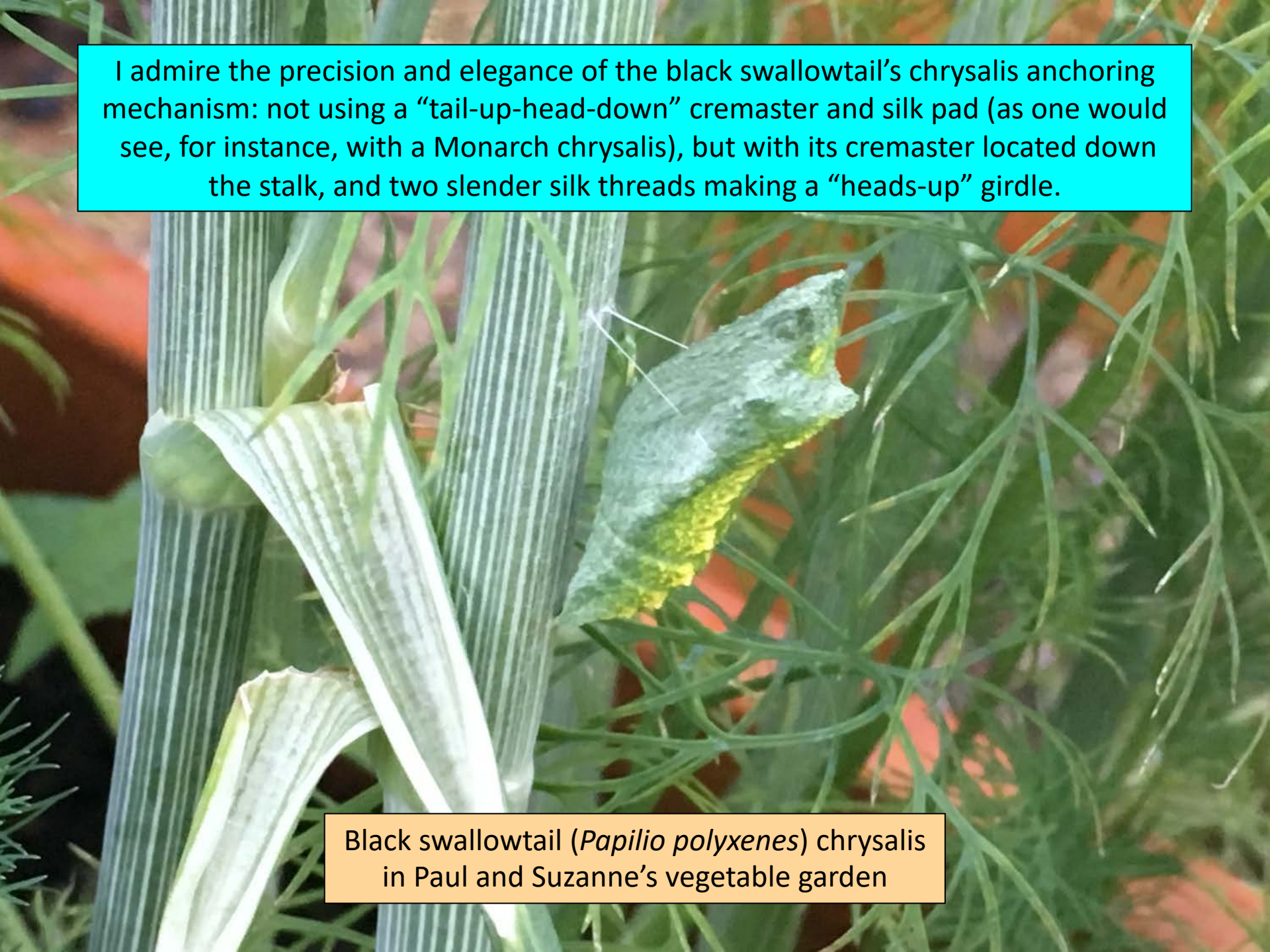

Monarch Life Cycle

Monarch egg capsule on the underside of a milkweed leaf.

Close-up of Monarch egg.

This and the next two slides are here for no better reason than, for several summers now, my wife and I have planted dill in our backyard garden boxes – and female black swallowtail butterflies show up faithfully to lay their eggs on it. Dill is one of the black swallowtail's preferred (host) plants. *Biology is the best.* It really is.

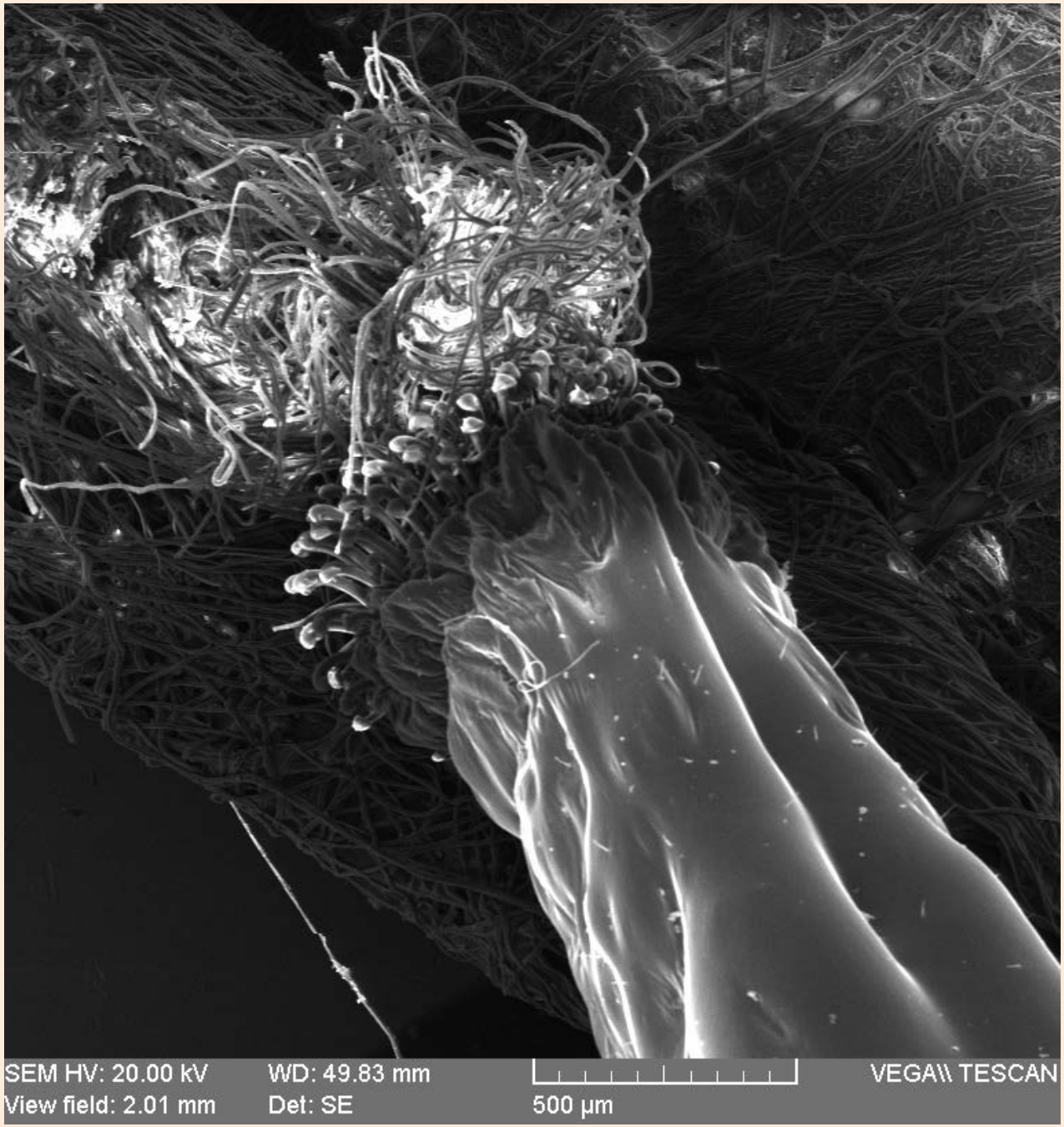
Black swallowtail (*Papilio polyxenes*) caterpillar feeding on dill in Paul's backyard vegetable garden.



I admire the precision and elegance of the black swallowtail's chrysalis anchoring mechanism: not using a "tail-up-head-down" cremaster and silk pad (as one would see, for instance, with a Monarch chrysalis), but with its cremaster located down the stalk, and two slender silk threads making a "heads-up" girdle.

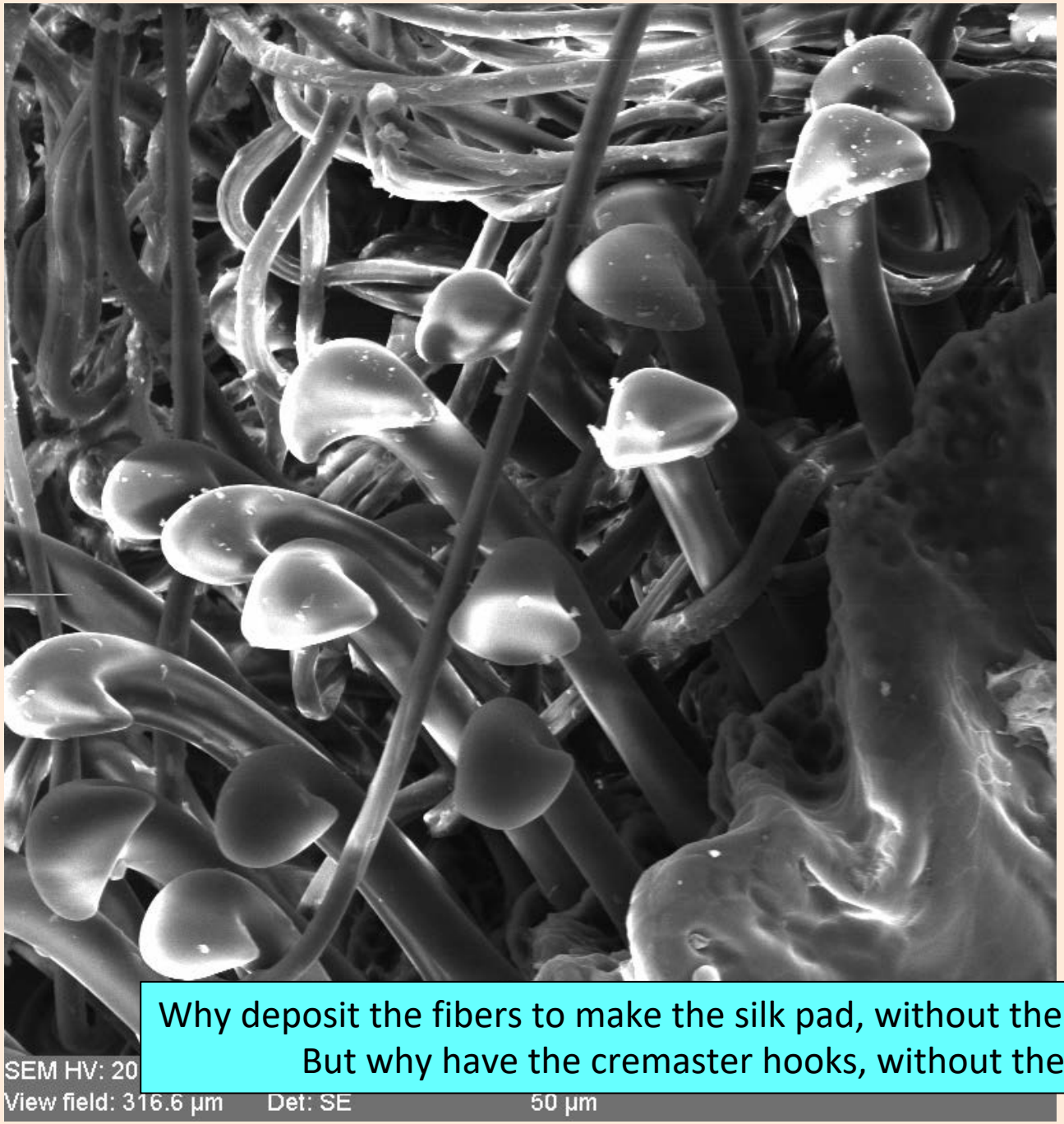
Black swallowtail (*Papilio polyxenes*) chrysalis in Paul and Suzanne's vegetable garden

Black swallowtail male in Paul’s office. This little fellow came in as a chrysalis stowaway when the plant he is standing on (in the photo) was moved indoors.



SEM of a Monarch cremaster, showing its "grasping" tip embedded in the silk pad spun by the caterpillar.

(Images, courtesy of Dr. Timothy Standish, Loma Linda University, and Illustra Media)



SEM of a Monarch cremaster, showing its "grasping" tip embedded in the silk pad spun by the caterpillar.

(Images, courtesy of Dr. Timothy Standish, Loma Linda University, and Illustra Media)

Why deposit the fibers to make the silk pad, without the cremaster hooks?
But why have the cremaster hooks, without the silk pad?

What is happening in the pupal (chrysalis) stage?

The caterpillar (larval) tissues are being destroyed by massive cell death, through processes of apoptosis and autophagy.

Cell populations sequestered in the larva then develop into the structures – legs, wings, genitalia, etc. – of the adult butterfly.

Houdini entering his “water torture” box:

Houdini wouldn't dare to enter the box, however, *without a plan to get out:*

(going in)

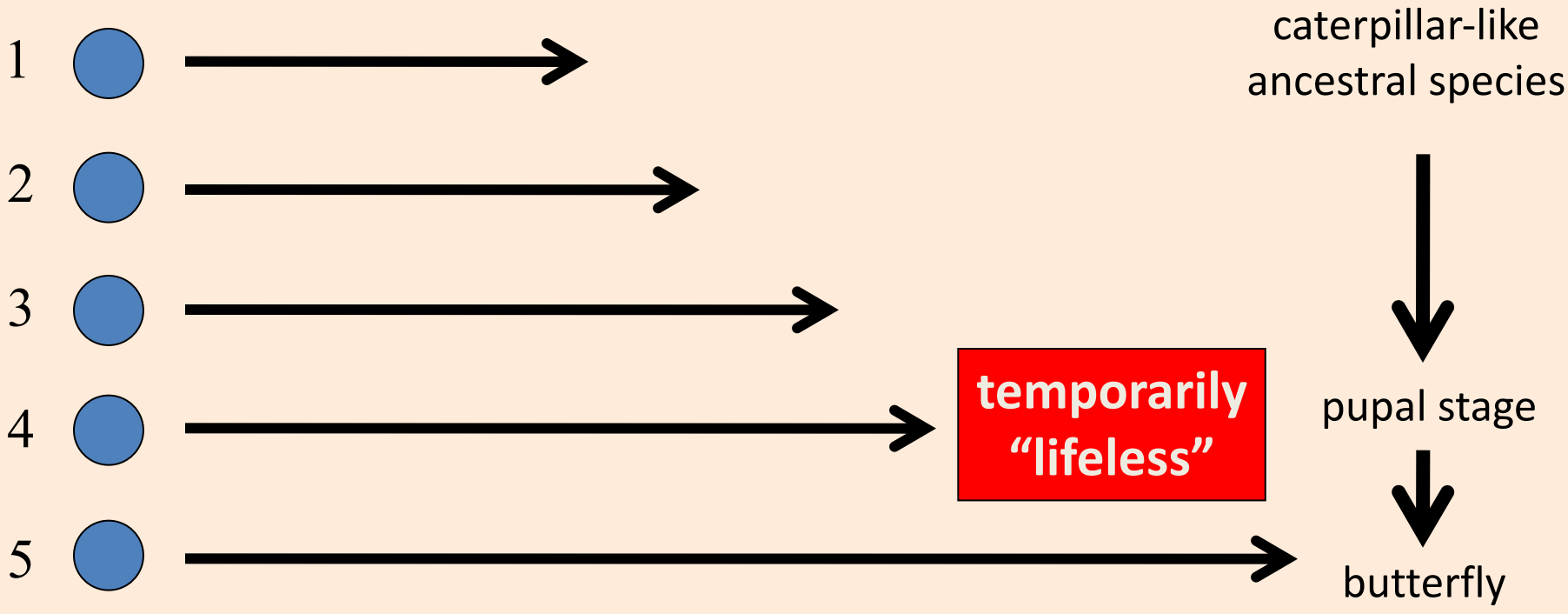
DEATH

(getting out)

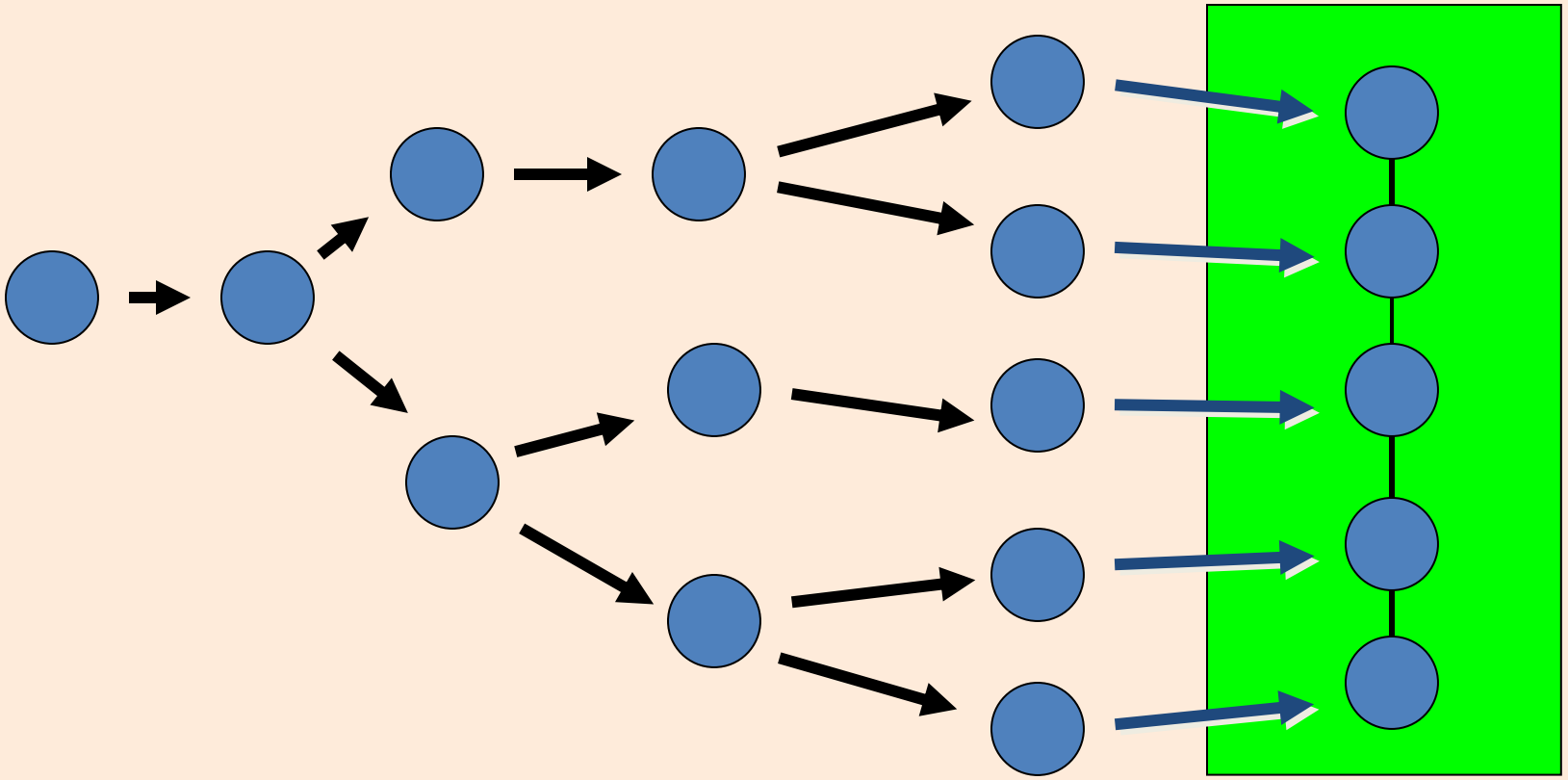
(going in)

(getting out)

Problem: How, then, did the pupal stage originate? – as evolving it would seem to require that natural selection simultaneously "knew how to get out."



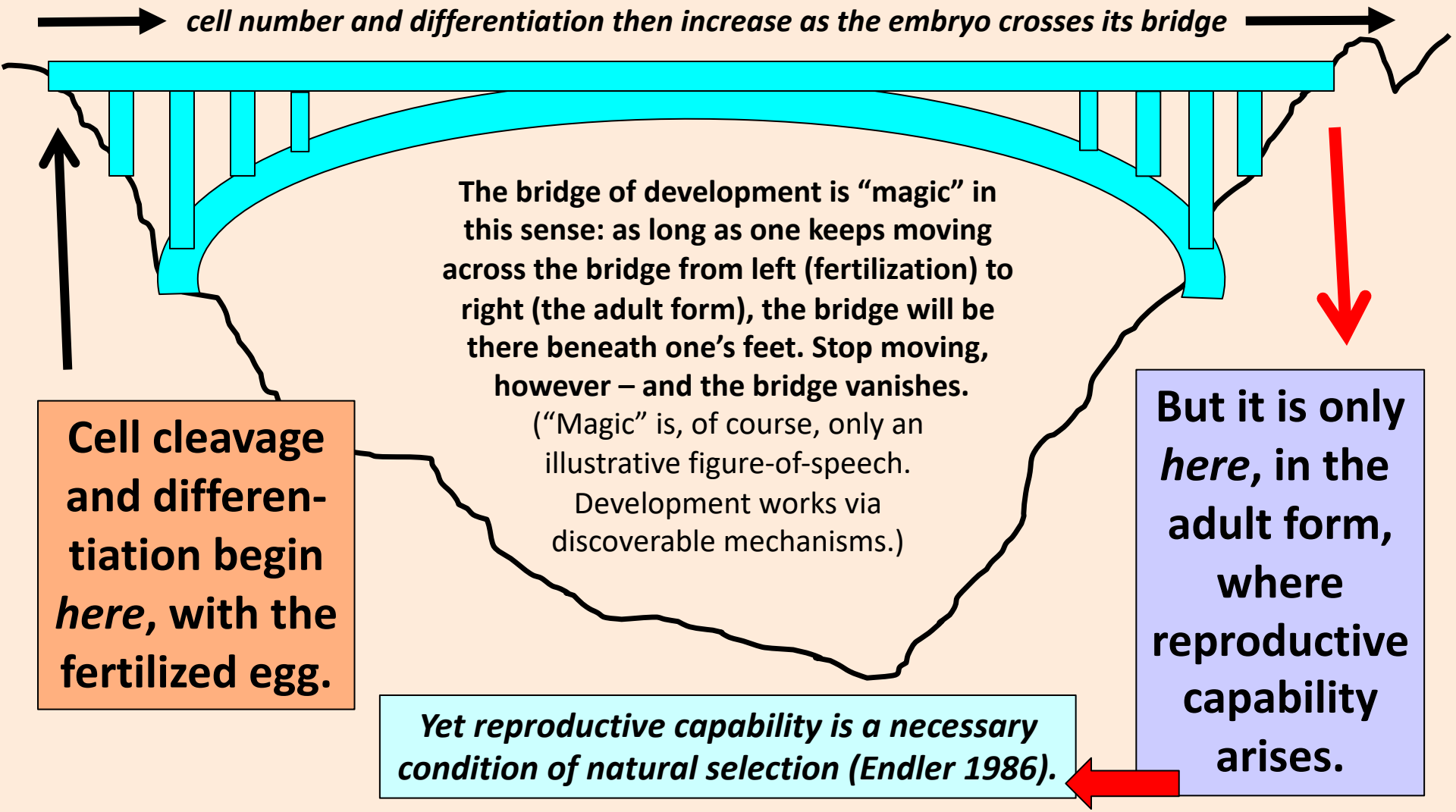
Can natural selection build complex developmental networks?



Natural selection only “sees” reproductive output.

The “magic bridge” of animal development

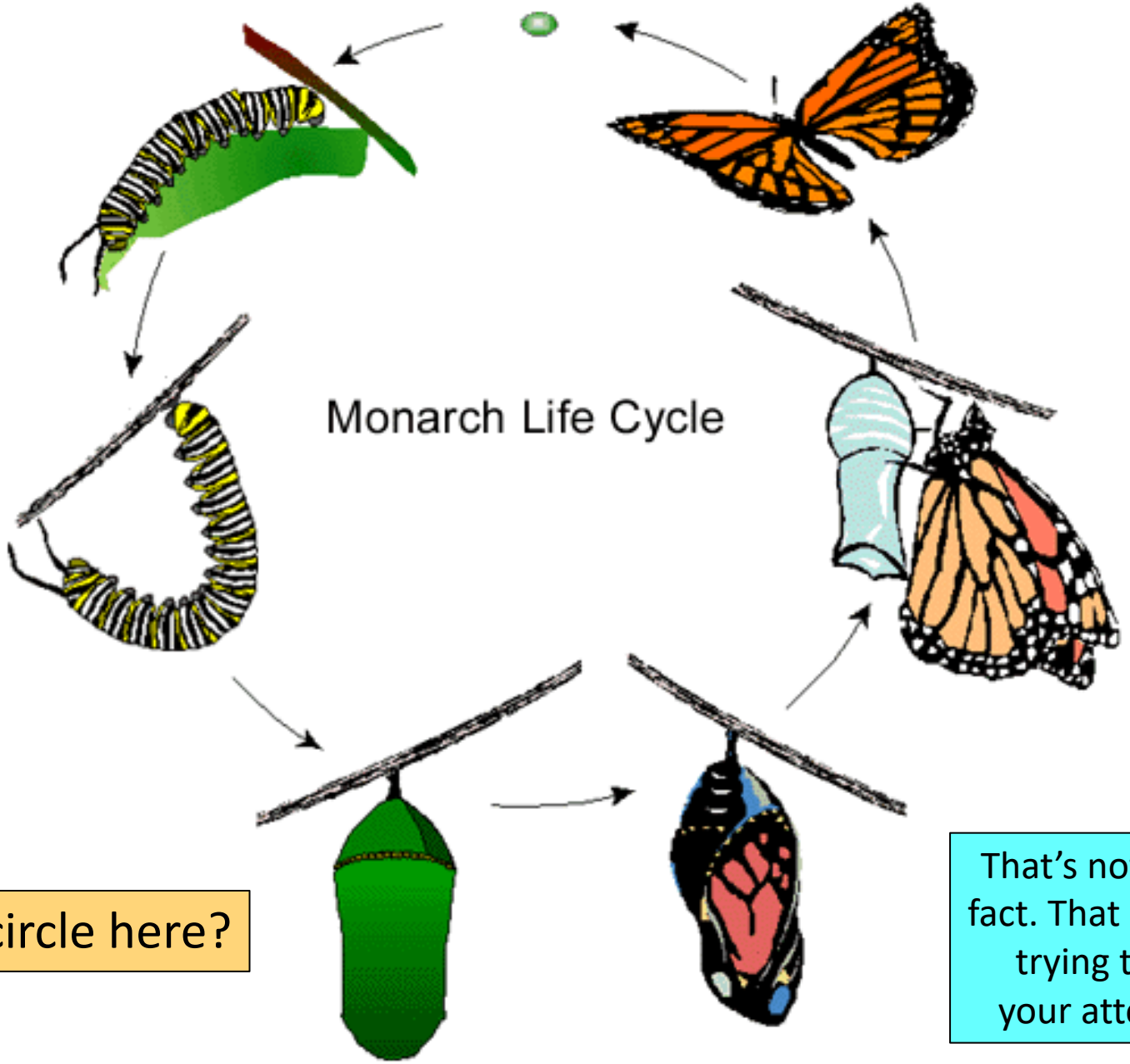
(Nelson & Gauger 2011:28)



So how do organisms solve this problem – that is, obtaining the instructions to build an embryo?

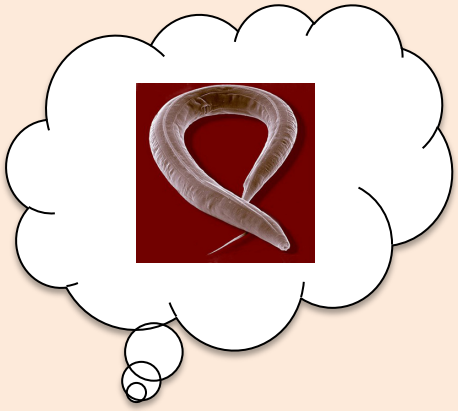
Answer:

They have *parents*.



See a circle here?

That's not a stray fact. That is nature trying to get your attention.

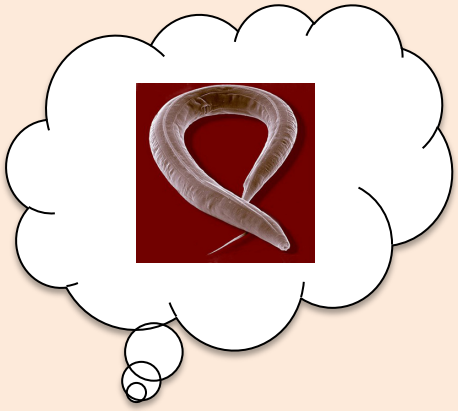


foresight

design

C. elegans

This would be an unobjectionable, even obvious causal inference...except for philosophy.



foresight

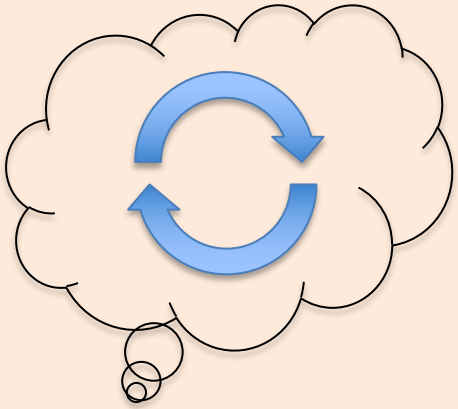
design

C. elegans

This explanatory option will take a major hit from the rise of naturalism in the 19th century.

After 1859: sorry, *but there is no such thing* as foresight in biological explanation.

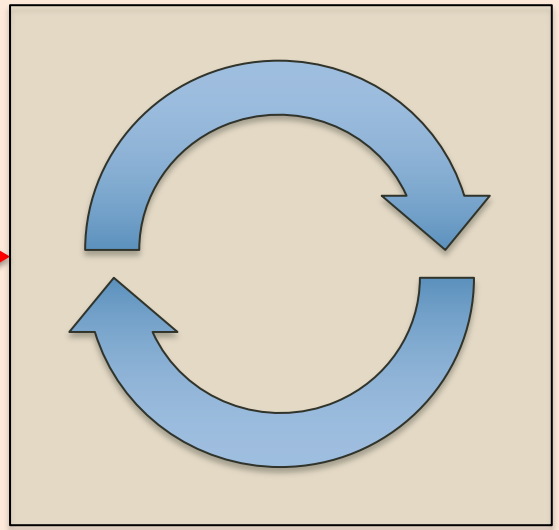
C. elegans



Recovering this inference looks to be *necessary* for advancing biological understanding.

foresight

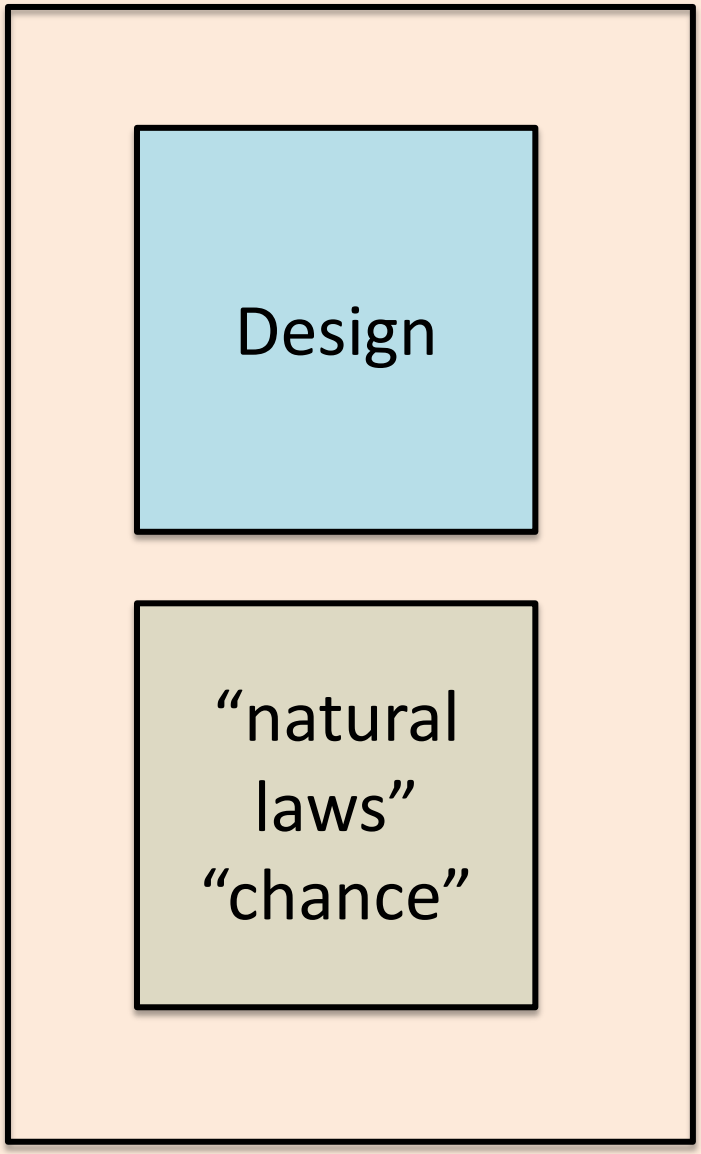
design



causal circularity

The *mental construct of the whole* is causally primary. Mind leads, seeing the target; realization follows.

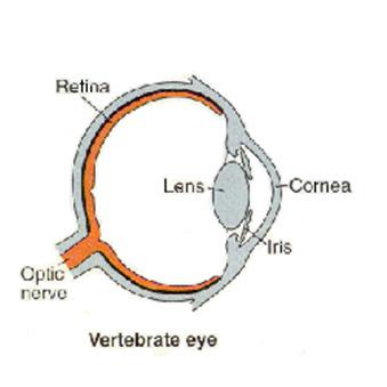
Any theory that *entails* its competitor, where the opposite is *not* the case, cannot lose when they go head-to-head in an explanatory context.



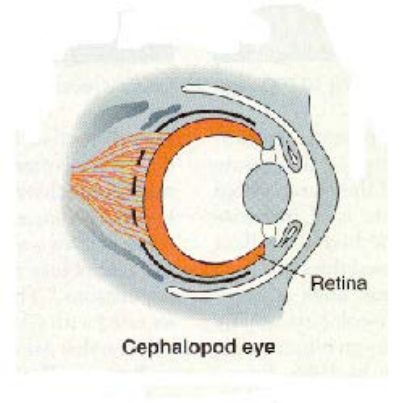
The explanatory toolkit of the **ID biologist** entails, or includes, *all* the causal possibilities of a philosphical naturalist.

But the opposite is not true. **This asymmetry favors ID inquiry in every circumstance.**

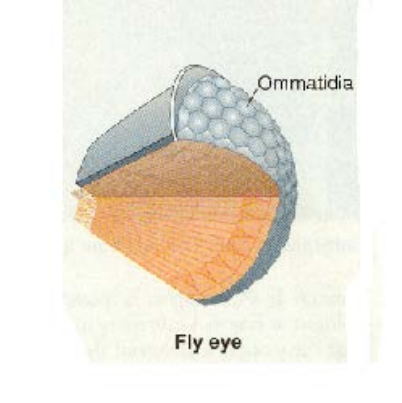
Three classically non-homologous eyes



Vertebrates



Molluscs



Arthropods

But the development of each of these eyes is regulated by the same so-called “master regulator” gene, Pax-6 (eyeless).

Let’s try a design-theoretic thought experiment – one which takes the reality of higher levels seriously.

Taking a lexicon from the Gettysburg Address:

"...that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion – that we here highly resolve that these dead shall not have died in vain – that this nation, under God, shall have a new birth of freedom – and that the government of the people, by the people, for the people, shall not perish from the earth."

Taking a lexicon from the Gettysburg Address:

a
and
birth
by
cause
dead
devotion
died
earth
for
freedom
from
full
gave
God
government
have
here
highly
honored
in
increased
last
measure
nation
new
not
of
people
perish
resolve
shall
take
that
the
these
they
this
to
under
vain
we
which

Taking a lexicon from the Gettysburg Address...*and writing a very different text – an anarchist's manifesto:*

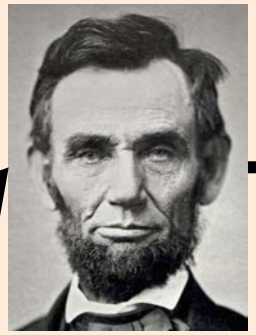
"...by this we highly resolve that we shall have freedom from this nation – that devotion shall perish. These people honored the last government, in vain. The dead increased. Measure that full devotion! The earth under here gave these people birth, not a dead God, and from that they shall take their new cause, for which people have not died."

common set of modules

author's intention

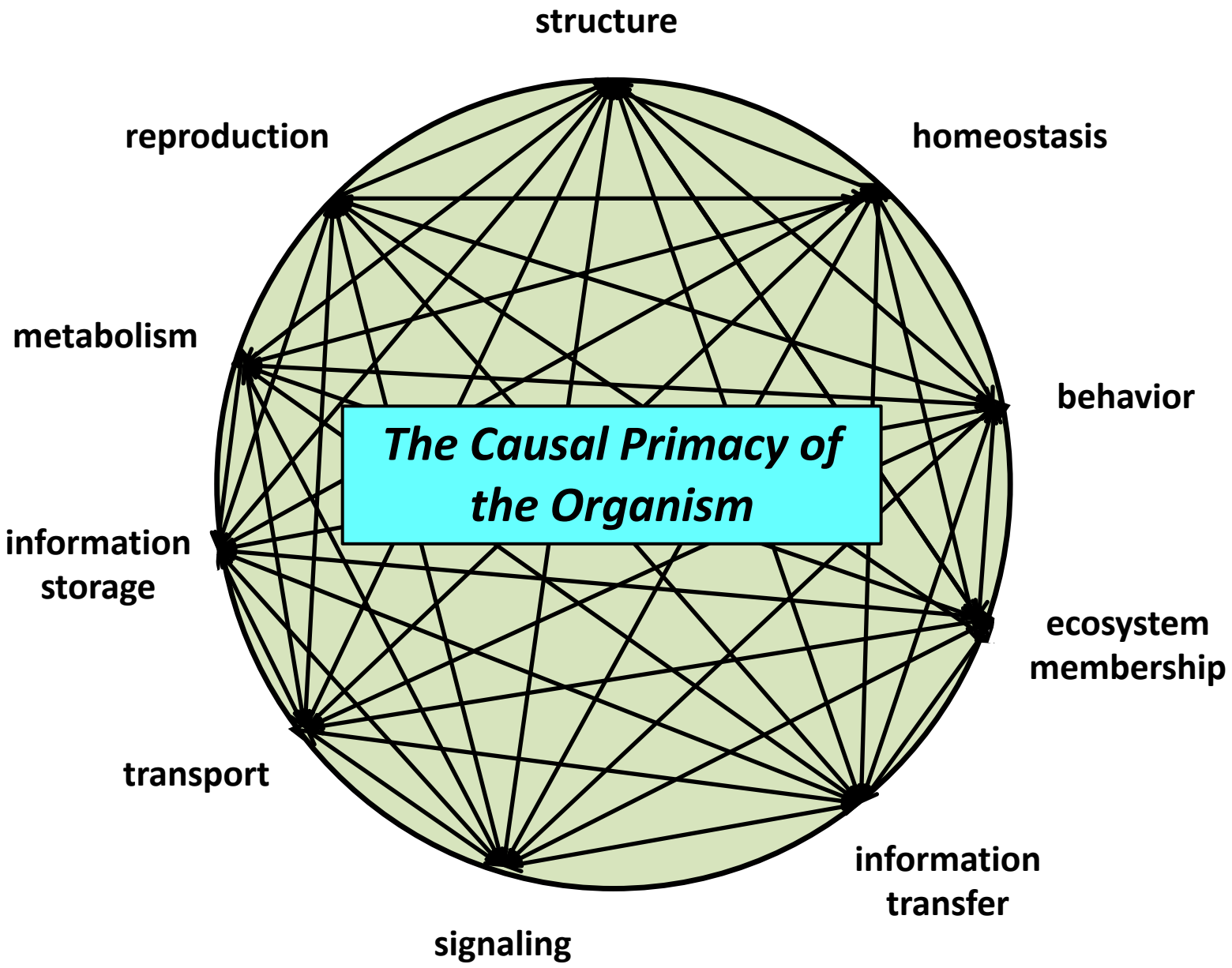
meaning (function)

lexicon



+

-


structure
reproduction
homeostasis
metabolism
behavior
The Causal Primacy of the Organism
information storage
ecosystem membership
transport
information transfer
signaling

Peter Tompa
VIB, Belgium

George Rose
Johns Hopkins

The Levinthal paradox of the interactome

Peter Tompa[1]* and George D. Rose[2]

[1]VIB Department of Structural Biology, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[2]Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland MD 21218

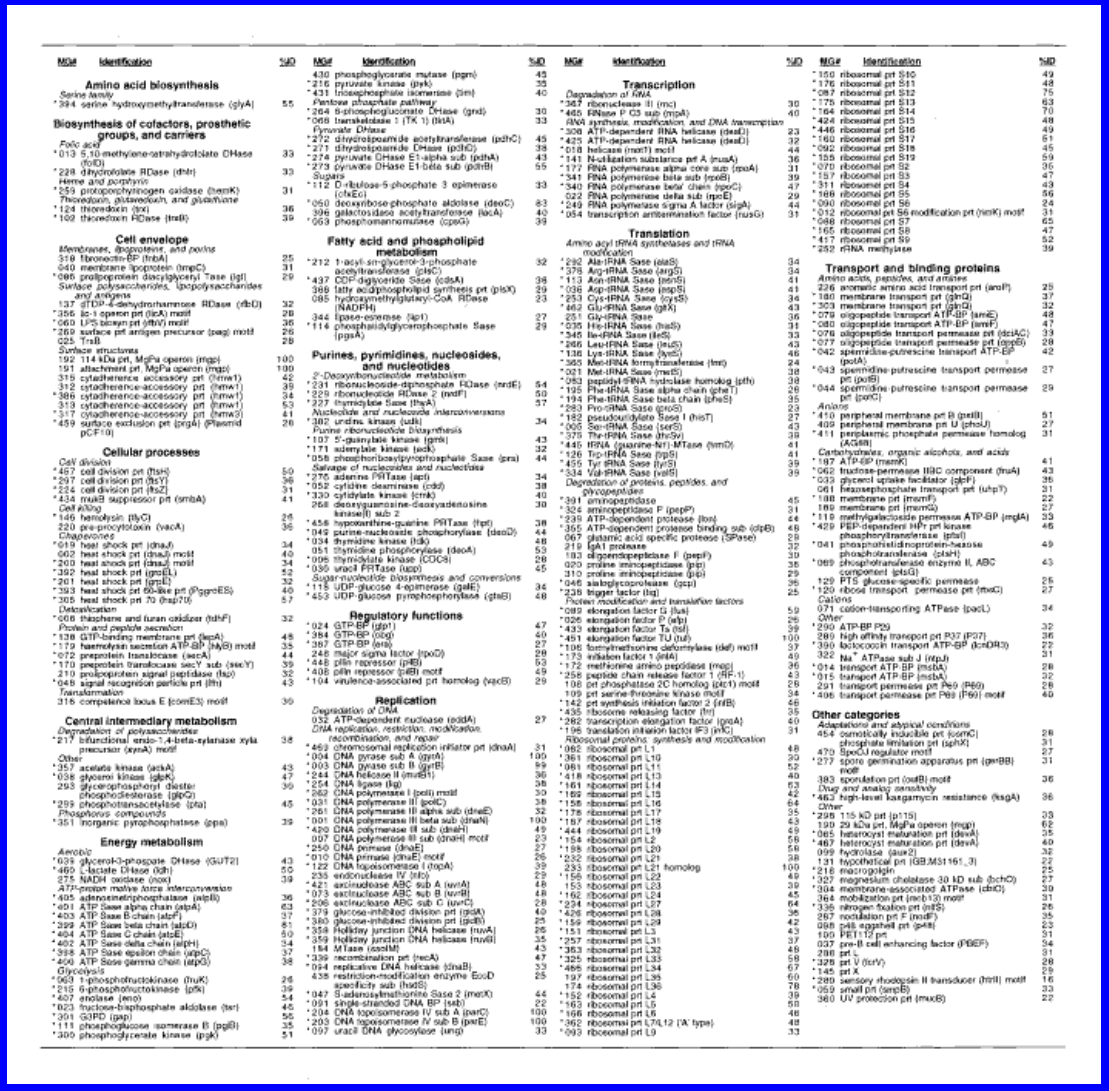
Received 6 September 2011; Revised 22 September 2011; Accepted 23 September 2011
DOI: 10.1002/pro.747
Published online 10 October 2011 proteinscience.org

Abstract: The central biological question of the 21st century is: how does a viable cell emerge from the bewildering combinatorial complexity of its molecular components? Here, we estimate the combinatorics of self-assembling the protein constituents of a yeast cell, a number so vast that the functional interactome could only have emerged by iterative hierarchic assembly of its component sub-assemblies. A protein can undergo both reversible denaturation and hierarchic self-assembly spontaneously, but a functioning interactome must expend energy to achieve viability. Consequently, it is implausible that a completely "denatured" cell could be reversibly renatured spontaneously, like a protein. Instead, new cells are generated by the division of pre-existing cells, an unbroken chain of renewal tracking back through contingent conditions and evolving responses to the origin of life on the prebiotic earth. We surmise that this non-deterministic temporal continuum could not be reconstructed *de novo* under present conditions.

PROTEIN SCIENCE 2011 VOL 20:2074—2079

This paper represents the shadow of a science yet to be born.

The minimal gene complement ("parts list") of *Mycoplasma genitalium* (Fraser *et al.*, 1995)



MGnl	Identification	%ID
	Amino acid biosynthesis	
	Serine family	
*394	serine hydroxymethyltransferase (glyA)	55
	Biosynthesis of cofactors, prosthetic groups, and carriers	
	Folic acid	
*013	5,10-methylene-tetrahydrofolate DHase (folD)	33
*228	dihydrofolate RDase (dhfr)	33
	Heme and porphyrin	
*255	protoporphyrinogen oxidase (hemK)	31
	Thioredoxin, glutaredoxin, and glutathione	
*124	thioredoxin (trx)	36
*102	thioredoxin RDase (trxB)	39
	Cell envelope	
	Membranes, lipoproteins, and porins	
519	fibronectin-BP (fnbA)	25
040	membrane lipoprotein (mpC)	31
*086	prolipoprotein diacylglyceryl Tase (lgt)	29
	Surface polysaccharides, lipopolysaccharides and antigens	
137	dTDP-4-dehydrorhamnose RDase (rfbD)	32
*356	lic-1 operon prt (licA) motif	26
*060	LPS biosyn prt (rfbV) motif	36
*269	surface prt antigen precursor (pag) motif	26
025	TraB	28
	Surface structures	
192	114 kDa prt, MgPa operon (mgp)	100
191	attachment prt, MgPa operon (mgp)	100
315	cytadherence accessory prt (hmw1)	42
312	cytadherence-accessory prt (hmw1)	35
*386	cytadherence-accessory prt (hmw1)	34
313	cytadherence-accessory prt (hmw1)	53
*317	cytadherence-accessory prt (hmw3)	41
*459	surface exclusion prt (prgA) (Plasmid pCF10)	26
	Cellular processes	
	Cell division	
*467	cell division prt (ftsH)	50
*297	cell division prt (ftsY)	36
*224	cell division prt (ftsZ)	31
*434	mukB suppressor prt (smbA)	41
	Cell killing	
*146	hemolysin (tlyC)	26
220	pre-procytotoxin (vacA)	36
	Chaperones	
*019	heat shock prt (dnaJ)	34
002	heat shock prt (dnaJ) motif	40
*200	heat shock prt (dnaJ) motif	34
*392	heat shock prt (groEL)	52
*201	heat shock prt (groES)	32
*393	heat shock prt 60-like prt (PggroES)	40
*305	heat shock prt 70 (hsp70)	57
	Detoxification	
*008	thiophene and furan oxidizer (thfF)	32
	Protein and peptide secretion	
*138	GTP-binding membrane prt (lepA)	48
*179	haemolysin secretion ATP-BP (hlyB) motif	35
*072	preprotein translocase (secA)	44
*170	preprotein translocase secY sub (secY)	39
210	prolipoprotein signal peptidase (lsp)	32
*049	signal recognition particle prt (ffh)	43
	Transformation	
318	competence locus E (comE3) motif	36
	Central intermediary metabolism	
	Degradation of polysaccharides	
*217	bifunctional endo-1,4-beta-xylanase xyla precursor (xynA) motif	38
	Other	
*357	acetate kinase (ackA)	43
*058	glycerol kinase (glpK)	47
293	glycerophosphoryl diester phosphodiesterase (glpQ)	36
*289	phosphotransacetylase (pta)	45
	Phosphorus compounds	
*351	inorganic pyrophosphatase (ppa)	39
	Energy metabolism	
	Aerobic	
*039	glycerol-3-phospate DHase (GUT2)	43
*460	L-lactate DHase (ldh)	50
275	NADH oxidase (nox)	39
	ATP-proton motive force interconversion	
*405	adenosinetriphosphatase (atpB)	36
401	ATP Sase alpha chain (atpA)	63
*403	ATP Sase B chain (atpF)	37
*399	ATP Sase beta chain (atpD)	61
*404	ATP Sase C chain (atpE)	50
*402	ATP Sase delta chain (atpH)	34
*398	ATP Sase epsilon chain (atpC)	37
*400	ATP Sase gamma chain (atpG)	38
	Glycolysis	
*069	1-phosphofructokinase (fruK)	26
*215	6-phosphofructokinase (pfk)	39
*407	enolase (eno)	54
*023	fructose-bisphosphate aldolase (tsr)	45
*301	G3PD (gap)	55
*111	phosphoglucose isomerase B (pgiB)	35
*305	phosphoglycerate kinase (pgk)	51
430	phosphoglycerate mutase (pgm)	45
*216	pyruvate kinase (pyk)	35
*431	triosephosphate isomerase (tim)	40
	Pentose phosphate pathway	
*264	6-phosphogluconate DHase (gnd)	30
*068	transketolase 1 (TK 1) (tktA)	33
	Pyruvate DHase	
*272	dihydrolipoamide acetyltransferase (pdhC)	45
*271	dihydrolipoamide DHase (pdhD)	38
*274	pyruvate DHase E1-alpha sub (pdhA)	43
*273	pyruvate DHase E1-beta sub (pdhB)	55
	Sugars	
*112	D-ribulose-5-phosphate 3 epimerase (cfxE)	33
022	RNA polymerase delta sub (rpsE)	29
*050	deoxyribose-phosphate aldolase (deoC)	83
396	galactosidase acetyltransferase (lacA)	40
*063	phosphomannomutase (cpeG)	39
	Fatty acid and phospholipid metabolism	
*212	1-acyl-sn-glycerol-3-phosphate acetyltransferase (plsC)	32
*437	CDP-diglyceride Sase (cdsA)	38
389	fatty acid/phospholipid synthesis prt (plsX)	29
085	hydroxymethylglutaryl-CoA RDase (NADPH)	23
344	lipase-esterase (lip1)	27
*114	phosphatidylglycerophosphate Sase (pgsA)	29
	Purines, pyrimidines, nucleosides, and nucleotides	
	2'-Deoxyribonucleotide metabolism	
*231	ribonucleoside-diphosphate RDase (nrdE)	54
*229	ribonucleotide RDase 2 (nrdF)	50
*227	thymidylate Sase (thyA)	57
	Nucleotide and nucleoside interconversions	
*382	uridine kinase (udk)	34
	Purine ribonucleotide biosynthesis	
*107	5'-guanylate kinase (gmk)	43
*171	adenylate kinase (adk)	32
*058	phosphoribosylpyrophosphate Sase (prs)	44
	Salvage of nucleotides and nucleotides	
*Val-tRNA	PRTase (apt)	34
*052	cytidine deaminase (cdd)	38
*330	cytidylate kinase (cmk)	40
268	deoxyguanosine-deoxyadenosine kinase(I) sub 2	30
*458	hypoxanthine-guanine PRTase (hpt)	38
*049	purine-nucleoside phosphorylase (deoD)	44
*034	thymidine kinase (tdk)	48
051	thymidine phosphorylase (deoA)	53
*095	thymidylate kinase (COG8)	28
*036	uracil PRTase (upp)	45
	Sugar-nucleotide biosynthesis and conversions	
*115	UDP-glucose 4-epimerase (galE)	34
*453	UDP-glucose pyrophosphorylase (galU)	48
	Regulatory functions	
*024	GTP-BP (gbp1)	47
*384	GTP-BP (obg)	48
*397	GTP-BP (era)	27
248	major sigma factor (rpoD)	28
*448	pilin repressor (pilB)	49
*408	pilin repressor (pilB) motif	49
*104	virulence-associated prt homolog (vacB)	29
	Replication	
	Degradation of DNA	
932	ATP-dependent nuclease (addA)	27
	DNA replication, restriction, modification, recombination, and repair	
*469	chromosomal replication initiator prt (dnaA)	31
*004	DNA gyrase sub A (gyrA)	100
*003	DNA gyrase sub B (gyrB)	99
*244	DNA helicase II (uvrB/H)	36
*254	DNA ligase (lig)	38
*262	DNA polymerase I (polI) motif	30
*031	DNA polymerase III (polC)	35
*261	DNA polymerase III alpha sub (dnaE)	32
*001	DNA polymerase III beta sub (dnaN)	100
*420	DNA polymerase III sub (dnaH)	49
007	DNA polymerase III sub (dnaH) motif	23
*250	DNA primase (dnaE)	27
*010	DNA primase (dnaE) motif	26
*122	DNA topoisomerase I (topA)	39
235	endonuclease IV (nfo)	29
*421	exonuclease ABC sub A (uvrA)	48
*073	exonuclease ABC sub B (uvrB)	48
*206	exonuclease ABC sub C (uvrC)	28
*379	glucose-inhibited division prt (gidA)	63
*380	glucose-inhibited division prt (gidB)	25
*358	Holliday junction DNA helicase (ruvA)	26
*359	Holliday junction DNA helicase (ruvB)	35
194	MTase (ssoIM)	43
*339	recombination prt (recA)	47
*094	replicative DNA helicase (dnaB)	33
438	restriction-modification enzyme EcoD specificity sub (hsdS)	25
*047	S-adenosylmethionine Sase 2 (metX)	44
*091	single-stranded DNA BP (ssb)	22
*204	DNA topoisomerase IV sub A (parC)	100
*203	DNA topoisomerase IV sub B (parE)	100
*097	uracil DNA glycosylase (ung)	33
	Transcription	
	Degradation of RNA	
*367	ribonuclease III (mc)	30
*465	RNase P G5 sub (rnpA)	40
	RNA synthesis, modification, and DNA transcription	
*308	ATP-dependent RNA helicase (deaD)	23
*425	ATP-dependent RNA helicase (deaD)	32
*018	helicase (motT) motif	44
*141	N-utilization substance prt A (nusA)	36
*177	RNA polymerase alpha core sub (rpoA)	31
*341	RNA polymerase beta sub (rpoB)	39
*340	RNA polymerase beta' chain (rpoC)	47
022	RNA polymerase delta sub (rpsE)	29
*249	RNA polymerase sigma A factor (sigA)	44
*054	transcription antitermination factor (nusG)	31
	Translation	
	Amino acyl tRNA synthetases and tRNA modification	
*292	Ala-tRNA Sase (alaS)	34
*378	Arg-tRNA Sase (argS)	34
*113	Asn-tRNA Sase (asnS)	41
*036	Asp-tRNA Sase (aspS)	41
*253	Cys-tRNA Sase (cysS)	34
*462	Glu-tRNA Sase (gltX)	43
251	Gly-tRNA Sase	36
*235	His-tRNA Sase (hisS)	31
*345	Ile-tRNA Sase (ileS)	33
*266	Leu-tRNA Sase (leuS)	43
*136	Lys-tRNA Sase (lysS)	46
*365	Met-tRNA formyltransferase (fmt)	24
*021	Met-tRNA Sase (metS)	38
*282	peptidyl-tRNA hydrolase homolog (pth)	38
*195	Phe-tRNA Sase alpha chain (pheT)	28
*194	Phe-tRNA Sase beta chain (pheS)	35
*283	Pro-tRNA Sase (proS)	23
*182	pseudouridylate Sase I (hisT)	27
*405	Ser-tRNA Sase (serS)	43
*375	Thr-tRNA Sase (thrS)	38
*445	tRNA (guanine-N1)-MTase (trmD)	41
*126	Trp-tRNA Sase (trpS)	41
*455	Tyr-tRNA Sase (tyrS)	39
*334	Val-tRNA Sase (valS)	39
	Degradation of proteins, peptides, and glycopeptides	
*391	aminopeptidase	45
*324	aminopeptidase P (pepP)	31
*239	ATP-dependent protease (lon)	44
*365	ATP-dependent protease binding sub (clpB)	48
067	glutamic acid specific protease (SPase)	28
219	IgA1 protease	32
183	oligopeptidase F (pepF)	30
920	proline iminopeptidase (pip)	29
310	proline iminopeptidase (pip)	29
*046	sialoglycoprotease (gcp)	36
*238	trigger factor (tig)	25
	Protein modification and translation factors	
*089	elongation factor G (fusA)	59
*026	elongation factor P (efp)	26
*433	elongation factor Ts (tsf)	39
*451	elongation factor TU (tuf)	100
*196	formylmethionine deformylase (def) motif	37
*173	initiation factor 1 (infA)	49
*172	methionine amino peptidase (map)	36
*256	peptide chain release factor 1 (RF-1)	43
108	prt phosphatase 2C homolog (ptc1) motif	28
109	prt serine-threonine kinase motif	34
*142	prt synthesis initiation factor 2 (infB)	46
*435	ribosome releasing factor (frr)	35
*262	transcription elongation factor (greA)	40
*196	translation initiation factor IF2 (infC)	31
	Ribosomal proteins: synthesis and modification	
*082	ribosomal prt L1	48
*361	ribosomal prt L10	30
*081	ribosomal prt L11	52
*418	ribosomal prt L13	49
*161	ribosomal prt L14	53
*159	ribosomal prt L15	42
*158	ribosomal prt L16	64
*178	ribosomal prt L17	35
*197	ribosomal prt L18	43
*444	ribosomal prt L19	49
*154	ribosomal prt L2	56
*198	ribosomal prt L20	56
*232	ribosomal prt L21	38
233	ribosomal prt L21 homolog	100
*156	ribosomal prt L22	49
*153	ribosomal prt L23	39
*162	ribosomal prt L24	45
*234	ribosomal prt L27	64
*426	ribosomal prt L28	35
*081	ribosomal prt L29	23
*151	ribosomal prt L3	43
*250	ribosomal prt L31	37
*363	ribosomal prt L32	46
*325	ribosomal prt L33	56
*466	ribosomal prt L34	67
197	ribosomal prt L35	60
174	ribosomal prt L36	78
*152	ribosomal prt L4	39
*163	ribosomal prt L5	58
*166	ribosomal prt L6	48
*362	ribosomal prt L7/L12 ('A' type)	48
*093	ribosomal prt L9	33
*150	ribosomal prt S10	49
*176	ribosomal prt S11	48
*087	ribosomal prt S12	75
*175	ribosomal prt S13	63
*164	ribosomal prt S14	70
*424	ribosomal prt S15	48
*446	ribosomal prt S16	49
*160	ribosomal prt S17	61
*092	ribosomal prt S18	45
*155	ribosomal prt S19	59
*070	ribosomal prt S2	35
*157	ribosomal prt S3	47
*311	ribosomal prt S4	43
*168	ribosomal prt S5	56
*090	ribosomal prt S6	24
*012	ribosomal prt S6 modification prt (himK) motif	31
*088	ribosomal prt S7	65
*165	ribosomal prt S8	47
*417	ribosomal prt S9	52
*252	rRNA methylase	39
	Transport and binding proteins	
	Amino acids, peptides, and amines	
226	aromatic amino acid transport prt (aroP)	25
*180	membrane transport prt (glnQ)	37
*303	membrane transport prt (glnQ)	32
*079	oligopeptide transport ATP-BP (amiE)	48
*080	oligopeptide transport ATP-BP (amiF)	47
*076	oligopeptide transport permease prt (dciAC)	33
*077	oligopeptide transport permease prt (oppB)	28
*042	spermidine-putrescine transport ATP-BP (potA)	42
*043	spermidine-putrescine transport permease prt (potB)	27
*044	spermidine-putrescine transport permease prt (potC)	29
	Anions	
*450	peripheral membrane prt B (pstB)	51
409	peripheral membrane prt U (phoU)	27
*411	periplasmic phosphate permease homolog (AG68)	31
	Carbohydrates, organic alcohols, and acids	
*187	ATP-BP (msmK)	41
*062	fructose-permease IIBC component (fruA)	43
*033	glycerol uptake facilitator (glpF)	36
061	hexosephosphate transport prt (uhpT)	31
*188	membrane prt (msmF)	22
189	membrane prt (msmG)	31
*119	methylgalactoside permease ATP-BP (mglA)	33
*429	PEP-dependent HPr prt kinase phosphorylransferase (ptsI)	46
*041	phosphatidylinoprotein-hexose phosphotransferase (ptsH)	49
*069	phosphofructokinase enzyme II, ABC component (ptsG)	43
129	PTS glucose-specific permease	25
*120	ribose transport permease prt (rbsC)	27
	Cations	
071	cation-transporting ATPase (pacL)	34
	Other	
*290	ATP-BP P29	32
289	high-affinity transport prt P37 (P37)	36
*390	lactococcin transport ATP-BP (lcnDR3)	22
322	Na^+ ATPase sub J (ntpJ)	31
*014	transport ATP-BP (msbA)	28
*015	transport ATP-BP (msbA)	32
291	transport permease prt P69 (P69)	28
*496	transport permease prt P69 (P69) motif	40
	Other categories	
	Adaptations and atypical conditions	
454	osmotically inducible prt (osmC)	28
	phosphate limitation prt (sphX)	31
479	SpoOJ regulator motif	27
*277	spore germination apparatus prt (gerBB) motif	31
383	sporulation prt (cotB) motif	36
	Drug and analog sensitivity	
*463	high-level kasgamycin resistance (ksgA)	36
	Other	
*295	115 kD prt (p115)	33
199	29 kDa prt, MgPa operon (mgp)	62
*065	heterocyst maturation prt (devA)	25
*467	heterocyst maturation prt (devA)	40
099	hydrolase (aue2)	32
131	hypothetical prt (GB:M31161_3)	22
*216	macrogolgin	25
*327	magnesium chelatase 38 kD sub (bchO)	27
*304	membrane-associated ATPase (chlC)	39
364	mobilization prt (mcb13) motif	31
*336	nitrogen fixation prt (nifS)	26
367	nodulation prt F (nodF)	35
088	p48 eggshell prt (p48)	23
100	PE112 prt	31
037	pre-B cell enhancing factor (PBEF)	34
286	prt	31
*326	prt V (tcrV)	28
*145	prt X	29
*289	sensory rhodopsin II transducer (htrII) motif	16
*059	small prt (smpB)	33
360	UV protection prt (muxB)	22

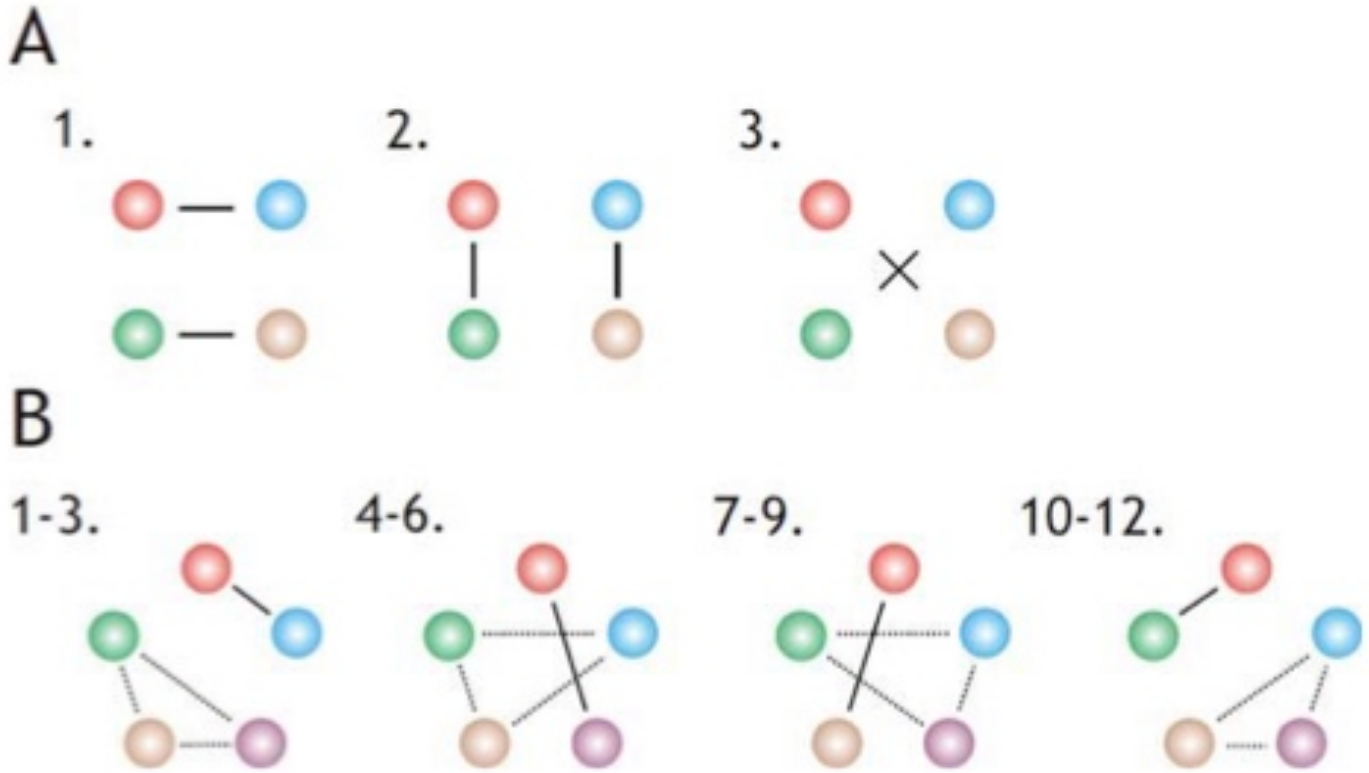


Figure 1. The number of possible interactomes increases exponentially with proteome size. The number of possible different states (patterns of pairwise interactions) of the interactome increases exponentially with the number of its constituent proteins.

Elements	Possible Pairwise Interactions
2	1
3	2
4	3
5	11
6	15
7	74
8	105
9	668
10	945
11	7,350
12	10,395
13	95,555
14	135,135
15	1,433,323
16	2,027,025
17	24,366,493
18	34,459,425

19	462,963,369
20	654,729,075
21	9,722,230,744
22	13,749,310,575
23	223,611,307,117
24	316,234,143,225
25	5,590,282,677,928
26	7,905,853,580,625
27	150,937,632,304,053
28	213,458,046,676,875
29	4,377,191,336,817,530
30	6,190,283,353,629,380
31	135,692,931,441,344,000
32	191,898,783,962,511,000
33	4,477,866,737,564,340,000
34	6,332,659,870,762,850,000
35	156,725,335,814,752,000,000
36	221,643,095,476,700,000,000
37	5,798,837,425,145,820,000,000
38	8,200,794,532,637,890,000,000
39	226,154,659,580,687,000,000,000
40	319,830,986,772,878,000,000,000
41	9,272,341,042,808,170,000,000,000
42	13,113,070,457,688,000,000,000,000
43	398,710,664,840,751,000,000,000,000
44	563,862,029,680,584,000,000,000,000
45	17,941,979,917,833,800,000,000,000,000
46	25,373,791,335,626,300,000,000,000,000
47	843,273,056,138,187,000,000,000,000,000
48	1,192,568,192,774,430,000,000,000,000,000
49	41,320,379,750,771,300,000,000,000,000,000

50	58,435,841,445,947,300,000,000,000,000,000
51	2,107,339,367,289,340,000,000,000,000,000,000
52	2,980,227,913,743,310,000,000,000,000,000,000
53	111,688,986,466,335,000,000,000,000,000,000,000
54	157,952,079,428,395,000,000,000,000,000,000,000
55	6,142,894,255,648,410,000,000,000,000,000,000,000
56	8,687,364,368,561,750,000,000,000,000,000,000,000
57	350,144,972,571,959,000,000,000,000,000,000,000,000
58	495,179,769,008,020,000,000,000,000,000,000,000,000
59	20,658,553,381,745,600,000,000,000,000,000,000,000,000
60	29,215,606,371,473,200,000,000,000,000,000,000,000,000
61	1,260,171,756,286,480,000,000,000,000,000,000,000,000,000
62	1,782,151,988,659,860,000,000,000,000,000,000,000,000,000
63	79,390,820,646,048,400,000,000,000,000,000,000,000,000,000
64	112,275,575,285,571,000,000,000,000,000,000,000,000,000,000
65	5,160,403,341,993,140,000,000,000,000,000,000,000,000,000,000
66	7,297,912,393,562,140,000,000,000,000,000,000,000,000,000,000
67	345,747,023,913,541,000,000,000,000,000,000,000,000,000,000,000
68	488,960,130,368,663,000,000,000,000,000,000,000,000,000,000,000
69	23,856,544,650,034,300,000,000,000,000,000,000,000,000,000,000
70	33,738,248,995,437,800,000,000,000,000,000,000,000,000,000,000
71	1,693,814,670,152,440,000,000,000,000,000,000,000,000,000,000,000
72	2,395,415,678,676,080,000,000,000,000,000,000,000,000,000,000,000
73	123,648,470,921,128,000,000,000,000,000,000,000,000,000,000,000,000
74	174,865,344,543,354,000,000,000,000,000,000,000,000,000,000,000,000
75	9,273,635,319,084,590,000,000,000,000,000,000,000,000,000,000,000,000
76	13,114,900,840,751,600,000,000,000,000,000,000,000,000,000,000,000,000
77	714,069,919,569,512,000,000,000,000,000,000,000,000,000,000,000,000,000
78	1,009,847,364,737,870,000,000,000,000,000,000,000,000,000,000,000,000,000
79	56,411,523,645,991,500,000,000,000,000,000,000,000,000,000,000,000,000,000
80	79,777,941,814,291,600,000,000,000,000,000,000,000,000,000,000,000,000,000

81	4,569,333,415,325,320,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
82	6,462,013,286,957,630,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
83	379,254,673,472,001,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
84	536,347,102,817,483,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
85	32,236,647,245,120,100,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
86	45,589,503,739,486,100,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
87	2,804,588,310,325,440,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
88	3,966,286,825,335,280,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
89	249,608,359,618,965,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
90	352,999,527,454,840,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
91	22,714,360,725,325,800,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
92	32,122,956,998,390,500,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
93	2,112,435,547,455,290,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
94	2,987,435,000,850,310,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
95	200,681,377,008,254,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
96	283,806,325,080,780,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
97	19,466,093,569,800,600,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
98	27,529,213,532,835,600,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
99	1,927,143,263,410,250,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000
100	2,725,392,139,750,730,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000

Bill Dembski stopped his calculations here,
at 100 proteins, even though real cells have at least
300 proteins. There was no point in going any further.

Nature is talking back to biologists. She insists on being heard.

PLOS ONE

RESEARCH ARTICLE

On an algorithmic definition for the components of the minimal cell

Octavio Martínez[1]*, M. Humberto Reyes-Valdés[2]

1 Unidad de Genómica Avanzada, Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO), Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, México, **2** Graduate Program on Plant Genetic Resources for Arid Lands, Universidad Autónoma Agraria Antonio Narro, Saltillo, Coahuila, México

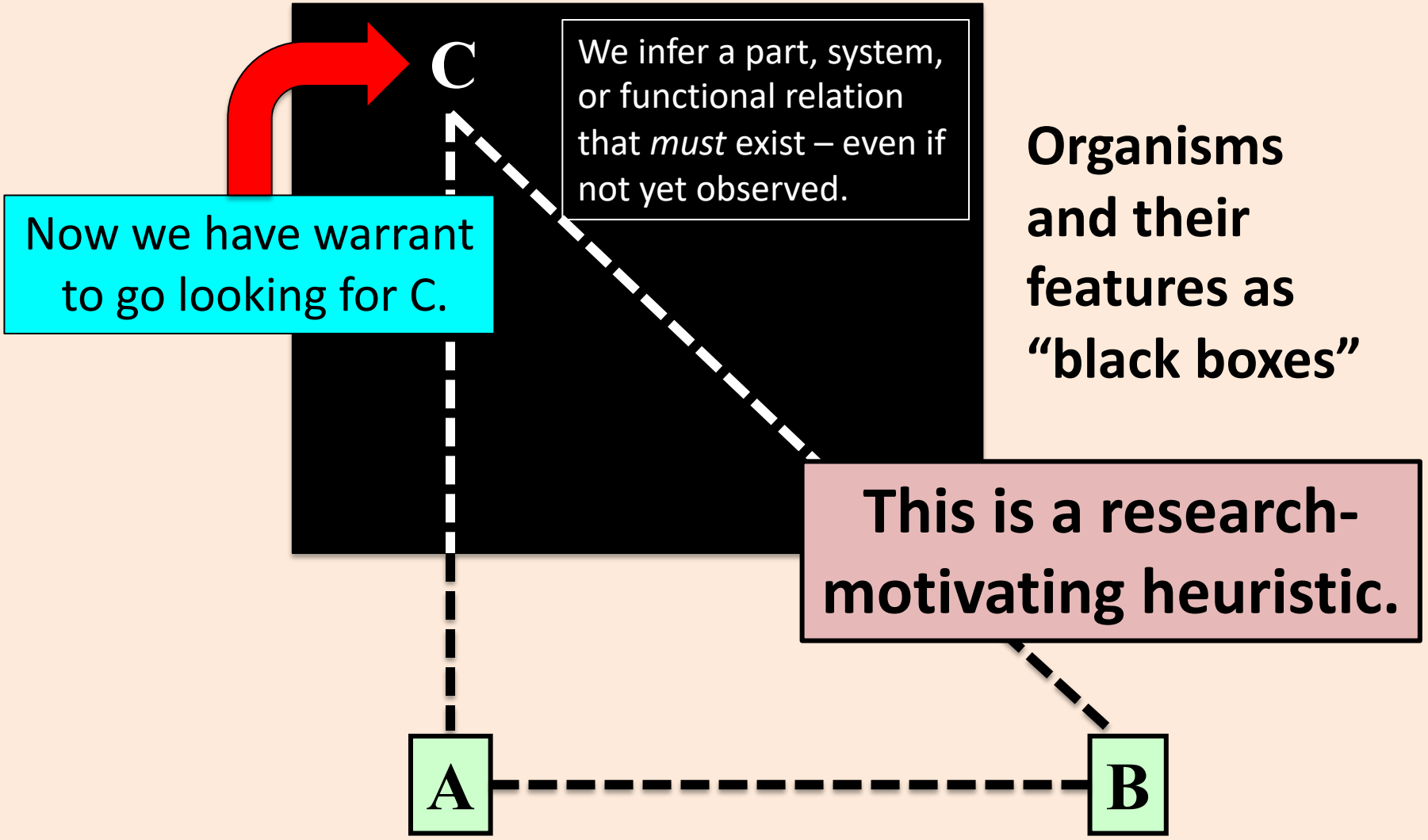
* octavio.martinez@cinvestav.mx

Nature is talking back to biologists. She insists on being heard.

Abstract

Living cells are highly complex systems comprising a multitude of elements that are engaged in the many convoluted processes observed during the cell cycle. However, not all elements and processes are essential for cell survival and reproduction under steady-state environmental conditions. To distinguish between essential from expendable cell components and thus define the 'minimal cell' and the corresponding 'minimal genome', we postulate that the synthesis of all cell elements can be represented as a finite set of binary operators, and within this framework we show that cell elements that depend on their previous existence to be synthesized are those that are essential for cell survival. An algorithm to distinguish essential cell elements is presented and demonstrated within an interactome. Data and functions implementing the algorithm are given as supporting information. We expect that this algorithmic approach will lead to the determination of the complete interactome of the minimal cell, which could then be experimentally validated. The assumptions behind this hypothesis as well as its consequences for experimental and theoretical biology are discussed.

What metaphysics of explanation enables one to start with the whole organism?



Design.

CODA:

I love magic. Mozart's Flute and Harp Concerto, the Stargate sequence in Kubrick's *2001*, Orwell's essays, the Narnia stories, Hopper's *Manhattan Bridge Loop*, Spanish red wine, the entirety of *Dark Side of the Moon*, my cats Newton and Einstein, Suzanne's jokes, Toews's 10/19/07 goal against Colorado (look it up on YouTube), mushroom tarts, a long walk by myself in the early spring...a hug from a grateful friend.

If "magic" means "this thing is too wonderful for words," then magic makes life worth living.

The magic fitting for biological understanding is *knowledge of mechanisms*, which has a subtle and mysterious beauty all its own.