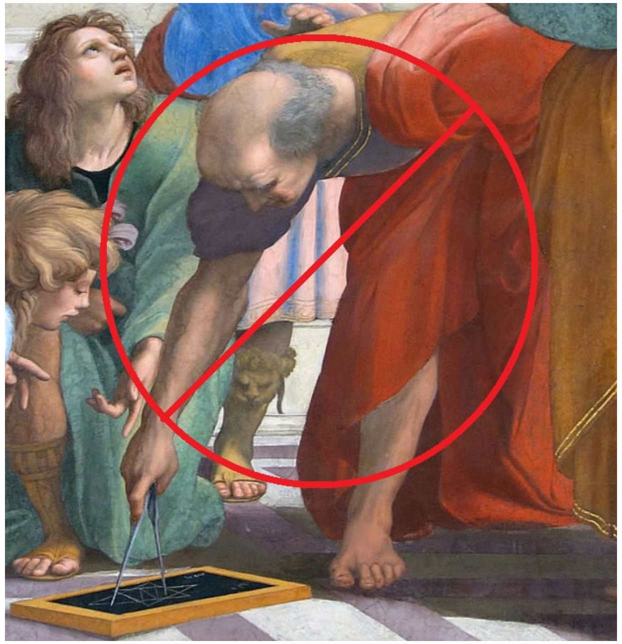
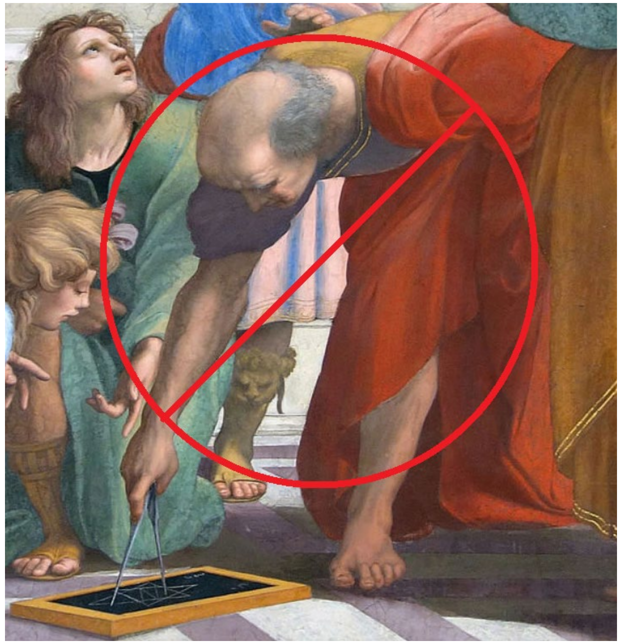
# The Archimedean trap: Why ~~traditional~~ *real-number-reward* reinforcement learning will probably not yield AGI




Samuel Alexander

U.S. Securities and Exchange Commission

samuelallenalexander@gmail.com

# Thesis

- The real numbers, $\mathbb{R}$, are too inflexible for RL rewards if we want RL to lead us to AGI.
- There are environments, which an AGI should be able to understand, with rewards from other number systems, not convertible into environments with rewards from $\mathbb{R}$.

# What is a "scalar"??

Artificial Intelligence

Volume 299, October 2021, 103535

**Reward is enough**

David Silver, Satinder Singh, Doina Precup, Richard S. Sutton

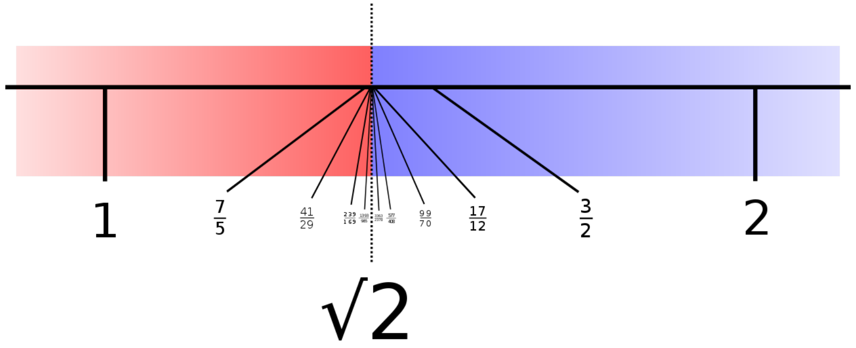
Silver et al (2021),

"Reward is enough":

> "A *reward* is a special scalar observation $R_t$, emitted at every time-step *t* by a *reward signal* in the environment..."

Silver et al conspicuously avoid committing to a specific number system (like $\mathbb{R}$ or $\mathbb{Q}$).

# What are the real numbers?




By definition, $\mathbb{R}$ is the unique (up to isomorphism) ordered field such that:

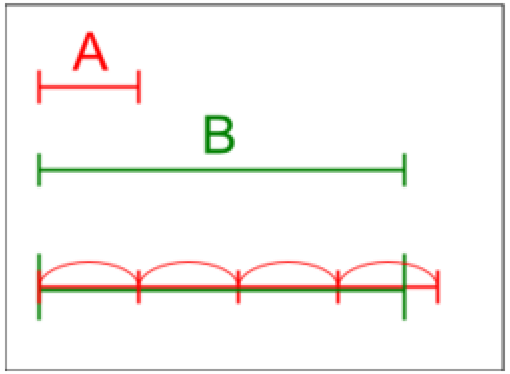
- (Completeness) For every subset S of $\mathbb{R}$, if S has an upper bound in $\mathbb{R}$, then S has a least upper bound in $\mathbb{R}$.

**Question**: Does anything about RL inherently imply that every set of rewards with an upper bound must have a reward least-upper-bound?

(If so, then let's get RL added to real analysis textbooks!!)

# Lemma (The Archimedean Property):

For every positive real number A, for every real number B, there exists a natural number n such that nA > B.

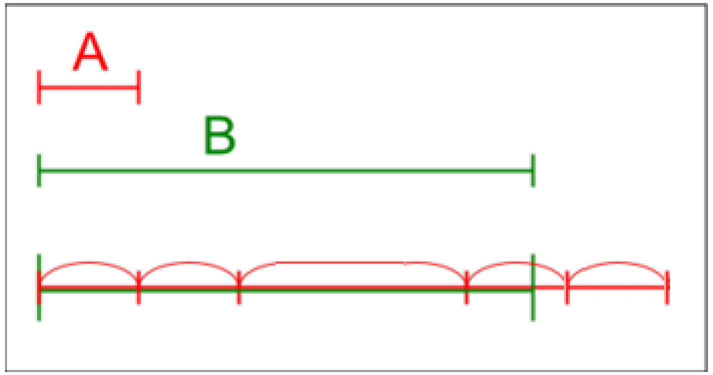


# Lemma (Generalized Archimedean Property):

- Suppose r is a positive real number.
- For all real x,y, say that x is *significantly less than* y if $x \leq y-r$.

Then:

- For every infinite sequence $x_0,x_1,x_2,\ldots$ of reals, if each $x_i$ is significantly less than $x_{i+1}$, then for every real y, there is some i such that y is significantly less than $x_i$.



# The Generalized Archimedean Property is More Adaptable

- The original Archimedean Property requires arithmetic. It requires the ability to multiply by natural number n.
- The generalized Archimedean Property does not require arithmetic. It can be adapted to any context where there's a notion of "significantly less than".

$3 \times$ 

$= ???$

# Example: "Is musical beauty Archimedean?"

That is:

If $x_0, x_1, x_2, \ldots$ are songs, each significantly less beautiful than the next, and if y is a song, is there necessarily some i such that y is significantly less beautiful than $x_i$?

This seems non-trivial. If "No", then real numbers are inadequate for measuring musical beauty (by a proposition on a later slide).

Need generalized Archimedean property: a priori, we can't multiply a musical beauty by n.

# A Musical RL Environment

- Consider an RL environment where the agent plays piano.
- Each song, the agent receives that song's musical beauty as a reward.
- Can this environment be converted to one with rewards from $\mathbb{R}$?

Saying "Yes" would suggest musical beauty is Archimedean
(a non-trivial assertion).

# Example: Asymptotic Complexity

```
for x in range(n):
  for y in range(n):
    for z in range(n):
      F(x,y,z)
```

Presumably $O(n^k)$ is significantly less complex than $O(n^{k+1})$ for each k.

Presumably $O(2^n)$ is *not* significantly less complex than any $O(n^k)$.

So complexity can keep growing significantly without exhausting the full range of all complexities.

So complexity is non-Archimedean in the generalized sense.

# A Code-Generation Environment

- Each episode, the agent is given a problem and a length k.
- Using k keystrokes, agent must write a program, and a proof that the program solves the problem in time $O(f(n))$ for some $f(n)$ of the agent's choosing.
- Agent receives reward $-O(f(n))$ (or $-\infty$ for not submitting a program and proof).

Non-Archimedeanness of complexity implies this environment cannot be converted to one with rewards from $\mathbb{R}$.

# What if we try to convert complexity to $\mathbb{R}$?

In previous example, suppose the agent also gets +1 reward for using fewer keystrokes than allotted.

Say we convert $O(2^n)$ into real number 100.

We're forced to convert $O(n)$, $O(n^2)$, $O(n^3)$, ... into reals closer and closer (they're crammed between 0 and 100).

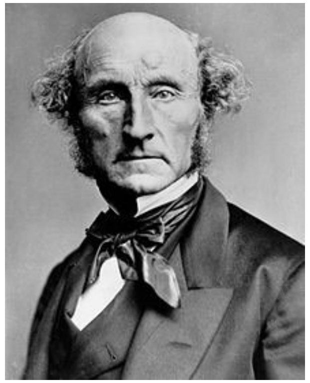
E.g., say $O(n^{100})$ and $O(n^{200})$ are converted to 99.98 and 99.99.

The agent would be *deceived* and would think a $O(n^{200})$ program is more rewarding than a slightly longer $O(n^{100})$ program!

# Example: Ethical Utility

"If one of [two pleasures] is ... placed so far above the other that they ... would not resign it for any quantity of the other pleasure which their nature is capable of, we are justified in ascribing to the preferred enjoyment a superiority in quality, so far outweighing quantity as to render it, in comparison, of small account"

---John Stuart Mill, *Utilitarianism* (1863)



This suggests said "pleasures" are non-Archimedean, thus too complicated to convert to rewards in $\mathbb{R}$.

# Example: True Arithmetical Theories

There are true arithmetical theories $x_0, x_1, x_2, \ldots$, each $x_{i+1}$ proving $CON(x_i)$, and a true arithmetical theory y proving $CON\left(\bigcup_i x_i\right)$.

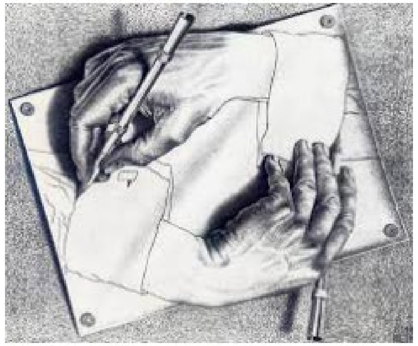
In a Gödelian sense, each $x_i$ is significantly weaker than $x_{i+1}$, yet y is not significantly weaker than any $x_i$.

$\mathbb{R}$ would be no good for incentivizing agents to invent theories.

Logicians measure such theories using *ordinal numbers*.

# Example: Intelligence Measurement

Silver et al ("Reward is enough", 2021) conjecture RL can lead to AGI.



In reply ("Can reinforcement learning learn itself?", 2022) I ask: can an RL environment incentivize RL agents to design good RL agents? (Apparently "yes" if RL can lead to AGI!)

Let's explore how this question relates to intelligence measurement...

# RL measures

Definition: An *RL measure* is a function f which takes as input (a source-code of) an RL agent $\pi$, and outputs a number $f(\pi)$.

The intention is that f should measure how good $\pi$ is.

# An RL-designing RL-environment

Definition of environment $M_f$ (for any RL measure f):

1. Pseudo-randomly generate a natural number k.
2. Prompt the agent to use exactly k actions to encode an RL agent child along with a proof that that child really is an RL agent.
3. Let r be the f-measure of the child that the agent designs in step 2 (or -1 if the agent does not prove the child is an RL agent).
4. Give the agent reward r.
5. Goto 1.

# Demo of $M_f$

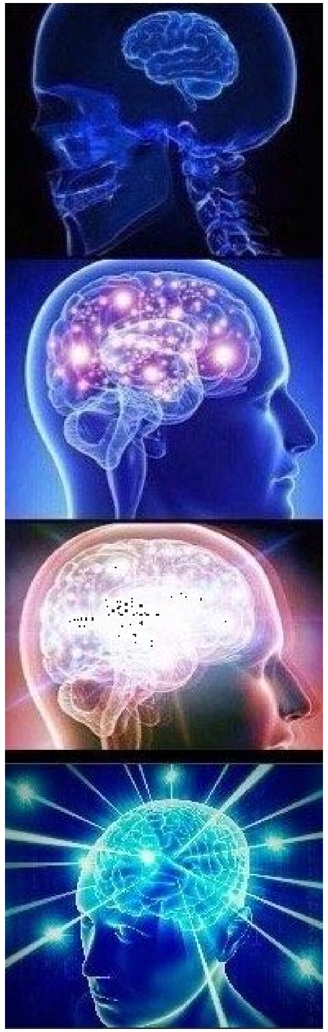
I will now demo a mockup of an environment $M_f$

# Is RL intelligence Archimedean?

If $x_0,x_1,x_2,...$ are RL agents, each one significantly less intelligent than the next, and $y$ is an RL agent, is it necessarily true that there is some $i$ such that $y$ is significantly less intelligent than $x_i$?

This seems non-trivial. If “No”, then real numbers are inadequate for measuring RL intelligence.

Need generalized Archimedean property: a priori, we can’t multiply an intelligence level by n.



# What about Legg-Hutter intelligence $\Upsilon$?

When I say,

"then real numbers are inadequate for measuring RL intelligence",

I mean,

"for the purpose of using RL intelligences as RL rewards."

If RL intelligence is non-Archimedean, $M_\Upsilon$ might fail to incentivize RL agents to design good RL agents. But $\Upsilon$ might still be a good RL intelligence measure for other purposes.

# Non-Archimedeanness in Plato

"The very day you start [as my student], you will go home a better man, and the same thing will happen the day after. Every day, day after day, you will get better and better."

---Protagoras (from Plato's *Protagoras*)

If Protagoras and his student live forever, and Protagoras' goodness stays constant, and the student never exceeds Protagoras, and if "better" means "significantly better"...

...then these goodnesses are non-Archimedean.

# Generalized Archimedean Structures

Definition:

- A *significantly ordered structure* is a collection X with an ordering << (pronounced: “much less than”).
- A significantly ordered structure (X,<<) is *Archimedean* if: for each infinite sequence $x_0 << x_1 << \ldots$ in X, for each y in X, there is some i such that $y << x_i$.

# Accurate real-valued measures

Definition: Suppose (X,<<) is a significantly-ordered structure. A function $f: X \to \mathbb{R}$ is said to *accurately measure* (X,<<) if there is some real r > 0 such that:

- For all $x_1$,$x_2$ in X, $x_1 << x_2$ iff $f(x_1) \leq f(x_2) - r$.

Proposition (Inadequacy of $\mathbb{R}$ for non-Archimedean structures): If (X,<<) is non-Archimedean, then no function $f: X \to \mathbb{R}$ accurately measures (X,<<).

Proposition (Inadequacy of $\mathbb{R}$ for non-Archimedean structures): If (X,<<) is non-Archimedean, then no function $f: X \to \mathbb{R}$ accurately measures (X,<<).

Proof: Assume $f: X \to \mathbb{R}$ accurately measures (X,<<).

- Thus there is some real $r > 0$ s.t. $x_1 << x_2$ iff $f(x_1) \leq f(x_2) - r$.
- Since (X,<<) is non-Archimedean, there exist $x_0 << x_1 << \ldots$ and y in X such that there is no i such that $y << x_i$.
- So each $f(x_i) \leq f(x_{i+1}) - r$, and there is no i such that $f(y) \leq f(x_i) - r$.

This contradicts the Generalized Archimedean Property.

# Consequences of the proposition

If (X,<<) is non-Archimedean and we attempt to map it to $\mathbb{R}$, while respecting <<, then there will be elements of X mapped arbitrarily close to each other even though one is significantly less than the other.

This suggests RL environments with rewards from such X cannot generally be converted to have rewards from $\mathbb{R}$ without agents being deceived (perceiving two near-equal rewards, when, in the original environment, one was significantly smaller than the other).

# Solutions?

1. Preference-based RL.
2. RL with rewards from a more flexible number system.

# Some extensions of $\mathbb{R}$

- Formal Laurent Series
- Hyperreals
- Surreals

# Formal Laurent Series

Formal Laurent Series are probably the most concrete, down-to-earth extension of $\mathbb{R}$.

They are non-Archimedean (when significantly-ordered in a natural way), so: maybe they're good enough for AGI?

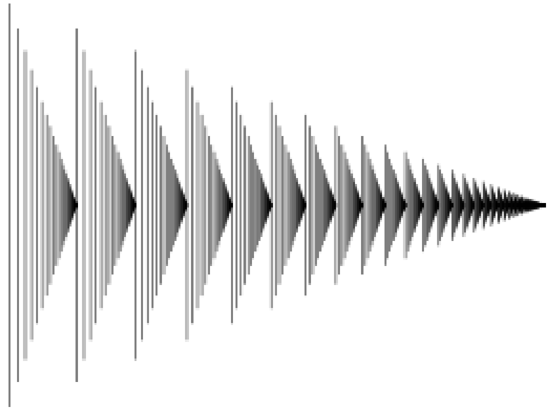
I'll sketch an argument suggesting "probably not".

# Semi-Archimedeanness

Definition: If (X,<<) is a significantly ordered structure, we define a new ordering <<' on X by declaring: x <<' y iff there are $x=x_0 << x_1 << ...$ such that each $x_i << y$.

Definition: (X,<<) is *semi-Archimedean* if:

- For all $x_0 <<' x_1 <<' ...$, for all y, there is some i such that $y << x_i$.



# Formal Laurent Series are semi-Archimedean

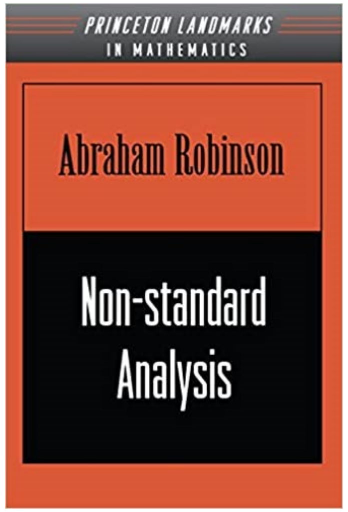
- Theorem: The Formal Laurent Series aren’t Archimedean, but they *are* semi-Archimedean.

Our reasoning can be modified, to argue Formal Laurent Series rewards are not enough for AGI. E.g....

Q: “Is musical beauty semi-Archimedean?” A: “Hard to say.”

Q: “Is big-O complexity semi-Archimedean?” A: “Definitely not!”

# Hyperreal numbers




- I almost independently discovered the hyperreals in my paper "Intelligence via Ultrafilters" (JAGI, 2019).

- The idea there: To compare RL agents $\pi$ and $\rho$, consider them to compete in an *election* where each environment casts a vote based on its rewards: $\mu$ votes for $\pi$ if $\pi$ outperforms $\rho$ in $\mu$.

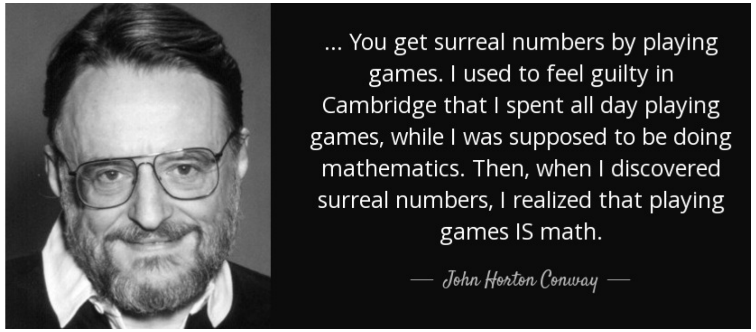
- This election idea comes close to the ultrafilter construction of the hyperreals.

# Surreal numbers

- Discovered by John Conway (of “Game of Life” fame)

The surreals are the most flexible number system possible without altering the foundations of mathematics. If *they* aren’t flexible enough for AGI rewards, nothing in mainstream 2022 mathematics is.

Surreals are *very* abstract. They could plausibly provide the fabled link between statistical and symbolic approaches to AI.



# Surreals as Statistico-Symbolic Bridge

The surreal numbers are *so* abstract, you’re basically *forced* to get symbolical just to work with them.

At the same time, they’re like a richer $\mathbb{R}$: hence, they’re statistical.

To illustrate how one is “forced to get symbolical”, I’d like to show you my *Library of Intuitive Ordinal Notations* (ordinals are to surreals just like $\mathbb{N}$ is to $\mathbb{R}$).

# Intuitive Ordinal Notations (IONs)

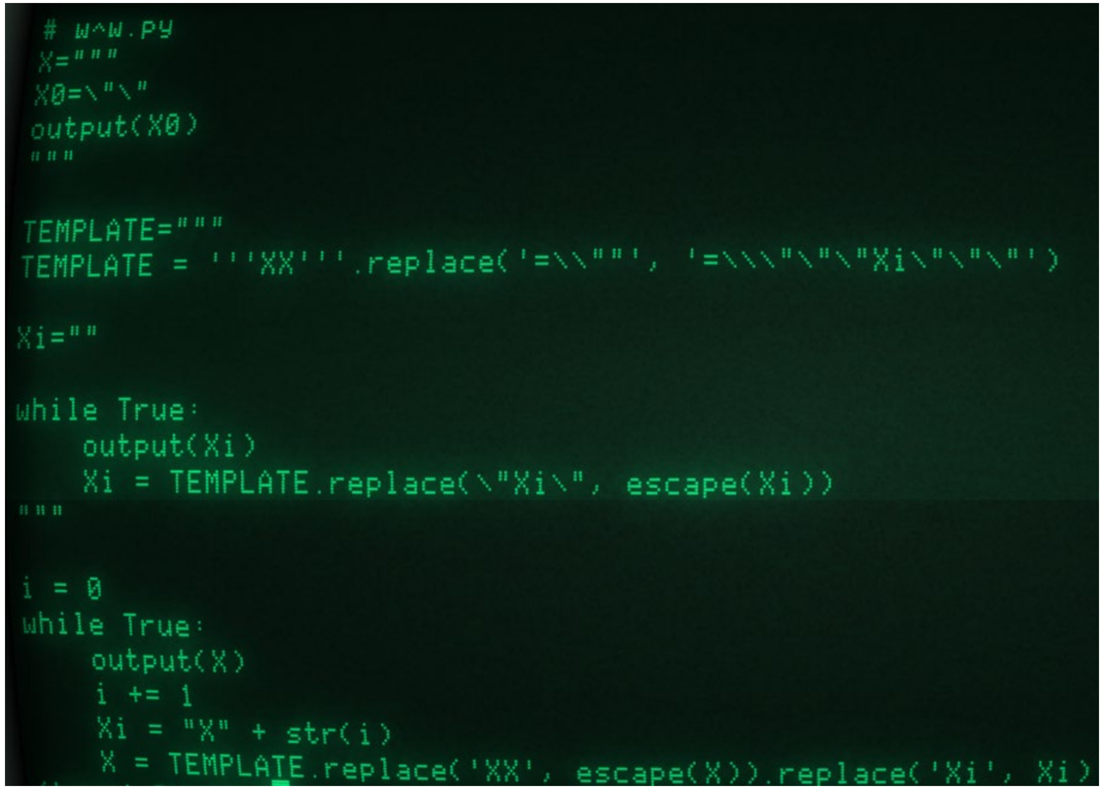
**Definition**: We define what it means for a computer program $p$ to be an *intuitive ordinal notation* (or *ION*):

1. If everything which $p$ outputs is an ION, then $p$ is an ION.
2. Nothing else is an ION.

# Examples of Intuitive Ordinal Notations

Live demo of the Library of Intuitive Ordinal Notations

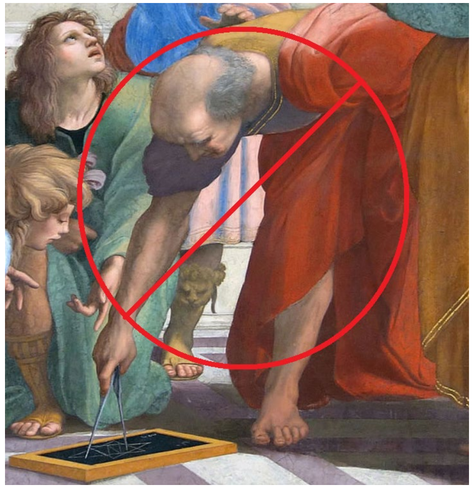
https://github.com/semitrivial/IONs



# An ordinal number environment

1. Randomly generate and display a number k of keystrokes.
2. Let agent use k keystrokes to write an ION and prove it's an ION.
3. If the agent's proof is valid, reward the agent with the ordinal notated by the ION the agent wrote. (Else, reward -1.)
4. Goto 1.

The ordinals (hence the surreals) are badly non-Archimedean. This environment can't be translated to $\mathbb{R}$ without misleading the agent.



# Summary

We tend to think of $\mathbb{R}$ as having unlimited flexibility, but that's not true. $\mathbb{R}$ is constrained by Archimedean properties.

- Thus $\mathbb{R}$ is unsuitable for some purposes. I argue $\mathbb{R}$ is unsuitable for certain RL environments, involving non-Archimedean phenomena, thus probably unsuitable for AGI via RL.

- One solution might be to do RL with a richer number system like the surreals (in which the ordinals are embedded). This might bridge the symbolic and the statistical.