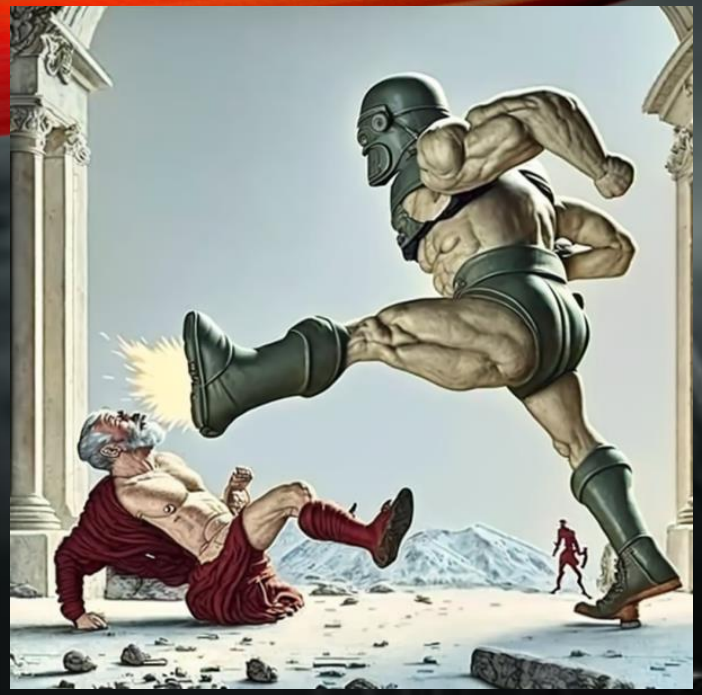


ALGORITHMIC INFORMATION AND FIRING PHILOSOPHERS

(FOR REAL THIS TIME)

ALGORITHMIC INFORMATION IS THE ROOT OF ALL OUR PROBLEMS

EFFICIENT MARKET HYPOTHESIS

- Economist: “If the market weren’t efficient, someone would exploit it and make it efficient. Thus the market must be efficient.”
- CSist: “That’s physically impossible.”

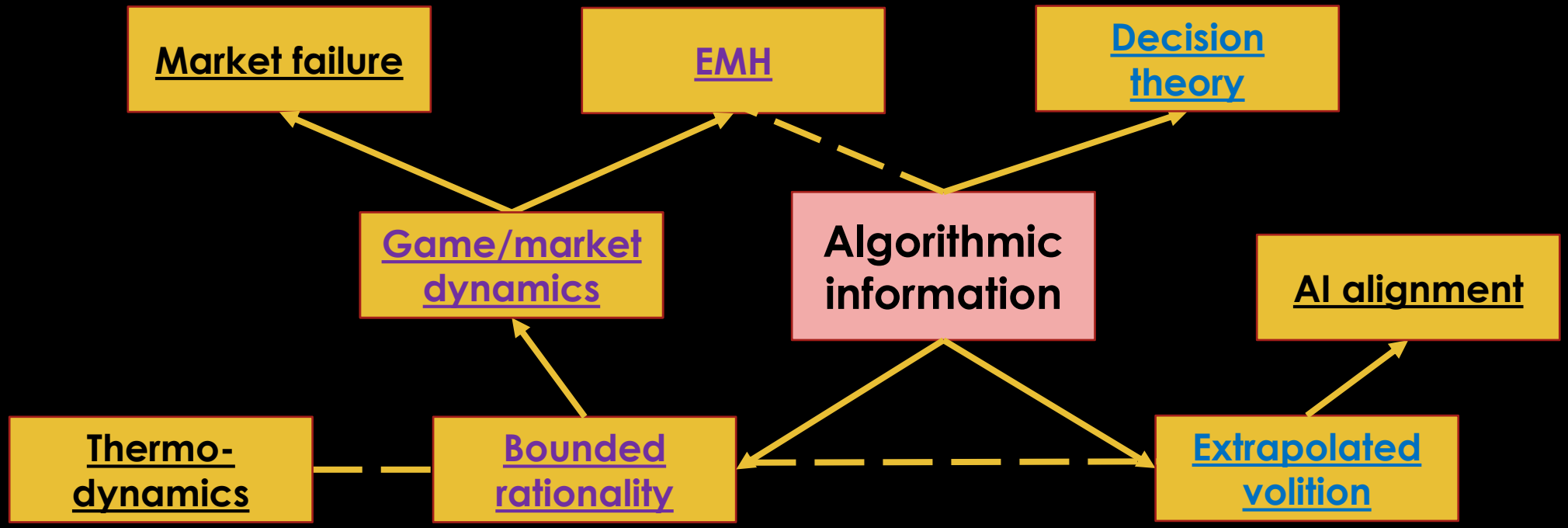
- The market is efficient “when accounting for imperfect logical knowledge”

BOUNDED RATIONALITY

- "Optimal when you account for imperfect logical knowledge" – where else have we heard that before?
- We assign a probability of 10% to the trillionth digit of pi, because that's the best algorithm we have (subject to some constraints)
- … or not necessarily the best algorithm. After all, we might not *know* the best algorithm (or that it is the best algorithm)

GAME DYNAMICS

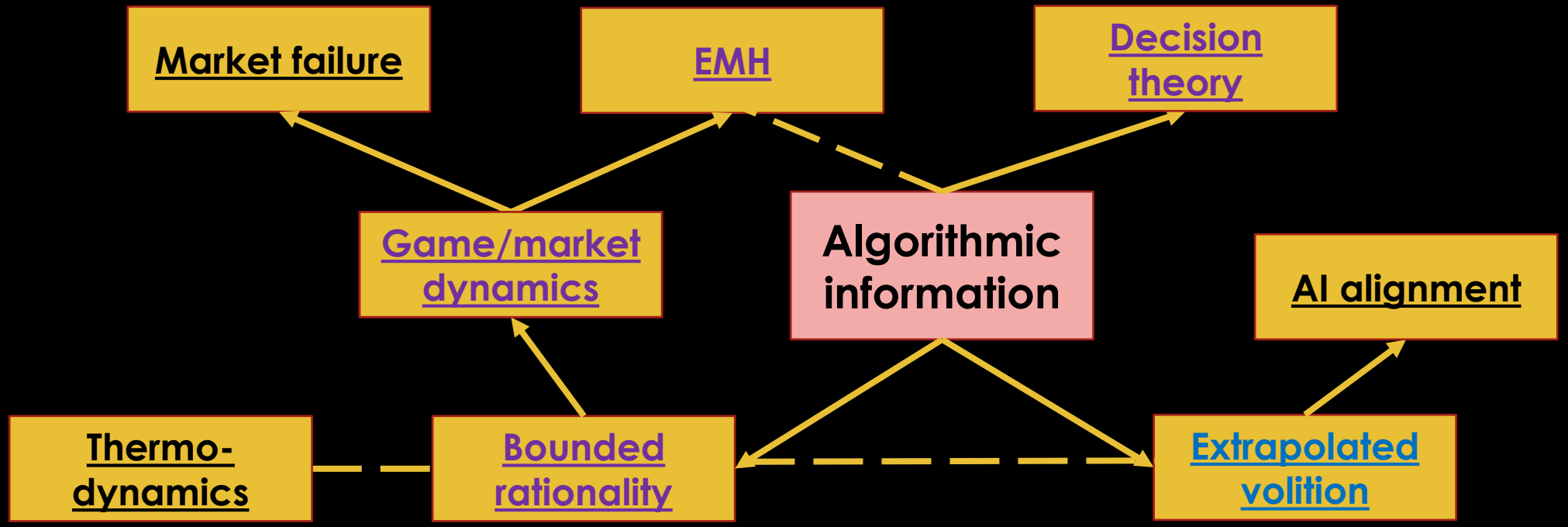
- Assuming perfect rationality is OK for equilibrium applications
- What if we want a fundamentally dynamical theory of games and markets?


Market failure
EMH
Decision theory
Game/market dynamics
Algorithmic information
AI alignment
Thermo-dynamics
Bounded rationality
Extrapolated volition

EMBEDDED AGENCY/ DECISION THEORY

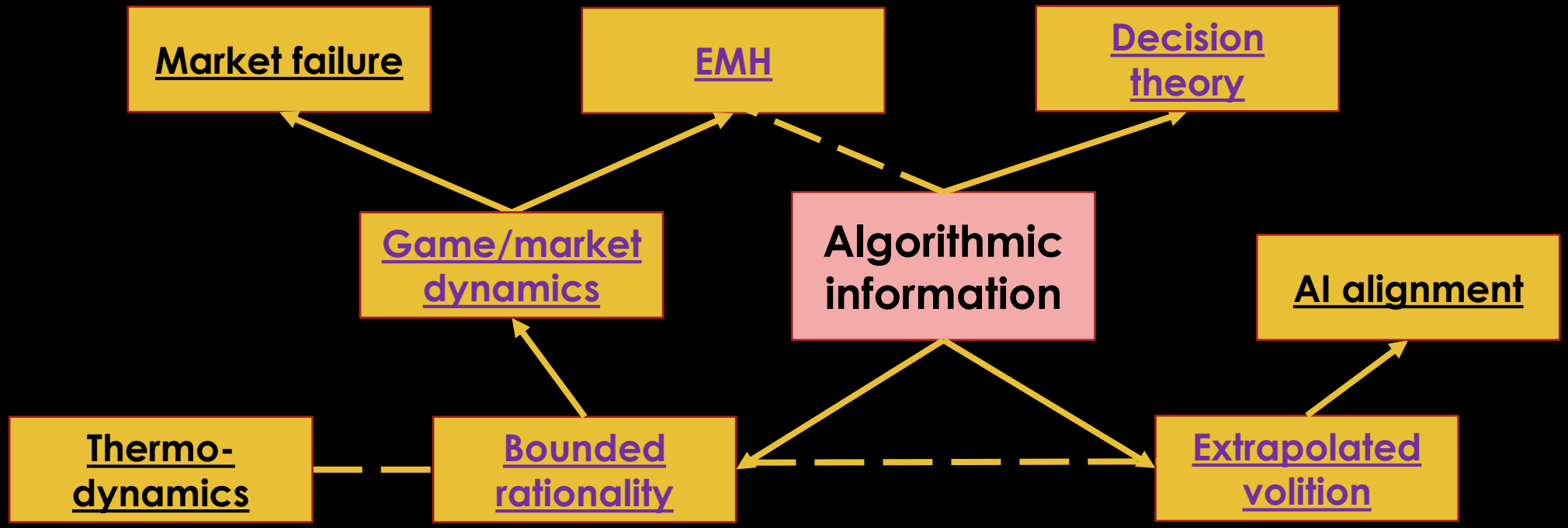
- Newcomb problem: Box A has $1M iff you reject Box B (which has $10)
- Newcomb soda: You drink a soda blindly, which gives you a strong urge to choose either Box A or Box B, and independently gives you $1M or $10

- Causal dependence: A = f(B) or B = f(A)
- Common cause: A = f(C), B = f(C)
- Algorithmic dependence: A = f(C), B = f(D)


Market failure
EMH
Decision theory
Game/market dynamics
Algorithmic information
AI alignment
Thermo-dynamics
Bounded rationality
Extrapolated volition

EXTRAPOLATED VOLITION

- What are is program's "utility function"?
- Cannot just fit an as-if theory: it is lacking in (algorithmic) information
- What would you want if you had all relevant algorithmic information?
- Relevant to AI alignment


Market failure
EMH
Decision theory
Game/market dynamics
Algorithmic information
AI alignment
Thermo-dynamics
Bounded rationality
Extrapolated volition

THIS IS THE MOST IMPORTANT THING IN THE WORLD

- The fundamental problem of philosophy
- Relevant to AI Notkilleveryoneism
- Lots of low-hanging fruit in the adjacent areas
- We can finally fire all the philosophers

IDEA: AGENTS *ARE* MARKETS!

- Garrabrant et al (2016) – Logical Induction