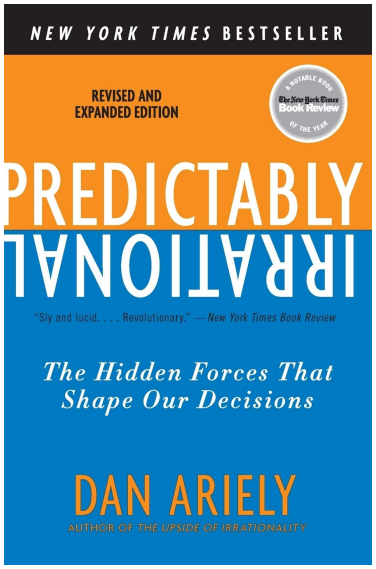
The big picture

Many studies have shown that humans are "predictably irrational"

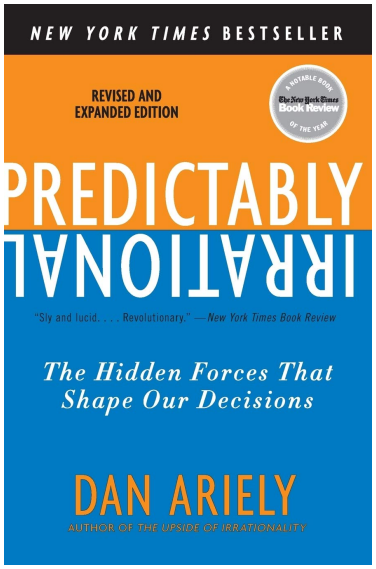
- they do not act in a fully rational way, as assumed by standard economic theory
- but their deviations from rational behavior are quite systematic



The big picture

Many studies have shown that humans are "predictably irrational"

- they do not act in a fully rational way, as assumed by standard economic theory
- but their deviations from rational behavior are quite systematic



Can we explain "predictably irrational" human behavior as the outcome of computational and cognitive constraints?

A motivating example

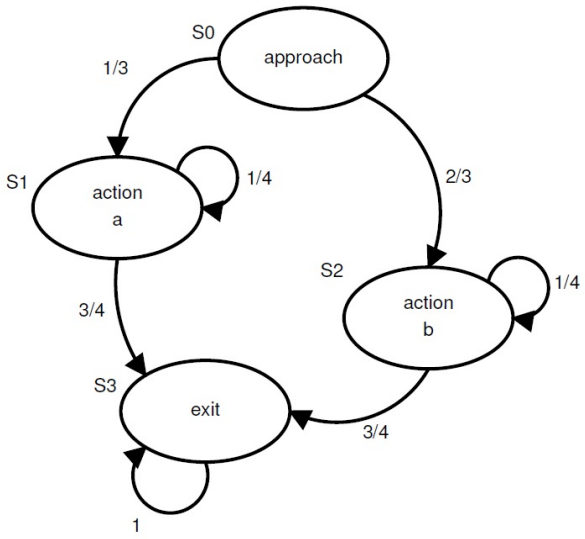
Wilson [2002/2014] considers a decision problem where an agent needs to make a single decision.

- Nature is in one of two states: 0, 1
- The DM (decision maker) wants to "match" nature's state
- Nature's state is static: it doesn't change
- The DM gets one of k independent signals, which are correlated with nature's state, at each time step
- The game ends at each step with some small constant probability. At that point the agent must make a decision.

Probabilistic finite automata (PFA)

To capture resource-boundedness, we want to model people as probabilistic finite automata (PFA).

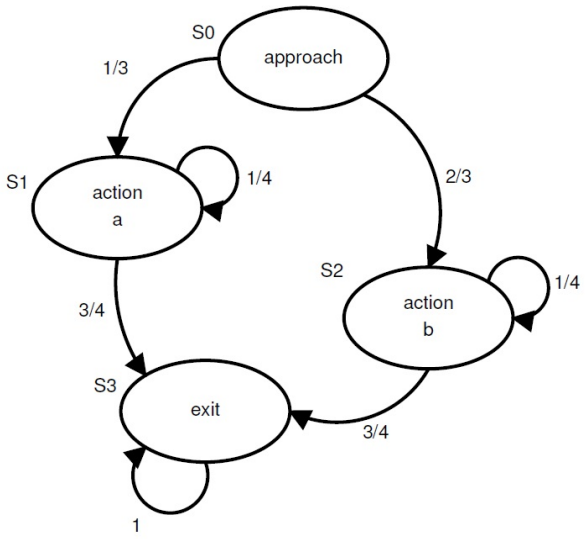
- Just like deterministic finite automata, except that we allow probabilistic state transitions.



Probabilistic finite automata (PFA)

To capture resource-boundedness, we want to model people as probabilistic finite automata (PFA).

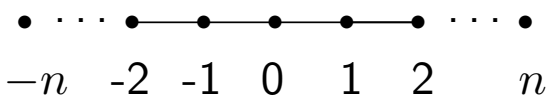
- Just like deterministic finite automata, except that we allow probabilistic state transitions.



Can we find good PFA for this problem?

The optimal automaton

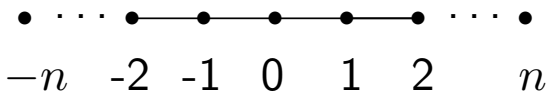
Wilson proves that the optimal PFA has the following structure:



- The states can be laid out "linearly": $-n, \ldots, 0, \ldots, n$
  - Intuitively, state 0 represents "indifference"
  - more positive/negative means more likely to be $1/0$
- The DM ignores all but the strongest signals for 0 and 1
- The automaton moves right/left with some probability iff it gets a strong signal for $1/0$.

The optimal automaton

Wilson proves that the optimal PFA has the following structure:



- The states can be laid out "linearly": $-n, \ldots, 0, \ldots, n$
  - Intuitively, state 0 represents "indifference"
  - more positive/negative means more likely to be $1/0$
- The DM ignores all but the strongest signals for 0 and 1
- The automaton moves right/left with some probability iff it gets a strong signal for $1/0$.
- Key point: The probability of moving left/right decreases the further out to the right/left the agent is.
  - "Don't bother me; I've made up my mind!"

The punch line

The optimal automaton with $2n + 1$ states has this structure:

- independent of n;
- transition probability depends on n and signal strength.

The optimal automaton exhibits "human-like" behavior:

- It ignore evidence
- It exhibits confirmation bias
- The order that evidence is received matters!
  - First-impression bias
- Belief polarization:
  - Two people that initially have have only slightly different beliefs can end up with very different beliefs

The punch line

The optimal automaton with $2n + 1$ states has this structure:

- independent of n;
- transition probability depends on n and signal strength.

The optimal automaton exhibits "human-like" behavior:

- It ignore evidence
- It exhibits confirmation bias
- The order that evidence is received matters!
  - First-impression bias
- Belief polarization:
  - Two people that initially have have only slightly different beliefs can end up with very different beliefs

Supposedly irrational behavior may be quite rational!