

RESEARCH

*“Once committed to an attack, fly in at full speed. After scoring crippling or disabling hits, I would clear myself and then repeat the process. I never pursued the enemy once they had eluded me. Better to break off and set up again for a new assault. I always began my attacks from full strength, if possible, my ideal flying height being 22,000 ft because at that altitude I could best utilize the performance of my aircraft. **Combat flying is based on the slashing attack and rough maneuvering. In combat flying, fancy precision aerobatic work is really not of much use. Instead, it is the rough maneuver which succeeds.**” - Colonel Erich ‘Bubi’ Hartmann*

Colonel Bubi Hartmann knew a thing or two about flying combat aircraft. He’s known as the world’s leading ace, scoring 352 air combat victories during World War II.

The above quote (emphasized by me) is a perfect explanation of the life of an investor. Read some of these quotes again ...

- ❖ *“Begin from full strength”*
- ❖ *“Find your ideal flying height”*
- ❖ *“Once committed, fly in at full speed”*
- ❖ *“Rough Maneuvering beats fancy precision”*

These tactics work in the clouds and on Wall Street.

Yet as important as learning how to fight, fighter pilots needed ways to avoid collision. Avoiding collision kept pilots “in the game” and out of catastrophic danger (i.e., death).

The same goes for investors. It’s not enough to know what to invest in and what types of companies we should buy. We want to avoid “collisions” with value traps, frauds and missed opportunities.

In fact, **a fundamental problem faced by all creatures is learning to select actions based on noisy sensory information and incomplete knowledge of the world.**

This essay dives deep into **The Partially Observable Markov Decision Process** or POMDP. The POMDP is one of four Markov Models developed by Russian Mathematician Andrey Markov.

Markov Models helps us think about avoiding collisions not only in planes but in life and investing. Distilling the noise. Creating an environment where we can make informed decisions with varying degrees of confidence.

These models allow us to have peace in our inability to forecast anything with precision. It frees us from the burden of acrobatic decimal places on cash flow models. Through Markov, we see the beauty in the imperfect. Elegance in robustness. Investing via rough maneuvering.

Before we begin I should let you know that I may have written this essay wearing a Russian Ben & Jerry’s Ice Cream shirt. And no, I don’t have an extra to send you.

The Markov Models

The POMDP is one of four Markov Models. Each model fits a different purpose or different state. In fact, we can think about the Models through the following decision matrix:

Markov models		
	System state is fully observable	System state is partially observable
System is autonomous	Markov chain	Hidden Markov model
System is controlled	Markov decision process	Partially observable Markov decision process

There are two inputs into the models: **systems** and **system states**

Systems are either autonomous (independent) or controlled (influenced by outside factors).

System states are either fully observable or partially observable. Think of this as having complete and incomplete information.

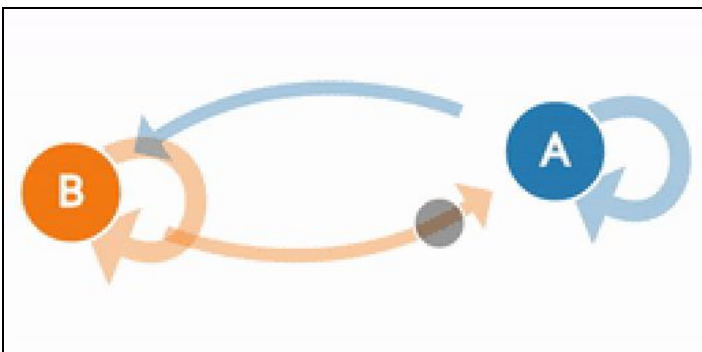
From these building blocks, we can form an understanding of each model. Let's start with Markov Chains.

Markov Chains

Markov chains are mathematical systems that jump from one 'state' to another. The simplest of the Markov Models, Markov Chains suggests that the distribution for each random variable depends **only on the distribution of the previous state**.

Put simply, it's a process in which predictions about future outcomes rely *solely on its present state*.

Most importantly, these predictions are **just as good** as predictions made knowing a process's full history.



A great example of a Markov Chain is weather forecasting. Let's say you live in a place with two weather climates: sunny and rainy. You know that every day the weather will be either sunny or rainy. But you also know that it's not a perfect 50% shot.

Weather, like everything in life, experiences transitions.

You collect weather data for the last few years and notice a pattern. Whenever there's a cloudy day in the current state, 75% of the time it's followed by another cloudy day.

This gets you the probability distribution that you can feed into your Markov Chain. When you feed the current day's weather into your Chain, you know that if today's weather is cloudy, there's a 75% chance that tomorrow will be cloudy too.

What makes Markov Chains unique is that they're memoryless. In this example, you don't need to know *anything* about historical data. You're forming probabilities based on analysis of current states moving to the next state.

Setosa.io explains this idea further:

"Did you notice how the above sequence doesn't look quite like the original? The second sequence seems to jump around, while the first one (the real data) seems to have a "stickyness". In the real data, if it's sunny (S) one day, then the next day is also much more likely to be sunny.

We can mimic this "stickyness" with a two-state Markov chain. When the Markov chain is in state "R", it has a 0.9 probability of staying put and a 0.1 chance of leaving for the "S" state. Likewise, "S" state has 0.9 probability of staying put and a 0.1 chance of transitioning to the "R" state."

Investing is a fight to get to a Markov Chain of thinking. We see this with quantitative approaches where price is the only thing that matters. In fact, [momentum strategies](#) are based on Markov Chains.

If a stock rises today, there's a >50% probability that it will rise tomorrow. And if it rises tomorrow, there's a >50% probability that it rises the next day. Etcetera.

But humans aren't robots and when we trade/invest we bring our human memories into the equation. Let's take a patented "burned trader" fallacy example. Someone loses big in one industry and complains, "I've lost money so many times in {insert industry -- likely oil -- here} there's no way I'm buying this stock!"

That's a violation of a Markov Chain. Markov Chains don't care about what happened in the past. They care about what is happening now and the states that are changing *now*.

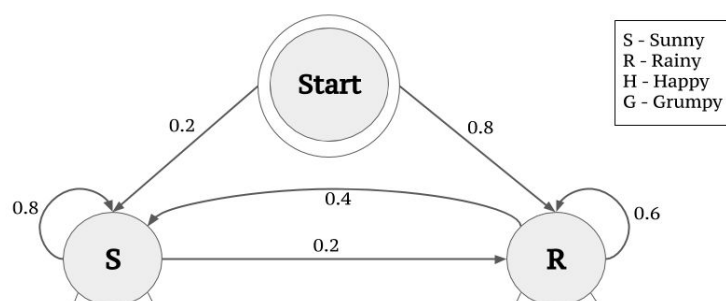
It's why Blackjack is such a great analogy to trading. In Blackjack, players can remember which cards they've already seen. This impacts their decisions in the next "state" (or hand). Investors know which companies have made them money in the past. Specifically which *attributes* from companies that made them money in the past. Those memories influence the decisions they make today with each new idea.

We don't operate in pure Markov Chains. That's where Hidden Markov Models come in.

Hidden Markov Models (HMMs)

HMMs are Markov Chains in which the "state" is only partially observable. This gets us closer to investing. We never have complete information on company. We only have what we what the company discloses.

In light of this partial information, assume that there's another



a stock or
can see and

HMMs
process (y)

whose behavior “depends” on the state of the system (x).

In other words, HMMs allow us to take what we see (observables) and guess the state we're in (x).

We can see this in action through another weather-based example (via TowardsDataScience):

“Consider the example given below in Fig. 3 which elaborates how a person feels on different climates.”

*“The feeling that you understand from a person emoting is called the **observations** since you observed them.”*

*“The weather that influences the feeling of a person is called the **hidden state** since you can't observe it.”*

In the above example, we're trying to guess the **type** of climate **given** what we can observe in a person's mood.

Investing is one giant Hidden Markov Model. We don't have perfect information. Because of that, we have to make the best *guesses* with the information we do have.

Like weather and mood patterns, we know *general trends* about markets. We know it's easier to make money going long in bull markets and harder to make money going long in bear markets. We know companies *tend* to outperform if they can grow **revenue**, **earnings**, and **multiples**. Just like we know a person *tends* to have a better mood during a sunny day.

Companies (like the person in the above example) reveal observations about the current state (weather) of the company. Higher revenues and margin expansion allows us to weigh with a greater probability long term success.

That doesn't mean they will. But it allows us to increase our probability and thus our confidence in the consequences of what we're observing.

We can view stock prices as the reflection of all market participants guessing about the observations they're receiving. Price movements are numerical expressions of the crowd's Hidden Markov Model at work!

This is why divorcing short-term price movements from long-term business performance is important. The stock price tells us nothing about the long-term probabilities of a given set of observations.

HMMs get us closer in understanding how to interact with the partially observable investment landscape. But we can do better. Enter the Markov Decision Process (MDP) and Partially Observed Markov Decision Process (POMDP).

Markov Decision Process: The Holy Grail of Investor Probability Models

An MDP provides a mathematical framework for modeling decision-making in situations where outcomes are *partly* random and *partly* under the control of the decision-maker.

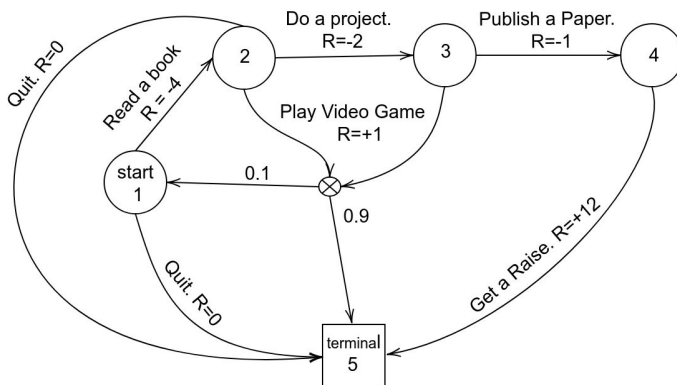
The model goes like this:

- ❖ At each step, the process is in some state (s)
- ❖ Decision-maker may choose any action (a) that is available in that state (s)
- ❖ Process responds by randomly moving to a new state (s_1)
- ❖ The new state gives the decision-maker a reward (Ra)

The important thing to remember with MDPs is that probabilities are influenced by the chosen action (and thus, the decision-maker). This *also* means that the next state depends on the current state *and* the decision-maker's choice in that current state.

So you've got these state spaces, action spaces, probabilities, and rewards. What's the goal with the model?

The goal in an MDP is to find an optimal "policy" that decision-makers use when faced with various actions at different state spaces.



In other words, the goal is to turn the partially observable world -- with its myriad variables and random outcomes -- into a **simple Markov Chain**.

Markov All The Way Down!

Investing is the perfect playground for experimenting with MDPs.

Generally, investors only make three decisions:

1. Buy
2. Sell
3. Do nothing

Using an MDP, investors can determine the optimal "policy" for their decisions in each state s (in this case: buy, sell, do nothing). The goal is to turn decisions into simple Markov chains. Each time state s_1 occurs, the investor makes d_1 decision. This fixes the action for each state, simplifying the job of the decision-maker.

This is what traders do *so well* compared to long-term investors. Traders develop, backtest and quantify their "policies". They **know** that when state s triggers they'll make decision d every time.

They do this through profit targets, buy rules, and sell rules.

In other words, they turn complex, MDP decisions into simple, robotic Markov chains.

Further complicating things, investors operate both in discrete-time and continuous-time environments. Quarterly portfolio rebalancing is an example of a discrete-time MDP. Yet finding a new investment idea that's better than one you currently own (on a random Tuesday) represents a continuous-time MDP.

In other words, there are certain investment actions we can delegate to discrete-time states. But there's an equal (if not greater) number of investment actions *we must accept* as elements of a continuous-time MDP.

Static MDPs works *great* when we can fully observe the state spaces we're in. Unfortunately, our craft isn't that simple.

Investing is like a Hell In The Cell cage-match between Expectations, Time, Technology, and Human Psychology.

We need a more robust model. A model that can handle the "slashing attack and rough maneuvering" of business valuation.

The Partially Observed Markov Decision Process (POMDP)

"Neurophysiological and psychophysical experiments suggest that the brain relies on probabilistic representations of the world and performs [Bayesian inference](#) using these representations to estimate task-relevant quantities (sometimes called "hidden or latent states") (Knill and Richards, 1996; Rao et al., 2002; Doya et al., 2007)."

The above quote is a fancy way of saying we make guesses about things based on what we've learned/experienced in the past.

That's where POMDP models come in.

A POMDP is an MDP in which the state of the system is only partially observed. In a POMDP, the agent **cannot directly observe the underlying state**. Instead, **it must maintain a probability distribution over a set of possible states**, based on a set of observation and observation probabilities.

How cool is it that the POMDP model *perfectly* describes what we do as investors?

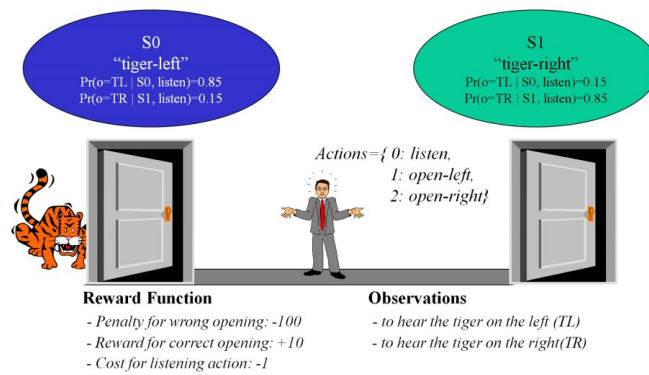
A POMDP model follows this basic formula:

1. At each time period, the environment is in some state s
2. The agent (decision maker) takes an action a , which causes the environment to transition to state s_1 with probability $T(s_1 | s, a)$
3. At the same time, the agent receives an observation o which depends on the new state of the environment s_1 , and on the just taken action a , with probability $O(o | s_1, a)$

4. Finally, the agent receives a reward equal to $R(s, a)$

The goal with any POMDP model is for (decision-maker) to choose actions at that maximize its **expected future rewards** (where r_t is the reward earned

The most important factor in this discount factor (γ). This discount factor whether the rewards gained from the short-term or long-term biased.



reward r

the agent each step **discounted** at time t).

model is the determines model are

Maximizing The Discount Factor Is The Most Important Lesson

When $\gamma = 0$ the agent *only cares about* which action will yield the largest expected *Immediate* reward; when $\gamma = 1$ the agent cares about maximizing the *expected sum of future* rewards.

The idea that to achieve greatness you need to maximize $\gamma = 1$ isn't new. But it's surprisingly rare when you look at the landscape of public equity executives.

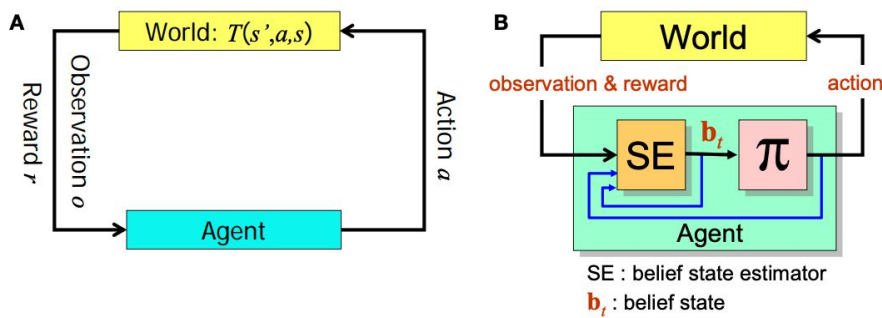
Take Amazon CEO Jeff Bezos for example. He's a product of maximizing his $\gamma = 1$ POMDP model. Check out an excerpt from his 2003 shareholder letter (emphasis mine):

“Long-term thinking is both a requirement and an outcome of true ownership. Owners are different from tenants. I know of a couple who rented out their house, and the family who moved in nailed their Christmas tree to the hardwood floors instead of using a tree stand. Expedient, I suppose, and admittedly these were particularly bad tenants, but no owner would be so short-sighted. Similarly, many investors are effectively short-term tenants, turning their portfolios so quickly they are really just renting the stocks that they temporarily ‘own.’”

Warren Buffett's also a product of maximizing his $\gamma = 1$ model. One of his most famous quotes hints at the idea:

“If you aren't thinking about owning a stock for 10 years, don't even think about owning it for 10 minutes.”

For POMDPs to work, we need to constantly update our beliefs. In turn, these updated beliefs impact the actions we take on the observations we see. Which impacts our beliefs and changes probabilities on future expected outcomes.



If you're thinking to yourself, "this reminds me of George Soros' Reflexivity Theory" you're right.

Alex (my partner at Macro Ops) [wrote a fantastic piece](#) on Soros' theory. Check out this excerpt (emphasis mine):

*“Objective realities are true regardless of what participants think about them. For example, if I remark that it’s snowing outside and it is in fact snowing outside, then that is an objective truth. It would be snowing outside whether I said or thought otherwise — I could say it’s sunny but that would not make it sunny, it would still be snowing. Subjective realities on the other hand are **affected by what participants think about them**. Markets fall into this category. **Since perfect information does not exist (ie, we can’t predict the future and it’s impossible to know all the variables moving markets at any given time) we make our best judgments as to what assets (stocks, futures, options, etc) should be valued at**. Our collective thinking is what moves markets and produces winners and losers. This means that what we think about reality affects reality itself. And that reality in turn affects our thinking once again.”*

Those revelations are nice, but how do we implement POMDPs in our investment decision-making process? Let's use an example highlighting each of the above five steps:

- **Step 1:** We find a new stock that looks promising and could generate market-beating returns over the next 5-10 years (environment in state s). At this stage, the company has a 50% probability of beating the market in our minds (because we haven't done any research on the name, it's a virtual coin toss/random walk).
- **Step 2:** After researching the company we realize the underlying business has very attractive qualities like:
 - Positive scale economics
 - Founder-led CEO
 - Lowest-cost provider
 - Shares cost savings with customers
 - High switching costs

Note: An investor's research process is what's known as "information-gathering" in the POMDP model. **This "information-gathering" allows the decision-maker to reduce the current uncertainty in their estimate of a belief state before committing to a decision.** In other words, researching a new idea allows us to narrow the cone of potential future outcomes (thus reducing uncertainty in the investment).

These observations, o , impact the action we take, a . Our observations make us **buy** that company with a $>50\%$ ($T(s^1 | s, a)$) probability that the stock will beat market returns over the next five-ten years (s^1).

It's also at this point in the process that we're *constantly* feeding new information into our Belief State Estimator ("SE").

- **Step 3:** Five years later we observe (o) that the company we bought is around our estimation of intrinsic business value (s^2).
- **Step 4:** We're "rewarded" for our long-term bet on the company with market-beating returns. And since we cared more about the future (5-10 year time horizon), our discount factor was closer to $\gamma = 1$.

As long-term investors in shares of businesses, we want to **maximize the discount factor** so that it's as close to 1 as possible. We also want to own businesses whose leaders want to maximize their $\gamma = 1$ factor!

It's Markov All The Way Down

The beauty of Markov's models (like all good mathematics) is that it is fractal in nature. The most complex models of understanding the world (POMDPs) have one goal: reduce decisions down to their most basic forms (Markov Chains).

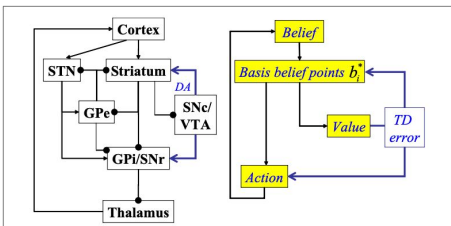


FIGURE 3 | Suggested mapping of elements of the model to components of the cortex-basal ganglia network. STN, subthalamic nucleus; GPe, globus pallidus, external segment; GPI, Globus pallidus, internal segment; SNc, substantia nigra pars compacta; SNr, substantia nigra pars reticulata; VTA, ventral tegmental area; DA, dopaminergic responses. The model suggests that cortical circuits compute belief states, which are provided as input to the striatum (and STN). Cortico-striatal synapses are assumed to maintain a compact learned representation of cortical belief space ("basis belief points"). The dopaminergic responses of SNc/VTA neurons are assumed to represent reward prediction (TD) errors based on striatal activations. The GPe-GPI/SNr network, in conjunction with the thalamus, is assumed to implement action selection based on the compact belief representation in the striatum/STN.

Markov helps the investor grip with forecasting uncertainty. The fighter pilot with its collision avoidance system. And the biologist with understanding cognitive function in the brain.

A model that penetrates all areas of life can't help but provide value to the area of investment management and business valuation.

What is a business but a collection of biological specimens? Serving the ever-changing needs of other biological specimens. It's Markov All The Way Down.