# Framing Practicum

1. Framing Practicum: Stable Equilibrium
2. Framing Practicum: Bistability
3. Framing Practicum: Dynamic Equilibrium
4. Framing Practicum: Timescale Separation
5. Framing Practicum: Turnover Time
6. Framing Practicum: Incentive
7. Framing Practicum: Selection Incentive
8. Framing Practicum: Comparative Advantage
9. Framing Practicum: Semistable Equilibrium
10. Framing Practicum: General Factor
11. Framing Practicum: Dynamic Programming

# Framing Practicum: Stable Equilibrium

This is a framing practicum post. We'll talk about what a stable equilibrium is, how to recognize stable equilibria in the wild, and what questions to ask when you find one. Then, we'll have a challenge to apply the idea.

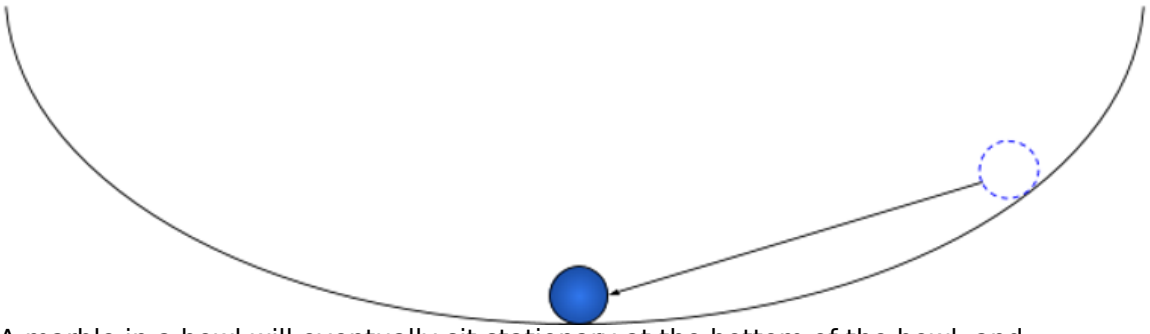
**Today's challenge: come up with 3 examples of stable equilibrium which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

Expected time: ~15-30 minutes at most, including the Bonus Exercise.

## What's a Stable Equilibrium?

Put a marble at the bottom of a round bowl, and it will just sit there without moving. Put it in the bowl but not quite at the bottom, and it will roll around a bit, but eventually settle at the bottom, and sit there without moving. Give it a poke, and it will roll around some more, but eventually it will again sit at the bottom without moving.

This is *stable equilibrium*: the system may start in different states, or it may be "perturbed" into different states by some external force, but eventually it settles back to the same state (assuming it isn't pushed *too* far away...).



A marble in a bowl will eventually sit stationary at the bottom of the bowl, and stay there.

## What To Look For

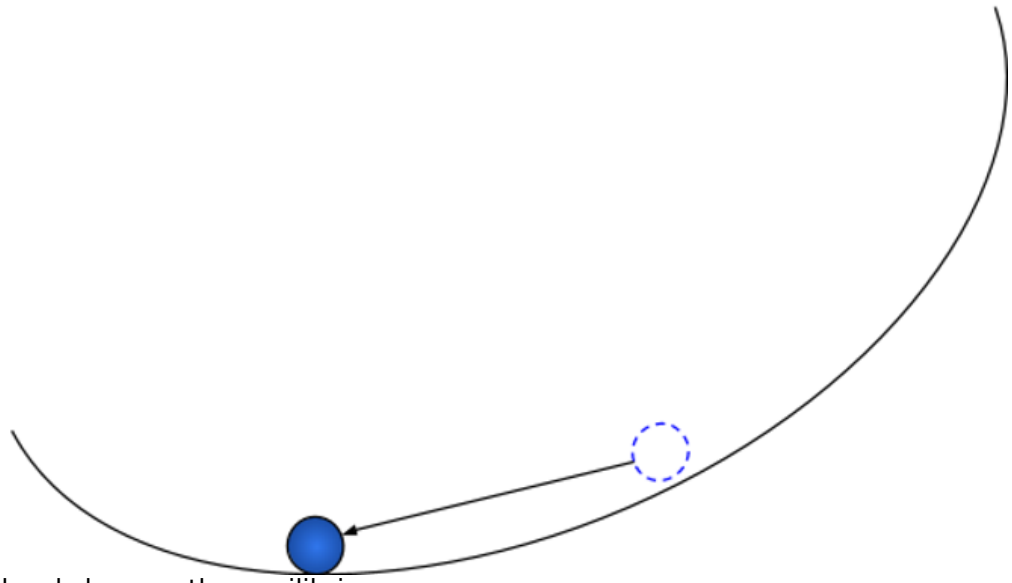
Stable equilibrium should spring to mind whenever a system tends to return to the same state. If you could "poke" it somehow, and the system would go back to normal eventually,

that’s probably a stable equilibrium. If a system tends to stay suspiciously the same over the long run, despite lots of short-run noise, that’s probably a stable equilibrium.

## Useful Questions To Ask

The marble always returns to the bottom of the bowl. If we push the marble away from the bottom, that’s only a short-term change - it will roll back down eventually. So, *if* we’re mainly interested in the *long run* behavior of the marble, then we can ignore such little pushes.

On the other hand, there may also be ways to change the equilibrium state itself. For instance, if we tip the bowl to the side slightly, then the equilibrium position of the marble will change. If we deform the bowl, that could change equilibrium position. If we charge the marble with a little static electricity, then place another charged object near the bowl, that could also change the equilibrium. Finally, very large changes to the system state could push it out of the bowl entirely.



Tipping the bowl changes the equilibrium.

**When we frame something as a stable equilibrium, we ignore temporary changes to the system state, and only pay attention to things which change the equilibrium.**

Two main ways this can apply:

- We want to change the long-term behavior of a system. So, we focus on things which can change the equilibrium, and ignore things which don’t.

- We see a change in the long-term behavior of a system, and want to know what caused it. So, we focus on things which can change the equilibrium, and ignore things which don't.

# The Challenge

(Rules adapted from the Babble Challenges)

**Come up with 3 examples of stable equilibrium which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you). I recommend mentioning what the equilibrium is, and a few ways you could "poke" the system for which it would return to equilibrium afterwards, so that everyone understands your example.

**Any answer must include at least 3 to count, and they must be novel to you**. That's the challenge. We're here to challenge ourselves, not just review examples we already know.

However, they don't have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

Reward people for babbling — don't punish them for not pruning.

I will remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for:**

- Systems which go back to normal after you poke them
- Systems which stay suspiciously the same over time despite lots of short-term noise

**Bonus Exercise**: for each of your three examples from the challenge, suppose you want to change the equilibrium, or you want to know what caused a change in the equilibrium. What factors should you pay attention to (since they can change the equilibrium)? What factors can you safely ignore (since they only affect the system in the short term)?

This bonus exercise is great blog-post fodder!

# Motivation

Much of the value I get from math is not from detailed calculations or elaborate models, but rather from *framing tools*: tools which suggest useful questions to ask, approximations to make, what to pay attention to and what to ignore.

Using a framing tool is sort of like using a trigger-action pattern: the hard part is to notice a pattern, a place where a particular tool can apply (the “trigger”). Once we notice the pattern, it suggests certain questions or approximations (the “action”). This challenge is meant to train the trigger-step: we look for novel examples to ingrain the abstract trigger pattern (separate from examples/contexts we already know).

The Bonus Exercise is meant to train the action-step: apply whatever questions/approximations the frame suggests, in order to build the reflex of applying them when we notice a stable equilibrium.

Hopefully, this will make it easier to notice when a stable equilibrium frame can be applied to a new problem you don’t understand in the wild, and to actually use it.

*Thankyou to Sisi, Eli, Adam and especially Jacob for beta-testing and feedback. Also thankyou to Aysajan for our daily discussions, which led to this concept.*

# Framing Practicum: Bistability

This is a framing practicum post. We'll talk about what bistability is, how to recognize bistability in the wild, and what questions to ask when you find it. Then, we'll have a challenge to apply the idea.

**Today's challenge: come up with 3 examples of bistability which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

Expected time: ~15-30 minutes at most, including the Bonus Exercise.

## What's Bistability?

The classic picture of bistability is a marble in a double bowl:

The marble has one stable equilibrium on the left, and another on the right. Although the marble has a whole continuum of possible positions, when left to its own devices for a while it will settle down to one of just two positions.

My own head-canonical examples of bistability come from digital electronics. One is the signal buffer: it turns a sorta-low voltage (like 1 V) and into an unambiguously low voltage (like 0.01 V), or a sorta-high voltage (like 4 V) into an unambiguously high voltage (like 4.99 V). One stable equilibrium is at 5 V, the other is at 0 V, and all other voltages get pushed toward one of those two. This is crucial to building large digital circuits: without buffering, 5 V would decay to 4 V then 3 V as we pass through one gate after another, and eventually we wouldn't be able to tell whether a voltage is supposed to be high or low.

Another electronic example is the latch, one of the standard low-level memory elements in digital circuits. You can think of a latch sort of like the marble-in-a-double-bowl, but with two extra features:

- The state of the marble can be read out. One "bowl" represents "0", and the other "1".
- An input signal can switch the "marble" from one state to the other.

So, to "write" a bit into the memory element, we push the system into the desired "bowl" (i.e. basin of attraction). It then stays in that state indefinitely, and we can read out the stored bit as many times as we like until it is "overwritten" (i.e. the state is set again).

## What To Look For

In general, bistability (and multiple stability) should come to mind whenever an analogue system (i.e. a system with continuous state variables) has discrete behavior. In particular, it's

usually necessary for lossless transmission/storage of discrete information - a system with a single stable equilibrium has no long-term memory, since it always returns to the same state.

# Useful Questions To Ask

In the double bowl picture earlier in the post, there's a hump between the two stable equilibria. The higher the hump, the harder it is to push the marble from one equilibrium to the other. In chemistry, we call that height the "activation energy" - the energy which must be provided to move from one stable state to another. If we want to switch the marble's state (e.g. to write a bit into memory), we need to provide that activation energy, and a "higher hump" makes it harder/more expensive to set the bit. On the other hand, a higher activation energy makes it less likely that random noise will accidentally push the marble from one side to the other, so a higher-hump memory element can store a bit for longer.

In general, other than the usual equilibrium questions, **in a bistable system we usually want to know what's required to change from one stable equilibrium to another**. A few ways this can apply:

- We want to set the system into one state or another. So, we need to know what kind of "kick" will do that.
- We want to make the system more or less likely to switch state. So, we need to know how to raise or lower the "hump" between states.
- We see the system change from one state to another, and we want to know what caused the switch. So, we look for kicks which are large enough to overcome the hump, and ignore smaller noise.
- We want to know how long it will take the system to accidentally change states due to random noise - i.e. how long the "memory" is reliable.

# The Challenge

**Come up with 3 examples of bistability which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

**Any answer must include at least 3 to count, and they must be novel to you**. That's the challenge. We're here to challenge ourselves, not just review examples we already know.

However, they don't have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

I may remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for:**

- Systems whose long run behavior can end up a few different possible ways, depending on the initial conditions
- Analogue systems with discrete behavior
- Long-range transmission or long-term storage of discrete information

**Bonus Exercise**: for each of your three examples from the challenge, what kinds of “kicks” could cause the system to switch state? What controls how strong the kick needs to be? Suppose you want to switch the state, or want to know what caused a state switch; what kicks could you rule out on the basis that they’re too small? Can you do a Fermi estimate for how often a kick large enough to force a state switch happens due to random noise?

This bonus exercise is great blog-post fodder!

# Motivation

Much of the value I get from math is not from detailed calculations or elaborate models, but rather from *framing tools*: tools which suggest useful questions to ask, approximations to make, what to pay attention to and what to ignore.

Using a framing tool is sort of like using a trigger-action pattern: the hard part is to notice a pattern, a place where a particular tool can apply (the “trigger”). Once we notice the pattern, it suggests certain questions or approximations (the “action”). This challenge is meant to train the trigger-step: we look for novel examples to ingrain the abstract trigger pattern (separate from examples/contexts we already know).

The Bonus Exercise is meant to train the action-step: apply whatever questions/approximations the frame suggests, in order to build the reflex of applying them when we notice bistability.

Hopefully, this will make it easier to notice when a bistability frame can be applied to a new problem you don’t understand in the wild, and to actually use it.

# Framing Practicum: Dynamic Equilibrium

This is a framing practicum post. We'll talk about what dynamic equilibrium is, how to recognize dynamic equilibrium in the wild, and what questions to ask when you find it. Then, we'll have a challenge to apply the idea.

**Today's challenge: come up with 3 examples of dynamic equilibrium which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

Expected time: ~15-30 minutes at most, including the Bonus Exercise.

## What's Dynamic Equilibrium?

Our picture of stable equilibrium was a marble sitting at the bottom of a bowl. This is a *static* equilibrium: the system's state doesn't change.

Picture instead a box full of air molecules, all bouncing around all over the place. The system's state constantly changes as the molecules move around. But if we look at the box at large scale, we don't really care about the motions of individual molecules, we just care about the overall *distribution* of molecule positions and velocities - the number of molecules in the upper-left quadrant, for instance. And this distribution may be (approximately) constant, even though the individual molecules are bouncing around.

This is *dynamic* equilibrium: even though the states of the system's components are constantly changing, the *distribution* of states of the system's components has a stable equilibrium.

The box-of-air example involves a distribution of similar physical parts (i.e. molecules), but we can also have a dynamic equilibrium with a Bayesian belief-distribution. For instance, I can think about which of my dishes I expect to be dirty tomorrow or next week or next month. I don't run the dishwasher every day, so in the short term I expect dirty dishes to pile up - a non-equilibrium expectation. But in the long run, I generally expect the distribution of dirty dishes to be roughly steady - I don't know exactly which days I'll wash them, but my expectations for 100 days from now are basically the same as my expectations for 101 days from now.

The dirty dishes themselves may pile up and then be cleaned and then pile up again, never reaching an equilibrium state. But my *expectations or forecasts* about the dishes do reach an equilibrium.

## What To Look For

In general, dynamic equilibrium should spring to mind in two situations:

- A system has parts which are constantly changing/moving, but some aggregate summary of those parts (like a total count or a distribution) tends to return to

approximately the same value.
- Our *expectations or forecasts* about a system's state tend to return to approximately the same thing when we forecast farther into the future.

# Useful Questions To Ask

If the number of air molecules in the upper-left quadrant of a box is roughly constant, then the rate at which molecules enter that quadrant must roughly equal the rate at which they leave. If the number of molecules in the quadrant is lower than usual, then molecules will enter faster than they leave until the number returns to equilibrium.

In general, dynamic equilibrium involves a balance: **the rate at which parts enter some state is roughly equal to the rate at which they leave that state**. Three key questions are:

- At what rate do parts enter each state?
- At what rate do parts leave each state?
- What does the state-distribution have to look like for those to balance?

In order for the equilibrium to be stable, we also need parts to enter a state a bit faster/leave it a bit slower when there are fewer-than-equilibrium-number of parts in that state. So, besides the rates, we also want some idea of how the rates *change* if there are slightly more or less parts in a given state.

We can also ask the corresponding questions for a dynamic equilibrium of expectations/forecasts. Rather than parts moving between states, we have probability-mass moving between states. If there's a chance that I wash the dishes in two days, then there's a flow of probability-mass from the "n dirty dishes" state to the "all dishes clean state" between two and three days in the future. If my expectations reach an equilibrium, then that means the rate of probability-mass-flow into each state equals the rate of probability-mass-flow out. So, the three key questions are:

- What processes cause probability-mass to flow into each state, and at what rate do those happen?
- What processes cause probability-mass to flow out of each state, and at what rate do those happen?
- What does the distribution have to look like for those to balance?

Once we know how the state-change rates work, we can also ask all the usual questions about stable equilibrium.

# The Challenge

**Come up with 3 examples of dynamic equilibrium which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

**Any answer must include at least 3 to count, and they must be novel to you**. That's the challenge. We're here to challenge ourselves, not just review examples we already know.

However, they don't have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

I will remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for:**

- Systems made of lots of similar parts which are all constantly in motion
- Systems with a stable equilibrium only if you "zoom out" from the details
- Systems for which your *expectations* are roughly constant sufficiently far into the future, even if the system itself is constantly in motion.

**Bonus Exercise**: for each of your three examples from the challenge, what are the relevant "parts", part-states and state-change rates (or probability-mass-flows, for expectations)? Can you do a Fermi estimate of the relevant rates, or estimate a rough (i.e. big-O) functional relationship between the state-change rates and the number of parts in each state (or probability-mass in each state)? How do the state-change rates change when the number of parts (or probability mass) in some state is higher/lower than its equilibrium value?

This bonus exercise is somewhat more abstract and conceptually tricky than previous exercises, especially for the probability-mass questions. I recommend it especially if you want some extra challenge.

# Motivation

Much of the value I get from math is not from detailed calculations or elaborate models, but rather from *framing tools*: tools which suggest useful questions to ask, approximations to make, what to pay attention to and what to ignore.

Using a framing tool is sort of like using a trigger-action pattern: the hard part is to notice a pattern, a place where a particular tool can apply (the "trigger"). Once we notice the pattern, it suggests certain questions or approximations (the "action"). This challenge is meant to train the trigger-step: we look for novel examples to ingrain the abstract trigger pattern (separate from examples/contexts we already know).

The Bonus Exercise is meant to train the action-step: apply whatever questions/approximations the frame suggests, in order to build the reflex of applying them when we notice dynamic equilibrium.

Hopefully, this will make it easier to notice when a dynamic equilibrium frame can be applied to a new problem you don’t understand in the wild, and to actually use it.

# Framing Practicum: Timescale Separation

This is a framing practicum post. We'll talk about what timescale separation is, how to recognize timescale separation in the wild, and what questions to ask when you find it. Then, we'll have a challenge to apply the idea.

**Today's challenge: pick 3 examples of equilibria from the previous three practica, and for each of them, give one use-case on a fast enough (or slow enough) timescale that we can treat the system as constant (or in equilibrium)**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

Expected time: ~15 minutes at most.

## What's Timescale Separation?

If I put a piece of iron in the ocean, its equilibrium state is rusted through. However, it takes a long time to reach that equilibrium. If I swim past and look at the piece of iron, I probably won't even see it gradually becoming rustier.

At the timescale of a person looking at the piece of iron on the way by, its state is roughly constant, even though it's out-of-equilibrium. The changes are so slow that we can ignore them.

On the other hand, consider pushing a wheelbarrow. The force isn't transmitted to the wheelbarrow instantaneously - for a brief fraction of a second after I push on the handle, the handle actually compresses a bit, and the compressed handle presses on the wheelbarrow frame, which compresses the frame a bit, which then presses on the load… the compression propagates as a wave, transmitting the force through the whole wheelbarrow. And the wave also bounces back, exerting pressure from the load back on my hands. But from my point of view, this whole process happens extremely quickly. Within a fraction of a second, the waves have settled down to an equilibrium force (and matching acceleration) between my hands and the wheelbarrow.

At the timescale of a person pushing the wheelbarrow, the forces are always roughly in equilibrium. The force-propagation process is so fast that we can ignore it.

This is timescale separation:

- If a system equilibrates on a timescale much *slower* than whatever-we're-interest-in, then we can approximate it as being in a constant non-equilibrium state.
- If a system equilibrates on a timescale much *faster* than whatever-we're-interested-in, then we can approximate it as always being in equilibrium (even if the equilibrium changes slowly over time).

Note that a system may involve multiple processes which equilibrate on different timescales. For instance, in the wheelbarrow example, there's a very fast equilibrium of forces between my hands and the wheelbarrow, but also a slower equilibrium in which I set a steady walking pace.

One common pattern to watch for: often a system doesn't return *exactly* to equilibrium, but exponentially decays toward equilibrium. (In general, this happens whenever the rate-at-which the system moves toward equilibrium is proportional to its "distance" from the equilibrium state.) In this case, the half-life is a good "equilibration timescale" for purposes of Fermi estimates and timescale separation.

# What To Look For

Timescale separation should come to mind whenever we have a stable equilibrium. For any equilibrium it's worth asking:

- Fermi estimate: how fast does the system equilibrate? If it decays to equilibrium exponentially, what's its half-life? (Note that there may be multiple processes in the same system which equilibrate on different timescales.)
- What things am I interested in which happen much faster than the equilibrium?
- What things am I interested in which happen much slower than the equilibrium?

# The Challenge

(Rules adapted from the Babble Challenges)

**Pick 3 examples of equilibria from the previous three practica, and for each of them, give one use-case on a fast enough (or slow enough) timescale that we can treat the system as constant (or in equilibrium)**.They don't need to be good, they don't need to be useful, they just need to be novel (to you).

**Any answer must include at least 3 to count, and they must be novel to you**. That's the challenge. We're here to challenge ourselves, not just review examples we already know.

However, they don't have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

I will remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck:**

- First, estimate how long it takes the system to equilibrate if it's "poked" somehow.
- Next, think about ways you might interact with the system much faster/slower than that.

# Motivation

Much of the value I get from math is not from detailed calculations or elaborate models, but rather from *framing tools*: tools which suggest useful questions to ask, approximations to make, what to pay attention to and what to ignore.

Using a framing tool is sort of like using a trigger-action pattern: the hard part is to notice a pattern, a place where a particular tool can apply (the "trigger"). Once we notice the pattern, it suggests certain questions or approximations (the "action"). This post is meant to practice the "action" step: once we recognize an equilibrium, what questions should we ask or what approximations should we test?

Hopefully, this will make it easier to notice when a timescale separation frame can be applied to a new problem you don't understand in the wild, and to actually use it.

# Framing Practicum: Turnover Time

This is a framing practicum post. We'll talk about what turnover time is, how to recognize when turnover time is relevant in the wild, and what questions to ask when you find it. Then, we'll have a challenge to apply the idea.

**Today's challenge: come up with 3 examples of approximately-independent turnover which do not resemble any you've seen before, and estimate their turnover times**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

Expected time: ~15-30 minutes at most, including the Bonus Exercise.

## What's Turnover Time?

The proteins in biological cells are often damaged by radiation or uncontrolled chemical reactions. To avoid damaged proteins building up, organisms turn over their proteins regularly: they're constantly broken down and replaced with fresh proteins. As long as the stressors which damage proteins are at a constant level, the proportion of damaged proteins reaches an equilibrium, at which turnover balances damage rate.

The typical time it takes a protein to turn over is its turnover time. (This might be an average time, median time, half-life, etc… it's a Fermi estimate, so it doesn't need to be exact.)

Another example, with a more abstract (but more general) notion of "turnover": air molecules in a box. We can draw an imaginary divider down the middle, and then think about the typical time it takes a molecule on one side to bounce over to the other side. This is a turnover time.

## What To Look For

In general, turnover time should come to mind whenever we have many parts which change, are created and destroyed, or which move between states roughly-independently over time.

## Useful Questions To Ask

Imagine a biological cell is hit with a pulse of radiation or a splash of harsh chemical, so that a bunch of its proteins are damaged. Assuming the cell is still functional, how long will it take the proportion-of-damaged-proteins to return to equilibrium? The turnover time is a good Fermi estimate: we're asking how long it will take for the damaged proteins to be replaced by new proteins, and the turnover time is the typical time it takes proteins to be replaced by new proteins.

Note that this is just a rough estimate - the cell might detect the extra damage and upregulate protein turnover, or a lot of the protein-turnover-machinery may itself be damaged; either of these would throw off the estimate, but probably not by many

orders of magnitude. If we just want to know whether the process will take milliseconds, minutes or months, then the turnover time estimate will likely be fine.

What about the box-of-air example? If there's a pulse of air on one side of the box, then the turnover time is potentially a good estimate of the time for pressures to equilibrate. BUT, this example highlights the key assumption: turnover time is a good estimate of equilibration time mainly when the parts turn over independently. If the air is at low enough pressure that the molecules aren't colliding too often (i.e. ideal gas approximation holds), then the estimate is probably good. But if the molecules are colliding often, then the equilibrium will mostly come from molecules pushing each other around, rather than moving around independently - and our turnover time estimate potentially breaks down.

**When parts change states roughly independently, turnover time tells us roughly the timescale on which the system equilibrates.**

# The Challenge

**Come up with 3 examples of approximately-independent turnover which do not resemble any you've seen before, and estimate their turnover times**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

**Any answer must include at least 3 to count**. That's the challenge. We're here to challenge ourselves, not just review examples we already know.

However, they don't have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

I will remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for:**

- Systems in which the parts change states more-or-less independently over time.
- Systems in which the parts are created and destroyed over time.
- If turnover is not independent, remember that it might be independent *conditional on* holding some variable fixed (like a feedback signal).

**Bonus Exercise**: for each of your three examples from the challenge, what kinds of things might interact with the system on timescales much shorter than the turnover time - i.e. on timescales at which the state-distribution is roughly constant? What kinds of things might interact with the system on timescales much longer than the turnover time - i.e. on timescales at which the system is (approximately) in equilibrium?

This bonus exercise is great blog-post fodder!

# Motivation

Much of the value I get from math is not from detailed calculations or elaborate models, but rather from *framing tools*: tools which suggest useful questions to ask, approximations to make, what to pay attention to and what to ignore.

Using a framing tool is sort of like using a trigger-action pattern: the hard part is to notice a pattern, a place where a particular tool can apply (the “trigger”). Once we notice the pattern, it suggests certain questions or approximations (the “action”). This challenge is meant to train the trigger-step: we look for novel examples to ingrain the abstract trigger pattern (separate from examples/contexts we already know).

The Bonus Exercise is meant to train the action-step: apply whatever questions/approximations the frame suggests, in order to build the reflex of applying them once we estimate a turnover time.

Hopefully, this will make it easier to notice when a turnover frame can be applied to a new problem you don’t understand in the wild, and to actually use it.

# Framing Practicum: Incentive

**Credit**

An enormous amount of credit goes to johnswentworth who made this new post possible.

This is a framing practicum post. We'll talk about what incentives are, how to recognize incentives in the wild, and what questions to ask when you find them. Then, we'll have a challenge to apply the idea.

**Today's challenge: come up with 3 examples of incentives which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

Expected time: ~15-30 minutes at most, including the Bonus Exercise.

# What Are Incentives?

At the beginning of a sowing season, the Government of India announces a list of guaranteed purchase prices for certain crops, e.g., rice, wheat, cotton, etc., to support farmers. In case the market price for a crop falls below the guaranteed purchase price, the government agencies purchase the entire quantity from farmers if the crop quality meets a minimum quality threshold. From an Indian farmer's perspective, the farmer is encouraged to produce price supported crops with quality just above the threshold level set by the government - and no higher.

This is an economic *incentive*:  There is a reward signal in the system. Farmers are rewarded for producing crops just above the quality threshold. On the other hand, they are not rewarded for producing higher quality crops. Here we see the defining features of incentives: A system (a farmer) "wants" some resource (money), and can get more of that resource in return for some actions (producing crops with quality just above the threshold level) than others (producing crops with quality well above the threshold level).

Another example, with a direct notion of "reward": cash incentive for taking Covid-19 vaccine shots. Some states in the US are offering rewards for Covid-19 vaccination in the form of direct cash or lottery programs. We can identify a clear reward signal in the system, which is people get rewarded for taking vaccines. Here again we see the defining features of incentives: A system (a human) "wants" some resource (money), and can get more of that resource in return for some actions (taking Covid-19 vaccines) than others (not taking Covid-19 vaccines).

# What To Look For

In general, incentives should come to mind whenever there is some kind of reward signal. A system "wants" some resource, and can get more of that resource in return for some actions  than others.

# Useful Questions To Ask

In the farmers support price example, the Government of India announces a minimum quality requirement for the crops to be purchased. Crops with lower quality will not qualify for the support program. On the other hand, farmers are not rewarded for having high quality crops. As a result, farmers will not only avoid the work required to produce higher quality crops, they will even make their crops' quality worse: farmers with high quality crops will mix small rocks, or leftover crops from previous years into their harvested crops to increase the total quantity of "crop", and thus total revenue. Obviously the Government of India did not intend for farmers to throw gravel into their crops, but they accidentally incentivized it anyway.

In general, whenever we see incentives, we should ask:

- **What actions are getting rewarded?**
- **What counterintuitive or unintended actions achieve high reward?**

What about the cash-reward-for-Covid-19-vaccine example? If there is someone who is urgently in need of money, that person might fake the vaccination status in order to receive rewards more than once.

# The Challenge

**Come up with 3 examples of incentives which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

**Any answer must include at least 3 to count, and they must be novel to you**. That's the challenge. We're here to challenge ourselves, not just review examples we already know.

However, they don't have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

I will remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for:**

- Environments in which there exists some kind of reward signal.
- Systems that “want” certain actions to be taken
- Agents that “want” some resource, and can get more of that resource in return for some actions than others.

**Bonus Exercise**: for each of your three examples from the challenge, explain:

- What other counterintuitive actions are getting rewarded?

This bonus exercise is great blog-post fodder!

# Motivation

Using a framing tool is sort of like using a trigger-action pattern: the hard part is to notice a pattern, a place where a particular tool can apply (the “trigger”). Once we notice the pattern, it suggests certain questions or approximations (the “action”). This challenge is meant to train the trigger-step: we look for novel examples to ingrain the abstract trigger pattern (separate from examples/contexts we already know).

The Bonus Exercise is meant to train the action-step: apply whatever questions/approximations the frame suggests, in order to build the reflex of applying them when we notice incentives.

Hopefully, this will make it easier to notice when an incentive frame can be applied to a new problem you don’t understand in the wild, and to actually use it.

# Framing Practicum: Selection Incentive

## Credit

An enormous amount of credit goes to johnswentworth who made this new post possible.

This is a framing practicum post. We'll talk about what selection incentives are, how to recognize selection incentives in the wild, and what questions to ask when you find them. Then, we'll have a challenge to apply the idea.

**Today's challenge: come up with 3 examples of selection incentives which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

Expected time: ~15-30 minutes at most, including the Bonus Exercise.

# What Are Selection Incentives?

Imagine trying to find great/popular posts on LessWrong. We look for things like high karma values, high number of comments, or a well-known writer. We don't really look at the contents of the individual posts (yet), we just look at an overall "score" that can help us to choose posts. This overall "score" mechanism encourages writers to write posts that could potentially achieve high "scores", for instance broad-interest posts, thought-provoking posts, controversial posts, etc, regardless of what the actual purpose of the writer is.

This is a *selection incentive*: Something is chosen based on some criteria or a known process. For instance, posts are chosen based on an overall "score": high karma values, high number of comments, etc. On the other hand, presenting ideas or transferring knowledge (not pursuing high karma values) might be what the writers actually want. But, the readers' selection criteria are there regardless of what writers actually want in the first place.

Another example is corporations maximizing profits. The founder of the corporation has something in mind, for instance sending humans to space, producing the most affordable cars to the mass population, etc., and they may or may not be trying to maximize profit.  What *happens* in the real business world, however, is that businesses live or die based on how well they maximize profits. Businesses are selected on the basis of how well they maximize profits, regardless of what the founders actually want.

# What To Look For

In general, selection incentives should spring to mind whenever something is chosen based on some criteria or a known process. We want to know what factors cause something to be more or less likely chosen. A few ways this can apply:

- We are selecting/choosing something based on some criteria.
- We see systems which grow or die, and we want to know what causes the system to grow/die faster or slower.

# Useful Questions To Ask

In the post selection example, posts with high karma values or high number of comments are more likely to be chosen than the ones with low karma values or low number of comments. But the post writers imagine a post that can present ideas/thoughts, transfer knowledge, or initiate a communication. High karma values may correlate with great posts, but may not align with what the writer actually wants, i.e., transfer knowledge and/or initiate conversation between the writer and the reader. What the writer actually wants diverges from what the selection criterion selects for/incentivizes.

In general, *if* an agent is involved, we want to know how the things the agent wants diverge from what the selection criteria "want".

# The Challenge

**Come up with 3 examples of selection incentives which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

**Any answer must include at least 3 to count, and they must be novel to you**. That's the challenge. We're here to challenge ourselves, not just review examples we already know.

However, they don't have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

I will remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for:**

- Systems in which something is chosen based on some criteria or a known process.
- Systems which grow/die and what makes them grow/die faster or slower.

**Bonus Exercise**: for each of your three examples from the challenge, explain:

- What strategies do the selection criteria incentivize for the things being selected?
- If an agent is being selected, how do the things the agent wants diverge from what the selection criteria are?
- If the agent is making the selection, how is it different from the selection criteria?

This bonus exercise is great blog-post fodder!

# Motivation

Using a framing tool is sort of like using a trigger-action pattern: the hard part is to notice a pattern, a place where a particular tool can apply (the “trigger”). Once we notice the pattern, it suggests certain questions or approximations (the “action”). This challenge is meant to train the trigger-step: we look for novel examples to ingrain the abstract trigger pattern (separate from examples/contexts we already know).

The Bonus Exercise is meant to train the action-step: apply whatever questions/approximations the frame suggests, in order to build the reflex of applying them when we notice selection incentives.

Hopefully, this will make it easier to notice when a selection incentive frame can be applied to a new problem you don’t understand in the wild, and to actually use it.

# Framing Practicum: Comparative Advantage

This is a framing practicum post. We'll talk about what comparative advantage is, how to recognize applications of comparative advantage in the wild, and what questions to ask when you find it. Then, we'll have a challenge to apply the idea.

**Today's challenge: come up with 3 examples of comparative advantage which you have *not* seen before**. **For each one, say what the different objectives are, and what the different components/subsystems are.** They don't need to be good, they don't need to be useful, they just need to be novel (to you).

("What do different objectives and subcomponents have to do with comparative advantage?" I hear you ask, "I don't remember anything about that from econ 101." This is framing practicum - we want to apply the frame of comparative advantage to new kinds of systems, not just trade-between-nations or whatever. So, we're going to present it a bit differently from what you're used to.)

Expected time: ~15-30 minutes at most, including the Bonus Exercise.

# What's Comparative Advantage?

Suppose we're running a fruit-growing company, Fruit Co, with many different orchards which can each grow apples or bananas. Each farm has different soil, different weather, different initial conditions, etc, and therefore faces different opportunity costs for growing each fruit. For instance, the Xenia site might be able to grow 1 extra unit of apples/yr at the cost of 0.5 units of bananas (by replacing their least-effective banana grove with apple trees) or vice versa, while the Yuma site may be able to grow 1 extra unit of apples/yr at the cost of 1 unit of bananas or vice versa.

| | ΔApples | ΔBananas |
|---|---|---|
| Xenia | 1 | : 0.5 |
| Yuma | 1 | : 1 |

Claim: given these numbers, the company can achieve a pareto increase in their fruit production. How? Well, they can produce one *more* unit of apples at the Xenia site (missing out on 0.5 units of bananas), and produce one *less* unit of apples at the Yuma site (using those resources to produce 1 extra unit of bananas). Overall, the amount of apples produced stays the same, but the amount of bananas produced increases by 0.5 units.

Intuitively: each site specializes a little more in whatever fruit they have a *comparative advantage* (aka *relative advantage*) in growing. We have multiple *goals* (growing more apples, and growing more bananas), and multiple *subsystems* which we can independently adjust to contribute to those goals (Xenia and Yuma orchards). Each subsystem faces different trade offs between the different goals, so we can make "trades" between subsystems with different trade off ratios in order to achieve

pareto gains. Each subsystem specializes a little more in whatever goal their trade off ratio favors, relative to the other subsystem.

We can also add more subsystems (e.g. Zion orchards), and more goals (e.g. coconut-growing). Maybe Xenia can trade off production in ratios of 1:0.5:3 (apples:bananas:coconuts), Yuma can trade off in ratios of 1:1:2, and Zion can trade off production in ratios of 1:0.5:1. We can pick *any* two sites, then pick *any* two fruits whose ratios differ between the sites, and do exactly the same sort of "trade" as before: each site specializes a bit more in whichever of the two fruits their ratio favors, compared to the other site. For instance, we could pick Xenia/Zion and apples/coconuts: Xenia could produce 3 more units of coconuts at the cost of 1 unit of apples, and Zion could replace those apples at the cost of just 1 unit of coconuts, so overall there's a gain of 2 coconuts.

| | ΔApples | ΔBananas | ΔCoconuts |
|---|---|---|---|
| Xenia | 1 | : 0.5 | : 3 |
| Yuma | 1 | : 1 | : 2 |
| Zion | 1 | : 0.5 | : 1 |

There are two ways this sort of "trade" *can't* be made:

- One site is already maximally specialized. For instance, if Zion is already fully specialized in growing apples, then there are no further banana or coconut groves to replace with apple trees.
- The two sites trade off in exactly the same ratios. For instance, Xenia and Zion both trade off apples:bananas at a ratio of 1:0.5, so we can't achieve a pareto gain with a little more specialization in those two fruits between those two sites.

# What To Look For

In general, comparative advantage should come to mind whenever we have an optimization problem with both

- Multiple goals/objectives (i.e. pareto optimality)
- Components/subsystems whose parameters can vary (approximately) independently

Note that "multiple goals" might really mean "multiple *sub*-goals" - e.g. Fruit Co might ultimately want to maximize profit, but producing more apples is a subgoal, producing more bananas is another subgoal, etc.

# Useful Questions To Ask

In the Fruit Co example, the key question is: what are the *ratios* at which different sites can trade off between production of different fruits? As long as those ratios are different, we can achieve a pareto gain.

More generally, we should ask: **what are the ratios at which different components/subsystems can trade off between different objectives?**

Another example: suppose we're designing a car. We have many objectives: speed, handling, cost, noise, comfort, etc. We have many subsystems which we can adjust approximately-independently: engine, transmission, body, seats, air conditioner, etc. So, we look at the ratios at which we can trade off speed:cost:noise by adjusting the engine, or the body, or the air conditioner. Can we achieve a 1-unit decrease in noise more cheaply by adjusting the engine or the air conditioner? Can we gain a bit of speed at the least noise-cost by adjusting the engine or the body? If these ratios differ, it often means we can achieve a pareto gain - e.g. maybe we can give the air conditioner team a bit of extra noise-budget (to make the air conditioner cheaper), and in exchange the engine team spends a little extra to cut back on engine noise, and that works out to a net decrease in both cost and noise.

# The Challenge

**Come up with 3 examples of comparative advantage which you have *not* seen before**. **For each one, say what the different objectives are, and what the different components/subsystems are.** They don't need to be good, they don't need to be useful, they just need to be novel (to you).

**Any answer must include at least 3 to count, and they must be novel to you**. That's the challenge. We're here to challenge ourselves, not just review examples we already know.

However, they don't have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in responses to other peoples' answers. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

~~**Post your answers inside of spoiler tags. (How do I do that?)**~~ [**EDIT:** I accidentally made this a normal post rather than a question post, and now there's responses so it's a bit late to switch. Ignore the spoiler requirement for this one.]

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

I will remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for optimization problems with *both* :**

- Multiple goals/objectives (i.e. pareto optimality)
- Components/subsystems whose parameters can vary independently

**Bonus Exercise**: for each of your three examples from the challenge, what are the ratios at which different components/subsystems can trade off between different objectives? I'm not looking for numerical values here, just a statement of what those

“ratios” mean within the context of your particular example. How might you measure the ratios?

This bonus exercise is great blog-post fodder!

# Framing Practicum: Semistable Equilibrium

*Thanks to John Wentworth for conceiving and executing the concept of a framing practicum, as well as much of the format and language of this post!*

This is a framing practicum post. We'll talk about what a semistable equilibrium is, how to recognize semistable equilibria in the wild, and what questions to ask when you find it. Then, we'll have a challenge to apply the idea.

**Today's challenge: come up with 3 examples of semistable equilibria which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

Expected time: ~15-30 minutes at most, including the Bonus Exercise.

## What's Semistable Equilibrium?

In Rebel Without a Cause, Jim Stark and Buzz Gunderson are racing their cars at top speed toward a cliff, with the gas pedal strapped down. The first to jump out of their car is chicken. Buzz's leather jacket gets caught on the door handle. He's unable to jump free, and plunges over the cliff to his death.

The cliff is a semistable equilibrium. The idea of the "chickie run" is that the cliff causes the racers to decelerate, so that their velocity approaches zero as they near the cliff's edge. That's how it works out for Jim. On the other side of the cliff, however, Buzz's velocity increases again as he hurtles toward the ground.

In general, a semistable equilibrium will approach an equilibrium point if it starts on one side, but will move away from the equilibrium point if it starts on the other side.

## What To Look For

A semistable equilibrium needs a threshold that attracts and slows things down if approached from one side, but repels or launches things away if they're on the other side. If there's a point "point of no return," that may be suggestive of a semistable equilibrium. There's a zone in which disturbances lead to a return to rest, and a second zone just beyond leading to ongoing activity. It's possible to have multiple equilibria in the system. All that's required is that there is a point that attracts things from one zone, but repels things in another adjacent zone.

Whether or not it is common to find the system at the equilibrium point will heavily depend on the direction and relative magnitude of disturbing forces.

# Useful Questions To Ask

Unlike with a stable equilibrium, the effect of nudges in a semistable equilibrium depend heavily on how close we are to the equilibrium point. If we're deep into the "zone of attraction" to the equilibrium point, nudges won't have much of an effect. But if we're near or at the equilibrium point, a small nudge could easily move us into the "zone of repulsion," leading to long-term instability in the system.

What happens if we change the equilibrium point? What are the disturbing forces in the system, and do they differ depending on where we are located relative to the equilibrium point? Can these disturbances "rescue us" from the zone of repulsion by bumping us back to the equilibrium point or into the zone of attraction? Does the zone of repulsion move us rapidly away from the equilibrium point, or is it a slower movement? How strong are the attractive and repulsive forces relative to any random disturbances in the system?

> Wearing a parachute significantly slows our movement through the "zone of instability!"

# The Challenge

**Come up with 3 examples of semistable equilibrium which do *not* resemble any you've seen before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

**Any answer must include at least 3 to count, and they must be novel to you**. That's the challenge. We're here to challenge ourselves, not just review examples we already know.

However, they don't have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

I will remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for:**

- Systems with a stopping or pause point, that is also a point of no return.
- Systems that show a combination of attraction and repulsion in a clearly directional manner.
- Systems that tend to slow us down to a stop as we approach a certain area, but move us faster if we go beyond it.

**Bonus Exercise**: for each of your three examples from the challenge, what forces might allow you to predict, measure or control the approach or repulsion from the equilibrium point? Is there some intervention or disturbance that might push us in one direction or the other if we are near the equilibrium point?

# Framing Practicum: General Factor

Credit to johnswentworth for being the general factor[1] of framing practicums.

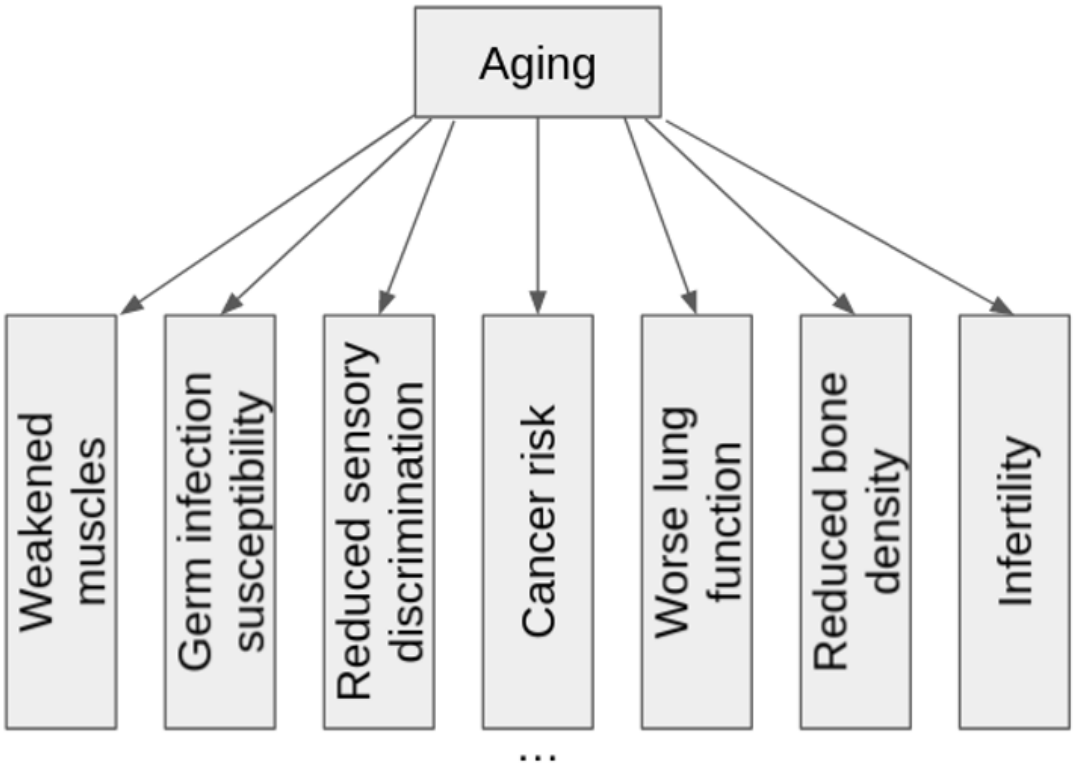
This is a framing practicum post. We'll talk about what general factors are, how to recognize general factors in the wild, and what questions to ask when you find them. Then, we'll have a challenge to apply the idea.

**Today's challenge: come up with 3 examples of general factors which you have *not* thought of as general factors before**. They don't need to be good, they don't need to be useful, they just need to be novel (to you).

Expected time: ~15-30 minutes at most, including the Bonus Exercise.

## What are general factors?

Consider the effects of aging on various bodily functions. Aging makes your muscles weaker, makes you more vulnerable to germs, makes your senses and mind worse, increases your risk of cancer, and so on. Aging is a common cause for a huge amount of health problems.




Any one of these effects would be relatively important, but the fact that aging has so many effects makes aging supremely important within contexts to do with health.

The importance of aging shows up in a lot of ways. One is a theoretical perspective; medical studies must make sure not to be confounded by aging, as otherwise their results will be completely uninterpretable. Another is an interventionist perspective; if we could control

aging, we could cure an enormous number of health problems. And there's also planning perspectives; it is more practical to do intense things when young than when older, as your body can better handle them while you are young.

Aging is not the traditional example of a general factor; a more canonical example would be the *g* factor of intelligence. *g* is an ability or set of abilities that people vary in, and it is required to various extents by just about any cognitive task you can come up with. When it comes to IQ tests, the specific skills that tap into *g* are often called **indicators** of *g*, though I will call them **outputs** to emphasize that they can be causally relevant too.

The cases where general factors become especially important are when their outputs have common effects. For instance, many outputs of aging, like weak immune systems or higher cancer rates, will also increase your mortality. These shared effects add up, and makes age the most important determinant of survival, as the likelihood of surviving the next year consistently drops with age.[2] Similarly, in order to perform well at school or at work, you need to solve a broad variety of cognitive tasks, making *g* particularly important for success.

# What to look for

Sometimes, it's fairly obvious that you have a general factor at hand, because you have some variable that you already know is an important influence on many other variables.

However, otherwise it is common when one sees a large set of correlated, related[3] variables to infer that their correlations are driven by one or more unobserved general factors.[4] Such a set of correlated variables is called a **positive manifold**.

Another good thing to be on the lookout for is when something is active in multiple contexts; if that is the case, then it will likely have similar effects in these multiple contexts, making it act like a general factor.

In general, if there is a variable that seems much more important than would be justified by its effects, and it feels like it is because of its *implications* or "the bigger picture", it may be that its importance stems from it being reflective of a general factor.

# Useful questions to ask

## Intervention

When someone claims to intervene on a domain, it is important not to mistake an intervention on the general factor from an intervention on its outputs. For instance, in the case of aging, some people say that exercise can reverse aging. And it's not entirely wrong that exercise can have widespread positive effects on your health, but at the same time the primary effects will be on the specific systems you exercise, such as the cardiovascular system and muscles, and will not "transfer" to your general health. It is, essentially, treating symptoms, and this makes it much less useful than it would be if it truly reversed aging itself.

## Correlation with outputs

Often, a general factor may not be directly or easily observable, while some of its outputs are readily observable. If one then wants to know about the factor, it may be useful to pay attention to its observable outputs. For instance, aging is associated with wrinkly skin and gray hair; these are not very important in and of themselves, but they provide a lot of information about a person's age, and therefore also about their general health.[5]

It may also be useful to think about the relationship between the general factor and the average of its outputs. Assuming that the general factor has a lot of independent outputs, one can for many purposes treat it as being *identical to* this average. However, some of the outputs of the general factor may not be known, or at least, not observed. Also, the general factor is strictly speaking not the same as its outputs; rather, it is the shared processes underlying the outputs. I often find that my understanding of some domain gets improved when I meditate on the distinction between the general factors and the averages of their outputs.

### Realism

If one has inferred the existence of a general factor from a widespread set of correlations between variables in a domain, and so doesn't know the root cause(s) of the factor, it can be enlightening to meditate on the realism of the general factor. Sometimes, there turns out to be a single core variable that mediates the common causes of everything, making the factor fully real.[6] On the other hand, sometimes there may be multiple overlapping wide-ranging root causes; then the general factor can be thought of as the sum of these causes.

But also commonly, people believe that the different outputs of the factor are mutually reinforcing, and that this is what is driving correlations. I think often people overestimate the relevance of mutual interactions. For instance, even if there are mutual interactions, there will often be a small "core" of interacting variables that drive most of the effect. And there may be factors influencing the general strengths of the mutual interactions, which may drive the overall dynamics.[7] And in the limit of a large number of homogeneous mutually interacting variables, each individual variable would only be able to have a small effect, while factors that influence many of the individual variables would be carried through, generating a true general factor.[8][9]

### Theory-building

Suppose you are studying human behavior. The problem is that behavior is highly chaotic and contextual; it's impossible to classify each individual interaction and model them all. But these chaotic interactions add up, and so things that are held in common across interactions become driving forces, which may be easier to classify and easier to model.

More generally, modelling something requires features, and factors provide a rich and convenient source of features for modelling.

## The Challenge

**Come up with 3 examples of general factors that you have not thought of as general factors before.** They don’t need to be good, they don’t need to be useful, they just need to be novel (to you). You can either take some observable variable that you know the effects of, where you thus *know* it functions as a general factor, or you can give examples of positive manifolds of correlated variables (which can be modelled as general factors, to various accuracies).

**Any answer must include at least 3 to count, and they must be novel to you**. That’s the challenge. We’re here to challenge ourselves, not just review examples we already know.

However, they don’t have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others' answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don't want people to leave this having back-chained the experience "If I go outside my comfort zone, people will look down on me". So be generous with those upvotes. I certainly will be.

If you comment on someone else's answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

Reward people for babbling — don't punish them for not pruning.

I will remove comments which I deem insufficiently kind, even if I believe they are otherwise valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for:**

- Cases where the same thing is present in multiple places or times
- Positive manifolds of consistently correlated variables

**Bonus Exercise**: for each of your three examples from the challenge, see if you can say something about one of the questions raised earlier in the post:

- Do people try to intervene on the variables, and if so do these interventions go through the general factor?
- Are there contexts where it seems reasonable to equate the factor with the average of its outputs? Or contexts where that could be misleading?
- Do some of the outputs provide a biased perspective of the underlying general factor value?
- If you've observed a positive manifold and inferred a general factor from this, do we have any knowledge or good guesses about how real the underlying general factor is?
- Are there chaotic interactions that can be approximated and simplified by understanding the underlying general factor?

You can pick and choose which question you want to answer for each of the examples you provide.

*Thanks to Justis Mills for proofreading and feedback.*

1. ^

   ;)

2. ^

   Further, many effects of aging will decrease your agency, which also adds up to make aging one of the most important determinants of agency.

3. ^

   I do not have a formal definition of relatedness; it is somewhat a matter of judgement. But I guess one example that can be said is, sometimes you "place" the factor in some physical location in the world; for instance, both aging and the *g* factor gets placed in an individual person. You'd then expect the outputs arrive from the same location as where you placed it.

4. ^

   A common critique here is that such a pattern of correlations could also be driven by mutual interactions; e.g. heat dissipates throughout objects, making the temperature

of one part of an object correlated with temperatures of other parts. This is sometimes an important point, but often there will be various things that make the system act as a general factor.

5. ^

   In some settings, it can also be useful to ask how the visible outputs you choose skew your perception of the factor. If there are certain groups where the outputs you look at act differently, then these groups might confuse you about their general factor. For instance in the case of aging, many people hide visible symptoms of aging to look more attractive. In psychometrics, there is a set of properties called "measurement invariance" which are designed to test for these exact problems.

6. ^

   In the case of aging, John Wentworth's overview points to a small core of aging, mainly related to reactive oxygen species.

7. ^

   For instance, with cognitive abilities, some people propose that being good at one thing provides a foundation, which gives one the resources and knowledge to transfer abilities to other abilities, e.g. via analogy. Well, maybe (as I understand, this notion is not well supported by research, but I am not an expert), but there are innate individual differences in e.g. ability to make analogies, which would become a common root cause driving these mutual interactions.

8. ^

   You can think of this as separation of scale being well-approximated by the large-scale influencing the small-scale. E.g. consider the temperature of different pieces of an object standing outside. The different pieces are mutually interacting, with heat dissipating around. However, mostly, the temperature is determined by the heat impacted by the sun, either by directly shining light on the object, or by heating up the overall surroundings.

9. ^

   One important context where this may fail is with long-tailed distributions or nonlinear interactions. Here, individual parts of a network can have strong effects in ways that are tightly related to the network shape.

# Framing Practicum: Dynamic Programming

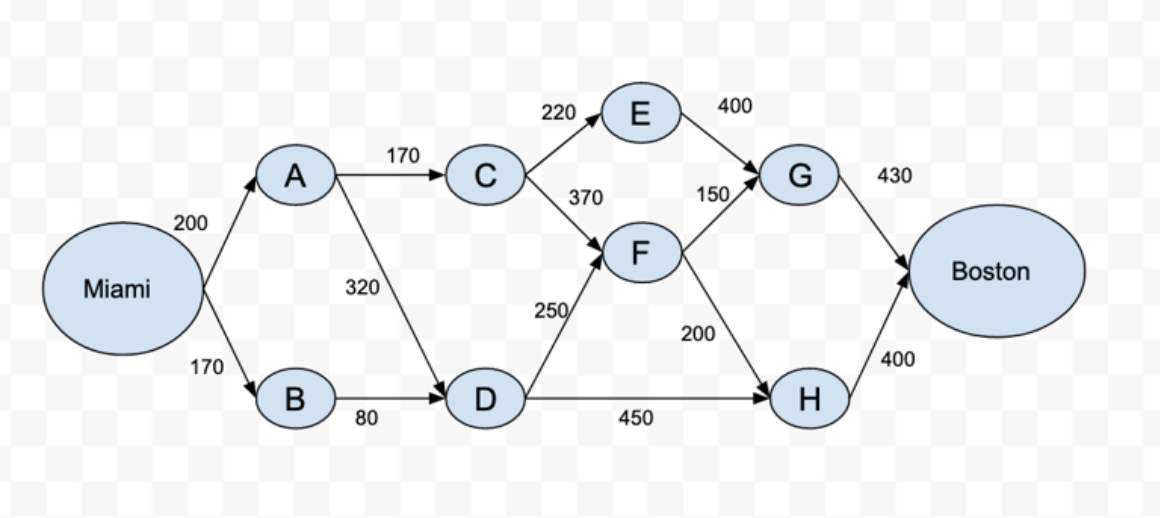
This is a framing practicum post. We'll talk about what dynamic programming (DP) is, how to recognize DP in the wild, and what questions to ask when you find it. Then, we'll have a challenge to apply the idea.

**Today's challenge: come up with 3 examples of DP which do *not* resemble any you've seen before.** They don't need to be good, they don't need to be useful, they just need to be novel (to you).

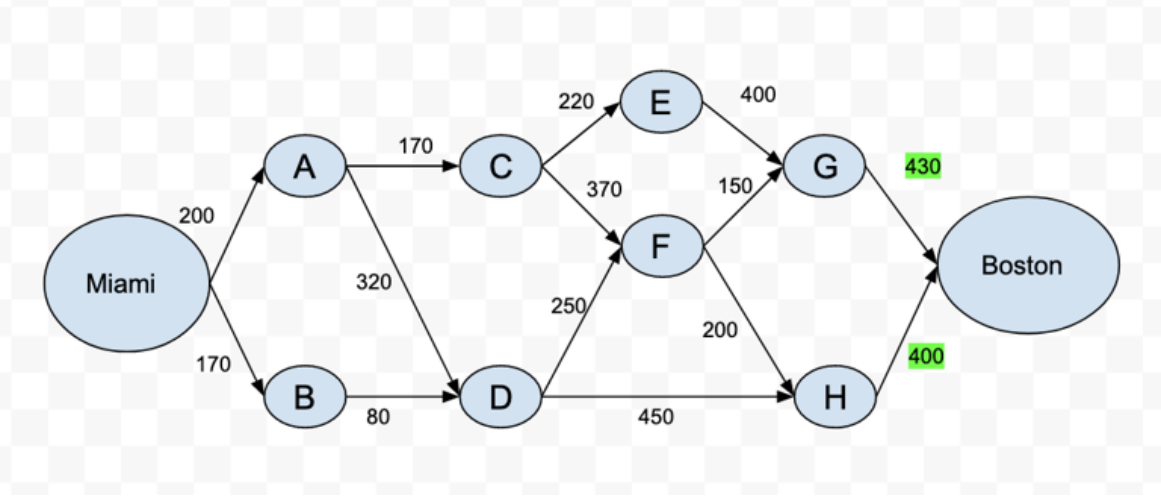
Expected time: ~15-30 minutes at most, including the Bonus Exercise.

## What is DP?

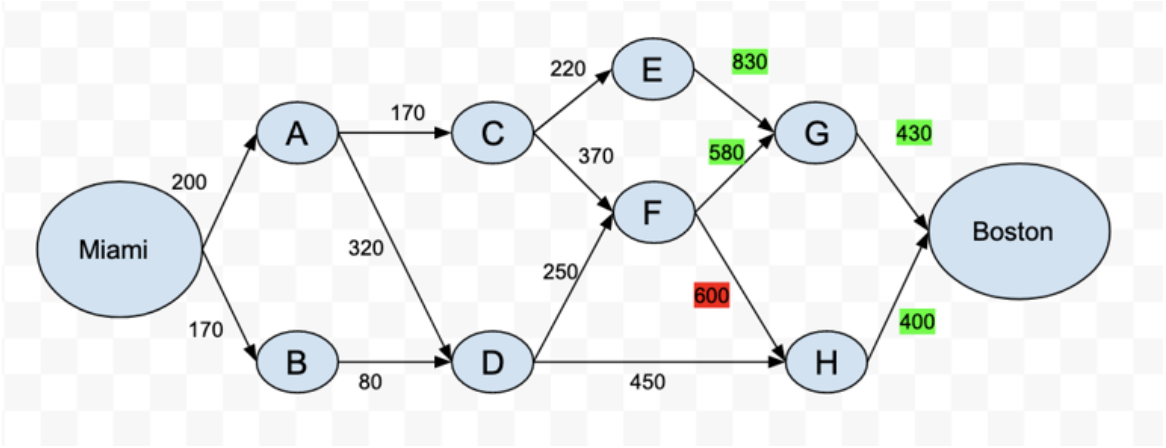
Suppose I am about to drive from Miami to Boston and I need to get to Boston as fast as possible. As a first step, I check the highway map and create a list of possible routes for this trip (let's assume "good" old times with no Google maps). For instance, looking at the imaginary routes in the figure below, the route at the top says I should take the "*Miami → A → C → E → G → Boston*" route and the total trip distance would be 200+170+220+400+430=1420 miles. Which specific route should I take to minimize the total distance, thus total travel time? I can, of course, calculate total travel distance for each possible route and pick the one with least-distance. But it could easily get very time consuming if there exist hundreds of thousands of possible routes to evaluate.



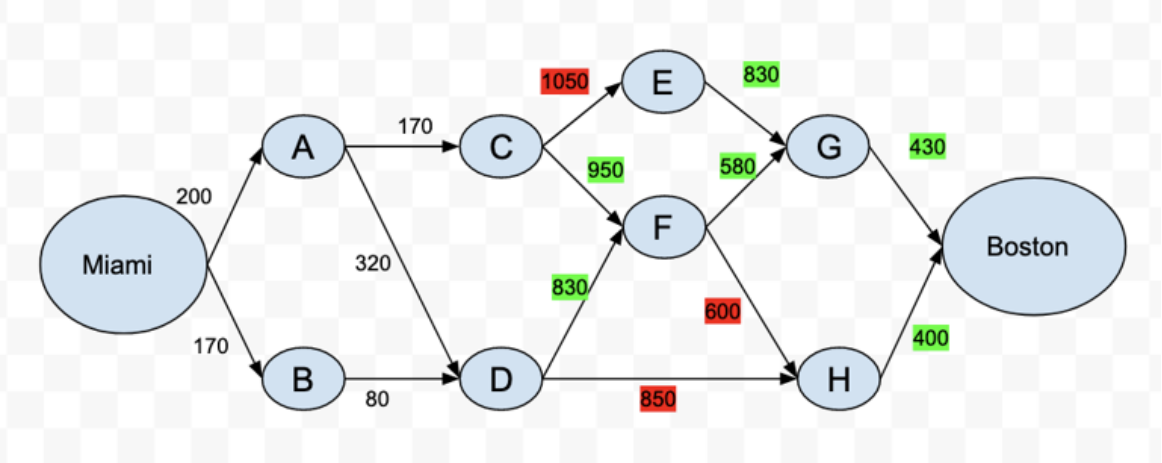

One alternative approach to identify a route with minimum distance is to use a backward method. Suppose I drive backward through the route map from right to left. First, I will start at the destination, Boston. If I am in city G or H, I have no further decision to make since there is only one route that leads me to Boston from either city. The number in green summarizes total distance with one last trip, or one stage, to go.



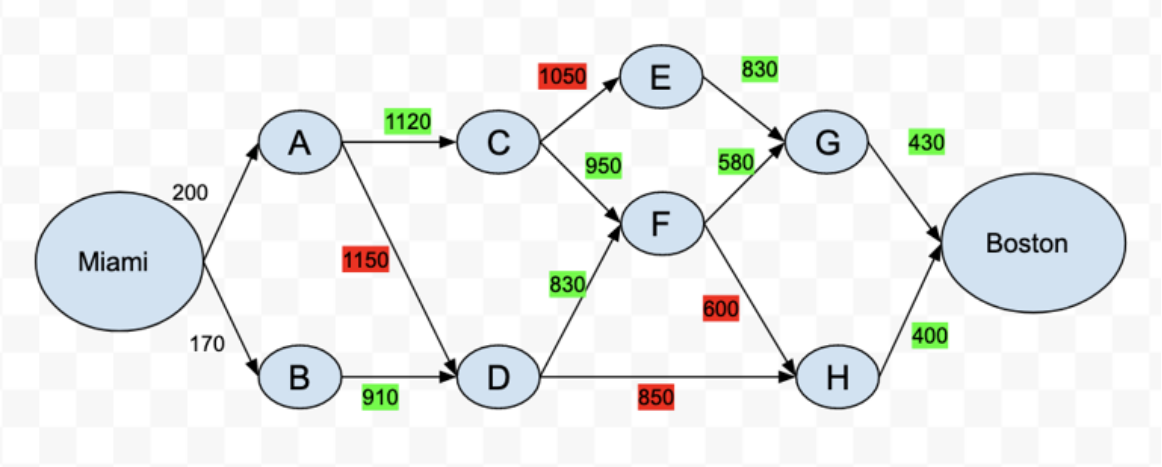

My first decision (from right to left) occurs with two trips, or stages to go. If, for example, I am in city F, I can either drive 150 miles to city G and another 430 miles to Boston - total 580 miles, *or* drive 200 miles to city H and another 400 miles to Boston - total 600 miles. Therefore, the shortest possible distance, or optimal route (in green), from city F is 580 miles (F → G → Boston). The alternative route from F (F → H → Boston) is suboptimal with 600 miles to go (in red).



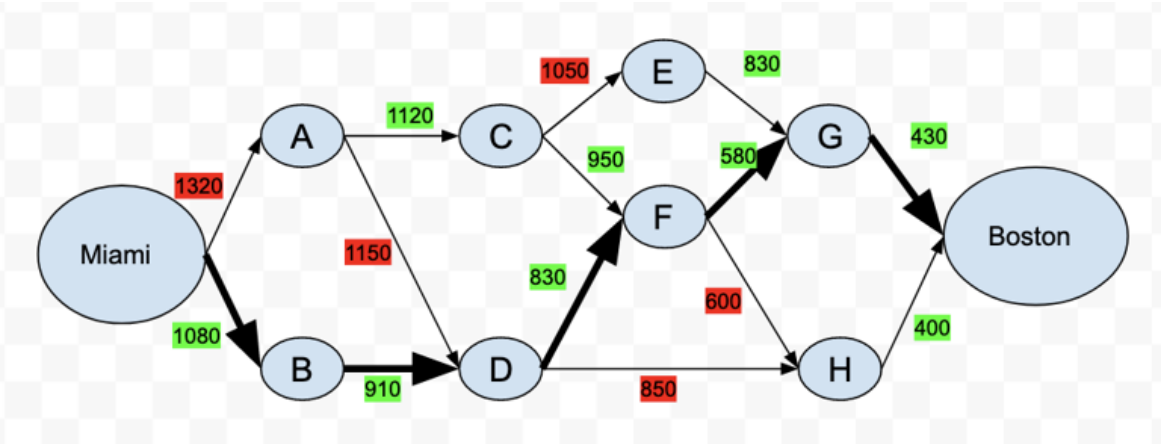

Let me back up one more city, or stage, and compute the least-distance from city C and D to Boston. Figure below summarizes these calculations.




Once I have computed the optimal route from city C and D onward to Boston, I can again move back one city and determine the optimal route from city A and B onward.




I continue this process and will end up with an optimal route with least-distance of 1080 miles to the problem (highlighted in bold arrows):




*This is DP:* An agent faces a multistage optimization problem (travelling from Miami to Boston by travelling through multiple cities). At each stage (e.g., I have one more trip to go to Boston), the agent might be in a different state (e.g., I am currently in city A or city B).

According to the current state (e.g., I am in city A), the agent takes a specific action (e.g., I will drive to city C) and as a result of that action, the system transitions to a different state in the next stage (e.g., now I am in city C). We solve the multistage optimization problem by working backwards, at each step computing the best reward we can get from that stage onward by considering each of the possible “next/child” stages.

# What To Look For?

DP should come to mind whenever an agent faces a problem with multi-stages nature and the agent takes a series of actions. Another defining feature of DP is that the original multi-stage complex problem can be dismantled into a sequence of simpler and smaller problems. The action the agent takes in a particular stage depends on the current state and the reward the agent would receive by taking that specific action. In addition, the action the agent takes impacts the state of the system, causing the system to transition to a new state.

# Useful Questions to Ask?

In the shortest driving time example, the ultimate goal is to minimize total driving time such that I can arrive at Boston as fast as possible. At any given time in my trip, I might be in a different state - the city I am in at that time. For instance, on the second day of the trip, I might be in city C or city D. Given my state,  I have a simpler problem to solve: What is the shortest travel time from city C or city D to Boston?

The system may start in different states. The agent takes a series of actions to optimize an objective across multiple stages. Each stage also has multiple states. The specific action an agent can take is a function of the state the agent currently in.

In general, whenever we see problems where DP is applicable, we should ask:

- What is the objective?
- What are the stages and states of the system?
- What are the actions the agent can take at any given state?
- How does a specific action change the state of the system?
- What is the value function? How is an agent rewarded for taking a particular action at the current stage?

# The Challenge

**Come up with 3 examples of DP which do *not* resemble any you’ve seen before**. They don’t need to be good, they don’t need to be useful, they just need to be novel (to you).

**Any answer must include at least 3 to count, and they must be novel to you**. That’s the challenge. We’re here to challenge ourselves, not just review examples we already know.

However, they don’t have to be very good answers or even correct answers. Posting wrong things on the internet is scary, but a very fast way to learn, and I will enforce a high bar for kindness in response-comments. I will personally default to upvoting every complete answer, even if parts of it are wrong, and I encourage others to do the same.

**Post your answers inside of spoiler tags. ( How do I do that? )**

**Celebrate others’ answers.** This is really important, especially for tougher questions. Sharing exercises in public is a scary experience. I don’t want people to leave this having

back-chained the experience “If I go outside my comfort zone, people will look down on me”. So be generous with those upvotes. I certainly will be.

If you comment on someone else’s answers, focus on making exciting, novel ideas work — instead of tearing apart worse ideas. Yes, And is encouraged.

I will remove comments which I deem insufficiently kind, even if I believe they are valuable comments. I want people to feel encouraged to try and fail here, and that means enforcing nicer norms than usual.

**If you get stuck, look for:**

- Problems with multistage nature.
- Problems that can be dismantled into a sequence of simpler problems.

**Bonus Exercise**: for each of your three examples from the challenge, explain:

- What are the simpler and smaller problems in the DP example?
- What are the states and how taking a specific action alters the state?
- what is the reward for taking a specific action on a given state?

This bonus exercise is great blog-post fodder!

# Motivation

Using a framing tool is sort of like using a trigger-action pattern: the hard part is to notice a pattern, a place where a particular tool can apply (the “trigger”). Once we notice the pattern, it suggests certain questions or approximations (the “action”). This challenge is meant to train the trigger-step: we look for novel examples to ingrain the abstract trigger pattern (separate from examples/contexts we already know).

The Bonus Exercise is meant to train the action-step: apply whatever questions/approximations the frame suggests, in order to build the reflex of applying them when we notice DP.

Hopefully, this will make it easier to notice when an DP frame can be applied to a new problem you don’t understand in the wild, and to actually use it.