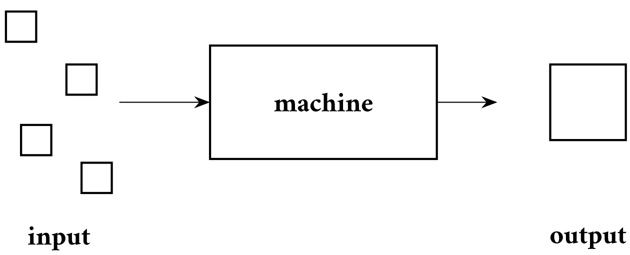
Logic, Reasoning, and Persuasion, Week 3-1

Today: Truth-Preservation and the Logic Machines.

1 | The Logic Machine

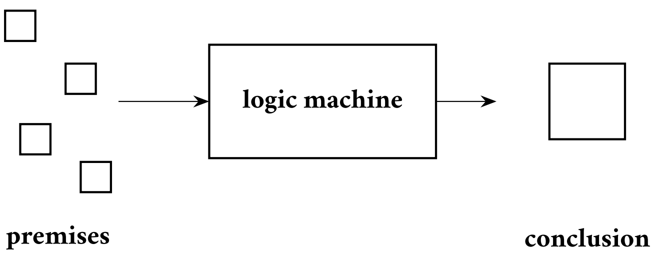
A machine is producing tools. You're on the quality assurance team. How do you know if the tool is good quality and should pass quality control? One thing you could do is just try it out. But sometimes it's really hard to figure out whether the tool is good quality just by observing it.

Here's another way to go about it. Look inside the machine and see if the machine is assembling tools correctly, and see if the materials that were put into the machine are high quality (after all, if someone put defective material into the machine, it might give defective outputs even if it functioned correctly).



Now, if for some of the building blocks, we also don't know if they pass quality control, we could go check and see if it itself comes out of a logic machine that is functioning properly, and is made out of high quality building blocks. Of course, at some point we have to be able to take the quality of some of the building blocks on trust, or by inspection: we can't check machines all the way down.

In the reasoning case, our building blocks are **premises**, our output is the **conclusion**, and the machine is a **logic machine**.



I said above that one way to check if the output is good is to check if the inputs are good and the machine is good. In the reasoning case, what does "good" mean?

- What makes a premise good quality is that it is **true**.
- What makes a logic machine good quality is that is **truth-preserving**: if you give it only true premises, it always gives you a true conclusion. If a logic machine is truth-preserving, then it *guarantees* a true conclusion when given true premises, no matter what fantastical hypotheticals we consider.

When we're evaluating an argument, we can think of ourselves as evaluating the quality of a conclusion in the same way as the quality-control inspector: by evaluating the quality of the logic machine and the premises. We want to check if the premises are *true*, and if the logic machine is *truth-preserving*.

1.1 | *Validity and Soundness*

You'll often hear arguments being called *sound* or *valid*, or talk of premises *entailing* conclusions. I mostly will *not* use this terminology, and talk about truth-preservation instead. But it's useful to be able to translate between different terminologies. So what do these mean?

1. An argument (a list of premises and a conclusion) is **valid** if there is a truth-preserving logic machine that builds the conclusion from the premises, so that if the premises are true, the conclusion is *guaranteed* to be true.
2. An argument is **sound** if it is valid and the premises are true (so if there is a truth preserving logic machine, and the premises are true).
3. The premises **entail** the conclusion if there is no way for the premises to be true and the conclusion to be false at the same time. Saying that the premises entail the conclusion is the same thing as saying there is a truth-preserving logic machine that builds the conclusion from the premises.

2 | The Truth-Preserving Logic Machines

As it turns out, every truth-preserving argument can be built up from true premises and strings of these eight truth-preserving machines.[1] You don't need to memorize these for class: I will write them out again if you need them. But it will be very useful to have an intuition about them and to know how to identify them.

2.1 | *Modus Ponens, the Implication Machine*

The implication machine takes the phrases "if … then …" and "… only if …" and spits out a conclusion. This should be familiar:

The Fed will lower interest rates.
*The Fed will lower interest rates **only if** economic growth has been slowing.*
*(**Therefore**,) economic growth has been slowing.*

While "if … then …" and "… only if …" have different contextual implications, they have the same logical format for the purposes of the implication machine. For example, here is a use of the "if … then …" formulation:

I'm going to the party.
***If** I'm going to the party, **then** my partner will come as a +1.*
*(**Therefore**,) my partner will come as a +1.*

Notice that this *feels* different than the following:

I'm going to the party.
*I'm going to the party, **only if** my partner will come as a +1.*
*(**Therefore**,) my partner will come as a +1.*

1. Later on we'll talk about arguments that can be good in other ways than being truth-preserving.

in the 'if … then …" formulation, it seems to be a statement of what will happen if the speaker goes to the party. In the "…only if…" formulation, it seems like the partner's coming along is a *condition* on going to the party. Nonetheless, for the purposes of the implication machine, we can infer the same things: either they both go to the party, or the speaker does not go.[2]

In general: let P and Q be placeholders for statements, and let $\implies$ abbreviate "Therefore". Then the implication machine permits the following inference:

$$\begin{bmatrix} P \\ \text{if } P \text{ then } Q \end{bmatrix} \implies Q \qquad \begin{bmatrix} P \\ P \text{ only if } Q \end{bmatrix} \implies Q \tag{1}$$

Exercise: illustrate how the following arguments could be produced by the implication machine, given the right premises. Example: suppose the sentence is

> Ray can't claim to have written his paper, because he used ChatGPT to generate the entire essay.

Then we can get "Ray can't claim to have written his paper" from the premise "he used ChatGPT to generate the entire essay" by adding an "if…then…" premise:

> (1) Ray used ChatGPT to generate his entire essay.
> (2) if Ray were to use ChatGPT to generate his entire essay, then he could not claim to have written it.[3]
> $\implies$ Ray can't claim to have written his paper.[4]

Exercises:

1. We can see smoke, so there must be a fire on the other side of the mountain.
2. If you're using AI to learn how to be an AI whisperer or prompt engineer, you can do that for free at home. So, why are you still here, if you're just cruising through school with AI?[5]
3. If you're the student, a key learning objective is for you to think critically about ethics (or whatever your course is about) and produce analysis *yourself* (Lin).
4. LLMs can't even get basic math right or accurately count. The usual response is that they're not designed to be calculators, which is true. But since mathematical reasoning appears closely related to critical thinking or logical reasoning, that's a big reason to be concerned about how well AI can actually "reason" and therefore throws every claim an AI makes into question (Lin).

2.2 | Modus Tollens, the Reverse Implication Machine

Suppose we know that if I'm going to the party, then my partner will come. If we also know that I'm going to the party, we can use modus ponens to conclude that my partner will come. But what if instead we know that my partner is *not* coming to the party. Then we can reason as follows:

2. Question: could the partner go by himself? Or does something in the statement preclude this?
3. a quote from Aylsworth and Castro, p. 5
4. Note that the wording is not exactly the same in the first and second premise
5. Patrick Lin, Why We're Not Using AI in This Course, Despite Its Obvious Benefits, https://emergingethics.substack.com/p/why-were-not-using-ai-in-this-course.

1. *If I'm going to the party, then my partner will come.*
2. *My partner is* not *coming.*
3. *So I must not be going to the party (since if I were, then my partner would come)*

The implication machine formalizes this. It takes an "if P then Q" statement and, if P is true, concludes that Q is true. The reverse implication machine takes that same "if P then Q" statement and reverses two things:

1. It changes truth to falsity
2. It switches P and Q.

This means that it takes that "if P then Q" and, if Q is *false*, concludes that P is also false.

Let P and Q be placeholders for statements, and let NOT mean "it is not the case that" or "it is not true that". So NOT[Adrian is in the classroom] is "It is not the case that Adrian is in the classroom." Then the reverse implication machine permits the following inference:

$$\begin{bmatrix} \textsc{not}Q \\ \text{if } P \text{ then } Q \end{bmatrix} \implies \textsc{not}P \qquad \begin{bmatrix} \textsc{not}Q \\ P \text{ only if } Q \end{bmatrix} \implies \textsc{not}P \tag{2}$$

2.3 | The Implication Chain Machine

Intuitively, you can chain true implications, like so:

1. If Gabrielle gets ice cream, then Lorenzo will get ice cream.
2. And if Lorenzo gets ice cream, then Brett will get ice cream.
3. So if Gabrielle gets ice cream, then Brett will get ice cream.

If implications "chain together", where each one implies the next one, then we can directly infer from the first one to the last one.

Let P, Q, R be placeholders for statements. Then the implication chain machine permits the following inference:

$$\begin{bmatrix} \text{if } P \text{ then } Q \\ \text{if } Q \text{ then } R \end{bmatrix} \implies \text{if } P \text{ then } R \qquad \begin{bmatrix} P \text{ only if } Q \\ Q \text{ only if } R \end{bmatrix} \implies P \text{ only if } R \tag{3}$$

Exercises: For each of the following, identify an intermediate premise that can be used in the implication chain machine to create an implication chain. For instance, suppose the sentence is:

If I have coffee so late, I will be sleepy tomorrow!

What does having coffee today have to do with being sleepy tomorrow? Perhaps it's an explanation like that given by these two premises:

If I have coffee so late, I won't be able to sleep tonight.

If I am not able to sleep tonight, I will be sleepy tomorrow.

So that using the implication machine, we can write:

If I have coffee so late, **then** I won't be able to sleep tonight.

If I am not able to sleep tonight, **then** I will be sleepy tomorrow.

So, **if** I have coffee so late, **then** I will be sleepy tomorrow.

Exercises:

1. Gabrielle will go to the party only if they have ice cream there.[6]
2. If a future employer treats your bachelor's degree in English or philosophy as evidence of your writing ability, your wanton use of ChatGPT makes you (and perhaps the university) guilty of false advertising.[7]
3. If you learn with AI, or have AI whisper in your ear, you could have more interesting conversations with other people.[8]

2.4 | The Gluestick Machine

If you have several true statements, and you stick them together to make a bigger statement with "and"s, the bigger statement is also true. So if I like purple, and Gabrielle likes blue, and Clem likes orange, then the three statements about what colors we like can be stuck together to make a bigger statement:

> *I like purple, and Gabrielle likes blue, and Clem likes orange.*

If you're noting that it looks exactly the same: *that's how good the glue is*. It also means the gluestick machine is one of those ones that seems so obvious we shouldn't need a machine for it (we'll see more of these below). But we need a machine for everything!

Here's an implementation of the gluestick, from A&C. First, we have three statements:

- *You have a duty to cultivate your autonomy.*
- *Writing your own college humanities papers is a good and unique opportunity to cultivate your autonomy.*
- *You do not have a good reason to pass on the opportunity to cultivate your autonomy by writing your own humanities papers.*

The gluestick machine allows us to stick these together into one big statement that says the same thing:

> *You have a duty to cultivate your autonomy, writing your own college humanities papers is a good and unique opportunity to cultivate your autonomy, and you do not have a good reason to pass on the opportunity to cultivate your autonomy by writing your own humanities papers.*

In general, let $P, Q, R, \ldots$ be placeholders for statements. Then the gluestick machine allows the following inference:

$$\begin{bmatrix} P \\ Q \\ R \\ \vdots \end{bmatrix} \implies P \text{ and } Q \text{ and } R \cdots \tag{4}$$

Exercise: Show how these statements could be derived from the gluestick machine:

1. While outlines and bullet points are faster to read and make it easier to see how the discussion flows, stripping down an article to its bare bones can flatten

6. My friend Gabrielle, who loves ice cream.
7. Aylsworth and Castro, 8.
8. Lin, "Why We're Not Using AI In This Course."

the discussion too much, causing nuance or details to be lost and moving too quickly to be digested.[9]

2. Although we may not remember quite as much as our ancestors who participated in oral traditions, reading and writing allow us to produce and pass on a far greater quantity of knowledge.[10]

Challenge: The gluestick machine can technically create sentences that involve words like "but", "however," and "although." It treats them all as "and." Intuitively, this loses some of the meaning of the sentences. What is lost?

2.5 | The Double Implication Machine

We have to either take the Rutgers bus or call an Uber. If we take the Rutgers bus, we'll be late. If we call an Uber, it will cost money. So either we'll be late, or we'll spend money.

In this sort of utterly familiar reasoning, we take one "this or that" statement and we derive another "this or that" statement, like this:

$$\begin{array}{ccccc} P & \text{if } P \text{ then } Q & \Longrightarrow & Q \\ \text{or} & & & \text{or} \\ R & \text{if } R \text{ then } S & \Longrightarrow & S \end{array}$$

In general, let P, Q, R and S be placeholders for statements. Then the double implication machine permits the following inference:

$$\begin{bmatrix} \text{either } P \text{ or } R, \text{ or both} \\ \text{if } P, \text{then } Q \\ \text{if } R, \text{then } S \end{bmatrix} \Longrightarrow \text{either } Q \text{ or } S, \text{ or both} \tag{5}$$

2.6 | The Process of Elimination Machine

The Process of Elimination Machine works like answering a multiple-choice question via process of elimination. If in a group of statements, we know that *at least one* is true, then if we've determined for all but one that they are false, then we can conclude that the remaining one is true. This is simple: you all do it in multiple-choice exams. If we know that at least one of Emily, Fernanda, George, and Harry signed up for the union, then we can reason as follows:

1. *Emily didn't sign up for the union*
2. *George didn't sign up for the union*
3. *Fernanda didn't sign up for the union.*
4. *Therefore, Harry must have signed up for the union.*

Here's the schematic for two choices: Let P and Q be placeholders for statements.

$$\begin{bmatrix} \text{either } P \text{ or } Q, \text{ or both} \\ \textsc{not } P \end{bmatrix} \Longrightarrow Q \tag{6}$$

9. Lin, "Why We're Not Using AI in This Course."
10. Aylsworth and Castro.

2.7 | *The Forgetting Machine*

Arjun told you yesterday that he likes cheesecake and chocolate cake. Today Priya asks you whether Arjun likes chocolate cake. Obviously yes: you know from yesterday that Arjun likes chocolate cake. How do we know? Well, Arjun likes cheesecake and chocolate cake. So for the purposes of Priya's question, we can forget about the cheesecake and just remember the chocolate cake part:

Arjun likes cheesecake and chocolate cake.
So Arjun likes chocolate cake.

This is what the forgetting machine does. In general, let P and Q be placeholders for statements. Then the forgetting machine allows the following inferences:

$$[P \text{ and } Q] \implies Q \qquad\qquad [P \text{ and } Q] \implies P \tag{7}$$

Like the redundancy machine below, this is one of the machines that gives us a "complete set" of machines. But we won't use it much in argument analysis because it does not tend to be used explicitly.

2.8 | *The Redundancy Machine*

The Redundancy Machine takes a statement, and then adds *any other statement* and says that one of the two statements is true. For example:

Ani is at the party.
***Therefore**, either Ani is at the party or Freya is at the party, or both.*

To be honest, you won't see this one much, since it's usually used implicitly. You probably don't have to remember it.

In general, let P and Q be placeholders for statements. Then the redundancy machine allows the following inferences:

$$[P] \implies \text{either } P \text{ or } Q, \text{ or both} \tag{8}$$

Exercise: come up with a real-world use case in which the redundancy machine is relevant. Do the same for the forgetting machine.

3 | Normative Implication Machines

Often arguments are "normative". They are about what you *should* or *ought* to do or believe, where this sense can be *moral* or *nonmoral*. Whenever words like "ought" or "should" or "duty" pop up, we might have a normative argument on our hands.

3.1 | *The Normative Implication Machine*

Consider the following reasoning:

You have a duty to cultivate your autonomy.
You cultivate your autonomy only if you write your own papers.
Therefore, you have a duty to write your own papers.

Question: Is this an instance of the implication machine? **Answer:** No. (Why not?)

If we were to straightforwardly implement the implication machine, we would have to have something like the following:

> *You have a duty to cultivate your autonomy.*
> *You **have a duty to** cultivate your autonomy only if you **have a duty to** write your own papers.*
> *Therefore, you have a duty to write your own papers.*

But this is a very different argument! The second premise in the new argument relates two things that you have a duty to do, whereas what we originally wanted was something relating two things that you might *actually* do.

When we have normative language, we need another version of the implication machine, which tells us where the normative language is allowed to be.
Let A and B be placeholders for some actions. Then the normative implication machine permits the following inference:

$$\begin{bmatrix} \text{You ought to do } A \\ \text{You do } A \text{ only if you do } B \end{bmatrix} \implies \text{ you ought to do } B$$

Throughout, you can replace "You ought to" with "You should" or "You have a moral duty to", or "You are morally required to".

3.2 | The Normative Reverse Implication Machine

If we again let A and B be placeholders for some actions, then the analogue for "ought" is as follows (although it may not be completely obvious why this is true):

$$\begin{bmatrix} \text{You ought not do } B \\ \text{If you do } A\text{, then you do } B \end{bmatrix} \implies \text{ you ought not do } A$$

3.3 | The Normative Implication Chain Machine

We don't need a normative implication chain machine. All we need is the normative implication machine and the normal implication chain machine. **Why is this?**

3.4 | The Normative Glue Stick Machine

Arguably, the normative gluestick machine is *not* truth-preserving. Can you think of a counterexample?

3.5 | Other Normative Implication Machines

Exercise: Do the following exist? Do they need to?

1. The Normative Double Implication Machine
2. The Normative Process of Elimination Machine
3. The Normative Forgetting Machine
4. The Normative Redundancy Machine