PHIL 222
Philosophical Foundations of Computer Science
Week 13, Tuesday

Nov. 19, 2024

Basic Probability Theory

A **probability space** consists of three components:

As an example let's take throwing a die.

A **probability space** consists of three components:

As an example let's take throwing a die.

1. **Sample space**. This is the set of all possible outcomes.
   The sample space is the set of all outcomes: $\{1, 2, 3, 4, 5, 6\}$.

A **probability space** consists of three components:

As an example let's take throwing a die.

1. **Sample space**. This is the set of all possible outcomes.
   The sample space is the set of all outcomes: $\{1, 2, 3, 4, 5, 6\}$.
2. **Events**. All subsets of the sample space. Typically represented by a common property.
   The event "The outcome is even" corresponds to the set $\{2, 4, 6\}$.

A **probability space** consists of three components:

As an example let's take throwing a die.

1. **Sample space**. This is the set of all possible outcomes.
   The sample space is the set of all outcomes: $\{1, 2, 3, 4, 5, 6\}$.
2. **Events**. All subsets of the sample space. Typically represented by a common property.
   The event "The outcome is even" corresponds to the set $\{2, 4, 6\}$.
3. **Probability measure**. This is a function which assigns to each event a probability between 0 and 1.
   If the die is fair, the probability measure is "uniform", meaning that each outcome $i = 1, \ldots, 6$ (or a one-element event $\{i\}$) has equal probability, i.e. $1/6$.

A probability measure P must satisfy the following axiom.

- For every event A (a subset of the sample space S), the probability of A is the sum of probabilities of outcomes in A.

Example: throwing a fair die. The event “You throw an even number” has probability $P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = 3 \cdot \frac{1}{6} = \frac{1}{2}$.

A probability measure P must satisfy the following axiom.

- For every event A (a subset of the sample space S), the probability of A is the sum of probabilities of outcomes in A.

Example: throwing a fair die. The event "You throw an even number" has probability $P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = 3 \cdot \frac{1}{6} = \frac{1}{2}$.

In general, two events A and B are called **disjoint** if no outcome belongs to both A and B (or, in notation, $A \cap B$).

E.g., "You throw an even number" and "You throw a 1" are disjoint.

A probability measure P must satisfy the following axiom.

- For every event A (a subset of the sample space S), the probability of A is the sum of probabilities of outcomes in A.

Example: throwing a fair die. The event "You throw an even number" has probability $P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = 3 \cdot \frac{1}{6} = \frac{1}{2}$.

In general, two events A and B are called **disjoint** if no outcome belongs to both A and B (or, in notation, $A \cap B$).
E.g., "You throw an even number" and "You throw a 1" are disjoint.

Then the axiom can also be stated as follows:

- For pairwise disjoint events $A_1, \ldots, A_n$, the probability that A_1 or ... or A_n happens (or, in notation, $A_1 \cup \cdots \cup A_n$) equals the sum

$$P(A_1) + \cdots + P(A_n).$$

E.g., "You throw either an even number or a 1" has probablility $\frac{1}{2} + \frac{1}{6} = \frac{2}{3}$.

Now watch the following YouTube videos (in this order):

1. *Basics of Probability: Unions, Intersections, and Complements*
   https://www.youtube.com/watch?v=B1v9OeCTlu0&list=PLvxOuBpazmsOGOursPoofaHyz_1NpxbhA
2. *An Introduction to Conditional Probability*
   https://www.youtube.com/watch?v=bgCMjHzXTXs&list=PLvxOuBpazmsOGOursPoofaHyz_1NpxbhA&index=2
3. *Independent Events (Basics of Probability: Independence of Two Events)*
   https://www.youtube.com/watch?v=1wuRV5z0PPE&list=PLvxOuBpazmsOGOursPoofaHyz_1NpxbhA&index=3

Metaphysics (1)
Time and (In/non)determinism

AMIR PNUELI

United States – 1996

CITATION

For seminal work introducing temporal logic into computing science and for outstanding contributions to program and system verification.

SHORT ANNOTATED BIBLIOGRAPHY

ACM TURING AWARD LECTURE

RESEARCH SUBJECTS

BIRTH:

April 22, 1941, Nahalal, Israel.

DEATH:

November 2, 2009 (aged 68) New York, USA.

EDUCATION:

B.Sc. (with distinction) Mathematics (Technion, 1962); Ph.D (with distinction), Applied Mathematics

Amir Pnueli (pronounced: p'new-EL-ee) was born on April 22, 1941, in Nahalal, Israel. His parents, Henya and Prof. Shmuel Yeshayahu ("Shay") Pnueli, immigrated to Israel, which was then Palestine, in 1936. They settled in Nahala, a cooperative agricultural community, where Shay Pnueli was the principal of the local school and Henya Pnueli was a teacher. In 1945, when Shay was appointed to teach in a teachers' college in the kibbutz Giva'at Hashlosha, the family relocated to Hulon. In the 1960s Prof. Shay Pnueli became one of the founders of Tel-Aviv University, and chaired the Hebrew literature department there until his death in 1965. Henya Pnueli continued teaching in Gordon school at Hulon until she died at the age of 75 in 1996.

https://amturing.acm.org/award_winners/pnueli_4725172.cfm/

Amir left the Weizmann Institute to found, and then chair, the department of computer science at Tel Aviv University, where he stayed until 1980. It was during that period that Amir got deeply involved in logics and deductive methods. During a sabbatical at the University of Pennsylvania he was introduced to the work of the philosopher Arthur Prior, who had developed "tense logic" to evaluate statements whose truthfulness changes over time. Amir was the first to realize the potential implications of applying Prior's work to computer programs. Amir's 1977 seminal paper "The Temporal Logic of Programs" [1] revolutionized the way computer programs are analyzed. At the time, practical program verification was widely considered to be hopeless. The main methodologies considered all possible pairs of program states. Amir's paper introduced the notion of reasoning about programs as execution paths, which breathed new life into the field of program verification.

https://amturing.acm.org/award_winners/pnueli_4725172.cfm/

To quote Amir from his talk after receiving the Israel Prize:

> "In mathematics, logic is static. It deals with connections among entities that exist in the same time frame. When one designs a dynamic computer system that has to react to ever changing conditions, ... one cannot design the system based on a static view. It is necessary to characterize and describe dynamic behaviors that connect entities, events, and reactions at different time points. Temporal Logic deals therefore with a dynamic view of the world that evolves over time." (Translated from the original Hebrew)

A. N. Prior, *Past, Present and Future* (1967):

> The usefulness of systems of this sort [that uses discrete time] does not depend on any serious metaphysical assumption that time *is* discrete; they are applicable in limited fields of discourse in which we are concerned only with what happens next in a sequence of discrete states, e.g. in the working of a digital computer. [p. 67]

Prior's **temporal logic** or **tense logic**:

Prior's **temporal logic** or **tense logic**:

It introduces the following phrases:

- "It has always been the case that φ",
- "It will always be the case that φ",
- "It has at some time been the case that φ",
- "It will at some time be the case that φ", or
  "It may be the case at some future point that φ".

Prior's **temporal logic** or **tense logic**:

It introduces the following phrases:

- "It has always been the case that φ",
- "It will always be the case that φ",
- "It has at some time been the case that φ",
- "It will at some time be the case that φ", or
  "It may be the case at some future point that φ".

Then we can entertain such axioms as:

a. $\varphi \implies$ it will always be the case that it has at some time been the case that φ.

b. $\varphi \implies$ it has always been the case that it would at some time be the case that φ.

The Problem of Future Contingents

In his *De Interpretatione* (or *On Interpretation*), Aristotle discusses whether we can coherently maintain that there are contingent things that may or may not occur in the future.

The Problem of Future Contingents

In his *De Interpretatione* (or *On Interpretation*), Aristotle discusses whether we can coherently maintain that there are contingent things that may or may not occur in the future.

First, a bit of reflection on logic:

> I call an affirmation and a negation contradictory opposites when what one signifies universally the other signifies not universally, e.g. every man is white — not every man is white, no man is white — some man is white. [17b17–21]

> It is evident that a single affirmation has a single negation. For the negation must deny the same thing as the affirmation affirmed, and of the same thing, whether a particular or a universal (taken either universally or not universally). I mean, for example, Socrates is white — Socrates is not white. [17b38–18a3]

The Problem of Future Contingents

In his *De Interpretatione* (or *On Interpretation*), Aristotle discusses whether we can coherently maintain that there are contingent things that may or may not occur in the future.

First, a bit of reflection on logic:

> I call an affirmation and a negation contradictory opposites when what one signifies universally the other signifies not universally, e.g. every man is white — not every man is white, no man is white — some man is white. [17b17–21]

> It is evident that a single affirmation has a single negation. For the negation must deny the same thing as the affirmation affirmed, and of the same thing, whether a particular or a universal (taken either universally or not universally). I mean, for example, Socrates is white — Socrates is not white. [17b38–18a3]

Affirmative and negative sentences may form a pair in which they are negations of each other (in modern terms), as in φ and not-φ.

Then, in chapter 9, Aristotle gives a controversial thesis and a difficult argument. (Lots of discussion by ancient + medieval + contemporary commentators, even as to what Aristotle really meant to conclude!)

Then, in chapter 9, Aristotle gives a controversial thesis and a difficult argument. (Lots of discussion by ancient + medieval + contemporary commentators, even as to what Aristotle really meant to conclude!)

First, the thesis:

> With regard to what is and what has been it is necessary for the affirmation or the negation to be true or false. [. . .] But with particulars that are going to be it is different. [18a28–33]

Then, in chapter 9, Aristotle gives a controversial thesis and a difficult argument. (Lots of discussion by ancient + medieval + contemporary commentators, even as to what Aristotle really meant to conclude!)

First, the thesis:

> With regard to what is and what has been it is necessary for the affirmation or the negation to be true or false. [. . .] But with particulars that are going to be it is different. [18a28–33]

Then the proof starts by supposing that the thesis does not hold.

> For if every affirmation or negation is true or false [. . .] [18a34]

Then, in chapter 9, Aristotle gives a controversial thesis and a difficult argument. (Lots of discussion by ancient + medieval + contemporary commentators, even as to what Aristotle really meant to conclude!)

First, the thesis:

> With regard to what is and what has been it is necessary for the affirmation or the negation to be true or false. [. . .] But with particulars that are going to be it is different. [18a28–33]

Then the proof starts by supposing that the thesis does not hold.

> For if every affirmation or negation is true or false [. . .] [18a34]

The bad thing that follows from the supposition is that

> it is necessary for everything either to be the case or not to be the case. For [. . .]. [18a35–36]

Let's keep reading . . .

> For if one person says that something will be and another denies this same thing, it is clearly necessary for one of them to be saying what is true — if every affirmation is true or false; for both will not be the case together under such circumstances.

Let's keep reading . . .

> For if one person says that something will be and another denies this same thing, it is clearly necessary for one of them to be saying what is true — if every affirmation is true or false; for both will not be the case together under such circumstances. For if it is true to say that it is white or is not white, it is necessary for it to be white or not white; and if it is white or is not white, then it was true to say or deny this. If it is not the case it is false, if it is false it is not the case. So it is necessary for the affirmation or the negation to be true.

Let's keep reading . . .

> For if one person says that something will be and another denies this same thing, it is clearly necessary for one of them to be saying what is true — if every affirmation is true or false; for both will not be the case together under such circumstances. For if it is true to say that it is white or is not white, it is necessary for it to be white or not white; and if it is white or is not white, then it was true to say or deny this. If it is not the case it is false, if it is false it is not the case. So it is necessary for the affirmation or the negation to be true. It follows that nothing either is or is happening, or will be or will not be, by chance or as chance has it, but everything of necessity and not as chance has it (since either he who says or he who denies is saying what is true). For otherwise it might equally well happen or not happen, since what is as chance has it is no more thus than not thus, nor will it be. [18a36–18b9]

The core part of Aristotle's argument may be:

> Again, if it is white now it was true to say earlier that it would be white; so that it was always true to say of anything that has happened that it would be so. [18b10-18b12]

The core part of Aristotle's argument may be:

> Again, if it is white now it was true to say earlier that it would be white; so that it was always true to say of anything that has happened that it would be so. [18b10-18b12]

Note that this is exactly the axiom b we mentioned earlier!

The core part of Aristotle's argument may be:

> Again, if it is white now it was true to say earlier that it would be white; so that it was always true to say of anything that has happened that it would be so. [18b10-18b12]

Note that this is exactly the axiom b we mentioned earlier!

> But if it was always true to say that it was so, or would be so, it could not not be so, or not be going to be so. But if something cannot not happen it is impossible for it not to happen; and if it is impossible for something not to happen it is necessary for it to happen. Everything that will be, therefore, happens necessarily. So nothing will come about as chance has it or by chance; for if by chance, not of necessity. [18b12-18b16]

The core part of Aristotle's argument may be:

> Again, if it is white now it was true to say earlier that it would be white; so that it was always true to say of anything that has happened that it would be so. [18b10-18b12]

Note that this is exactly the axiom b we mentioned earlier!

> But if it was always true to say that it was so, or would be so, it could not not be so, or not be going to be so. But if something cannot not happen it is impossible for it not to happen; and if it is impossible for something not to happen it is necessary for it to happen. Everything that will be, therefore, happens necessarily. So nothing will come about as chance has it or by chance; for if by chance, not of necessity. [18b12-18b16]

Thus there can be no future contingency: if φ will be the case then it will necessarily be the case.

But then absurdity follows:

> Nor, however, can we say that neither is true — that it neither will be nor will not be so. For, firstly, though the affirmation is false the negation is not true, and though the negation is false the affirmation, on this view, is not true. Moreover, if it is true to say that something is white and large, both have to hold of it, and if true that they will hold tomorrow, they will have to hold tomorrow; and if it neither will be nor will not be the case tomorrow, then there is no 'as chance has it'. Take a sea-battle: it would have neither to happen nor not to happen. [18b17-18b25]

Aristotle *seems* to say this (again, a lot of interpretations!):

Aristotle *seems* to say this (again, a lot of interpretations!):

Write

φ for "There will be a sea battle tomorrow",
ψ for "There will not be a sea battle tomorrow".

Aristotle *seems* to say this (again, a lot of interpretations!):

Write

φ for “There will be a sea battle tomorrow”,
ψ for “There will not be a sea battle tomorrow”.

Then ψ is not-φ.

Aristotle *seems* to say this (again, a lot of interpretations!):

Write

φ for "There will be a sea battle tomorrow",
ψ for "There will not be a sea battle tomorrow".

Then ψ is not-φ.

Moreover,

i φ is true and ψ is false $\Longrightarrow$ iii φ is necessarily true,

ii φ is false and ψ is true $\Longrightarrow$ iv ψ is necessarily true.

Aristotle *seems* to say this (again, a lot of interpretations!):

Write

$$\varphi \text{ for "There will be a sea battle tomorrow",}$$
$$\psi \text{ for "There will not be a sea battle tomorrow".}$$

Then ψ is not-φ.

Moreover,

$$\text{(i) } \varphi \text{ is true and } \psi \text{ is false} \implies \text{(iii) } \varphi \text{ is necessarily true,}$$
$$\text{(ii) } \varphi \text{ is false and } \psi \text{ is true} \implies \text{(iv) } \psi \text{ is necessarily true.}$$

So
$$\text{(i) or (ii)} \implies \text{(iii) or (iv),}$$

Aristotle *seems* to say this (again, a lot of interpretations!):

Write

φ for "There will be a sea battle tomorrow",
ψ for "There will not be a sea battle tomorrow".

Then ψ is not-φ.

Moreover,

$$\text{(i) } \varphi \text{ is true and } \psi \text{ is false} \implies \text{(iii) } \varphi \text{ is necessarily true,}$$

$$\text{(ii) } \varphi \text{ is false and } \psi \text{ is true} \implies \text{(iv) } \psi \text{ is necessarily true.}$$

So

$$\text{(i) or (ii)} \implies \text{(iii) or (iv),}$$

$$\text{neither (i) nor (ii)} \impliedby \text{neither (iii) nor (iv).}$$

Aristotle *seems* to say this (again, a lot of interpretations!):

Write

$$\begin{aligned} &\varphi \text{ for "There will be a sea battle tomorrow"},\\ &\psi \text{ for "There will not be a sea battle tomorrow"}. \end{aligned}$$

Then ψ is not-φ.

Moreover,

$$\begin{aligned} \text{(i) } \varphi \text{ is true and } \psi \text{ is false} &\implies \text{(iii) } \varphi \text{ is necessarily true},\\ \text{(ii) } \varphi \text{ is false and } \psi \text{ is true} &\implies \text{(iv) } \psi \text{ is necessarily true}. \end{aligned}$$

So

$$\begin{aligned} \text{(i) or (ii)} &\implies \text{(iii) or (iv)},\\ \text{neither (i) nor (ii)} &\impliedby \text{neither (iii) nor (iv)}. \end{aligned}$$

But neither φ nor ψ is necessarily true.

Therefore neither φ nor ψ is true.

$$\begin{aligned}
&\text{(i) } \varphi \text{ is true and } \psi \text{ is false} &&\Longrightarrow&& \text{(iii) } \varphi \text{ is necessarily true,}\\
&\text{(ii) } \varphi \text{ is false and } \psi \text{ is true} &&\Longrightarrow&& \text{(iv) } \psi \text{ is necessarily true.}\\
&\text{(i) or (ii)} &&\Longrightarrow&& \text{(iii) or (iv),}\\
&\text{neither (i) nor (ii)} &&\Longleftarrow&& \text{neither (iii) nor (iv).}
\end{aligned}$$

$$\begin{aligned}
\text{(i) } \varphi \text{ is true and } \psi \text{ is false} &\implies \text{(iii) } \varphi \text{ is necessarily true,} \\
\text{(ii) } \varphi \text{ is false and } \psi \text{ is true} &\implies \text{(iv) } \psi \text{ is necessarily true.} \\
\text{(i) or (ii)} &\implies \text{(iii) or (iv),} \\
\text{neither (i) nor (ii)} &\impliedby \text{neither (iii) nor (iv).}
\end{aligned}$$

A lot of interpretations of and reactions to this argument, including a lot of similar but different arguments showing the above.

$$
\begin{aligned}
\text{(i) } \varphi \text{ is true and } \psi \text{ is false} &\implies \text{(iii) } \varphi \text{ is necessarily true,}\\
\text{(ii) } \varphi \text{ is false and } \psi \text{ is true} &\implies \text{(iv) } \psi \text{ is necessarily true.}\\
\text{(i) or (ii)} &\implies \text{(iii) or (iv),}\\
\text{neither (i) nor (ii)} &\impliedby \text{neither (iii) nor (iv).}
\end{aligned}
$$

A lot of interpretations of and reactions to this argument, including a lot of similar but different arguments showing the above. E.g.,

- The Stoics adhered to the principle that every proposition is either true or false, and accepted that the future is predetermined.

$$
\begin{aligned}
\text{(i) } \varphi \text{ is true and } \psi \text{ is false} &\implies \text{(iii) } \varphi \text{ is necessarily true,}\\
\text{(ii) } \varphi \text{ is false and } \psi \text{ is true} &\implies \text{(iv) } \psi \text{ is necessarily true.}\\
\text{(i) or (ii)} &\implies \text{(iii) or (iv),}\\
\text{neither (i) nor (ii)} &\impliedby \text{neither (iii) nor (iv).}
\end{aligned}
$$

A lot of interpretations of and reactions to this argument, including a lot of similar but different arguments showing the above. E.g.,

- The Stoics adhered to the principle that every proposition is either true or false, and accepted that the future is predetermined.
- Medieval Christian philosophers faced great difficulty because

$$
\begin{aligned}
\text{(i) } \varphi \text{ is true and } \psi \text{ is false} &\implies \text{(iii) } \varphi \text{ is necessarily true,}\\
\text{(ii) } \varphi \text{ is false and } \psi \text{ is true} &\implies \text{(iv) } \psi \text{ is necessarily true.}\\
\text{(i) or (ii)} &\implies \text{(iii) or (iv),}\\
\text{neither (i) nor (ii)} &\impliedby \text{neither (iii) nor (iv).}
\end{aligned}
$$

A lot of interpretations of and reactions to this argument, including a lot of similar but different arguments showing the above. E.g.,

- The Stoics adhered to the principle that every proposition is either true or false, and accepted that the future is predetermined.
- Medieval Christian philosophers faced great difficulty because
  - predetermination ((iii) or (iv)) apparently precludes free will,

$$\begin{aligned}
\text{(i) } \varphi \text{ is true and } \psi \text{ is false} &\implies \text{(iii) } \varphi \text{ is necessarily true,}\\
\text{(ii) } \varphi \text{ is false and } \psi \text{ is true} &\implies \text{(iv) } \psi \text{ is necessarily true.}\\
\text{(i) or (ii)} &\implies \text{(iii) or (iv),}\\
\text{neither (i) nor (ii)} &\impliedby \text{neither (iii) nor (iv).}
\end{aligned}$$

A lot of interpretations of and reactions to this argument, including a lot of similar but different arguments showing the above. E.g.,

- The Stoics adhered to the principle that every proposition is either true or false, and accepted that the future is predetermined.
- Medieval Christian philosophers faced great difficulty because
  - predetermination ((iii) or (iv)) apparently precludes free will,
  - God is supposed to know whether φ is true or not ((i) or (ii)).

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

1. Either E is going to take place tomorrow or non-E is going to take place tomorrow.
2. If a proposition about the past is true, then it is now necessary, i.e., inescapable or unpreventable.
3. If E is going to take place tomorrow, then it is true that yesterday it was the case that E would take place in two days.
4. If it is now necessary that yesterday E would take place in two days, then it is now necessary that E is going to take place tomorrow.

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

1. Either E is going to take place tomorrow or non-E is going to take place tomorrow.
2. If a proposition about the past is true, then it is now necessary, i.e., inescapable or unpreventable.
3. If E is going to take place tomorrow, then it is true that yesterday it was the case that E would take place in two days.
4. If it is now necessary that yesterday E would take place in two days, then it is now necessary that E is going to take place tomorrow.

I. If E is going to take place tomorrow, then it is now necessary that yesterday E would take place in two days. (By 2 + 3)

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

1. Either *E* is going to take place tomorrow or non-*E* is going to take place tomorrow.
2. If a proposition about the past is true, then it is now necessary, i.e., inescapable or unpreventable.
3. If *E* is going to take place tomorrow, then it is true that yesterday it was the case that *E* would take place in two days.
4. If it is now necessary that yesterday *E* would take place in two days, then it is now necessary that *E* is going to take place tomorrow.

I. If *E* is going to take place tomorrow, then it is now necessary that yesterday *E* would take place in two days. (By 2 + 3)

II. If *E* is going to take place tomorrow, then *E* is necessarily going to take place tomorrow. (By 4 + I)

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

1. Either E is going to take place tomorrow or non-E is going to take place tomorrow.
2. If a proposition about the past is true, then it is now necessary, i.e., inescapable or unpreventable.
3. If E is going to take place tomorrow, then it is true that yesterday it was the case that E would take place in two days.
4. If it is now necessary that yesterday E would take place in two days, then it is now necessary that E is going to take place tomorrow.

- I. If E is going to take place tomorrow, then it is now necessary that yesterday E would take place in two days. (By 2 + 3)
- II. If E is going to take place tomorrow, then E is necessarily going to take place tomorrow. (By 4 + I)
- III. If non-E is going to take place tomorrow, then non-E is necessarily going to take place tomorrow. (Similar to II)

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

1. Either E is going to take place tomorrow or non-E is going to take place tomorrow.
2. If E is going to take place tomorrow, then E is necessarily going to take place tomorrow.
3. If non-E is going to take place tomorrow, then non-E is necessarily going to take place tomorrow.

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

1. Either E is going to take place tomorrow or non-E is going to take place tomorrow.
2. If E is going to take place tomorrow, then E is necessarily going to take place tomorrow.
3. If non-E is going to take place tomorrow, then non-E is necessarily going to take place tomorrow.
4. Either E is necessarily going to take place tomorrow or non-E is necessarily going to take place tomorrow. (By 1 + II + III)

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

1. Either E is going to take place tomorrow or non-E is going to take place tomorrow.
2. If E is going to take place tomorrow, then E is necessarily going to take place tomorrow.
3. If non-E is going to take place tomorrow, then non-E is necessarily going to take place tomorrow.
4. Either E is necessarily going to take place tomorrow or non-E is necessarily going to take place tomorrow. (By 1 + 2 + 3)
5. Therefore, what is going to happen tomorrow is going to happen with necessity. (By 4)

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

1. Either *E* is going to take place tomorrow or non-*E* is going to take place tomorrow.
2. If a proposition about the past is true, then it is now necessary, i.e., inescapable or unpreventable.
3. If *E* is going to take place tomorrow, then it is true that yesterday it was the case that *E* would take place in two days.
4. If it is now necessary that yesterday *E* would take place in two days, then it is now necessary that *E* is going to take place tomorrow.

One of those arguments (formulated by Richard Lavenham, c. 1380, as summarized by Øhrstrøm & Hasle):

1. Either *E* is going to take place tomorrow or non-*E* is going to take place tomorrow.
2. If a proposition about the past is true, then it is now necessary, i.e., inescapable or unpreventable.
3. If *E* is going to take place tomorrow, then it is true that yesterday it was the case that *E* would take place in two days.
4. If it is now necessary that yesterday *E* would take place in two days, then it is now necessary that *E* is going to take place tomorrow.

IV. Either *E* is necessarily going to take place tomorrow or non-*E* is necessarily going to take place tomorrow.

V. Therefore, what is going to happen tomorrow is going to happen with necessity.

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that ..."
- Was_d for "It was the case d (duration) ago that ..."
- $\mathrm{Nec}_{\mathrm{now}}$ for "It is now necessary that ..."

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that ..."
- Was_d for "It was the case d (duration) ago that ..."
- Nec_{now} for "It is now necessary that ..."

1. $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.
2. If-$(\text{Was}_d : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.
3. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.
4. If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

IV. $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that ..."
- Was_d for "It was the case d (duration) ago that ..."
- Nec_{now} for "It is now necessary that ..."

1. $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.
2. If-$(\text{Was}_d : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.
3. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.
4. If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

IV. $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that ..."
- Was_d for "It was the case d (duration) ago that ..."
- Nec_{now} for "It is now necessary that ..."

1. $(\text{Will}_{\text{1-day}} : \varphi)\text{-or-}(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.
2. $\text{If-}(\text{Was}_d : \psi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.
3. $\text{If-}(\text{Will}_{\text{1-day}} : \varphi)\text{-then-}(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.
4. $\text{If-}(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

I. $\text{If-}(\text{Will}_{\text{1-day}} : \varphi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (2 + 3).

IV. $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)\text{-or-}(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that . . ."
- Was_d for "It was the case d (duration) ago that . . ."
- Nec_{now} for "It is now necessary that . . ."

1. $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.
2. If-$(\text{Was}_d : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.
3. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.
4. If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

I. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (2 + 3).

IV. $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that . . . "
- Was_d for "It was the case d (duration) ago that . . . "
- Nec_{now} for "It is now necessary that . . . "

1 $(\text{Will}_{\text{1-day}} : \varphi)\text{-or-}(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.

2 $\text{If-}(\text{Was}_d : \psi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.

3 $\text{If-}(\text{Will}_{\text{1-day}} : \varphi)\text{-then-}(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.

4 $\text{If-}(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

I $\text{If-}(\text{Will}_{\text{1-day}} : \varphi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (2 + 3).

II $\text{If-}(\text{Will}_{\text{1-day}} : \varphi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$ (4 + I).

IV $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)\text{-or-}(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that ..."
- Was_d for "It was the case d (duration) ago that ..."
- Nec_{now} for "It is now necessary that ..."

❶ $(\text{Will}_{\text{1-day}} : \varphi)\text{-or-}(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.

❷ $\text{If-}(\text{Was}_d : \psi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.

❸ $\text{If-}(\text{Will}_{\text{1-day}} : \varphi)\text{-then-}(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.

❹ $\text{If-}(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

Ⓘ $\text{If-}(\text{Will}_{\text{1-day}} : \varphi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (❷ + ❸).

Ⓘⓘ $\text{If-}(\text{Will}_{\text{1-day}} : \varphi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$ (❹ + Ⓘ).

Ⓘⓘⓘ $\text{If-}(\text{Will}_{\text{1-day}} : \text{not-}\varphi)\text{-then-}(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to Ⓘⓘ).

Ⓘⓥ $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)\text{-or-}(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that . . . "
- Was_d for "It was the case d (duration) ago that . . . "
- Nec_{now} for "It is now necessary that . . . "

❶ $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.

❷ If-$(\text{Was}_d : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.

❸ If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.

❹ If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

Ⅰ If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (❷ + ❸).

Ⅱ If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$ (❹ + Ⅰ).

Ⅲ If-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to Ⅱ).

Ⅳ $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that ..."
- Was_d for "It was the case d (duration) ago that ..."
- Nec_{now} for "It is now necessary that ..."

❶ $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.

❷ If-$(\text{Was}_d : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.

❸ If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.

❹ If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

Ⓘ If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (❷ + ❸).

Ⓘⓘ If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$ (❹ + Ⓘ).

Ⓘⓘⓘ If-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to Ⓘⓘ).

Ⓘⓥ $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$

Temporal Logic Analysis

We can formalize this argument with the vocabulary of temporal logic:

- Will_d for "It will be the case in d (duration) that ..."
- Was_d for "It was the case d (duration) ago that ..."
- $\mathrm{Nec}_{\mathrm{now}}$ for "It is now necessary that ..."

❶ $(\mathrm{Will}_{\text{1-day}} : \varphi)$-or-$(\mathrm{Will}_{\text{1-day}} : \text{not-}\varphi)$.

❷ If-$(\mathrm{Was}_d : \psi)$-then-$(\mathrm{Nec}_{\mathrm{now}} : \mathrm{Was}_d : \psi)$.

❸ If-$(\mathrm{Will}_{\text{1-day}} : \varphi)$-then-$(\mathrm{Was}_{\text{1-day}} : \mathrm{Will}_{\text{2-days}} : \varphi)$.

❹ If-$(\mathrm{Nec}_{\mathrm{now}} : \mathrm{Was}_{\text{1-day}} : \mathrm{Will}_{\text{2-days}} : \varphi)$-then-$(\mathrm{Nec}_{\mathrm{now}} : \mathrm{Will}_{\text{1-day}} : \varphi)$.

Ⅰ If-$(\mathrm{Will}_{\text{1-day}} : \varphi)$-then-$(\mathrm{Nec}_{\mathrm{now}} : \mathrm{Was}_{\text{1-day}} : \mathrm{Will}_{\text{2-days}} : \varphi)$ (❷ + ❸).

Ⅱ If-$(\mathrm{Will}_{\text{1-day}} : \varphi)$-then-$(\mathrm{Nec}_{\mathrm{now}} : \mathrm{Will}_{\text{1-day}} : \varphi)$ (❹ + Ⅰ).

Ⅲ If-$(\mathrm{Will}_{\text{1-day}} : \text{not-}\varphi)$-then-$(\mathrm{Nec}_{\mathrm{now}} : \mathrm{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to Ⅱ).

Ⅳ $(\mathrm{Nec}_{\mathrm{now}} : \mathrm{Will}_{\text{1-day}} : \varphi)$-or-$(\mathrm{Nec}_{\mathrm{now}} : \mathrm{Will}_{\text{1-day}} : \text{not-}\varphi)$ (❶ + Ⅱ + Ⅲ).

1 $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.

2 If-$(\text{Was}_{d} : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_{d} : \psi)$.

3 If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.

4 If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

I If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (2 + 3).

II If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$ (4 + I).

III If-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to II).

IV $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (1 + II + III).

❶ $(\mathrm{Will}_{\text{1-day}} : \varphi)$-or-$(\mathrm{Will}_{\text{1-day}} : \text{not-}\varphi)$.

❷ If-$(\mathrm{Was}_{d} : \psi)$-then-$(\mathrm{Nec}_{\text{now}} : \mathrm{Was}_{d} : \psi)$.

❸ If-$(\mathrm{Will}_{\text{1-day}} : \varphi)$-then-$(\mathrm{Was}_{\text{1-day}} : \mathrm{Will}_{\text{2-days}} : \varphi)$.

❹ If-$(\mathrm{Nec}_{\text{now}} : \mathrm{Was}_{\text{1-day}} : \mathrm{Will}_{\text{2-days}} : \varphi)$-then-$(\mathrm{Nec}_{\text{now}} : \mathrm{Will}_{\text{1-day}} : \varphi)$.

Ⅰ If-$(\mathrm{Will}_{\text{1-day}} : \varphi)$-then-$(\mathrm{Nec}_{\text{now}} : \mathrm{Was}_{\text{1-day}} : \mathrm{Will}_{\text{2-days}} : \varphi)$ (❷ + ❸).

Ⅱ If-$(\mathrm{Will}_{\text{1-day}} : \varphi)$-then-$(\mathrm{Nec}_{\text{now}} : \mathrm{Will}_{\text{1-day}} : \varphi)$ (❹ + Ⅰ).

Ⅲ If-$(\mathrm{Will}_{\text{1-day}} : \text{not-}\varphi)$-then-$(\mathrm{Nec}_{\text{now}} : \mathrm{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to Ⅱ).

Ⅳ $(\mathrm{Nec}_{\text{now}} : \mathrm{Will}_{\text{1-day}} : \varphi)$-or-$(\mathrm{Nec}_{\text{now}} : \mathrm{Will}_{\text{1-day}} : \text{not-}\varphi)$ (❶ + Ⅱ + Ⅲ).

To the argument above, there are roughly four schools of response:

1. $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.
2. If-$(\text{Was}_d : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.
3. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.
4. If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

I. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (2 + 3).

II. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$ (4 + I).

III. If-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to II).

IV. $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (1 + II + III).

To the argument above, there are roughly four schools of response:

A. Accept all the premises and conclusions.

1. $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.
2. If-$(\text{Was}_d : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.
3. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.
4. If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

I. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (2 + 3).
II. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$ (4 + I).
III. If-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to II).
IV. $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (1 + II + III).

To the argument above, there are roughly four schools of response:

A. Accept all the premises and conclusions.
B. Reject just 1, that statements about the future have truth values, accept 2–4. (Aristotle himself seems to be taking this position.)

❶ $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.

❷ If-$(\text{Was}_d : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.

❸ If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.

❹ If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

Ⓘ If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (❷ + ❸).

Ⓘⓘ If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$ (❹ + Ⓘ).

Ⓘⓘⓘ If-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to Ⓘⓘ).

Ⓘⓥ $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (❶ + Ⓘⓘ + Ⓘⓘⓘ).

To the argument above, there are roughly four schools of response:

Ⓐ Accept all the premises and conclusions.

Ⓑ Reject just ❶, that statements about the future have truth values, accept ❷–❹. (Aristotle himself seems to be taking this position.)

Ⓒ **"Peircean"**: Accept ❷ & ❹, reject ❸ in addition to ❶.

1. $(\text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$.
2. If-$(\text{Was}_d : \psi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_d : \psi)$.
3. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$.
4. If-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$.

I. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Was}_{\text{1-day}} : \text{Will}_{\text{2-days}} : \varphi)$ (2 + 3).
II. If-$(\text{Will}_{\text{1-day}} : \varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$ (4 + I).
III. If-$(\text{Will}_{\text{1-day}} : \text{not-}\varphi)$-then-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (similar to II).
IV. $(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \varphi)$-or-$(\text{Nec}_{\text{now}} : \text{Will}_{\text{1-day}} : \text{not-}\varphi)$ (1 + II + III).

To the argument above, there are roughly four schools of response:

A. Accept all the premises and conclusions.
B. Reject just 1, that statements about the future have truth values, accept 2–4. (Aristotle himself seems to be taking this position.)
C. **"Peircean"**: Accept 2 & 4, reject 3 in addition to 1.
D. **"Ockhamist"**: Accept 1 & 3 & 4, reject 2.

So both “Peircean” and “Ockhamist” accounts avoid this particular argument. But about each account you may ask:

So both “Peircean” and “Ockhamist” accounts avoid this particular argument. But about each account you may ask:

- Surely there can be other arguments to the same conclusion from premises that the account accepts, no?

So both "Peircean" and "Ockhamist" accounts avoid this particular argument. But about each account you may ask:

- Surely there can be other arguments to the same conclusion from premises that the account accepts, no?
- In fact, is the account consistent? E.g., premise x may perhaps entail premise y. In that case you cannot consistently reject premise y while accepting premise x.

So both “Peircean” and “Ockhamist” accounts avoid this particular argument. But about each account you may ask:

- Surely there can be other arguments to the same conclusion from premises that the account accepts, no?
- In fact, is the account consistent? E.g., premise x may perhaps entail premise y. In that case you cannot consistently reject premise y while accepting premise x.
- Even if the account is consistent in the sense above, the premise which the account rejects seems plausible. Can the account show in what way the premise can fail?

So both “Peircean” and “Ockhamist” accounts avoid this particular argument. But about each account you may ask:

- Surely there can be other arguments to the same conclusion from premises that the account accepts, no?
- In fact, is the account consistent? E.g., premise x may perhaps entail premise y. In that case you cannot consistently reject premise y while accepting premise x.
- Even if the account is consistent in the sense above, the premise which the account rejects seems plausible. Can the account show in what way the premise can fail?

To answer these questions, the truth-condition method is a great help!

So both "Peircean" and "Ockhamist" accounts avoid this particular argument. But about each account you may ask:

- Surely there can be other arguments to the same conclusion from premises that the account accepts, no?
- In fact, is the account consistent? E.g., premise x may perhaps entail premise y. In that case you cannot consistently reject premise y while accepting premise x.
- Even if the account is consistent in the sense above, the premise which the account rejects seems plausible. Can the account show in what way the premise can fail?

To answer these questions, the truth-condition method is a great help!

By the way, the truth-condition method is often called **semantics**.

Standard Semantics

As a warm-up, let's consider a special case of what we need,

Standard Semantics

As a warm-up, let's consider a special case of what we need, viz., the (instant-based) **linear time model**:

- There are states, called "moments / instants / points of time". At each moment, atomic sentences may be true or false.

Standard Semantics

As a warm-up, let's consider a special case of what we need, viz., the (instant-based) **linear time model**:

- There are states, called "moments / instants / points of time". At each moment, atomic sentences may be true or false.
- Given two moments $m_1 \neq m_2$, there are two possibilities:
  - (a) m_1 is in the future of m_2. In symbol, $m_2 \rightarrow m_1$.
  - (b) m_2 is in the future of m_1. In symbol, $m_1 \rightarrow m_2$.

Standard Semantics

As a warm-up, let's consider a special case of what we need, viz., the (instant-based) **linear time model**:

- There are states, called "moments / instants / points of time". At each moment, atomic sentences may be true or false.
- Given two moments $m_1 \neq m_2$, there are two possibilities:
  - (a) m_1 is in the future of m_2. In symbol, $m_2 \to m_1$.
  - (b) m_2 is in the future of m_1. In symbol, $m_1 \to m_2$.

  We may even specify durations, as in

$$m_1 \xrightarrow{\text{1 day}} m_2 \xrightarrow{\text{1 day}} m_3$$

Standard Semantics

As a warm-up, let's consider a special case of what we need, viz., the (instant-based) **linear time model**:

- There are states, called "moments / instants / points of time". At each moment, atomic sentences may be true or false.
- Given two moments $m_1 \neq m_2$, there are two possibilities:
  - (a) m_1 is in the future of m_2. In symbol, $m_2 \to m_1$.
  - (b) m_2 is in the future of m_1. In symbol, $m_1 \to m_2$.

  We may even specify durations, as in

$$m_1 \xrightarrow{\text{1 day}} m_2 \xrightarrow{\text{1 day}} m_3$$

(i) "$\text{Will}_{\text{always}} : \varphi$" is true at a moment m

$\iff$ "φ" is true at every moment m' s.th. $m \to m'$.

Standard Semantics

As a warm-up, let's consider a special case of what we need, viz., the (instant-based) **linear time model**:

- There are states, called "moments / instants / points of time". At each moment, atomic sentences may be true or false.
- Given two moments $m_1 \neq m_2$, there are two possibilities:
  - (a) m_1 is in the future of m_2. In symbol, $m_2 \to m_1$.
  - (b) m_2 is in the future of m_1. In symbol, $m_1 \to m_2$.

  We may even specify durations, as in

$$m_1 \xrightarrow{\text{1 day}} m_2 \xrightarrow{\text{1 day}} m_3$$

(i) "$\text{Will}_{\text{always}} : \varphi$" is true at a moment m

$\iff$ "φ" is true at every moment m' s.th. $m \to m'$.

(ii) "$\text{Will}_{\text{sometimes}} : \varphi$" is true at a moment m

$\iff$ "φ" is true at some moment m' s.th. $m \to m'$.

1. "Will$_{\text{always}}$: φ" is true at m

   $\iff$ "φ" is true at every m' s.th. $m \to m'$.

2. "Will$_{\text{sometimes}}$: φ" is true at m

   $\iff$ "φ" is true at some m' s.th. $m \to m'$.

i "$\text{Will}_{\text{always}} : \varphi$" is true at m

$\iff$ "φ" is true at every m' s.th. $m \to m'$.

ii "$\text{Will}_{\text{sometimes}} : \varphi$" is true at m

$\iff$ "φ" is true at some m' s.th. $m \to m'$.

iii "$\text{Was}_{\text{always}} : \varphi$" is true at m

$\iff$ "φ" is true at every m' s.th. $m' \to m$.

iv "$\text{Was}_{\text{sometimes}} : \varphi$" is true at m

$\iff$ "φ" is true at some m' s.th. $m' \to m$.

1. "$\text{Will}_{\text{always}} : \varphi$" is true at m

   $\iff$ "φ" is true at every m' s.th. $m \to m'$.

2. "$\text{Will}_{\text{sometimes}} : \varphi$" is true at m

   $\iff$ "φ" is true at some m' s.th. $m \to m'$.

3. "$\text{Was}_{\text{always}} : \varphi$" is true at m

   $\iff$ "φ" is true at every m' s.th. $m' \to m$.

4. "$\text{Was}_{\text{sometimes}} : \varphi$" is true at m

   $\iff$ "φ" is true at some m' s.th. $m' \to m$.

5. "$\text{Will}_{d} : \varphi$" is true at m

   $\iff$ "φ" is true at the m' s.th. $m \xrightarrow{d} m'$.

6. "$\text{Was}_{d} : \varphi$" is true at m

   $\iff$ "φ" is true at the m' s.th. $m' \xrightarrow{d} m$.

Pnueli also defined:

- vii "Will-φ-until-ψ" is true at m

  $\iff$ there is an m' s.th. $m \to m'$, ψ is true at m', and φ is true at every m'' between m and m'.

Pnueli also defined:

vii "Will-φ-until-ψ" is true at m

$\iff$ there is an m' s.th. $m \to m'$, ψ is true at m', and

φ is true at every m'' between m and m'.

E.g., "the screen will be off until the button is pushed" true at m.

$$\cdots \longrightarrow m \longrightarrow \underset{\text{off}}{m''_1} \longrightarrow \underset{\text{off}}{m''_2} \longrightarrow \underset{\text{off}}{m''_3} \longrightarrow \overset{\text{push}}{m'} \longrightarrow \cdots$$

Pnueli also defined:

vii "Will-φ-until-ψ" is true at m

$$\iff \text{there is an } m' \text{ s.th. } m \to m', \psi \text{ is true at } m', \text{ and } \varphi \text{ is true at every } m'' \text{ between } m \text{ and } m'.$$

E.g., "the screen will be off until the button is pushed" true at m.

$$\begin{array}{ccccccccccccc} \cdots & \longrightarrow & m & \longrightarrow & m''_1 & \longrightarrow & m''_2 & \longrightarrow & m''_3 & \longrightarrow & \overset{\text{push}}{m'} & \longrightarrow & \cdots \\ & & & & \text{off} & & \text{off} & & \text{off} & & & & \end{array}$$

"Will$_{\text{always}}$: (if-push-then-(Will$_1$: on))" true at m.

Pnueli also defined:

vii "Will-φ-until-ψ" is true at m

$\iff$ there is an m' s.th. $m \to m'$, ψ is true at m', and
φ is true at every m'' between m and m'.

E.g., "the screen will be off until the button is pushed" true at m.

$$\cdots \longrightarrow m \longrightarrow \underset{\text{off}}{m''_1} \longrightarrow \underset{\text{off}}{m''_2} \longrightarrow \underset{\text{off}}{m''_3} \longrightarrow \overset{\text{push}}{m'} \longrightarrow \cdots$$

"Will$_{\text{always}}$: (if-push-then-(Will$_1$: on))" true at m.

$$\cdots \longrightarrow m \xrightarrow{1} \overset{\text{not-push}}{m''_1} \xrightarrow{1} \overset{\text{not-push}}{m''_2} \xrightarrow{1} \overset{\text{push}}{m'} \xrightarrow{1} \underset{\text{on}}{m'''} \longrightarrow \cdots$$

PHIL 222
Philosophical Foundations of Computer Science
Week 13, Thursday

Nov. 21, 2024

Basic Probability Theory
(cont'd)

Bayes' Rule

Given two events A and B (with non-zero probabilities), two conditional probabilities $P(A|B)$ and $P(B|A)$ are defined — so we have

$$P(B)P(A|B) = P(A \cap B) = P(A)P(B|A).$$

This implies a theorem called **Bayes' theorem** or **Bayes' rule**:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

Bayes' Rule

Given two events A and B (with non-zero probabilities), two conditional probabilities $P(A|B)$ and $P(B|A)$ are defined — so we have

$$P(B)P(A|B) = P(A \cap B) = P(A)P(B|A).$$

This implies a theorem called **Bayes' theorem** or **Bayes' rule**:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

An important interpretation of this rule is

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}.$$

E.g., classification in machine learning ("naive Bayes classifier"):

- A is "The email is a spam",
- B is "The subject line contains the '\$' sign".

Bayes' Rule

Given two events A and B (with non-zero probabilities), two conditional probabilities $P(A|B)$ and $P(B|A)$ are defined — so we have

$$P(B)P(A|B) = P(A \cap B) = P(A)P(B|A).$$

This implies a theorem called **Bayes' theorem** or **Bayes' rule**:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

An important interpretation of this rule is

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}.$$

E.g., classification in machine learning ("naive Bayes classifier"):

- A is "The email is a spam",
- B is "The subject line contains the '\$' sign".
- $P(A) = 10\%$;

Bayes' Rule

Given two events A and B (with non-zero probabilities), two conditional probabilities $P(A|B)$ and $P(B|A)$ are defined — so we have

$$P(B)P(A|B) = P(A \cap B) = P(A)P(B|A).$$

This implies a theorem called **Bayes' theorem** or **Bayes' rule**:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

An important interpretation of this rule is

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}.$$

E.g., classification in machine learning ("naive Bayes classifier"):

- A is "The email is a spam",
- B is "The subject line contains the '\$' sign".
- $P(A) = 10\%$; $P(B) = 1.25\%$; $P(B|A) = 10\%$;

Bayes' Rule

Given two events A and B (with non-zero probabilities), two conditional probabilities $P(A|B)$ and $P(B|A)$ are defined — so we have

$$P(B)P(A|B) = P(A \cap B) = P(A)P(B|A).$$

This implies a theorem called **Bayes' theorem** or **Bayes' rule**:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

An important interpretation of this rule is

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}.$$

E.g., classification in machine learning ("naive Bayes classifier"):

- A is "The email is a spam",
- B is "The subject line contains the '\$' sign".
- $P(A) = 10\%$; $P(B) = 1.25\%$; $P(B|A) = 10\%$; $P(A|B) = 80\%$.

Epistemology (3)
Learning

Here is Aaronson again:

> Centuries ago, David Hume [77] famously pointed out that learning from the past (and, by extension, science) seems logically impossible. For example, if we sample 500 ravens and every one of them is black, why does that give us any grounds — even probabilistic grounds — for expecting the 501^{st} raven to be black also? Any modern answer to this question would probably refer to *Occam's razor*, the principle that simpler hypotheses consistent with the data are more likely to be correct. [p. 286]

Here is Aaronson again:

> Centuries ago, David Hume [77] famously pointed out that learning from the past (and, by extension, science) seems logically impossible. For example, if we sample 500 ravens and every one of them is black, why does that give us any grounds — even probabilistic grounds — for expecting the 501^{st} raven to be black also? Any modern answer to this question would probably refer to *Occam's razor*, the principle that simpler hypotheses consistent with the data are more likely to be correct. [p. 286]

Philosophers have a *lot* to say about this.

Let's keep reading . . .

> So for example, the hypothesis that all ravens are black is "simpler" than the hypothesis that most ravens are green or purple, and that only the 500 we happened to see were black. Intuitively, it seems Occam's razor *must* be part of the solution to Hume's problem; the difficulty is that such a response leads to questions of its own:
>
> 1. What do we mean by "simpler"?
> 2. *Why* are simple explanations likely to be correct? Or, less ambitiously: what properties must reality have for Occam's Razor to "work"?
> 3. How much data must we collect before we can find a "simple hypothesis" that will probably predict future data? How do we go about finding such a hypothesis?
>
> [p. 286]

https://amturing.acm.org/award_winners/valiant_2612174.cfm

MORE ACM AWARDS

A.M. TURING AWARD

A.M. TURING AWARD LAUREATES BY...

ALPHABETICAL LISTING | **YEAR OF THE AWARD** | **RESEARCH SUBJECT**

LESLIE GABRIEL VALIANT

United States – 2010

CITATION

For transformative contributions to the theory of computation, including the theory of probably approximately correct (PAC) learning, the complexity of enumeration and of algebraic computation, and the theory of parallel and distributed computing.

SHORT ANNOTATED BIBLIOGRAPHY | ACM TURING AWARD LECTURE VIDEO | RESEARCH SUBJECTS | ADDITIONAL MATERIALS | VIDEO INTERVIEW

BIRTH:

28 March 1949, Budapest, Hungary

EDUCATION:

Latymer Upper School, London England; King's College, Cambridge, England (BA, Mathematics, 1970); Imperial College, London, England (DIC in Computing Science); University of Warwick, England (PhD, Computer Science, 1974)

Les Valiant has had an extraordinarily productive career in theoretical computer science producing results of great beauty and originality. His research has opened new frontiers and has resulted in a transformation of many areas. His work includes the study of both natural and artificial phenomena. The natural studies encompass the algorithms used by computing objects such as the human brain while the artificial include computers and their capabilities. In the case of computers the limitations of these devices are only beginning to be understood while

We are interested in learning, or making an AI learn, things like these:

We are interested in learning, or making an AI learn, things like these:

a. Given an email, whether it is a spam or not.

We are interested in learning, or making an AI learn, things like these:

a. Given an email, whether it is a spam or not.
b. Given a color (or its RGB code), whether it is "warm".
c. Given a person (or their height and weight), whether they are "medium-built".

We are interested in learning, or making an AI learn, things like these:

a. Given an email, whether it is a spam or not.
b. Given a color (or its RGB code), whether it is “warm”.
c. Given a person (or their height and weight), whether they are “medium-built”.
d. Given a raven, whether it is black or white.

We are interested in learning, or making an AI learn, things like these:

(a) Given an email, whether it is a spam or not.

(b) Given a color (or its RGB code), whether it is "warm".

(c) Given a person (or their height and weight), whether they are "medium-built".

(d) Given a raven, whether it is black or white.

All these can be abstractly expressed by functions $f : X \to \{0, 1\}$. E.g.,

(b) $X = 256^3$ and

$$f(c) = \begin{cases} 1 & \text{if the color } c \text{ is warm,} \\ 0 & \text{otherwise.} \end{cases}$$

We are interested in learning, or making an AI learn, things like these:

(a) Given an email, whether it is a spam or not.

(b) Given a color (or its RGB code), whether it is “warm”.

(c) Given a person (or their height and weight), whether they are “medium-built”.

(d) Given a raven, whether it is black or white.

All these can be abstractly expressed by functions $f : X \to \{0, 1\}$. E.g.,

(b) $X = 256^3$ and

$$f(c) = \begin{cases} 1 & \text{if the color } c \text{ is warm,} \\ 0 & \text{otherwise.} \end{cases}$$

(d) “All ravens are black” corresponds to f s.th. $f(x) = 1$ (black) for all x.

We are interested in learning, or making an AI learn, things like these:

ⓐ Given an email, whether it is a spam or not.

ⓑ Given a color (or its RGB code), whether it is "warm".

ⓒ Given a person (or their height and weight), whether they are "medium-built".

ⓓ Given a raven, whether it is black or white.

All these can be abstractly expressed by functions $f : X \to \{0, 1\}$. E.g.,

ⓑ $X = 256^3$ and

$$f(c) = \begin{cases} 1 & \text{if the color } c \text{ is warm,} \\ 0 & \text{otherwise.} \end{cases}$$

ⓓ "All ravens are black" corresponds to f s.th. $f(x) = 1$ (black) for all x.

There are many functions $f : X \to \{0, 1\}$, each expressing a hypothesis — but one of them is the correct one, and we want (the AI) to learn it.

We write k for the correct hypothesis.

We write k for the correct hypothesis.

The simple inductive reasoning Aaronson mentioned is

- d We observe 500 ravens and they are all black, i.e., $k(x) = 1$ for all of them. After this we reach the hypothesis h that all ravens are black, i.e., $h(x) = 1$ for all x in the set X of all ravens.

We write k for the correct hypothesis.

The simple inductive reasoning Aaronson mentioned is

- ⓓ We observe 500 ravens and they are all black, i.e., $k(x) = 1$ for all of them. After this we reach the hypothesis h that all ravens are black, i.e., $h(x) = 1$ for all x in the set X of all ravens.

In **PAC (“probably approximately correct”) learning**,

- ⓑ We pick a certain number of colors and for each c of them we teach the AI whether it is warm ($k(c) = 1$) or not. After this training, the AI reaches (by using some algorithm or other) a hypothesis $h : X \to \{0, 1\}$ that may be different from k but that is still “**probably approximately correct**”.

- ⓑ We pick a certain number of colors and for each c of them we teach the AI whether it is warm ($k(c) = 1$) or not. After this training, the AI reaches a hypothesis $h : X \to \{0, 1\}$ that may be different from k but that is still "probably approximately correct".

What does "probably approximately correct" mean?

- (b) We pick a certain number of colors and for each c of them we teach the AI whether it is warm ($k(c) = 1$) or not. After this training, the AI reaches a hypothesis $h : X \to \{0, 1\}$ that may be different from k but that is still "probably approximately correct".

What does "probably approximately correct" mean?

— It acknowledges two types of errors:

1. "Rarity errors": h may be incorret in the sense that $h \neq k$. But it may still be approximately correct, in the sense that $h(x) = k(x)$ with high probability.

ⓑ We pick a certain number of colors and for each c of them we teach the AI whether it is warm ($k(c) = 1$) or not. After this training, the AI reaches a hypothesis $h : X \to \{0, 1\}$ that may be different from k but that is still "probably approximately correct".

What does "probably approximately correct" mean?

— It acknowledges two types of errors:

1. "Rarity errors": h may be incorret in the sense that $h \neq k$. But it may still be approximately correct, in the sense that $h(x) = k(x)$ with high probability.
2. "Misfortune errors": If the set of colors used in the training is unrepresentative (e.g. only colors from the red family are chosen), then the AI may end up with a hypothesis that is not approximately correct. But this happens with low probability — i.e., it is probable that an approximately correct hypothesis has been reached.

We can prove, as a mathematical theorem, that PAC learning works.
(For simplicity's sake, we assume X to be finite.)

We can prove, as a mathematical theorem, that PAC learning works. (For simplicity's sake, we assume X to be finite.)

One important assumption is that we entertain (or the AI entertains) a small subset H of hypotheses rather than all the possible hypotheses.

We can prove, as a mathematical theorem, that PAC learning works. (For simplicity's sake, we assume X to be finite.)

One important assumption is that we entertain (or the AI entertains) a small subset H of hypotheses rather than all the possible hypotheses.

Now let's say we want the rate of rarity errors to be at most ε, so that

- h is **approx. correct** $\iff h(x) = k(x)$ with probability $\geqslant 1 - \varepsilon$.

We can prove, as a mathematical theorem, that PAC learning works. (For simplicity's sake, we assume X to be finite.)

One important assumption is that we entertain (or the AI entertains) a small subset H of hypotheses rather than all the possible hypotheses.

Now let's say we want the rate of rarity errors to be at most ε, so that

- h is **approx. correct** $\iff$ $h(x) = k(x)$ with probability $\geqslant 1 - \varepsilon$.

Moreover, let's say

- training data points $x_1, \ldots, x_m$ are **representative** $\iff$
  every h in H s.th. $h(x_i) = k(x_i)$ for all $i = 1, \ldots, m$ is approx. correct.

We can prove, as a mathematical theorem, that PAC learning works. (For simplicity's sake, we assume X to be finite.)

One important assumption is that we entertain (or the AI entertains) a small subset H of hypotheses rather than all the possible hypotheses.

Now let's say we want the rate of rarity errors to be at most ε, so that

- h is **approx. correct** $\iff$ $h(x) = k(x)$ with probability $\geqslant 1 - \varepsilon$.

Moreover, let's say

- training data points $x_1, \ldots, x_m$ are **representative** $\iff$ every h in H s.th. $h(x_i) = k(x_i)$ for all $i = 1, \ldots, m$ is approx. correct.

Then

- The rate of misfortune errors is the probability that randomly chosen data points fail to be representative.

Let's say we want this rate to be at most δ, so that a randomly chosen set of data points is representative with probability $\geqslant 1 - \delta$.

- We entertain (or the AI does) a small subset H of hypotheses.
- h is **approx. correct** $\iff$ $h(x) = k(x)$ with probability $\geqslant 1 - \varepsilon$.
- Training data points $x_1, \ldots, x_m$ are **representative** $\iff$
  every h in H s.th. $h(x_i) = k(x_i)$ for all $i = 1, \ldots, m$ is approx. correct.
- We want a randomly chosen set of data points to be representative with probability $\geqslant 1 - \delta$.

- We entertain (or the AI does) a small subset H of hypotheses.
- h is **approx. correct** $\iff$ $h(x) = k(x)$ with probability $\geqslant 1 - \varepsilon$.
- Training data points $x_1, \ldots, x_m$ are **representative** $\iff$ every h in H s.th. $h(x_i) = k(x_i)$ for all $i = 1, \ldots, m$ is approx. correct.
- We want a randomly chosen set of data points to be representative with probability $\geqslant 1 - \delta$.
- Let's say H contains n hypotheses.

- We entertain (or the AI does) a small subset H of hypotheses.
- h is **approx. correct** $\iff$ $h(x) = k(x)$ with probability $\geqslant 1 - \varepsilon$.
- Training data points $x_1, \ldots, x_m$ are **representative** $\iff$ every h in H s.th. $h(x_i) = k(x_i)$ for all $i = 1, \ldots, m$ is approx. correct.
- We want a randomly chosen set of data points to be representative with probability $\geqslant 1 - \delta$.
- Let's say H contains n hypotheses.

Theorem (Valiant). There is a number $m(n, \varepsilon, \delta)$

- that is linear in $1/\varepsilon$ and logarithmic in n and $1/\delta$, and
- s.th. any training data points $x_1, \ldots, x_{m(n,\varepsilon,\delta)}$ are representative with probability $\geqslant 1 - \delta$, assuming that they are chosen from X randomly and independently.

Valiant, §5.2 (we read some of these passages before):

> The main paradox of induction is the apparent contradiction between the following two of its facets. On the one hand, if no assumptions are made about the world, then clearly induction cannot be justified, because the world could conceivably be adversarial enough to ensure that the future is exactly the opposite of whatever prediction has just been made. [...]
>
> On the other hand, and in apparent contradiction to this argument, successful induction abounds all around us. [...]
>
> There may exist some acceptable assumptions that hold for the reproducible, naturally occurring form of induction, and under which induction is rigorously justifiable. I argue that this is exactly the case, and that just two assumptions are sufficient to give a quantitatively compelling account of induction. Further, these two particular assumptions are also necessary and unavoidable.
>
> [pp. 59f.]

The first assumption is the Invariance Assumption: The context in which the generalization is to be applied cannot be fundamentally different from that in which it was made. [. . .]

The second assumption is the Learnable Regularity Assumption. [. . .] These criteria can be viewed as regularities in the world. Such regularities have been discussed as such by philosophers, notably by David Hume in the eighteenth century.

The first assumption is the Invariance Assumption: The context in which the generalization is to be applied cannot be fundamentally different from that in which it was made. [. . .]

The second assumption is the Learnable Regularity Assumption. [. . .] These criteria can be viewed as regularities in the world. Such regularities have been discussed as such by philosophers, notably by David Hume in the eighteenth century. Computer science adds at least two further levels to this discussion. First, it is essential to require that any useful criterion or regularity be detectable: Whether the criterion applies to an instance should be resolvable by a feasible computation. [. . .]

However, the induction phenomenon has a second, even more severe, further constraint on it. [. . .] To explain induction it is also necessary to explain how an individual can acquire the detection algorithm for the regularity in the first place. In particular, this acquisition must be feasible, requiring only realistic resources and only a modest number of interactions with the world. [pp. 61f.]

Let's take the example from §5.5 of Valiant:

- Want to distinguish animals from plants using 20 features (can move, is green, etc.) that animals and plants may or may not have.

Let's take the example from §5.5 of Valiant:

- Want to distinguish animals from plants using 20 features (can move, is green, etc.) that animals and plants may or may not have.
- X = the set of all the $2^{20} = 1{,}048{,}576$ combinations of having or not having the 20 features (e.g. "can move & is not green & . . . ").

Let's take the example from §5.5 of Valiant:

- Want to distinguish animals from plants using 20 features (can move, is green, etc.) that animals and plants may or may not have.
- X = the set of all the $2^{20} = 1{,}048{,}576$ combinations of having or not having the 20 features (e.g. "can move & is not green & . . . ").
- I.e., want to learn $k : 2^{20} \to \{\, 0 \text{ (plant)}, 1 \text{ (animal)} \,\}$.

Let's take the example from §5.5 of Valiant:

- Want to distinguish animals from plants using 20 features (can move, is green, etc.) that animals and plants may or may not have.
- X = the set of all the $2^{20} = 1{,}048{,}576$ combinations of having or not having the 20 features (e.g. "can move & is not green & . . . ").
- I.e., want to learn $k : 2^{20} \to \{\, 0 \text{ (plant)}, 1 \text{ (animal)} \,\}$.

E.g. an "elimination algorithm" as follows:

Let's take the example from §5.5 of Valiant:

- Want to distinguish animals from plants using 20 features (can move, is green, etc.) that animals and plants may or may not have.
- X = the set of all the $2^{20} = 1{,}048{,}576$ combinations of having or not having the 20 features (e.g. "can move & is not green & . . . ").
- I.e., want to learn $k : 2^{20} \to \{ 0 \text{ (plant)}, 1 \text{ (animal)} \}$.

E.g. an "elimination algorithm" as follows:

- Prepare a list "can move; cannot move; is green; is not green, . . . ".

Let's take the example from §5.5 of Valiant:

- Want to distinguish animals from plants using 20 features (can move, is green, etc.) that animals and plants may or may not have.
- X = the set of all the $2^{20} = 1{,}048{,}576$ combinations of having or not having the 20 features (e.g. "can move & is not green & . . . ").
- I.e., want to learn $k : 2^{20} \to \{\, 0 \text{ (plant)}, 1 \text{ (animal)} \,\}$.

E.g. an "elimination algorithm" as follows:

- Prepare a list "can move; cannot move; is green; is not green, . . . ".
- Prepare 100 (say) animals $x_1, \ldots, x_{100}$ as training data points.

Let's take the example from §5.5 of Valiant:

- Want to distinguish animals from plants using 20 features (can move, is green, etc.) that animals and plants may or may not have.
- X = the set of all the $2^{20} = 1{,}048{,}576$ combinations of having or not having the 20 features (e.g. "can move & is not green & …").
- I.e., want to learn $k : 2^{20} \to \{\, 0 \text{ (plant)}, 1 \text{ (animal)} \,\}$.

E.g. an "elimination algorithm" as follows:

- Prepare a list "can move; cannot move; is green; is not green, …".
- Prepare 100 (say) animals $x_1, \ldots, x_{100}$ as training data points.
- For each x_i, cross out from the list the properties that do not fit it. E.g., if x_i can move, then cross out "cannot move".

Let's take the example from §5.5 of Valiant:

- Want to distinguish animals from plants using 20 features (can move, is green, etc.) that animals and plants may or may not have.
- X = the set of all the $2^{20} = 1{,}048{,}576$ combinations of having or not having the 20 features (e.g. "can move & is not green & . . . ").
- I.e., want to learn $k : 2^{20} \to \{\, 0 \text{ (plant)}, 1 \text{ (animal)} \,\}$.

E.g. an "elimination algorithm" as follows:

- Prepare a list "can move; cannot move; is green; is not green, . . . ".
- Prepare 100 (say) animals $x_1, \ldots, x_{100}$ as training data points.
- For each x_i, cross out from the list the properties that do not fit it. E.g., if x_i can move, then cross out "cannot move".
- After the training, the conjunction of the remaining properties, e.g. "x can move **and** x is not green", is the hypothesis h that we reach:

$$h(x) = 1 \text{ (i.e. } x \text{ is an animal)} \iff x \text{ can move } \textbf{and } x \text{ is not green.}$$

This algorithm gives us a probably approx. correct h — *if* we use the invariance assumption and the learnable regularity assumption.

This algorithm gives us a probably approx. correct h — *if* we use the invariance assumption and the learnable regularity assumption.

1. The invariance assumption: "[T]he examples encountered in the testing phase come from the same source as in the learning phase. Examples rarely seen during learning will be equally rare during testing . . . ." [p. 70]

This algorithm gives us a probably approx. correct h — *if* we use the invariance assumption and the learnable regularity assumption.

1. The invariance assumption: "[T]he examples encountered in the testing phase come from the same source as in the learning phase. Examples rarely seen during learning will be equally rare during testing ...." [p. 70]

This assumption is used in the theorem as follows. The theorem has

a. h is approx. correct $\iff h(x) = k(x)$ with probability $\geqslant 1 - \varepsilon$.
b. ... any training data points $x_1, \ldots, x_m$ are representative with probability $\geqslant 1 - \delta$, assuming that they are chosen from X randomly and independently.

This algorithm gives us a probably approx. correct h — *if* we use the invariance assumption and the learnable regularity assumption.

1. The invariance assumption: “[T]he examples encountered in the testing phase come from the same source as in the learning phase. Examples rarely seen during learning will be equally rare during testing . . . .” [p. 70]

This assumption is used in the theorem as follows. The theorem has

a. h is approx. correct $\iff$ $h(x) = k(x)$ with probability $\geqslant 1 - \varepsilon$.

b. . . . any training data points $x_1, \ldots, x_m$ are representative with probability $\geqslant 1 - \delta$, assuming that they are chosen from X randomly and independently.

Each of a and b involves a probability distribution (b involves random choosing of data points), but they must be using the same distribution and the choice of x in a and x_i in b must be independent — or else the theorem fails.

This assumption often fails, or the hypothesis comes with a restricted range of application for the assumption to hold, both in AIs' learning and in our learning (or science).

This assumption often fails, or the hypothesis comes with a restricted range of application for the assumption to hold, both in AIs' learning and in our learning (or science).

- Data sets used in the training / learning phase may be biased.

  E.g., incidents in the facial recognition industry: white people being the majority in the learning phase vs. a lot of people of color in the market use.

This assumption often fails, or the hypothesis comes with a restricted range of application for the assumption to hold, both in AIs' learning and in our learning (or science).

- Data sets used in the training / learning phase may be biased.

  E.g., incidents in the facial recognition industry: white people being the majority in the learning phase vs. a lot of people of color in the market use.

- Technological development may make previously unavailable experimental conditions / environments available.

  E.g., particle / high-energy physics: hypotheses learned in low-energy conditions may (of course) not be applicable in high-energy conditions.

2. The learnable regularity assumption: "[F]or the given features a criterion for distinguishing animals from plants could be expressed as a conjunction. This was sufficient because conjunctions can be shown to be learnable, as we have just seen." [p. 70]

2. The learnable regularity assumption: "[F]or the given features a criterion for distinguishing animals from plants could be expressed as a conjunction. This was sufficient because conjunctions can be shown to be learnable, as we have just seen." [p. 70]

For the sake of the theorem, the algorithm sets

- H = the set of hypotheses based on conjunctions, stating, e.g.,

$$x \text{ is an animal} \iff x \text{ can move } \textbf{and } x \text{ is not green.}$$

2. The learnable regularity assumption: "[F]or the given features a criterion for distinguishing animals from plants could be expressed as a conjunction. This was sufficient because conjunctions can be shown to be learnable, as we have just seen." [p. 70]

For the sake of the theorem, the algorithm sets

- H = the set of hypotheses based on conjunctions, stating, e.g.,

$$x \text{ is an animal} \iff x \text{ can move } \textbf{and } x \text{ is not green.}$$

- There are $2^{2^{20}} = 2^{1{,}048{,}576} \approx 6.7 \times 10^{315{,}652}$ possible hypotheses, but H contains *only* $n = 3^{20} - 1 = 3{,}486{,}784{,}400$.

Hence we have a small enough $m(n, \varepsilon, \delta)$,

2. The learnable regularity assumption: "[F]or the given features a criterion for distinguishing animals from plants could be expressed as a conjunction. This was sufficient because conjunctions can be shown to be learnable, as we have just seen." [p. 70]

For the sake of the theorem, the algorithm sets

- H = the set of hypotheses based on conjunctions, stating, e.g.,

$$x \text{ is an animal} \iff x \text{ can move } \textbf{and } x \text{ is not green.}$$

- There are $2^{2^{20}} = 2^{1{,}048{,}576} \approx 6.7 \times 10^{315{,}652}$ possible hypotheses, but H contains *only* $n = 3^{20} - 1 = 3{,}486{,}784{,}400$.

Hence we have a small enough $m(n, \varepsilon, \delta)$, but remember

- Training data points $x_1, \ldots, x_m$ are representative $\iff$ every h in H that has survived the training is approx. correct.

2. The learnable regularity assumption: "[F]or the given features a criterion for distinguishing animals from plants could be expressed as a conjunction. This was sufficient because conjunctions can be shown to be learnable, as we have just seen." [p. 70]

For the sake of the theorem, the algorithm sets

- H = the set of hypotheses based on conjunctions, stating, e.g.,

$$x \text{ is an animal} \iff x \text{ can move } \textbf{and } x \text{ is not green.}$$

- There are $2^{2^{20}} = 2^{1{,}048{,}576} \approx 6.7 \times 10^{315{,}652}$ possible hypotheses, but H contains *only* $n = 3^{20} - 1 = 3{,}486{,}784{,}400$.

Hence we have a small enough $m(n, \varepsilon, \delta)$, but remember

- Training data points $x_1, \ldots, x_m$ are representative $\iff$ every h in H that has survived the training is approx. correct.

This give no guarantee that at least one h survives the training (in the running example, at least one property in the list survives)

❷ The learnable regularity assumption: "[F]or the given features a criterion for distinguishing animals from plants could be expressed as a conjunction. This was sufficient because conjunctions can be shown to be learnable, as we have just seen." [p. 70]

For the sake of the theorem, the algorithm sets

- H = the set of hypotheses based on conjunctions, stating, e.g.,

$$x \text{ is an animal} \iff x \text{ can move } \textbf{and } x \text{ is not green.}$$

- There are $2^{2^{20}} = 2^{1{,}048{,}576} \approx 6.7 \times 10^{315{,}652}$ possible hypotheses, but H contains *only* $n = 3^{20} - 1 = 3{,}486{,}784{,}400$.

Hence we have a small enough $m(n, \varepsilon, \delta)$, but remember

- Training data points $x_1, \ldots, x_m$ are representative $\iff$
  every h in H that has survived the training is approx. correct.

This give no guarantee that at least one h survives the training (in the running example, at least one property in the list survives) — unless we assume ❷ (or maybe a slightly weaker assumption).

But this assumption can easily fail. E.g., the correct criterion may be

$$x \text{ is an animal} \iff$$
$$x \text{ can move } \textbf{and } (x \text{ is not green } \textbf{or } x \text{ has feathers}).$$

But this assumption can easily fail. E.g., the correct criterion may be

$$x \text{ is an animal } \iff$$
$$x \text{ can move } \textbf{and } (x \text{ is not green } \textbf{or } x \text{ has feathers}).$$

But once we let such hypotheses in H, the set H may no longer be PAC learnable (i.e. a PAC hypothesis cannot be reached efficiently).

E.g., if we let H contain every conjunction of disjunctions (or every disjunction of conjunctions), it ends up containing all $2^{2^{20}}$ of the possible hypotheses $h : 2^{20} \to \{0, 1\}$.

Valiant regards PAC learning as a response to the problem of induction.

Valiant regards PAC learning as a response to the problem of induction.

He needs the invariance and learnable regularity assumptions:

1. The context in which the generalization is to be applied cannot be fundamentally different from that in which it was made.
2. The correct regularity is efficiently learnable. (E.g. conjucntions vs. conjunctions + disjunctions.)

Valiant regards PAC learning as a response to the problem of induction.

He needs the invariance and learnable regularity assumptions:

1. The context in which the generalization is to be applied cannot be fundamentally different from that in which it was made.
2. The correct regularity is efficiently learnable. (E.g. conjucntions vs. conjunctions + disjunctions.)

1 is a variant of Hume's uniformity principle:

i. [I]nstances, of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same.

Valiant regards PAC learning as a response to the problem of induction.

He needs the invariance and learnable regularity assumptions:

1. The context in which the generalization is to be applied cannot be fundamentally different from that in which it was made.
2. The correct regularity is efficiently learnable. (E.g. conjucntions vs. conjunctions + disjunctions.)

1 is a variant of Hume's uniformity principle:

- i [I]nstances, of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same.

But is "The induction is possible, once we assume 1 (or i) and 2" or "We need 1 (or i) and 2 to make induction possible" a good response to Hume?

Valiant regards PAC learning as a response to the problem of induction.

He needs the invariance and learnable regularity assumptions:

1. The context in which the generalization is to be applied cannot be fundamentally different from that in which it was made.
2. The correct regularity is efficiently learnable. (E.g. conjucntions vs. conjunctions + disjunctions.)

1 is a variant of Hume's uniformity principle:

i. [I]nstances, of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same.

But is "The induction is possible, once we assume 1 (or i) and 2" or "We need 1 (or i) and 2 to make induction possible" a good response to Hume?

Can there be any separate justifications for 1 (or i) and 2?

Epistemology (3)
Learning and Simplicity

Philosophers have various accounts of how to justify induction / learning.

Philosophers have various accounts of how to justify induction / learning.

- One account highlights the role of a background system of beliefs and rules.

Philosophers have various accounts of how to justify induction / learning.

- One account highlights the role of a background system of beliefs and rules. It points out that there are two different questions.
    a. Given this evidence, whether that conclusion follows or not.

Philosophers have various accounts of how to justify induction / learning.

- One account highlights the role of a background system of beliefs and rules. It points out that there are two different questions.
  a. Given this evidence, whether that conclusion follows or not.
  b. Given the background system and new evidence, whether it is rational to infer such and such.

Philosophers have various accounts of how to justify induction / learning.

- One account highlights the role of a background system of beliefs and rules. It points out that there are two different questions.
  - (a) Given this evidence, whether that conclusion follows or not.
  - (b) Given the background system and new evidence, whether it is rational to infer such and such.

  E.g., in terms of Bayes' rule,

$$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}.$$

  The same evidence revises different priors to different posteriors, but both revisions can be rational.

Philosophers have various accounts of how to justify induction / learning.

- One account highlights the role of a background system of beliefs and rules. It points out that there are two different questions.
  - a Given this evidence, whether that conclusion follows or not.
  - b Given the background system and new evidence, whether it is rational to infer such and such.

  E.g., in terms of Bayes' rule,

  $$\text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Evidence}}.$$

  The same evidence revises different priors to different posteriors, but both revisions can be rational.

  Induction does not concern a (which deduction is about) but b.

Recall this passage from Aaronson:

> Intuitively, it seems Occam's razor *must* be part of the solution to Hume's problem; the difficulty is that such a response leads to questions of its own:
>
> 1. What do we mean by "simpler"?
> 2. *Why* are simple explanations likely to be correct? Or, less ambitiously: what properties must reality have for Occam's Razor to "work"?
> 3. How much data must we collect before we can find a "simple hypothesis" that will probably predict future data? How do we go about finding such a hypothesis?
>
> [p. 286]

Recall this passage from Aaronson:

> Intuitively, it seems Occam's razor *must* be part of the solution to Hume's problem; the difficulty is that such a response leads to questions of its own:
>
> 1. What do we mean by "simpler"?
> 2. *Why* are simple explanations likely to be correct? Or, less ambitiously: what properties must reality have for Occam's Razor to "work"?
> 3. How much data must we collect before we can find a "simple hypothesis" that will probably predict future data? How do we go about finding such a hypothesis?
>
> [p. 286]

Some philosophers discuss how to justify the "Ockham's razor" principle, and what it means for a hypothesis to be simple (Question 1) — although they may not accept the assumption behind Question 2.

Valiant regards PAC learning as a response to the problem of induction.

He needs the invariance and learnable regularity assumptions:

1. The context in which the generalization is to be applied cannot be fundamentally different from that in which it was made.
2. The correct regularity is efficiently learnable. (E.g. conjucntions vs. conjunctions + disjunctions.)

1 is a variant of Hume's uniformity principle:

i. [I]nstances, of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same.

But is "The induction is possible, once we assume 1 (or i) and 2" or "We need 1 (or i) and 2 to make induction possible" a good response to Hume?

Can there be any separate justifications for 1 (or i) and 2?

Valiant discusses a related aspect of this question in §5.7.

Valiant discusses a related aspect of this question in §5.7.

> However, it is often the case that we are confronted with a hypothesis about the provenance of which we know nothing. [p. 72]

The hypothesis may be generated by an algorithm that is not proven to be a PAC learning algorithm. Or the hypothesis may come from cherry-picked data; even an adversary may be trying to trick us.

Valiant discusses a related aspect of this question in §5.7.

> However, it is often the case that we are confronted with a hypothesis about the provenance of which we know nothing. [p. 72]

The hypothesis may be generated by an algorithm that is not proven to be a PAC learning algorithm. Or the hypothesis may come from cherry-picked data; even an adversary may be trying to trick us.

> The answer lies with Occam algorithms: They provide a rigorous approach, even in such cases of total ignorance about the origins of a hypothesis, and exemplify the role of purely statistical arguments in machine learning. What this approach provides are some conditions under which an unfamiliar hypothesis can be trusted. These conditions make concrete and rigorous the intuition sometimes attributed to the fourteenth-century logician William of Ockham that all things being equal, simpler hypotheses are more likely to be valid than complex ones. [p. 73]

> Suppose that you are trying to predict horse races, and that someone gives you data from a hundred past races in which every time the heaviest horse won. Discerning whether the heaviest horse is the sure thing it might appear to be requires several steps. [p. 73]

We still need the invariance assumption, but once we have it,

> Third, you would need to assess the complexity of the hypothesis. It is tempting to bet on the heaviest horse because of the simplicity of the rule "the heaviest will win." It seems unlikely that 100 races would all satisfy such a simple rule just by accident. If the rule were much more complex, for example that the horse's height, the owner's weight, and the trainer's age (all in appropriate units) added up to a prime number, then you would be a little more skeptical, and justifiably so. Even if the winners were totally unpredictable and arbitrary, some prediction rule could always be engineered to match them after the fact, if the rule is allowed to be complicated enough. [p. 73]

- **Formal learning theory** justifies Ockham's razor without assuming that a simpler hypothesis is more likely to be true. (Indeed the theory does not involve any probability!)

- **Formal learning theory** justifies Ockham's razor without assuming that a simpler hypothesis is more likely to be true. (Indeed the theory does not involve any probability!)

Consider "All ravens are black".

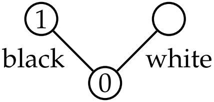
- **Formal learning theory** justifies Ockham's razor without assuming that a simpler hypothesis is more likely to be true. (Indeed the theory does not involve any probability!)

Consider "All ravens are black".

0

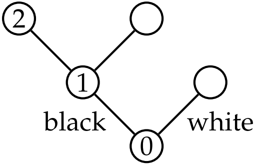
- **Formal learning theory** justifies Ockham's razor without assuming that a simpler hypothesis is more likely to be true. (Indeed the theory does not involve any probability!)

Consider "All ravens are black".



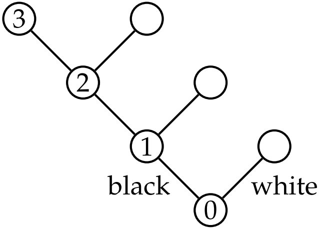
- **Formal learning theory** justifies Ockham's razor without assuming that a simpler hypothesis is more likely to be true. (Indeed the theory does not involve any probability!)

Consider "All ravens are black".



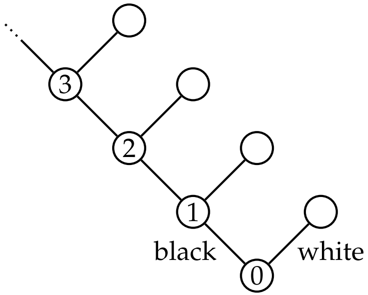
- **Formal learning theory** justifies Ockham's razor without assuming that a simpler hypothesis is more likely to be true. (Indeed the theory does not involve any probability!)

Consider "All ravens are black".



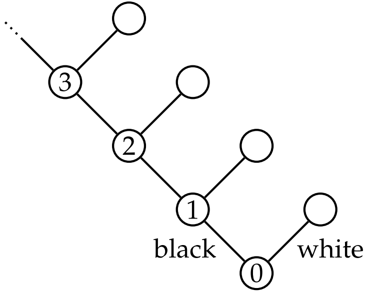
- **Formal learning theory** justifies Ockham's razor without assuming that a simpler hypothesis is more likely to be true. (Indeed the theory does not involve any probability!)

Consider "All ravens are black".



- **Formal learning theory** justifies Ockham's razor without assuming that a simpler hypothesis is more likely to be true. (Indeed the theory does not involve any probability!)

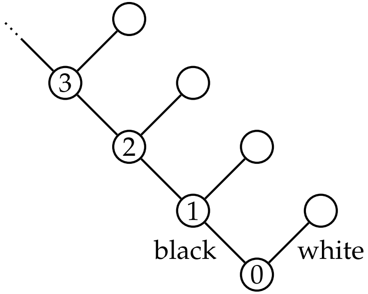
Consider "All ravens are black".



"All ravens are black" is simpler (and better) than "Some are black and others are white" in the following sense.

- **Formal learning theory** justifies Ockham's razor without assuming that a simpler hypothesis is more likely to be true. (Indeed the theory does not involve any probability!)

Consider "All ravens are black".



"All ravens are black" is simpler (and better) than "Some are black and others are white" in the following sense.

Think of the situation as a game between us and Nature, where we hypothesize about ravens and Nature may try to trick us.

The game goes in turns, by

- us deciding whether to stay with a current hypothesis or to adopt a new one,
- Nature, or the Adversary, showing us a new raven.

Our goal is to become correct eventually. Nature's goal is to trick us.

The game goes in turns, by

- us deciding whether to stay with a current hypothesis or to adopt a new one,
- Nature, or the Adversary, showing us a new raven.

Our goal is to become correct eventually. Nature's goal is to trick us. Nature can give us any new raven at any stage, as long as it does not violate the law of nature, which may be ❶ or may be ❷.

❶ All ravens are indeed black.

❷ Some ravens are white.

The game goes in turns, by

- us deciding whether to stay with a current hypothesis or to adopt a new one,
- Nature, or the Adversary, showing us a new raven.

Our goal is to become correct eventually. Nature's goal is to trick us.

Nature can give us any new raven at any stage, as long as it does not violate the law of nature, which may be 1 or may be 2.

1. All ravens are indeed black.
2. Some ravens are white.

We have two strategies:

i. We start with the simple "All ravens are black".
   Revise after seeing a lot of ravens suggesting the other hypothesis.
ii. We start with the less simple "Some ravens are white".
   Revise after seeing a lot of ravens suggesting the other hypothesis.

Two laws 1 & 2 × two strategies i & ii. What happens in each case?

1. All ravens are indeed black.
   - i. We start with the simple "All ravens are black".
   - ii. We start with the less simple "Some ravens are white".
2. Some ravens are white.
   - i. We start with the simple "All ravens are black".
   - ii. We start with the less simple "Some ravens are white".

Two laws 1 & 2 × two strategies i & ii. What happens in each case?

1. All ravens are indeed black.
   - i. We start with the simple "All ravens are black". Then we never need to change our hypothesis / belief.
   - ii. We start with the less simple "Some ravens are white".

2. Some ravens are white.
   - i. We start with the simple "All ravens are black".
   - ii. We start with the less simple "Some ravens are white".

Two laws ❶ & ❷ × two strategies ⓘ & ⓘⓘ. What happens in each case?

1. All ravens are indeed black.
   - i. We start with the simple "All ravens are black". Then we never need to change our hypothesis / belief.
   - ii. We start with the less simple "Some ravens are white". Then, to attain true belief, we need to change our belief eventually.
2. Some ravens are white.
   - i. We start with the simple "All ravens are black".
   - ii. We start with the less simple "Some ravens are white".

Two laws ❶ & ❷ × two strategies ⓘ & ⓘⓘ. What happens in each case?

❶ All ravens are indeed black.

- ⓘ We start with the simple "All ravens are black". Then we never need to change our hypothesis / belief.
- ⓘⓘ We start with the less simple "Some ravens are white". Then, to attain true belief, we need to change our belief eventually.

❷ Some ravens are white.

- ⓘ We start with the simple "All ravens are black". This is not true, so Nature must eventually show us a white raven. Then (and only then) we change our belief.
- ⓘⓘ We start with the less simple "Some ravens are white".

Two laws ❶ & ❷ × two strategies ⓘ & ⓘⓘ. What happens in each case?

1. All ravens are indeed black.
   - i. We start with the simple "All ravens are black". Then we never need to change our hypothesis / belief.
   - ii. We start with the less simple "Some ravens are white". Then, to attain true belief, we need to change our belief eventually.
2. Some ravens are white.
   - i. We start with the simple "All ravens are black". This is not true, so Nature must eventually show us a white raven. Then (and only then) we change our belief.
   - ii. We start with the less simple "Some ravens are white". Then Nature can withhold a white raven and keep showing black ravens, until we decide to change our belief to "All ravens are black" as in ii of 1. Then Nature reveals a white raven, and we are forced to change our belief the second time.

Two laws 1 & 2 × two strategies i & ii. What happens in each case?

1 All ravens are indeed black.

  - i We start with the simple "All ravens are black". Then we never need to change our hypothesis / belief.
  - ii We start with the less simple "Some ravens are white". Then, to attain true belief, we need to change our belief eventually.

2 Some ravens are white.

  - i We start with the simple "All ravens are black". This is not true, so Nature must eventually show us a white raven. Then (and only then) we change our belief.
  - ii We start with the less simple "Some ravens are white". Then Nature can withhold a white raven and keep showing black ravens, until we decide to change our belief to "All ravens are black" as in ii of 1. Then Nature reveals a white raven, and we are forced to change our belief the second time.

In each case, strategy i involves fewer belief-changes than ii.

More generally, we have the following results:

More generally, we have the following results:

- There is a hierarchy of hypotheses s.th. the strategy of always choosing a not-yet-refuted hypothesis higher in this hierarchy guarantees that we converge to a true hypothesis with the fewest belief-changes (in the worst case).

More generally, we have the following results:

- There is a hierarchy of hypotheses s.th. the strategy of always choosing a not-yet-refuted hypothesis higher in this hierarchy guarantees that we converge to a true hypothesis with the fewest belief-changes (in the worst case).
- A certain theorem justifies calling hypotheses higher in this hierarchy simpler.

More generally, we have the following results:

- There is a hierarchy of hypotheses s.th. the strategy of always choosing a not-yet-refuted hypothesis higher in this hierarchy guarantees that we converge to a true hypothesis with the fewest belief-changes (in the worst case).
- A certain theorem justifies calling hypotheses higher in this hierarchy simpler.

So the philosophical upshot is this:

More generally, we have the following results:

- There is a hierarchy of hypotheses s.th. the strategy of always choosing a not-yet-refuted hypothesis higher in this hierarchy guarantees that we converge to a true hypothesis with the fewest belief-changes (in the worst case).
- A certain theorem justifies calling hypotheses higher in this hierarchy simpler.

So the philosophical upshot is this:

- There is no guarantee that a simple hypothesis is more likely to be true — but that is not the guarantee we are after.

More generally, we have the following results:

- There is a hierarchy of hypotheses s.th. the strategy of always choosing a not-yet-refuted hypothesis higher in this hierarchy guarantees that we converge to a true hypothesis with the fewest belief-changes (in the worst case).
- A certain theorem justifies calling hypotheses higher in this hierarchy simpler.

So the philosophical upshot is this:

- There is no guarantee that a simple hypothesis is more likely to be true — but that is not the guarantee we are after.
- The Ockham strategy, of always choosing a simpler hypothesis, guarantees the shortest path to truth.