LING3804: Lecture Notes 9
Complex Reasoning

Humans seem to intend precise meanings for sentences they utter, with well-defined entailments.

Linguists usually represent these meanings using unambiguous formal logical expressions.

This lecture defines a very common logic and maps it to a neural mechanism for logical entailment.

Contents

9.1 Reasoning about generalizations

Recall our two sources of knowledge:

- from other people, via language (we'll discuss that in the next lecture),
- from existing knowledge (other logical propositions), via entailment.

Entailment $\pi_1, \pi_2, ... \Rightarrow \kappa$ is a process of inferring conclusion statements κ from premises π_1, π_2, ...

Here are the two most common entailments:

1. **conjunction elimination (CE)** — for all φ, ψ of type $\mathtt{t}$:

$$\varphi \wedge \psi \;\Rightarrow\; \varphi \qquad \varphi \wedge \psi \;\Rightarrow\; \psi$$

   (we saw this in the last lecture notes),

2. **universal modus ponens (UMP)** — for all ρ, σ of type $(\alpha){\rightarrow}\beta$ and χ of type α:

$$\mathsf{All}\ \rho\ \sigma,\ \ \rho\ \chi \;\Rightarrow\; \sigma\ \chi$$

   (this has two premises: the first is the **major premise** and the second is the **minor premise**)

As an example, if we have a theory $\mathsf{Theory1}$:

$$\mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x)\ (\lambda_x\ \mathsf{Animal}\ x) \wedge$$
$$\mathsf{Dog}\ \mathsf{Woofy}$$

we can conclude the following, using (conjuncts of) our theory as premises:

$$\begin{aligned}
\overbrace{\mathsf{Theory1}}^{\text{major premise}}, \quad \overbrace{\mathsf{Theory1}}^{\text{minor premise}} \quad &\Rightarrow \overbrace{\mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x)\ (\lambda_x\ \mathsf{Animal}\ x)}^{\text{major premise}},\ \overbrace{\mathsf{Dog\ Woofy}}^{\text{minor premise}} && \text{(CE)}\\
&\Rightarrow \mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x)\ (\lambda_x\ \mathsf{Animal}\ x),\ (\lambda_x\ \mathsf{Dog}\ x)\ \mathsf{Woofy} && \text{(beta reduction)}\\
&\Rightarrow (\lambda_x\ \mathsf{Animal}\ x)\ \mathsf{Woofy} && \text{(UMP)}\\
&\Rightarrow \mathsf{Animal\ Woofy} && \text{(beta reduction)}
\end{aligned}$$

(Note that beta reduction is an equation, so we can reverse it in the second line.)

These entailments can be chained up for nested quantifiers, so if we have a theory $\mathsf{Theory2}$:

$$\begin{aligned}
&\mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x)\ (\lambda_x\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{Chase}\ y\ x)) \wedge\\
&\mathsf{Dog\ Woofy}\\
&\mathsf{Bird\ Chirpy}
\end{aligned}$$

we can conclude:

$$\begin{aligned}
&\overbrace{\mathsf{Theory2}}^{\text{major premise}}, \quad \overbrace{\mathsf{Theory2}}^{\text{minor premise}}\\
&\quad \Rightarrow \mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x)\ (\lambda_x\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{Chase}\ y\ x)),\ \mathsf{Dog\ Woofy} && \text{(CE)}\\
&\quad \Rightarrow \mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x)\ (\lambda_x\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{Chase}\ y\ x)),\ (\lambda_x\ \mathsf{Dog}\ x)\ \mathsf{Woofy} && \text{(beta reduction)}\\
&\quad \Rightarrow (\lambda_x\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{Chase}\ y\ x))\ \mathsf{Woofy} && \text{(UMP)}\\
&\quad \Rightarrow \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{Chase}\ y\ \mathsf{Woofy})) && \text{(beta reduction)}
\end{aligned}$$

then we can use that conclusion as a major premise in another universal modus ponens:

$$\begin{aligned}
&\overbrace{\mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{Chase}\ x\ \mathsf{Woofy})}^{\text{major premise}}, \quad \overbrace{\mathsf{Theory2}}^{\text{minor premise}}\\
&\quad \Rightarrow \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{Chase}\ x\ \mathsf{Woofy}),\ \mathsf{Bird\ Chirpy} && \text{(CE)}\\
&\quad \Rightarrow \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{Chase}\ y\ \mathsf{Woofy}),\ (\lambda_y\ \mathsf{Bird}\ y)\ \mathsf{Chirpy} && \text{(beta reduction)}\\
&\quad \Rightarrow (\lambda_y\ \mathsf{Chase}\ y\ \mathsf{Woofy})\ \mathsf{Chirpy} && \text{(UMP)}\\
&\quad \Rightarrow \mathsf{Chase\ Chirpy\ Woofy} && \text{(beta reduction)}
\end{aligned}$$

Practice 9.1:

If we have a theory:

$$\begin{aligned}
&\mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x)\ (\lambda_x\ \mathsf{Scratch}\ x\ x) \wedge\\
&\mathsf{Dog\ Woofy}
\end{aligned}$$

show a derivation for:

$$\mathsf{Scratch\ Woofy\ Woofy}.$$

9.2 Reasoning about intensions

Intensions (using the '↑' operator) allow us to treat expressions as proposition objects (theories).

This means we can quantify over them and run universal modus ponens on them as well.

For example, if we have this Theory3:

$$
\begin{array}{l}
\mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x) \\
\quad (\lambda_x\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y) \\
\qquad (\lambda_y\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase}\ y\ x)) \\
\qquad\quad (\lambda_i\ \mathsf{Want}\ i\ x)))\ \wedge \\
\mathsf{Dog\ Woofy}\ \wedge \\
\mathsf{Bird\ Chirpy}
\end{array}
$$

then we can conclude:

$$
\overbrace{\mathsf{Theory3}}^{\text{major premise}},\ \overbrace{\mathsf{Theory3}}^{\text{minor premise}}
$$

$$
\begin{array}{ll}
\Rightarrow\ \mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x)\ (\lambda_x\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase}\ y\ x)) & \text{(CE)} \\
\qquad (\lambda_i\ \mathsf{Want}\ i\ x))),\ \ \mathsf{Dog\ Woofy} & \\
\Rightarrow\ \mathsf{All}\ (\lambda_x\ \mathsf{Dog}\ x)\ (\lambda_x\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase}\ y\ x)) & \text{(beta reduction)} \\
\qquad (\lambda_i\ \mathsf{Want}\ i\ x))),\ \ (\lambda_x\ \mathsf{Dog}\ x)\ \mathsf{Woofy} & \\
\Rightarrow\ (\lambda_x\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase}\ y\ x)) & \text{(UMP)} \\
\qquad (\lambda_i\ \mathsf{Want}\ i\ x)))\ \mathsf{Woofy} & \\
\Rightarrow\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase}\ y\ \mathsf{Woofy})) & \text{(beta reduction)} \\
\qquad (\lambda_i\ \mathsf{Want}\ i\ \mathsf{Woofy})) &
\end{array}
$$

Then we can use that conclusion as a major premise in another universal modus ponens:

$$
\overbrace{\mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase}\ y\ \mathsf{Woofy}))\ (\lambda_i\ \mathsf{Want}\ i\ \mathsf{Woofy}))}^{\text{major premise}},\ \overbrace{\mathsf{Theory3}}^{\text{minor premise}}
$$

$$
\begin{array}{ll}
\Rightarrow\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase}\ y\ \mathsf{Woofy})) & \text{(CE)} \\
\qquad (\lambda_i\ \mathsf{Want}\ i\ \mathsf{Woofy})),\ \ \mathsf{Bird\ Chirpy} & \\
\Rightarrow\ \mathsf{All}\ (\lambda_y\ \mathsf{Bird}\ y)\ (\lambda_y\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase}\ y\ \mathsf{Woofy})) & \text{(beta reduction)} \\
\qquad (\lambda_i\ \mathsf{Want}\ i\ \mathsf{Woofy})),\ \ (\lambda_y\ \mathsf{Bird}\ y)\ \mathsf{Chirpy} & \\
\Rightarrow\ (\lambda_y\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase}\ y\ \mathsf{Woofy})) & \text{(UMP)} \\
\qquad (\lambda_i\ \mathsf{Want}\ i\ \mathsf{Woofy}))\ \mathsf{Chirpy} & \\
\Rightarrow\ \mathsf{All}\ (\lambda_i\ \mathsf{Equal}\ i\ (\uparrow\ \mathsf{Chase\ Chirpy\ Woofy})) & \text{(beta reduction)} \\
\qquad (\lambda_i\ \mathsf{Want}\ i\ \mathsf{Woofy}) &
\end{array}
$$

Beta reduction can substitute variables in the intension (the desired proposition).

We then create an object Theory3a: Chase Chirpy Woofy for use in subsequent entailments.

Practice 9.2:

If we have a theory:

All (λ_x Dog x)
    (λ_x All (λ_i Equal i (↑ Scratch x x))
        (λ_i Believe i x)) ∧
Dog Woofy

show a derivation for:

All (λ_i Equal i (↑ Scratch Woofy Woofy) .
    (λ_i Believe i Woofy)

9.3 Example: interpreting a question

Now we can reason about what kind of answer a question requires, e.g. using these predicates:

- CurrentTheory : (p)→t, true for some current theory proposition in associative memory
- CurrentTimeInTheory : (p)→(e)→t, true for the current time entity in the provided theory
- ConsecutiveTo : (e)→(e)→t, true if the second time entity immediately follows the first
- Admin : (e)→t, true for the administrator agent
- AuthorizedUser : (e)→t, true for any authorized user
- Speaker : (e)→(e)→t, true for the current speaker at the provided time entity
- Student : (e)→t, true for any student
- Equal : (p)→(p)→t, true if the two propositions are identical
- Want : (p)→(e)→(e)→t, true if the first entity wants the proposition at the time entity
- Believe : (p)→(e)→(e)→t, true if the first entity believes the proposition at the time entity
- Tell : (p)→(e)→(e)→(e)→t, true if the first entity tells the second the proposition at the time

and the following conjunction as a current theory, Theory1:

Admin Person1 ∧
AuthorizedUser Person7 ∧
Speaker Person7 10:00InTheory1 ∧
Student Person43 ∧
Student Person58

and the following conjunction as a top-level theory, Theory0, to be combined with other theories:

CurrentTimeInTheory p 10:00Inp ∧	(current time in every possible proposition p)
ConsecutiveTo tInp $(t{+}1)$Inp ∧	(for every time t in every possible proposition p)
Equal p p ∧	(every possible proposition p is equal to itself)
CurrentTheory (↑ Theory1)	

If we then add a translation of *Who is a student?*:

$$
\begin{array}{l}
\textsf{All}\ (\lambda_i\ \textsf{CurrentTheory}\ i) \\
\quad (\lambda_i\ \textsf{All}\ (\lambda_t\ \textsf{CurrentTimeInTheory}\ i\ t) \\
\qquad (\lambda_t\ \textsf{All}\ (\lambda_a\ \textsf{Speaker}\ a\ t) \\
\qquad\quad (\lambda_a\ \textsf{All}\ (\lambda_x\ \textsf{Student}\ x) \\
\qquad\qquad (\lambda_x\ \textsf{All}\ (\lambda_j\ \textsf{Equal}\ j\ (\uparrow \textsf{Student}\ x)) \\
\qquad\qquad\quad (\lambda_j\ \textsf{All}\ (\lambda_k\ \textsf{Equal}\ k\ (\uparrow \textsf{All}\ (\lambda_u\ \textsf{CurrentTimeInTheory}\ k\ u) \\
\qquad\qquad\qquad\qquad (\lambda_u\ \textsf{All}\ (\lambda_v\ \textsf{ConsecutiveTo}\ u\ v) \\
\qquad\qquad\qquad\qquad\quad (\lambda_v\ \textsf{Believe}\ j\ a\ v)))) \\
\qquad\qquad\qquad (\lambda_k\ \textsf{Want}\ k\ a\ t))))
\end{array}
$$

then we apply universal modus ponens (UMP) using this translation as the major premise:

$$
\begin{array}{lr}
\overbrace{\textsf{Theory0} \wedge \textsf{Theory1}}^{\text{major premise}},\ \overbrace{\textsf{Theory0} \wedge \textsf{Theory1}}^{\text{minor premise}} & \\
\quad \Rightarrow \textsf{All}\ (\lambda_i\ \textsf{CurrentTheory}\ i) & \\
\qquad (\lambda_i\ \textsf{All}\ (\lambda_t\ \textsf{CurrentTimeInTheory}\ t\ i)\ (\lambda_t\ \ldots)), & \\
\quad\ \ \textsf{CurrentTheory}\ (\uparrow \textsf{Theory1}) & \text{(CE)} \\
\quad = \textsf{All}\ (\lambda_i\ \textsf{CurrentTheory}\ i) & \\
\qquad (\lambda_i\ \textsf{All}\ (\lambda_t\ \textsf{CurrentTimeInTheory}\ t\ i)\ (\lambda_t\ \ldots)), & \\
\quad\ \ (\lambda_i\ \textsf{CurrentTheory}\ i)\ (\uparrow \textsf{Theory1}) & \text{(beta reduction)} \\
\quad \Rightarrow (\lambda_i\ \textsf{All}\ (\lambda_t\ \textsf{CurrentTimeInTheory}\ t\ i)\ (\lambda_t\ \ldots))\ (\uparrow \textsf{Theory1}) & \text{(UMP)} \\
\quad = \textsf{All}\ (\lambda_t\ \textsf{CurrentTimeInTheory}\ t\ (\uparrow \textsf{Theory1}))\ (\lambda_t\ \ldots) & \text{(beta reduction)}
\end{array}
$$

and then repeat using each conclusion as the major premise of another universal modus ponens:

$$
\begin{array}{lr}
\overbrace{\textsf{All}\ (\lambda_t\ \textsf{CurrentTimeInTheory}\ t\ (\uparrow \textsf{Theory1}))\ (\lambda_t\ \textsf{All}\ (\lambda_a\ \textsf{Speaker}\ a\ t)\ (\lambda_a\ \ldots))}^{\text{major premise}},\ \overbrace{\textsf{Theory0} \wedge \textsf{Theory1}}^{\text{minor premise}} & \\
\quad \Rightarrow \textsf{All}\ (\lambda_t\ \textsf{CurrentTimeInTheory}\ t\ (\uparrow \textsf{Theory1}))\ (\lambda_t\ \textsf{All}\ (\lambda_a\ \textsf{Speaker}\ a\ t)\ (\lambda_a\ \ldots)), & \\
\quad\ \ \textsf{CurrentTimeInTheory 10:00InTheory1}\ (\uparrow \textsf{Theory1}) & \text{(CE)} \\
\quad \Rightarrow \textsf{All}\ (\lambda_a\ \textsf{Speaker}\ a\ \textsf{10:00InTheory1})\ (\lambda_a\ \ldots) & \text{(UMP)}
\end{array}
$$

$$
\begin{array}{lr}
\overbrace{\textsf{All}\ (\lambda_a\ \textsf{Speaker}\ a\ \textsf{10:00InTheory1})\ (\lambda_a\ \textsf{All}\ (\lambda_x\ \textsf{Student}\ x)\ (\lambda_x\ \ldots))}^{\text{major premise}},\ \overbrace{\textsf{Theory0} \wedge \textsf{Theory1}}^{\text{minor premise}} & \\
\quad \Rightarrow \textsf{All}\ (\lambda_a\ \textsf{Speaker}\ a\ \textsf{10:00InTheory1})\ (\lambda_a\ \textsf{All}\ (\lambda_x\ \textsf{Student}\ x)\ (\lambda_x\ \ldots)), & \\
\quad\ \ \textsf{Speaker Person7 10:00InTheory1} & \text{(CE)} \\
\quad \Rightarrow \textsf{All}\ (\lambda_x\ \textsf{Student}\ x)\ (\lambda_x\ \ldots) & \text{(UMP)}
\end{array}
$$

$$
\begin{array}{l}
\overbrace{\textsf{All}\ (\lambda_x\ \textsf{Student}\ x)\ (\lambda_x\ \textsf{All}\ (\lambda_j\ \textsf{Equal}\ j\ (\uparrow \textsf{Student}\ x))\ (\lambda_j\ \ldots))}^{\text{major premise}},\ \overbrace{\textsf{Theory0} \wedge \textsf{Theory1}}^{\text{minor premise}} \\
\quad \Rightarrow \textsf{All}\ (\lambda_x\ \textsf{Student}\ x)\ (\lambda_x\ \textsf{All}\ (\lambda_j\ \textsf{Equal}\ j\ (\uparrow \textsf{Student}\ x))\ (\lambda_j\ \ldots)),
\end{array}
$$

$$\text{Student Person43} \quad \text{(CE)}$$
$$\Rightarrow \text{All } (\lambda_j \text{ Equal } j\ (\uparrow \text{Student Person43}))\ (\lambda_j\ \ldots) \quad \text{(UMP)}$$

$$\overbrace{\text{All } (\lambda_j \text{ Equal } j\ (\uparrow \text{Student Person43}))\ (\lambda_j \text{ All } (\lambda_k \text{ Equal } k\ (\uparrow \ldots))\ (\lambda_k\ \ldots)),}^{\text{major premise}}\ \overbrace{\text{Theory0} \wedge \text{Theory1}}^{\text{minor premise}}$$
$$\Rightarrow \text{All } (\lambda_j \text{ Equal } j\ (\uparrow \text{Student Person43}))\ (\lambda_j \text{ All } (\lambda_k \text{ Equal } k\ (\uparrow \ldots))\ (\lambda_k\ \ldots)),$$
$$\text{Equal } (\uparrow \text{Theory1b43})\ (\uparrow \text{Student Person43}) \quad \text{(CE)}$$
$$\Rightarrow \text{All } (\lambda_k \text{ Equal } k\ (\uparrow \ldots))\ (\lambda_k\ \ldots) \quad \text{(UMP)}$$

$$\text{All } (\lambda_k \text{ Equal } k\ (\uparrow \text{All } (\lambda_u \text{ CurrentTimeInTheory } k\ u)$$
$$(\lambda_u \text{ All } (\lambda_v \text{ ConsecutiveTo } u\ v)$$
$$(\lambda_v \text{ Believe } (\uparrow \text{Theory1b43) Person7})\ v)))$$
$$(\lambda_k \text{ Want } k \text{ Person7 10:00InTheory1}),\ \ \text{Theory0} \wedge \text{Theory1}$$
$$\Rightarrow \text{All } (\lambda_k \text{ Equal } k\ (\uparrow \text{All } (\lambda_u \text{ CurrentTimeInTheory } k\ u)$$
$$(\lambda_u \text{ All } (\lambda_v \text{ ConsecutiveTo } u\ v)$$
$$(\lambda_v \text{ Believe } (\uparrow \text{Theory1b43) Person7 } v))))$$
$$(\lambda_k \text{ Want } k \text{ Person7 10:00InTheory1}),$$
$$\text{Equal } (\uparrow \text{Theory1a43})\ (\uparrow \text{All } (\lambda_u \text{ CurrentTimeInTheory } (\uparrow \text{Theory1a43})\ u) \quad \text{(CE)}$$
$$(\lambda_u \text{ All } (\lambda_v \text{ ConsecutiveTo } u\ v)$$
$$(\lambda_v \text{ Believe } (\uparrow \text{Theory1b43) Person7 } v)))$$
$$\Rightarrow \text{Want } (\uparrow \text{Theory1a43) Person7 10:00InTheory1} \quad \text{(UMP)}$$

to obtain a revised Theory1 with Want predicates:

$$\ldots \wedge \quad \text{(everything in the previous Theory1)}$$
$$\text{Want } (\uparrow \text{Theory1a43) Person7 10:00InTheory1} \wedge$$
$$\text{Want } (\uparrow \text{Theory1a58) Person7 10:00InTheory1}$$

Note above we just define Theory1a43 and Theory1b43 to be the expressions they are Equal to.

We can now simplify Theory1a43 using more universal modus ponens (UMP):

$$\overbrace{\text{Theory0} \wedge \text{Theory1a43},}^{\text{major premise}}\ \overbrace{\text{Theory0} \wedge \text{Theory1a43}}^{\text{minor premise}}$$
$$\Rightarrow \text{All } (\lambda_u \text{ CurrentTimeInTheory } (\uparrow \text{Theory1a43})\ u)$$
$$(\lambda_u \text{ All } (\lambda_v \text{ ConsecutiveTo } u\ v)$$
$$(\lambda_v \text{ Believe } (\uparrow \text{Theory1b43) Person7 } v)),$$
$$\text{CurrentTimeInTheory } (\uparrow \text{Theory1a43) 10:00InTheory1a43} \quad \text{(CE)}$$
$$\Rightarrow \text{All } (\lambda_v \text{ ConsecutiveTo 10:00InTheory1a43 } v) \quad \text{(UMP)}$$
$$(\lambda_v \text{ Believe } (\uparrow \text{Theory1b43) Person7 } v)$$

$$\text{All } (\lambda_v \text{ ConsecutiveTo 10:00InTheory1a43 } v)$$
$$(\lambda_v \text{ Believe } (\uparrow \text{Theory1b43) Person7 } v),\ \ \text{Theory0} \wedge \text{Theory1a43}$$

$$\Rightarrow \mathsf{All}\ (\lambda_v\ \mathsf{ConsecutiveTo}\ \mathsf{10{:}00InTheory1a43}\ v)\ (\lambda_v\ \mathsf{Believe}\ (\uparrow \mathsf{Theory1b43})\ \mathsf{Person7}\ v),$$

$$\mathsf{ConsecutiveTo}\ \mathsf{10{:}00InTheory1a43}\ \mathsf{10{:}01InTheory1a43} \qquad \text{(CE)}$$

$$\Rightarrow \mathsf{Believe}\ (\uparrow \mathsf{Theory1b43})\ \mathsf{Person7}\ \mathsf{10{:}01InTheory1a43} \qquad \text{(UMP)}$$

then Theory1a43 is:

$$\mathsf{Believe}\ (\uparrow \mathsf{Theory1b43})\ \mathsf{Person7}\ \mathsf{10{:}01InTheory1a43}$$

and Theory1b43 is:

$$\mathsf{Student}\ \mathsf{Person43}$$

and Theory1a58 and Theory1b58 are the same, but for Person58.

This establishes what, specifically, the authorized user wants to know, given who the students are.

9.4 Example: following a plan to answer

Now we consider a plan π to answer authorized users' queries, and then advance the clock:

$$\begin{array}{l}
\mathsf{All}\ (\lambda_i\ \mathsf{CurrentTheory}\ i)\\
\quad (\lambda_i\ \mathsf{All}\ (\lambda_t\ \mathsf{CurrentTimeInTheory}\ i\ t)\\
\qquad (\lambda_t\ \mathsf{All}\ (\lambda_u\ \mathsf{ConsecutiveTo}\ u\ t)\\
\qquad\quad (\lambda_u\ \mathsf{All}\ (\lambda_a\ \mathsf{Admin}\ a)\\
\qquad\qquad (\lambda_a\ \mathsf{All}\ (\lambda_b\ \mathsf{AuthorizedUser}\ b)\\
\qquad\qquad\quad (\lambda_b\ \mathsf{All}\ (\lambda_j\ \mathsf{Want}\ j\ b\ u)\\
\qquad\qquad\qquad (\lambda_j\ \mathsf{All}\ (\lambda_v\ \mathsf{CurrentTimeInTheory}\ j\ v)\\
\qquad\qquad\qquad\quad (\lambda_v\ \mathsf{All}\ (\lambda_k\ \mathsf{Believe}\ k\ b\ v)\\
\qquad\qquad\qquad\qquad (\lambda_k\ \mathsf{Tell}\ k\ b\ a\ t)))))))
\end{array}$$

then after more universal modus ponens:

$$\pi,\ \ \mathsf{Theory0} \wedge \mathsf{Theory1}$$

$$\Rightarrow \mathsf{All}\ (\lambda_i\ \mathsf{CurrentTheory}\ i)\ (\lambda_i\ \mathsf{All}\ (\lambda_t\ \mathsf{CurrentTimeInTheory}\ t\ i)\ (\lambda_t\ \ldots)),$$

$$\mathsf{CurrentTheory}\ (\uparrow \mathsf{Theory1}) \qquad \text{(CE)}$$

$$\Rightarrow \mathsf{All}\ (\lambda_t\ \mathsf{CurrentTimeInTheory}\ t\ (\uparrow \mathsf{Theory1}))\ (\lambda_t\ \ldots) \qquad \text{(UMP)}$$

$$\mathsf{All}\ (\lambda_t\ \mathsf{CurrentTimeInTheory}\ t\ (\uparrow \mathsf{Theory1}))\ (\lambda_t\ \mathsf{All}\ (\lambda_u\ \mathsf{ConsecutiveTo}\ t\ u)\ (\lambda_u\ \ldots)),\ \ \mathsf{Theory0} \wedge \mathsf{Theory1}$$

$$\Rightarrow \mathsf{All}\ (\lambda_t\ \mathsf{CurrentTimeInTheory}\ t\ (\uparrow \mathsf{Theory1}))\ (\lambda_t\ \mathsf{All}\ (\lambda_u\ \mathsf{ConsecutiveTo}\ t\ u)\ (\lambda_u\ \ldots)),$$

$$\mathsf{CurrentTimeInTheory}\ \mathsf{10{:}00InTheory1}\ (\uparrow \mathsf{Theory1}) \qquad \text{(CE)}$$

$$\Rightarrow \mathsf{All}\ (\lambda_t\ \mathsf{ConsecutiveTo}\ \mathsf{10{:}00InTheory1}\ u)\ (\lambda_t\ \ldots) \qquad \text{(UMP)}$$

$$\vdots$$

$$\begin{aligned}
&\mathsf{All}\ (\lambda_j\ \mathsf{Want}\ j\ \mathsf{Person7}\ \mathsf{10{:}00InTheory1})\\
&\quad(\lambda_j\ \mathsf{All}\ (\lambda_v\ \mathsf{CurrentTimeInTheory}\ v\ j)\ (\lambda_v\ \ldots)),\ \ \mathsf{Theory0} \wedge \mathsf{Theory1}\\
&\qquad \Rightarrow\ \mathsf{All}\ (\lambda_j\ \mathsf{Want}\ j\ \mathsf{Person7}\ \mathsf{10{:}00InTheory1})\\
&\qquad\qquad (\lambda_j\ \mathsf{All}\ (\lambda_v\ \mathsf{CurrentTimeInTheory}\ v\ j)\ (\lambda_v\ \ldots)),\\
&\qquad\quad\ \mathsf{Want}\ (\uparrow \mathsf{Theory1a43})\ \mathsf{Person7}\ \mathsf{10{:}00InTheory1} && \text{(CE)}\\
&\qquad \Rightarrow\ \mathsf{All}\ (\lambda_v\ \mathsf{CurrentTimeInTheory}\ v\ (\uparrow \mathsf{Theory1a43}))\ (\lambda_v\ \ldots) && \text{(UMP)}
\end{aligned}$$

$$\begin{aligned}
&\mathsf{All}\ (\lambda_v\ \mathsf{CurrentTimeInTheory}\ v\ (\uparrow \mathsf{Theory1a43}))\\
&\quad(\lambda_v\ \mathsf{All}\ (\lambda_k\ \mathsf{Believe}\ k\ \mathsf{Person7}\ v)\ (\lambda_k\ \ldots)),\ \ \mathsf{Theory0} \wedge \mathsf{Theory1}\\
&\qquad \Rightarrow\ \mathsf{All}\ (\lambda_v\ \mathsf{CurrentTimeInTheory}\ v\ (\uparrow \mathsf{Theory1a43}))\\
&\qquad\qquad (\lambda_v\ \mathsf{All}\ (\lambda_k\ \mathsf{Believe}\ k\ \mathsf{Person7}\ v)\ (\lambda_k\ \ldots)),\\
&\qquad\quad\ \mathsf{CurrentTimeInTheory}\ \mathsf{10{:}01InTheory1a43}\ (\uparrow \mathsf{Theory1a43}) && \text{(CE)}\\
&\qquad \Rightarrow\ \mathsf{All}\ (\lambda_k\ \mathsf{Believe}\ k\ \mathsf{Person7}\ \mathsf{10{:}01InTheory1a43})\ (\lambda_k\ \ldots) && \text{(UMP)}
\end{aligned}$$

$$\begin{aligned}
&\mathsf{All}\ (\lambda_k\ \mathsf{Believe}\ k\ \mathsf{Person7}\ \mathsf{10{:}01InTheory1a43})\\
&\quad(\lambda_k\ \mathsf{Tell}\ k\ \mathsf{Person7}\ \mathsf{Person1}\ \mathsf{10{:}01InTheory1}),\ \ \mathsf{Theory0} \wedge \mathsf{Theory1}\\
&\qquad \Rightarrow\ \mathsf{All}\ (\lambda_k\ \mathsf{Believe}\ k\ \mathsf{Person7}\ \mathsf{10{:}01InTheory1a43})\\
&\qquad\qquad (\lambda_k\ \mathsf{Tell}\ k\ \mathsf{Person7}\ \mathsf{Person1}\ \mathsf{10{:}01InTheory1}),\\
&\qquad\quad\ \mathsf{Believe}\ (\uparrow \mathsf{Theory1b43})\ \mathsf{Person7}\ \mathsf{10{:}01InTheory1a43} && \text{(CE)}\\
&\qquad \Rightarrow\ \mathsf{Tell}\ (\uparrow \mathsf{Theory1b43})\ \mathsf{Person7}\ \mathsf{Person1}\ \mathsf{10{:}01InTheory1} && \text{(UMP)}
\end{aligned}$$

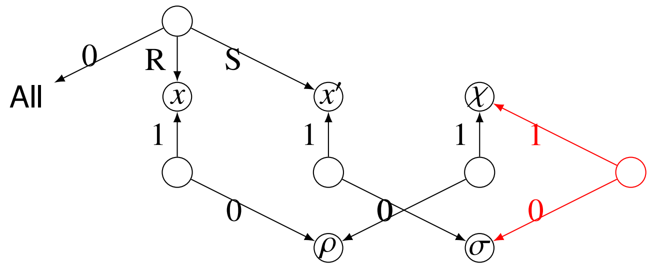
we obtain a revised Theory1 with Tell predicates:

$$\begin{aligned}
&\ldots\ \wedge\quad \text{(everything in the previous } \mathsf{Theory1}\text{)}\\
&\mathsf{Tell}\ (\uparrow \mathsf{Theory1b43})\ \mathsf{Person7}\ \mathsf{Person1}\ \mathsf{10{:}01InTheory1} \wedge\\
&\mathsf{Tell}\ (\uparrow \mathsf{Theory1b58})\ \mathsf{Person7}\ \mathsf{Person1}\ \mathsf{10{:}01InTheory1}
\end{aligned}$$

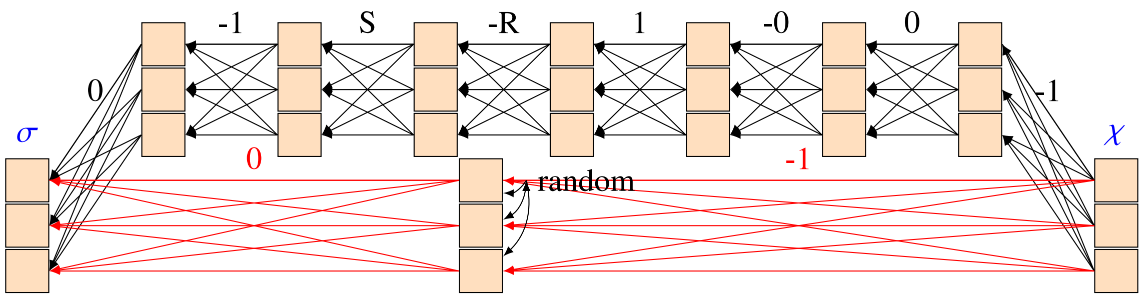
This conclusion is for the administrator to tell the user that Person43 and Person58 are students. It's a logical consequence of the current theory, the user's stated desire and the administrator plan.

9.5 Extra: A neural network for universal modus ponens

Here is a neural representation of universal modus ponens $\mathsf{All}\ (\lambda_{x:\mathsf{e}}\ \rho\ x)\ (\lambda_{x':\mathsf{e}}\ \sigma\ x'),\ \rho\,\chi \Rightarrow \sigma\,\chi$:



Starting at χ, at right, it cues existing associations to σ, then forms new associations from χ to σ:



(This requires inverse associations from targets to cues, represented here as negative labels.)

9.6 Extra: Decision theory (von Neumann & Morgenstern, 1944)

How do we choose a plan among several options? By estimating probabilities and rewards.

We first define a special probability over 'trials' υ, which are **possible successor** times to τ:

$$\mathsf{P}_\tau(\lambda_\upsilon\ \sigma\ \upsilon \mid \varphi) \quad = \quad \mathsf{P}(\lambda_\upsilon\ \sigma\ \upsilon \mid \lambda_\upsilon\ \varphi \wedge \mathsf{PossibleSuccessor}\ \tau\ \upsilon)$$

(We'll use $\mathsf{Suc}\ \tau$ to define a unique **actual successor** to τ, and $\mathsf{Cur}\ \tau$ to flag τ as **current**.)

We then use these probabilities in a decision process to calculate time-averaged expected utility:

$$\mathsf{U}(\pi,\ \theta,\ \tau) = \overbrace{\mathsf{R}_\theta(\tau)}^{\text{reward}} \cdot \begin{cases} \text{if } \exists_\kappa\ \overbrace{\pi,\ \theta \wedge \mathsf{Cur}\ \tau \Rightarrow \kappa}^{\text{next action } \kappa \text{ of } \pi}\text{:} & \sum_\omega \overbrace{\mathsf{P}_\tau(\omega \mid \theta \wedge \kappa)}^{\text{prob. of outcome } \omega \text{ of } \kappa} \cdot \overbrace{\mathsf{U}(\pi,\ \underbrace{\theta \wedge \kappa \wedge \omega\ (\mathsf{Suc}\ \tau)}_{\text{new } \theta \text{ at next } \tau},\ \mathsf{Suc}\ \tau)}^{\text{repeat with } \theta,\ \kappa \text{ and } \omega \text{ as new } \theta} \\ \text{otherwise:} & \dfrac{1}{\tau} \end{cases}$$

(It assumes the plan π is perfectly specific: at most one next action κ for each theory θ.)

This tells us how good a plan is. We can use it to choose among plans to maximize average reward!

For example, if the goal hill is one step away, we get a reward in one step, so $\mathsf{U}(\tau,\ \varphi,\ \pi) = 1$.

But if it's muddy and we slip half the time and don't go anywhere, then:

$$\mathsf{AEU}(\tau, \varphi, \pi) = \begin{cases} .5 \text{ (slip)} & \times \begin{cases} .5 \text{ (slip)} & \times \begin{cases} .5 \text{ (slip)} & \times \ \ldots \\ +.5 \text{ (no slip)} & \times \frac{1}{3} \text{ (arrive in 3 steps)} \end{cases} \\ +.5 \text{ (no slip)} & \times \frac{1}{2} \text{ (arrive in 2 steps)} \end{cases} \\ +.5 \text{ (no slip)} & \times \frac{1}{1} \text{ (arrive in 1 step)} \end{cases}$$
$$\approx .7$$

So if we have two plans (clear path and muddy path) and we know mud slows us, we can avoid it.

References

von Neumann, J. & Morgenstern, O. (1944). Theory of games and economic behavior. *Science and Society*, 9(4), 366–369.