

# Alignment-Based Compositional Semantics for Instruction Following



Jacob Andreas and Dan Klein

UC Berkeley

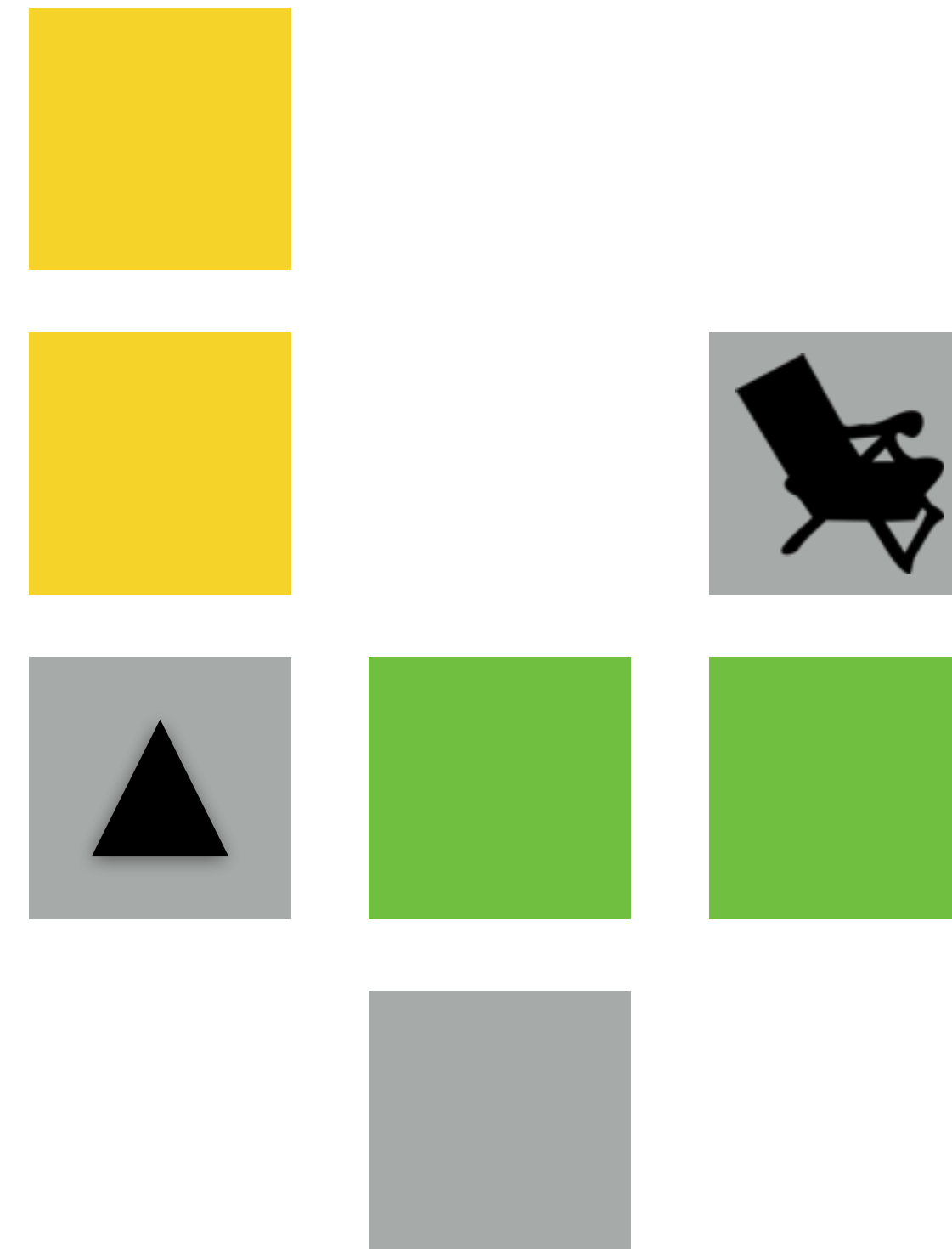


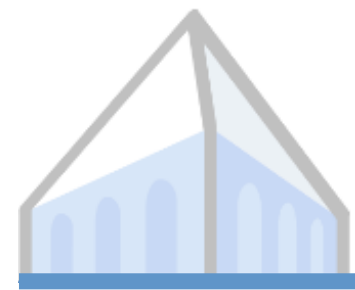
# Following instructions

*go down the green hall*

*go all the way*

*then face the chair*

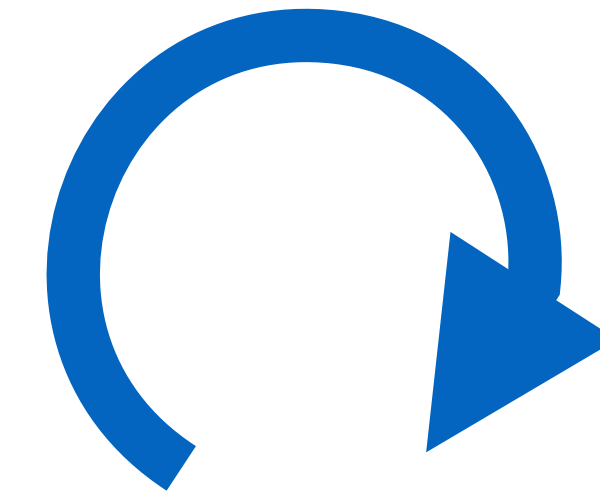




# Following instructions

---

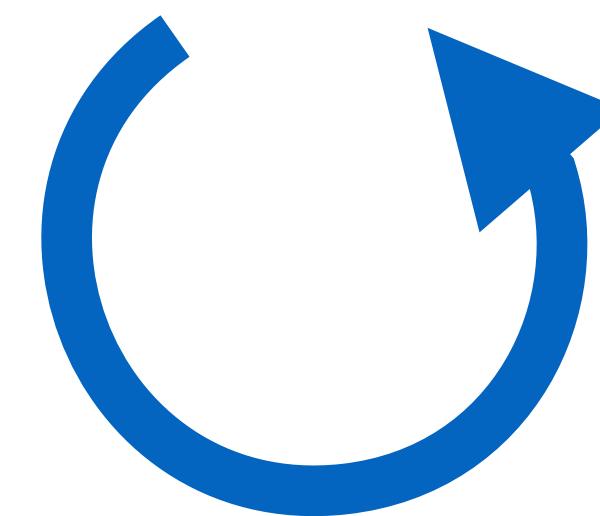
*go down the green hall*

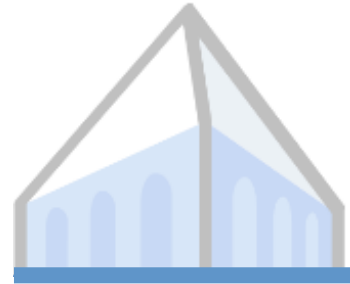


*go all the way*



*then face the chair*





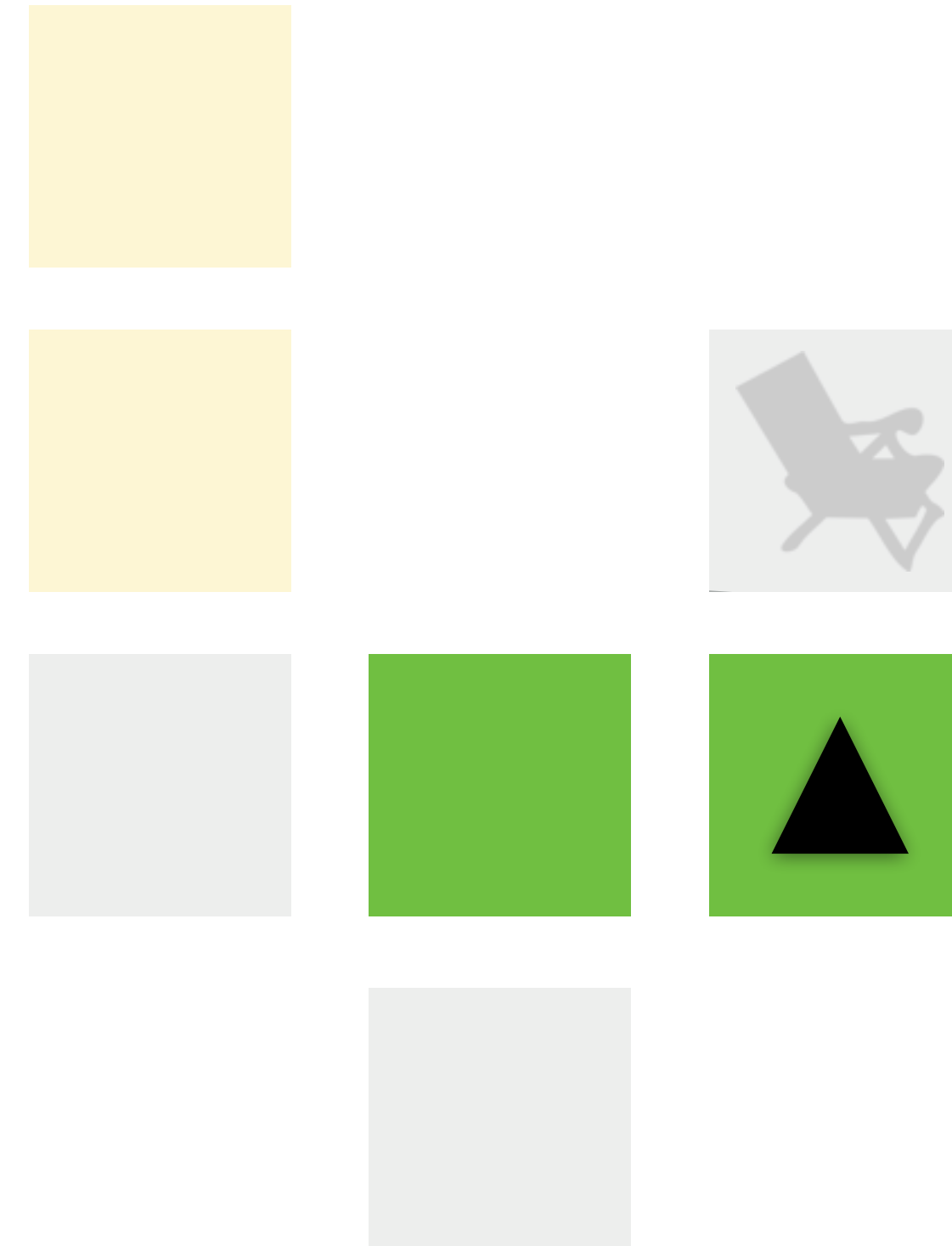
# Following instructions: parsing

---

*go down the **green** hall*

*go all the way*

*then face the chair*



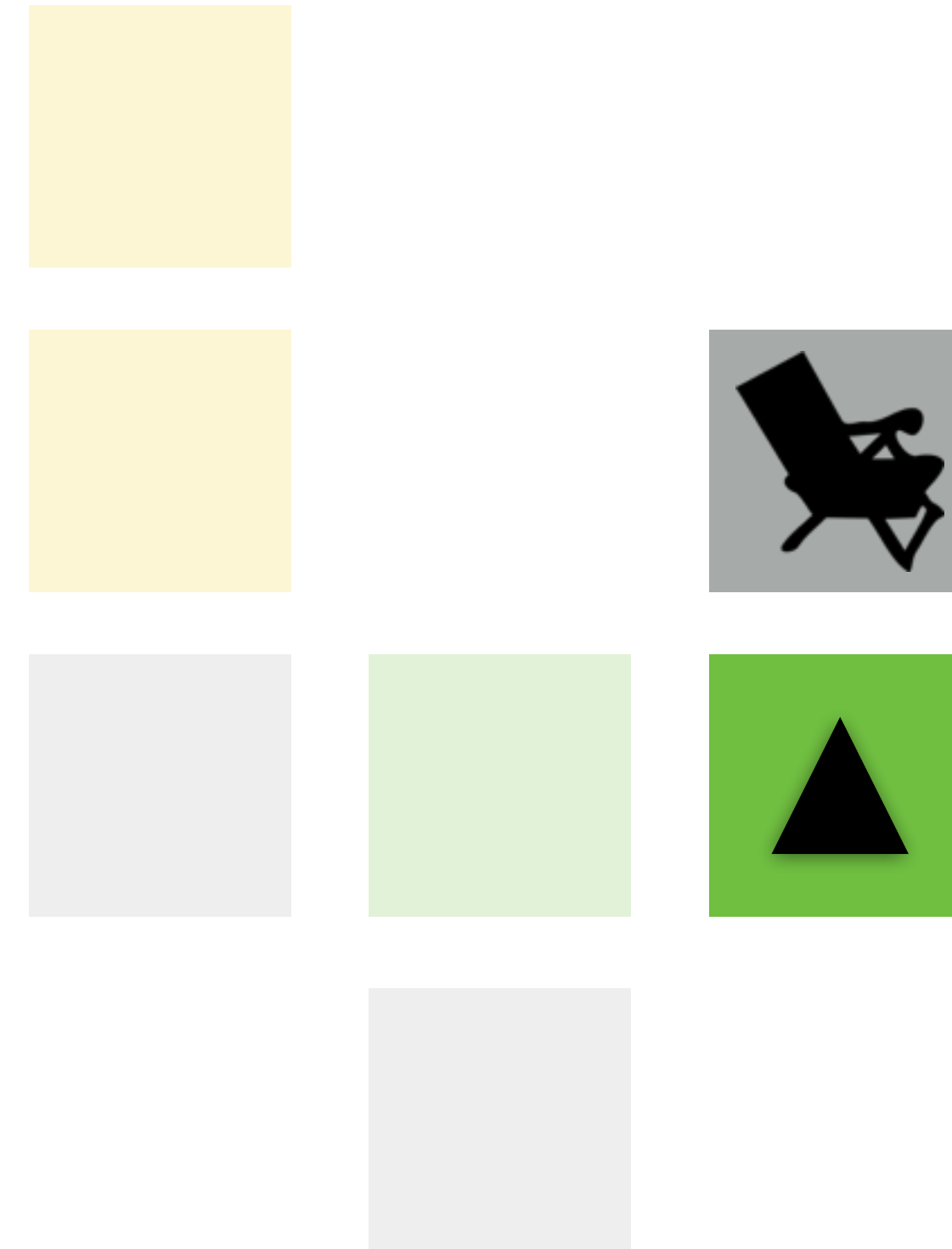


# Following instructions: parsing

*go down the green hall*

*go all the way*

*then face the chair*





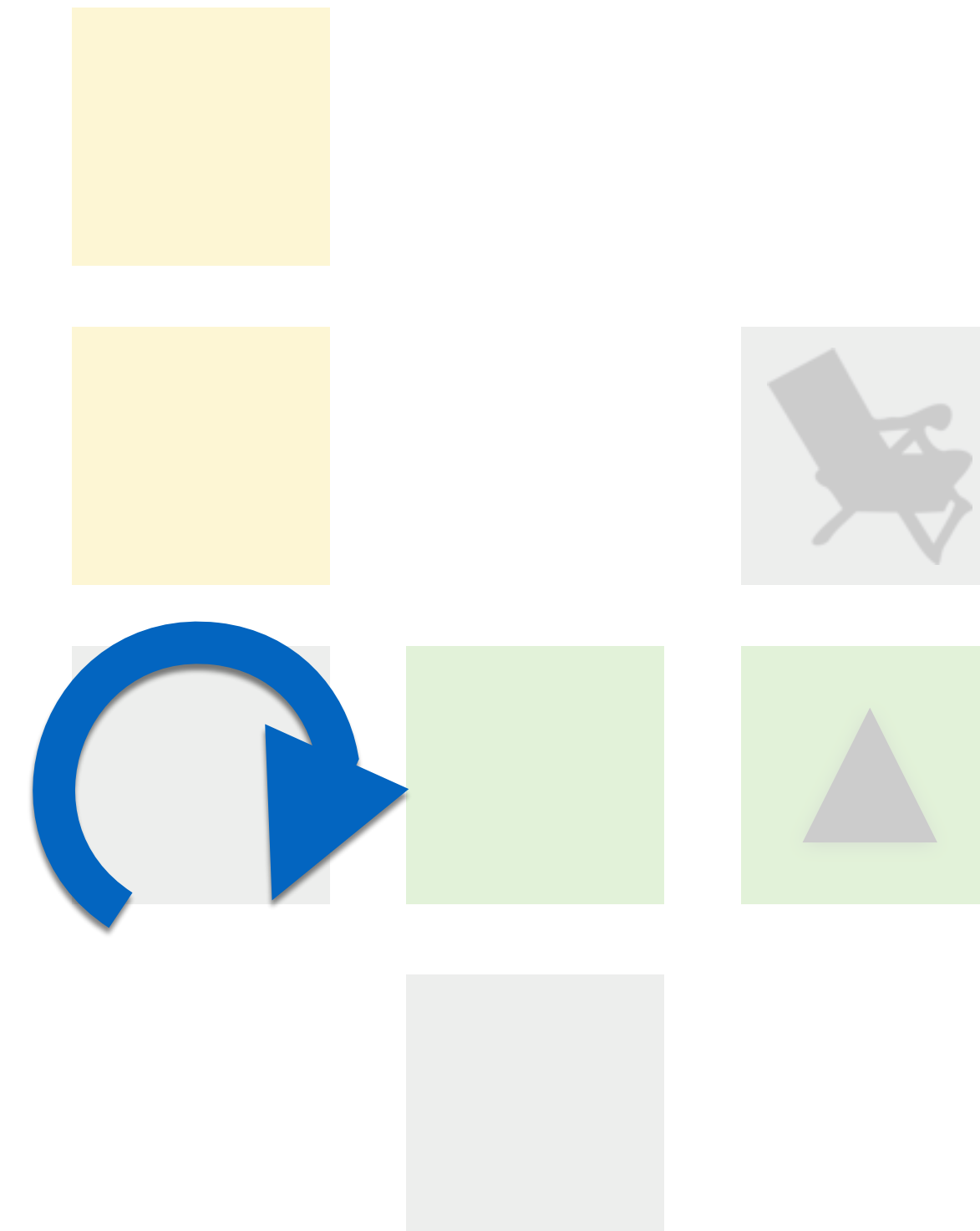
# Following instructions: planning

---

*go down the green hall*

*go all the way*

*then face the chair*





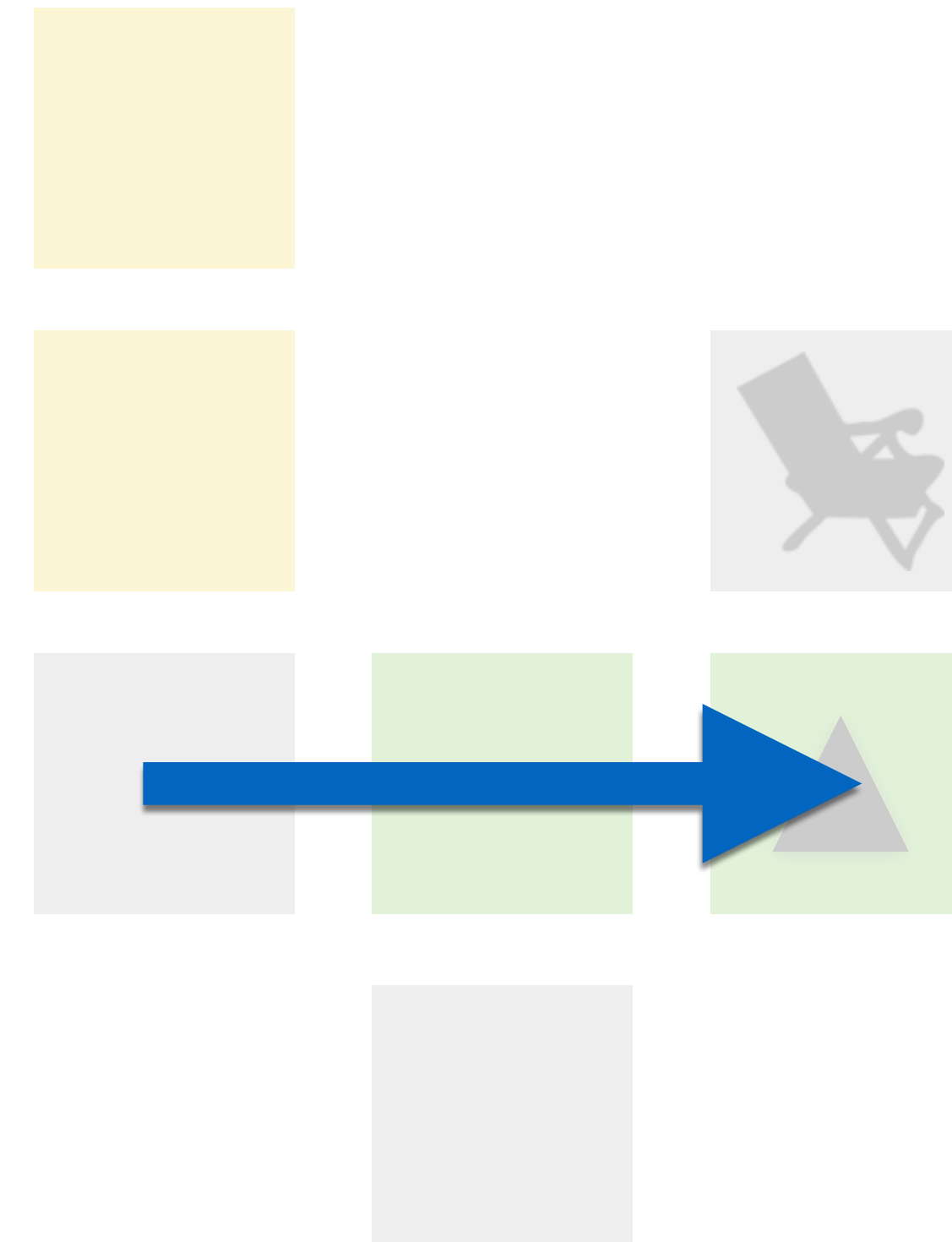
# Following instructions: planning

---

*go down the green hall*

*go all the way*

*then face the chair*





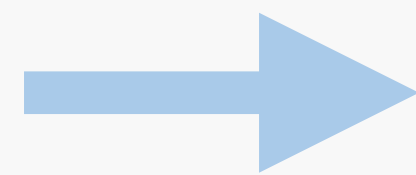
# Following instructions

## Planning

(no instruction)

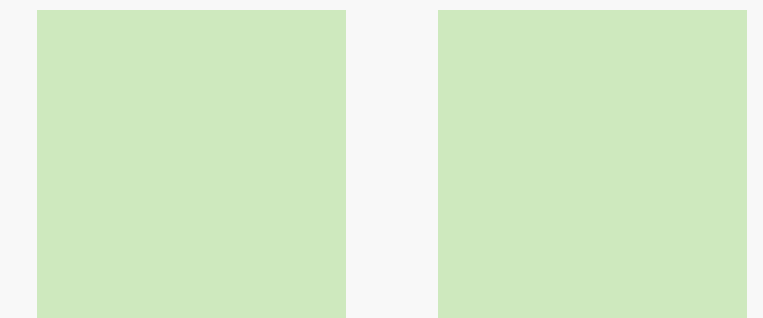


[ *go down the hall,  
go all the way* ]



## Parsing

*green hall*



*face the chair*







# Previous work

---

## Planning

Branavan et al. (2009)

Vogel and Jurafsky (2010)

## Parsing

Chen and Mooney (2011)

Artzi and Zettlemoyer (2013)

Kim and Mooney (2013)

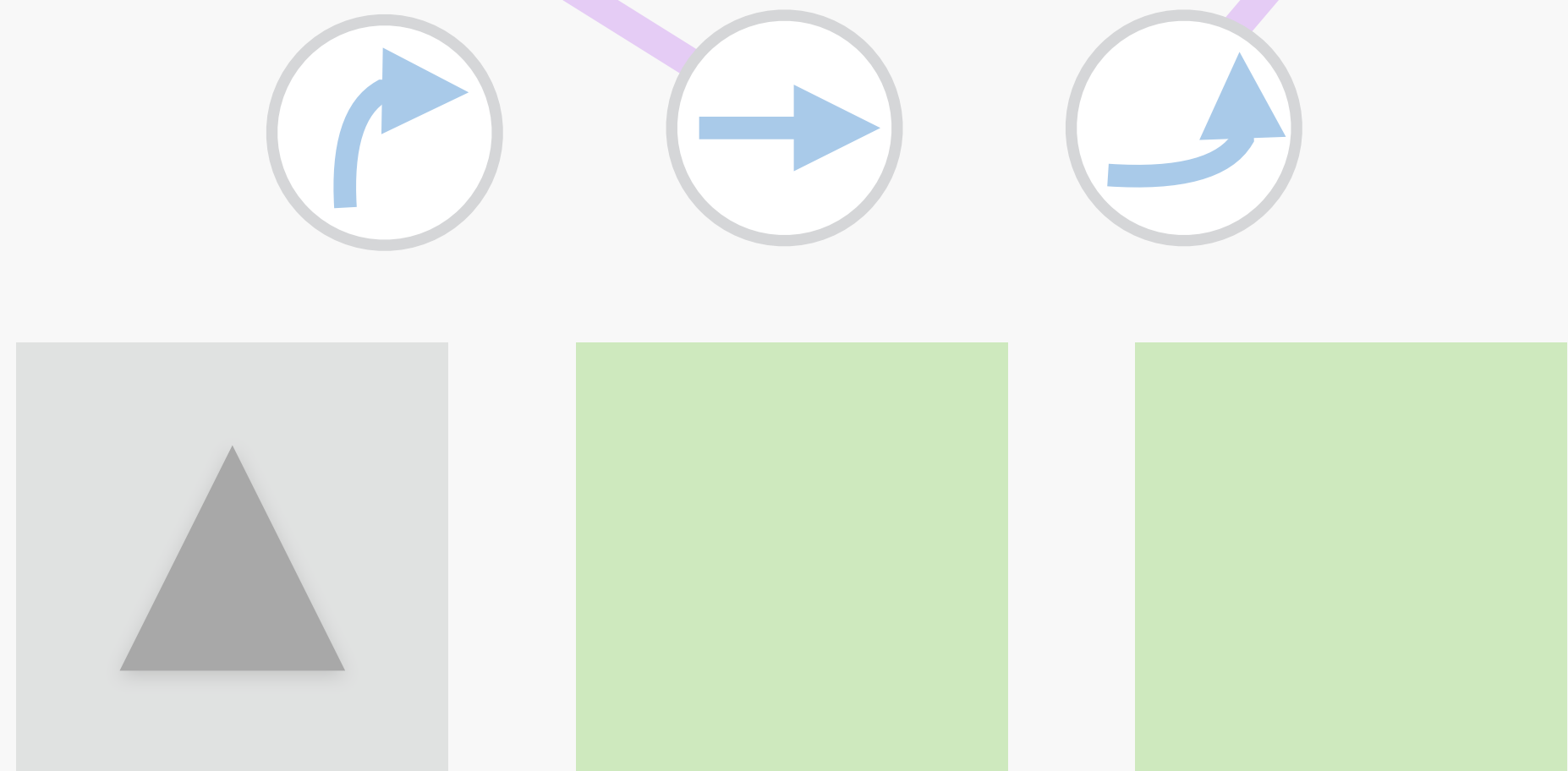
Tellex et al. (variously)



# Parsing and planning

## Sequence-to-sequence

*go down the green hall* *turn left*



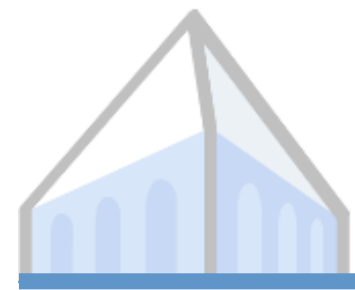
## Structure-to-structure

*go* *down* *the* *green* *hall*



# SEQUENCE-TO-SEQUENCE MODEL

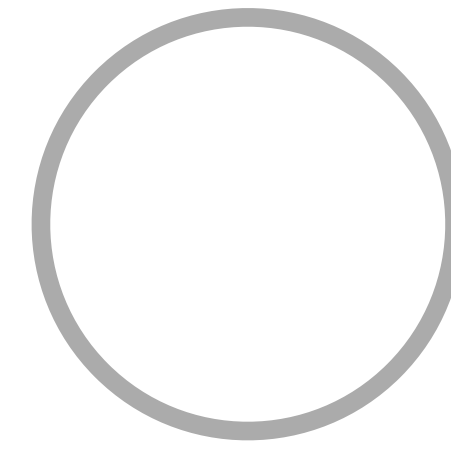
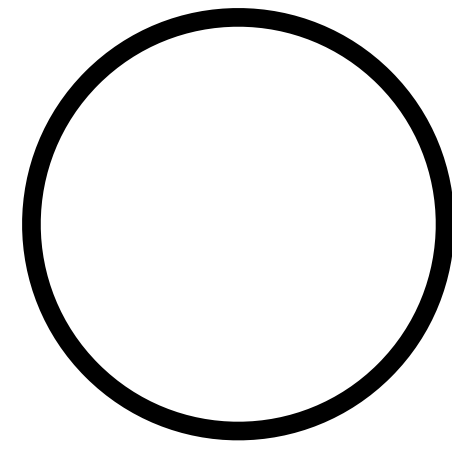




# Sequence-to-sequence

---

*go down the yellow hall*    *turn left*

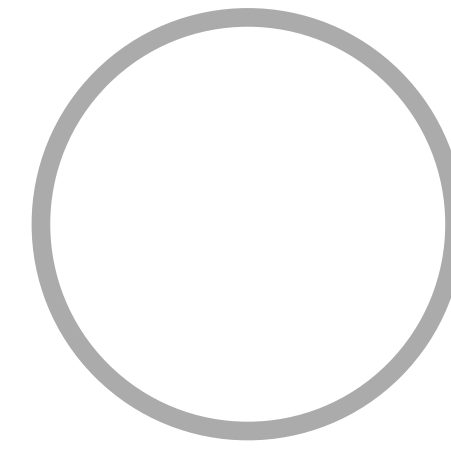
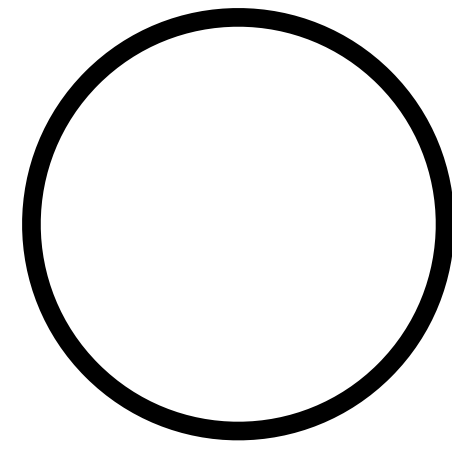




# Sequence-to-sequence

---

*go down the yellow hall*    *turn left*

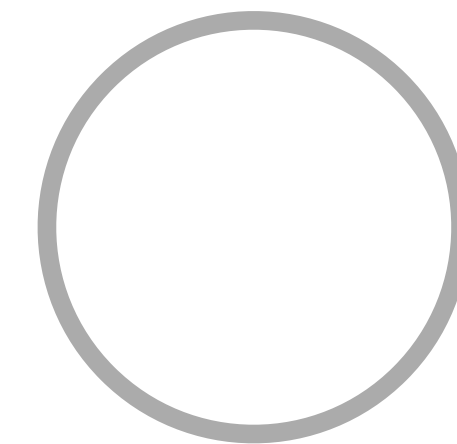
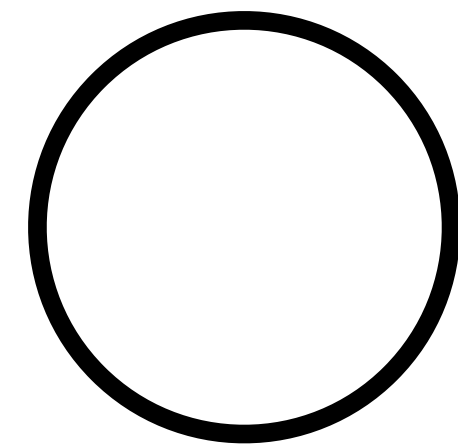


Alignments

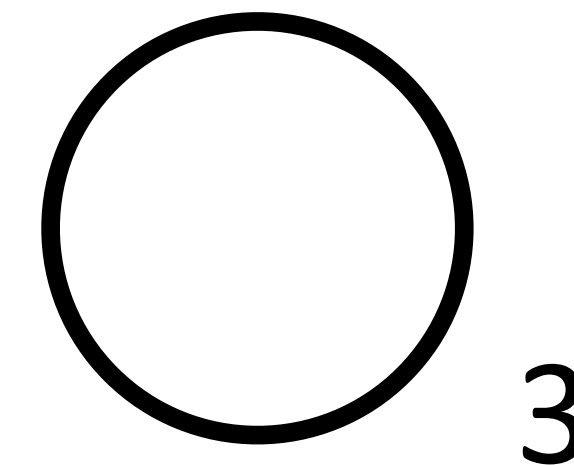
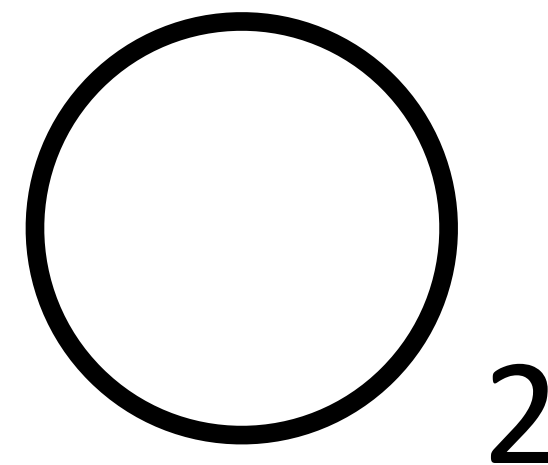
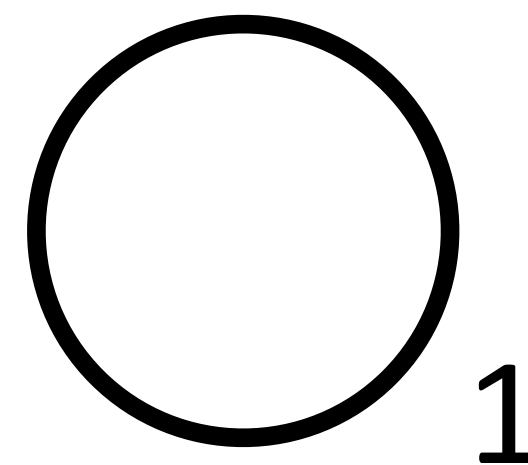


# Sequence-to-sequence

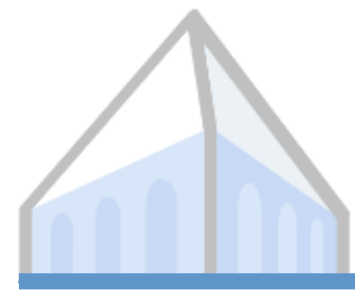
*go down the yellow hall*    *turn left*



Alignments

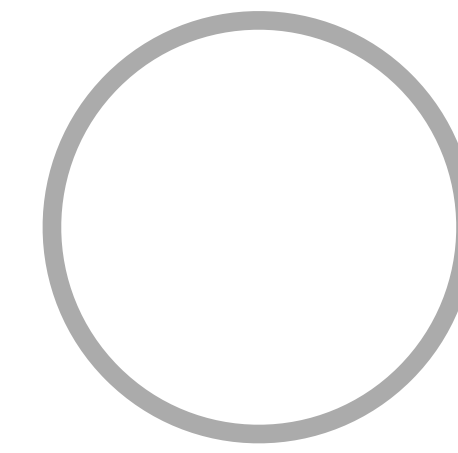
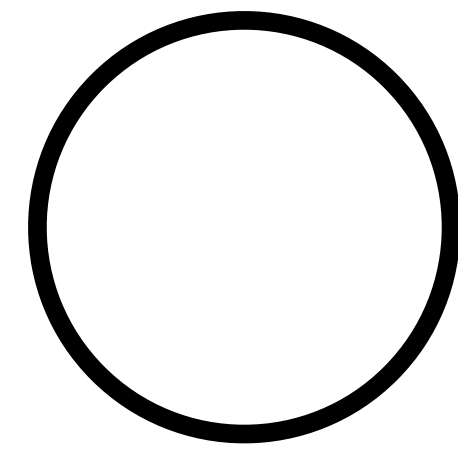


Actions

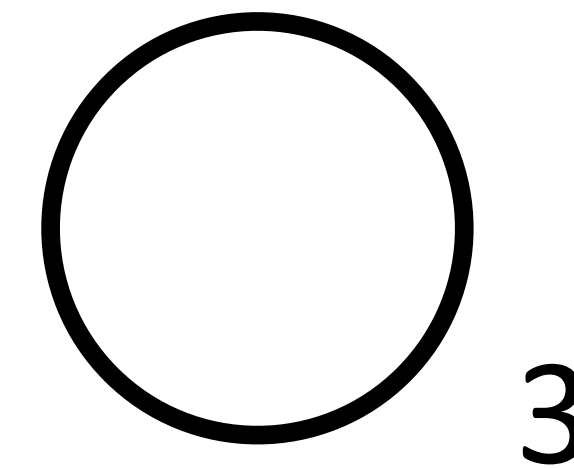
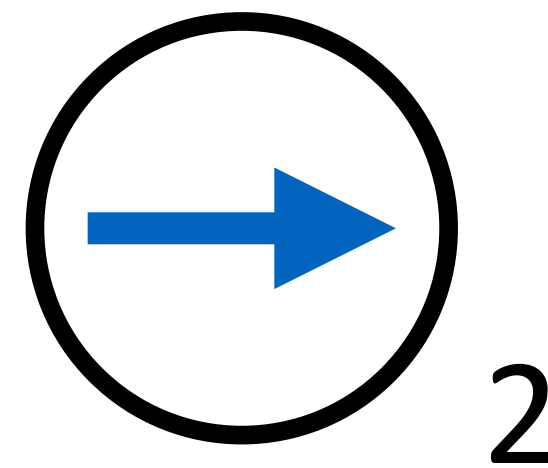
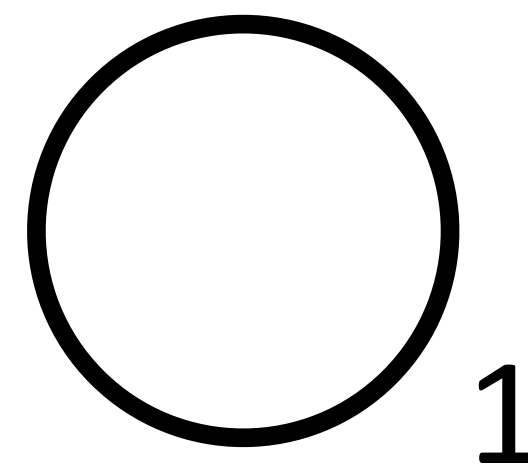


# Sequence-to-sequence

*go down the yellow hall*    *turn left*



Alignments



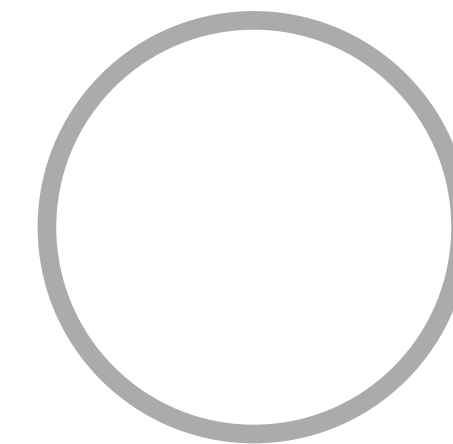
Actions



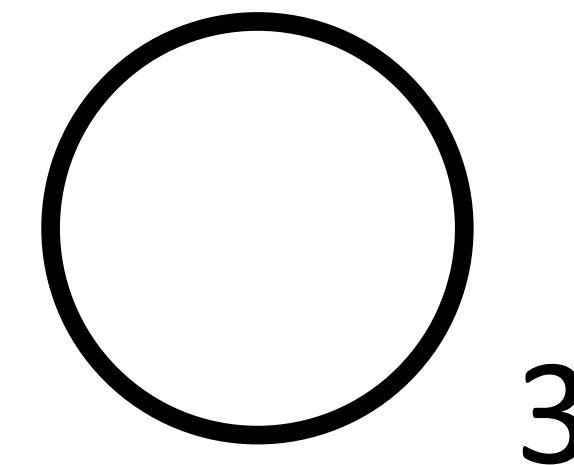
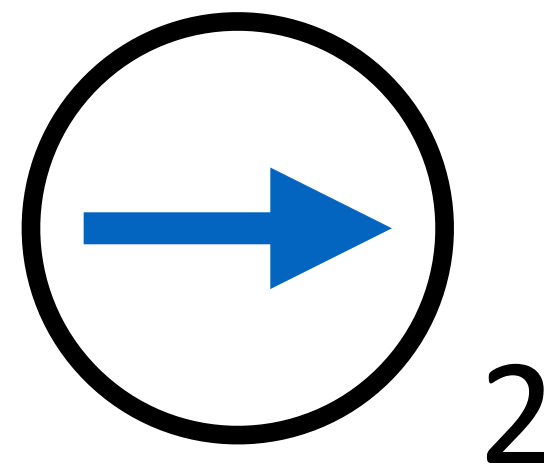
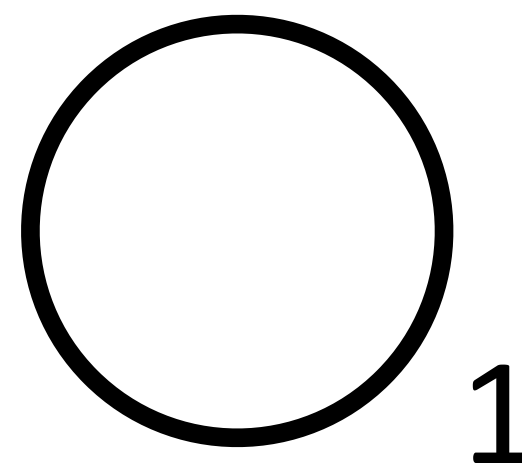
# Sequence-to-sequence

*go down the yellow hall*    *turn left*

2



Alignments



Actions

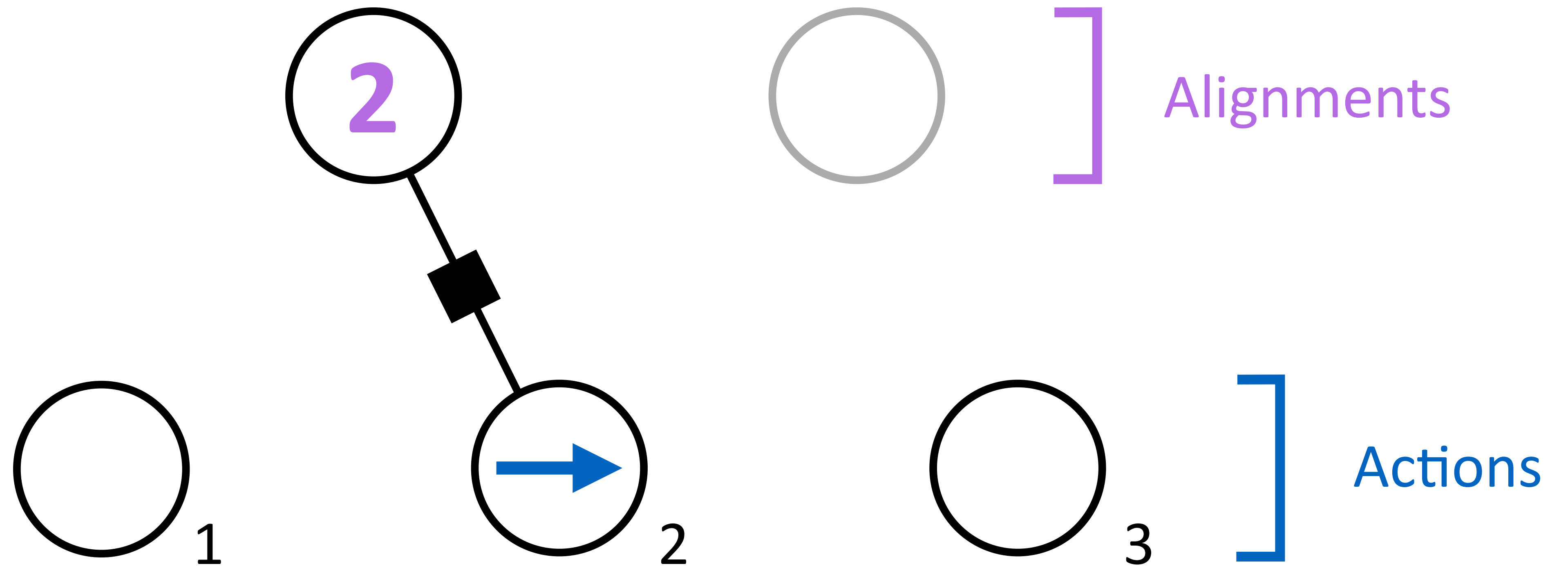




# Sequence-to-sequence

*go down the yellow hall*

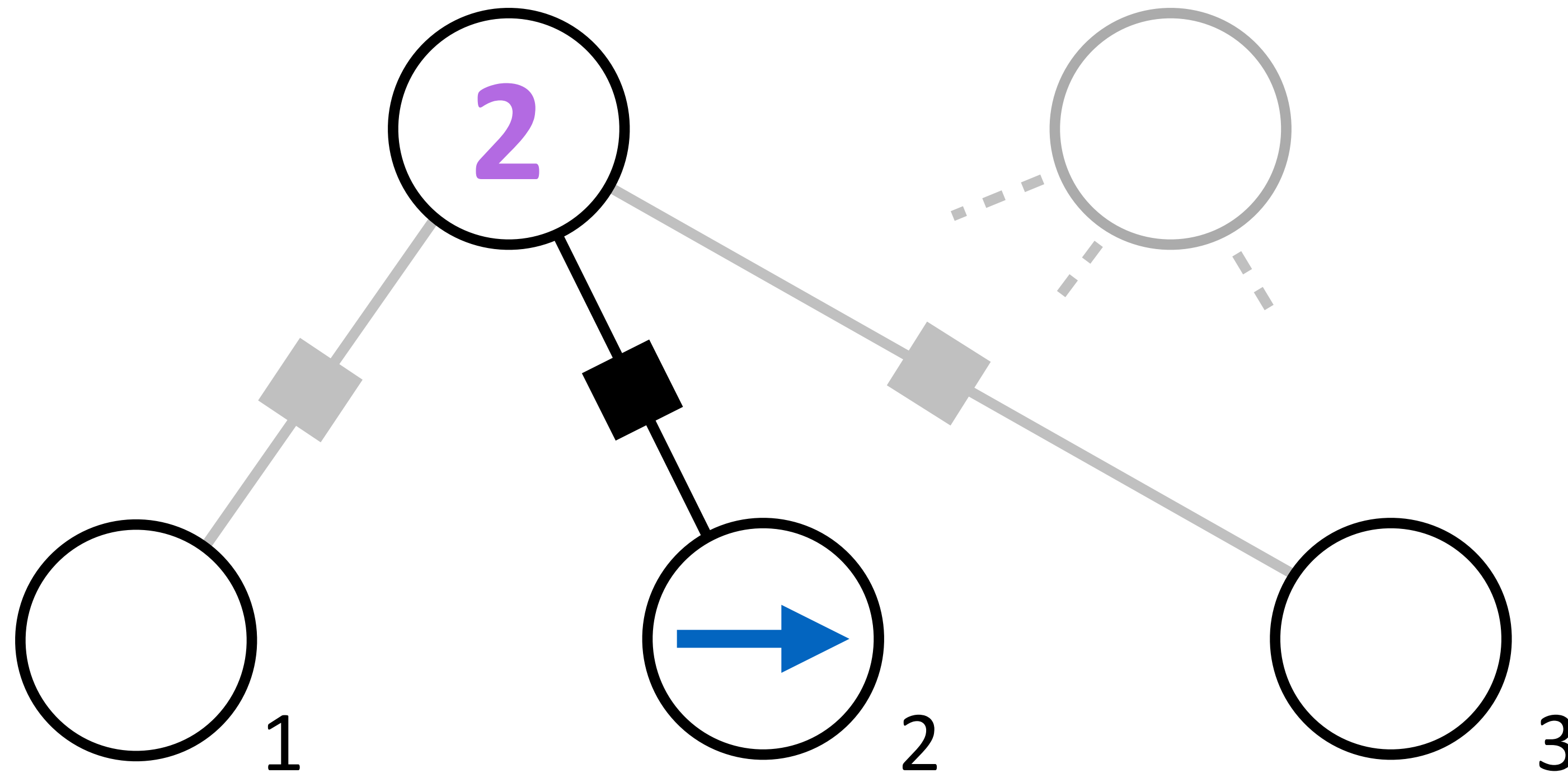
*turn left*

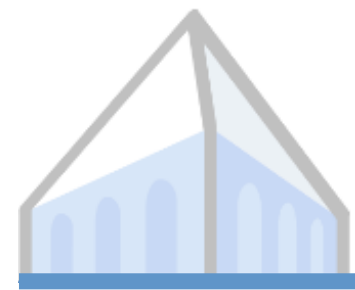




# Sequence-to-sequence

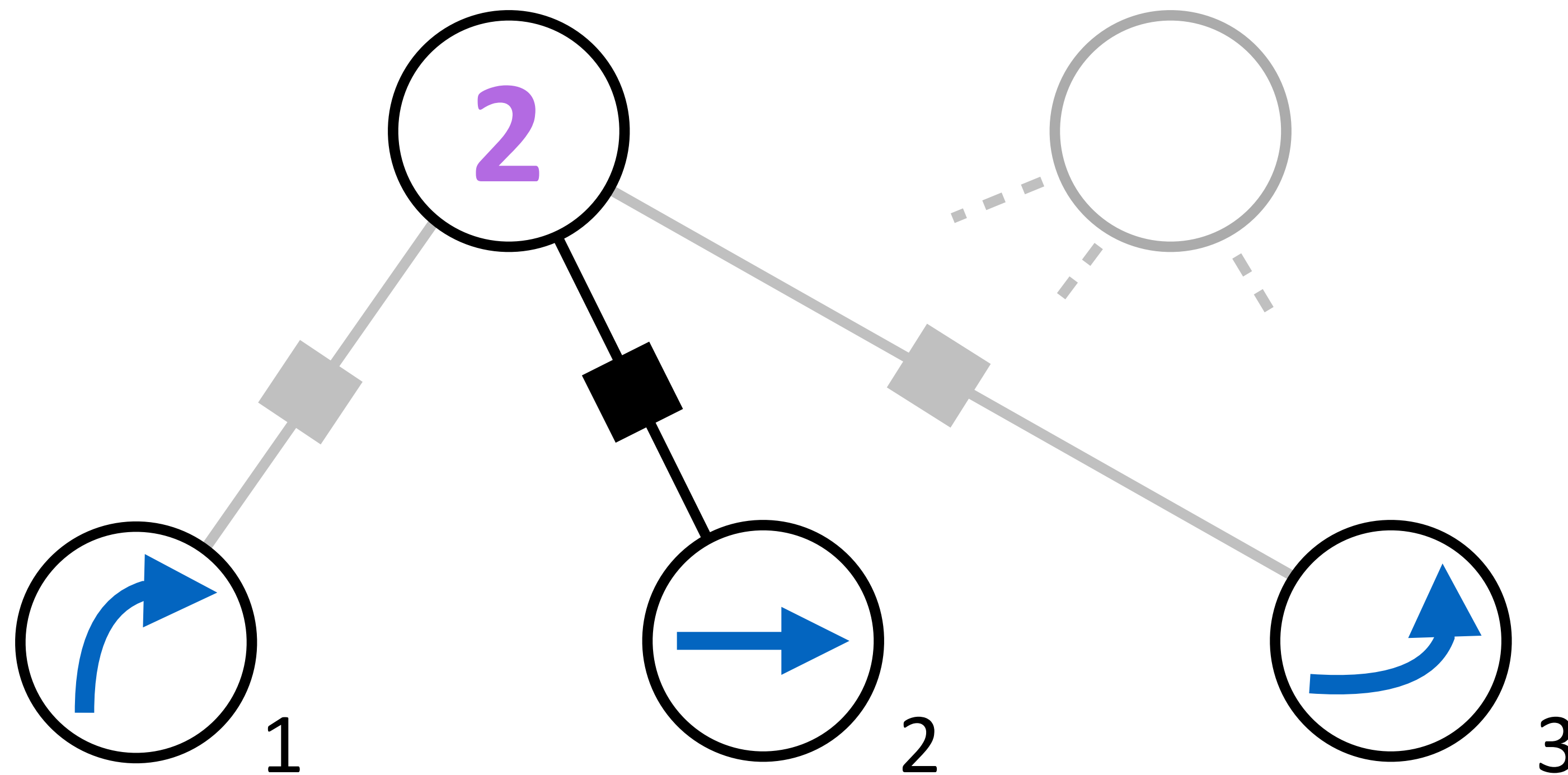
*go down the yellow hall*    *turn left*

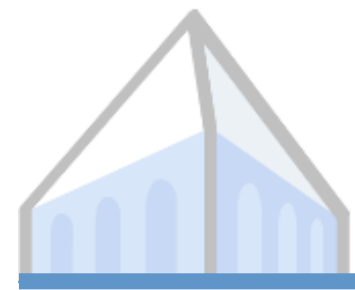




# Sequence-to-sequence

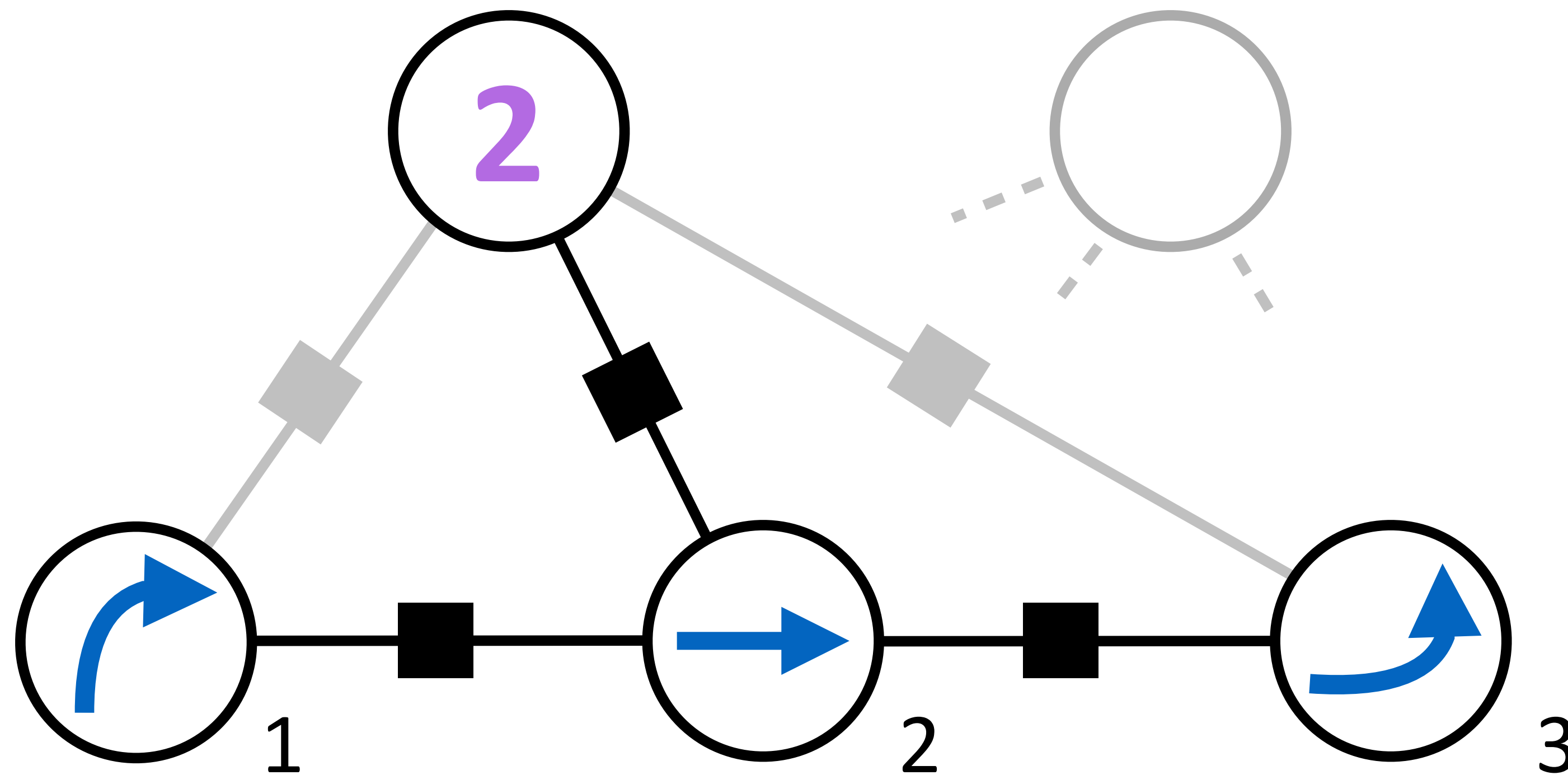
*go down the yellow hall*    *turn left*





# Sequence-to-sequence

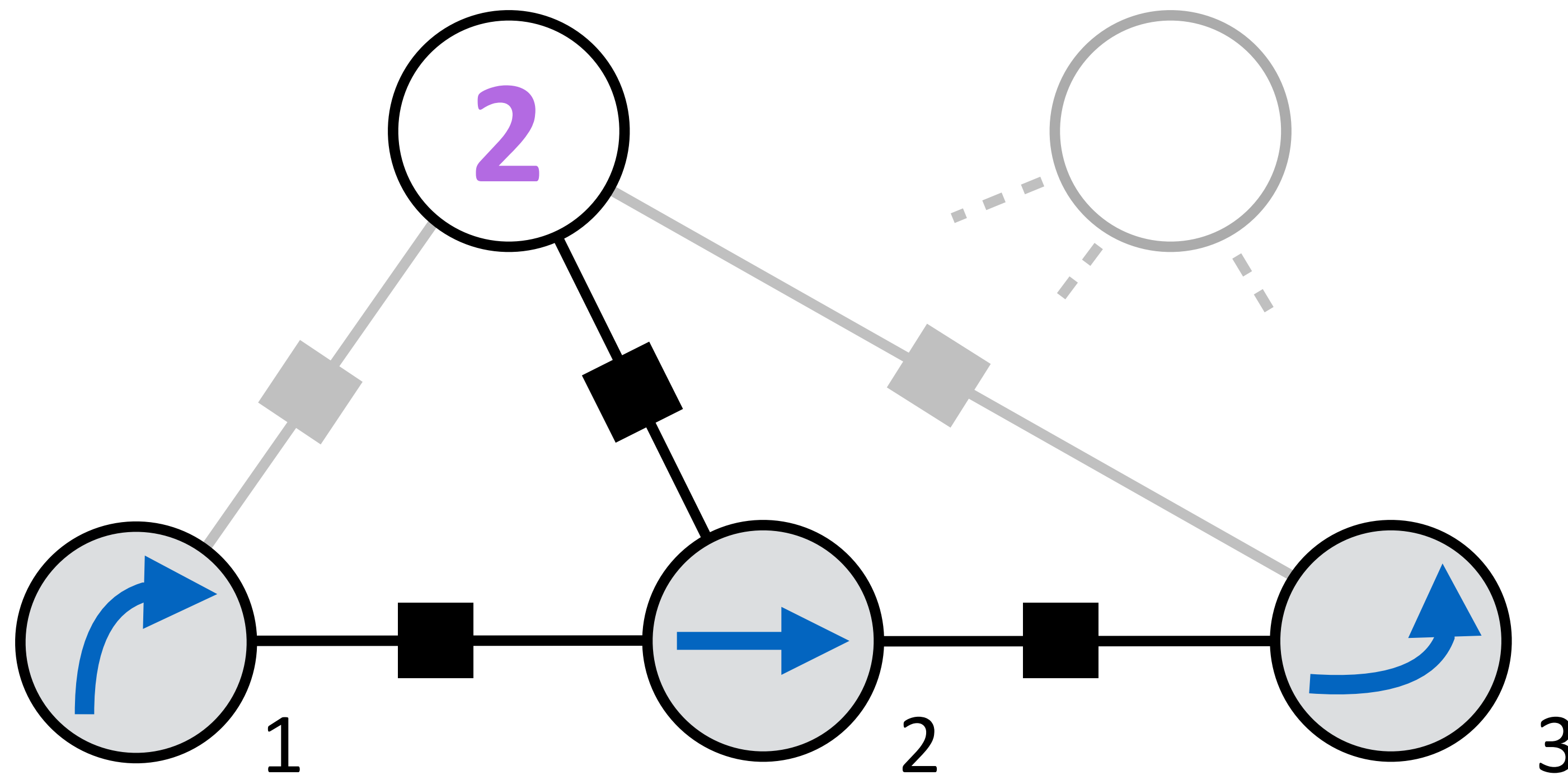
*go down the yellow hall*    *turn left*





# Sequence-to-sequence

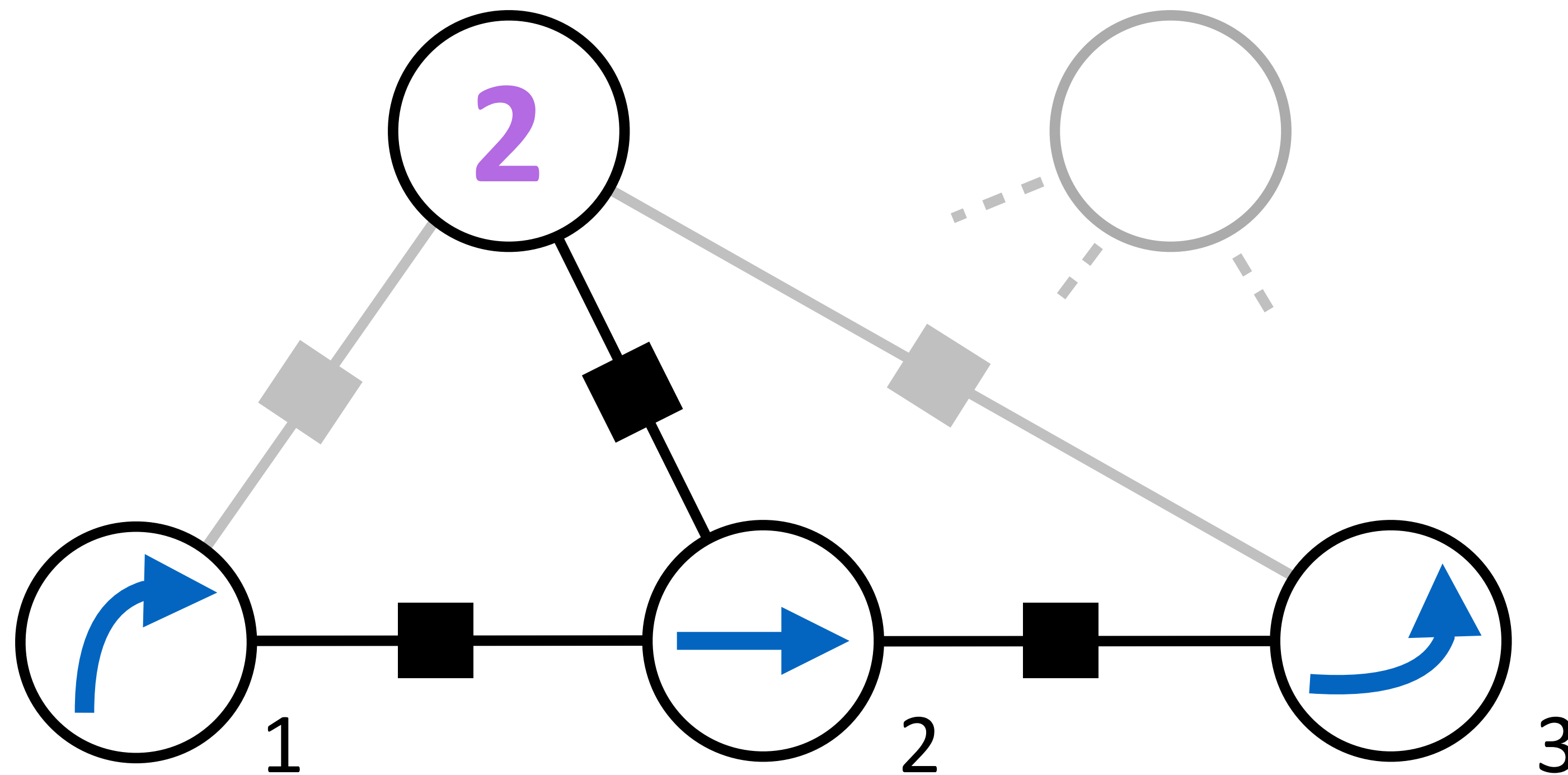
*go down the yellow hall*    *turn left*

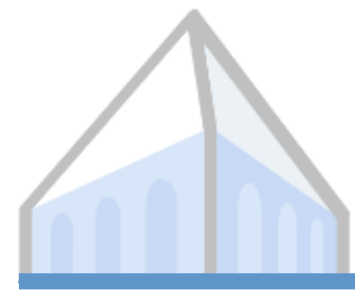




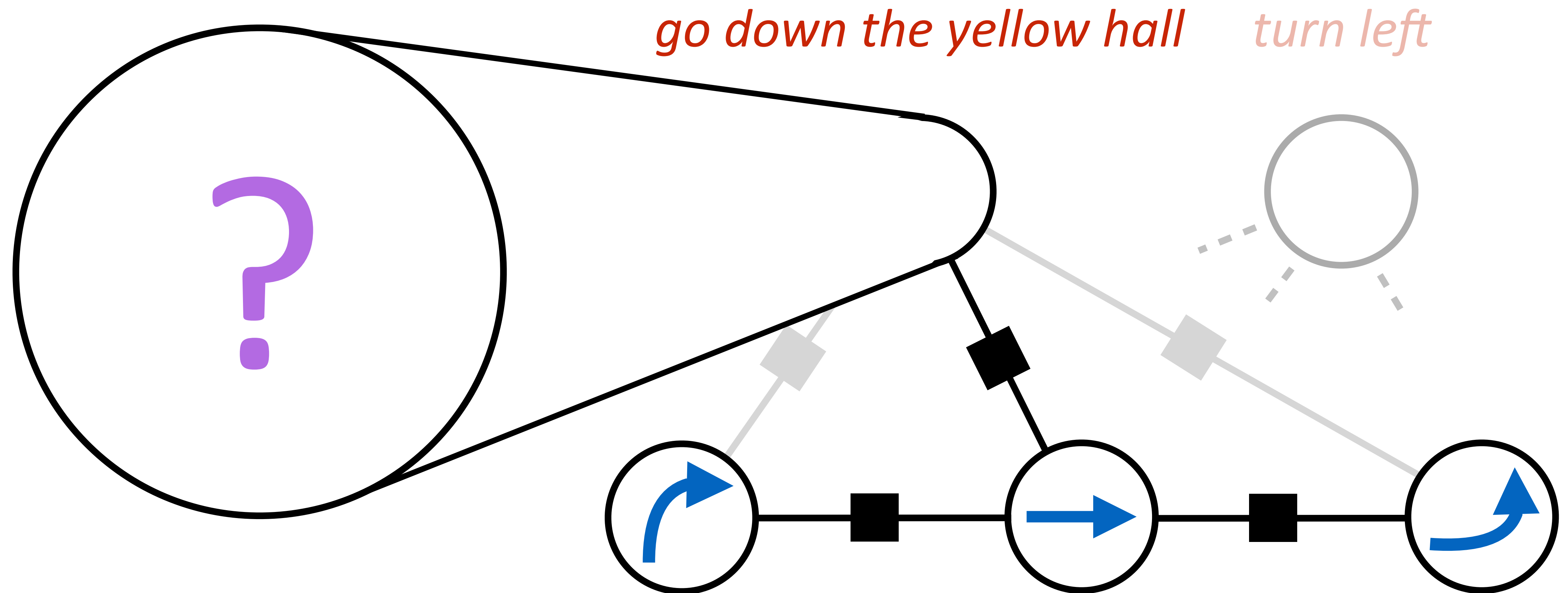
# Sequence-to-sequence

*go down the yellow hall*    *turn left*

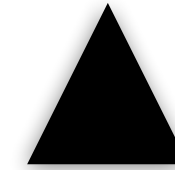




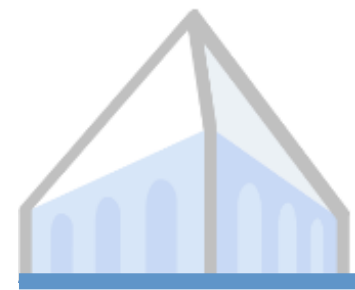
# Sequence-to-sequence



# STRUCTURE-TO-STRUCTURE MODEL



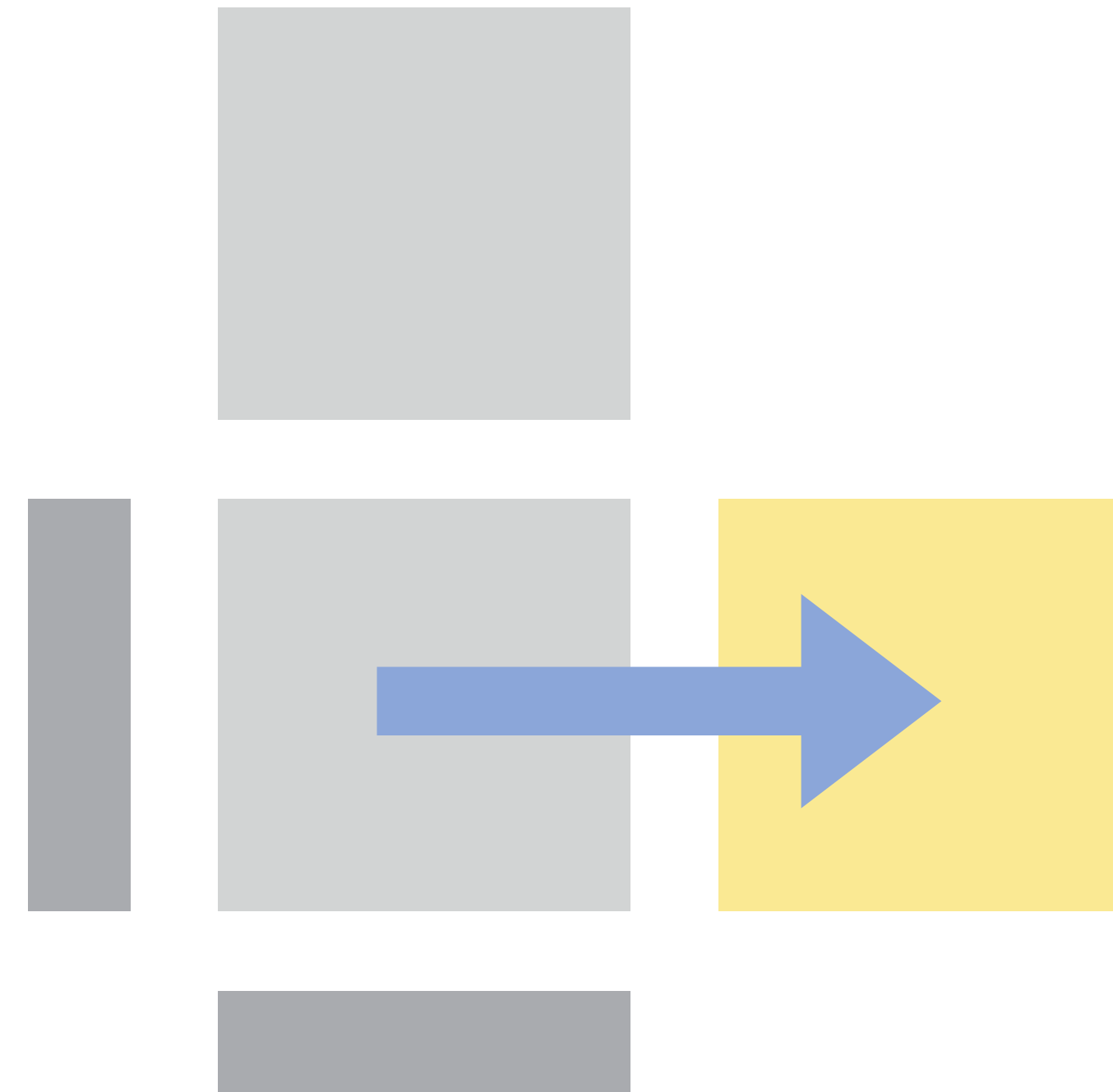




# Structure-to-structure

---

*go down the yellow hall*

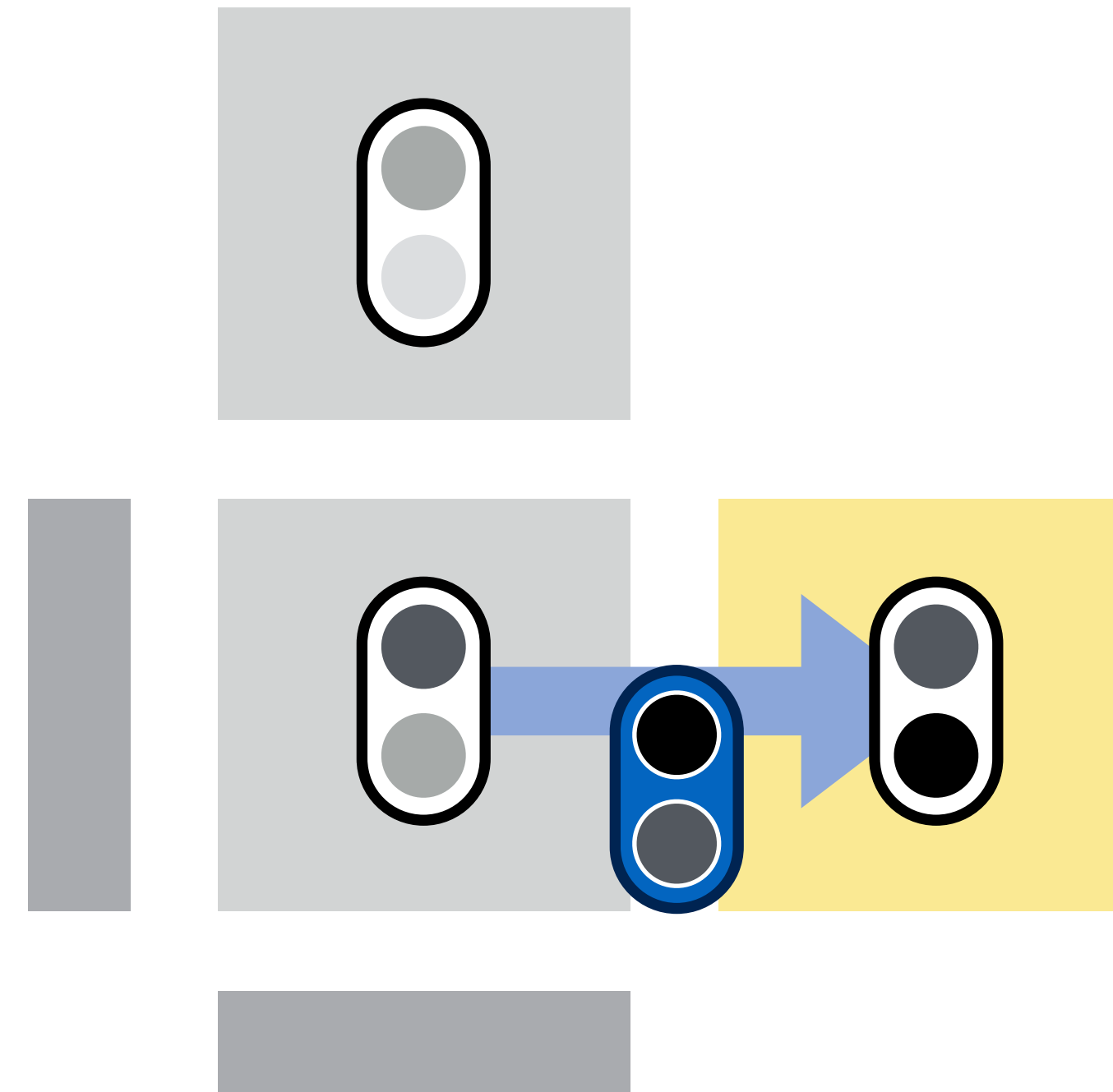




# Structure-to-structure

---

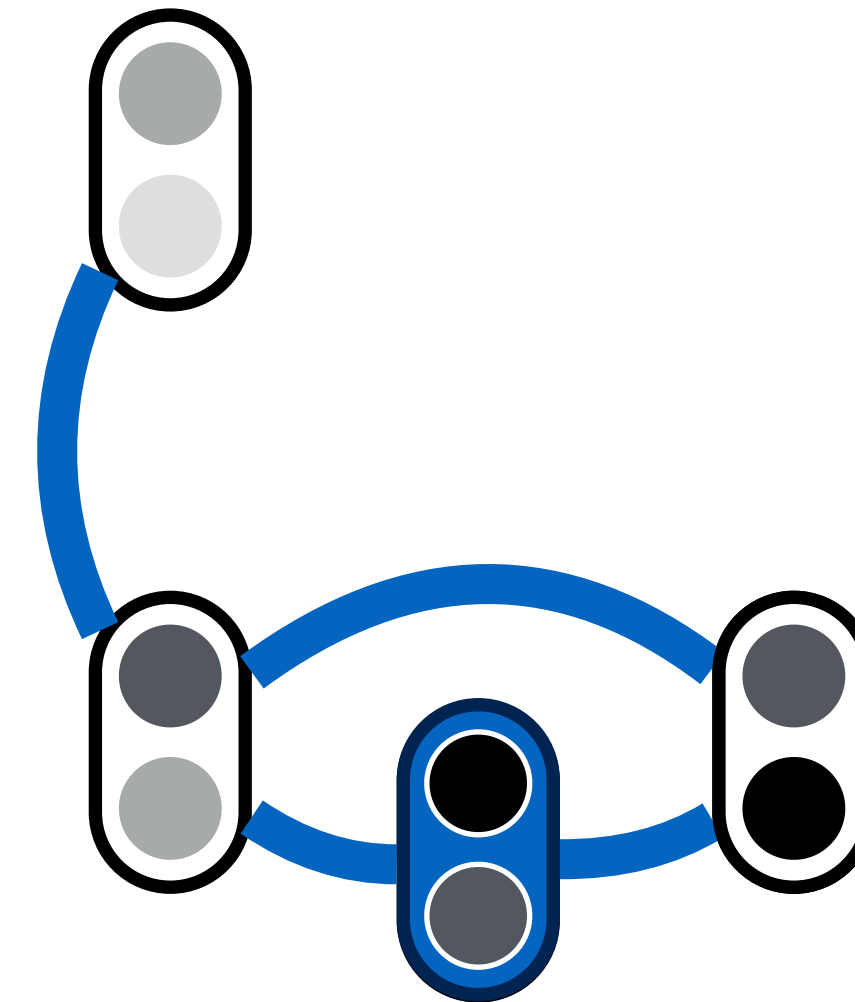
*go* → *down* → *hall* → *yellow*  
*the*

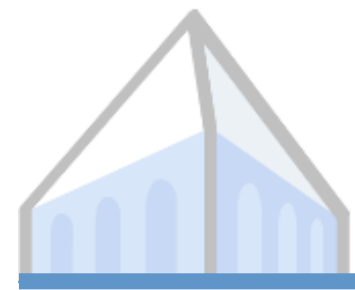




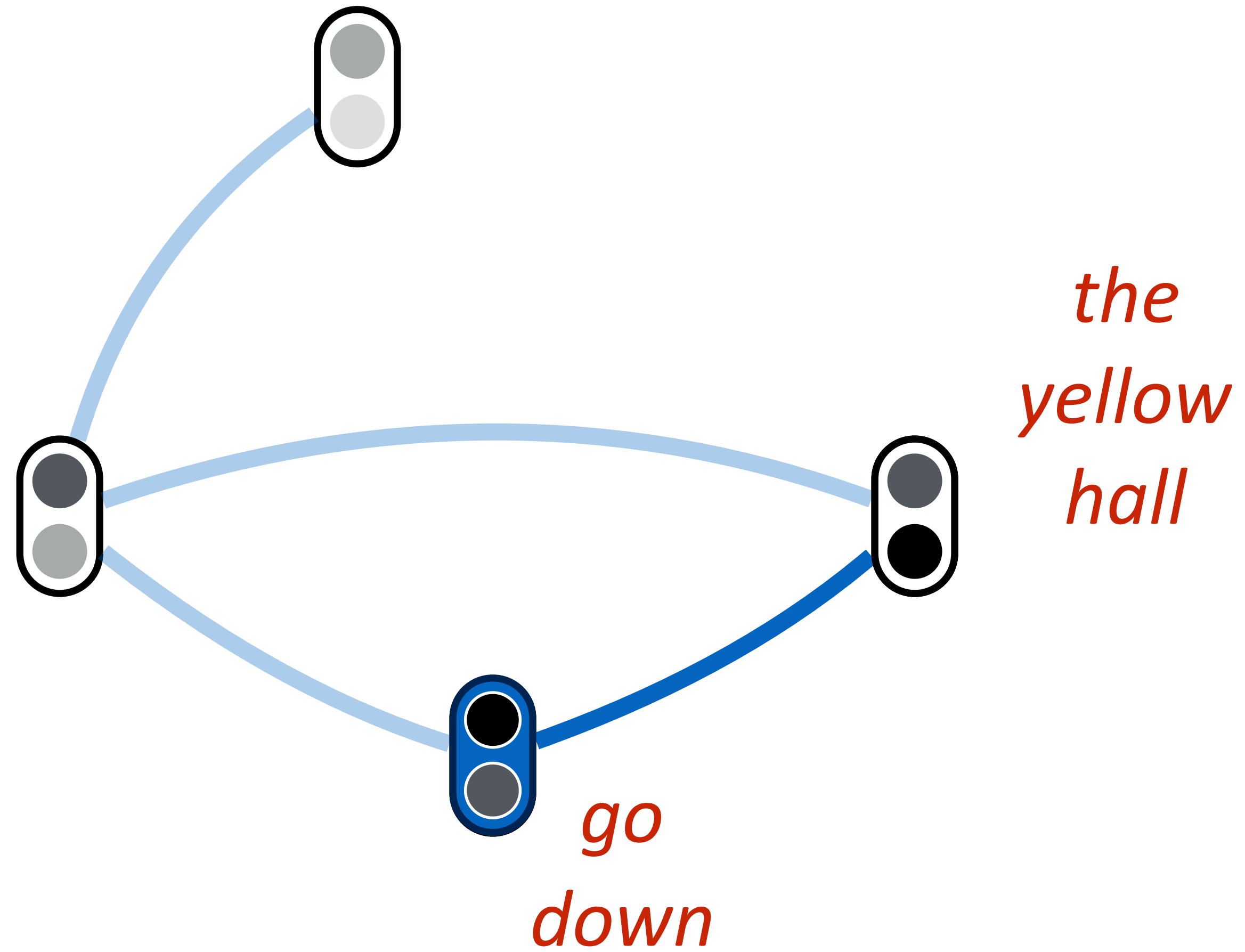
# Structure-to-structure

*go* → *down* → *hall* → *yellow*  
*the*



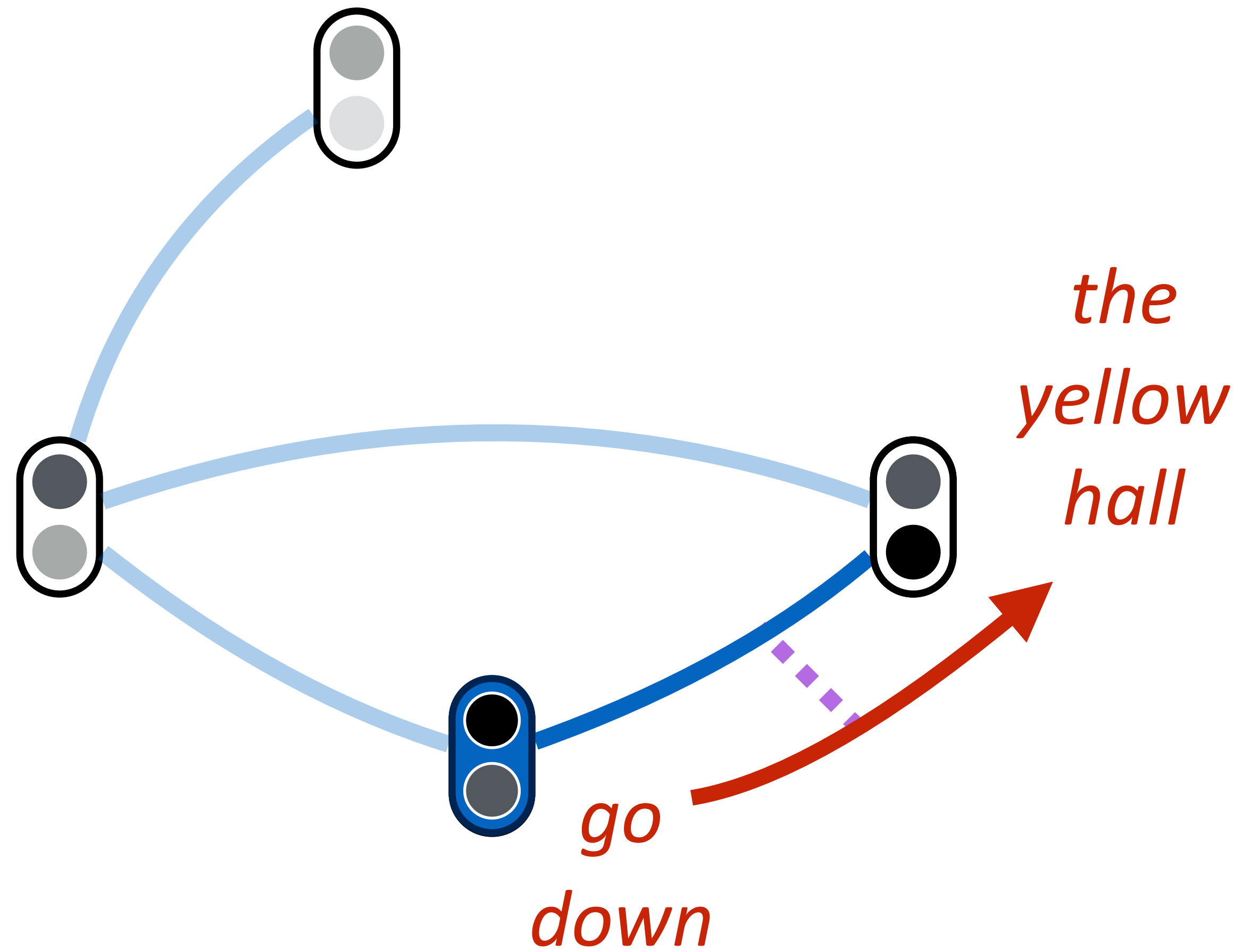


# Structure-to-structure



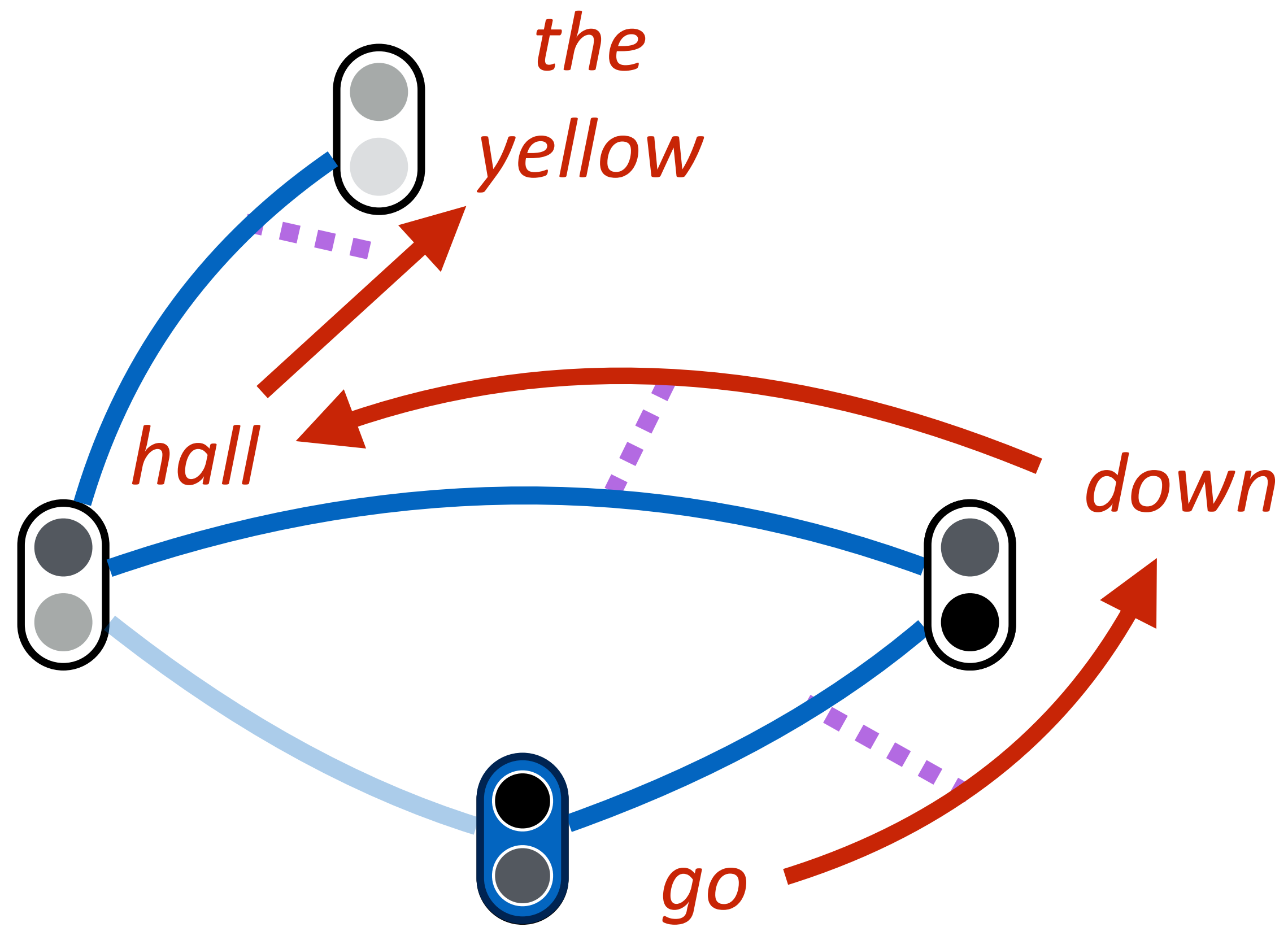


# Structure-to-structure





# Structure-to-structure

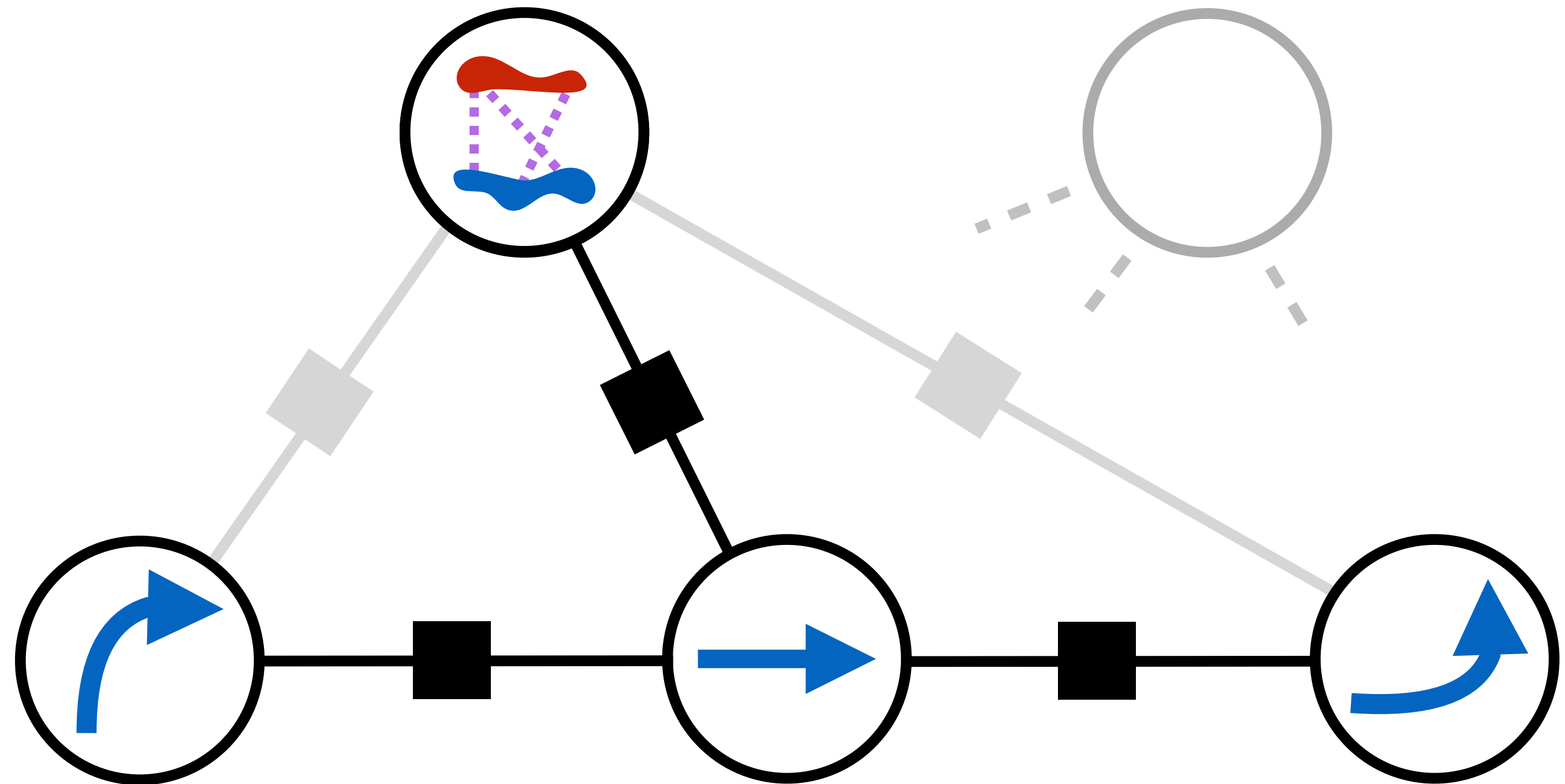


# JOINT MODEL



# Joint model

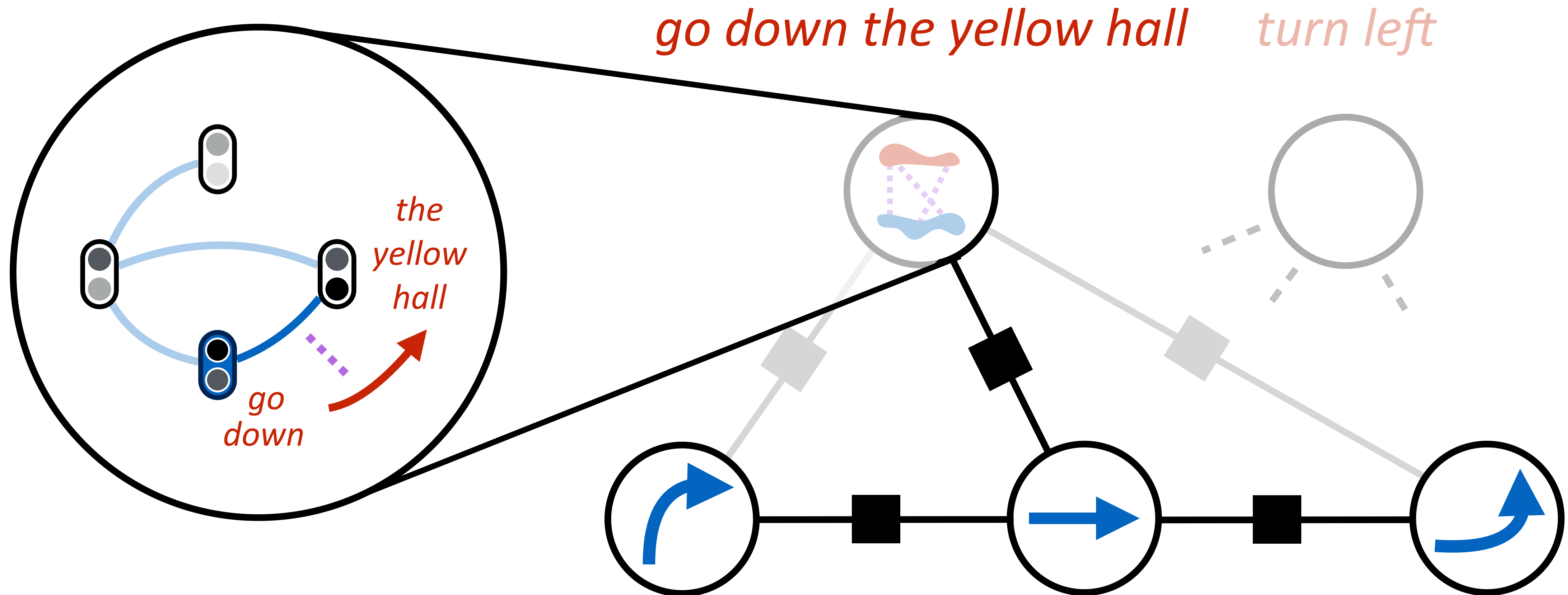
*go down the yellow hall    turn left*







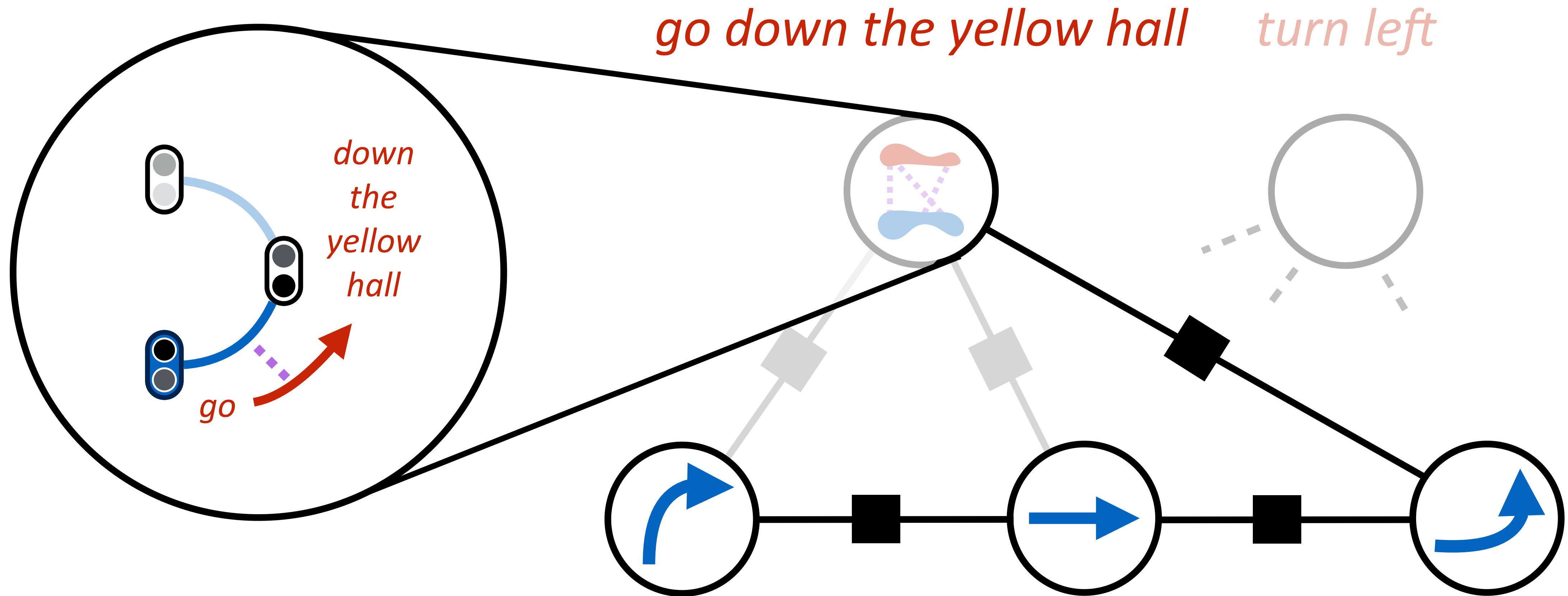
# Joint model

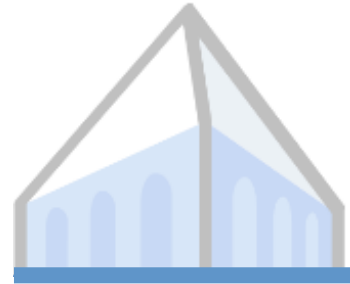






# Joint model





# Joint model

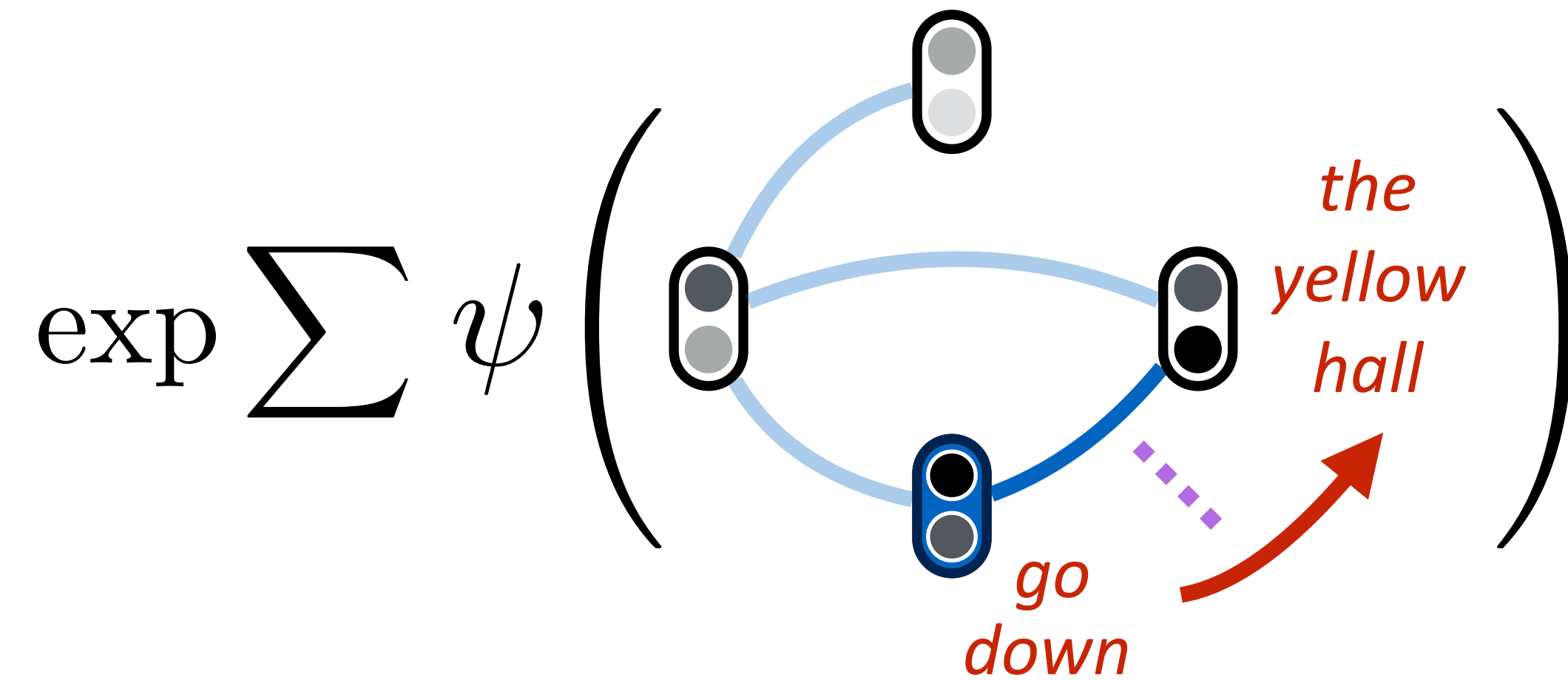
$$\psi \left( \begin{array}{c} \text{Diagram of a robot in a hallway with actions 'go down' and 'the yellow hall'} \end{array} \right) = \theta^\top \left( \phi(\textit{go}, \text{robot}) + \phi(\textit{down}, \text{robot}) + \dots \right)$$

The diagram shows a robot (blue) in a hallway with three elevators (grey). A blue path leads from the robot to the right elevator, with a blue arrow labeled "go down" pointing to the robot. A red arrow labeled "the yellow hall" points to the right elevator. A purple dashed line and a red arrow also point towards the right elevator.



# Joint model

$$p(\text{actions}, \text{alignment} \mid \text{text}; \theta) \propto$$





LEARNING / INFERENCE



# Learning: coordinate ascent

---

$$\max_{\theta, \text{alignments}} p(\text{actions}, \text{alignments} \mid \text{text}; \theta)$$

$$\max_{\text{alignments}}$$

using custom alignment dynamic program

$$\max_{\theta}$$

using L-BFGS





# Decoding: coordinate ascent

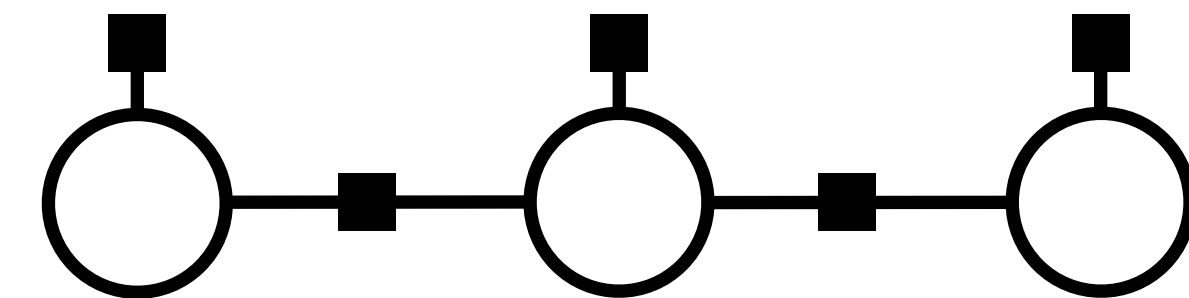
$$\boxed{\begin{array}{l} \max \\ \text{actions,} \\ \text{alignments} \end{array}} p(\text{actions, alignments} \mid \text{text}; \theta)$$

$\max$   
alignments

using custom alignment dynamic program

$\max$   
actions

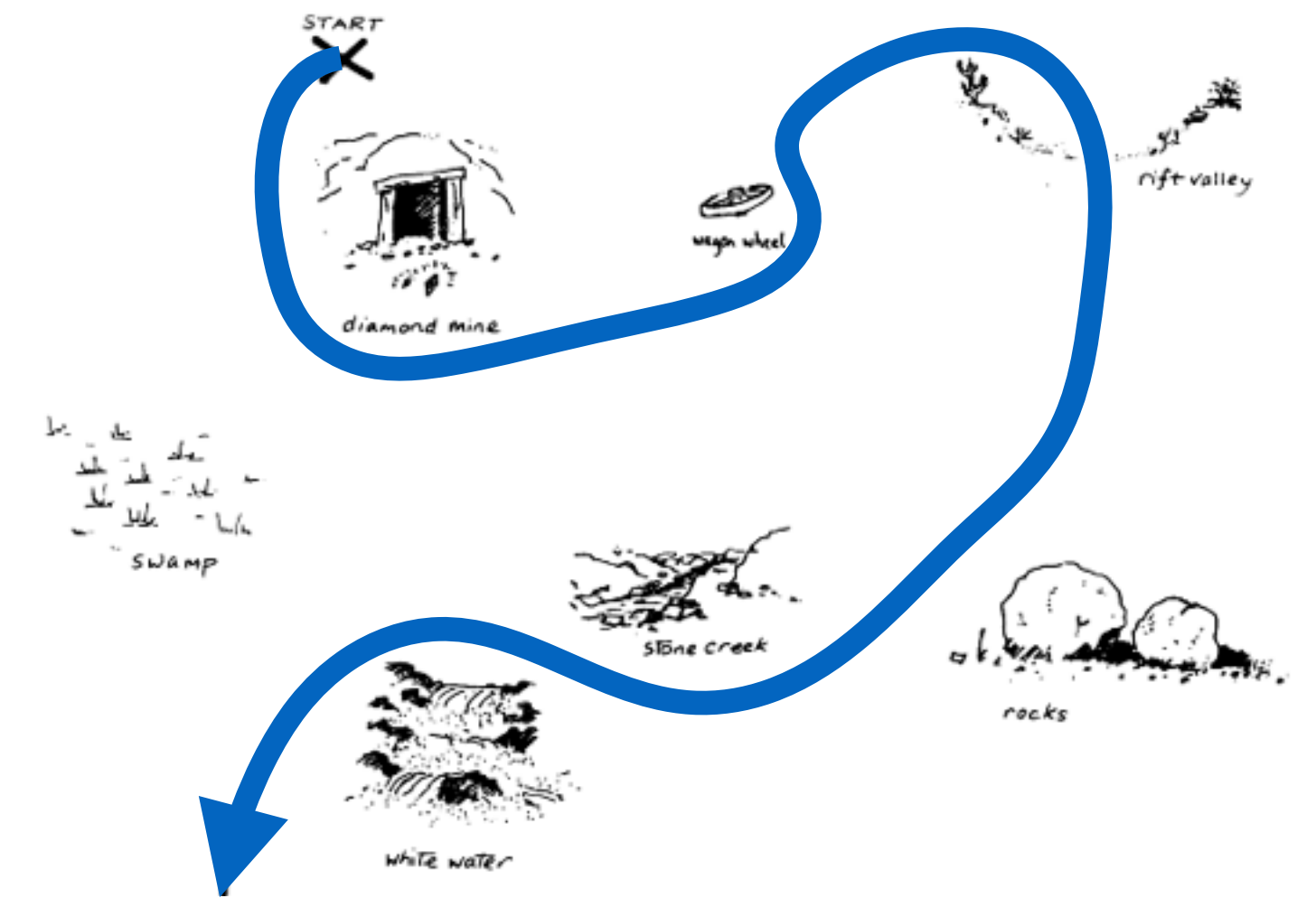
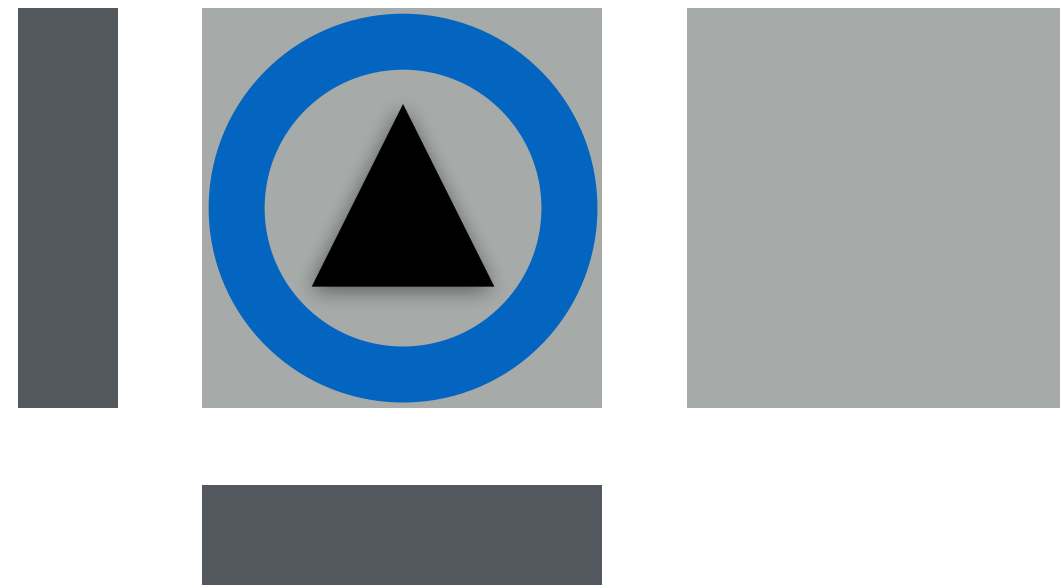
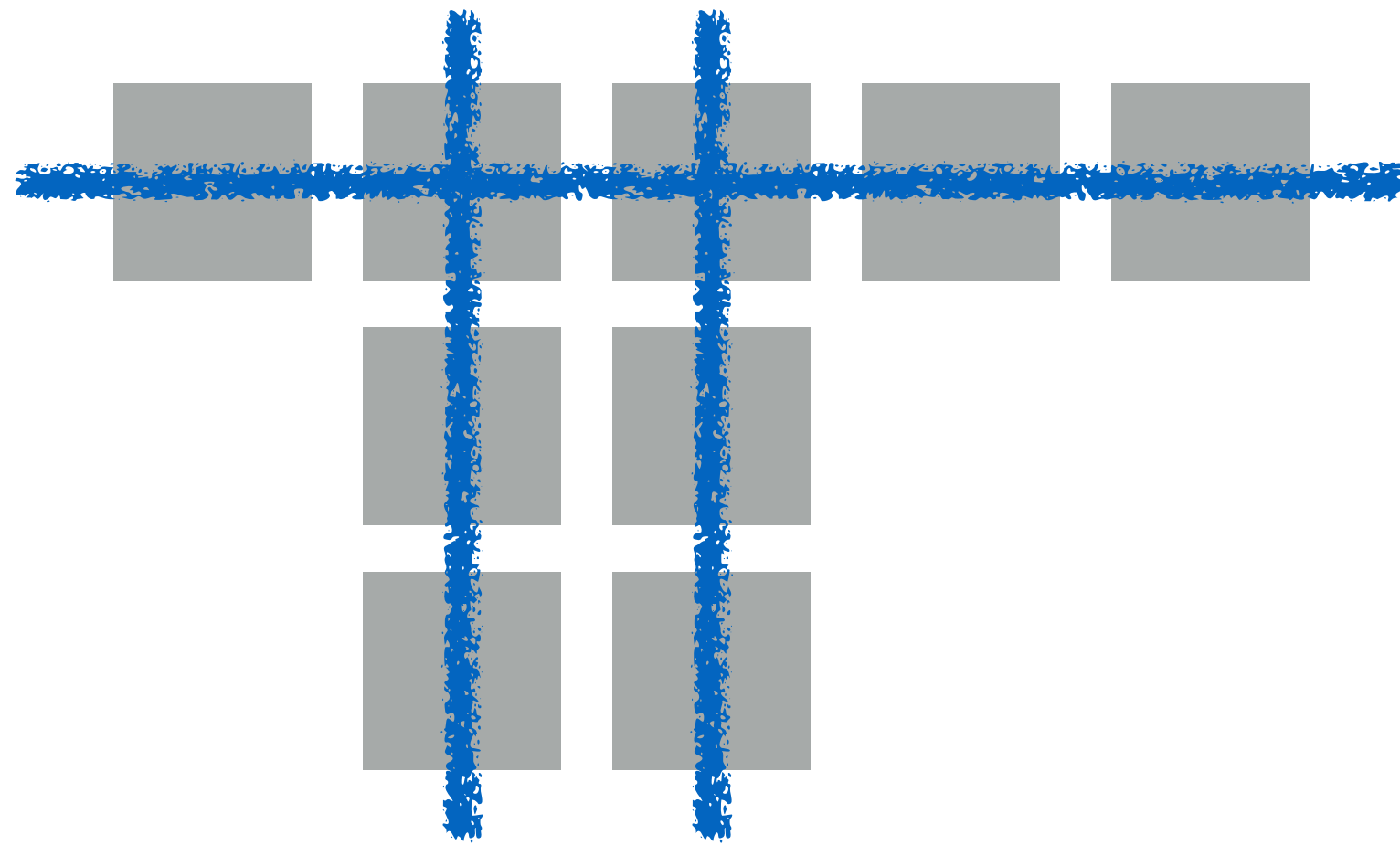
using a planner



# EXPERIMENTS



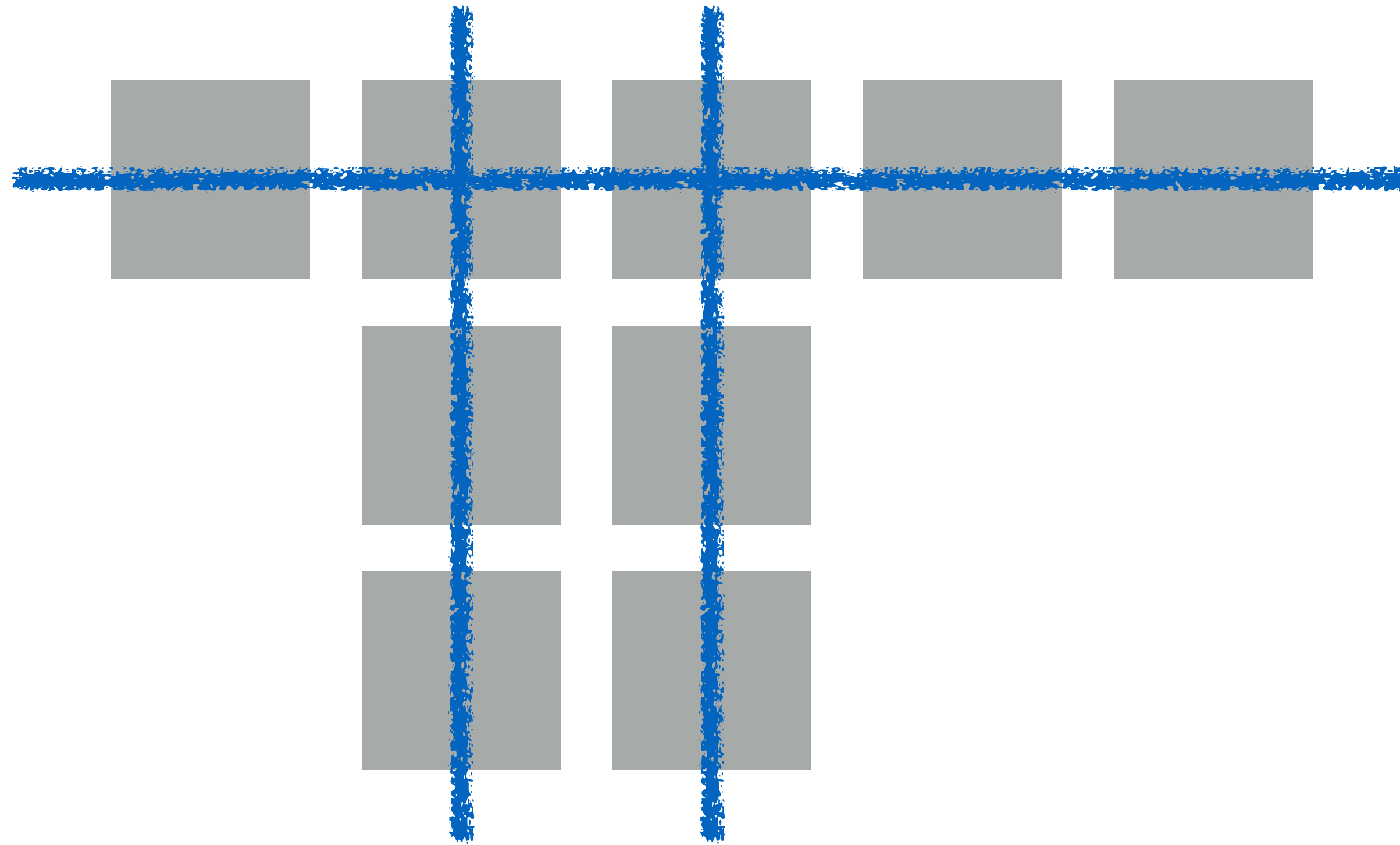
# Three tasks





# Puzzle solving

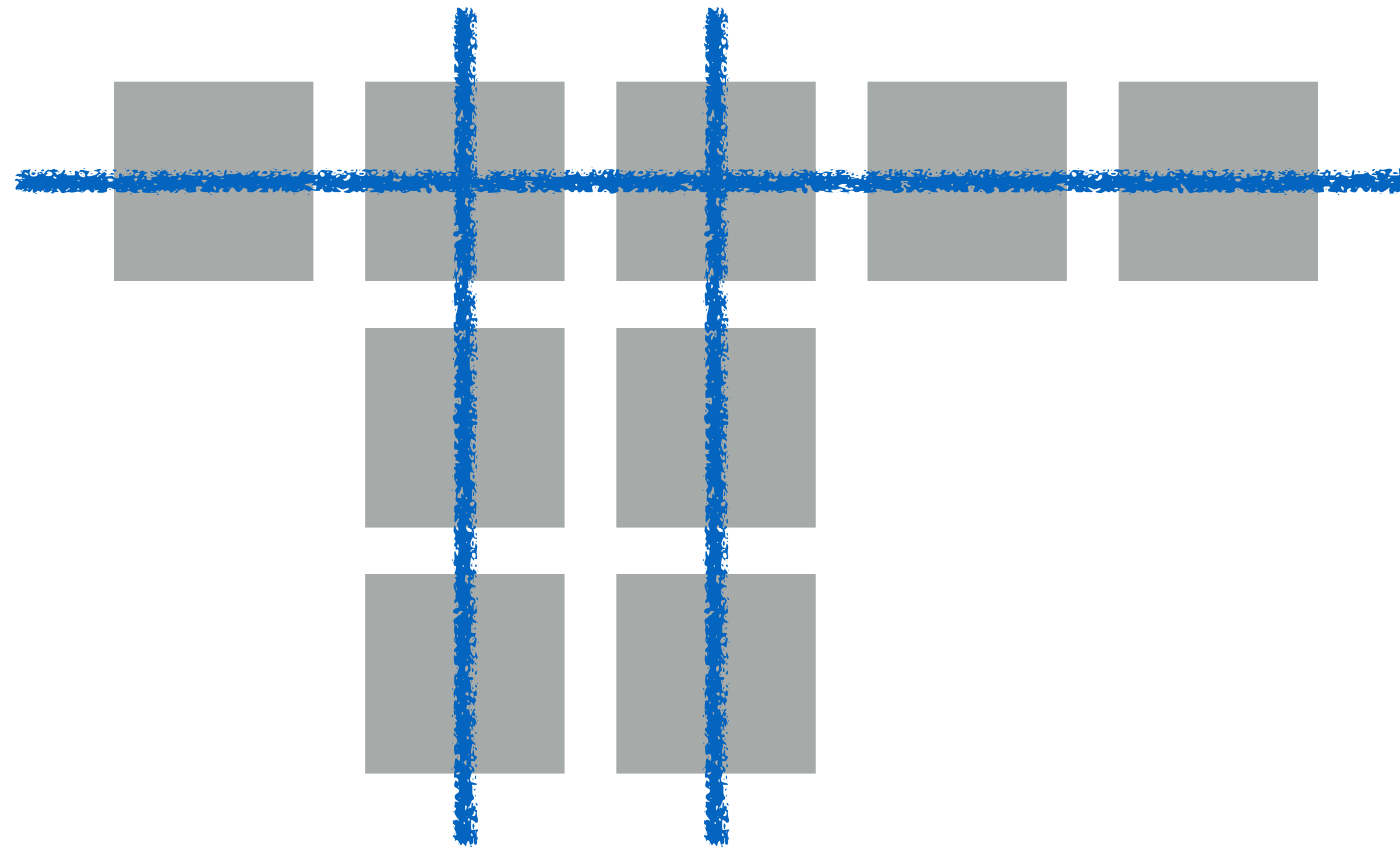
*clear the two long columns, and then the row*





# Puzzle solving

*clear the two long columns, and then the row*





# Puzzle solving

Plan execution (%)

90

60

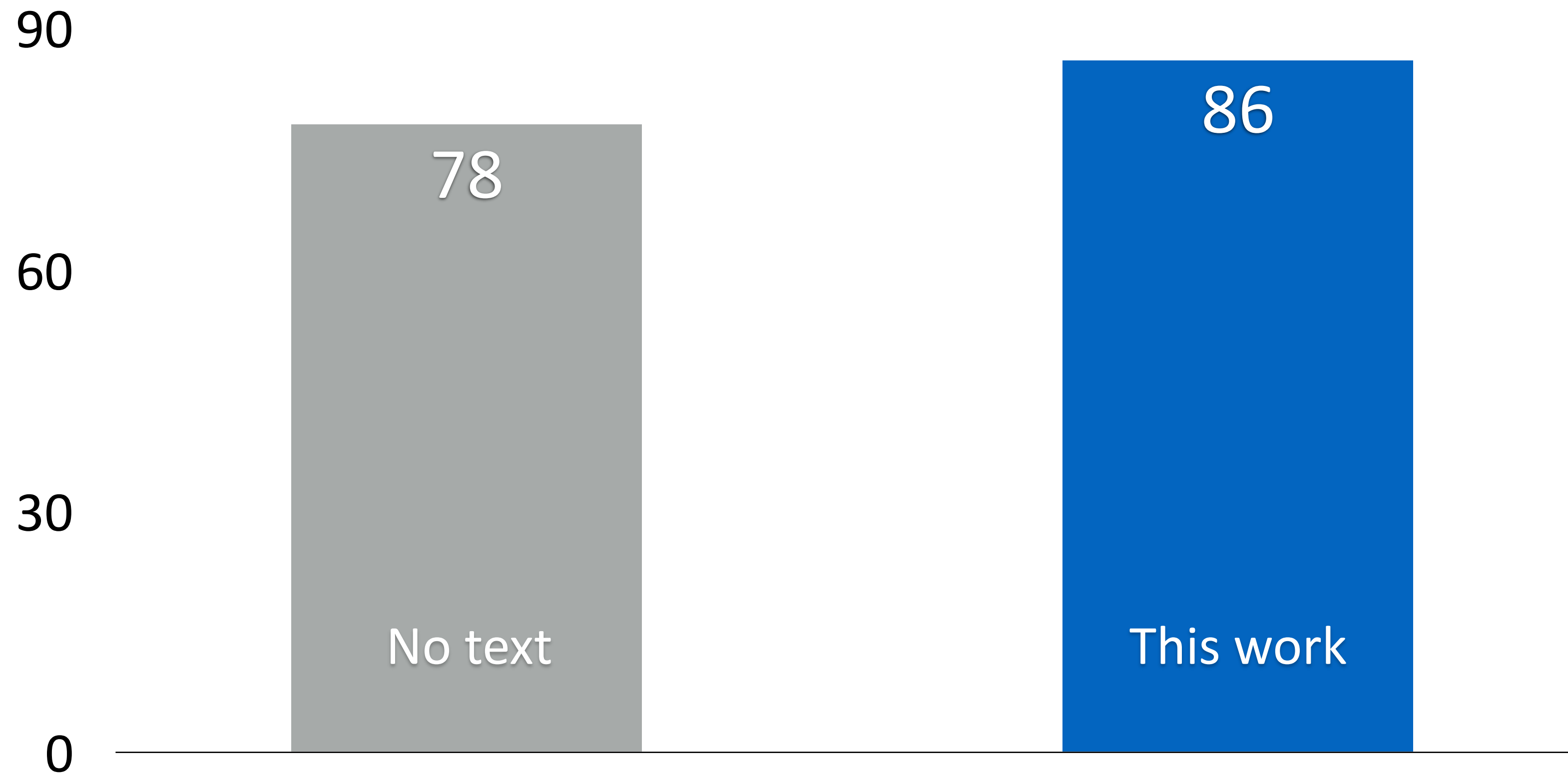
30

0



# Puzzle solving

Task completion (%)





# Maze navigation

*you should see a grey floor to your right and a chair in front of you*







# Maze navigation

Task completion (%)

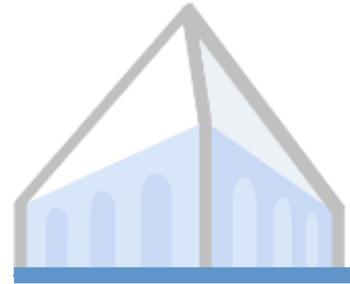
65

55

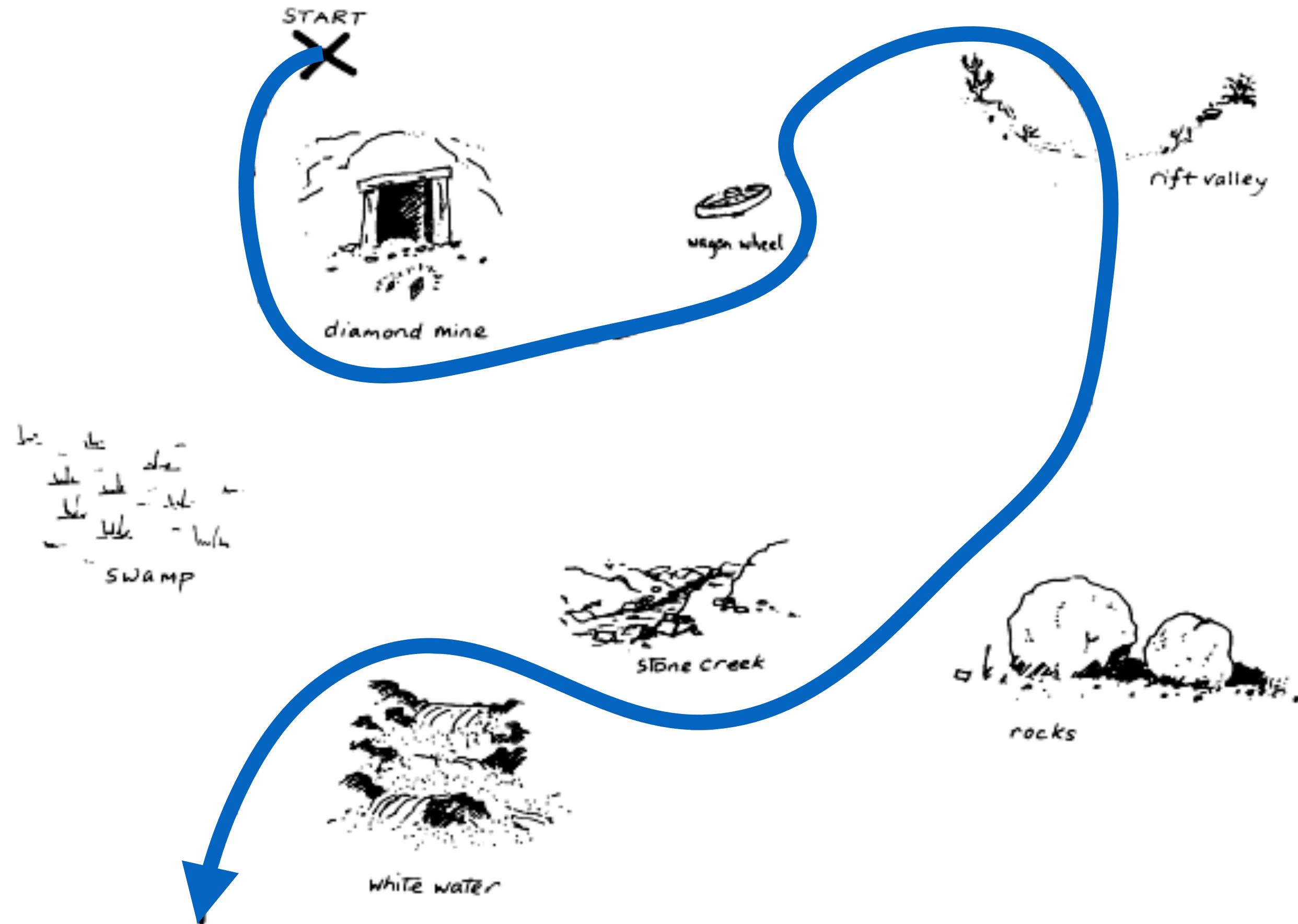
45

35





# Map reading



*right round the white water [...] but stay quite close 'cause you don't otherwise you're going to be in that stone creek*



# Map reading

Plan execution ( $F_0$ )

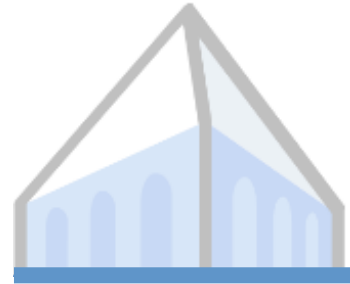
60

45

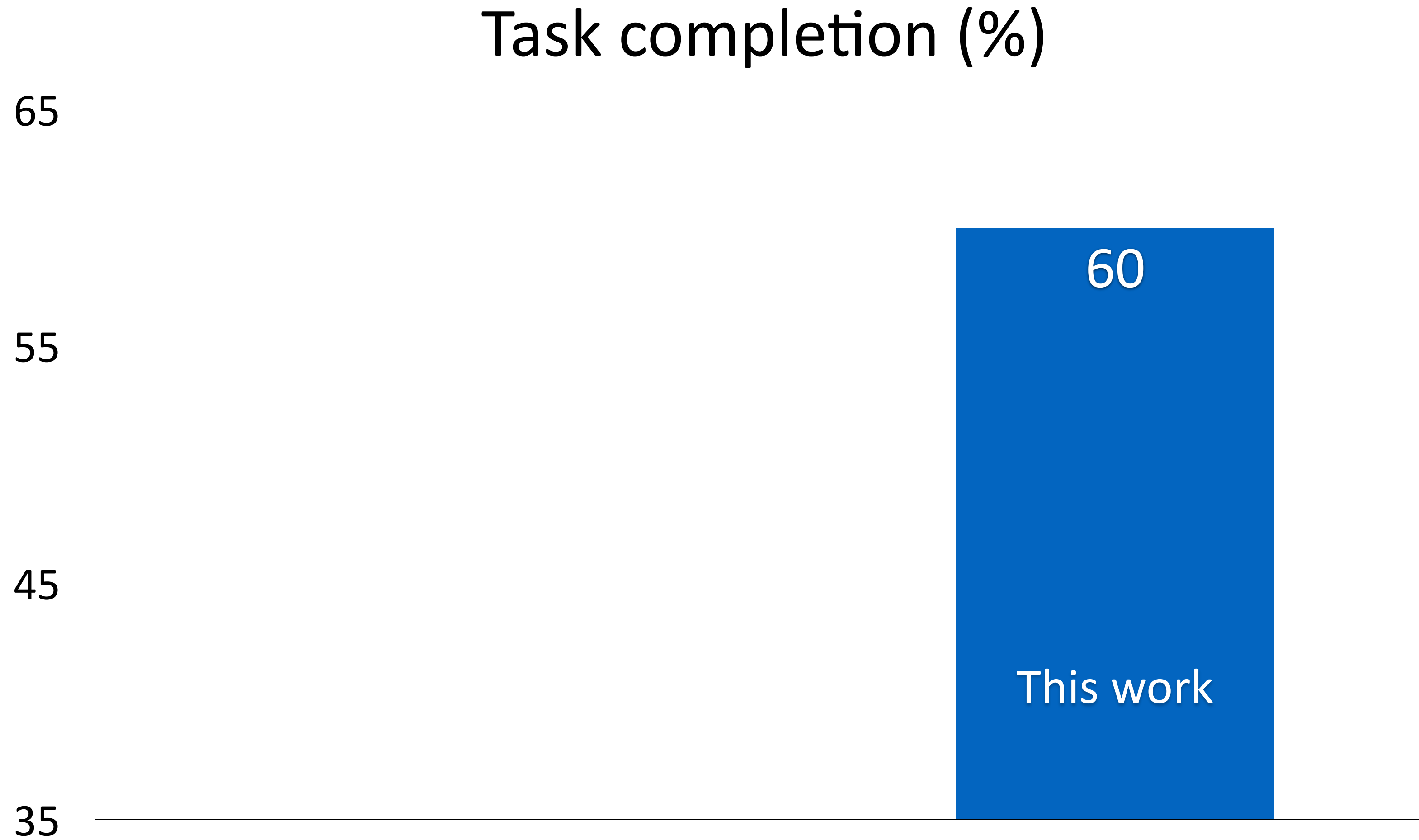
30

15

0



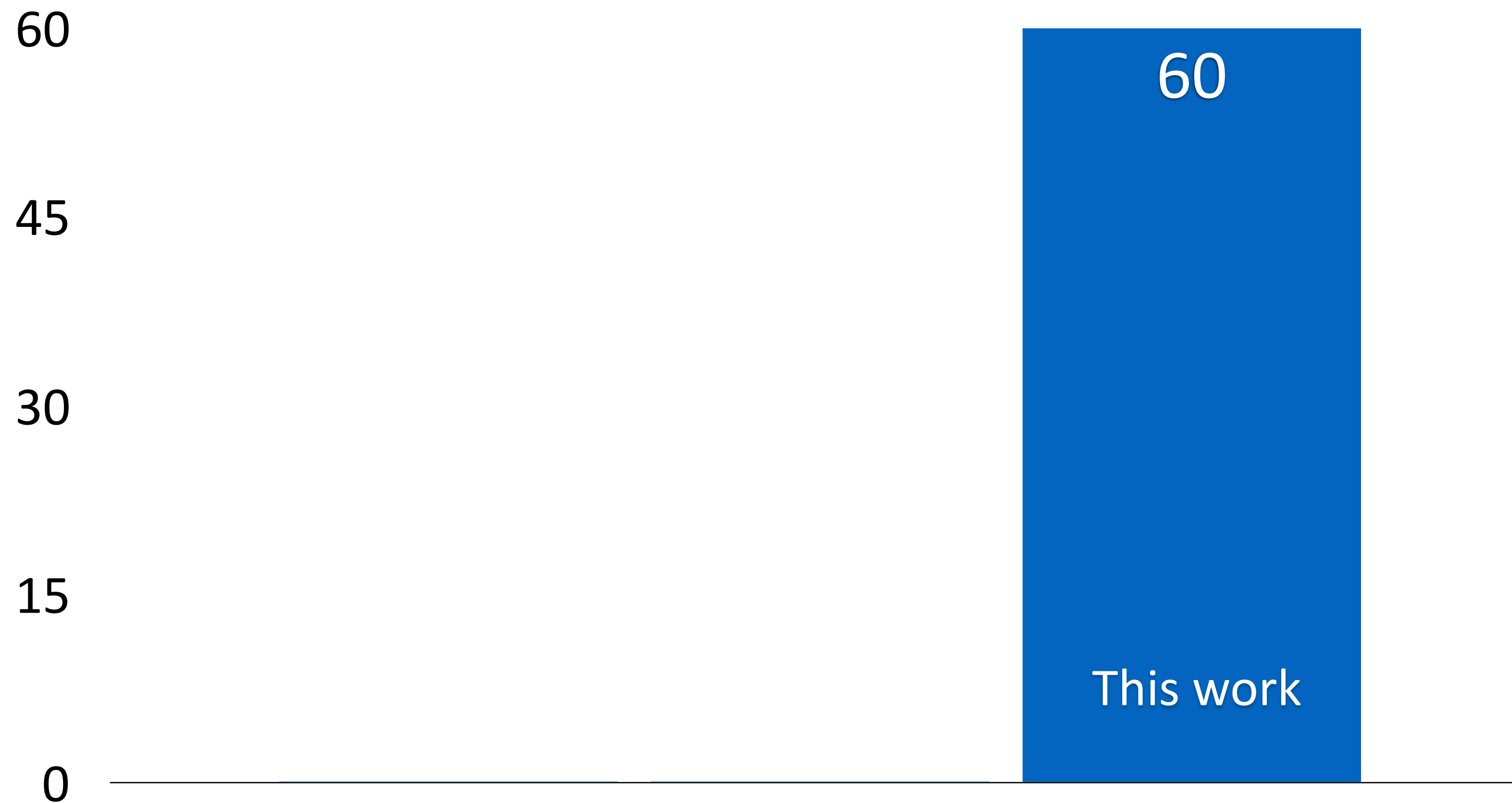
# Ablations: maze navigation





# Ablations: map reading

Plan execution ( $F_0$ )





# Conclusion

---

Structured alignment/decoding gives us best aspects of:

- ▶ Compositional semantics (like a parser)
- ▶ Sequence structure (like a planner)

THANK YOU