

Alignment-Based Compositional Semantics for Instruction Following



Jacob Andreas and Dan Klein
UC Berkeley

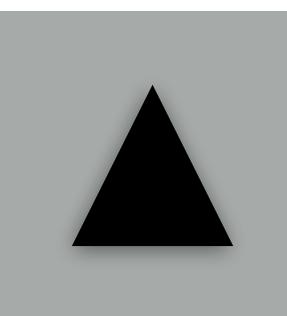


Following instructions

go down the green hall



go all the way



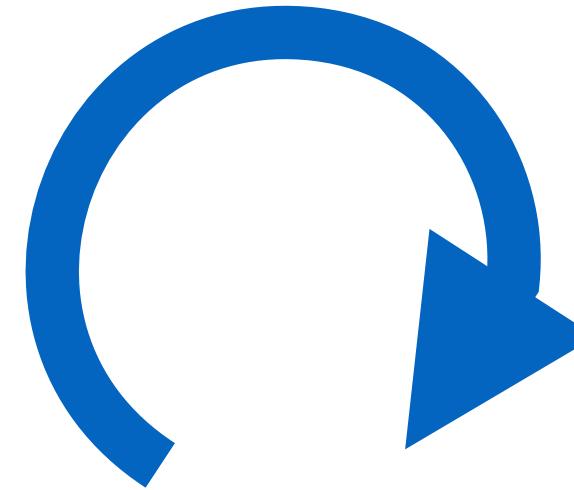
then face the chair



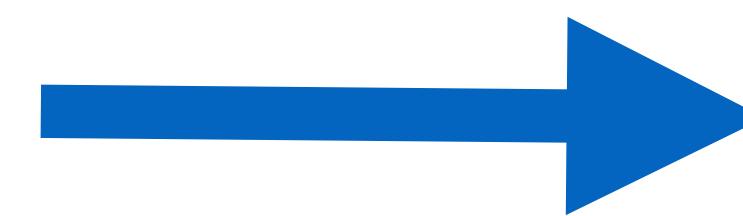


Following instructions

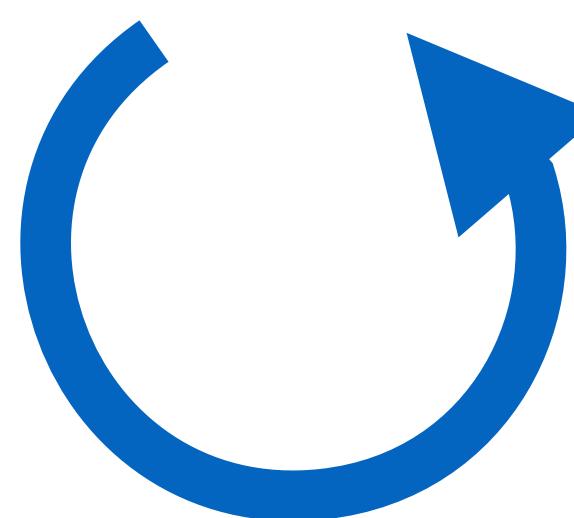
go down the green hall



go all the way



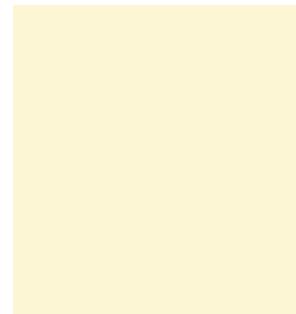
then face the chair



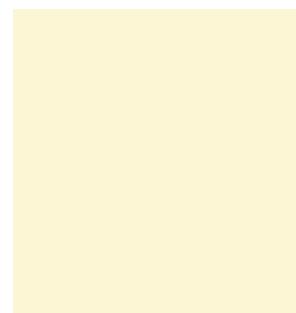


Following instructions: parsing

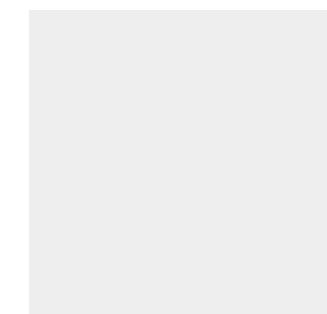
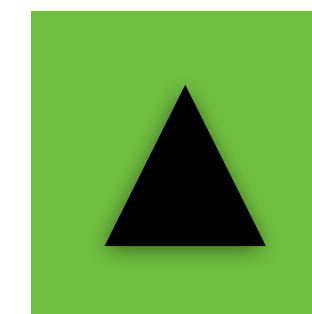
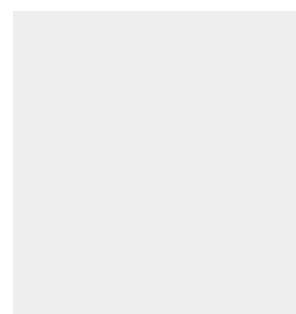
go down the green hall



go all the way



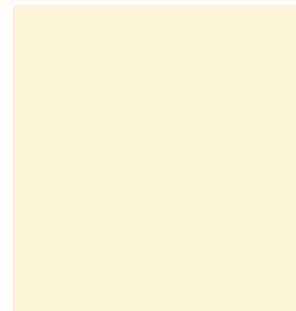
then face the chair



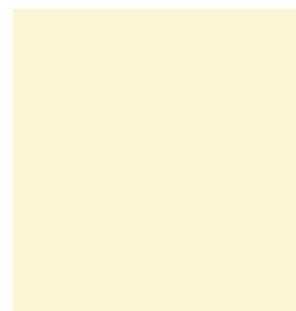


Following instructions: parsing

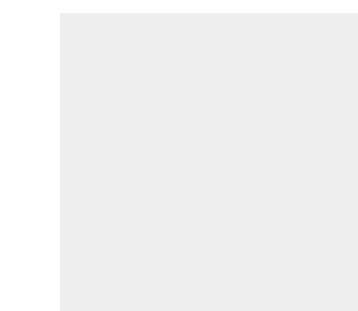
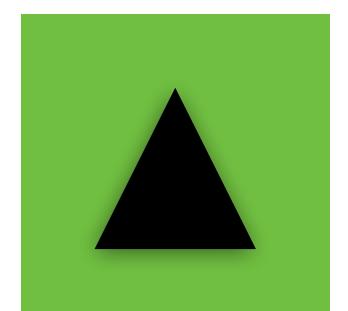
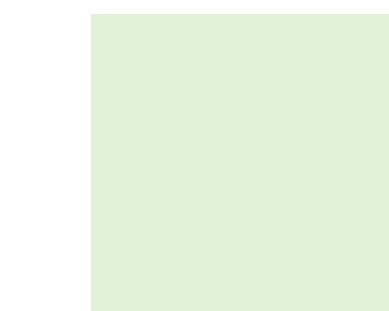
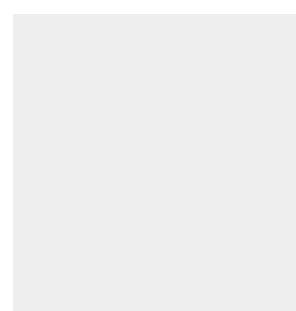
go down the green hall



go all the way



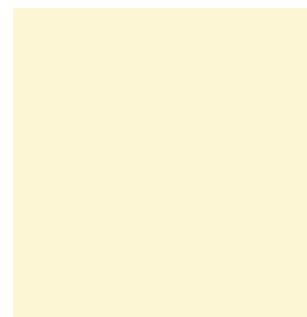
then face the chair





Following instructions: planning

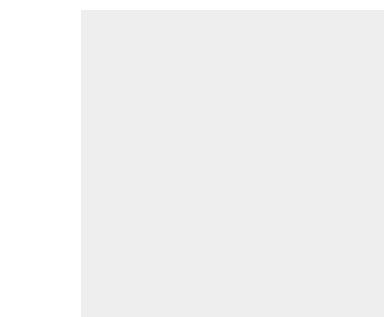
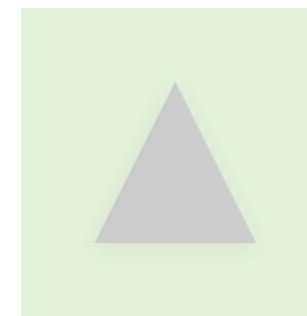
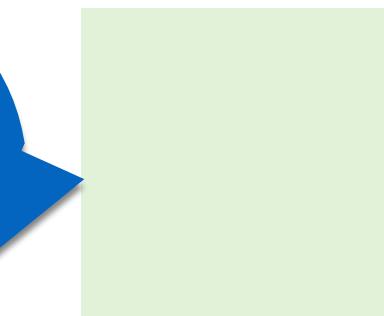
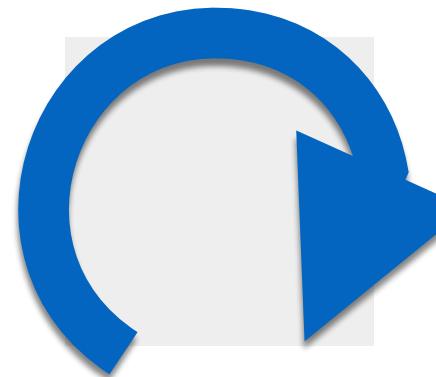
go down the green hall



go all the way



then face the chair



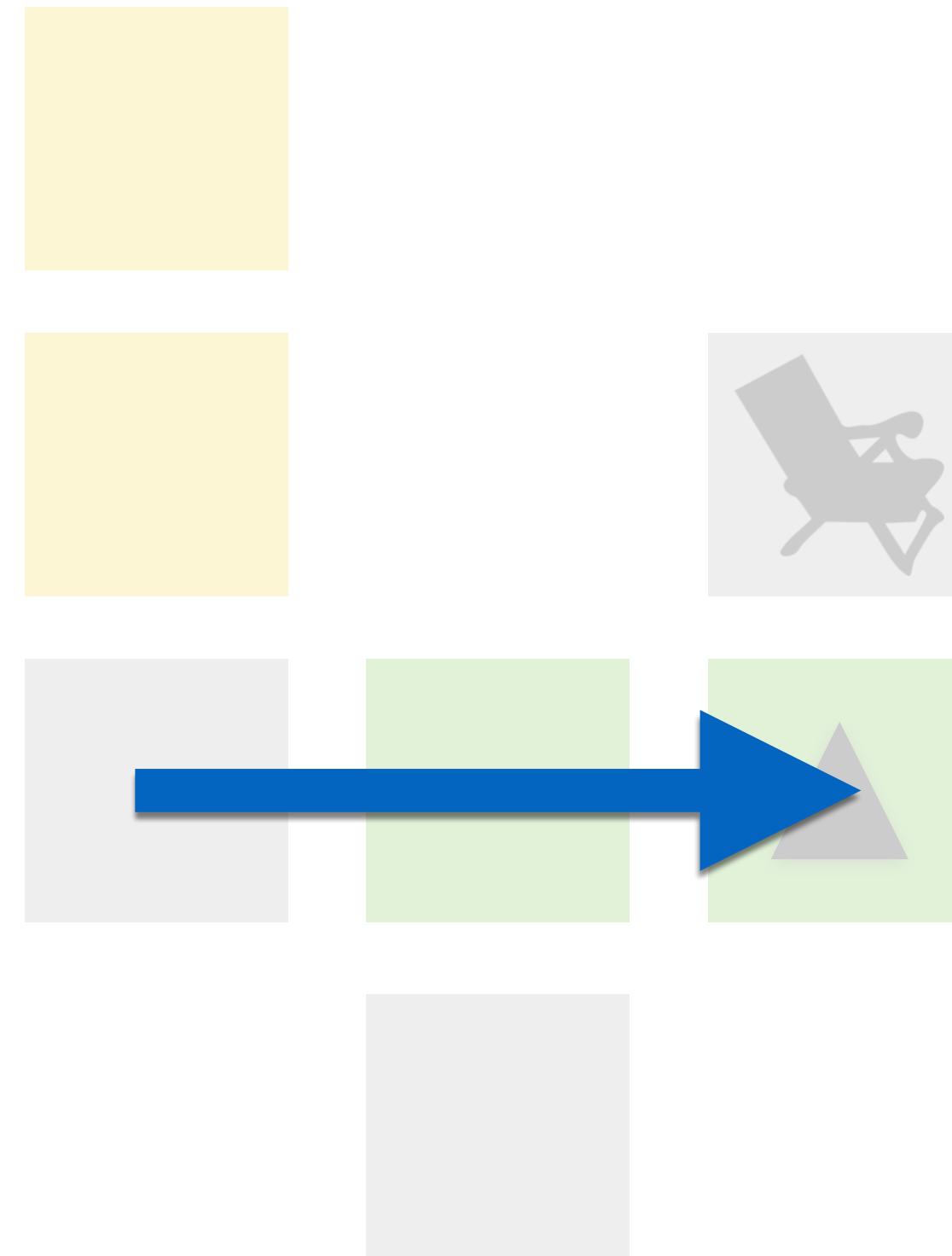


Following instructions: planning

go down the green hall

go all the way

then face the chair

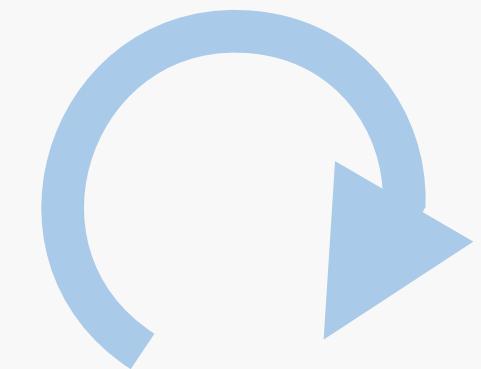




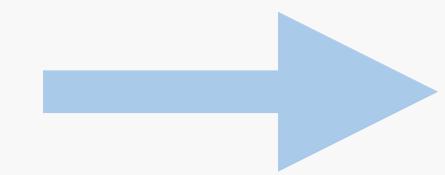
Following instructions

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)

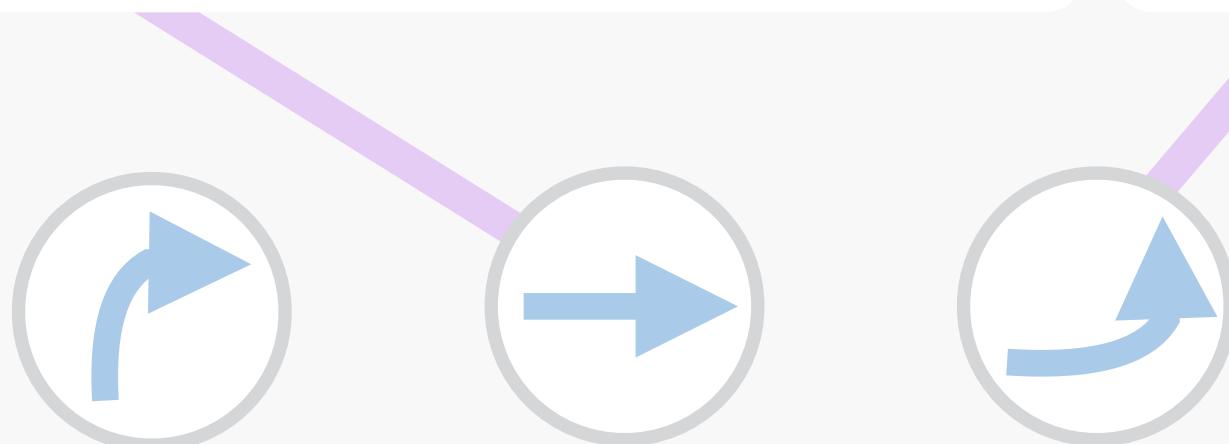
Kim and Mooney (2013)



Parsing and planning

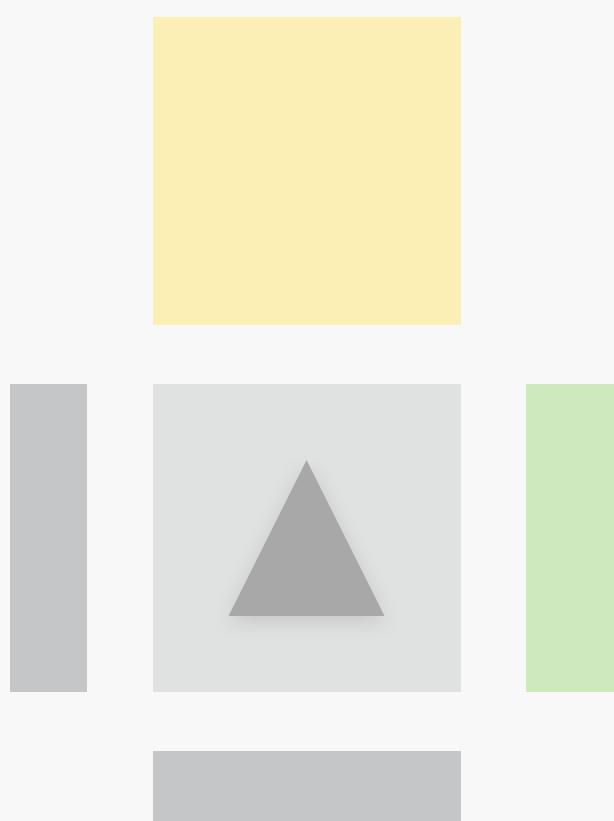
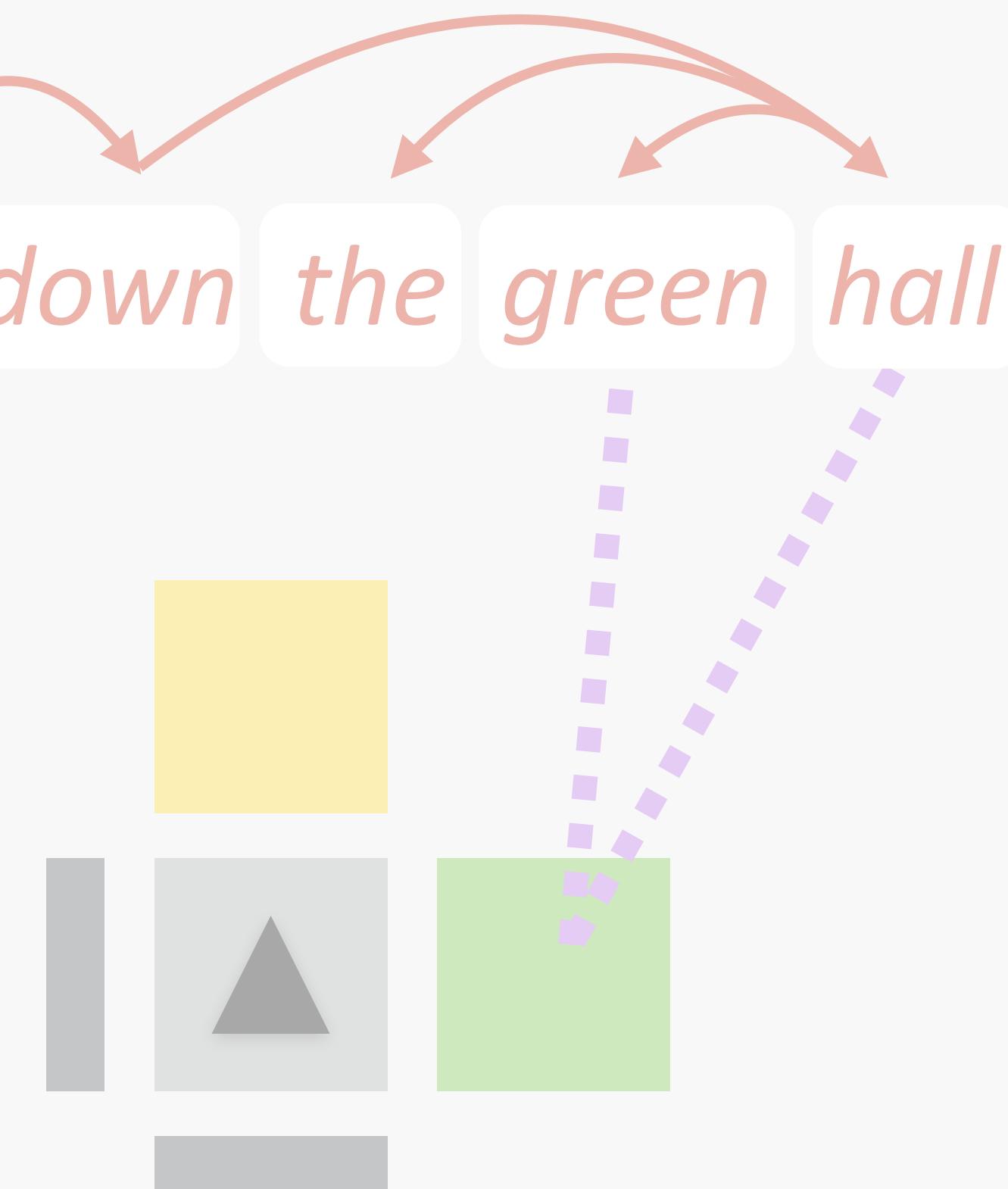
Sequence-to-sequence

go down the green hall turn left

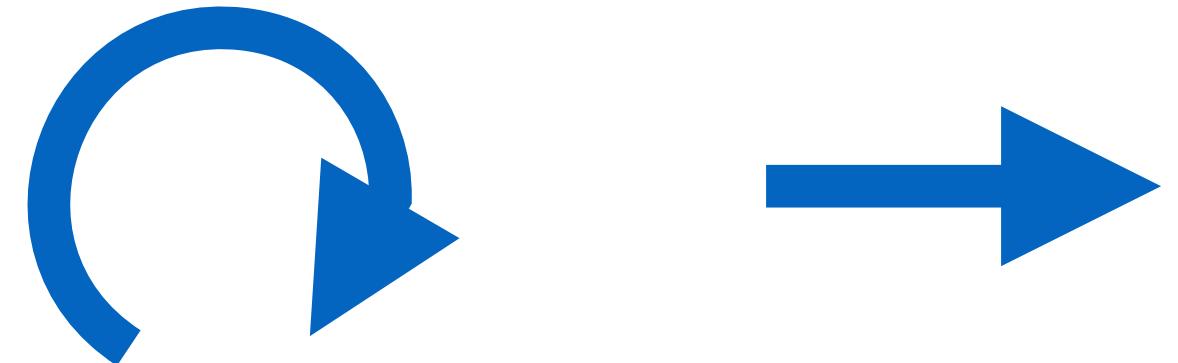


Structure-to-structure

go down the green hall



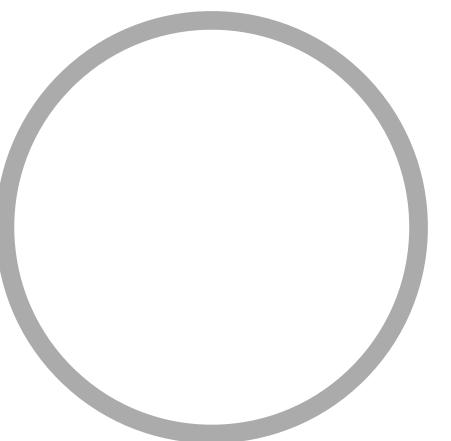
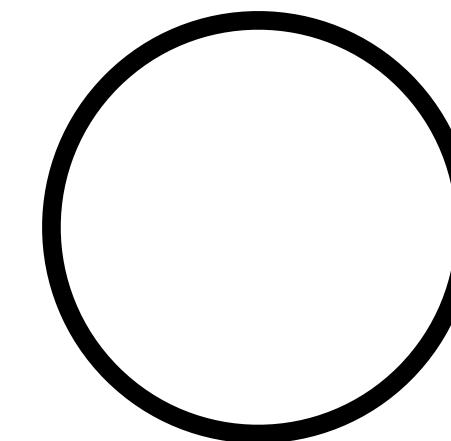
SEQUENCE-TO-SEQUENCE MODEL





Sequence-to-sequence

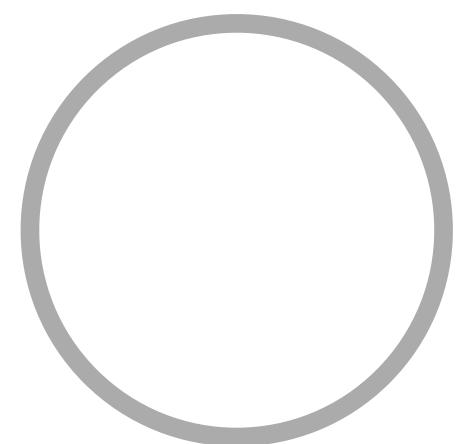
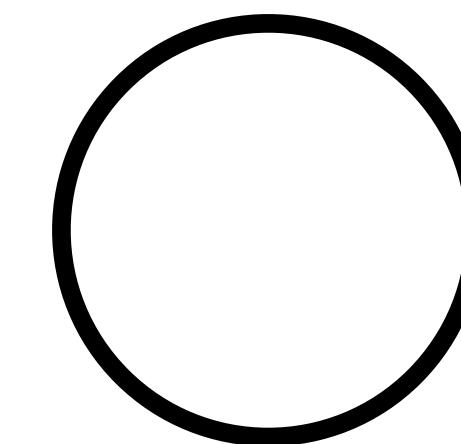
go down the yellow hall turn left





Sequence-to-sequence

go down the yellow hall turn left

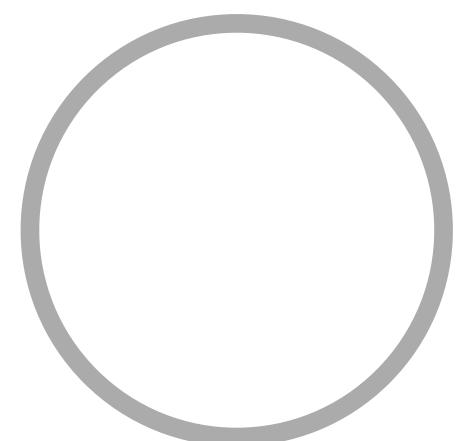
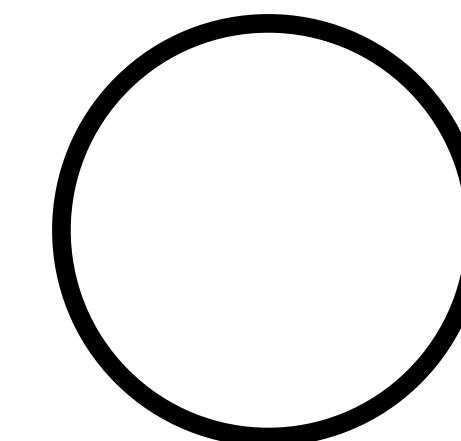


Alignments



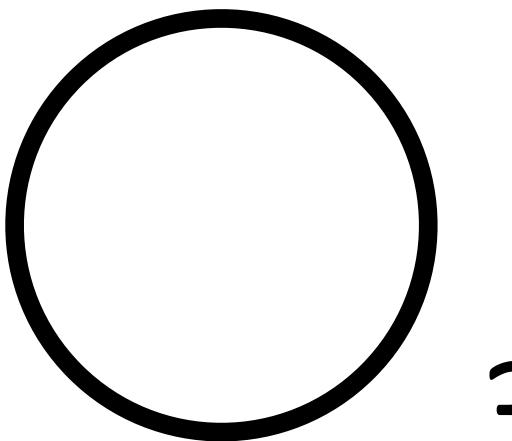
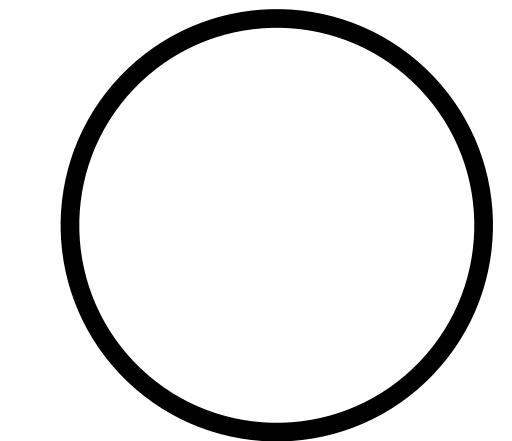
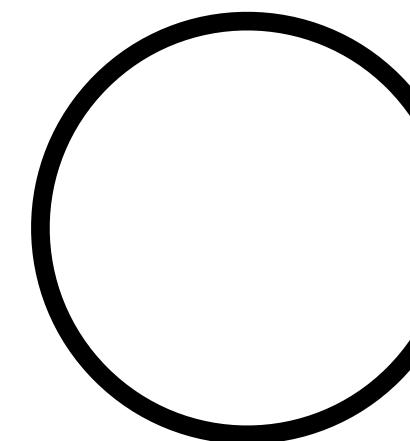
Sequence-to-sequence

go down the yellow hall turn left



]

Alignments



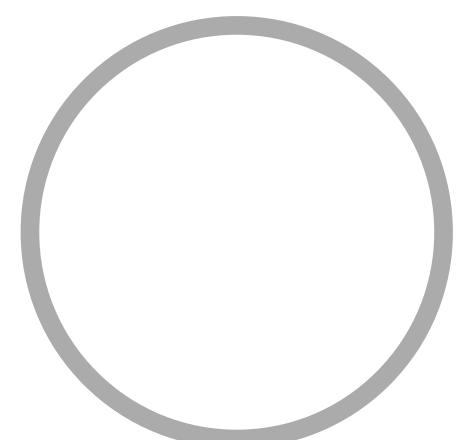
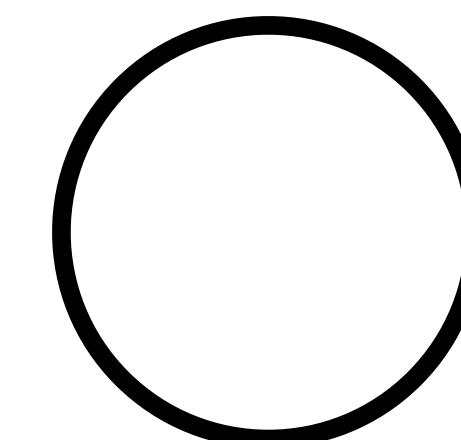
]

Actions



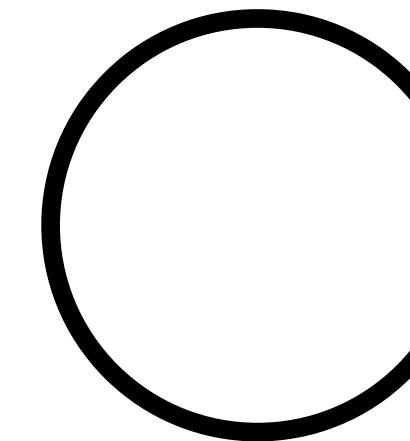
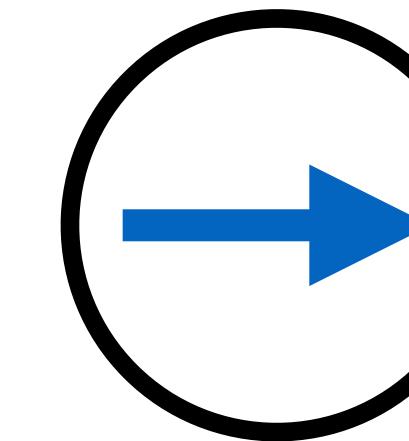
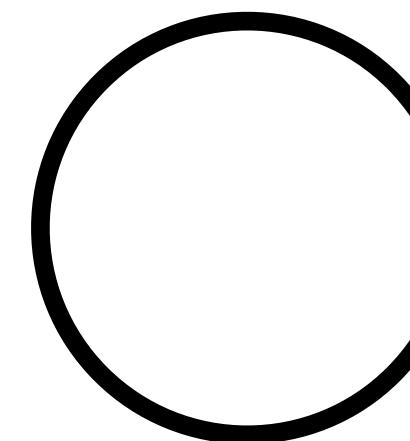
Sequence-to-sequence

go down the yellow hall turn left



]

Alignments



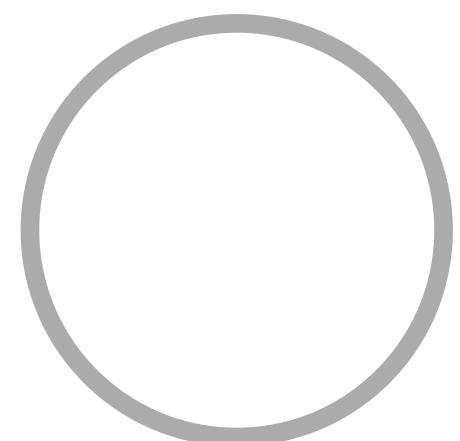
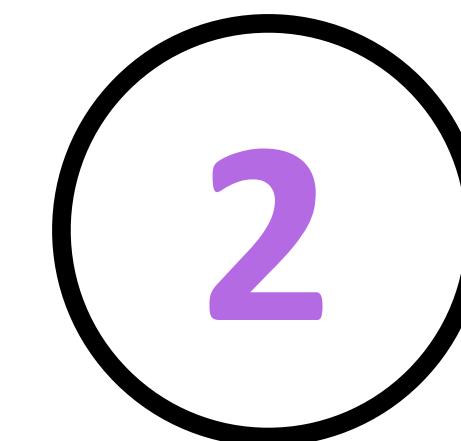
]

Actions

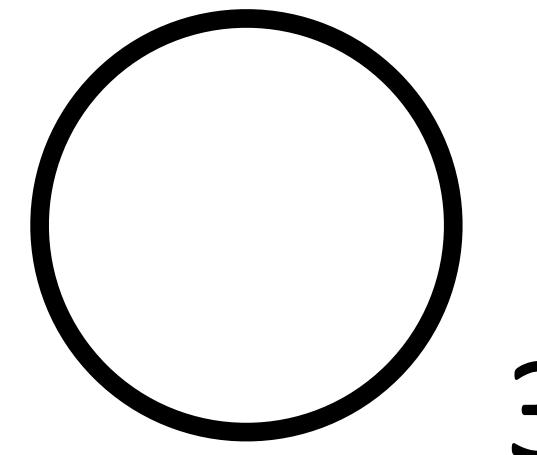
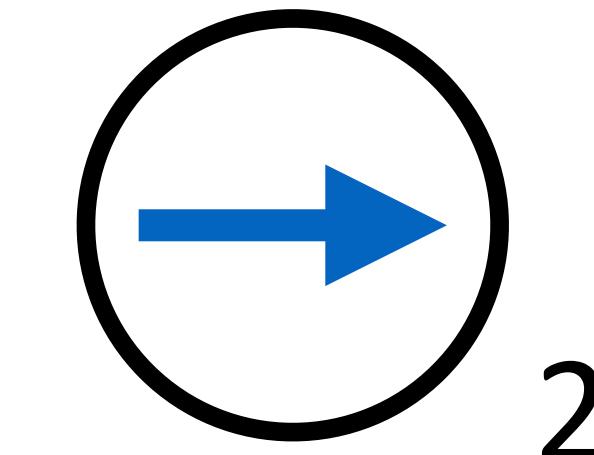
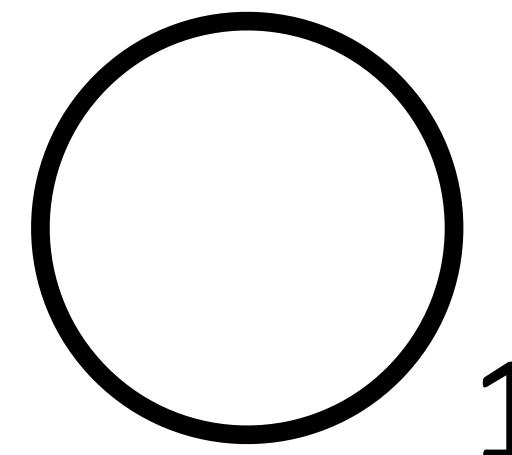


Sequence-to-sequence

go down the yellow hall turn left



Alignments

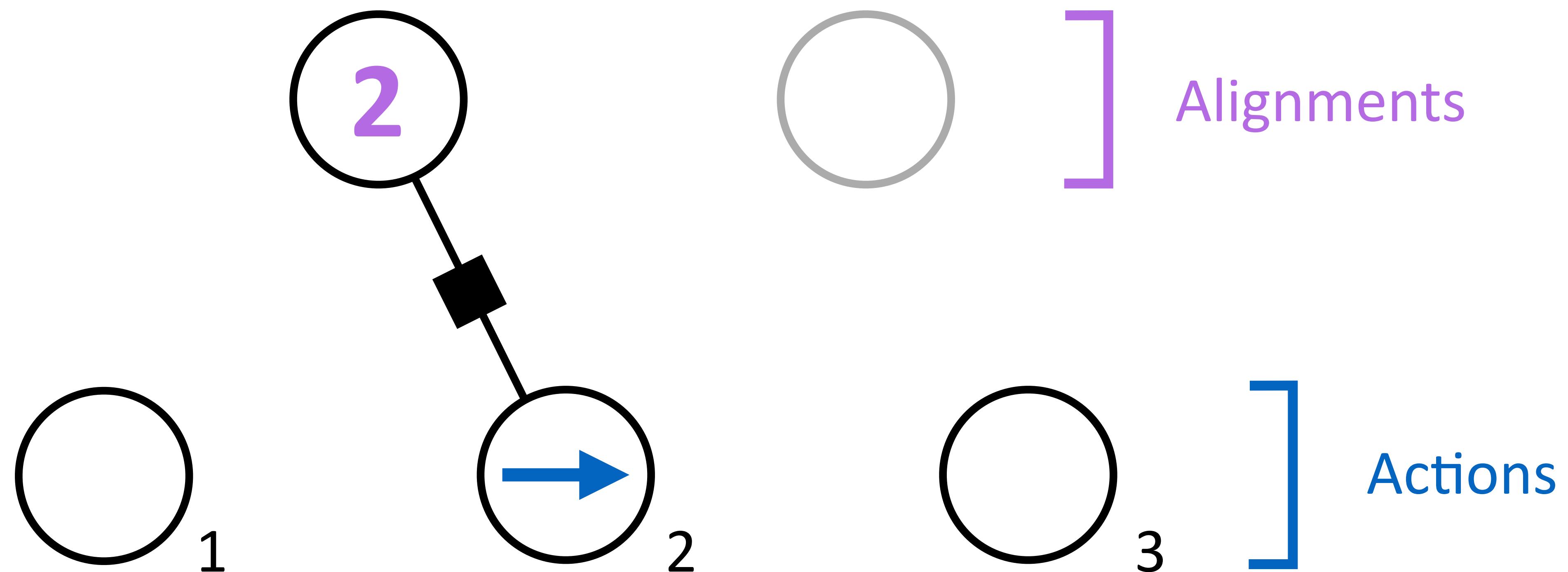


Actions



Sequence-to-sequence

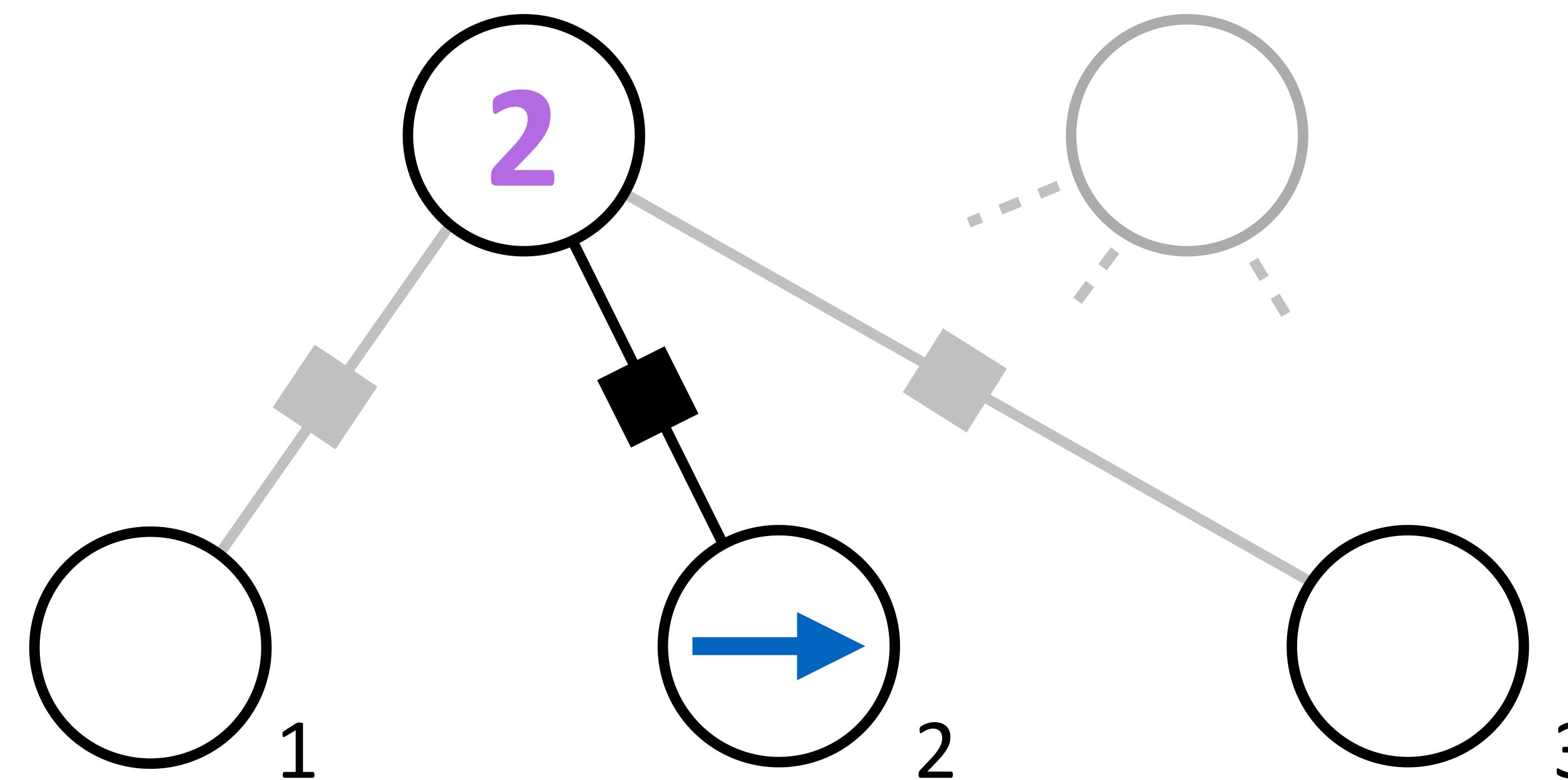
go down the yellow hall turn left





Sequence-to-sequence

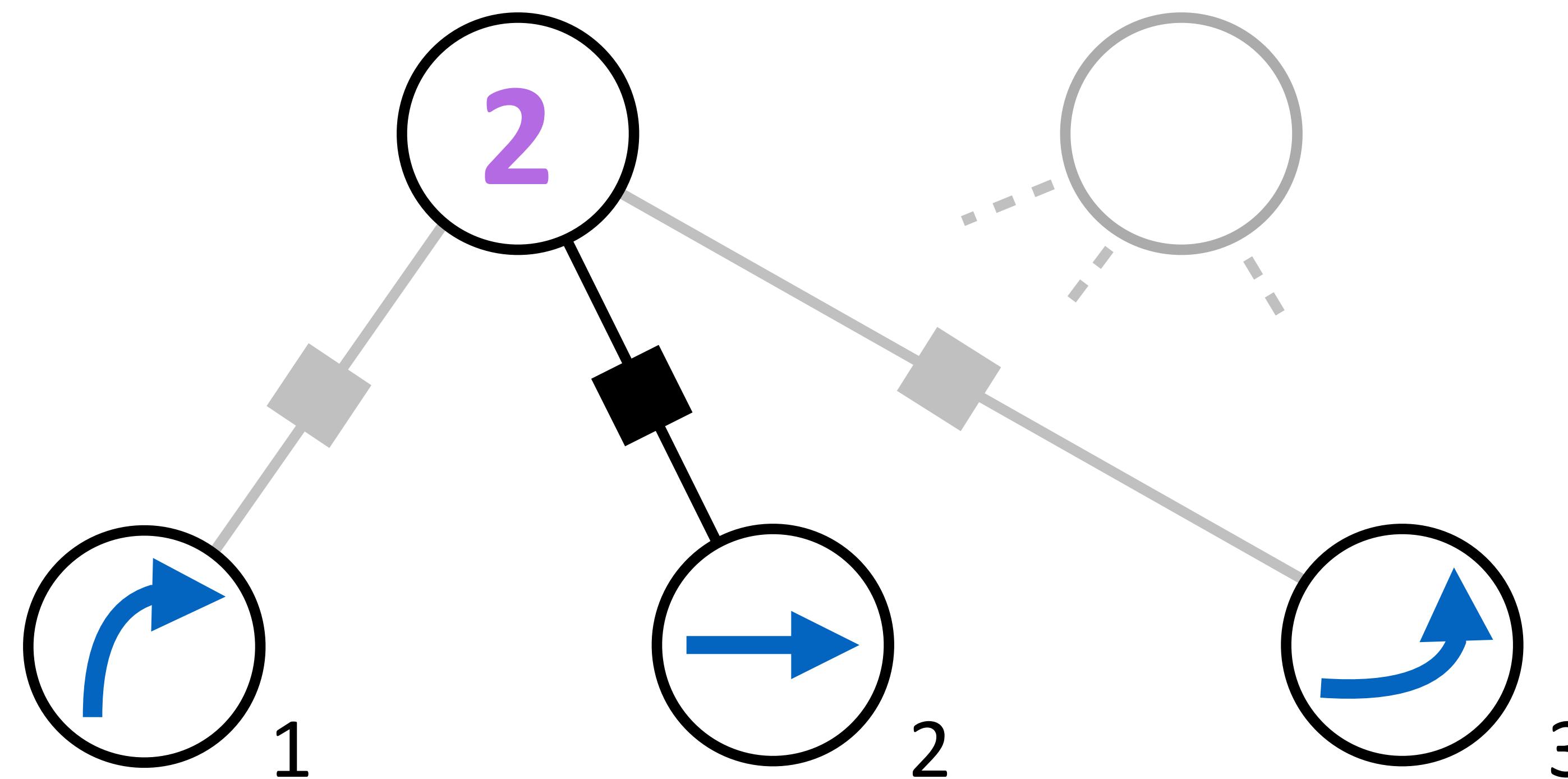
go down the yellow hall turn left





Sequence-to-sequence

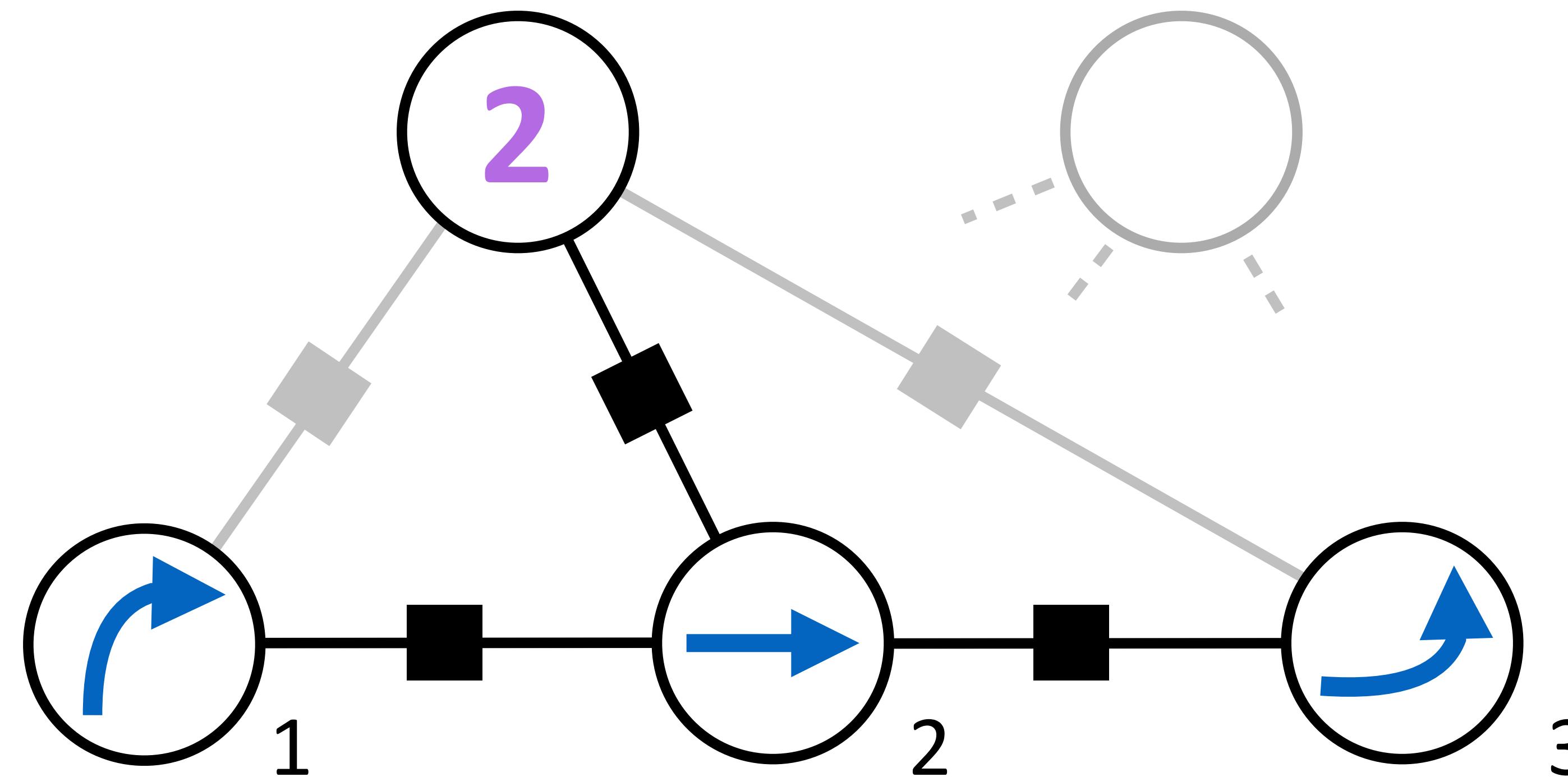
go down the yellow hall turn left





Sequence-to-sequence

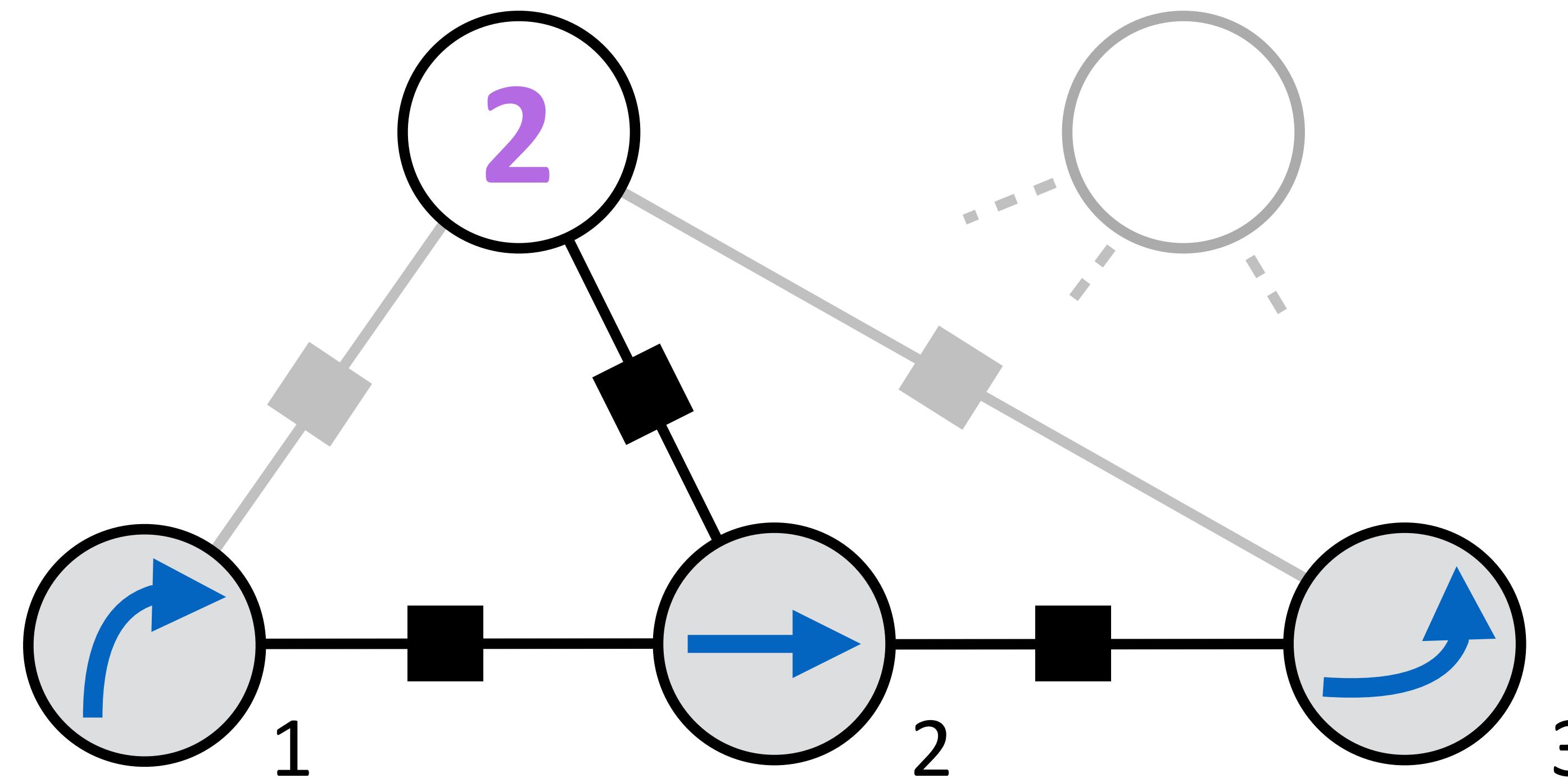
go down the yellow hall turn left





Sequence-to-sequence

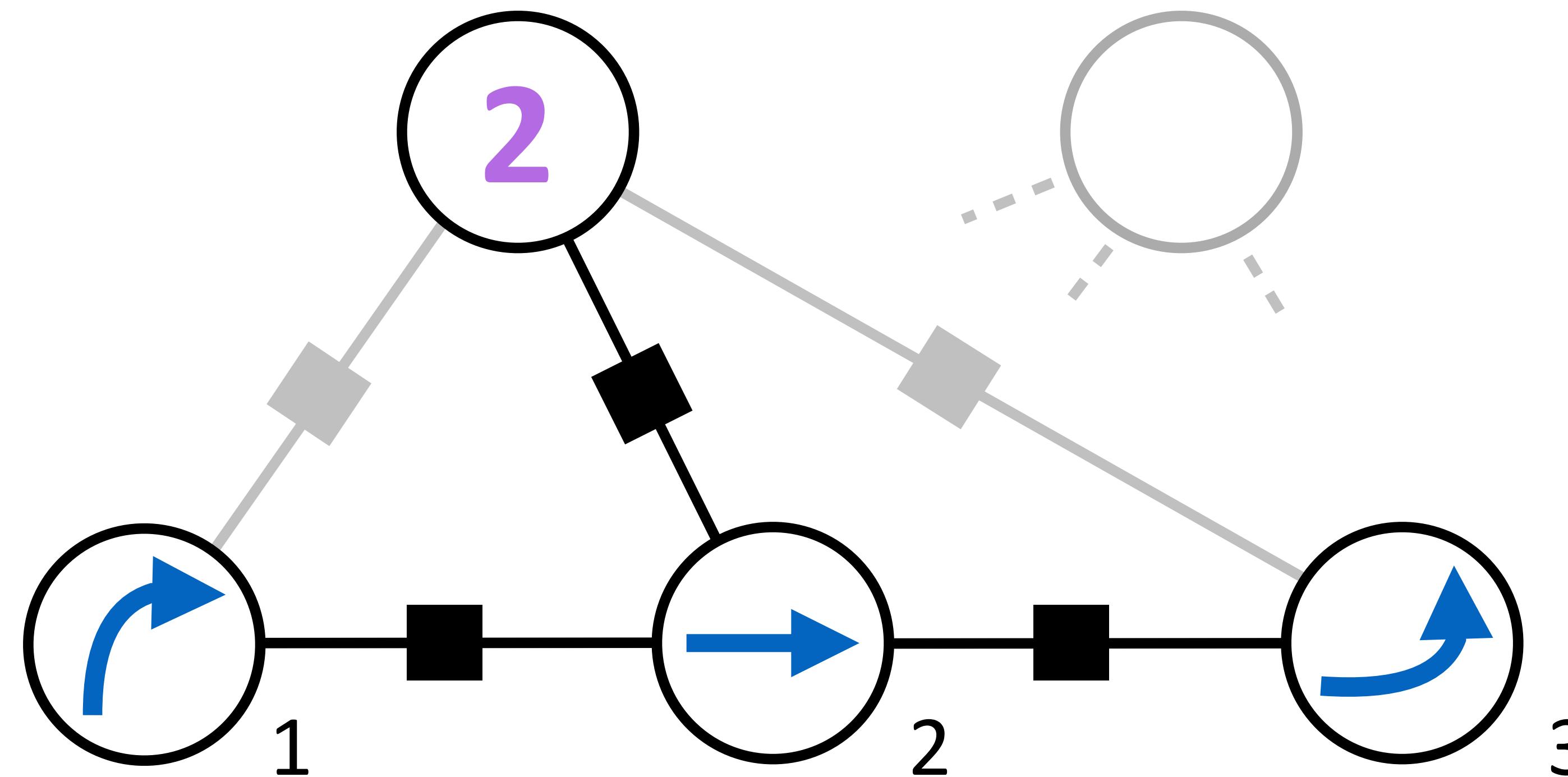
go down the yellow hall turn left





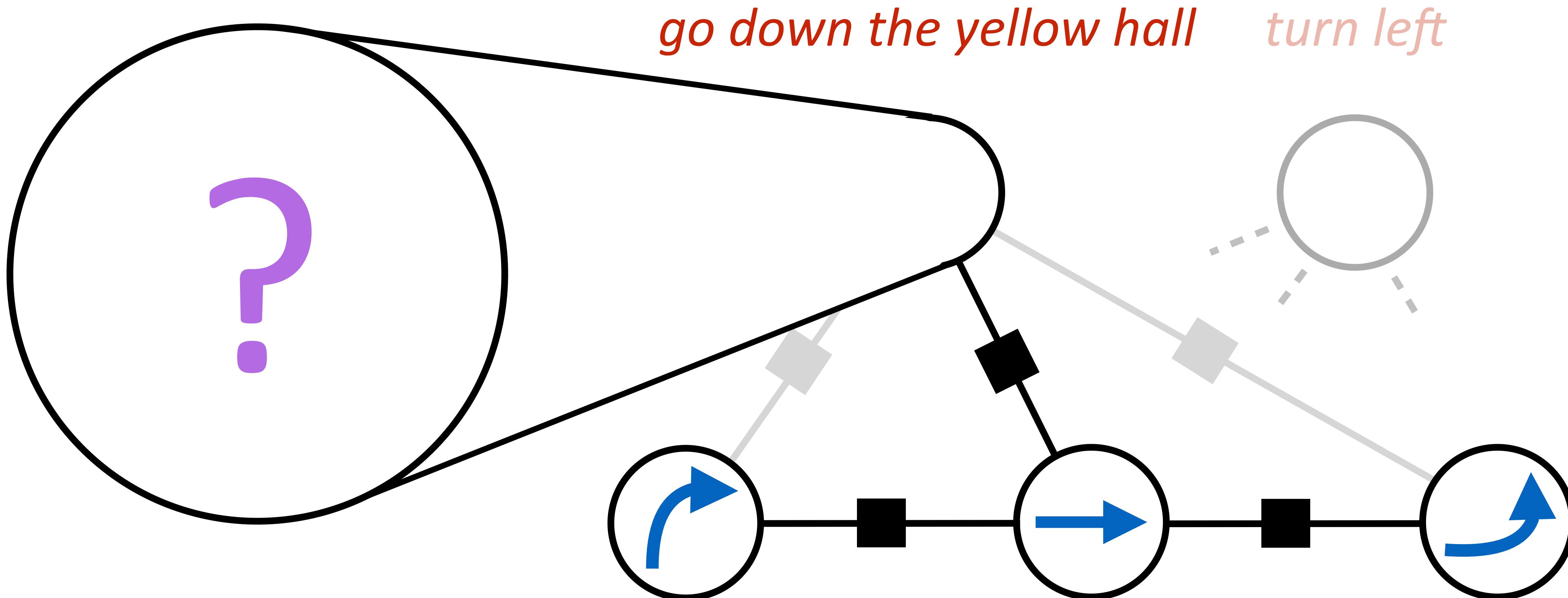
Sequence-to-sequence

go down the yellow hall turn left





Sequence-to-sequence



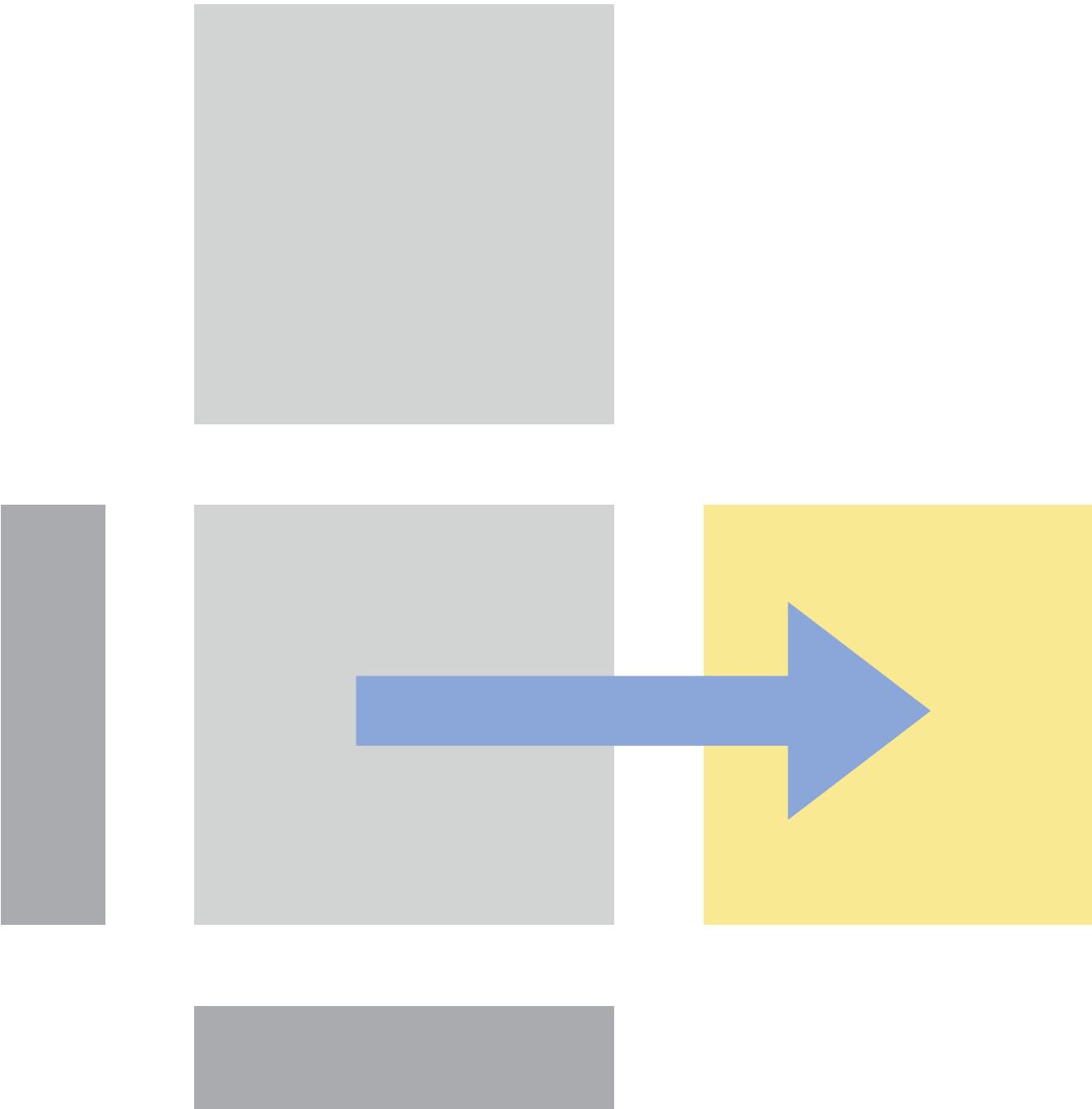
STRUCTURE-TO-STRUCTURE MODEL





Structure-to-structure

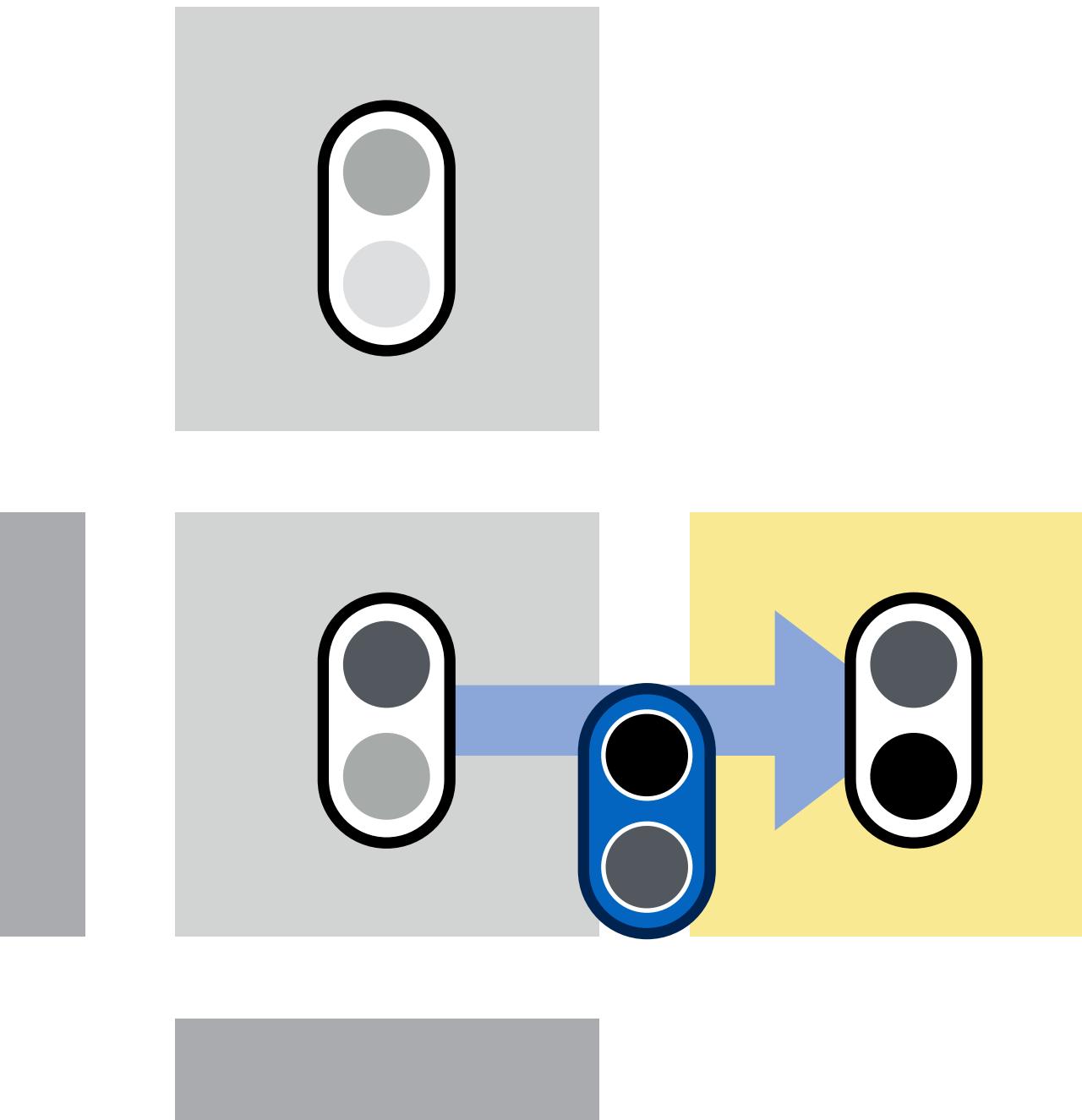
go down the yellow hall





Structure-to-structure

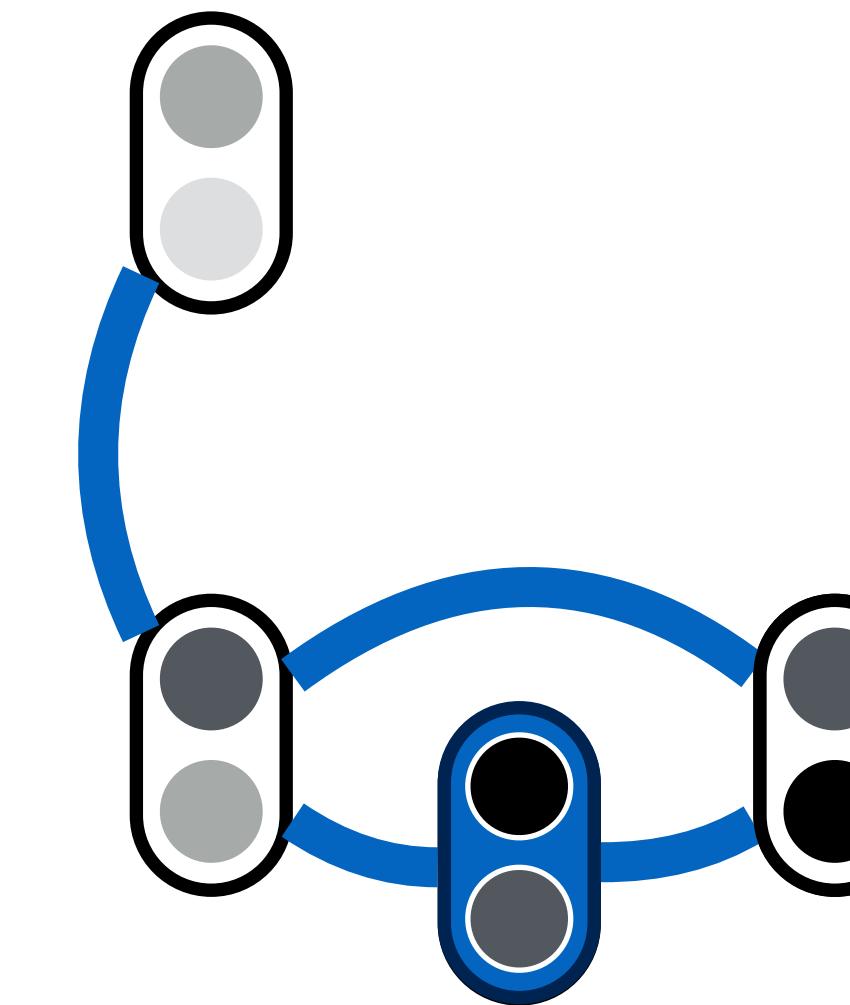
go → down → hall ↗ *yellow*
↗ *the*





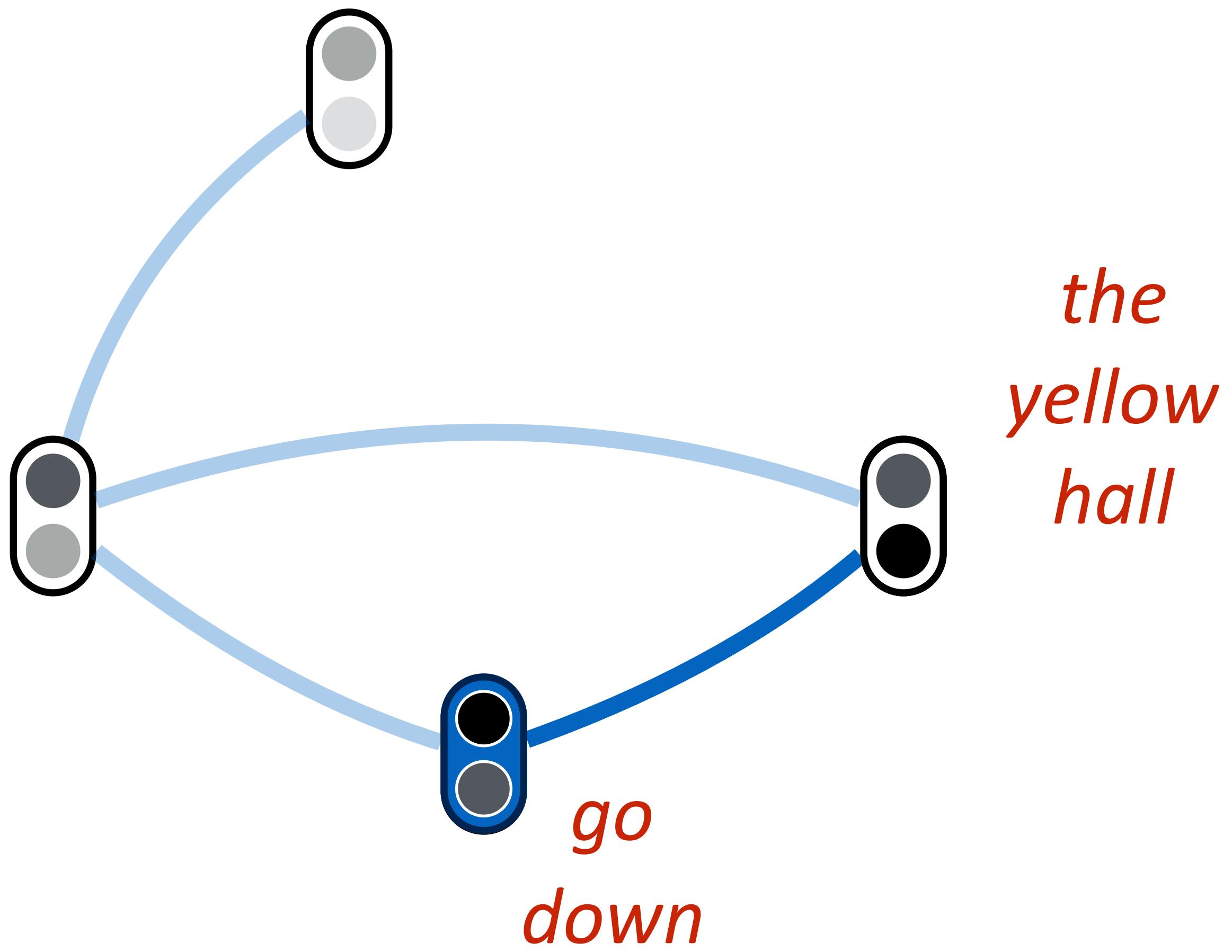
Structure-to-structure

go → down → hall ↗ yellow
↗ the



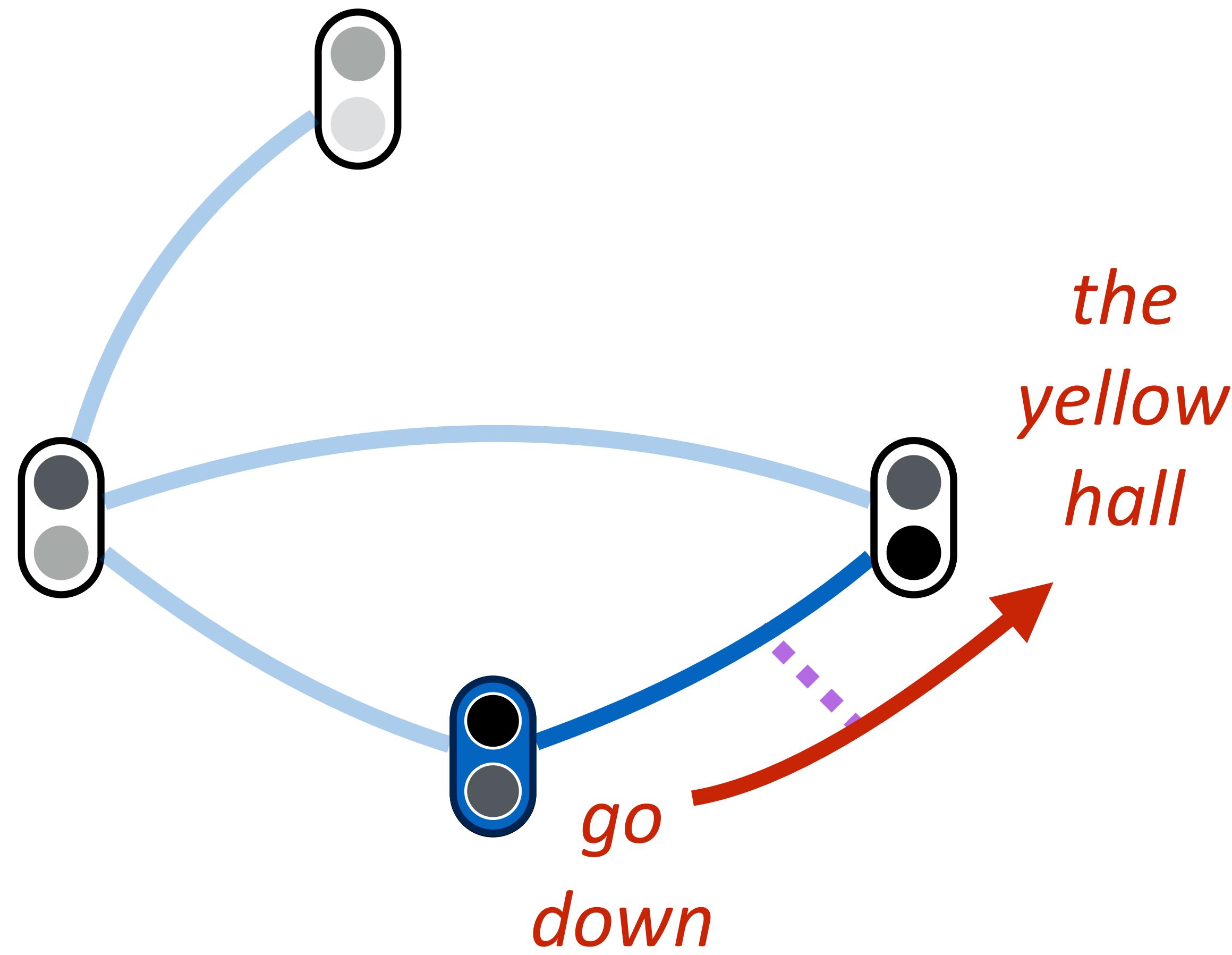


Structure-to-structure



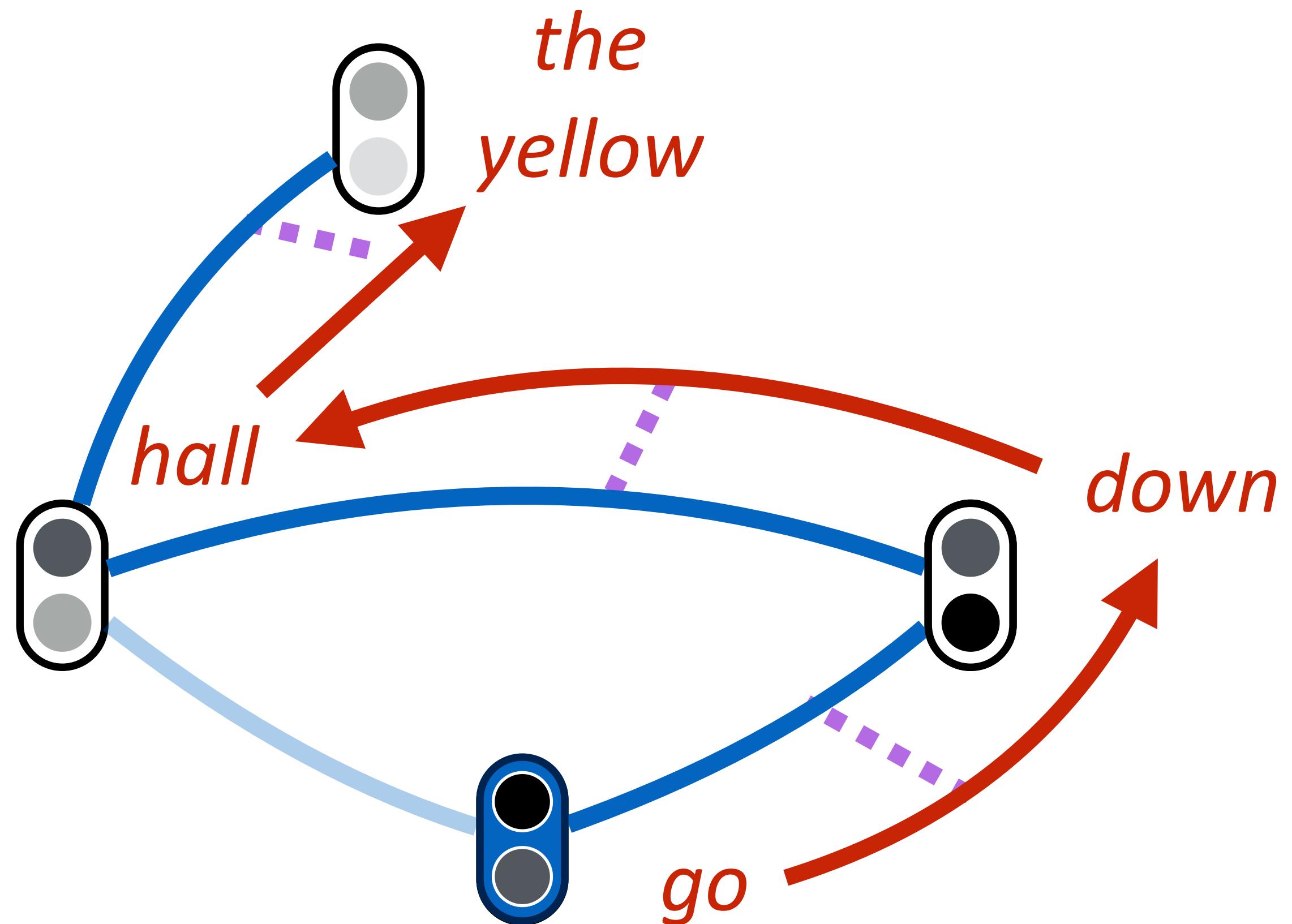


Structure-to-structure





Structure-to-structure

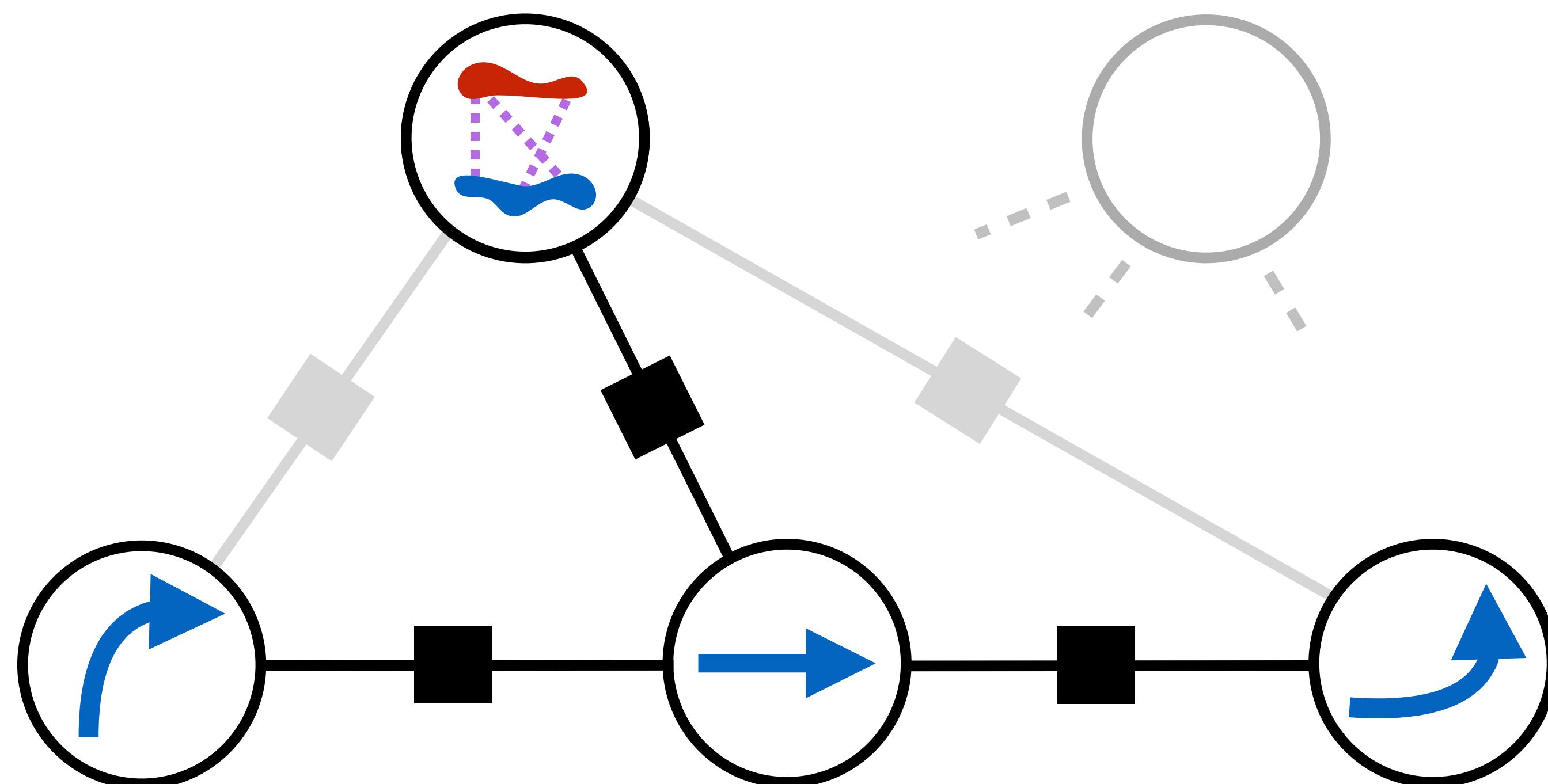


JOINT MODEL

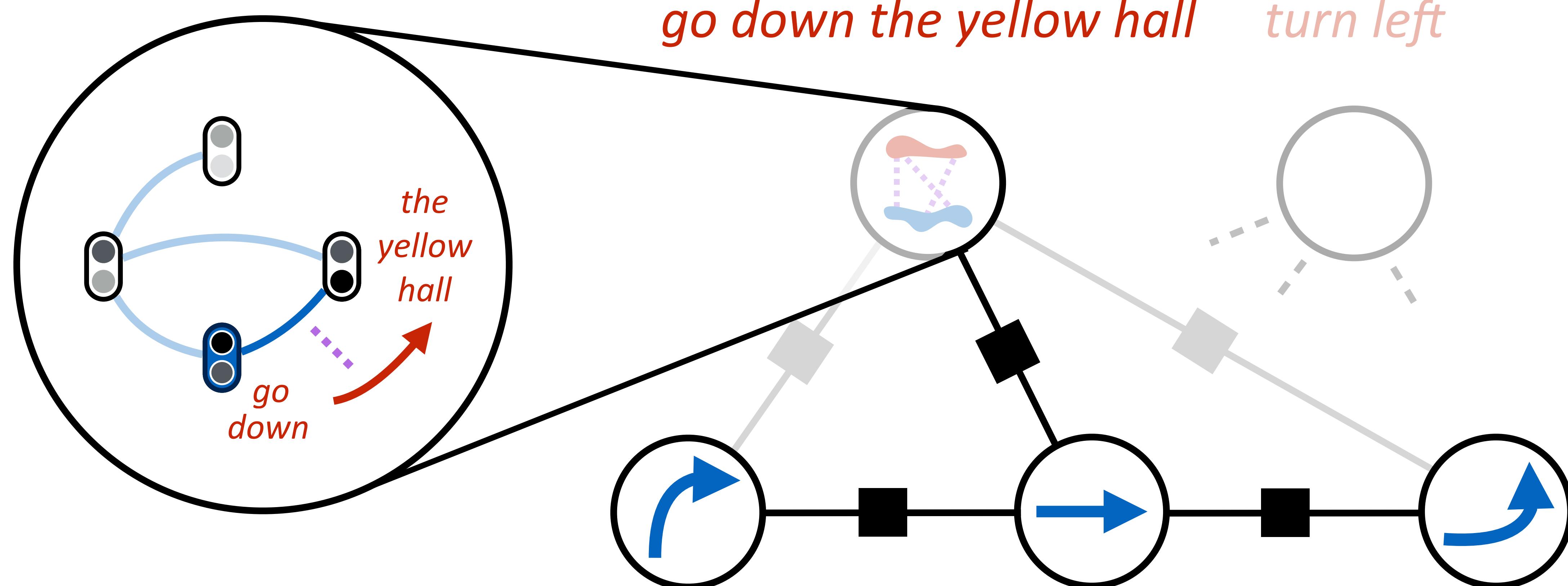


Joint model

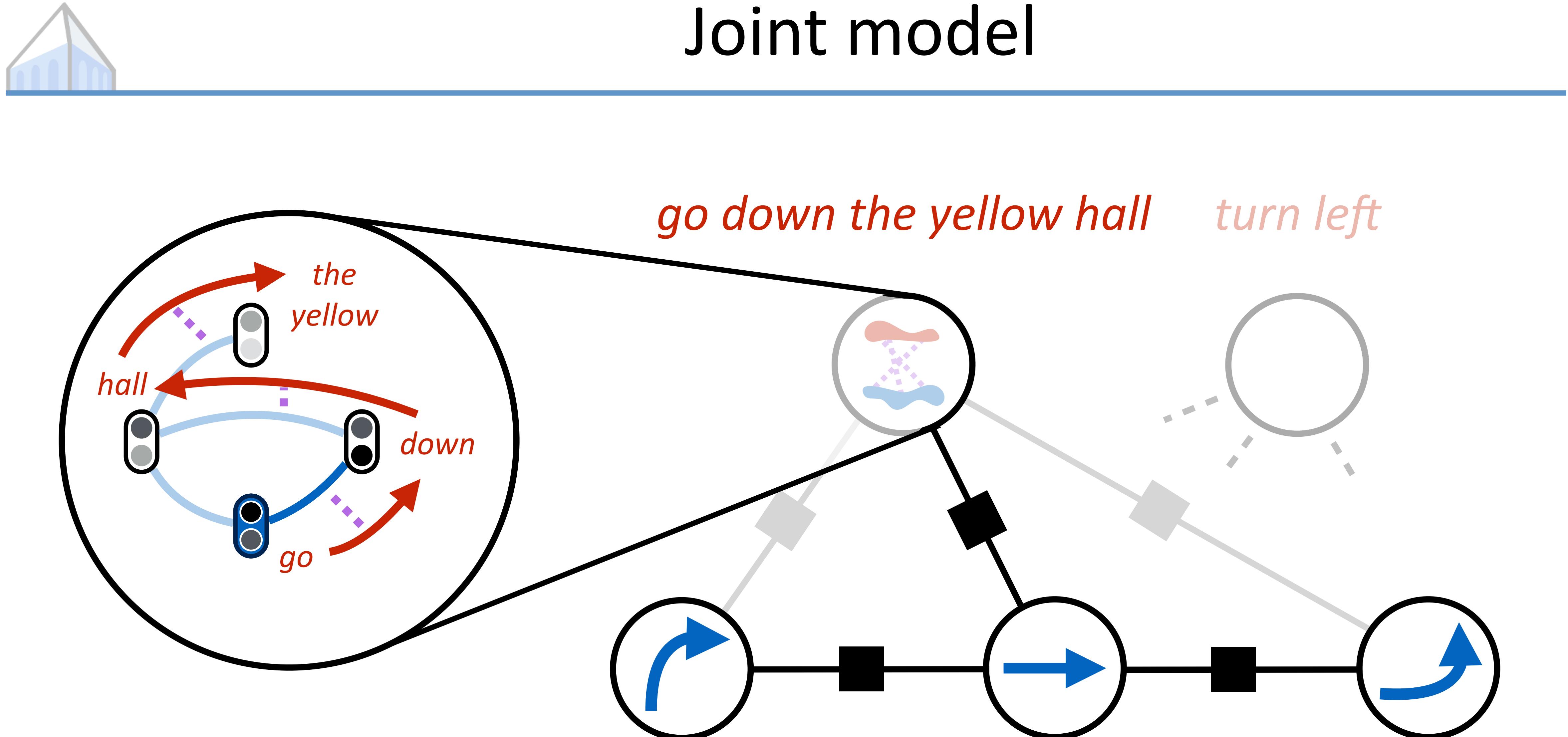
go down the yellow hall turn left



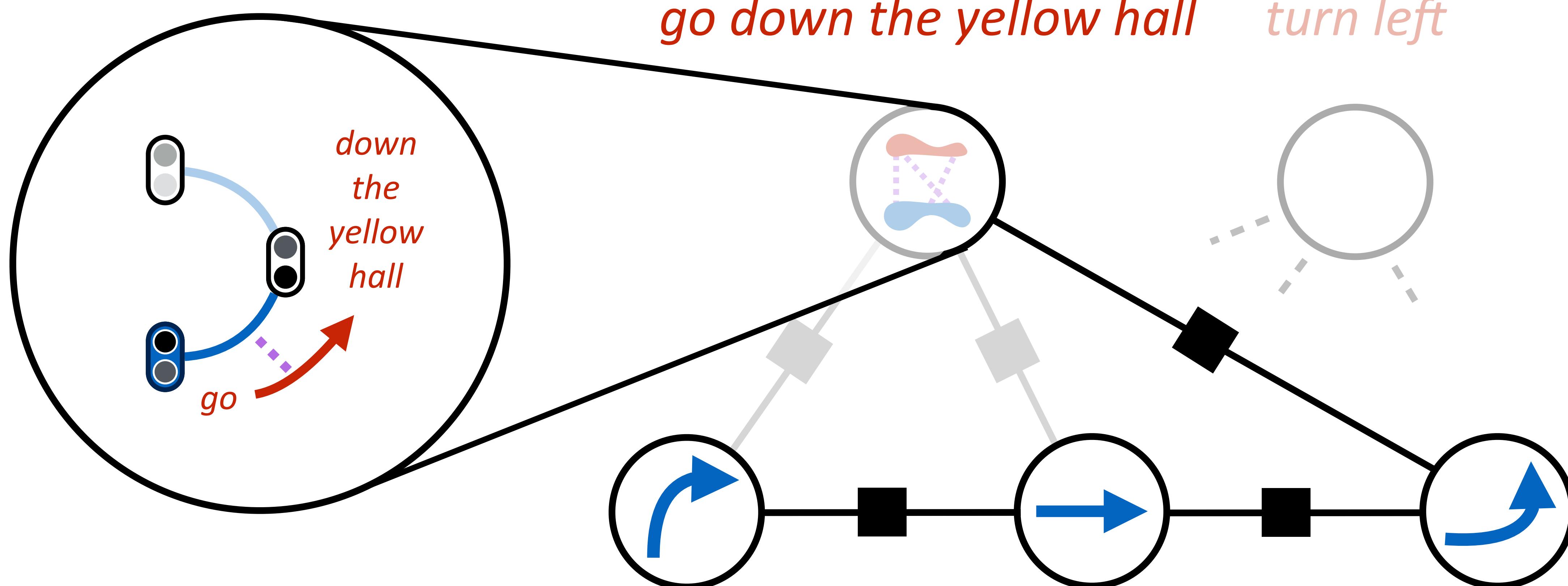
Joint model



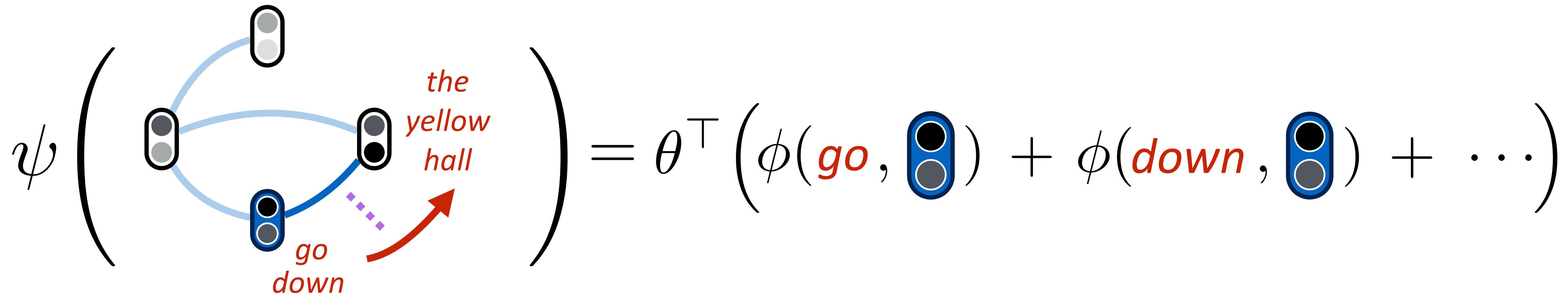
Joint model



Joint model



Joint model



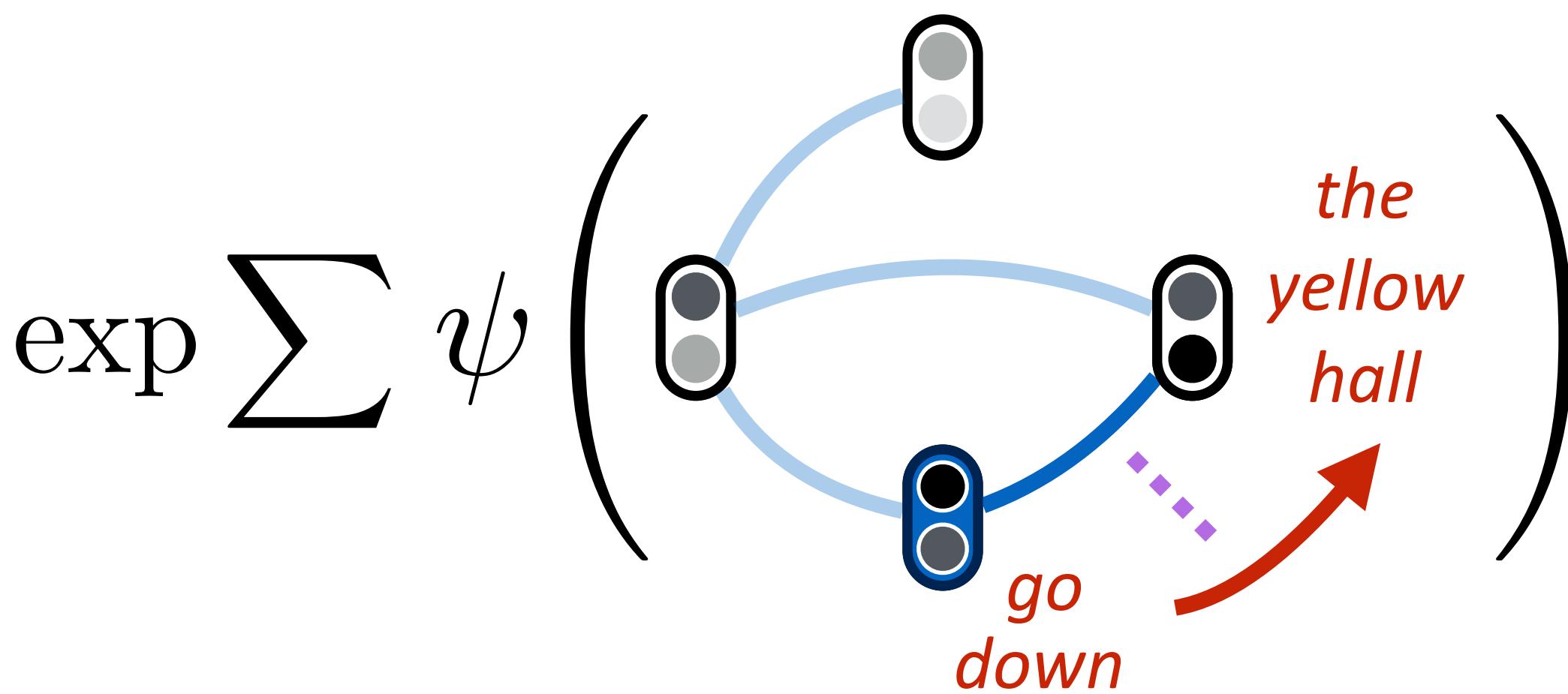
The diagram illustrates a joint model for a robot navigating through a yellow hall. On the left, a small icon of a building with a blue roof and grey walls is shown above a horizontal blue line. To the right of this line, a state transition graph is displayed. The graph consists of four states represented by ovals with two circles inside: a top-left state with one dark grey circle and one light grey circle; a top-right state with one light grey circle and one white circle; a bottom-left state with one dark grey circle and one light grey circle; and a bottom-right state with one dark grey circle and one black circle. Blue curved arrows connect the top-left state to the top-right state, the top-right state to the bottom-right state, and the bottom-left state to the bottom-right state. A red curved arrow labeled "go down" points from the bottom-left state to the bottom-right state. A red dashed arrow labeled "the yellow hall" points from the bottom-right state back to the bottom-left state. To the right of the graph, the equation $\psi \left(\dots \right) = \theta^\top \left(\phi(go, \dots) + \phi(down, \dots) + \dots \right)$ is shown, where $\phi(go, \dots)$ and $\phi(down, \dots)$ are highlighted in red, corresponding to the red arrows in the graph.

$$\psi \left(\dots \right) = \theta^\top \left(\phi(go, \dots) + \phi(down, \dots) + \dots \right)$$

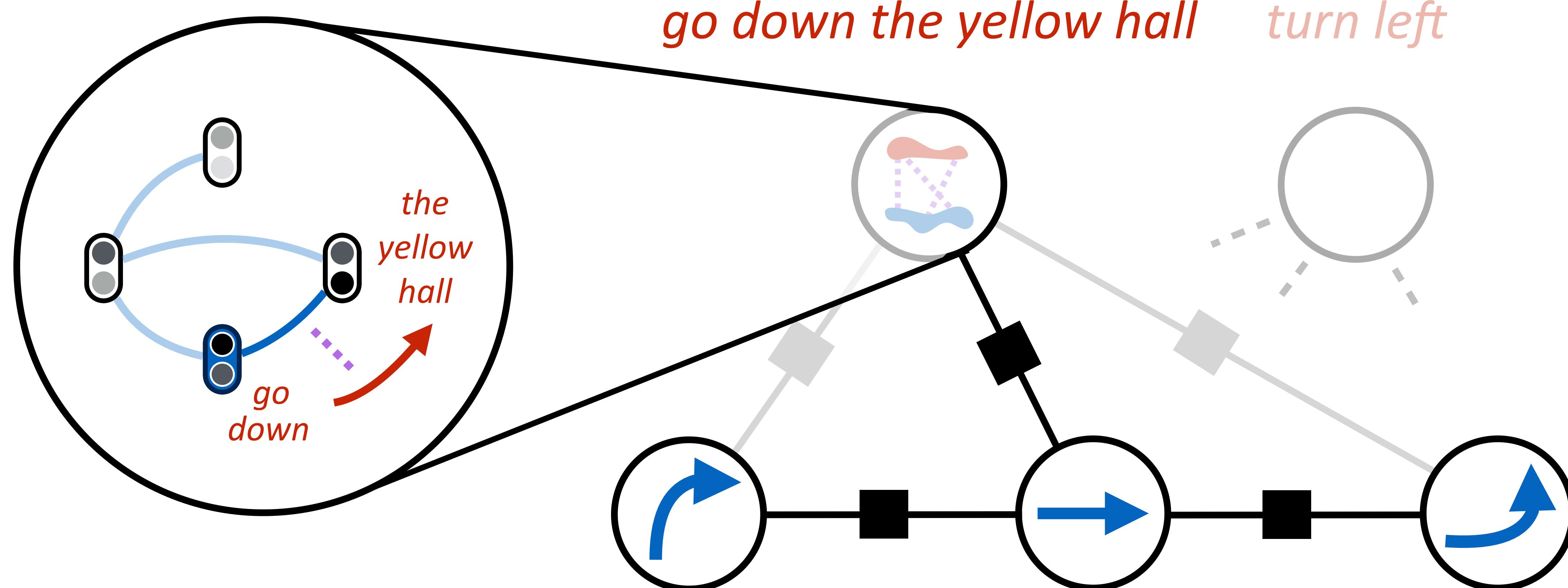


Joint model

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



Joint model



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

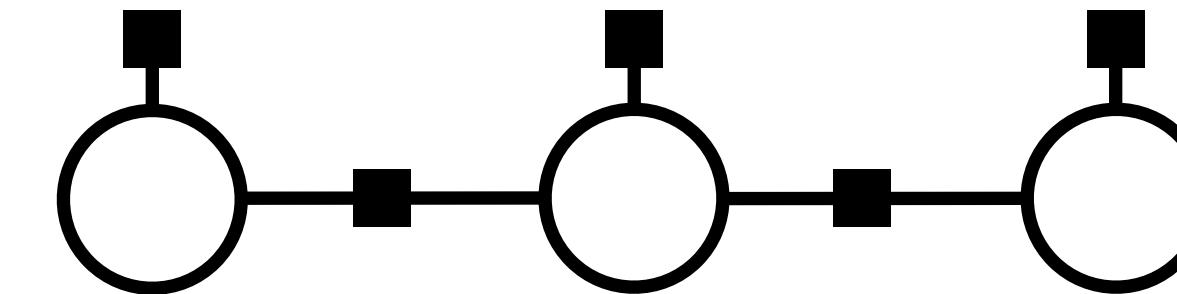
max
actions,
alignments

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

max
alignments

using custom alignment dynamic program

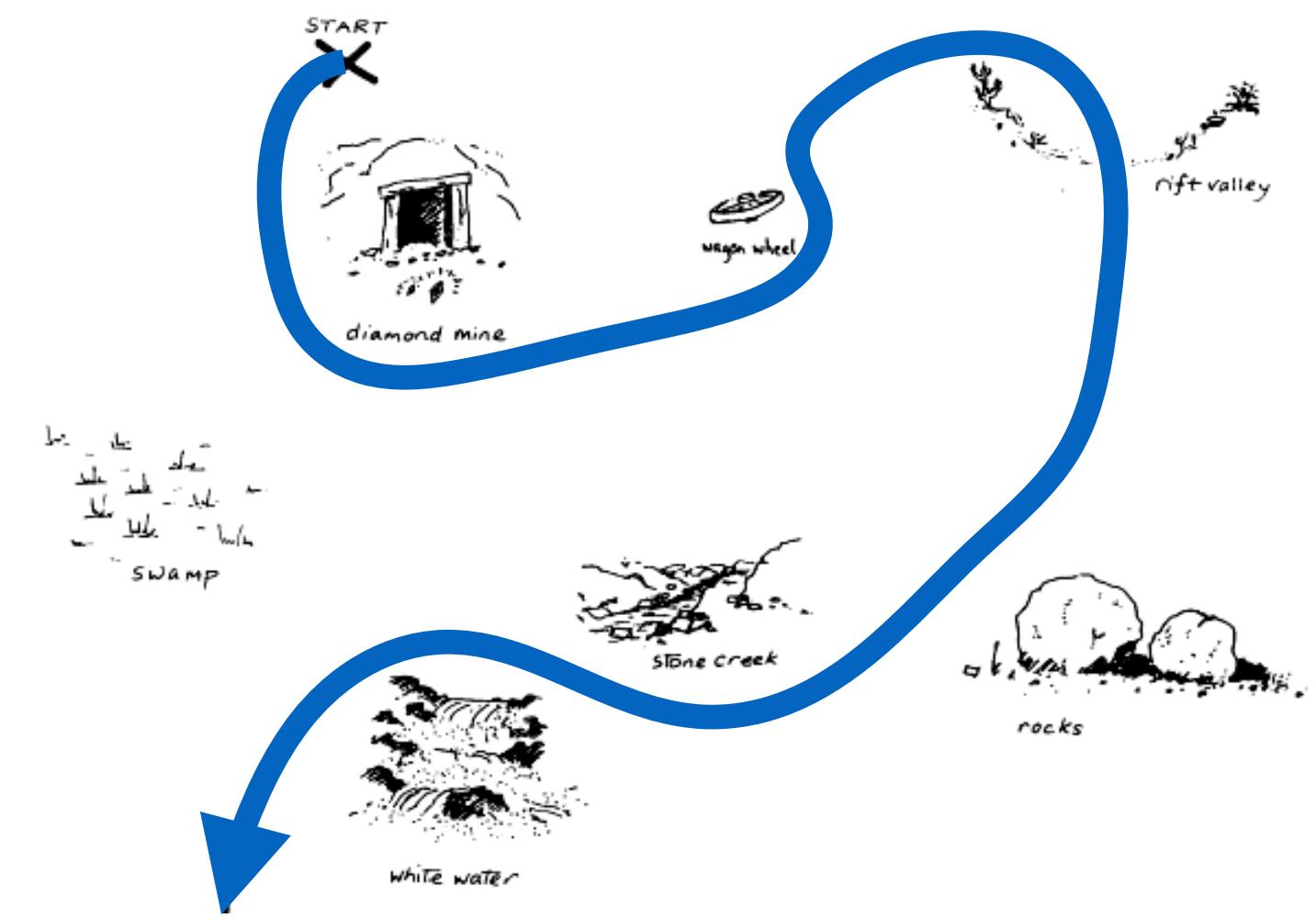
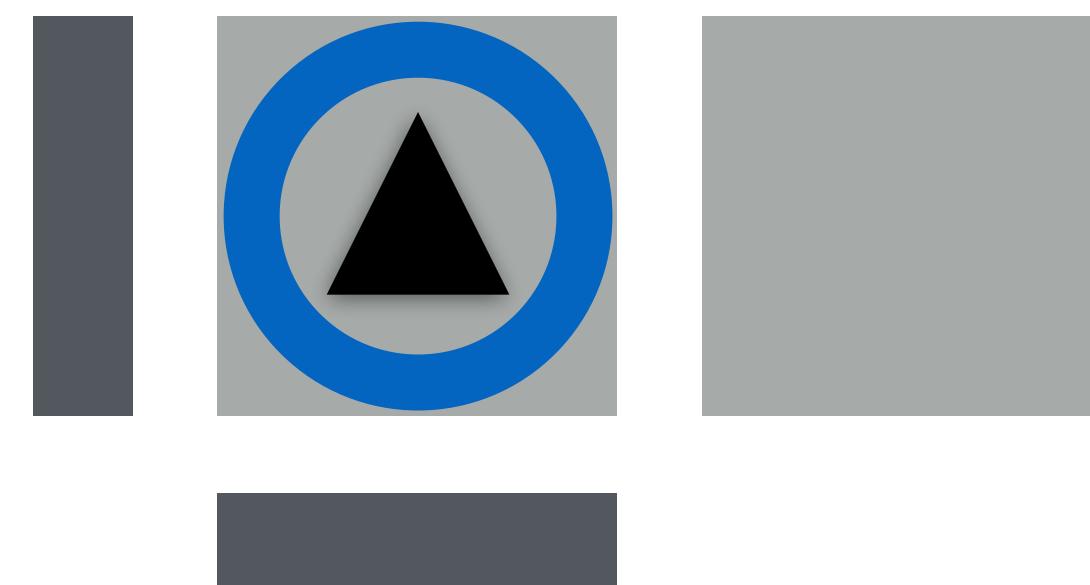
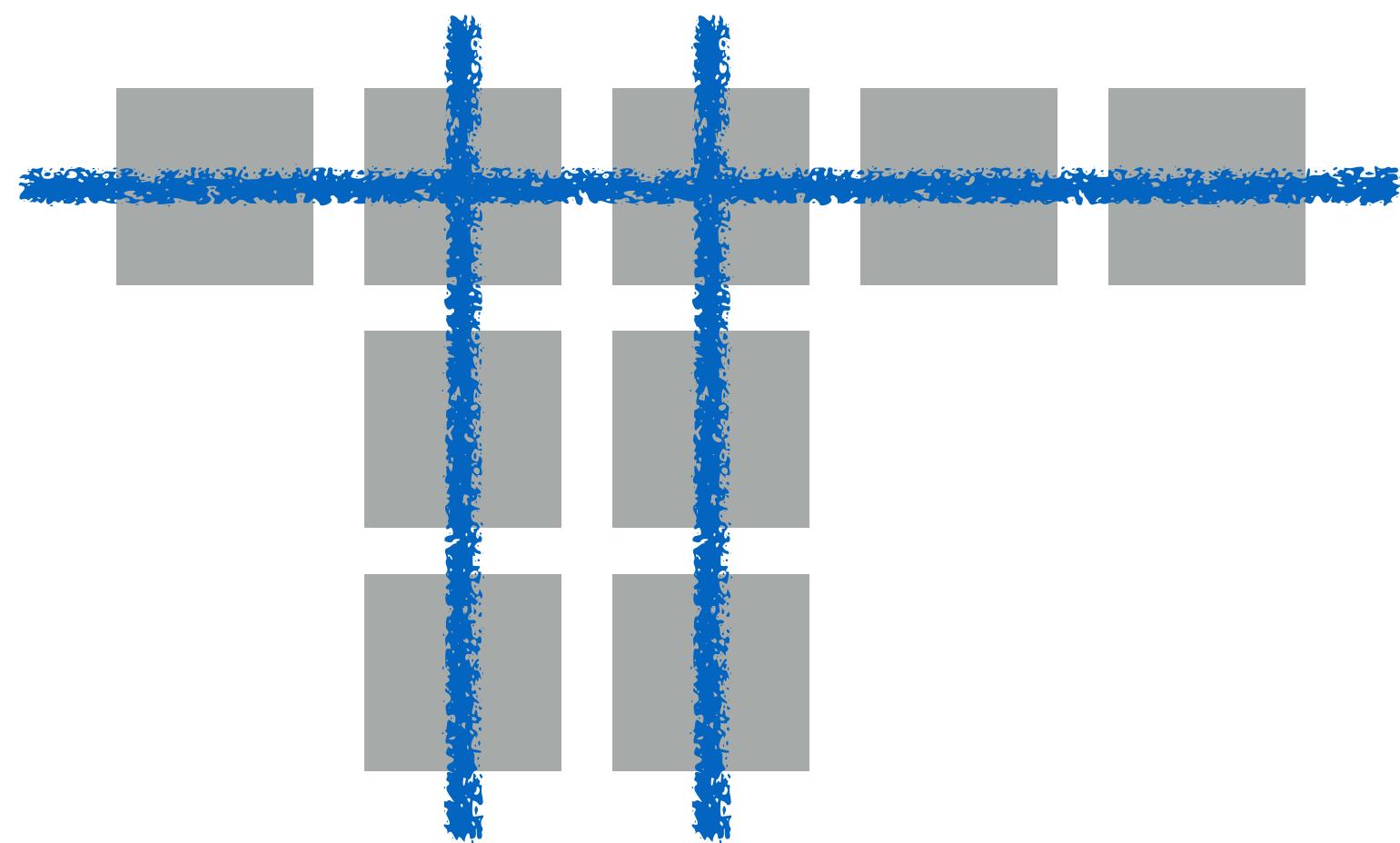
max
actions



using a planner

EXPERIMENTS

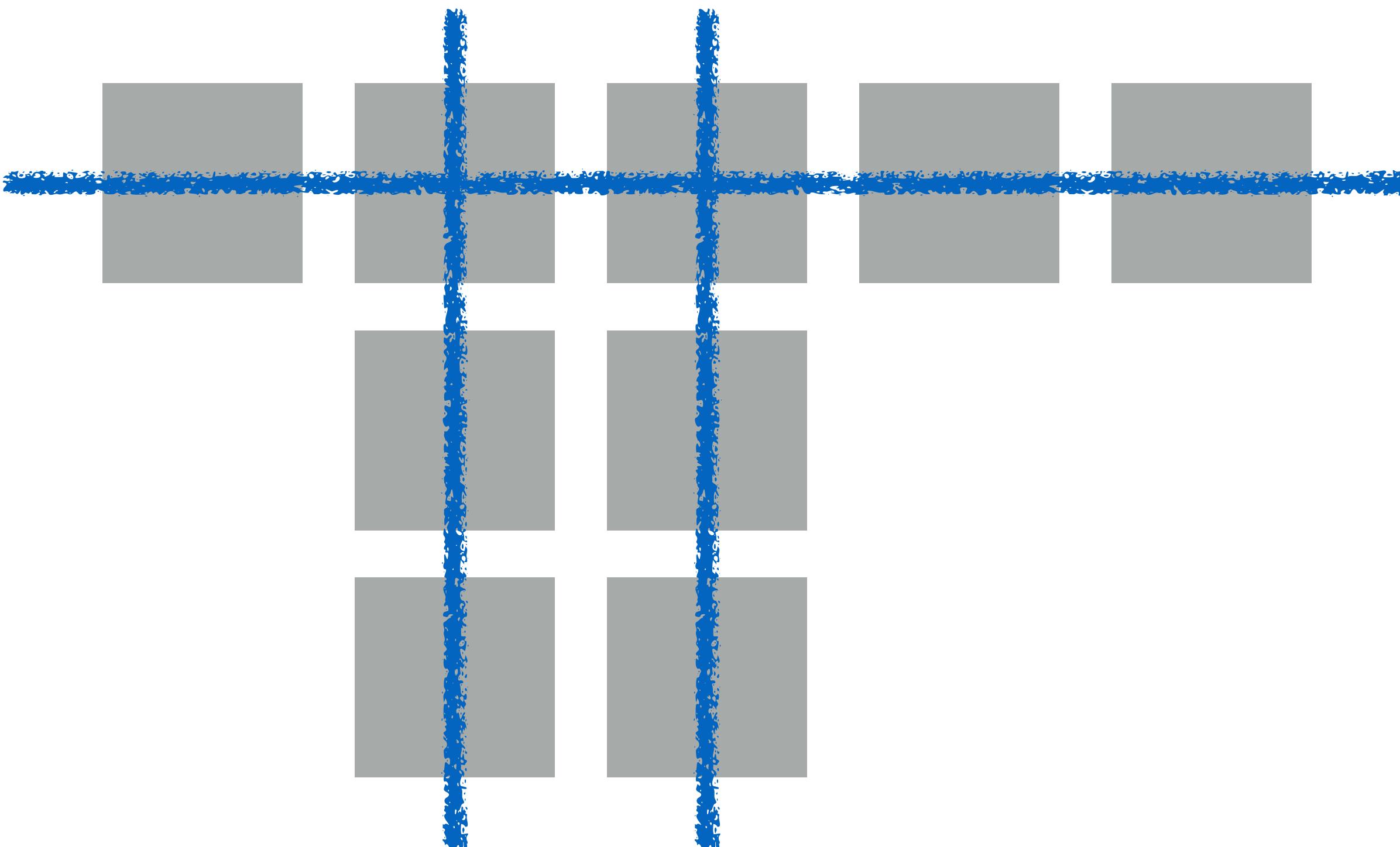
Three tasks





Puzzle solving

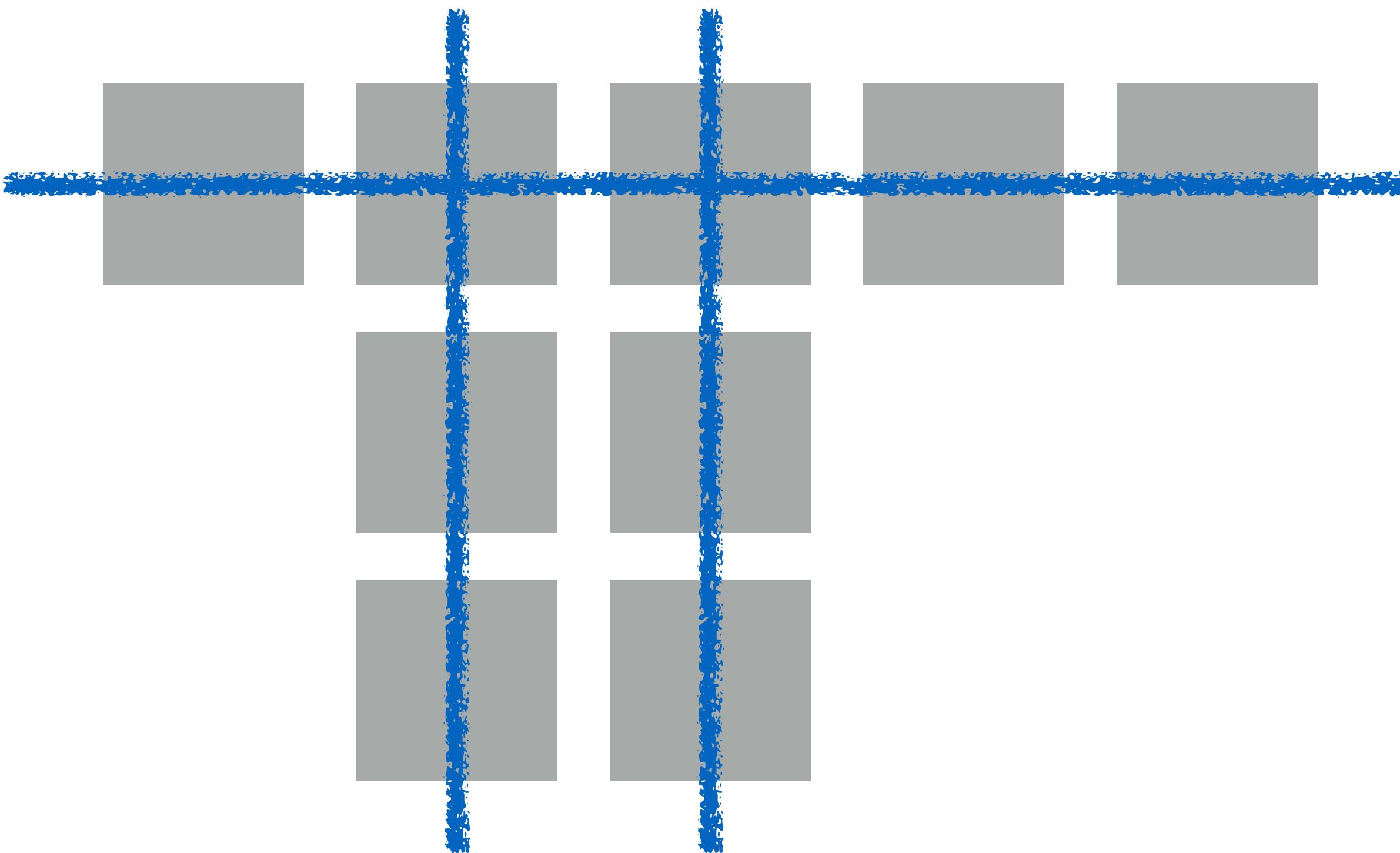
clear the two long columns, and then the row





Puzzle solving

clear the two long columns, and then the row





Puzzle solving

Plan execution (%)

90

60

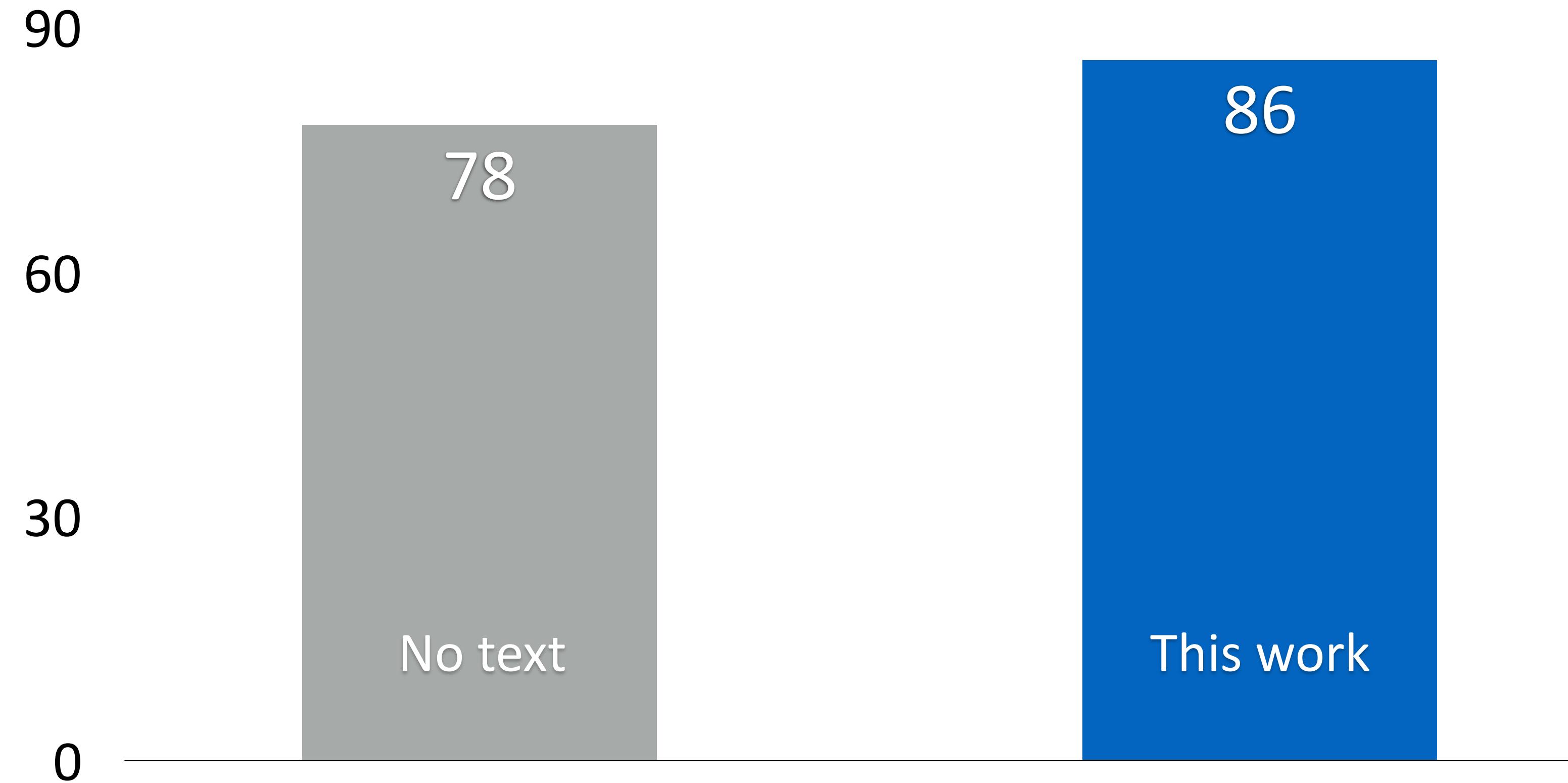
30

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 (%)

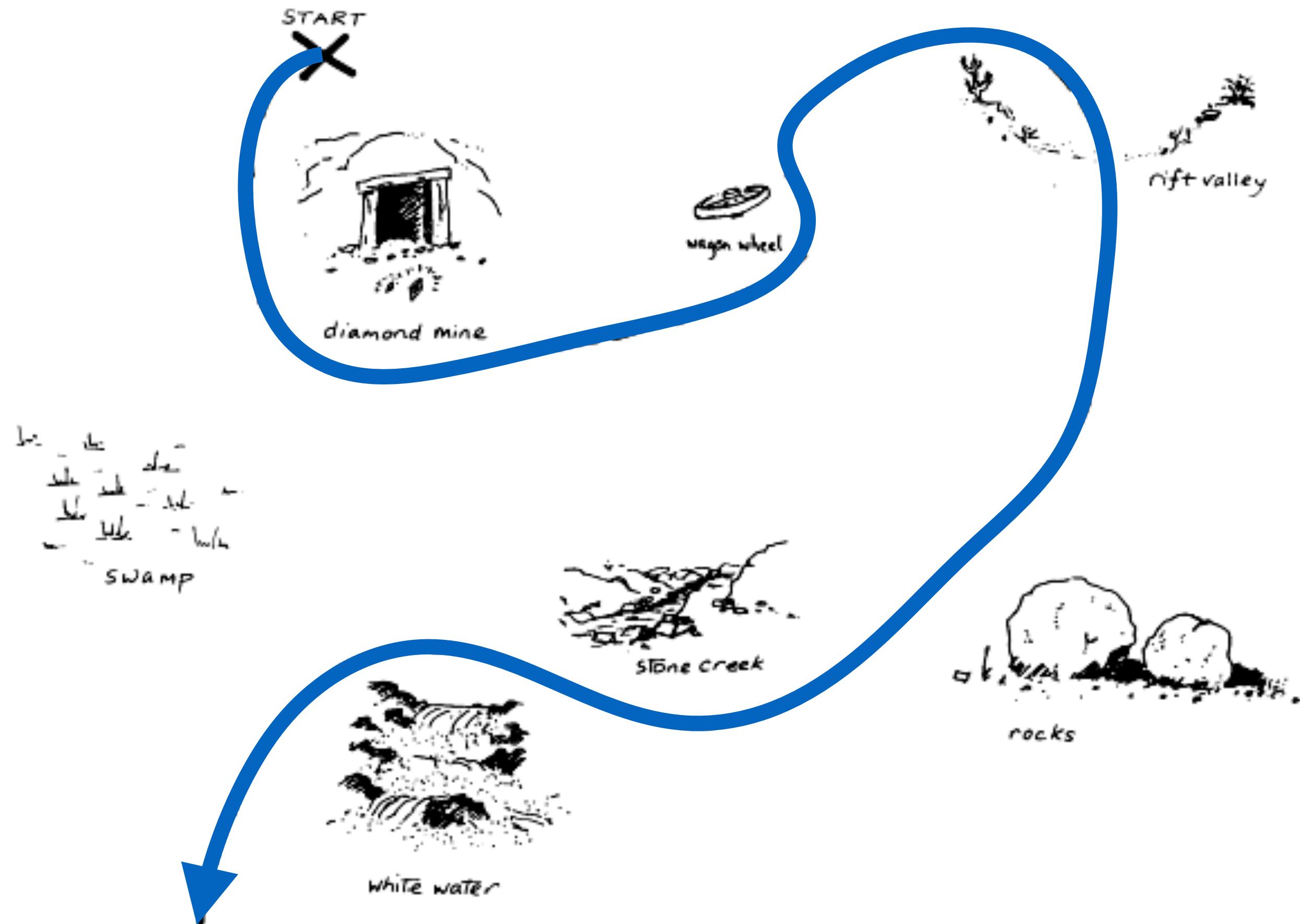
65

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

Task completion (%)

65

55

45

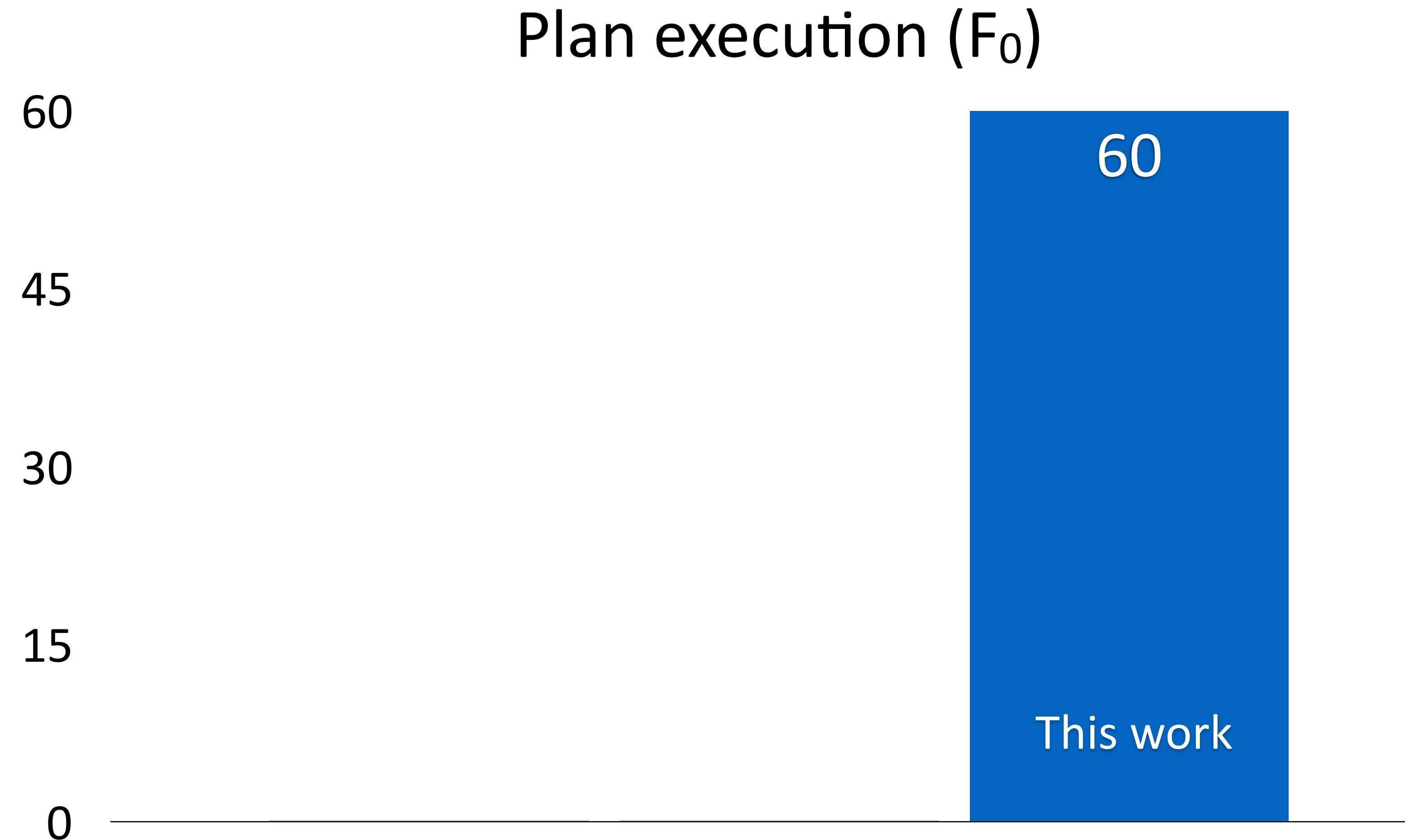
35

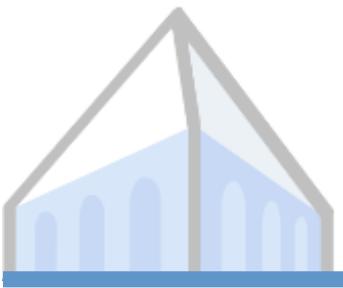
60

This work



Ablations: map reading





Conclusion

Structured alignment/decoding gives us best aspects of:

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

THANK YOU