

Observation (o_t)



Environment

Agent

Action (a_t)

$P(x)$

0.3

0.2

0.1

1

2

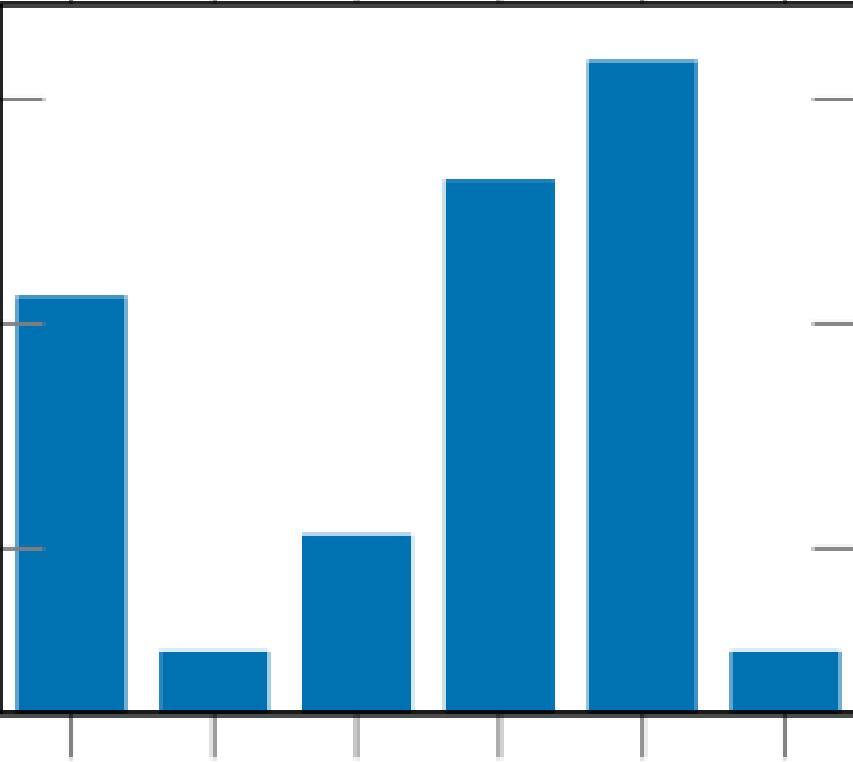
3

x

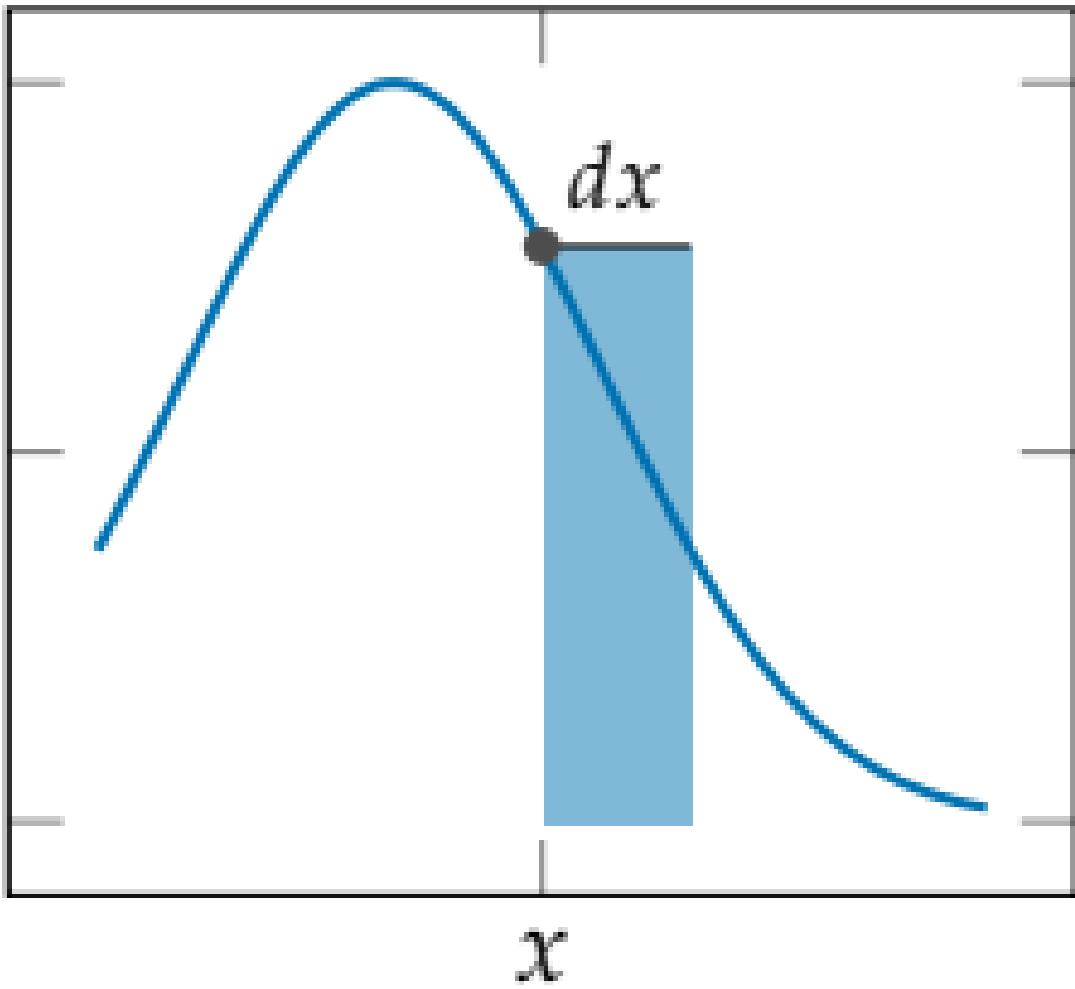
4

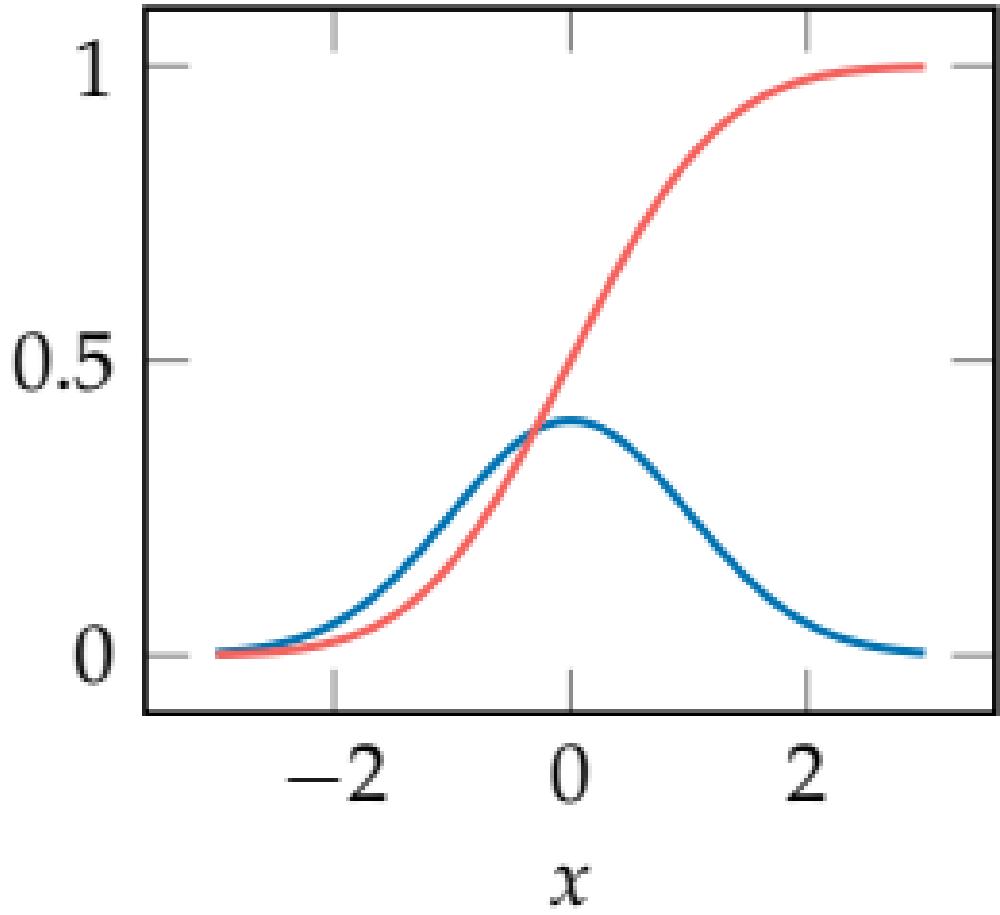
5

6



$p(x)$





— $p(x)$

— $\text{cdf}_X(x)$

quantile_X(α)

2

0

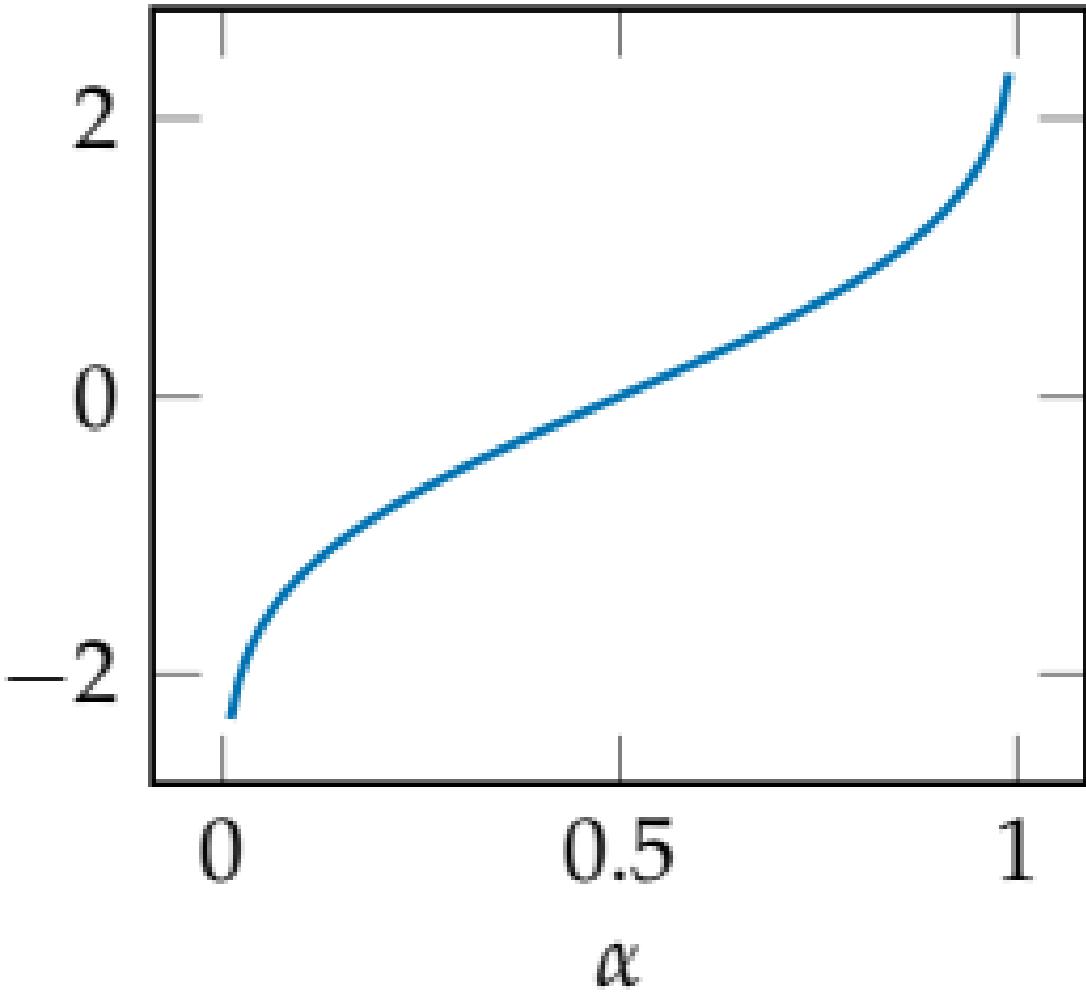
-2

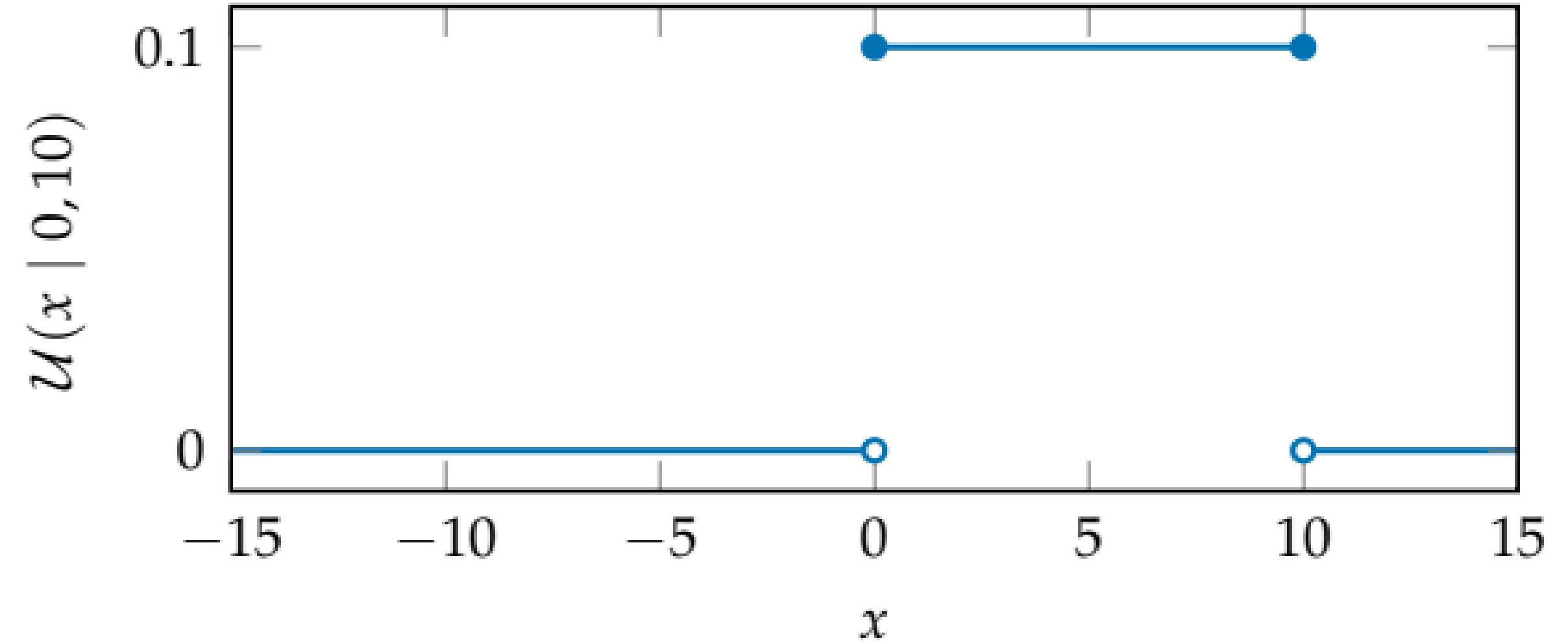
0

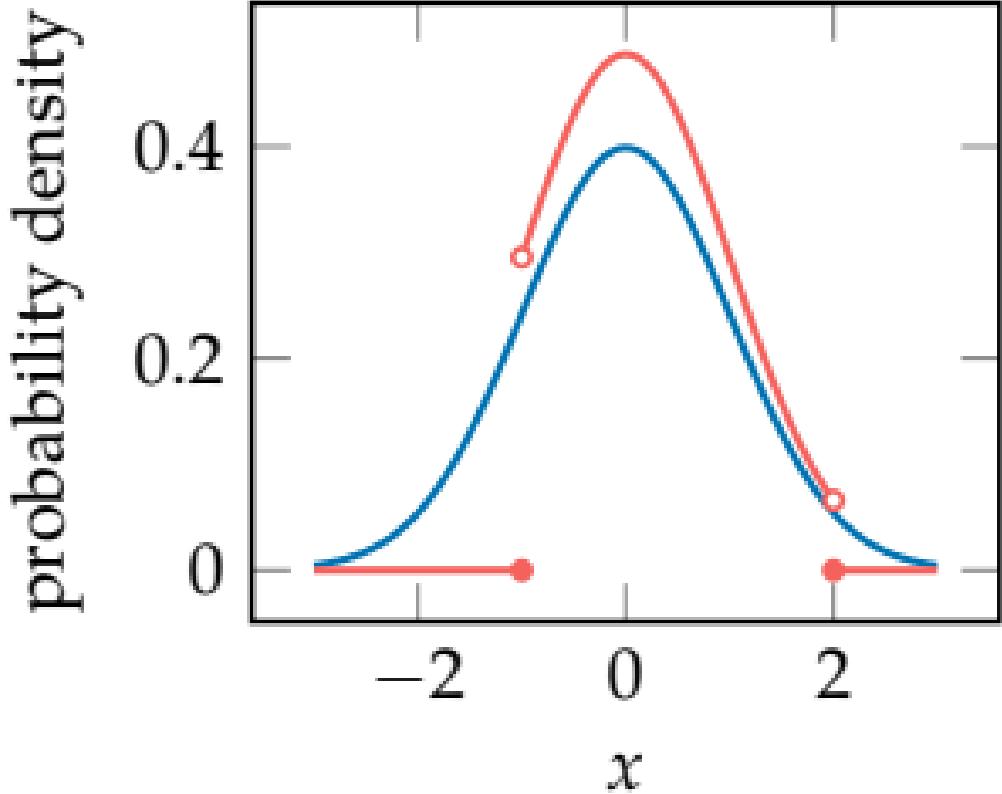
0.5

1

α

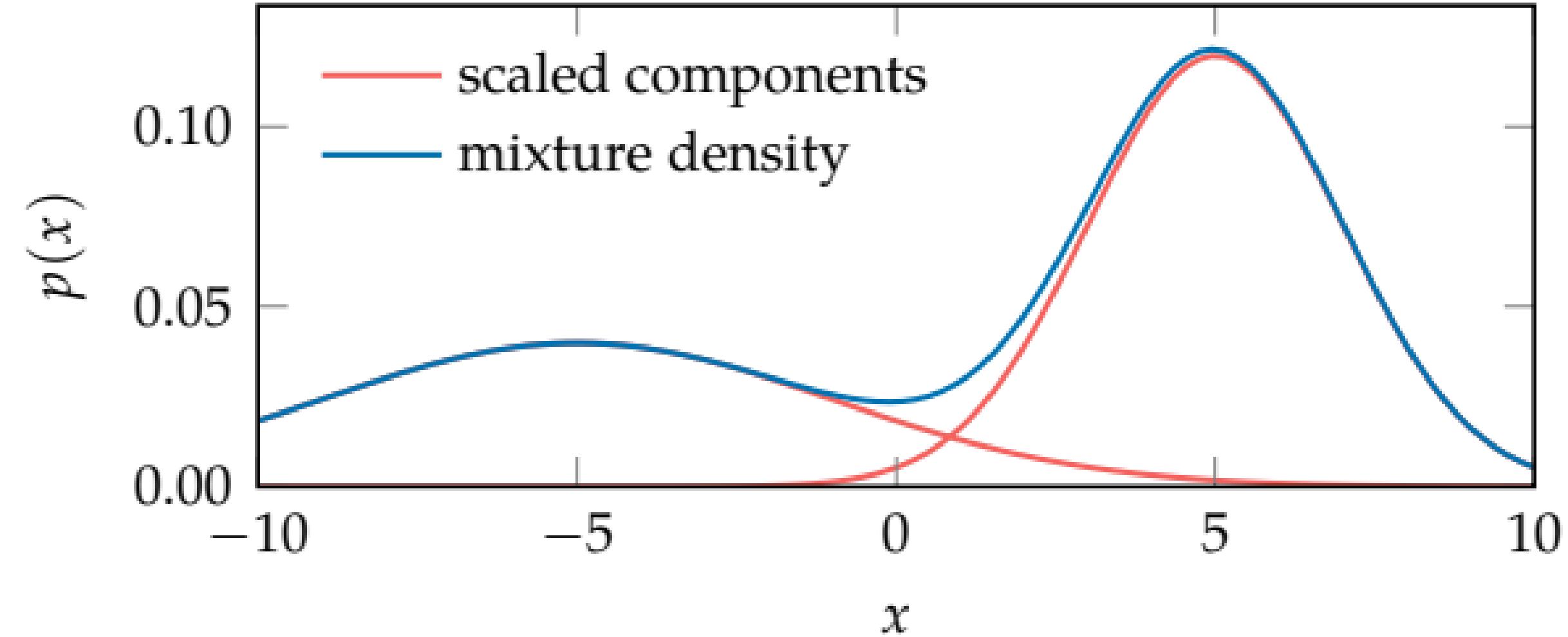


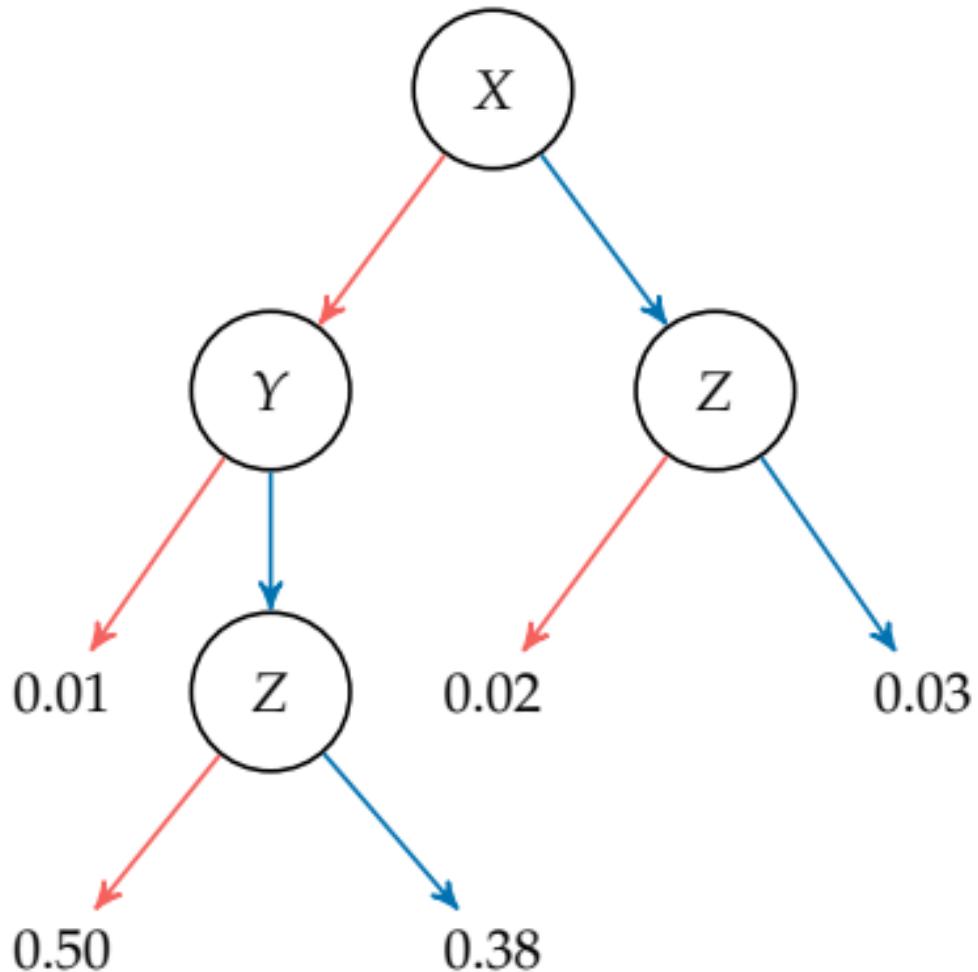


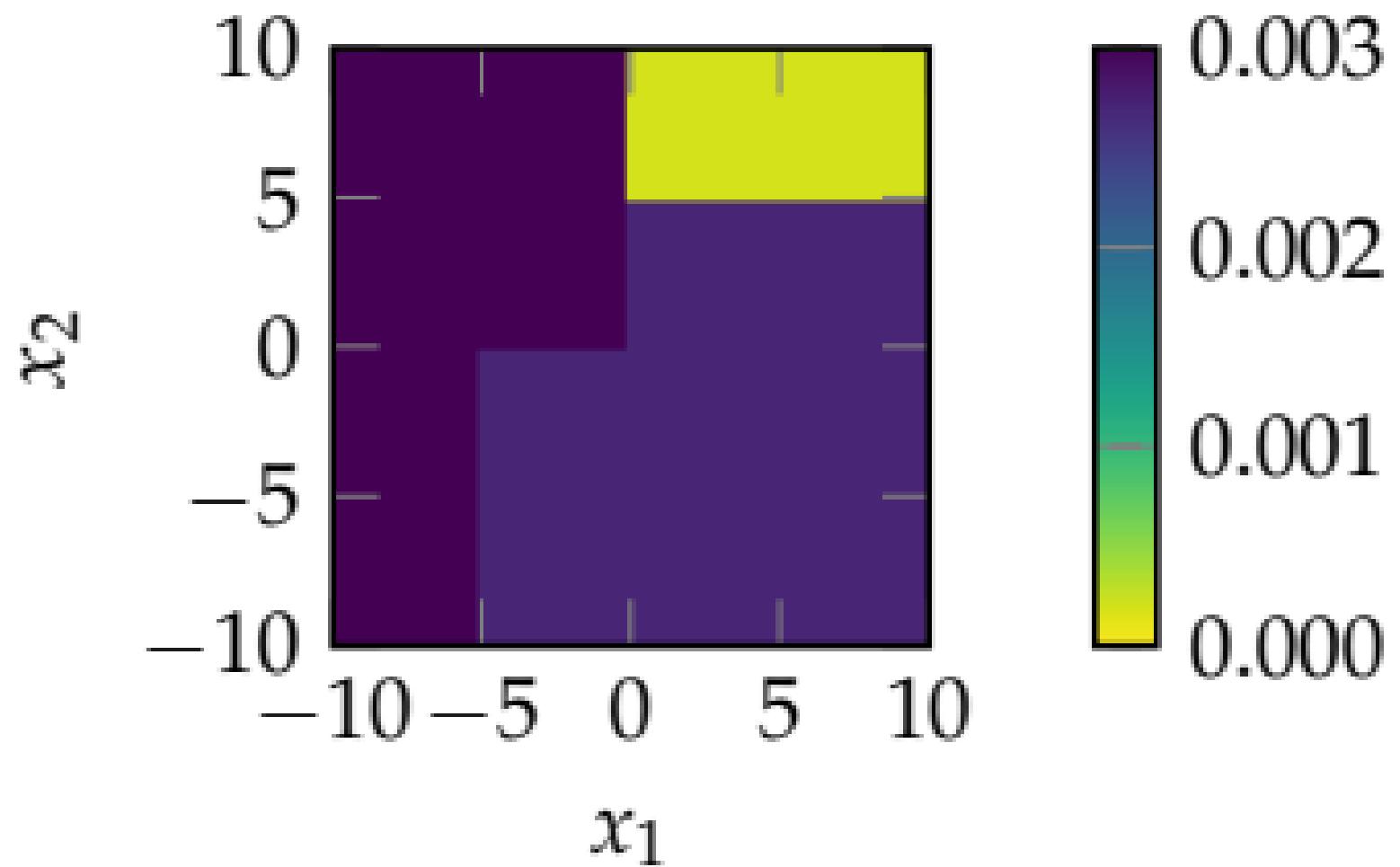


— full

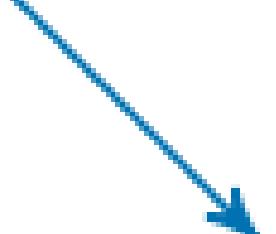
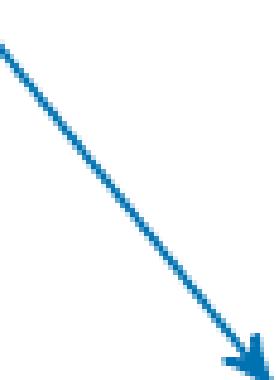
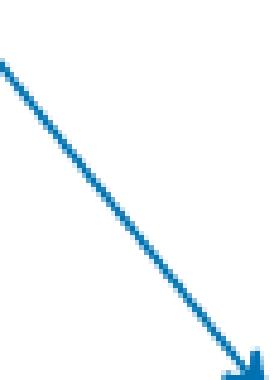
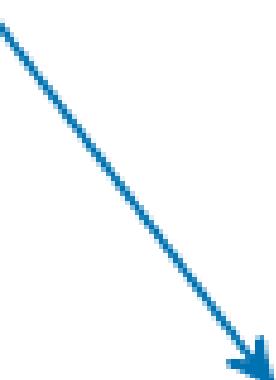
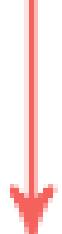
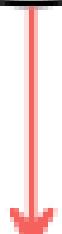
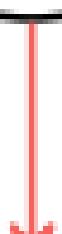
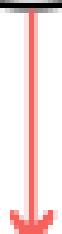
— truncated







$x_1 < -5$



0.003

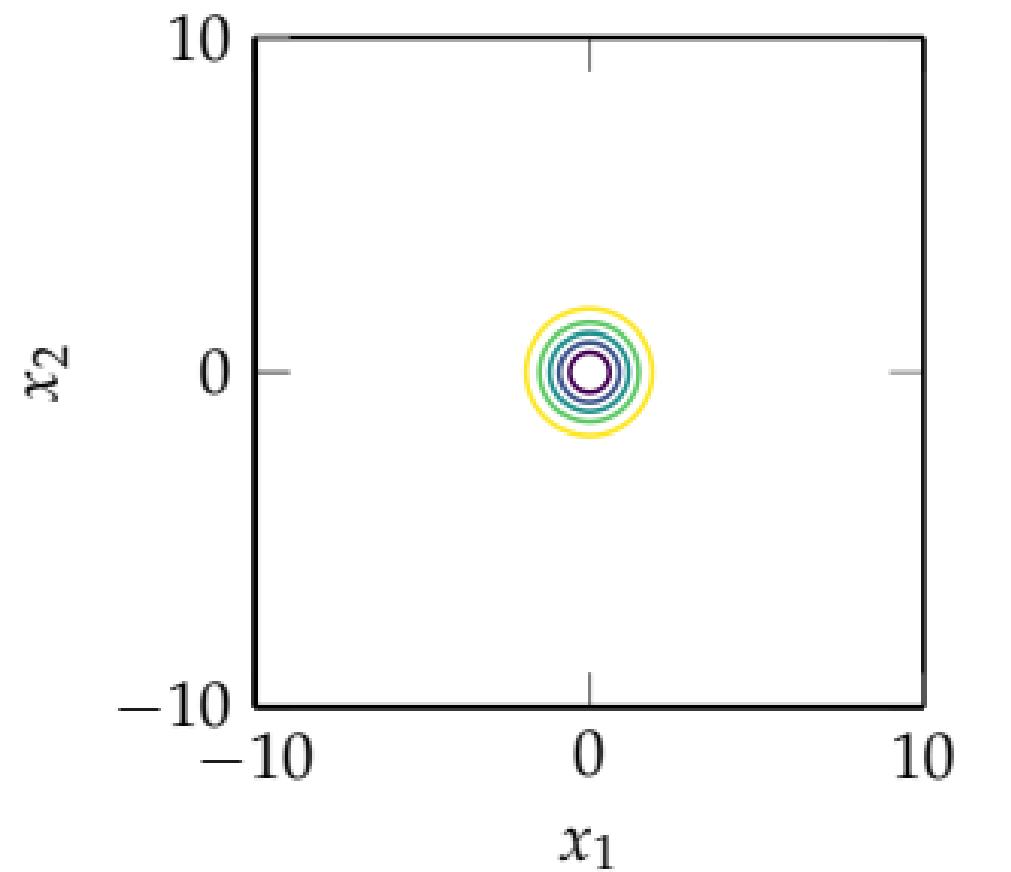
0.0027

0.003

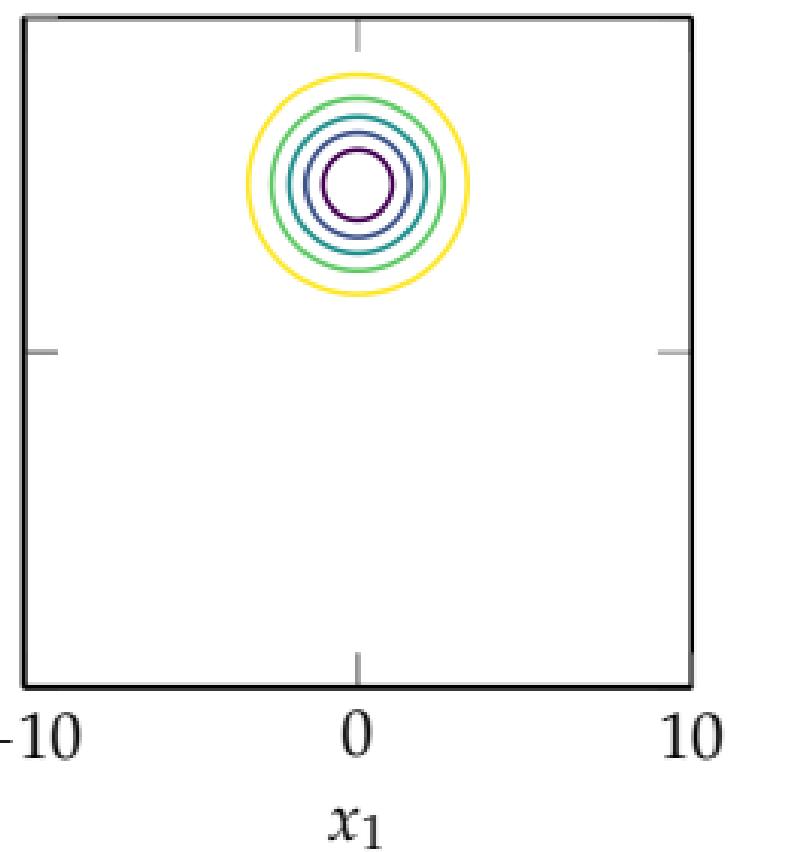
0.0027

0.0002

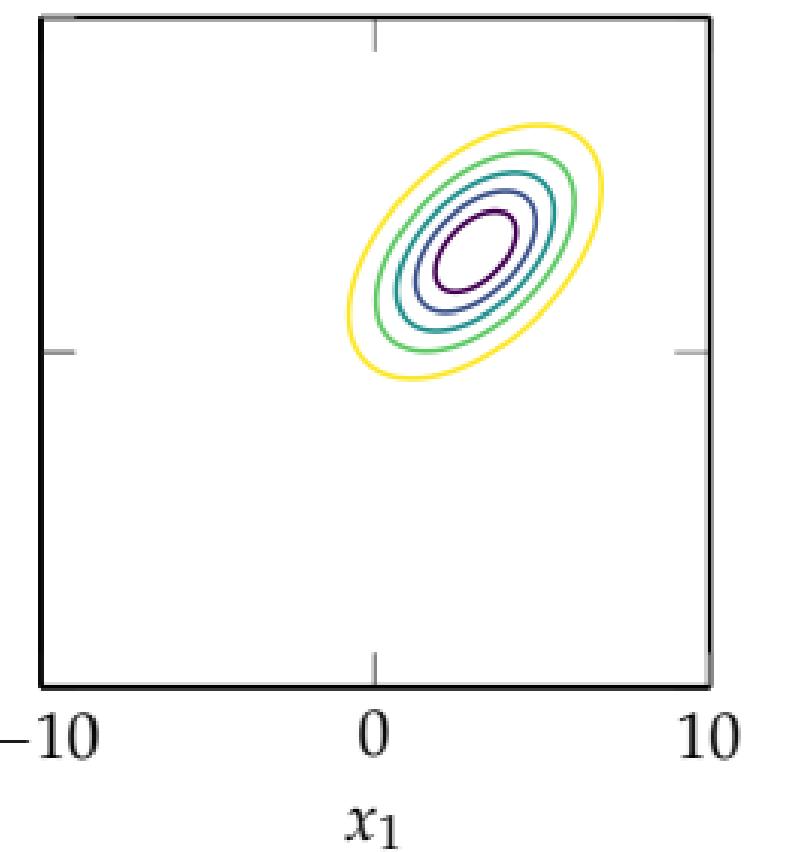
$$\mu = [0, 0], \Sigma = [1 \ 0; \ 0 \ 1]$$



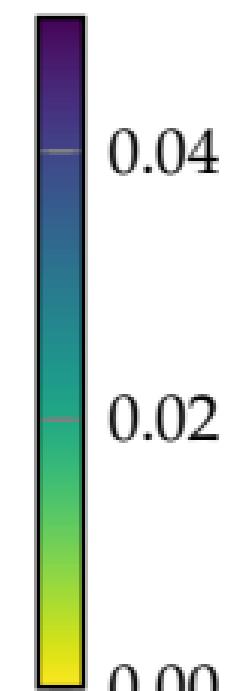
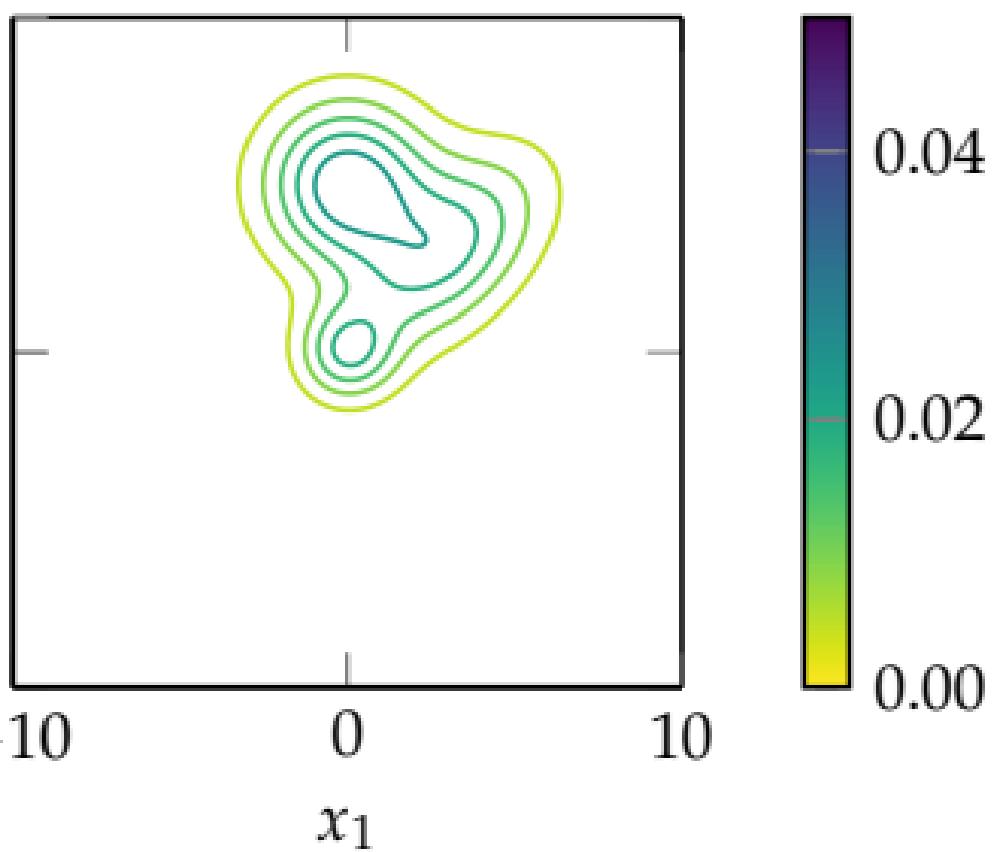
$$\mu = [0, 5], \Sigma = [3 \ 0; \ 0 \ 3]$$

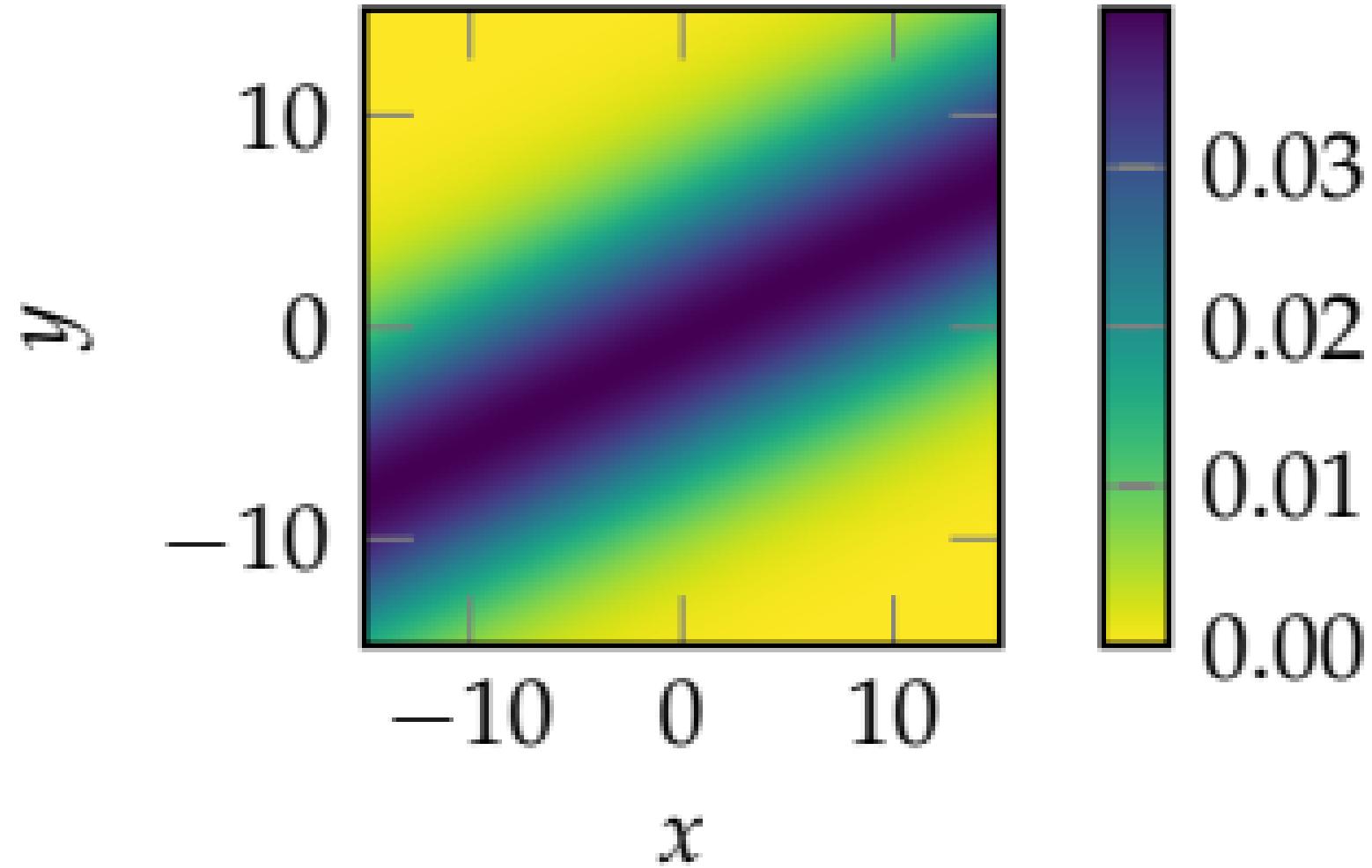


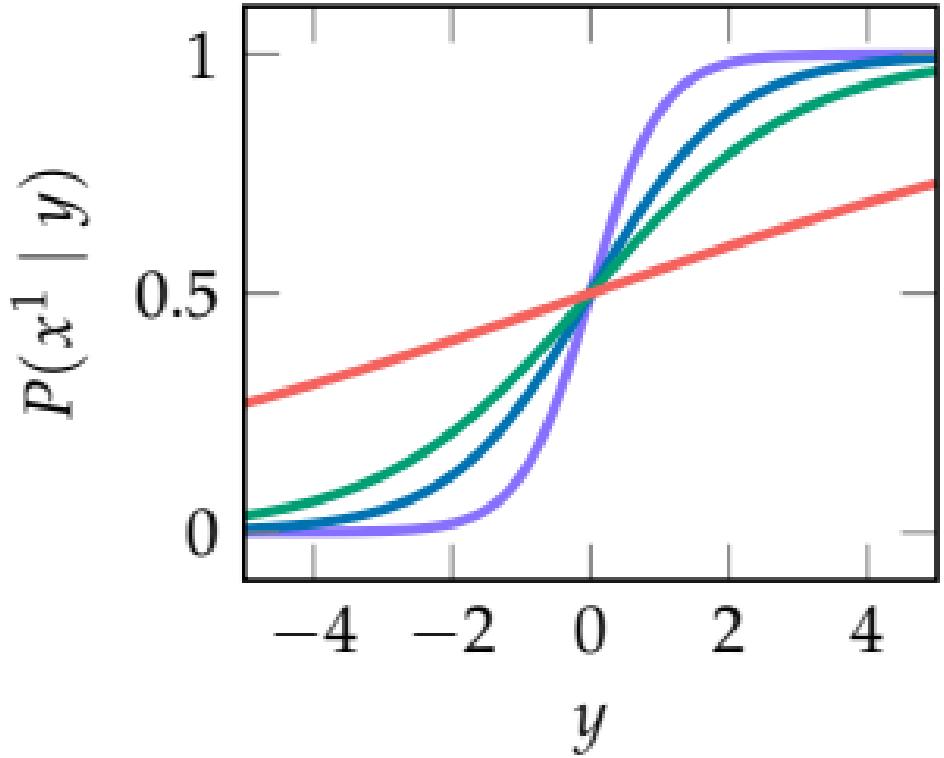
$$\mu = [3, 3], \Sigma = [4 \ 2; \ 2 \ 4]$$



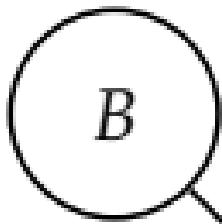
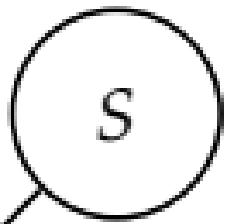
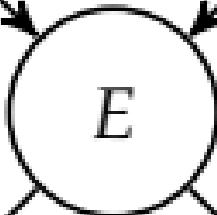
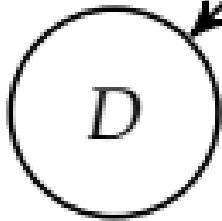
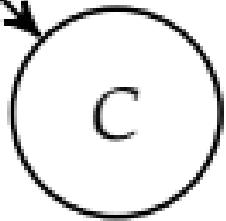
mixture

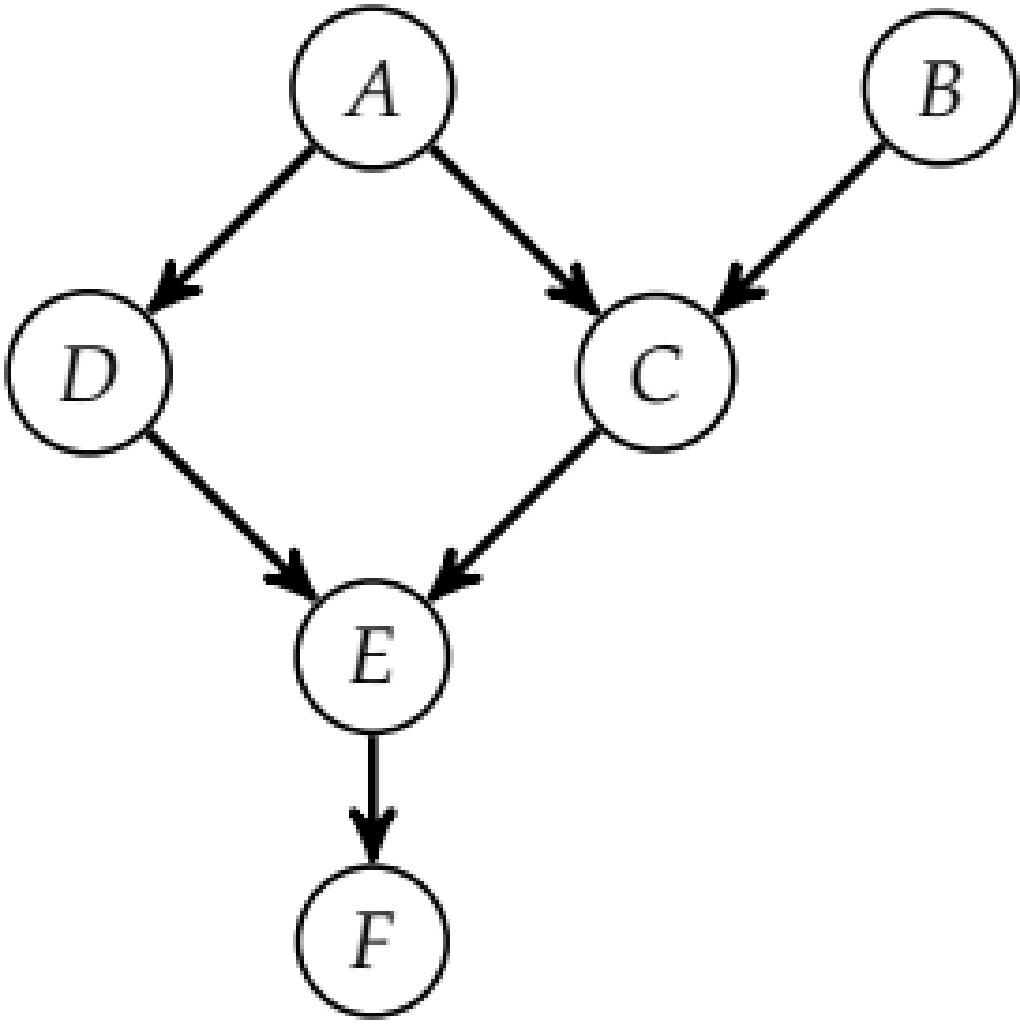


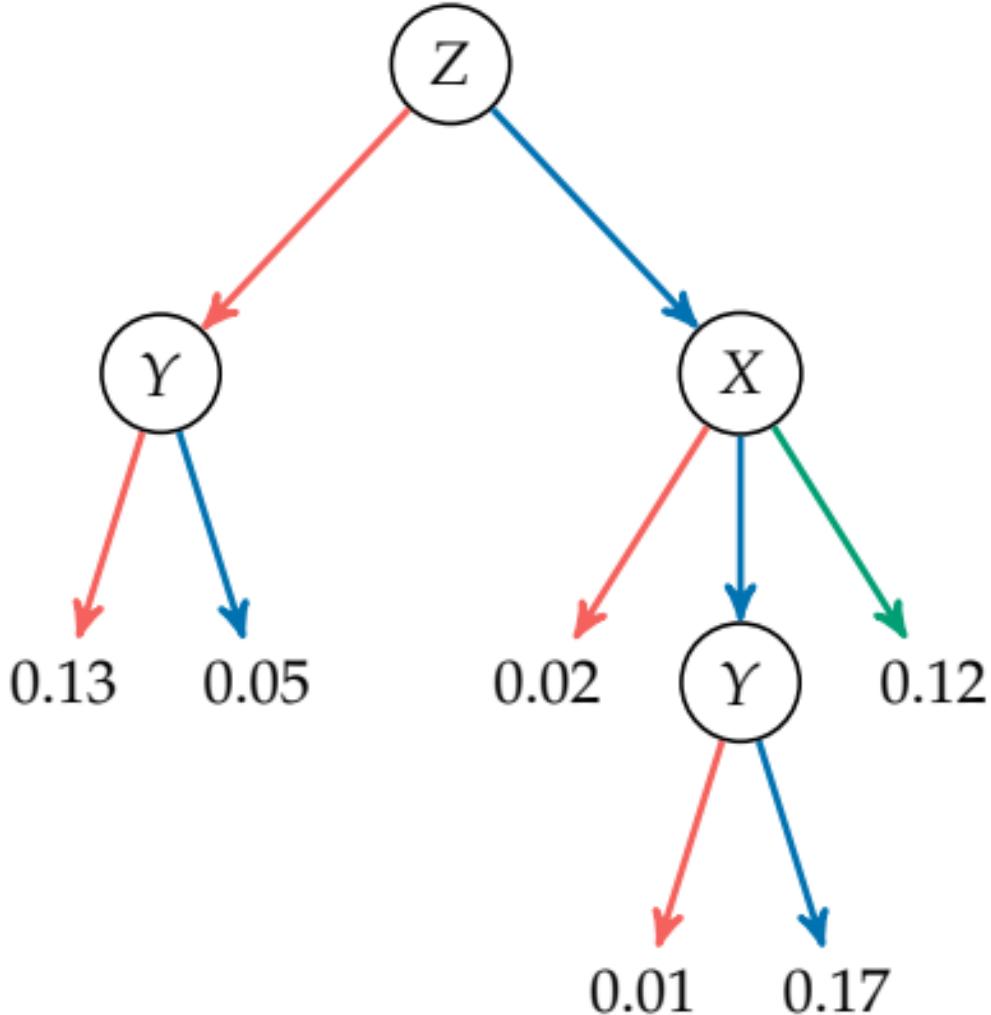


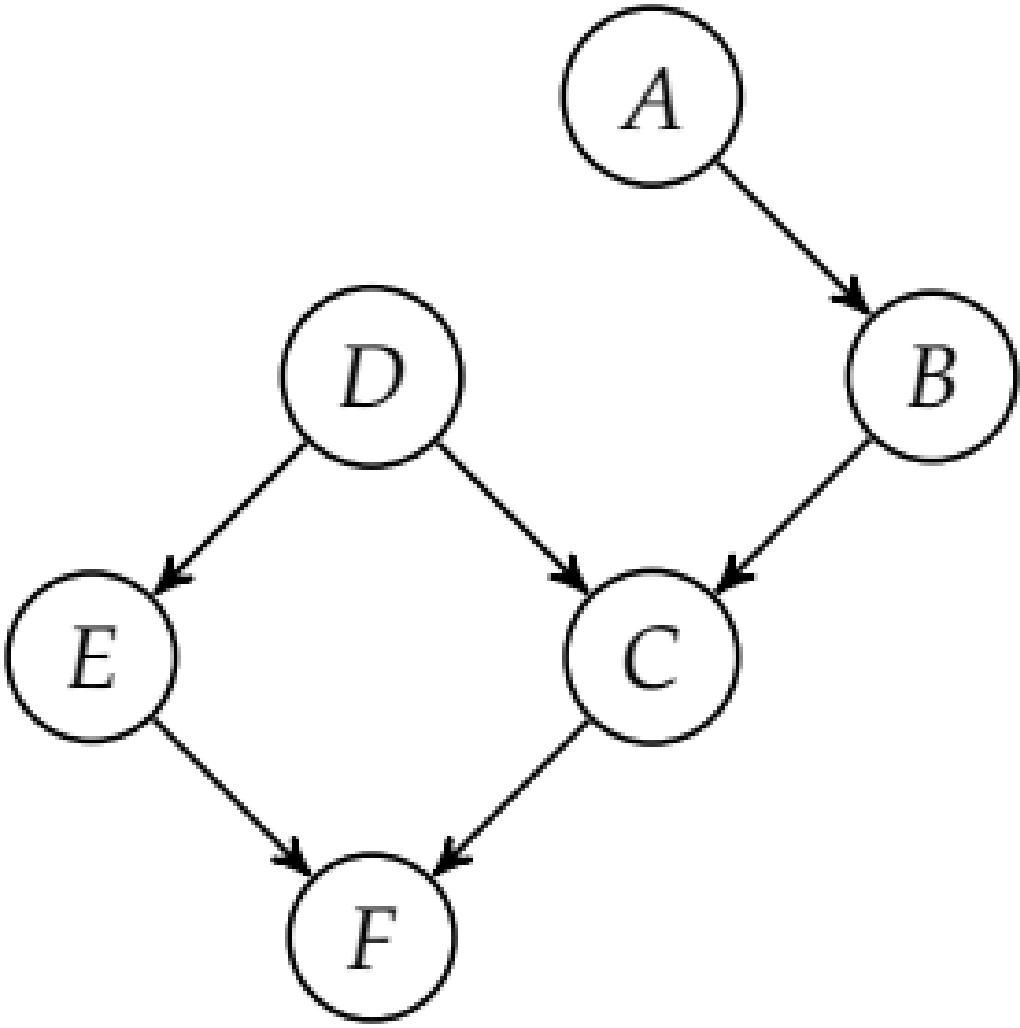


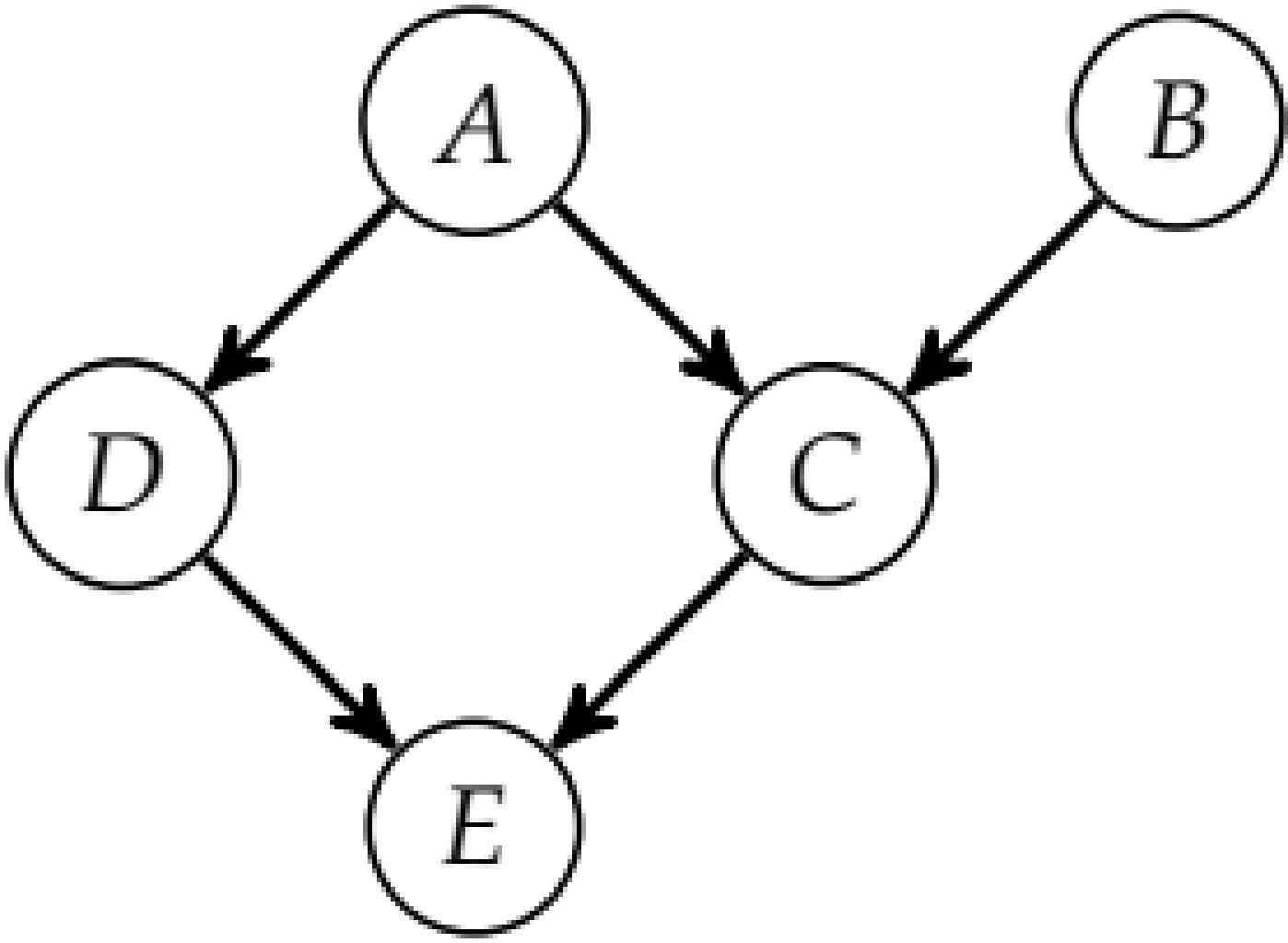
— $\theta_2 = 1$ — $\theta_2 = 2$
— $\theta_2 = 3$ — $\theta_2 = 10$

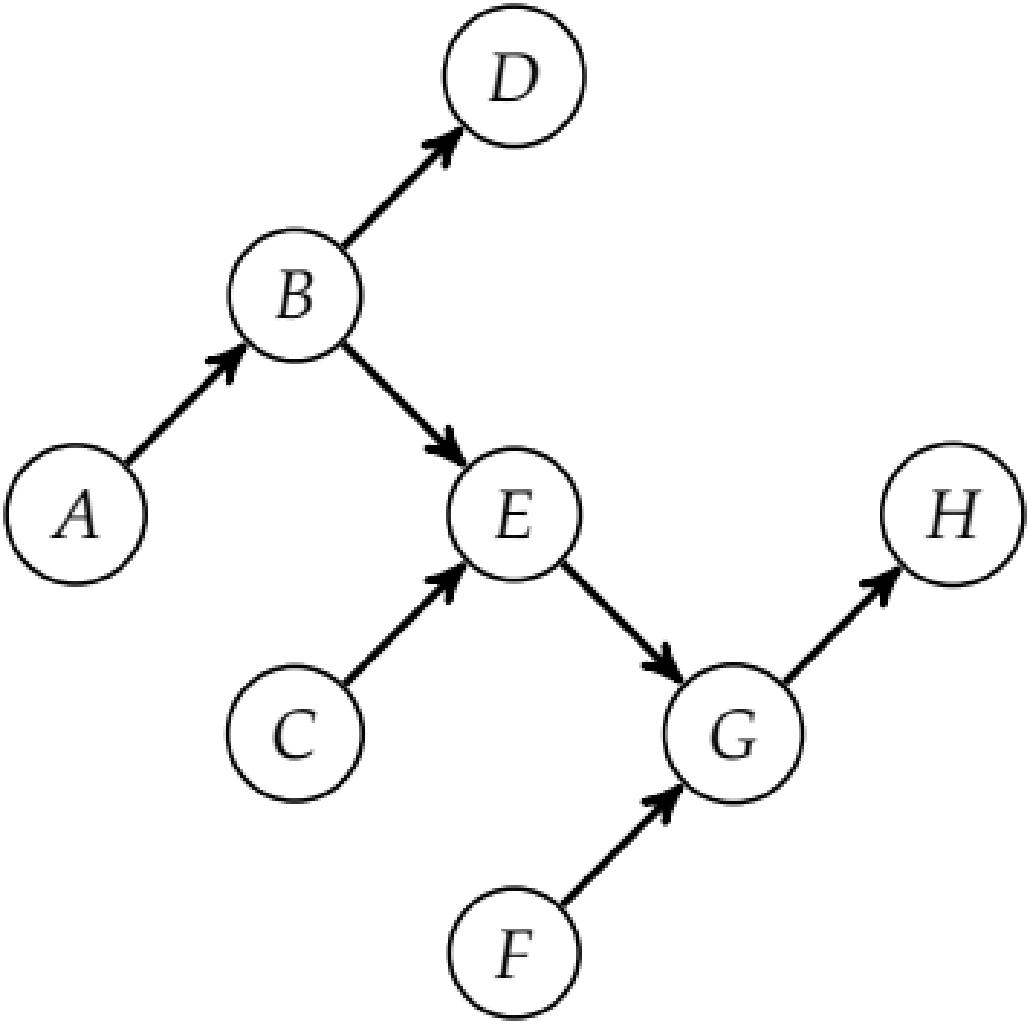
$P(B)$  $P(S)$  E  $P(E | B, S)$ D  $P(D | E)$ C  $P(C | E)$

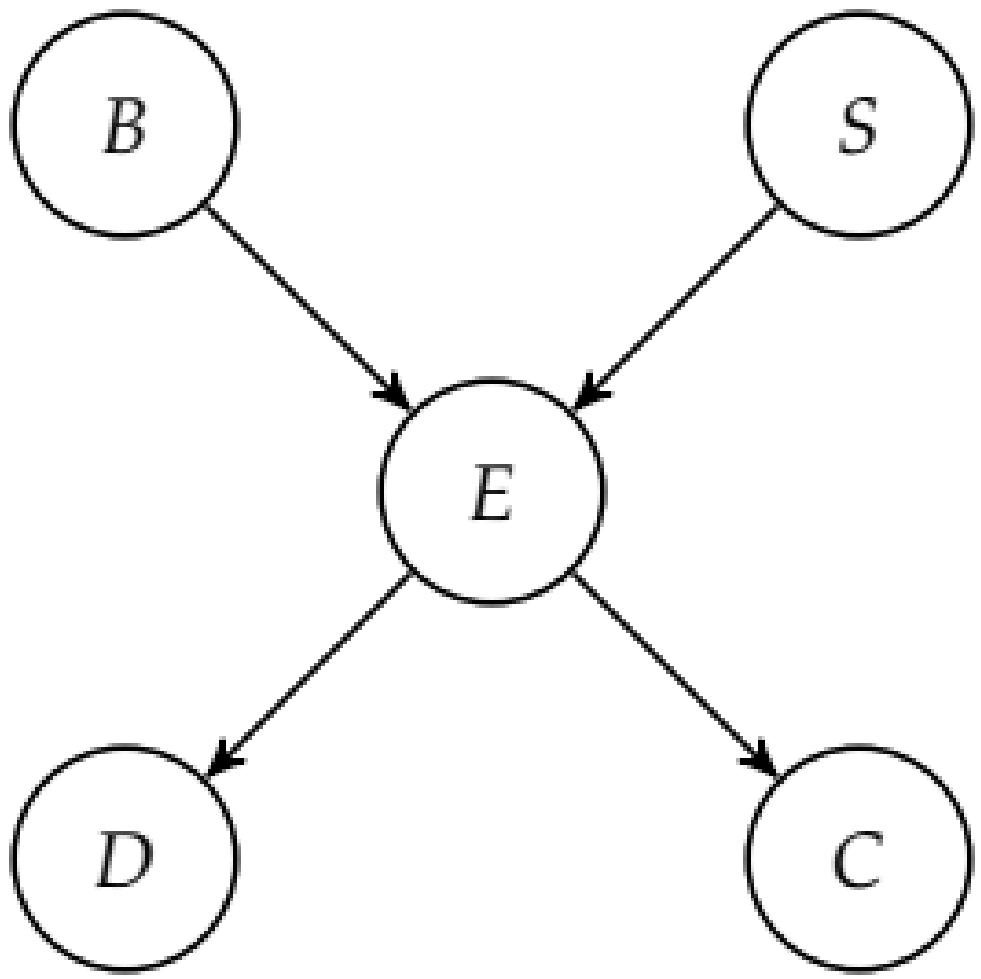




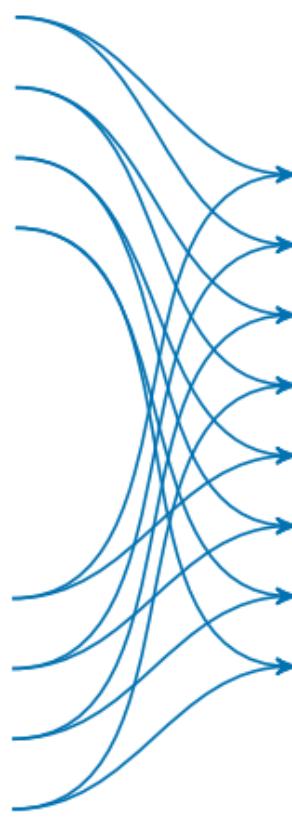






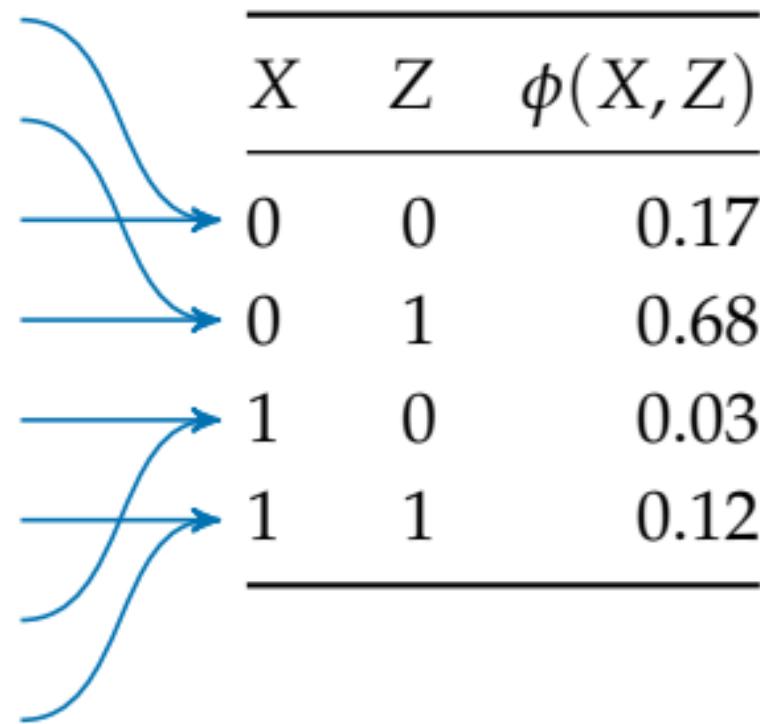


X	Y	$\phi_1(X, Y)$
0	0	0.3
0	1	0.4
1	0	0.2
1	1	0.1



X	Y	Z	$\phi_3(X, Y, Z)$
0	0	0	0.06
0	0	1	0.00
0	1	0	0.12
0	1	1	0.20
1	0	0	0.04
1	0	1	0.00
1	1	0	0.03
1	1	1	0.05

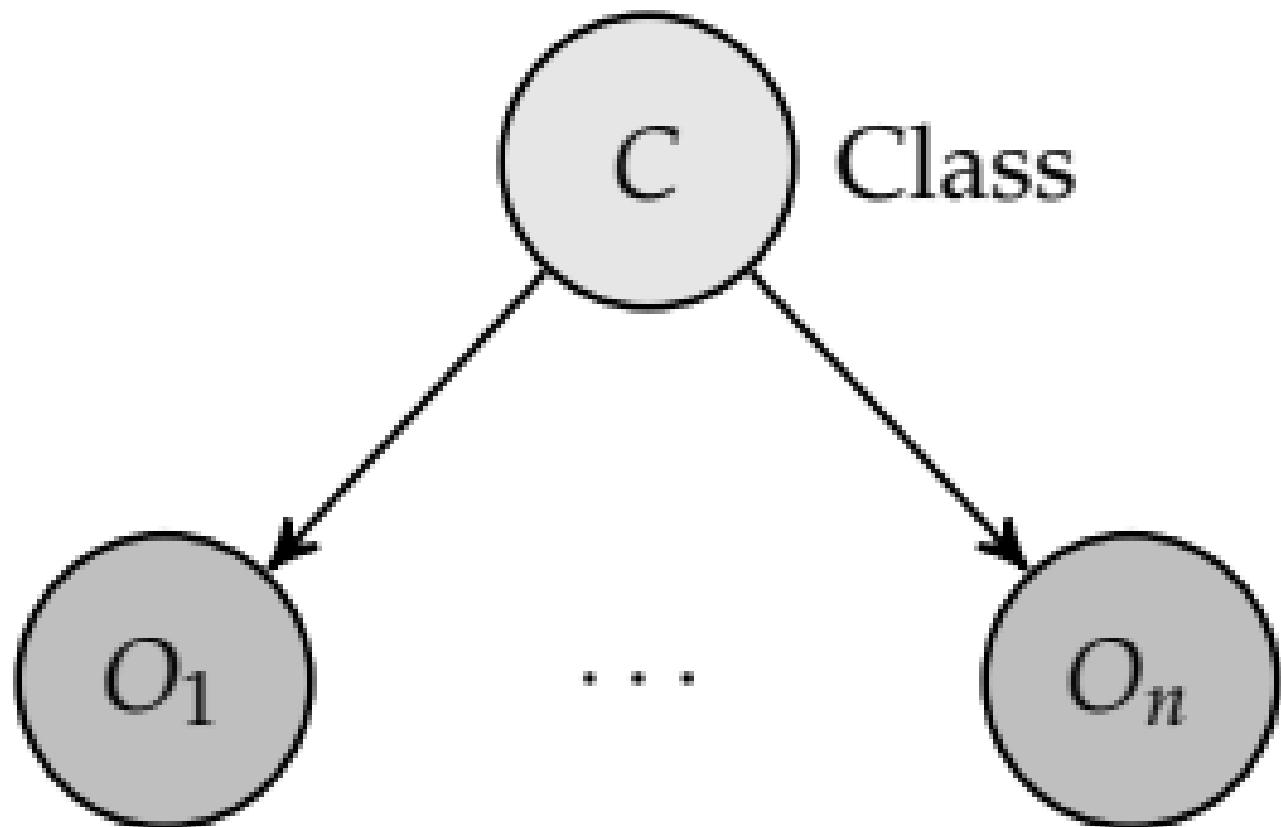
X	Y	Z	$\phi(X, Y, Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07



X	Y	Z	$\phi(X, Y, Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

$Y = 1$

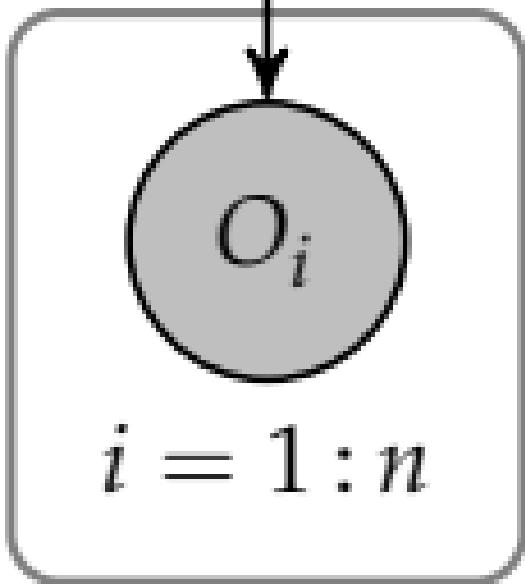
X	Z	$\phi(X, Z)$
0	0	0.09
0	1	0.37
1	0	0.02
1	1	0.07



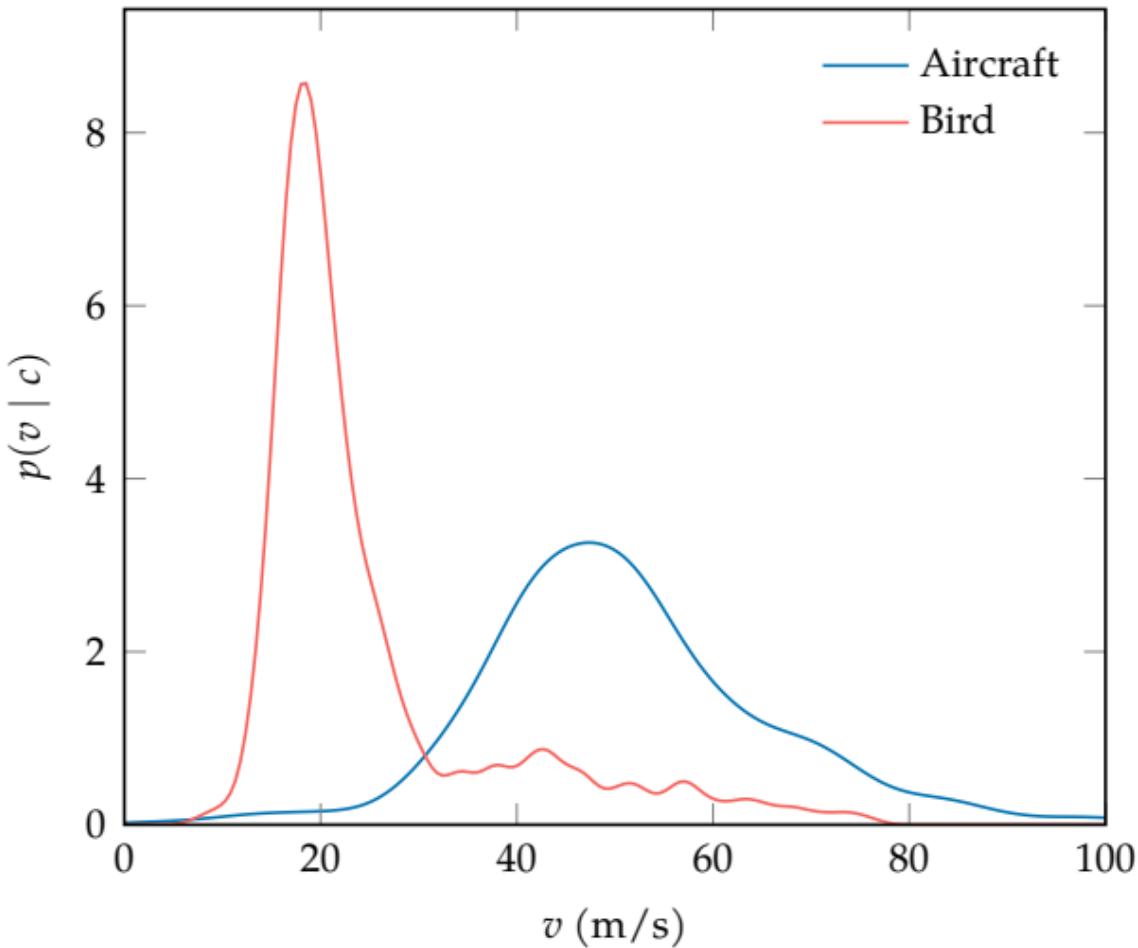
Observed features

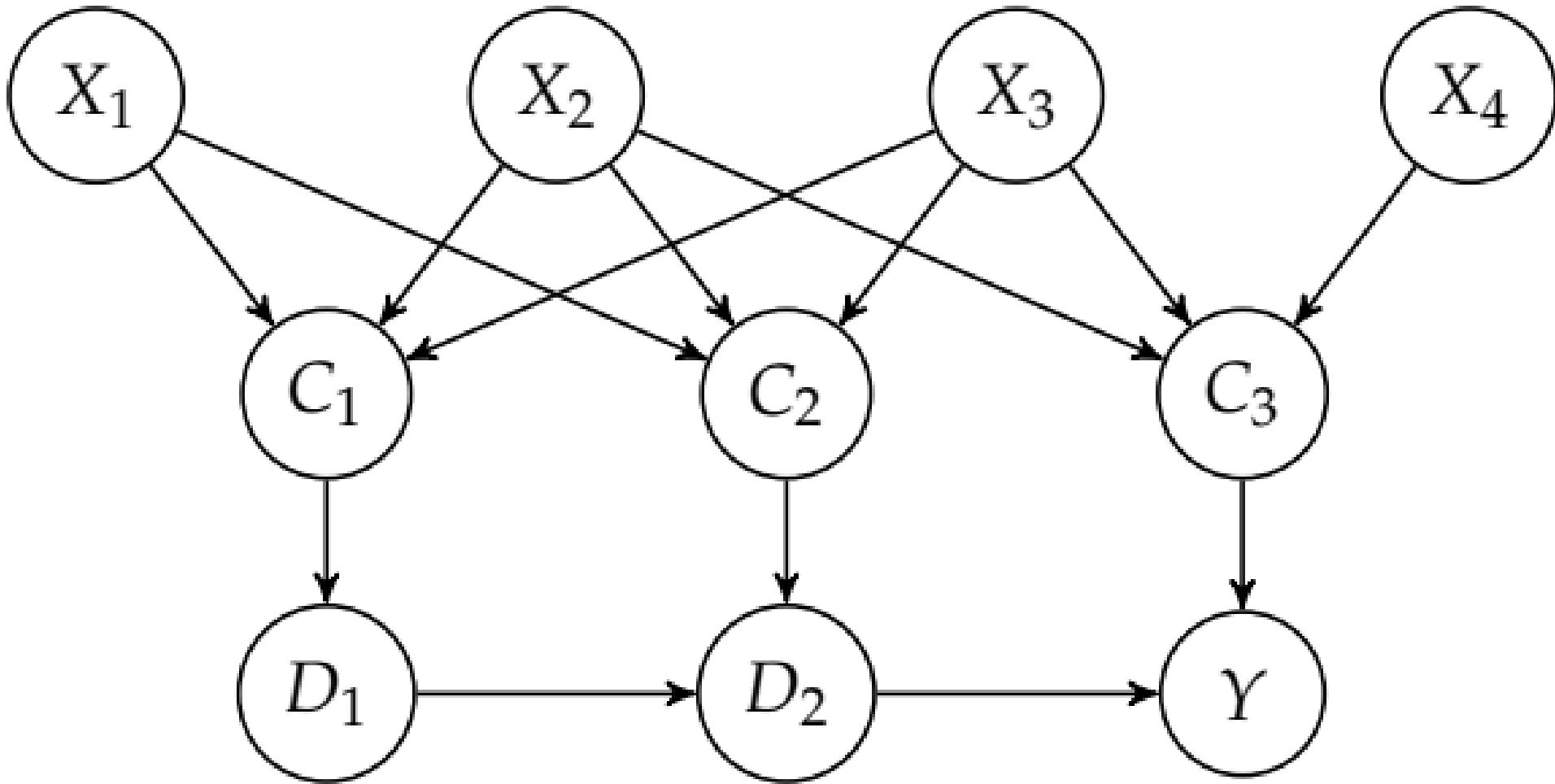


Class

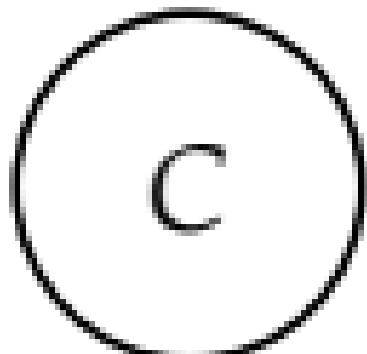


Observed features

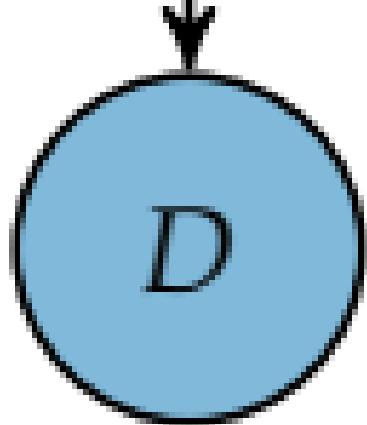
$\times 10^{-2}$ 



<i>B</i>	<i>S</i>	<i>E</i>	<i>D</i>	<i>C</i>	
0	0	1	1	0	
0	0	0	0	0	
1	0	1	0	0	
1	0	1	1	1	←
0	0	0	0	0	
0	0	0	1	0	
0	0	0	0	1	
0	1	1	1	1	←
0	0	0	0	0	
0	0	0	1	0	

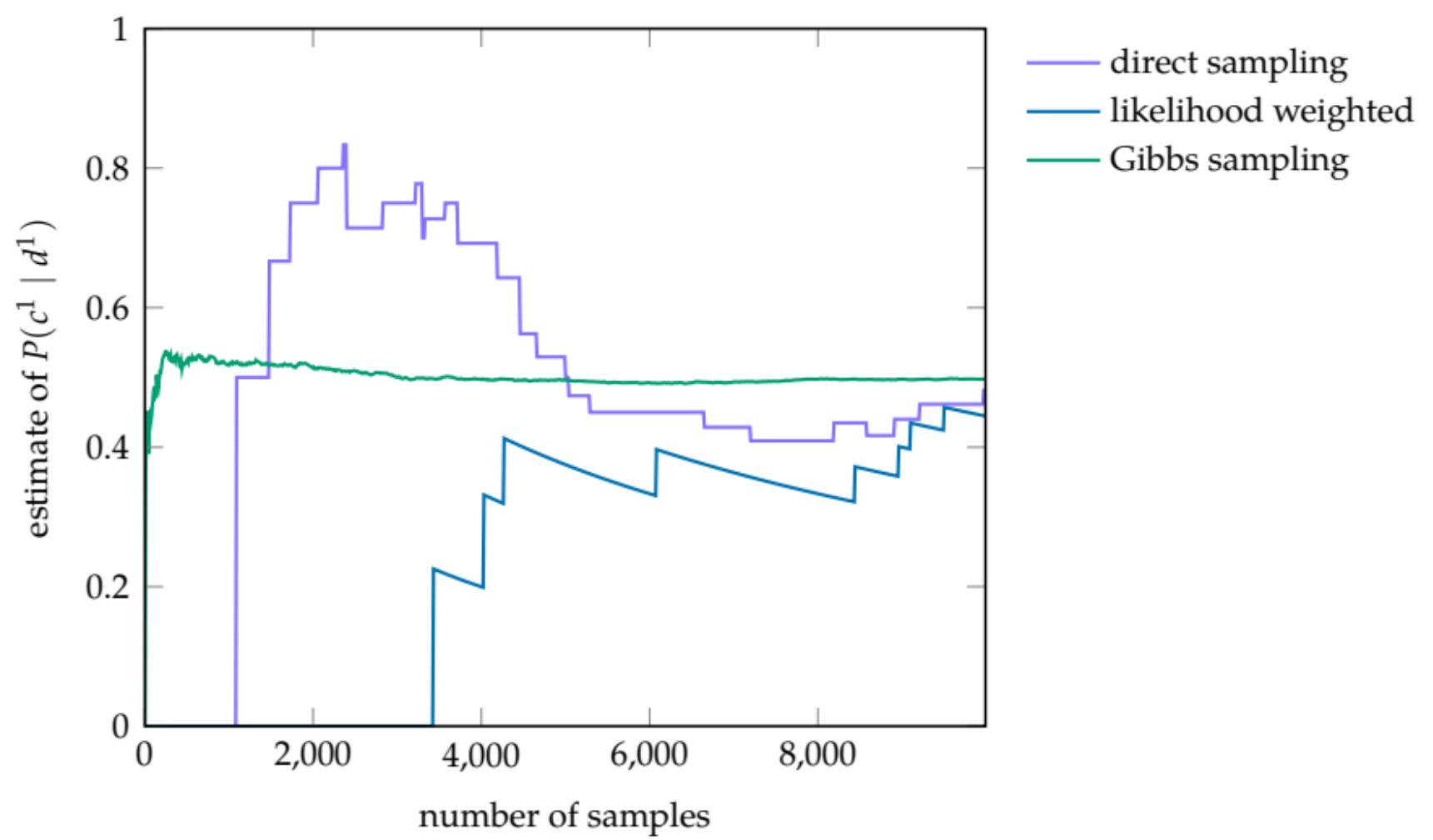


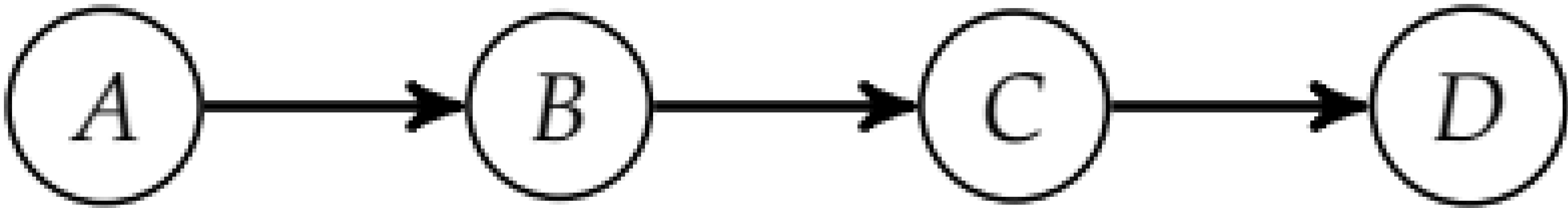
$$P(c^1) = 0.001$$

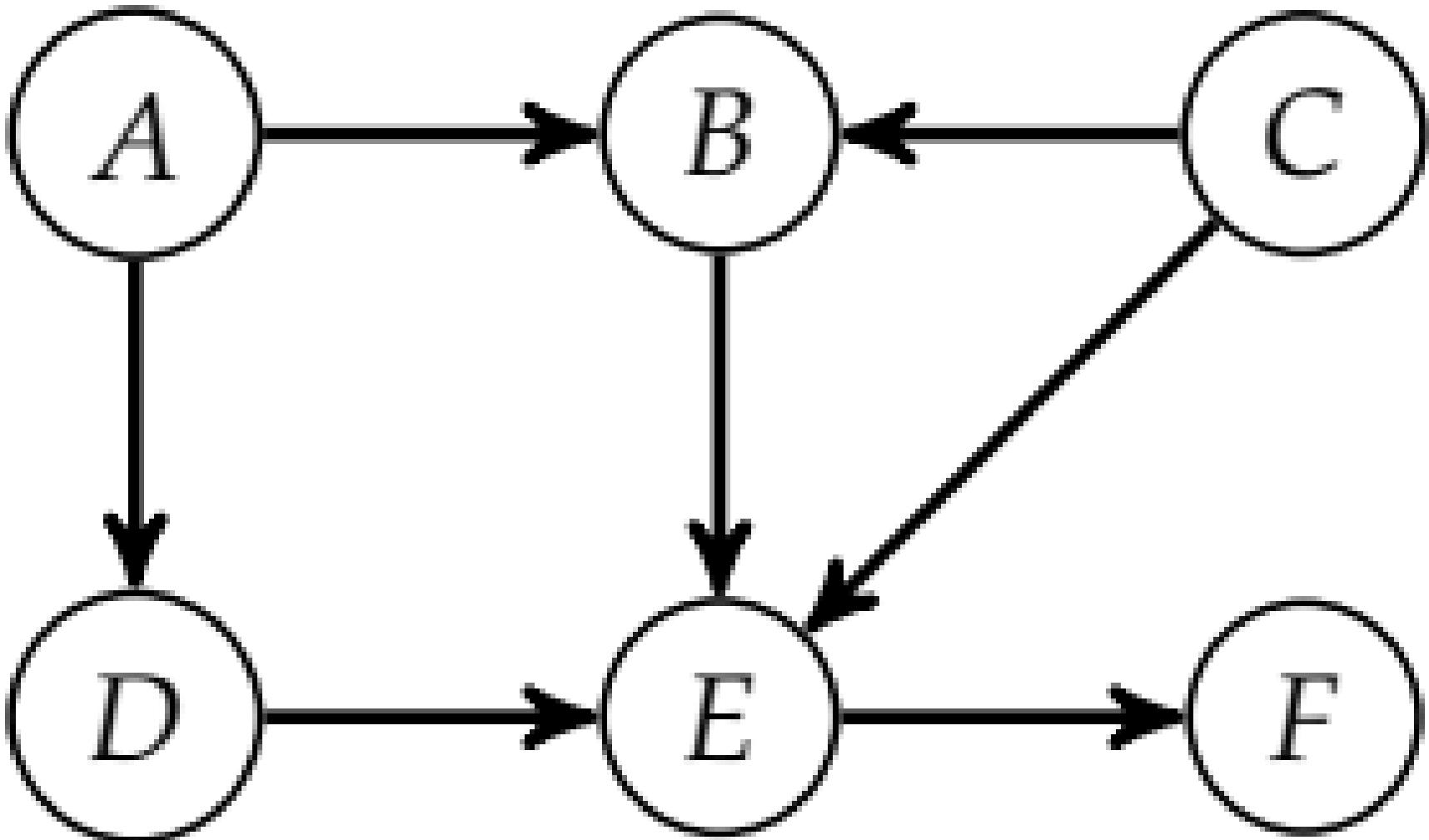


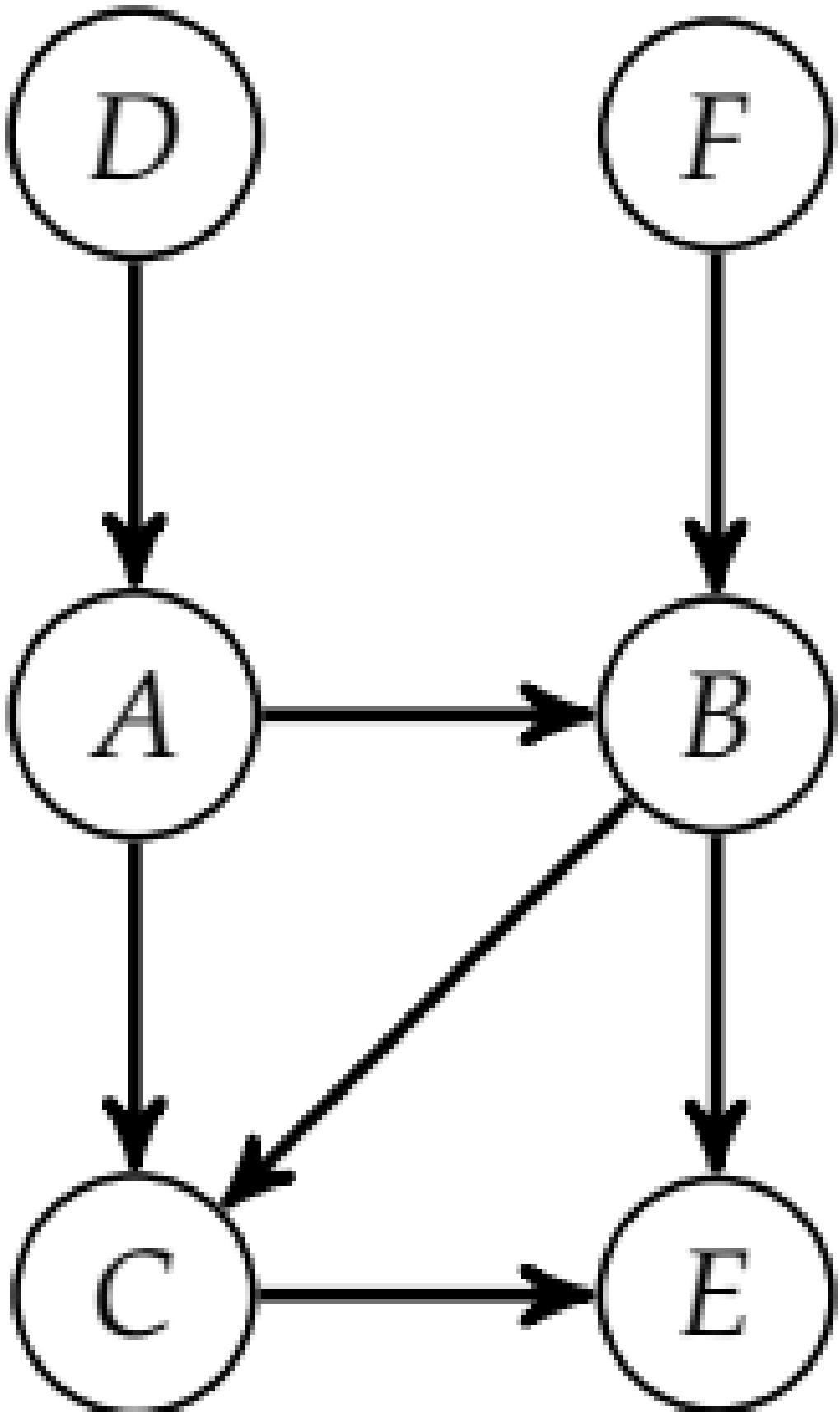
$$P(d^1 \mid c^0) = 0.001$$

$$P(d^1 \mid c^1) = 0.999$$

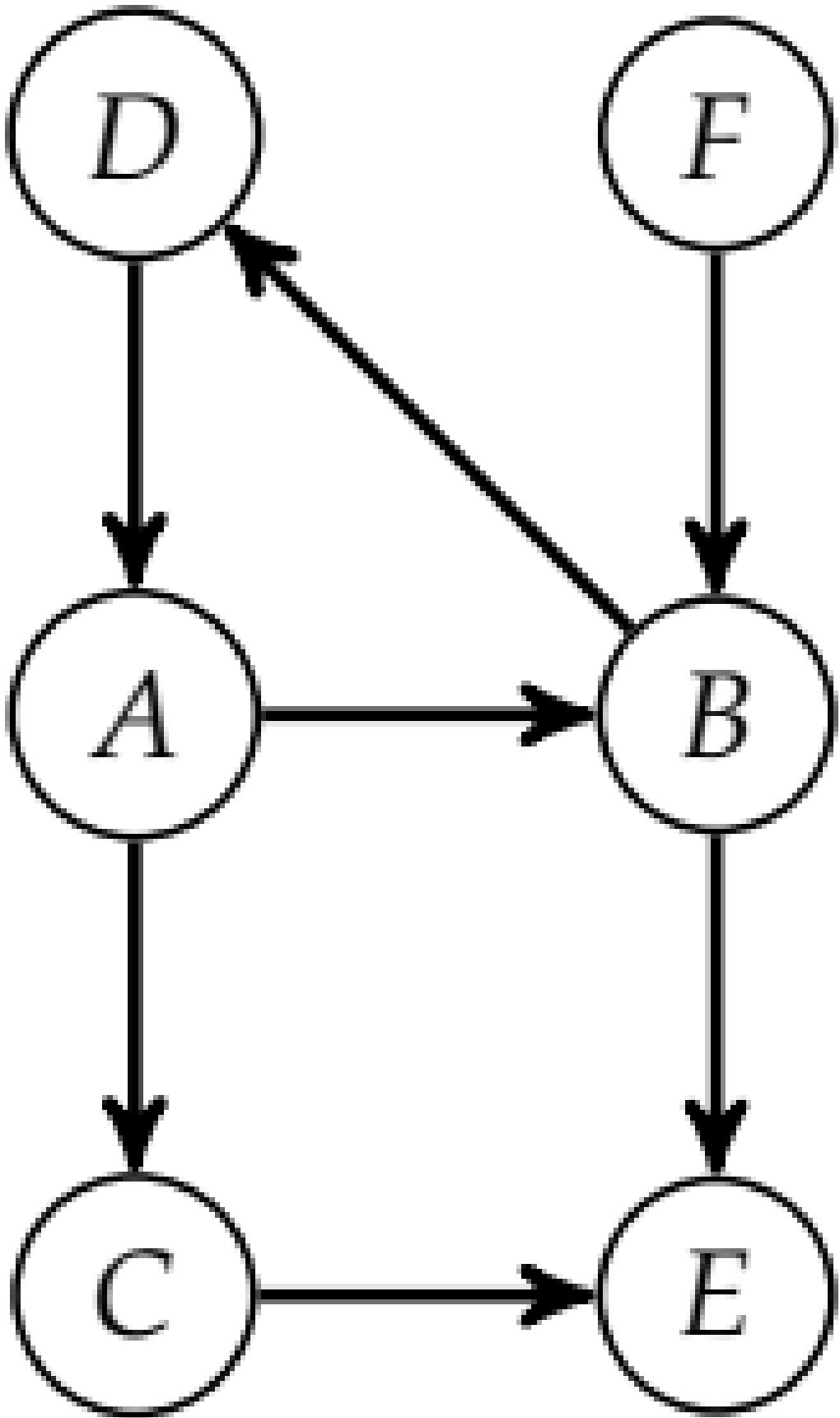




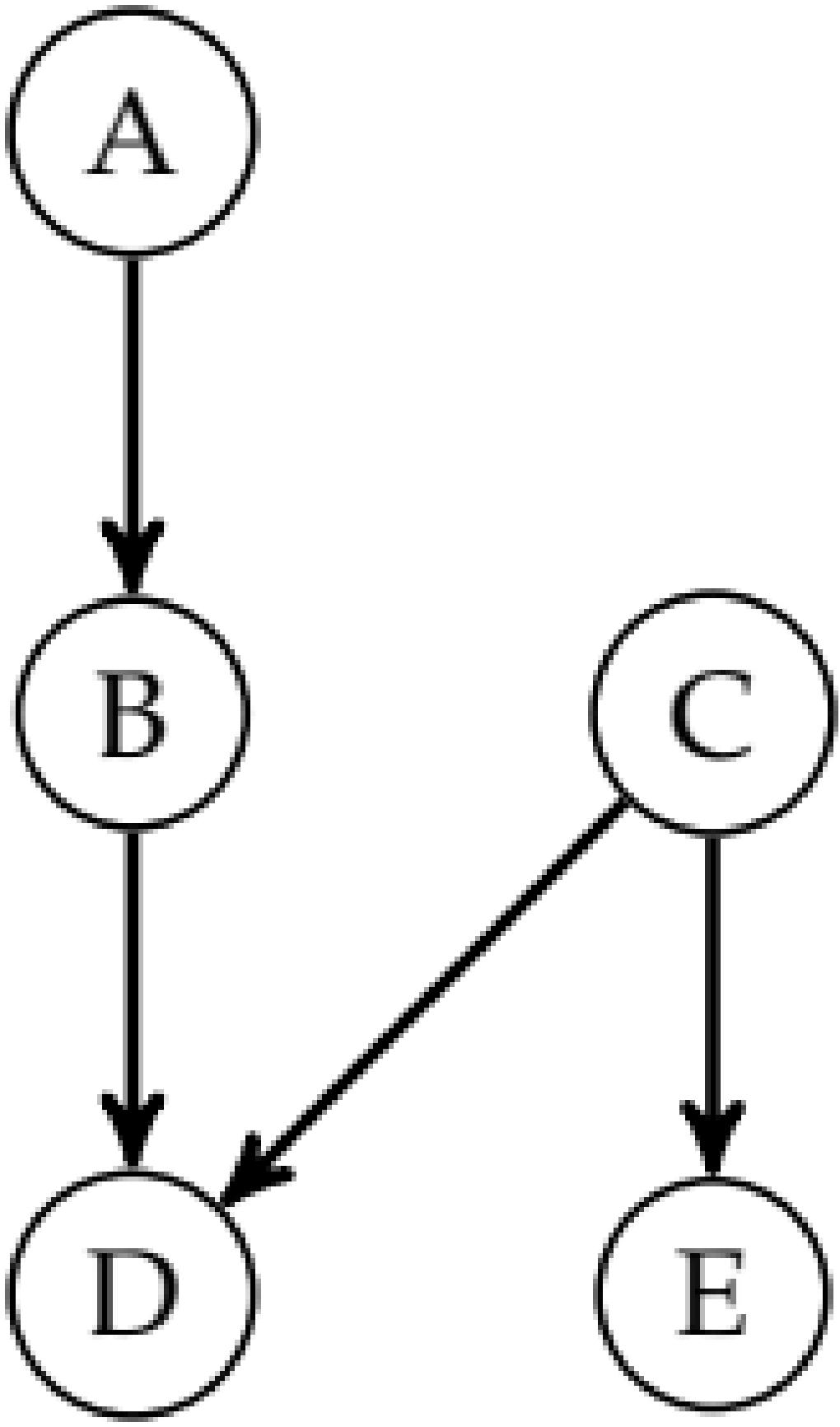


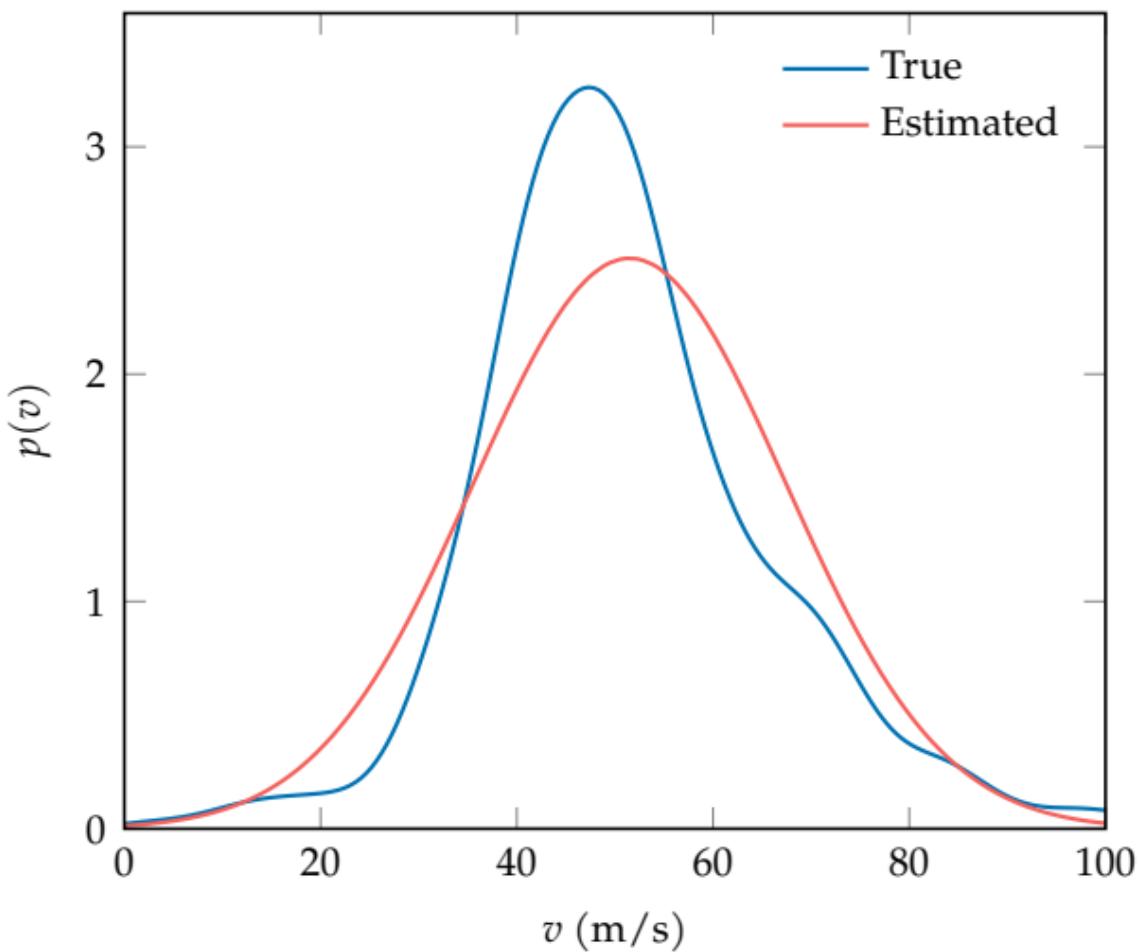


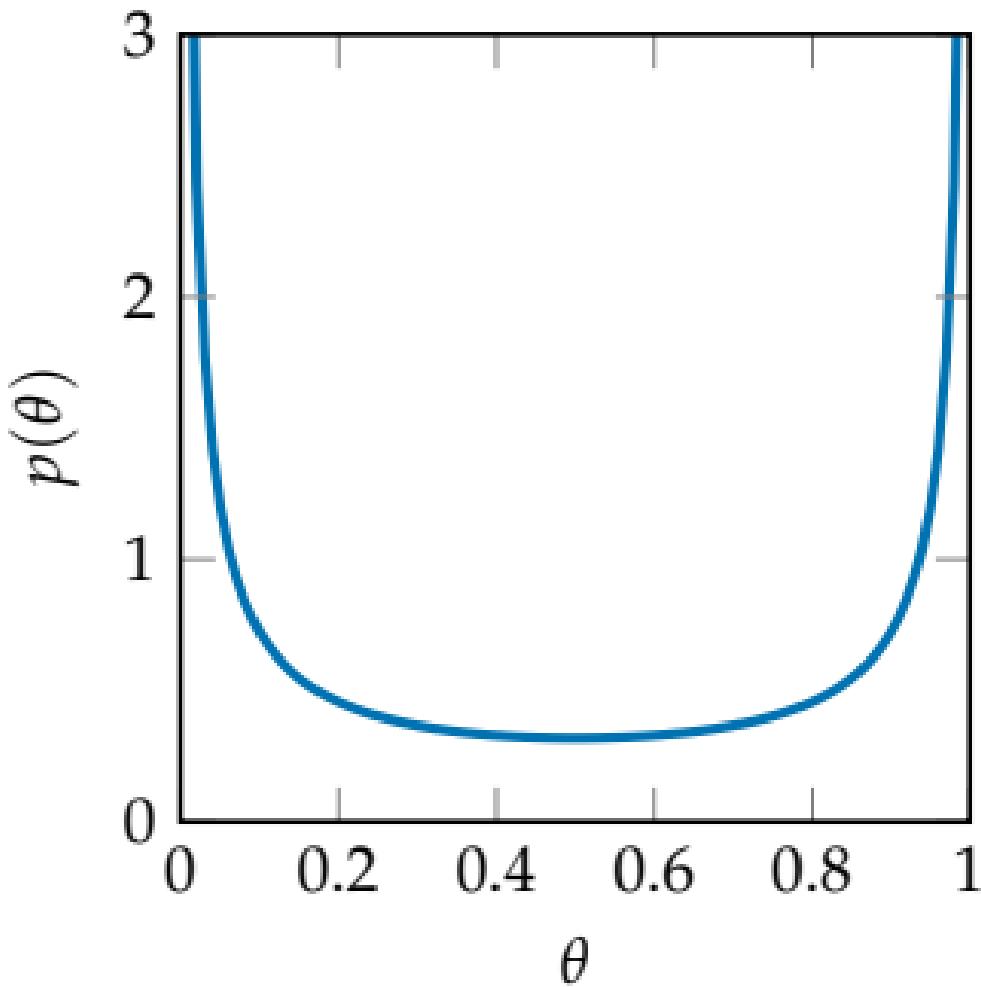
(1)

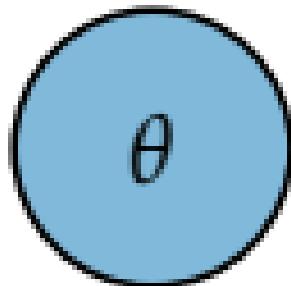


(2)

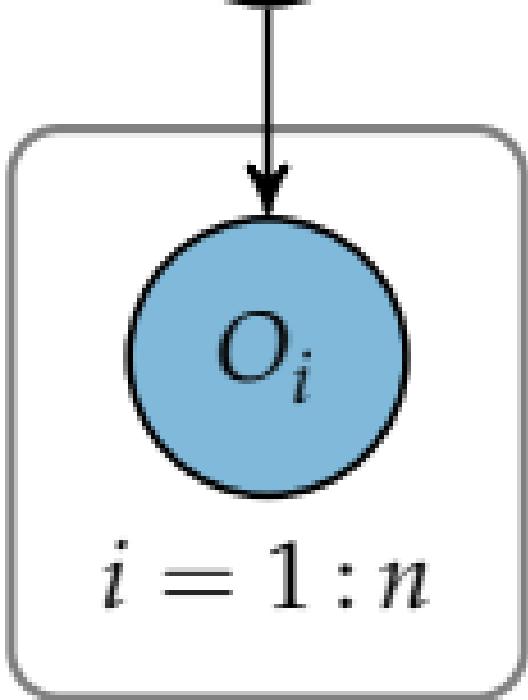


$\times 10^{-2}$ 



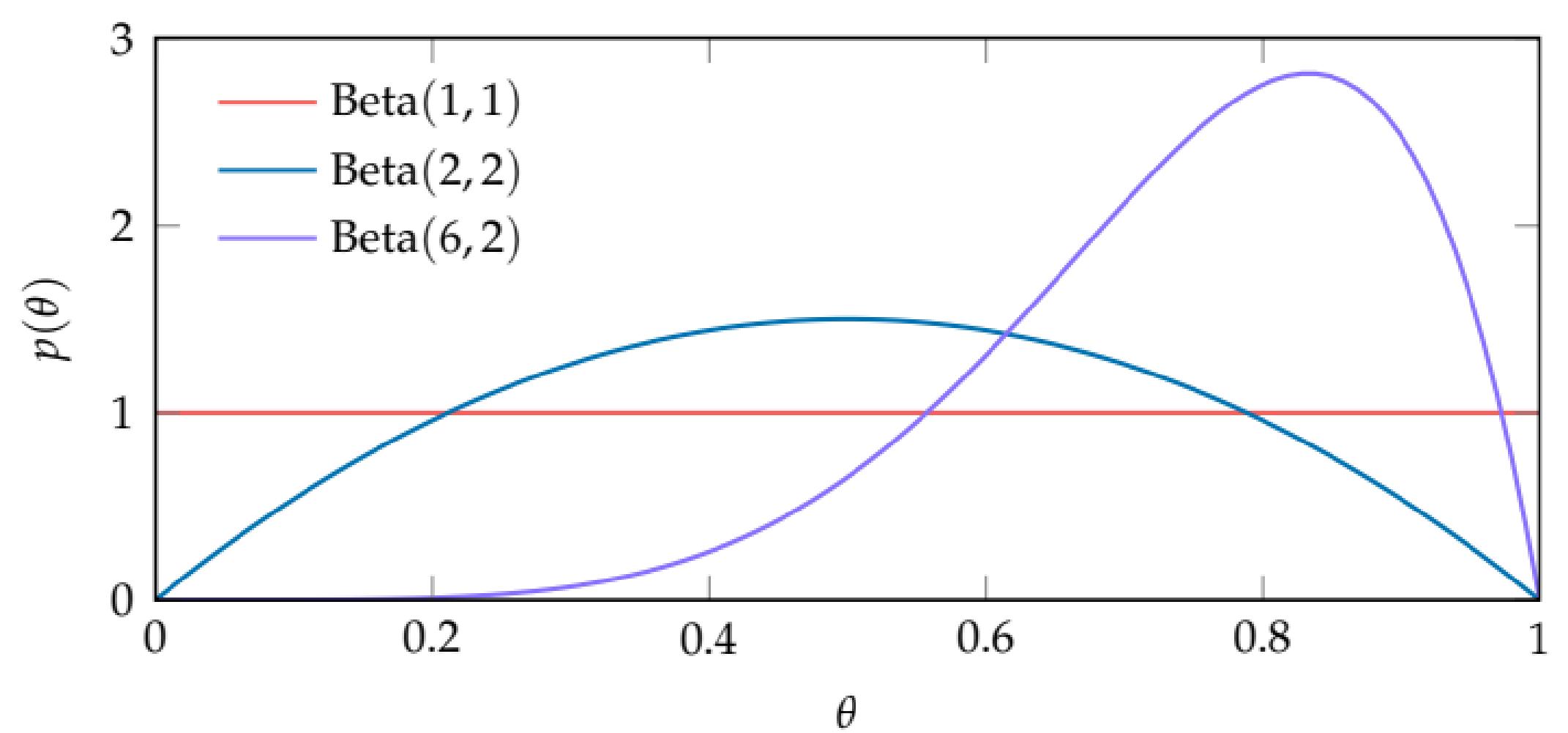


Parameter

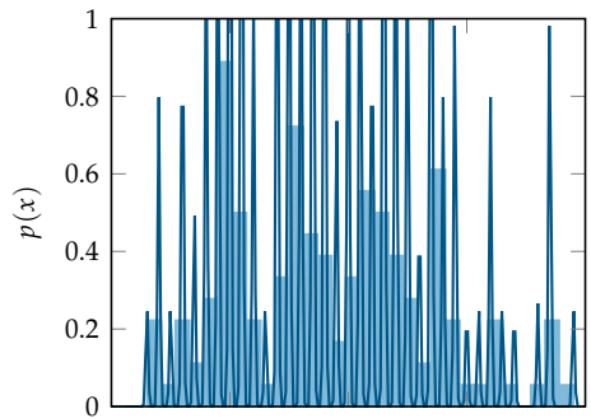


Observations

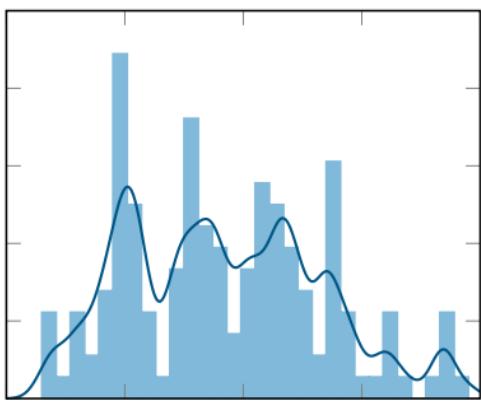
$$i = 1 : n$$



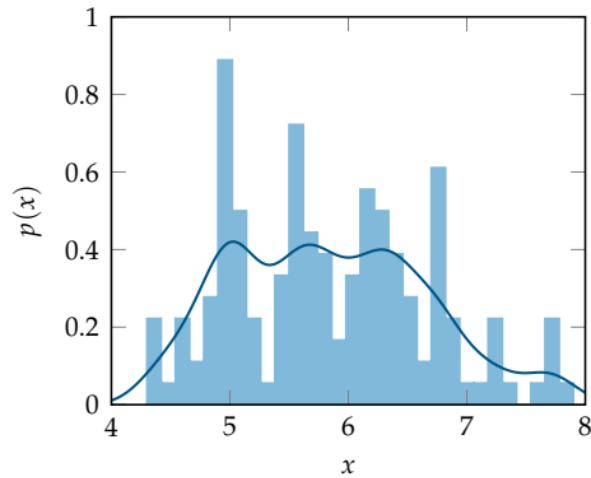
bandwidth = 0.01



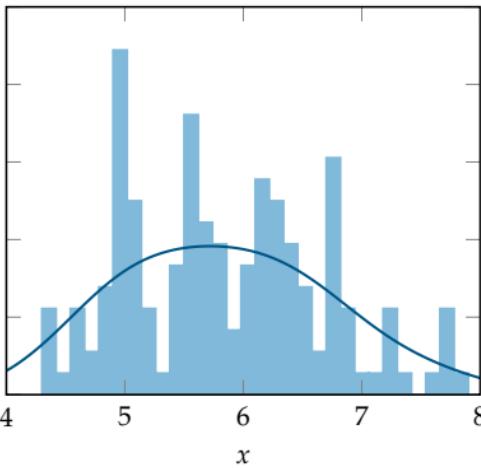
bandwidth = 0.1



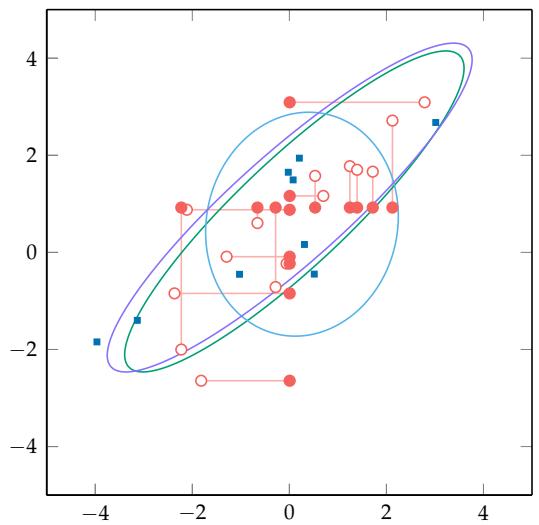
bandwidth = 0.2



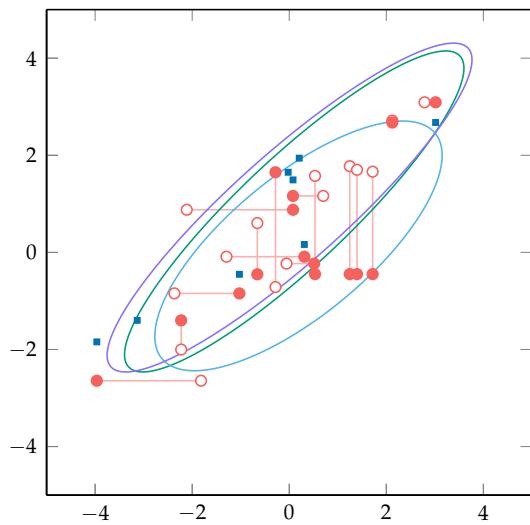
bandwidth = 0.5



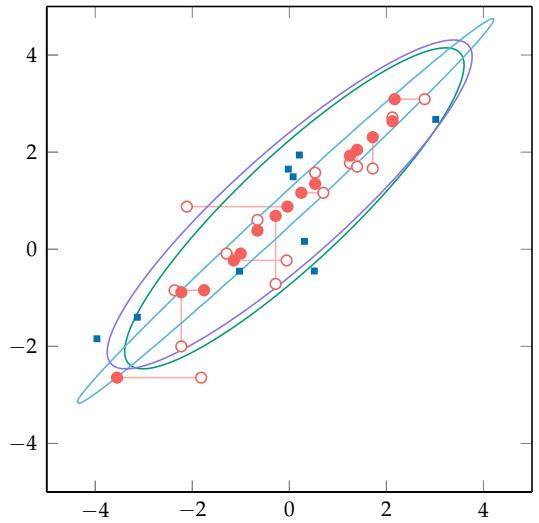
marginal mode



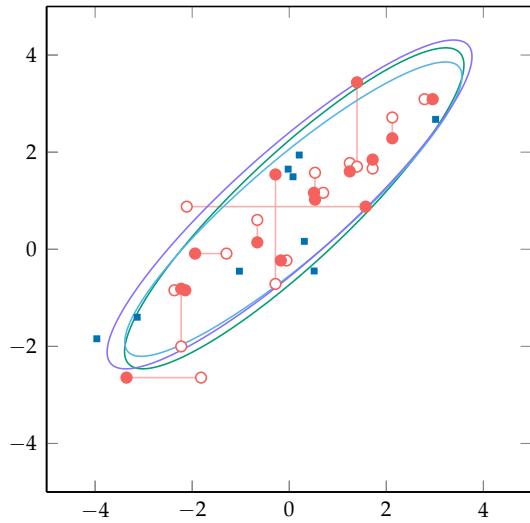
nearest



posterior mode



posterior sample



Marker is data that is:

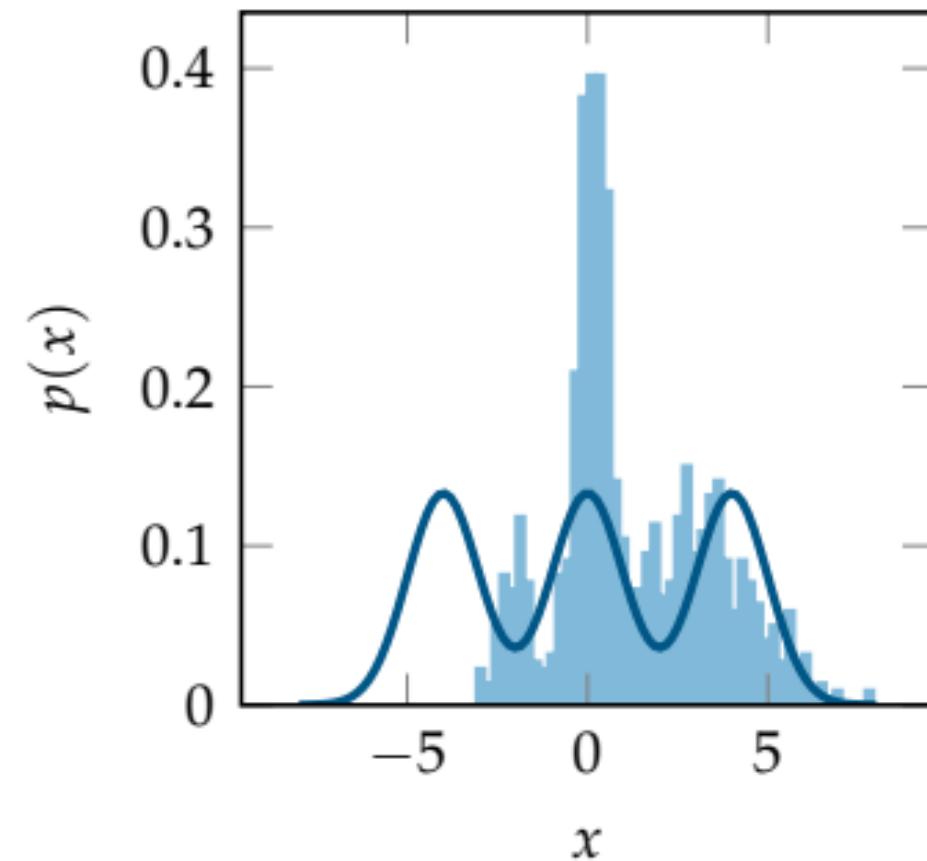
- observed
- missing
- imputed

Density ellipse estimated from:

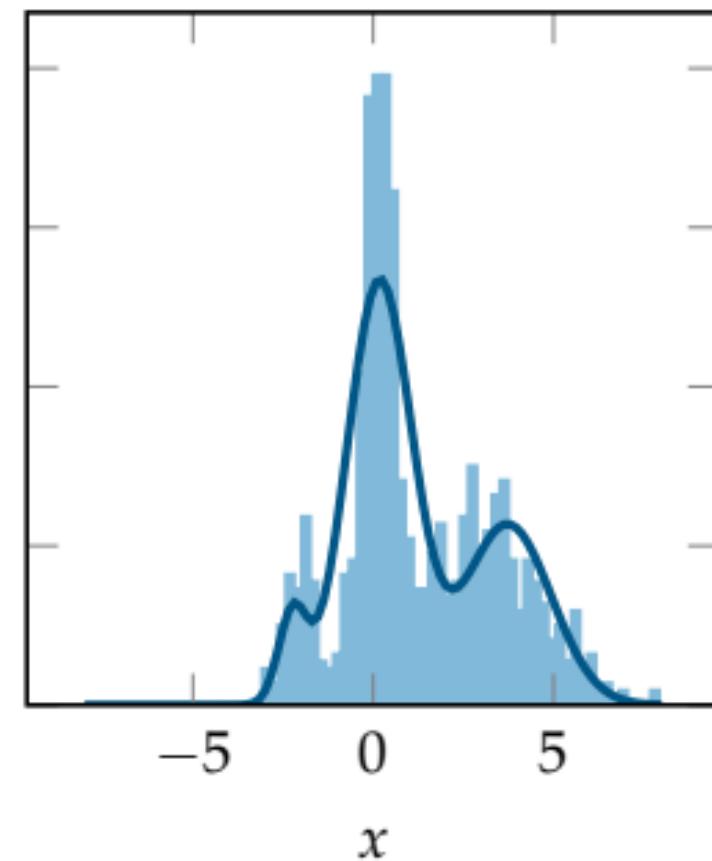
- all data (ground truth)
- observed and imputed
- observed only

<hr/>		<i>A</i>	<i>B</i>	weight	
<hr/>	<i>A</i>	<i>B</i>			
1	1	0	1	1	
0	1	0	0	$1 - P(b^1 \mid a^0) = 0.8$	
0	?	0	1	$P(b^1 \mid a^0) = 0.2$	
?	0	0	0	$\alpha P(a^0)P(b^0 \mid a^0) = \alpha 0.4 = 2/3$	
		1	0	$\alpha P(a^1)P(b^0 \mid a^1) = \alpha 0.2 = 1/3$	

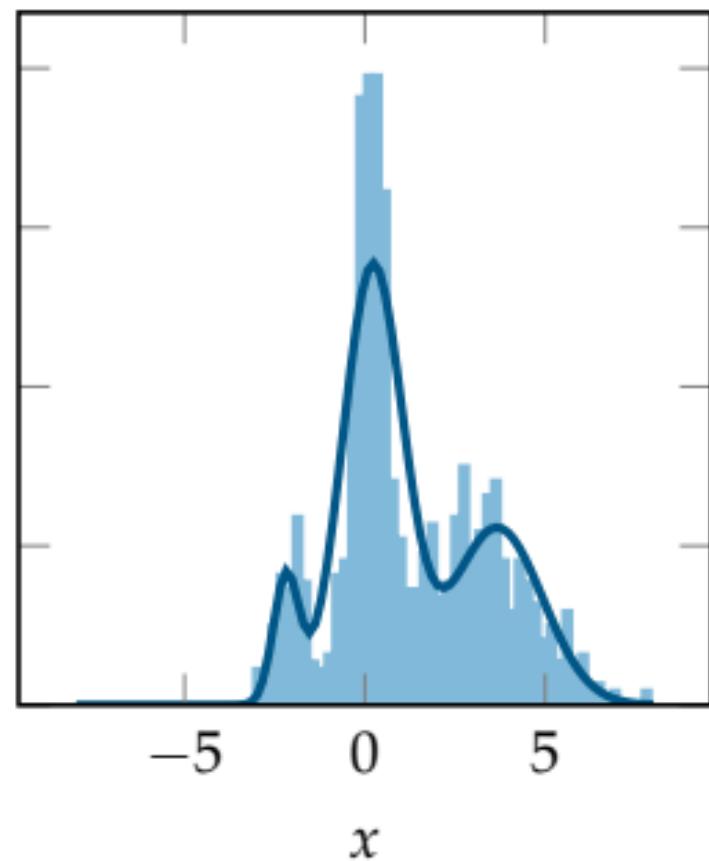
Iteration 1

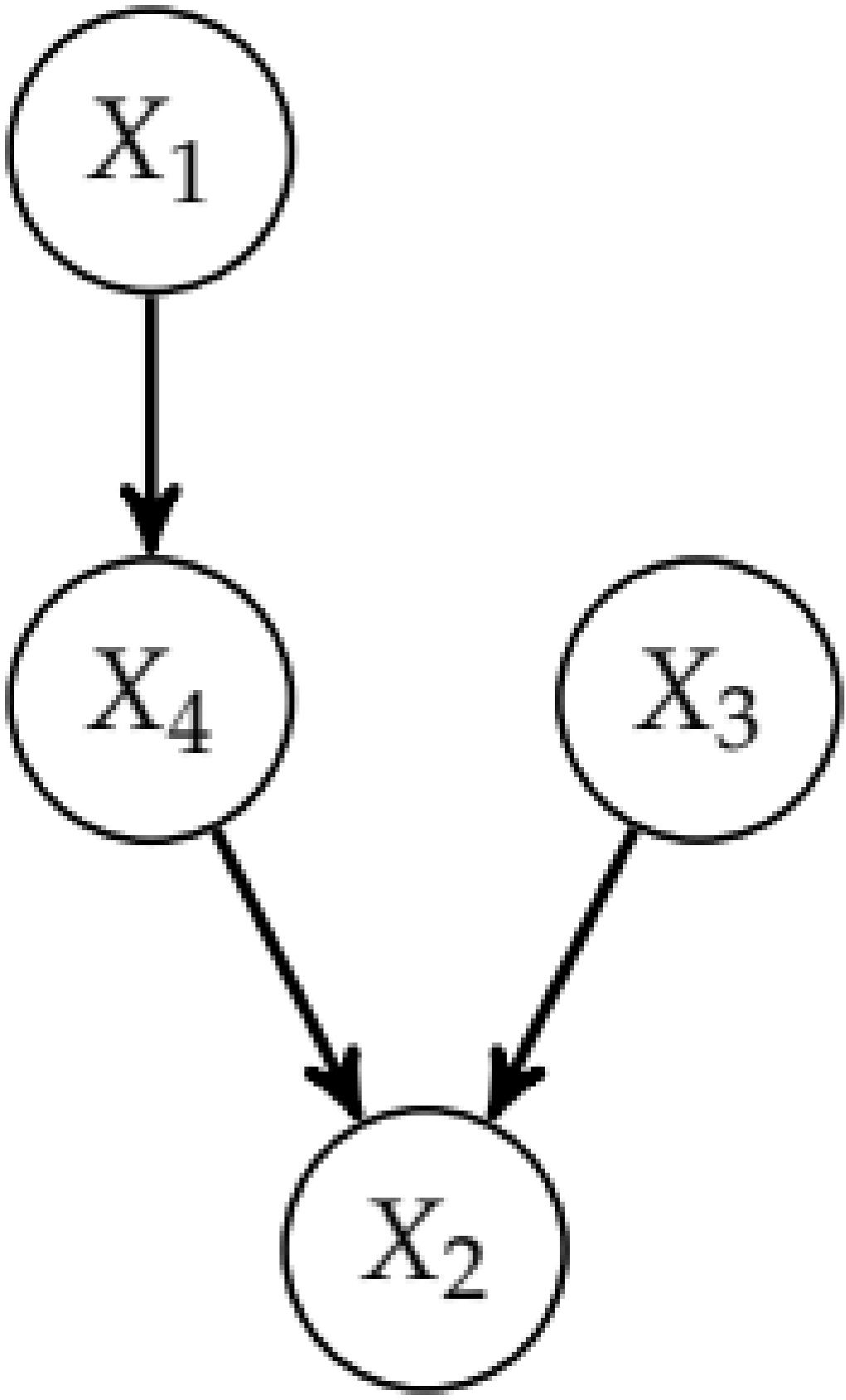


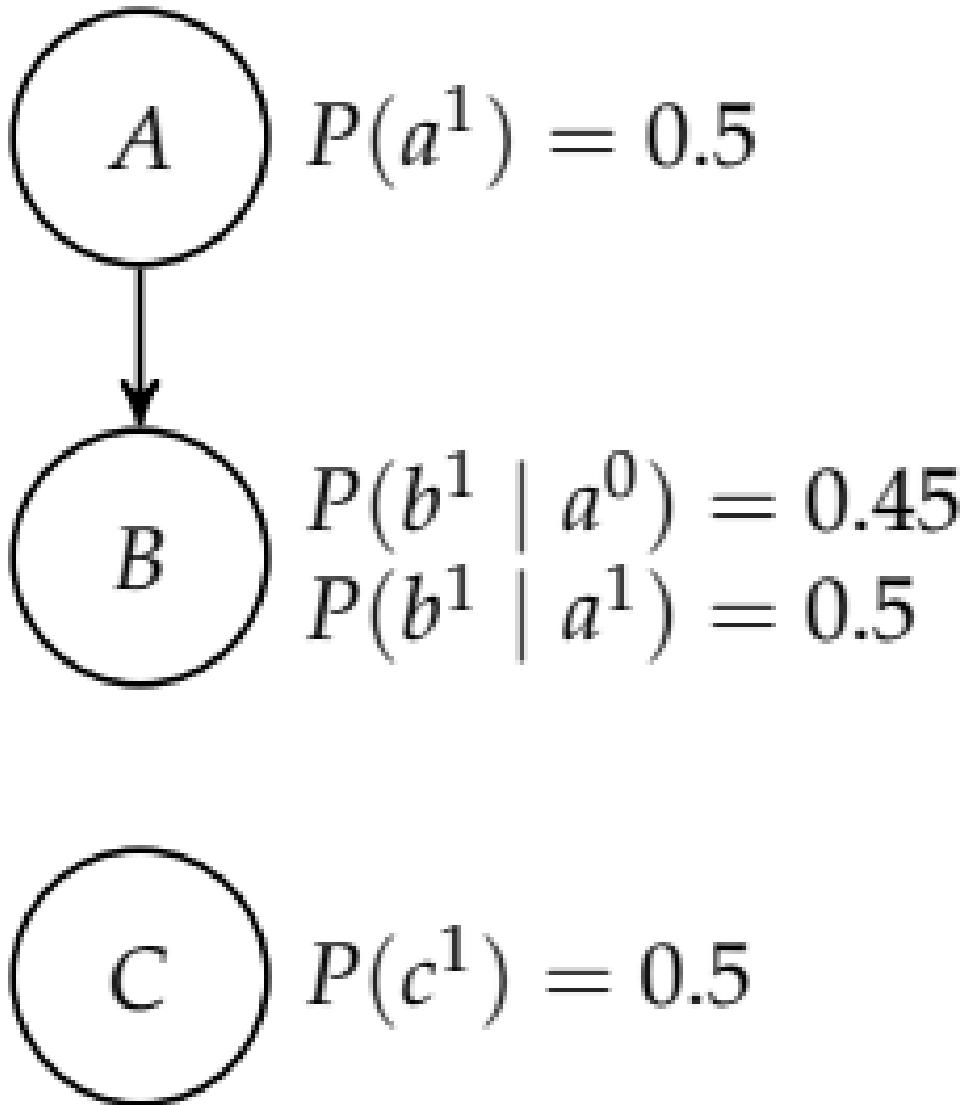
Iteration 2



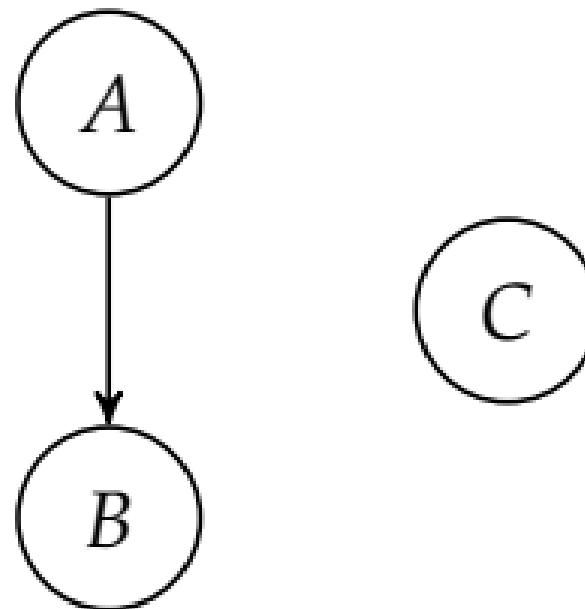
Iteration 3





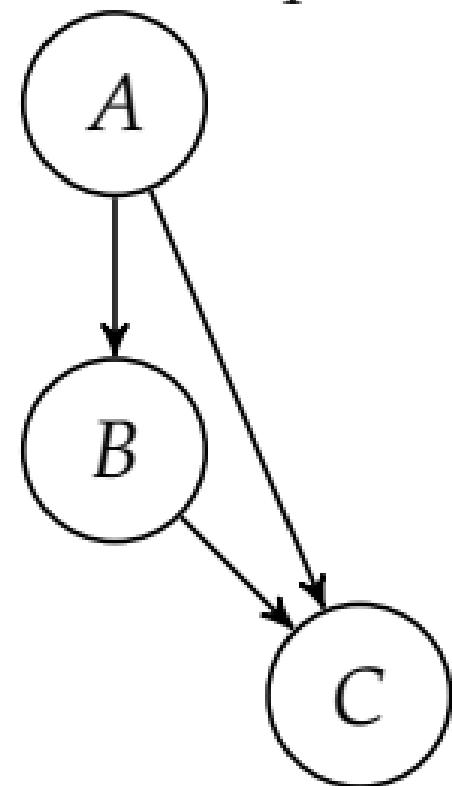


True model



$1 + 2 + 1 = 4$ parameters

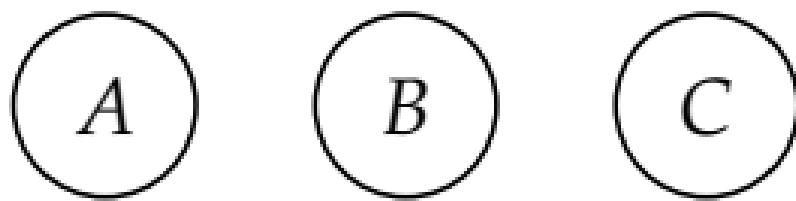
Completely connected



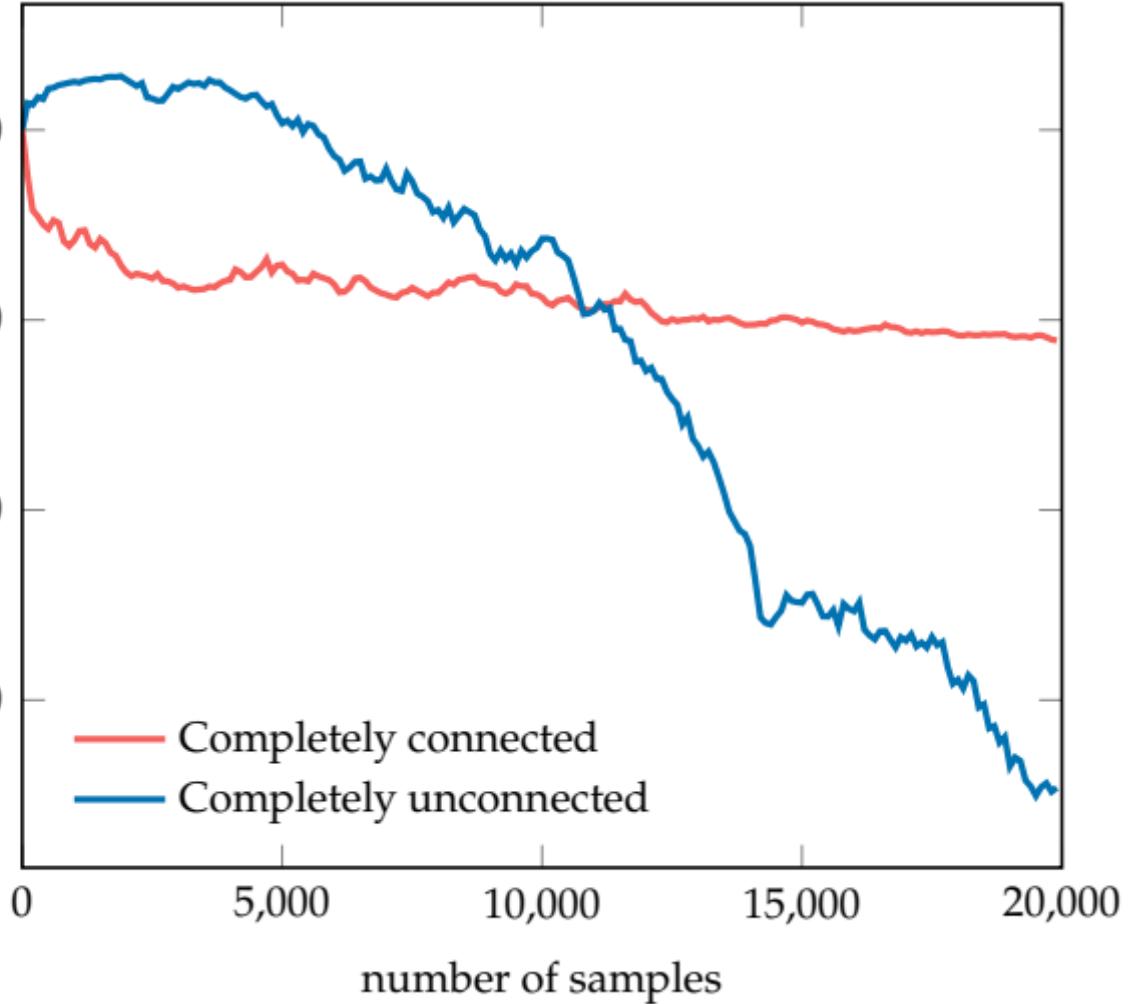
$1 + 2 + 4 = 7$ parameters

Completely unconnected

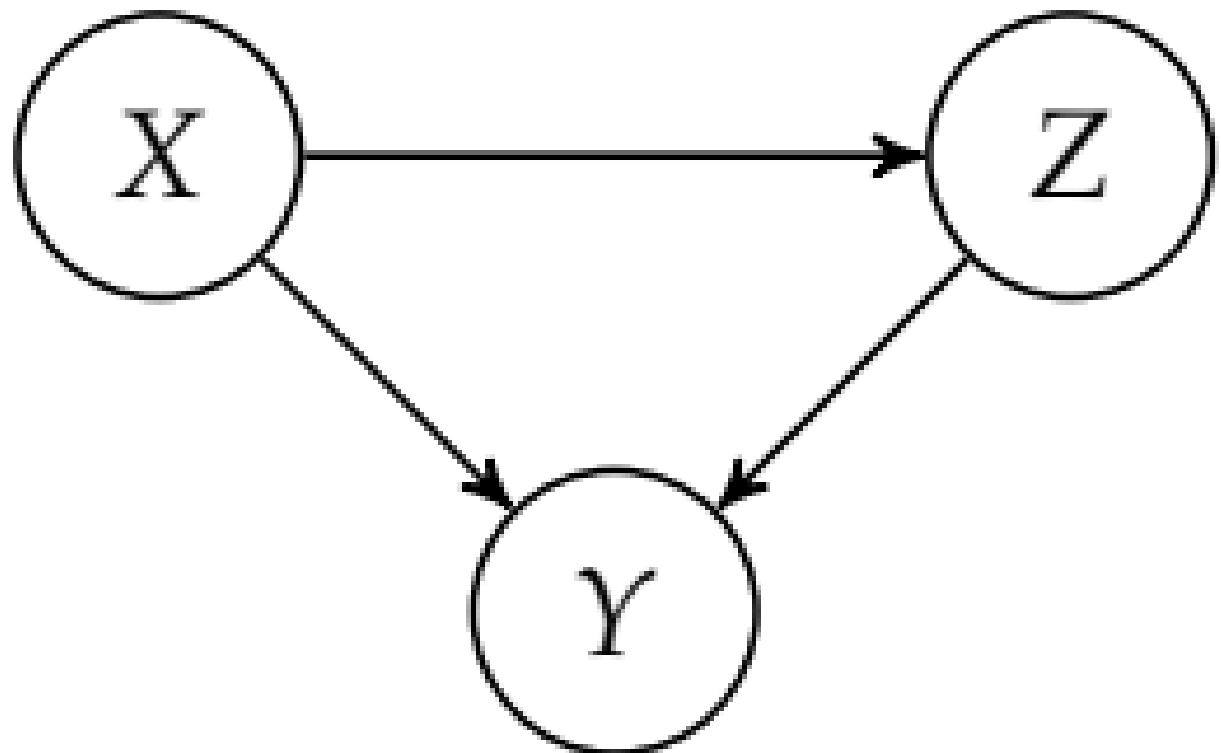
$1 + 1 + 1 = 3$ parameters



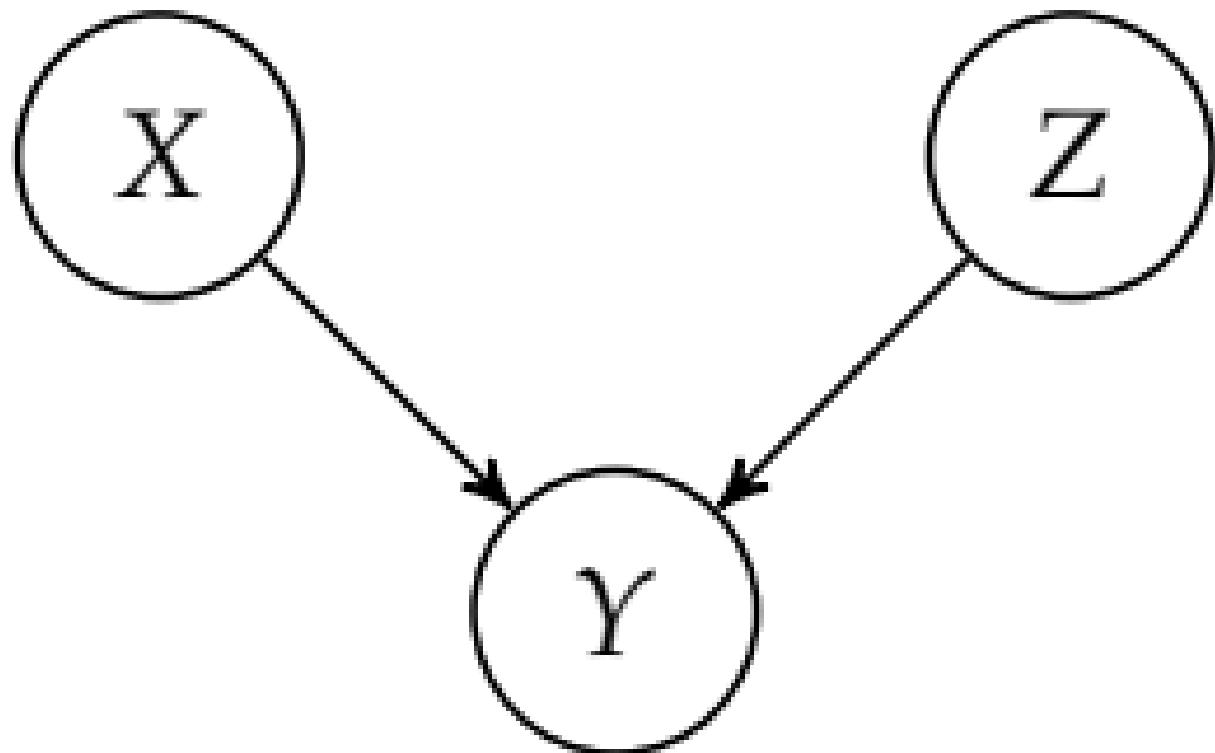
Bayesian score relative to true model



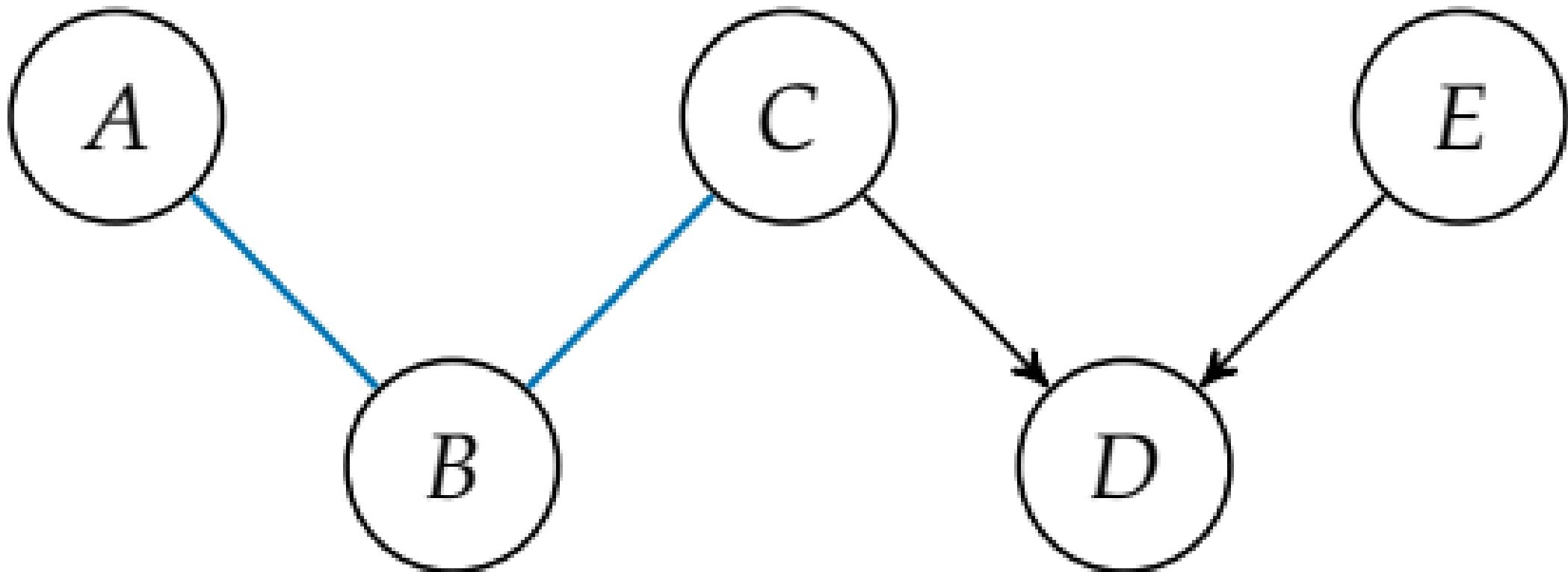
moral



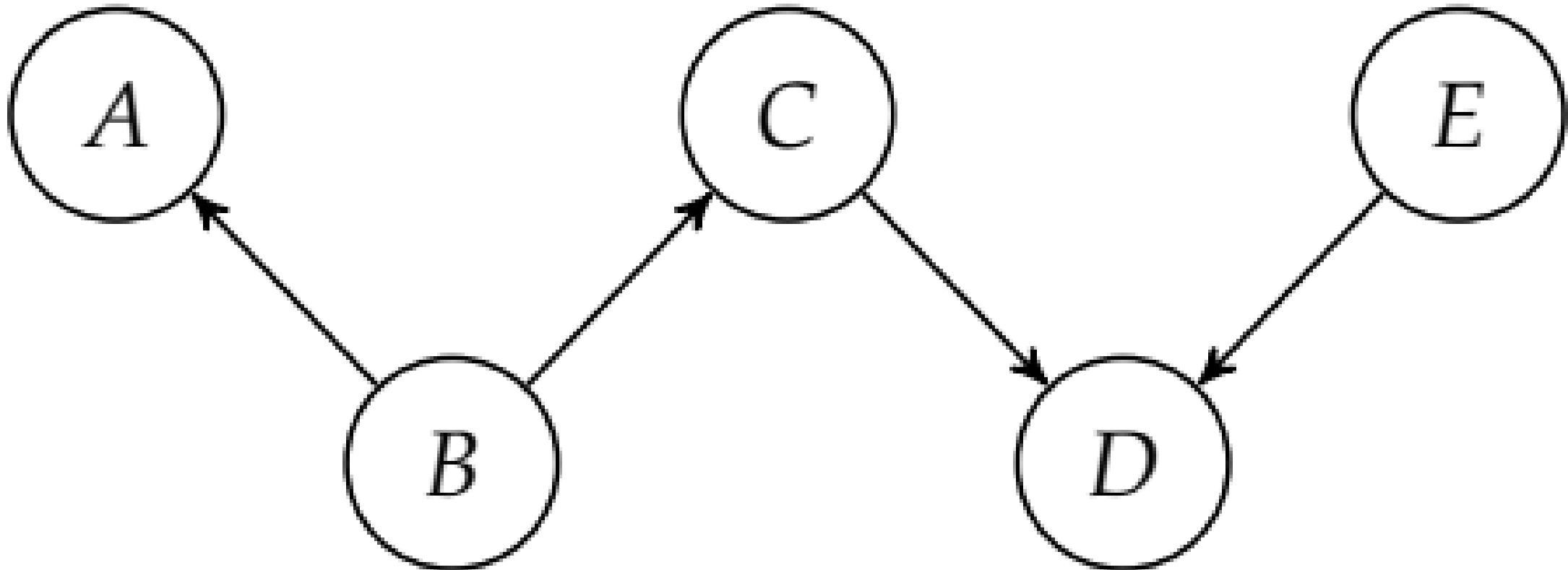
immoral



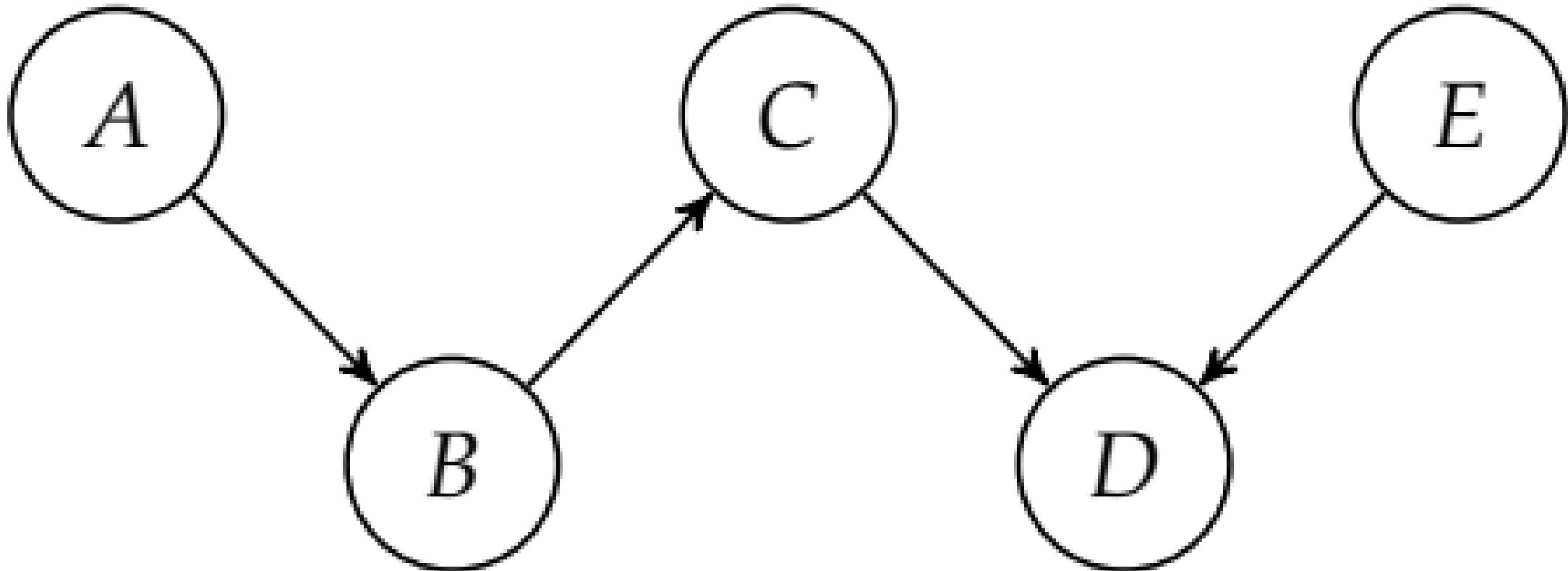
Markov equivalence class



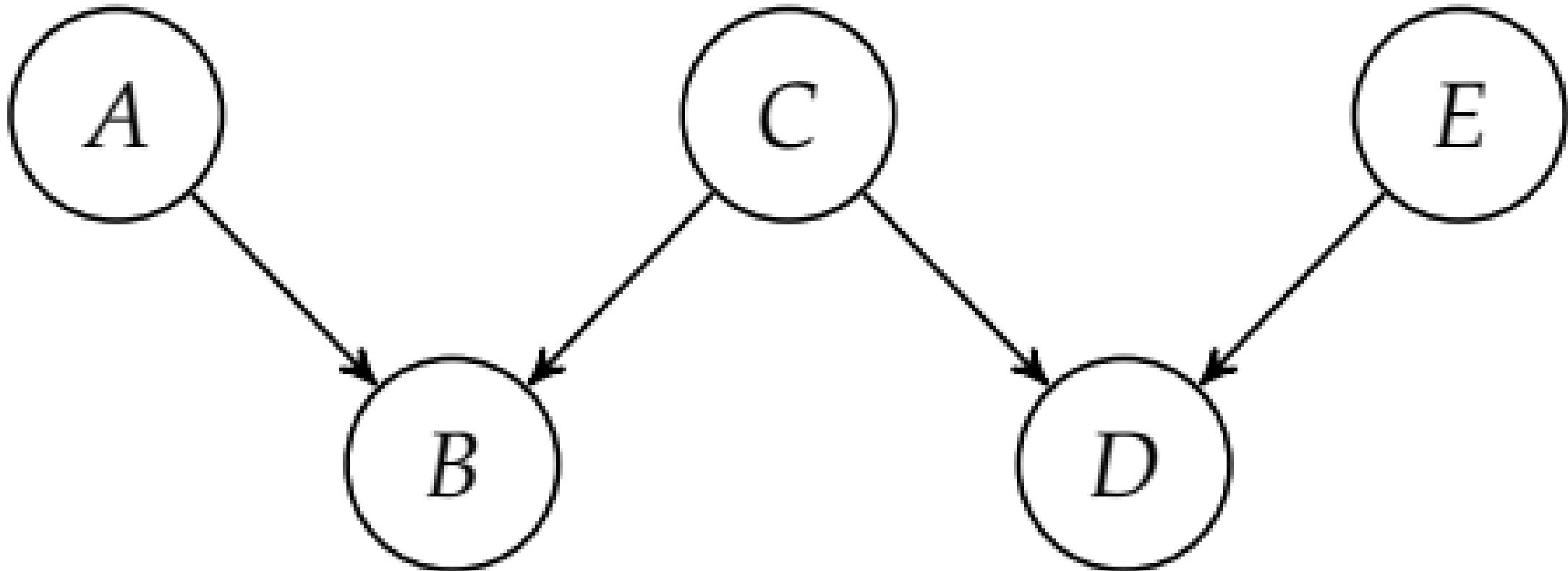
Member

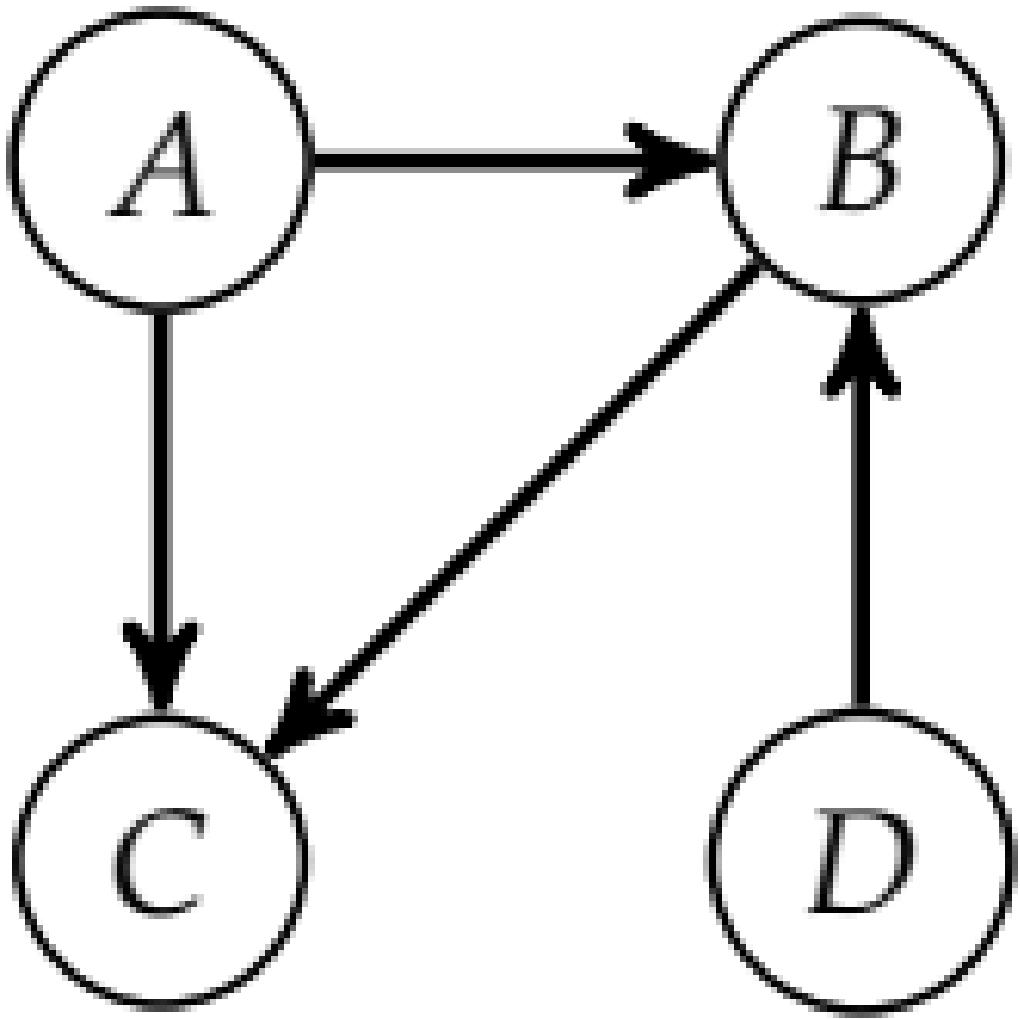


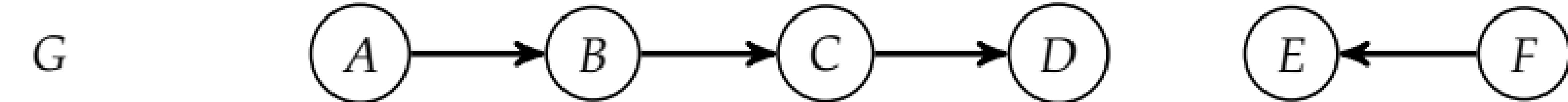
Member

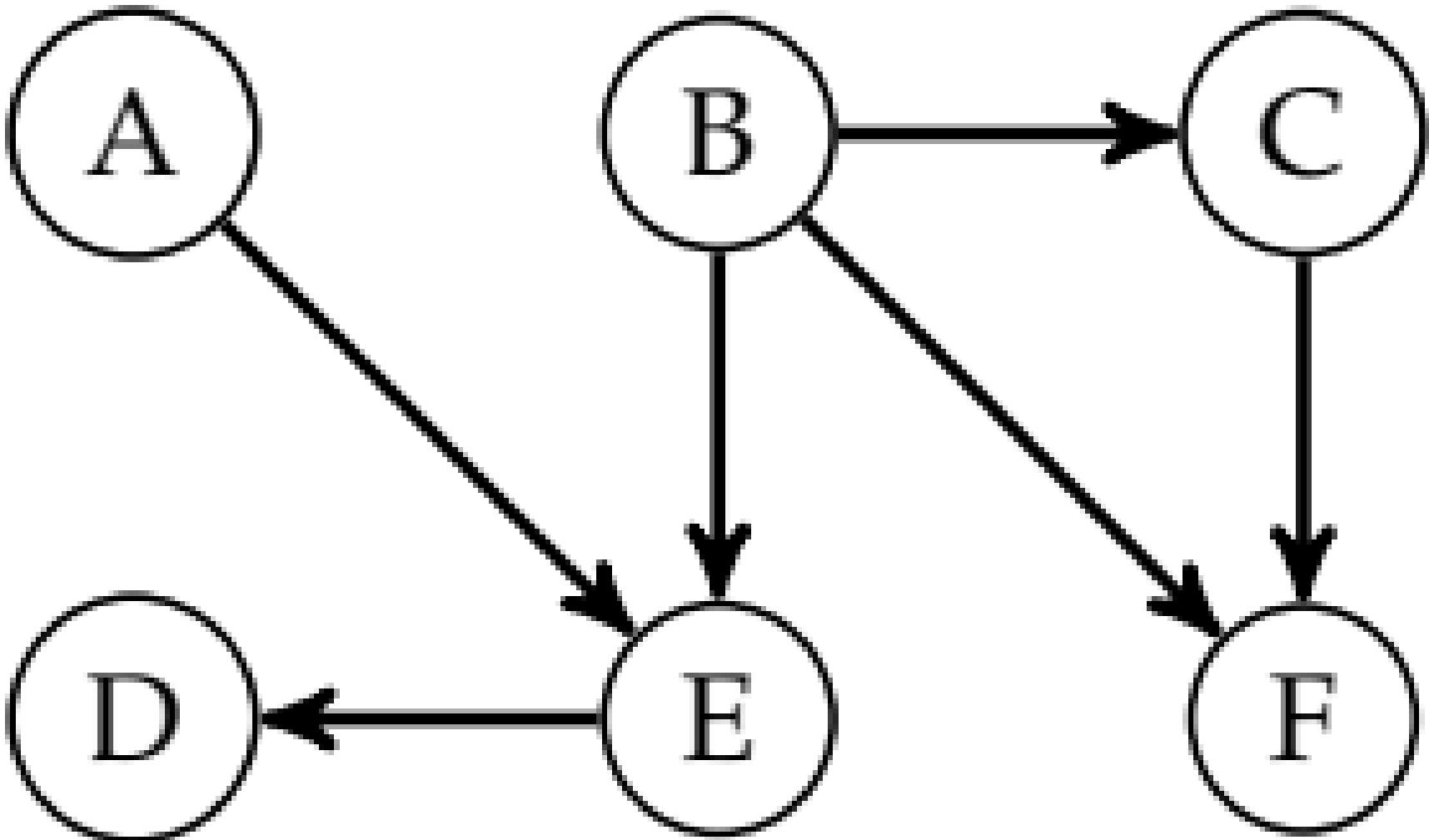


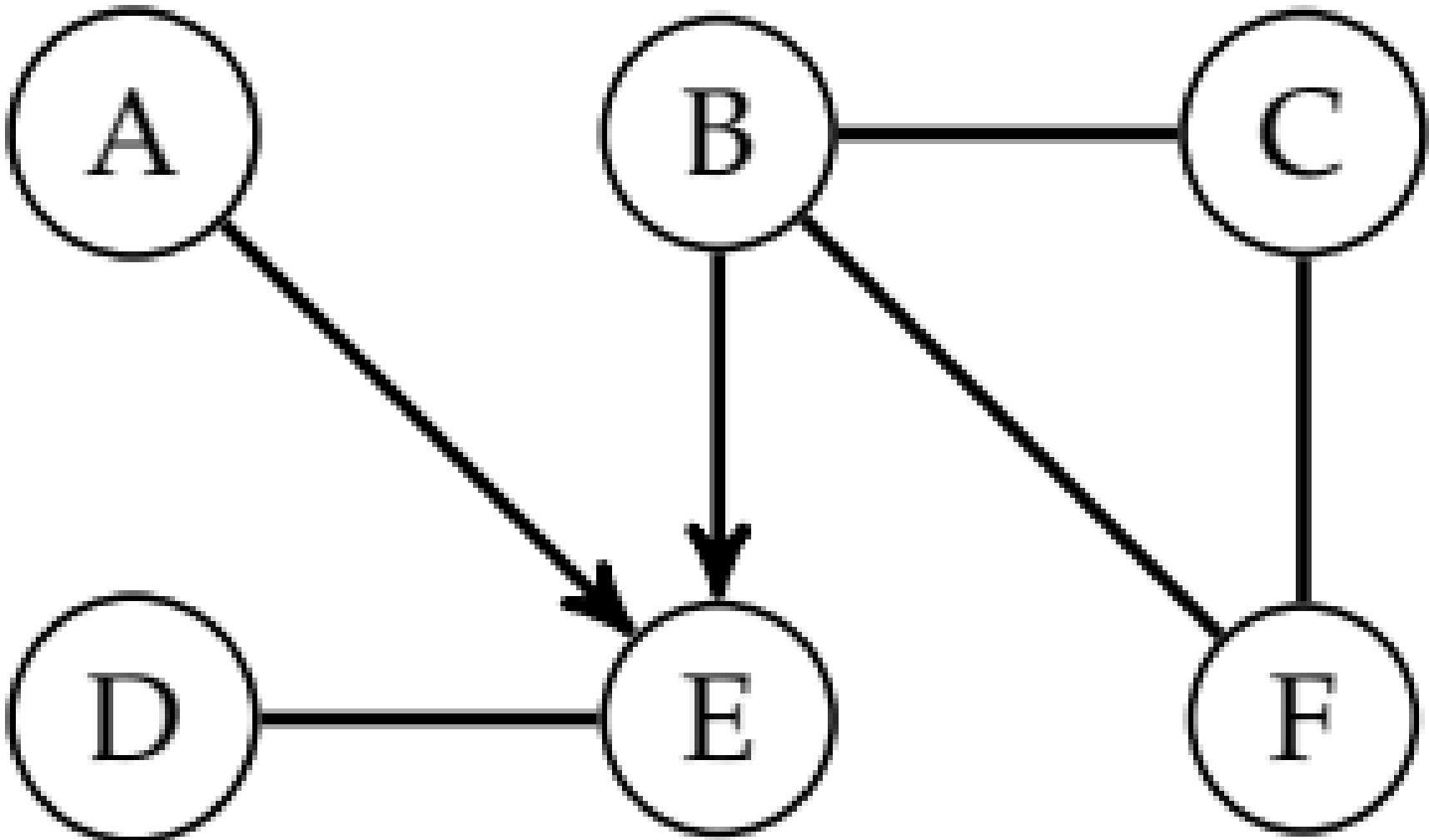
Nonmember

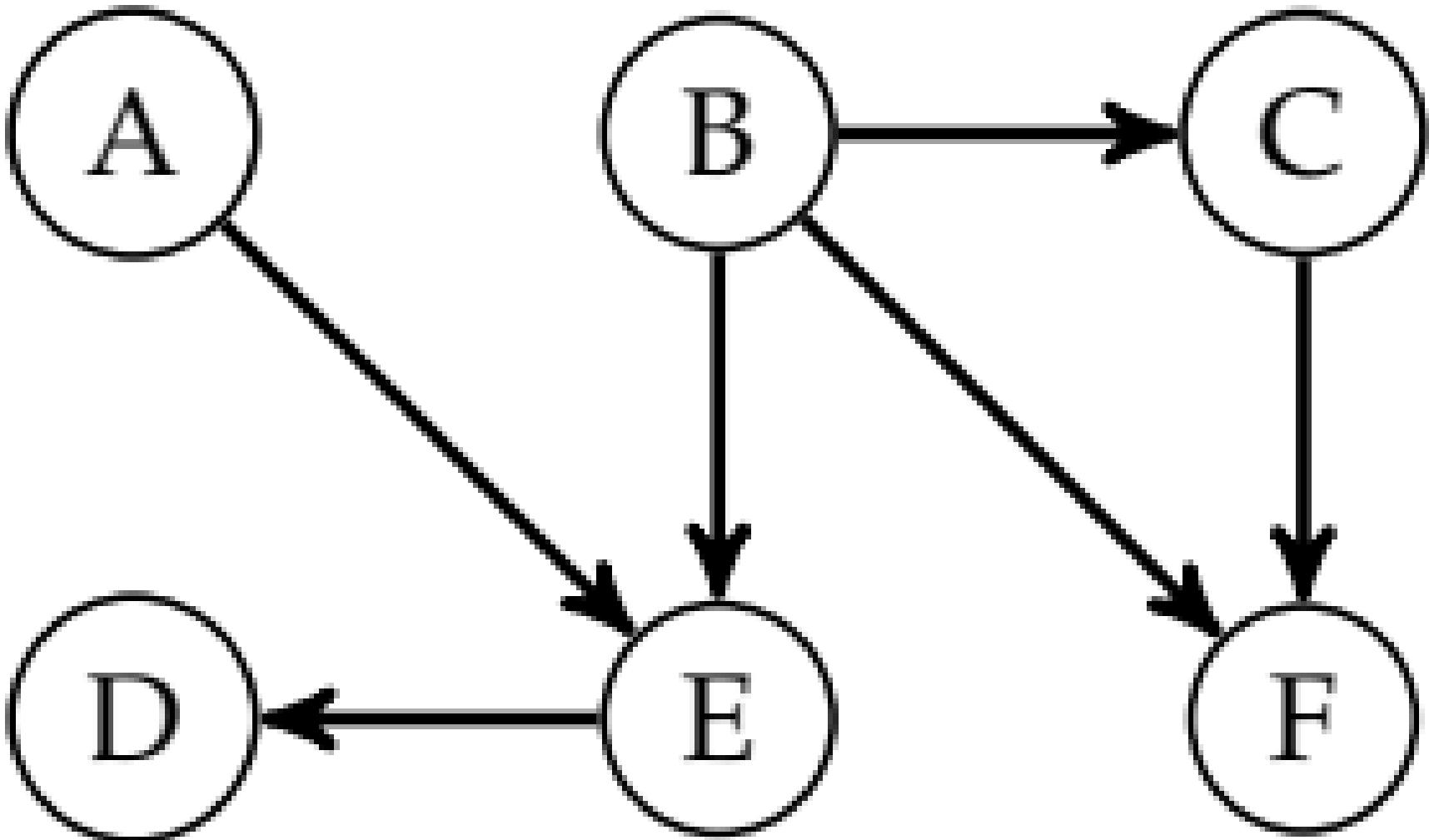


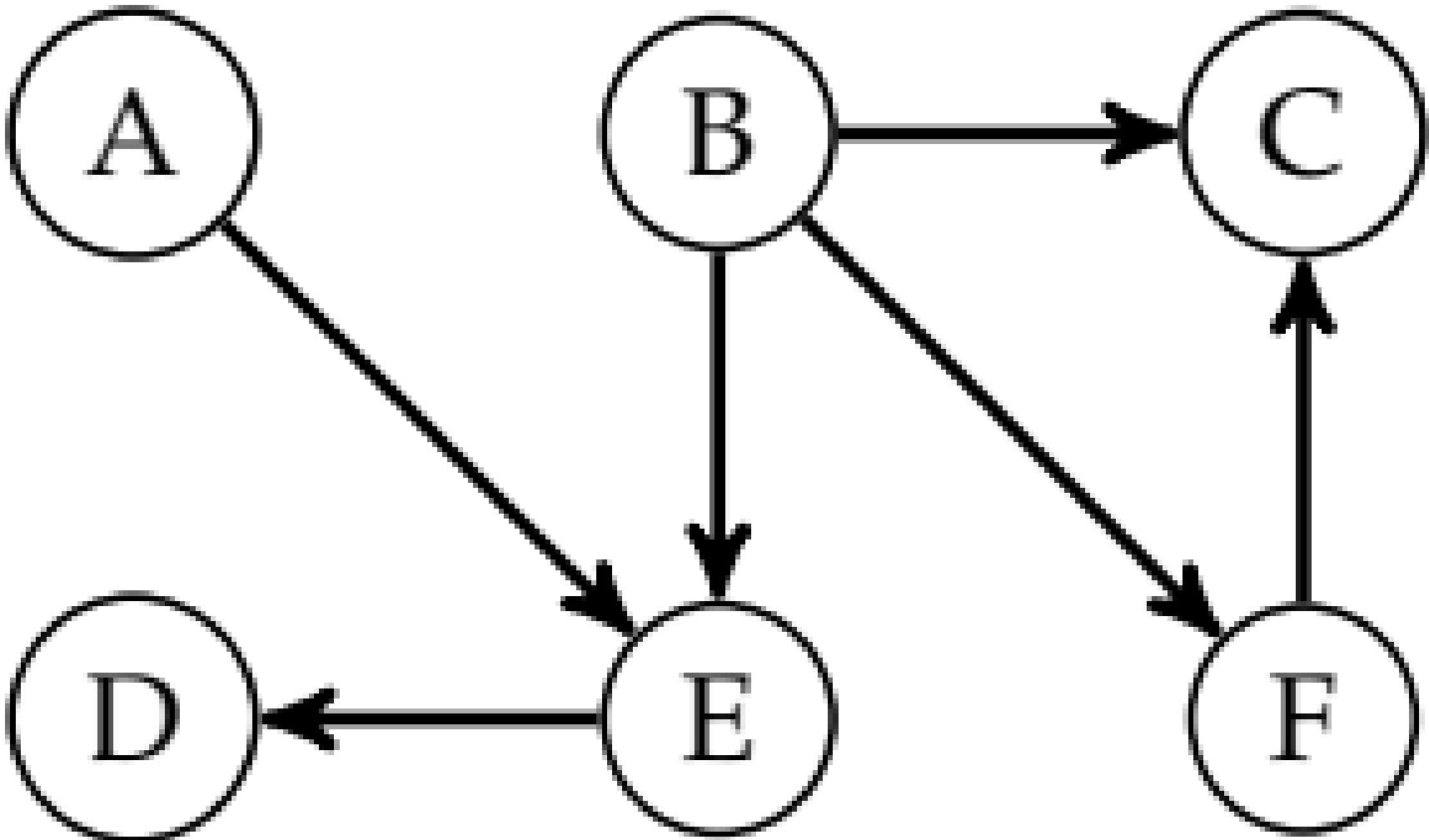


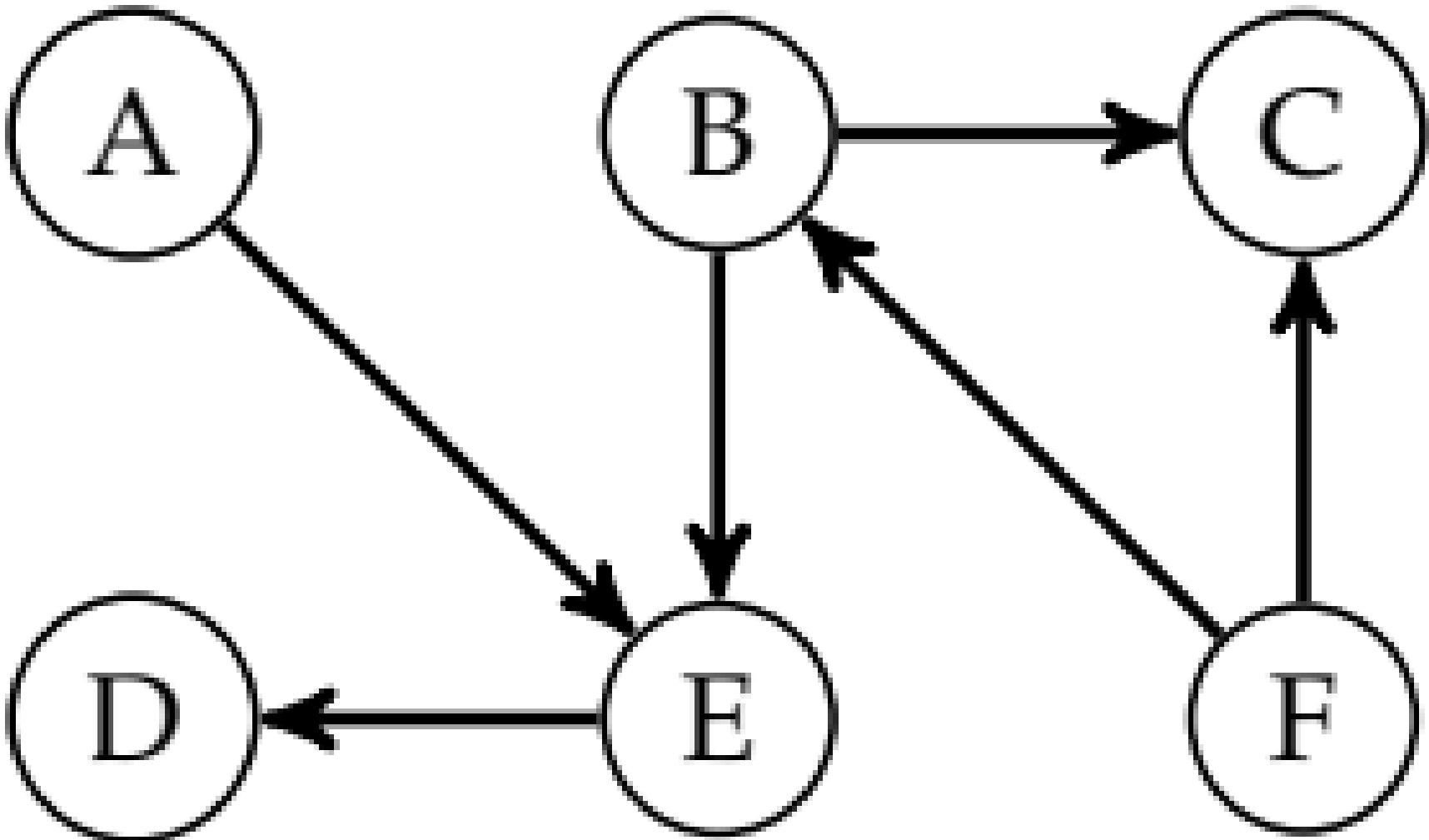


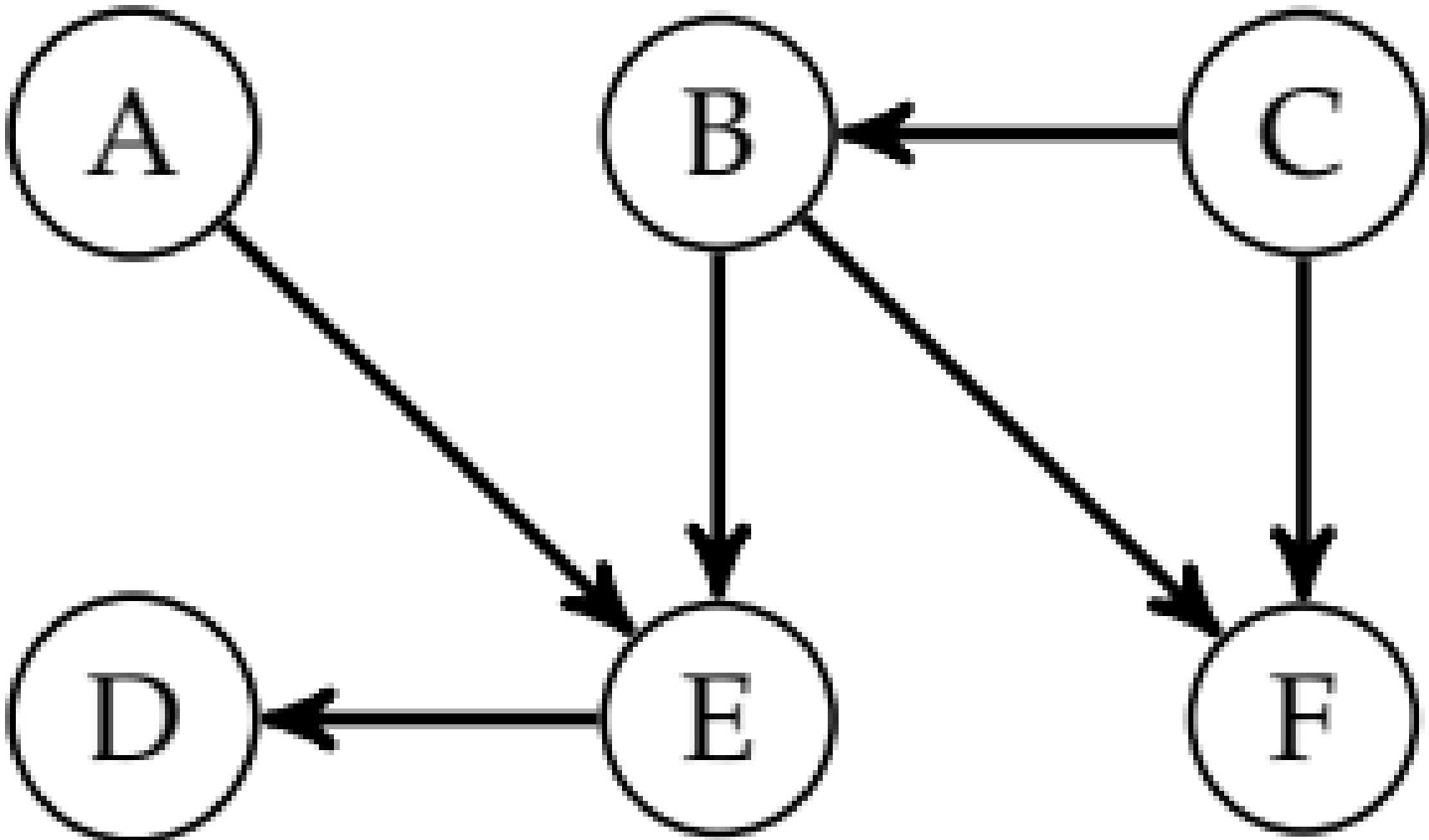


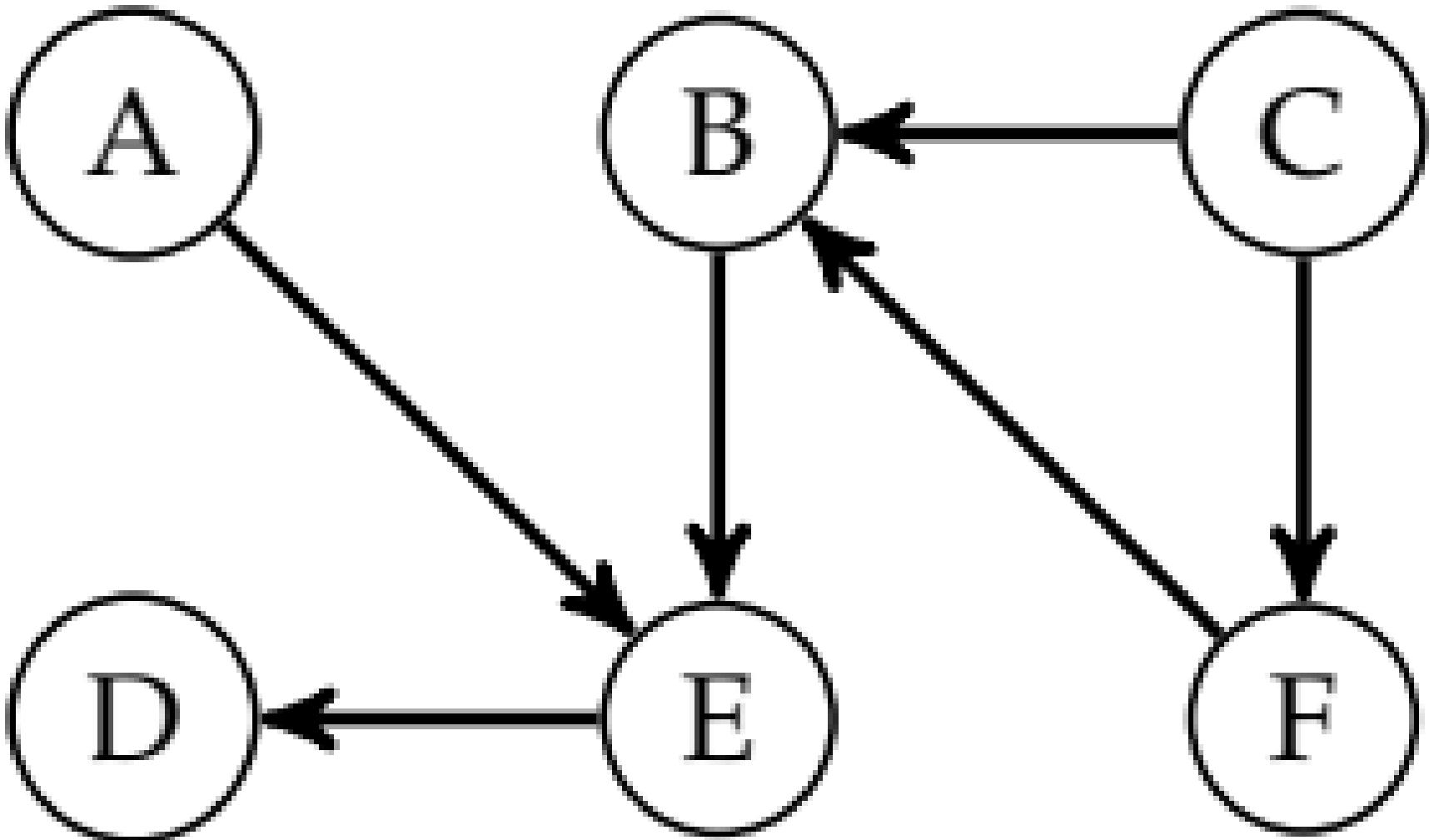


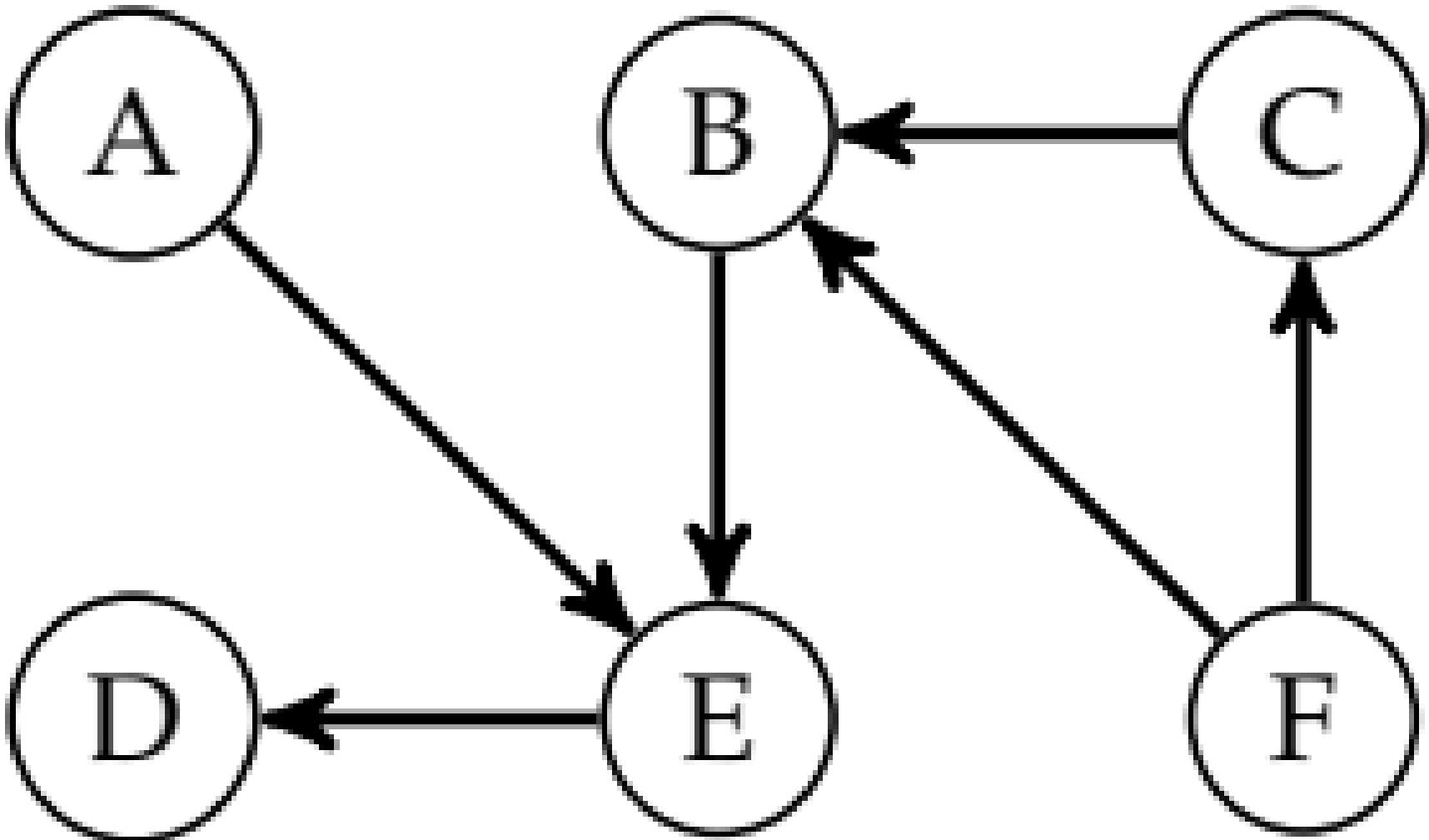


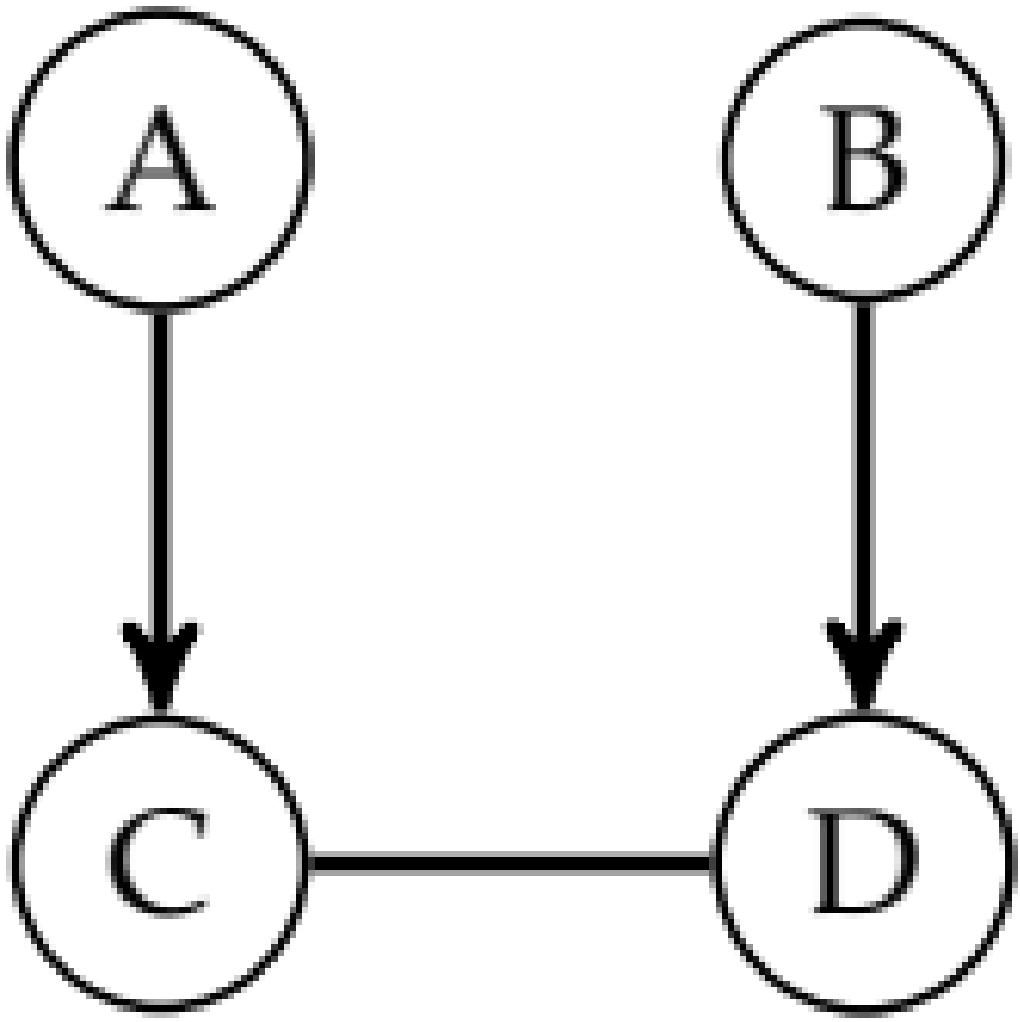


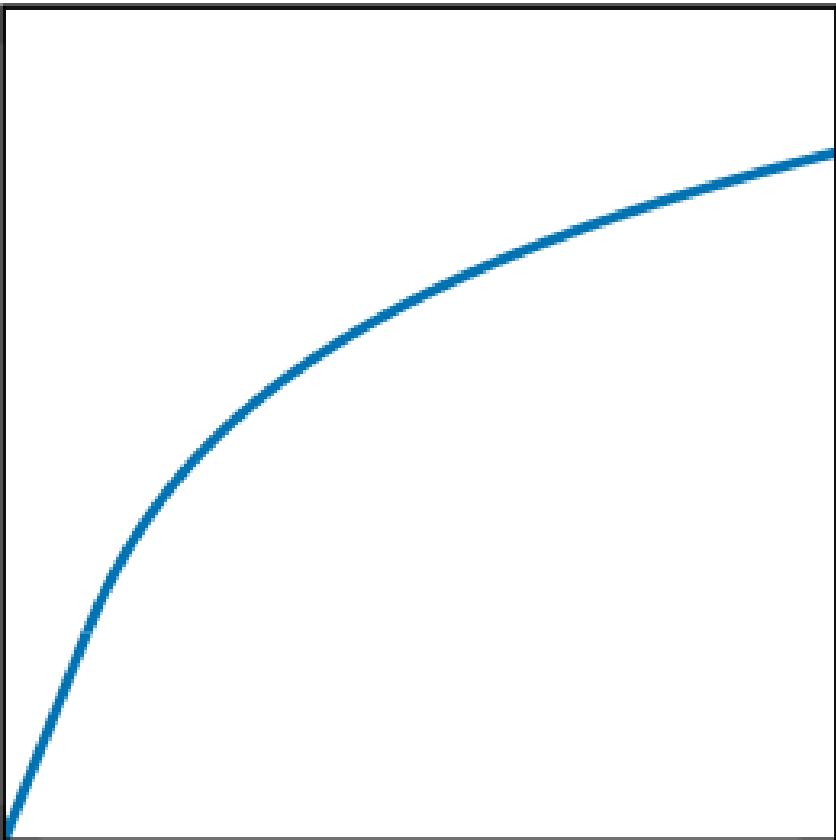


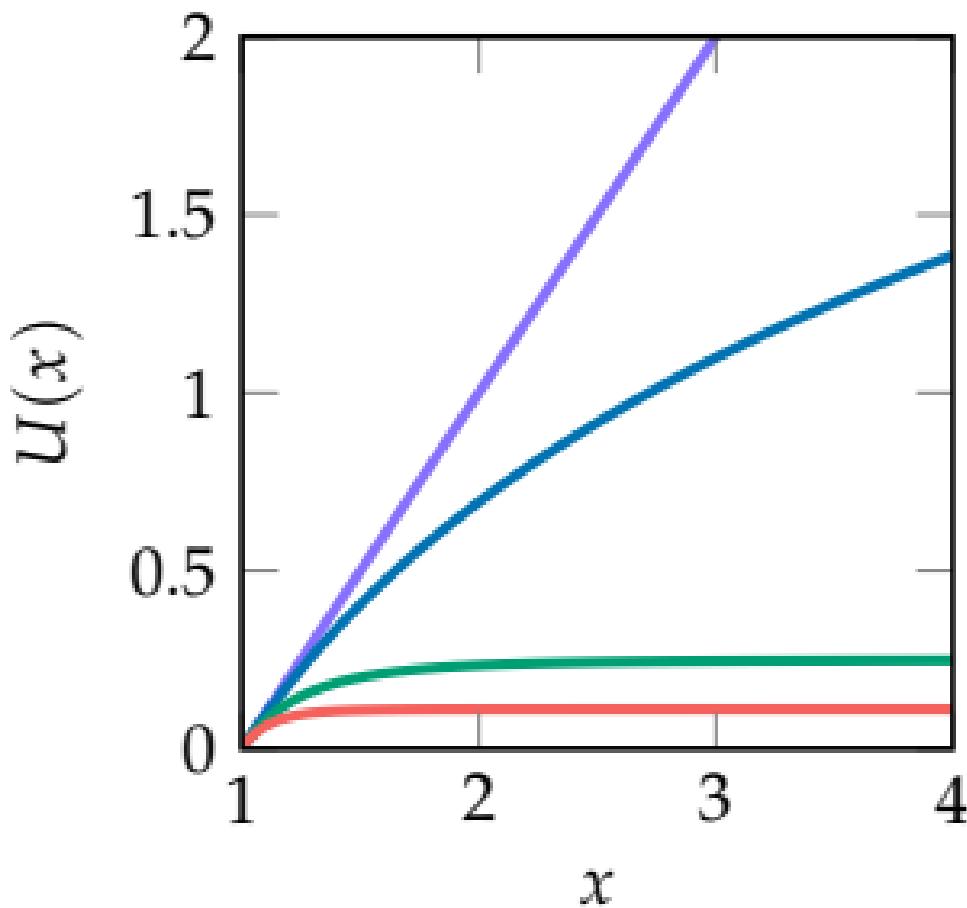




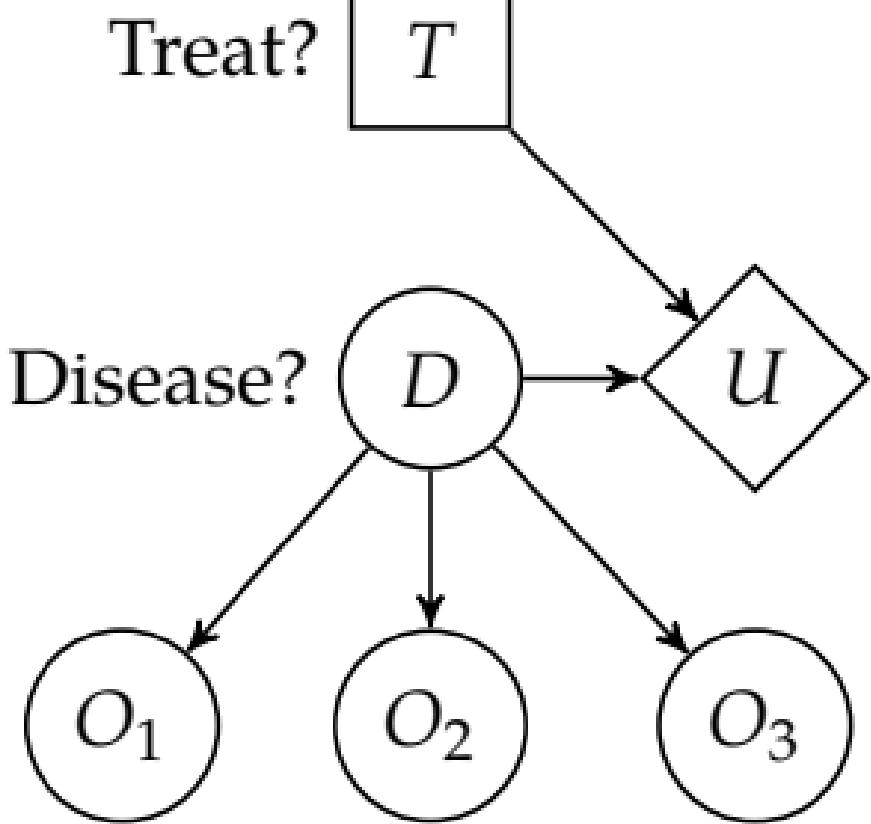




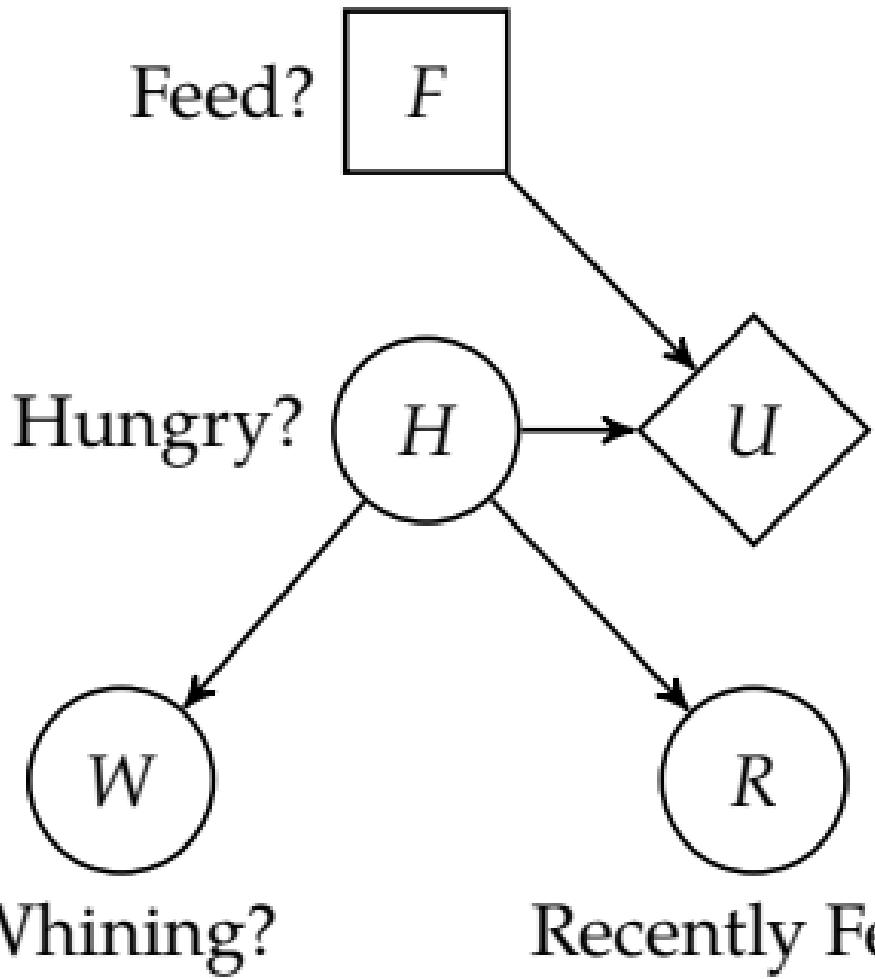
$U(x)$ 0 x 

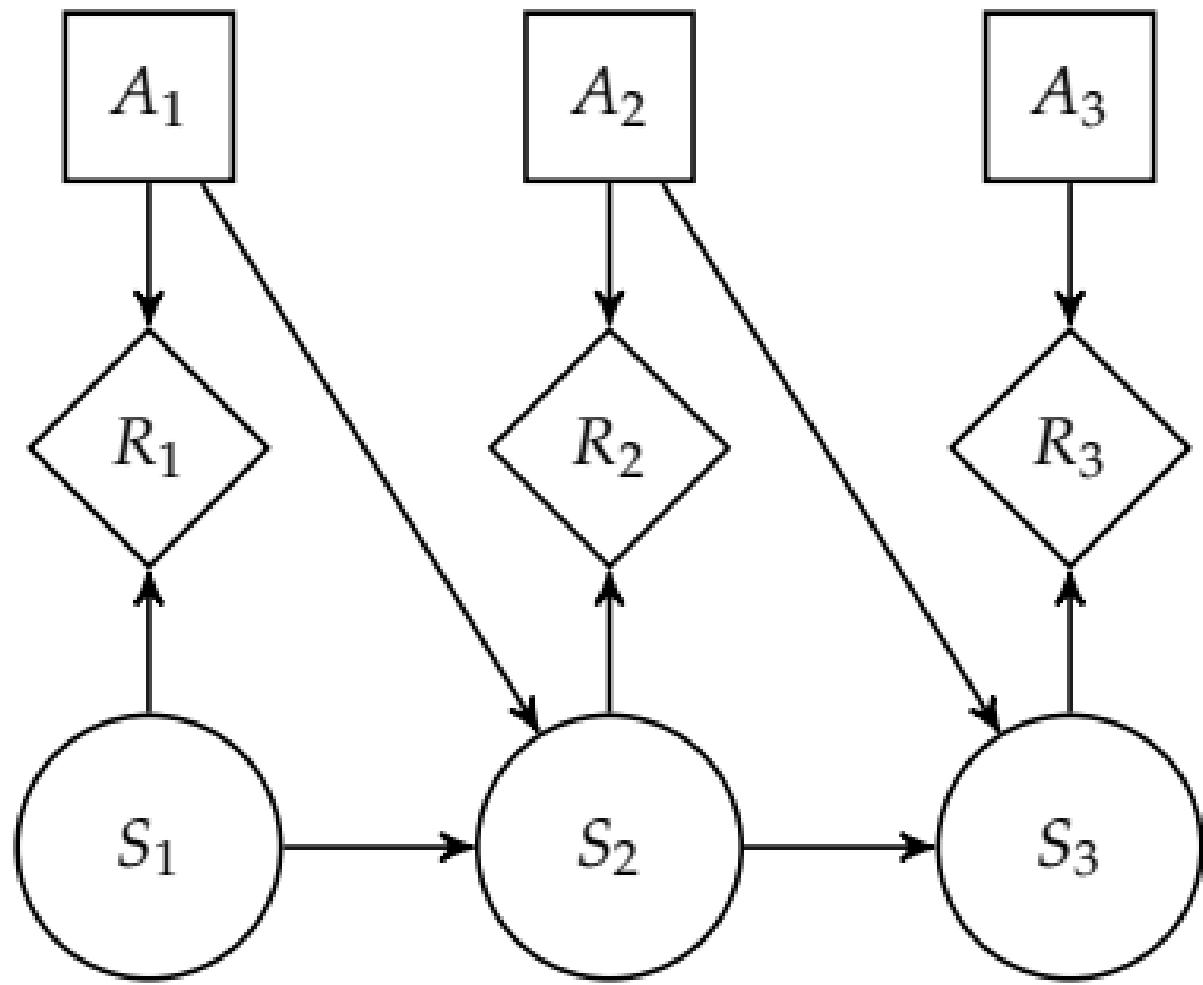


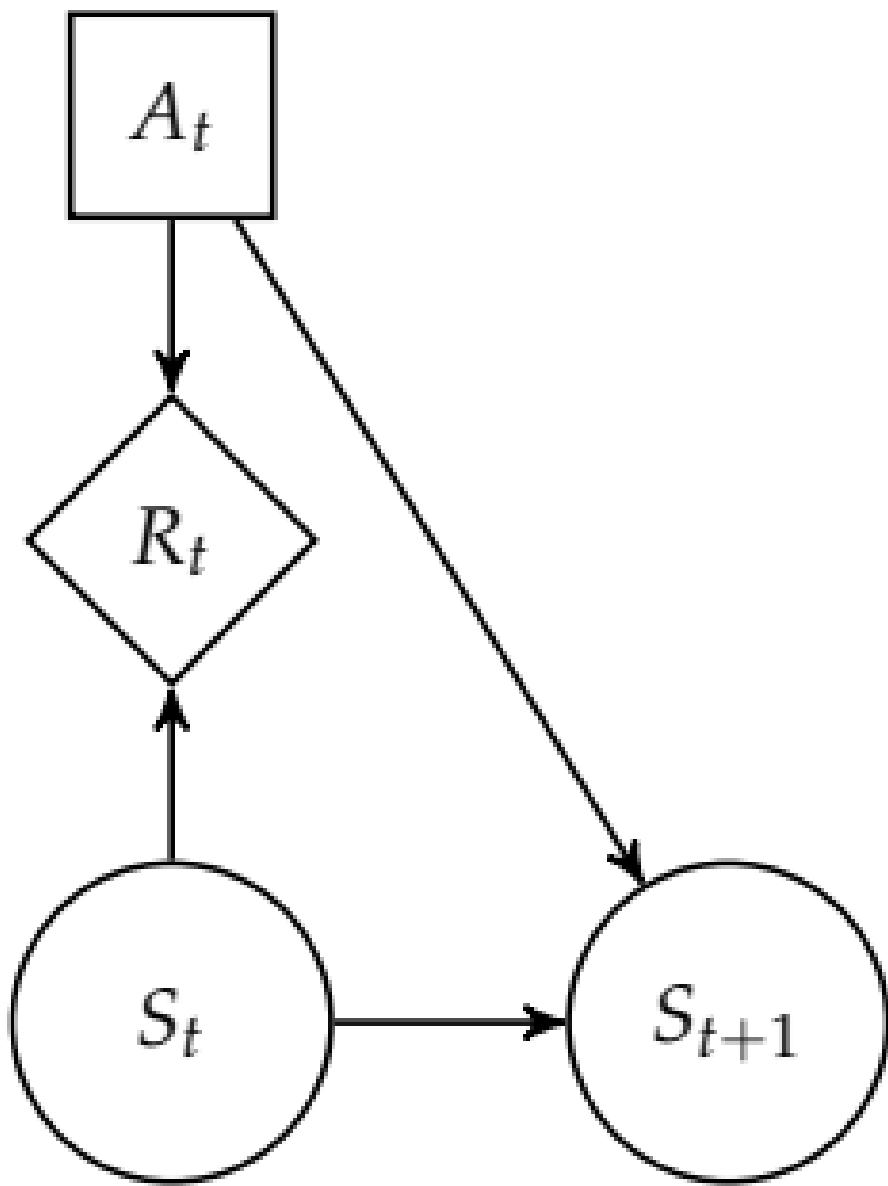
— $\lambda = 0$ — $\lambda \rightarrow 1$
— $\lambda = 5$ — $\lambda = 10$

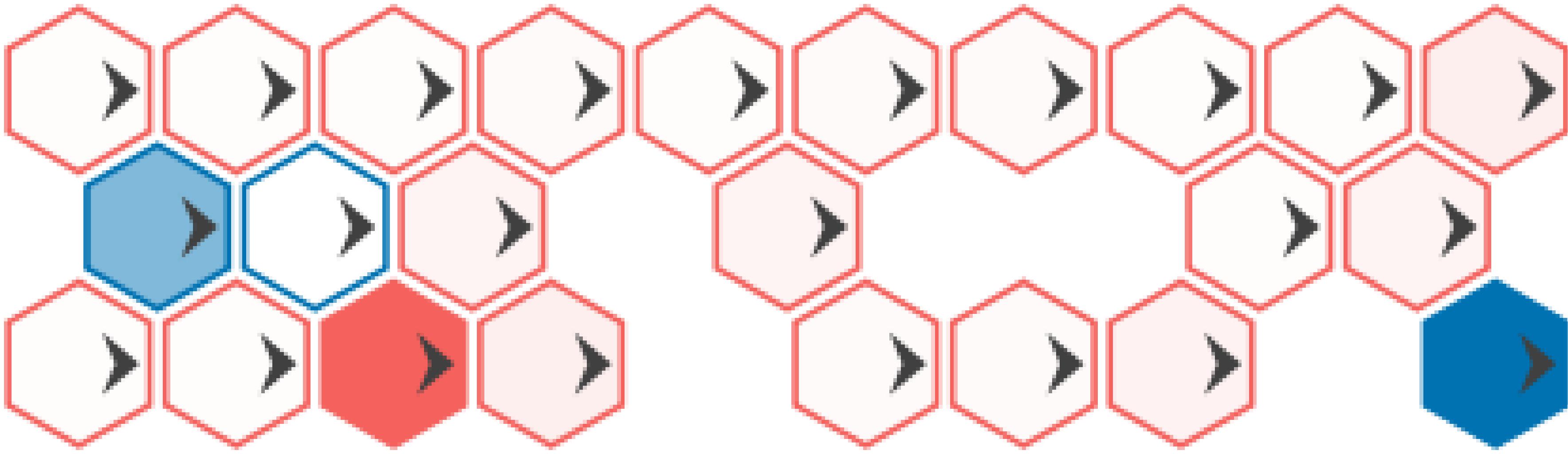


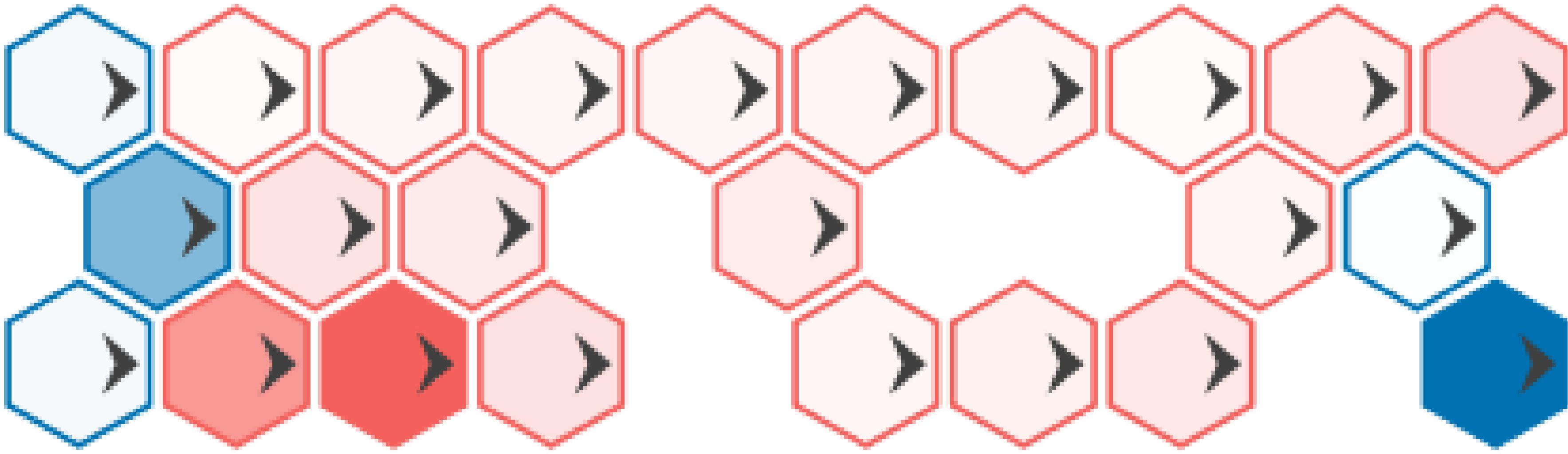
Results from diagnostic tests

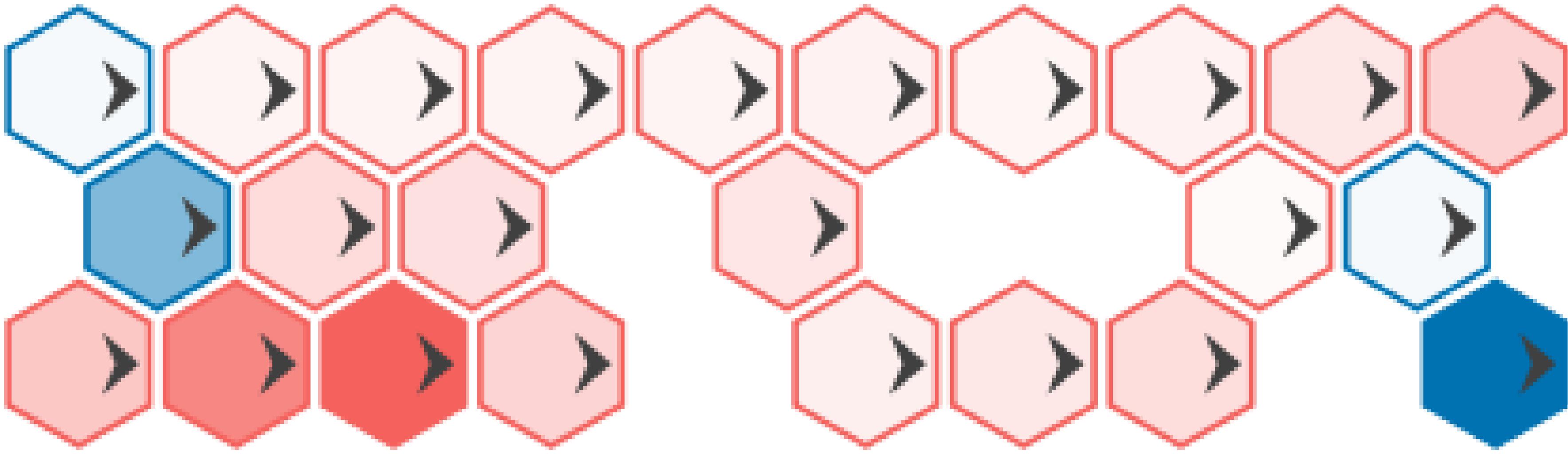


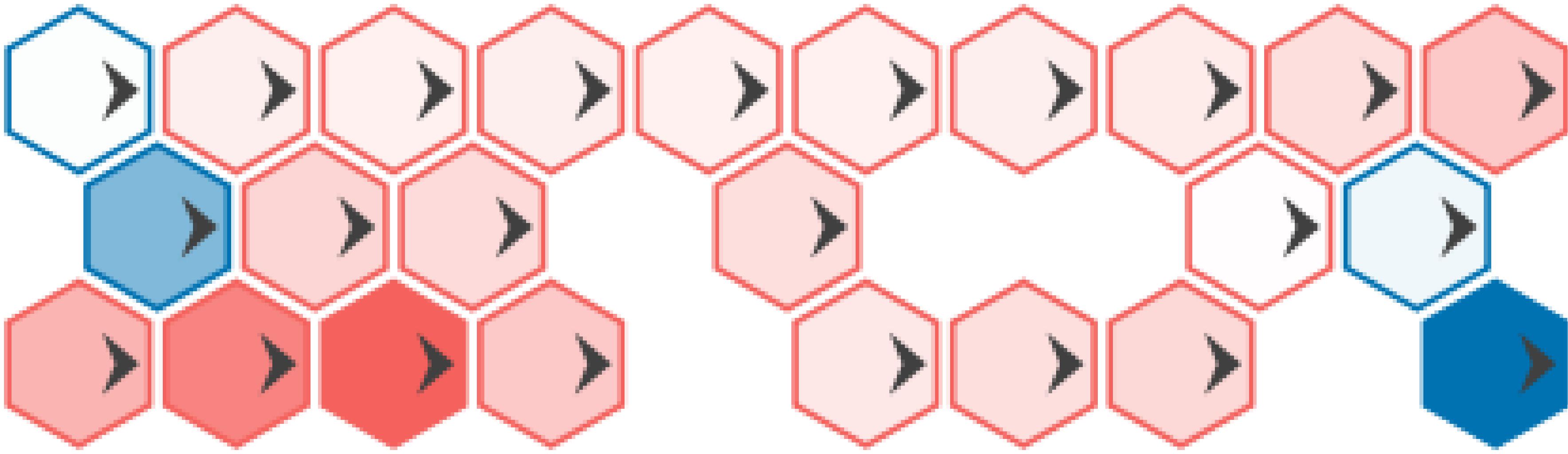


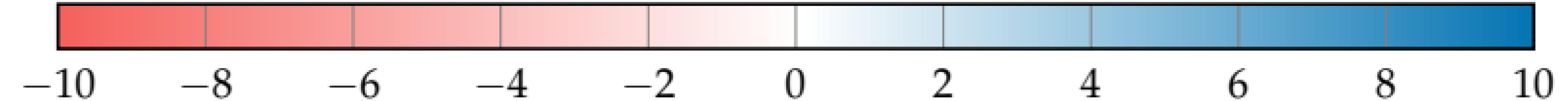


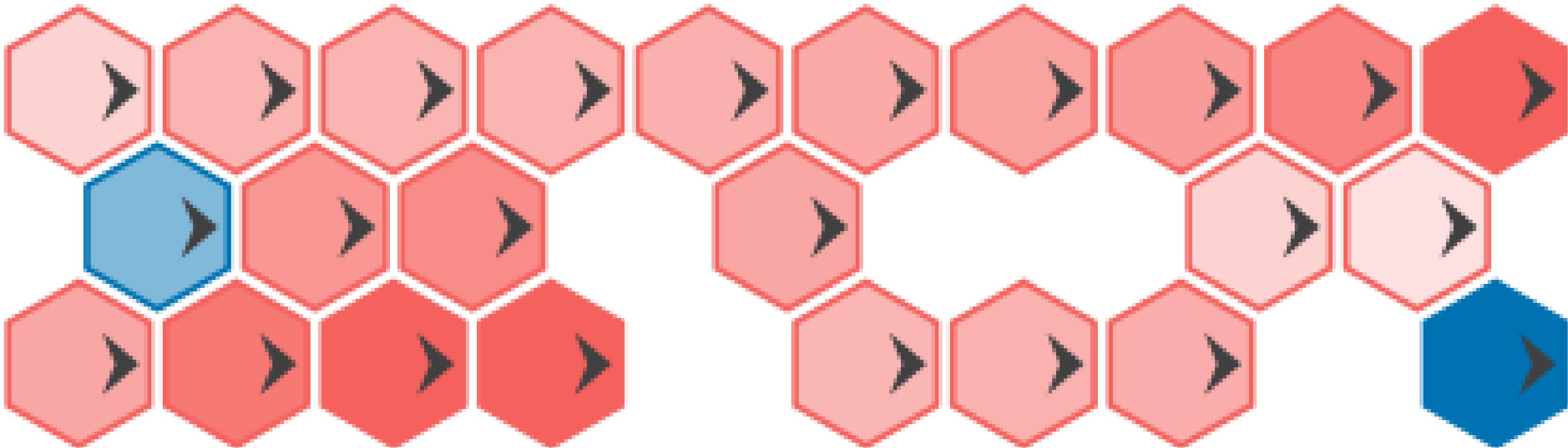


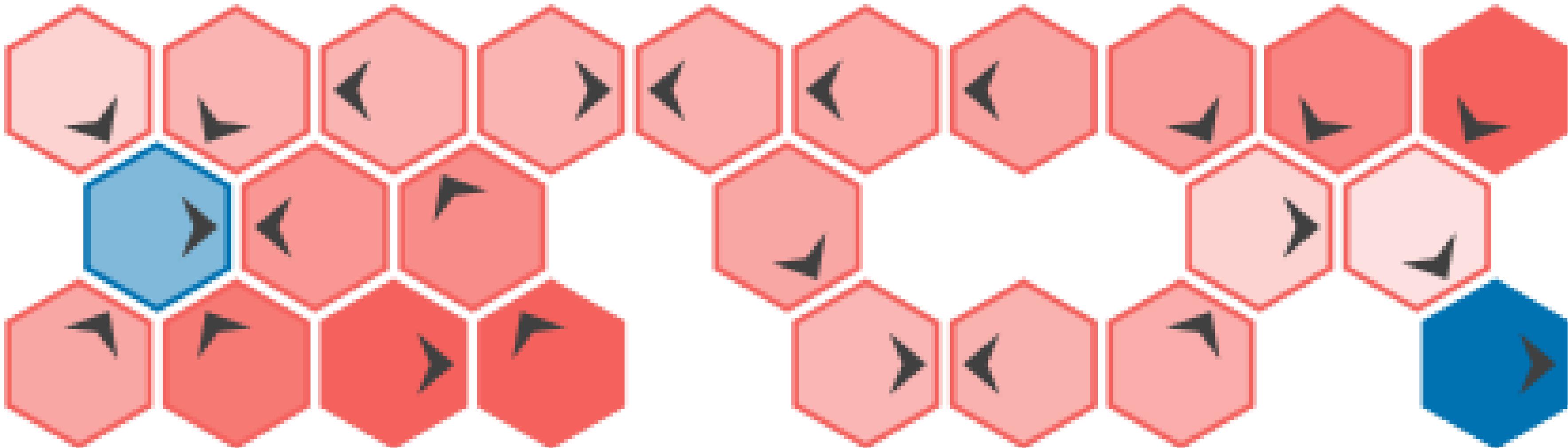


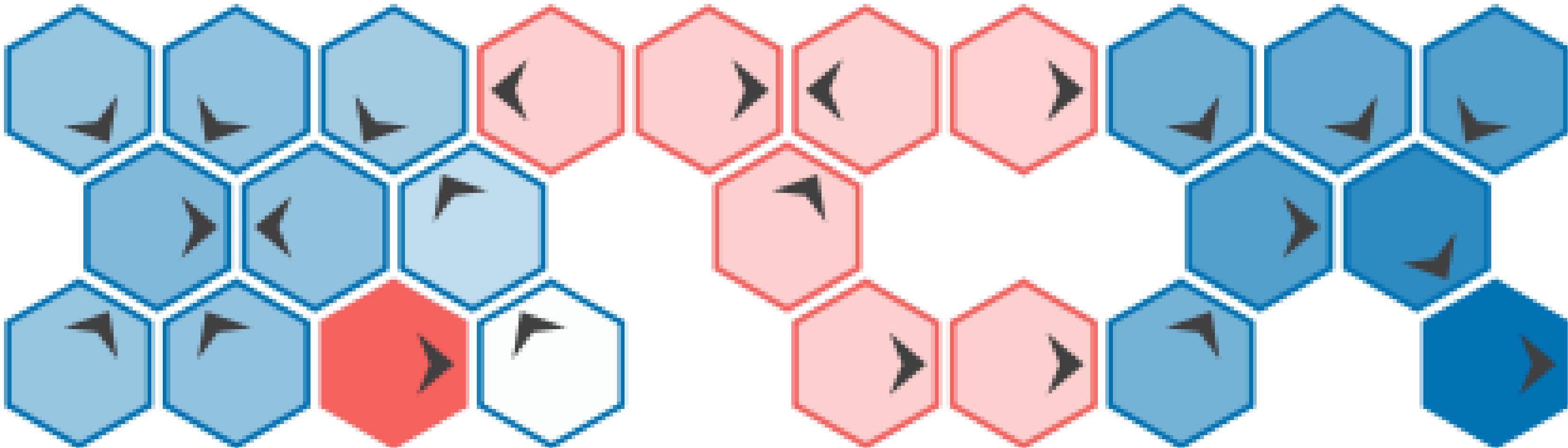


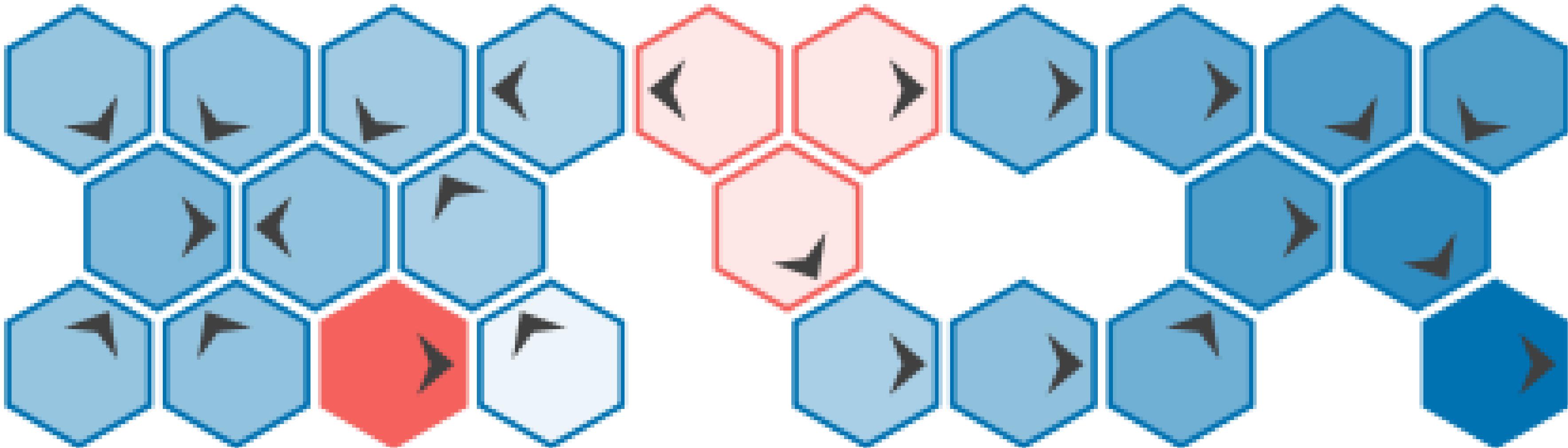


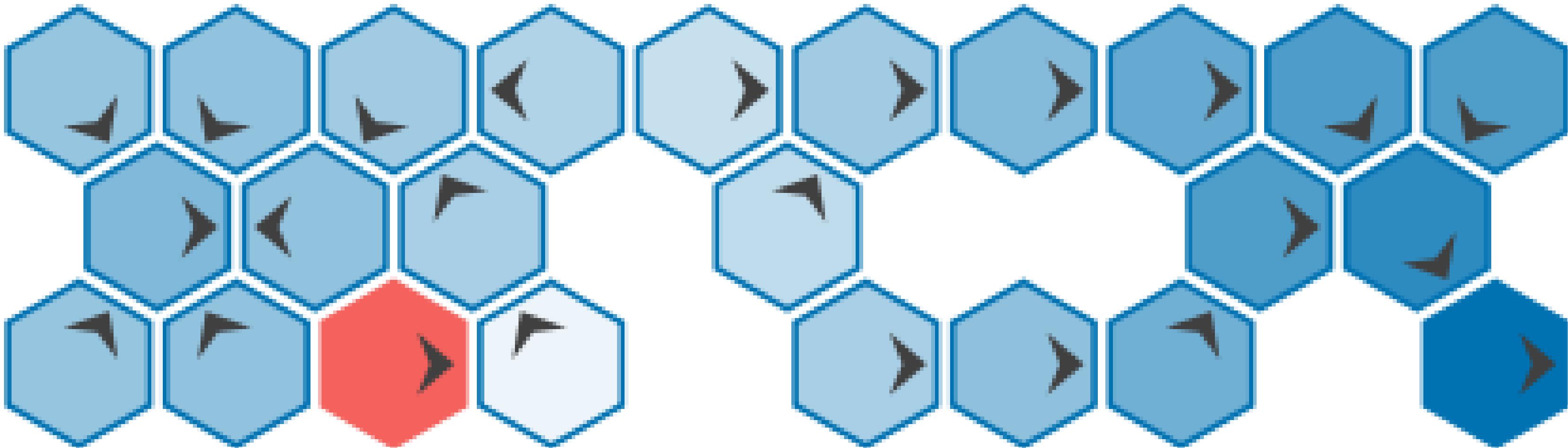


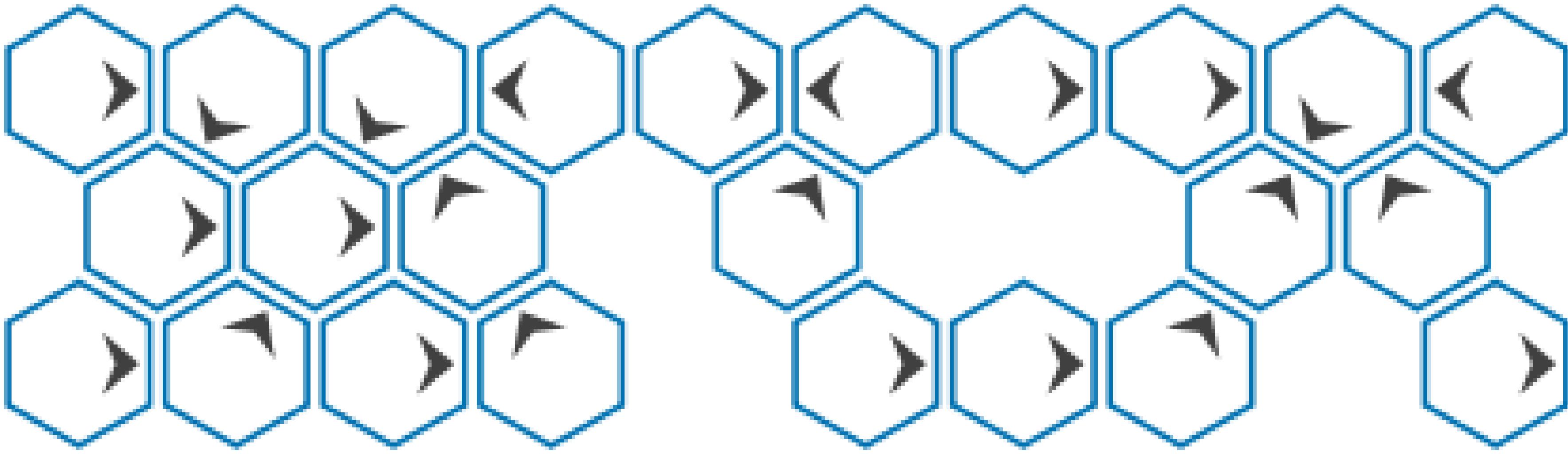


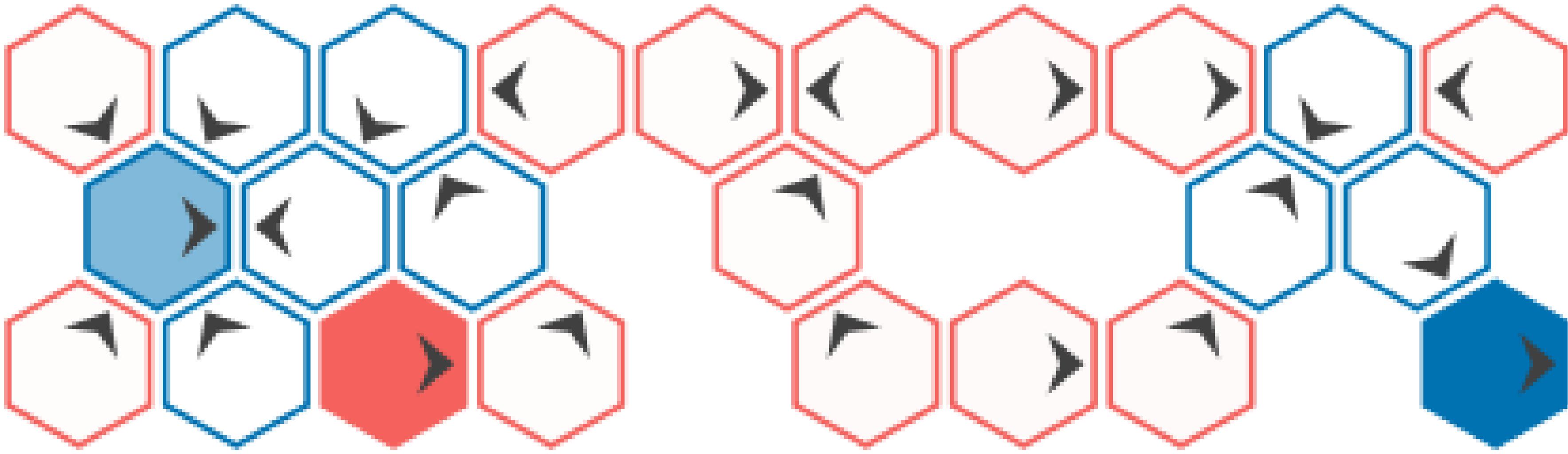


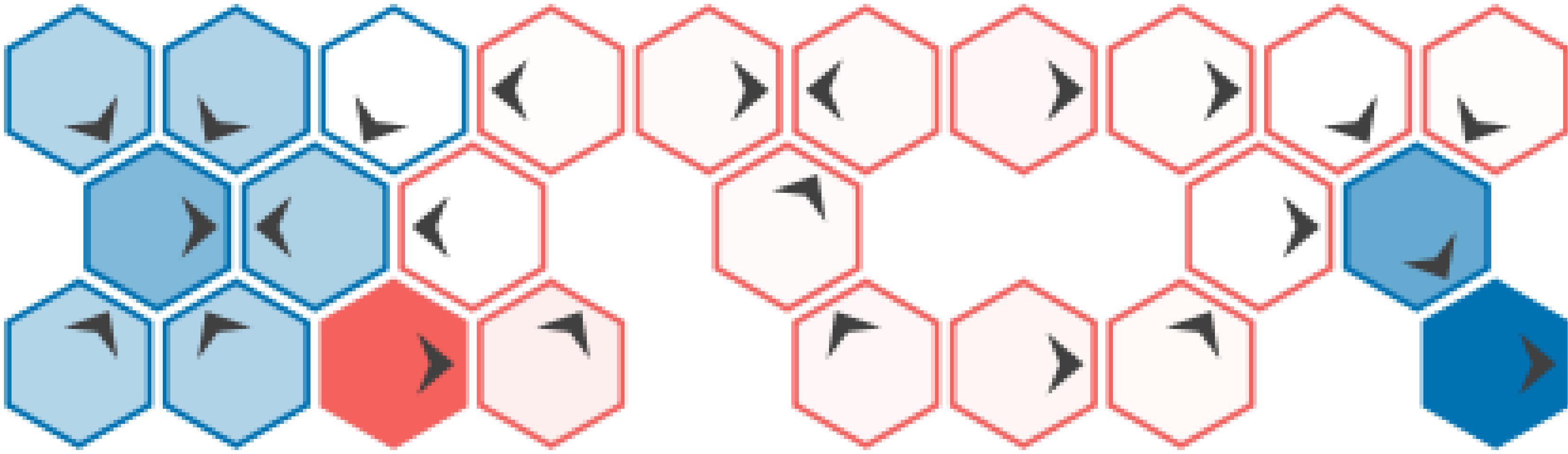


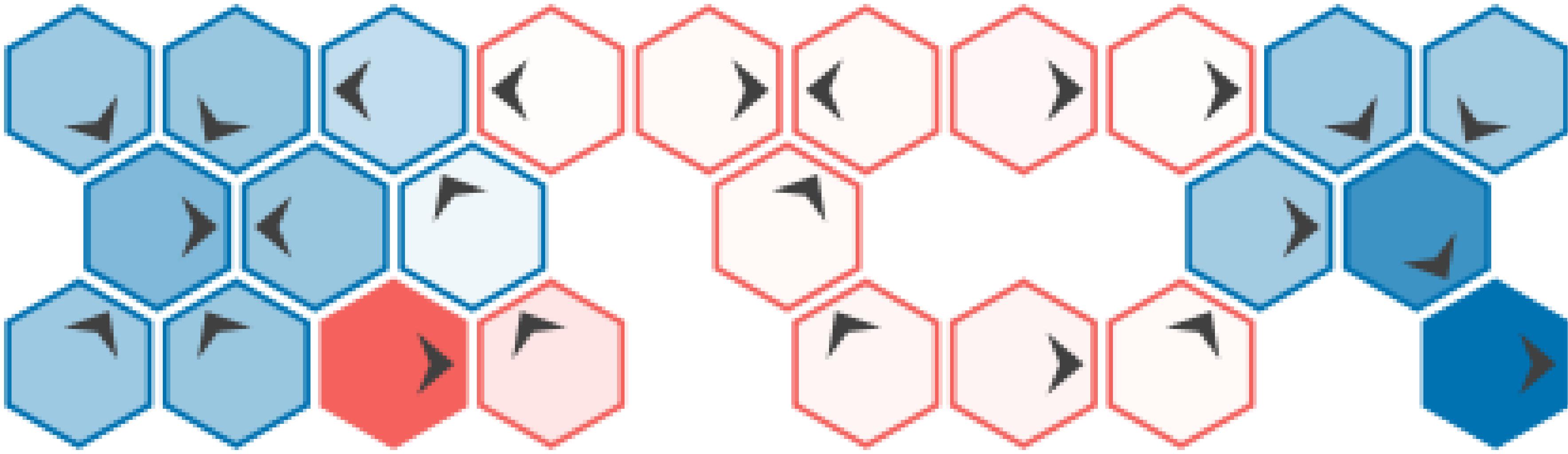


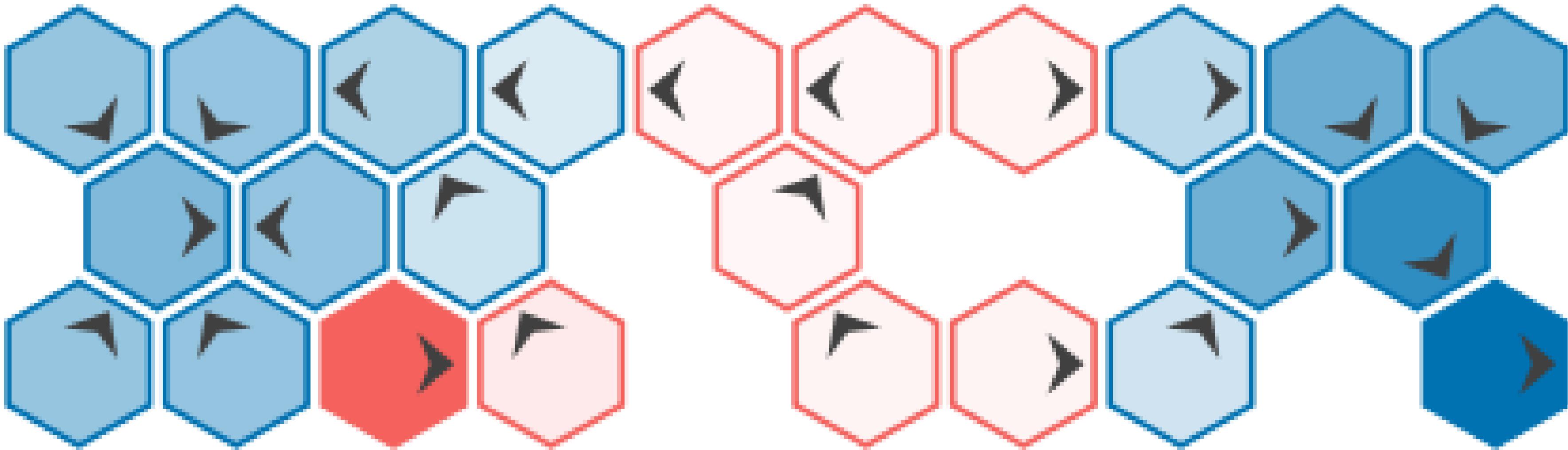


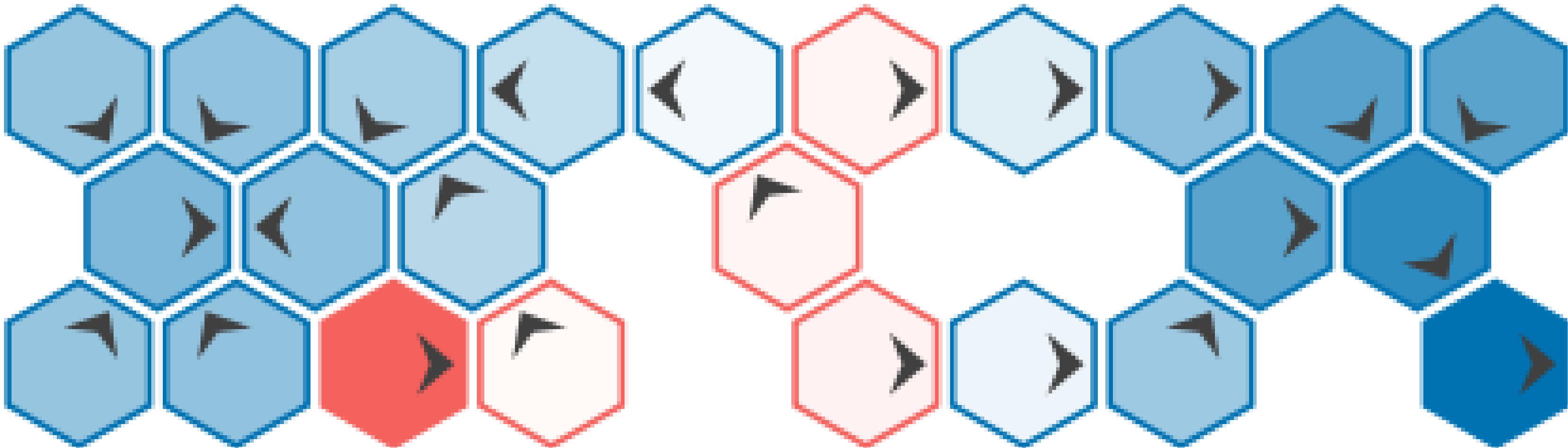


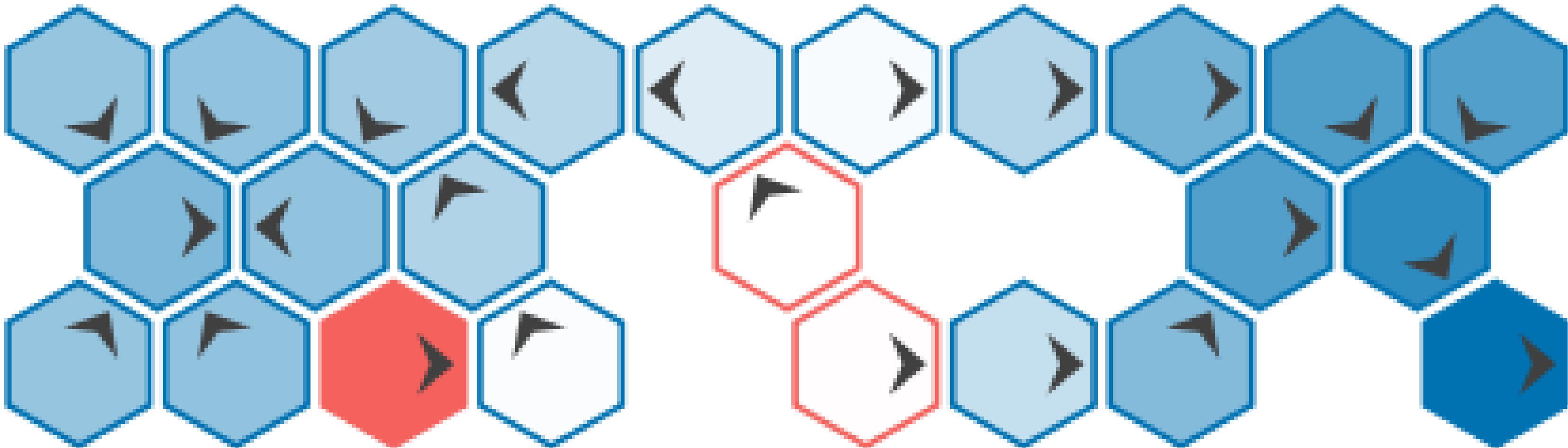


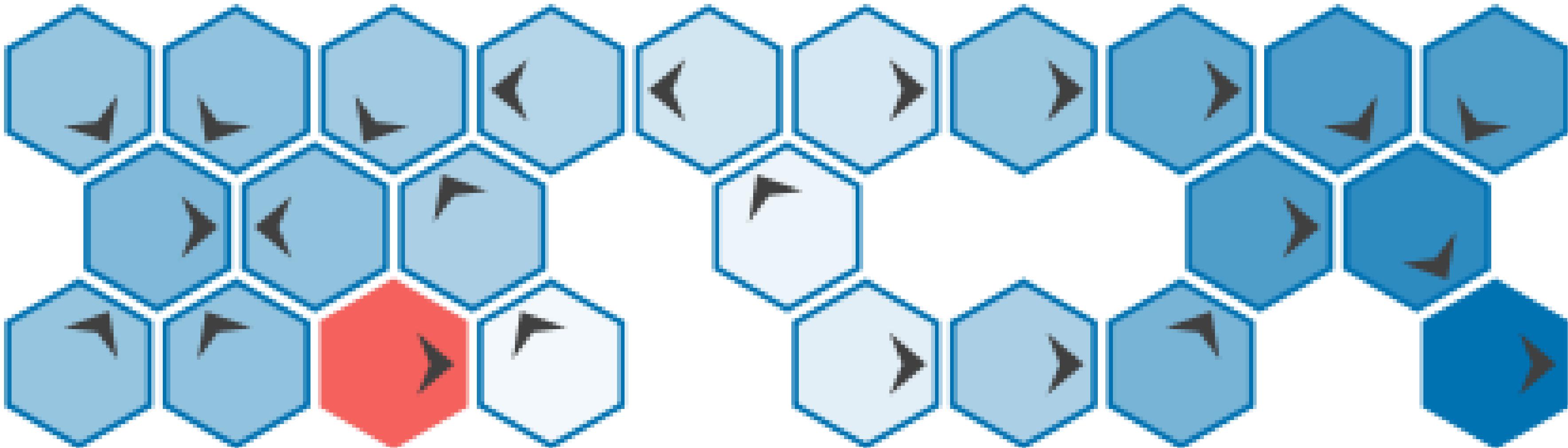












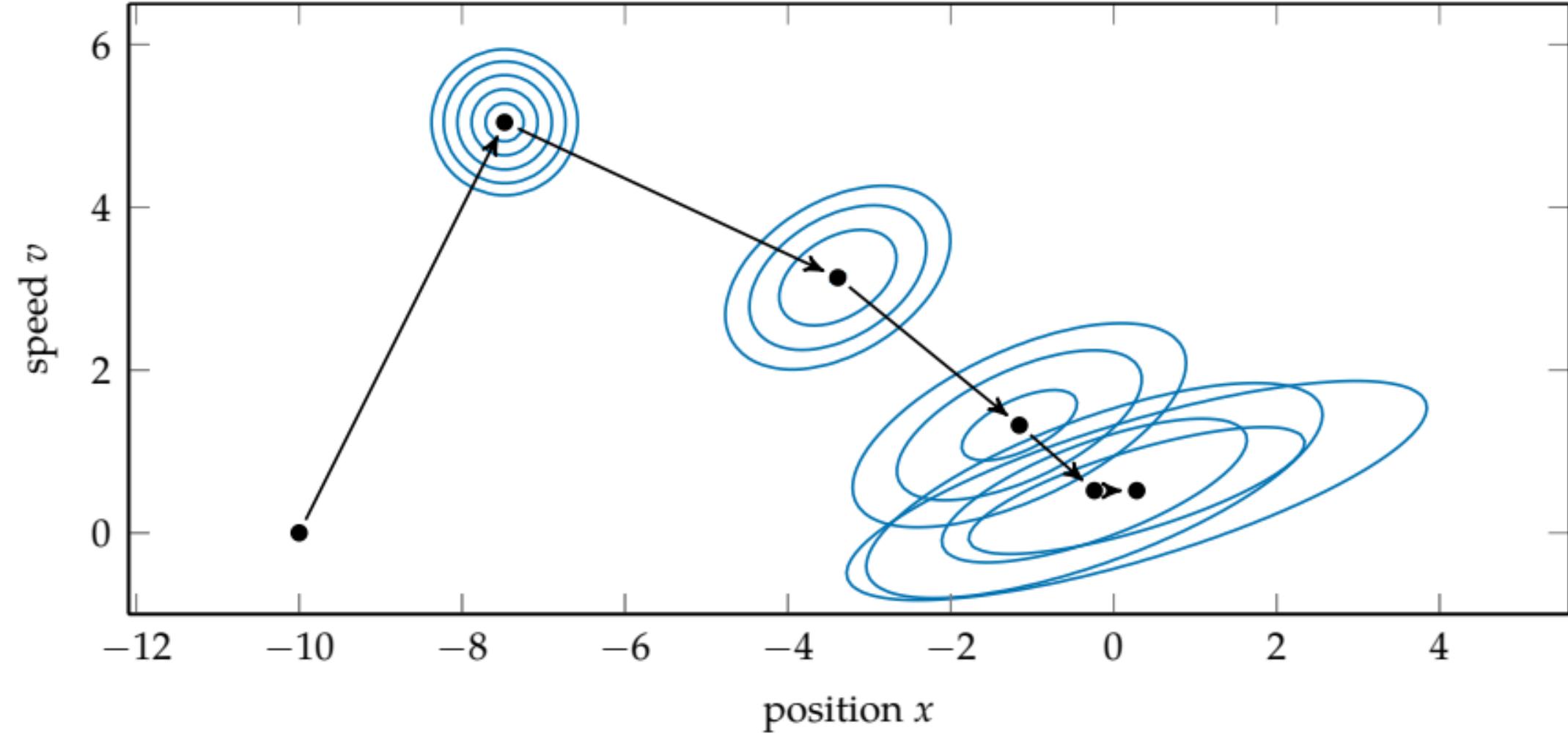
$\gamma = 0.9$ $\gamma = 0.5$ 

left to right

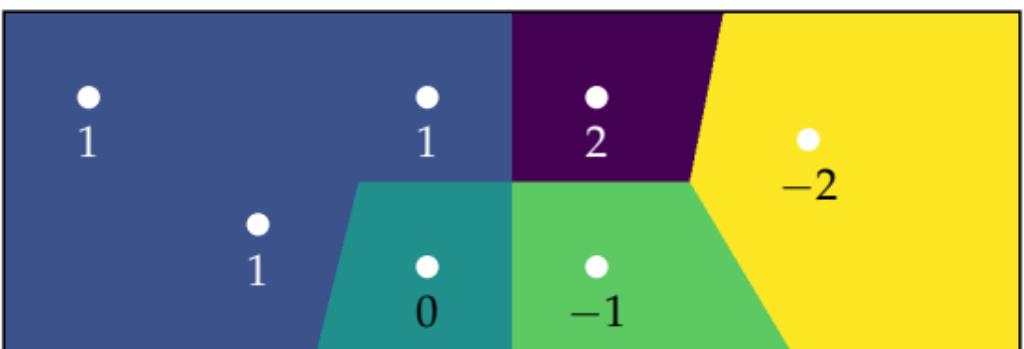


right to left

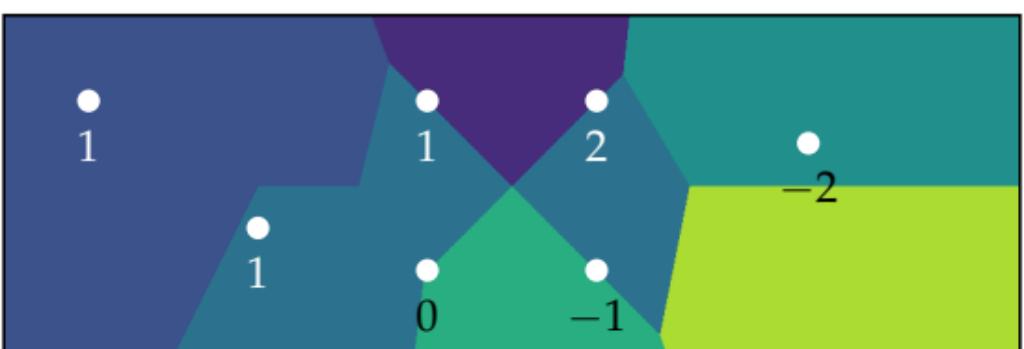




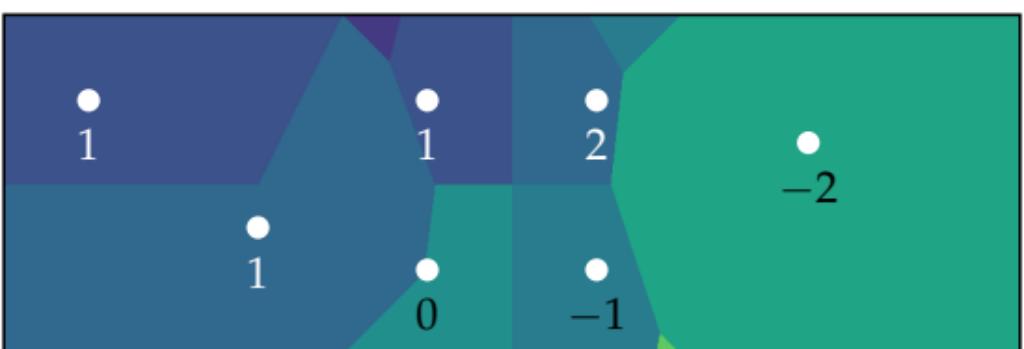
nearest neighbor ($k = 1$)



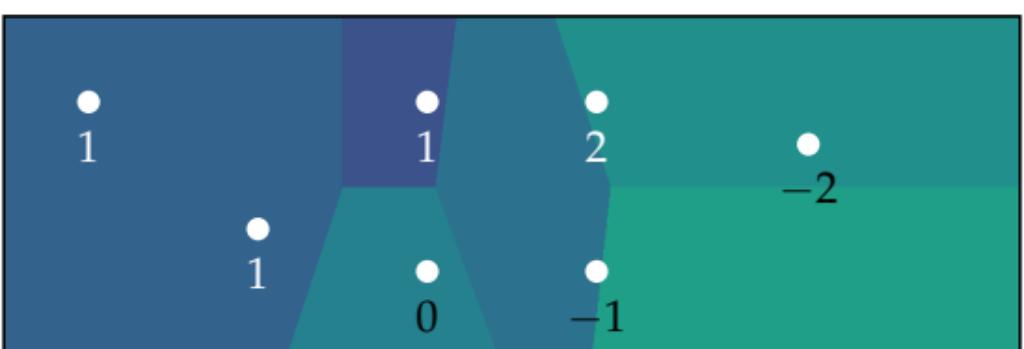
$k = 2$



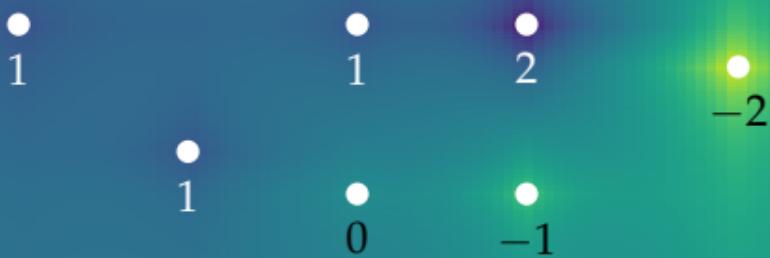
$k = 3$



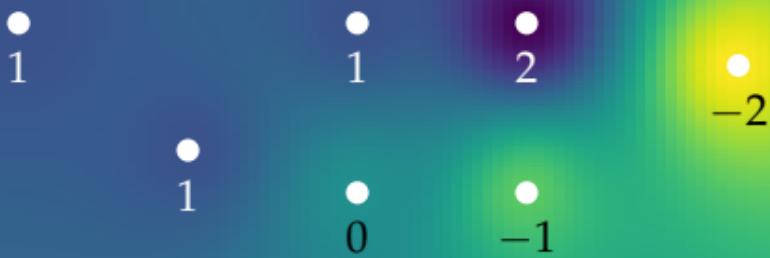
$k = 4$



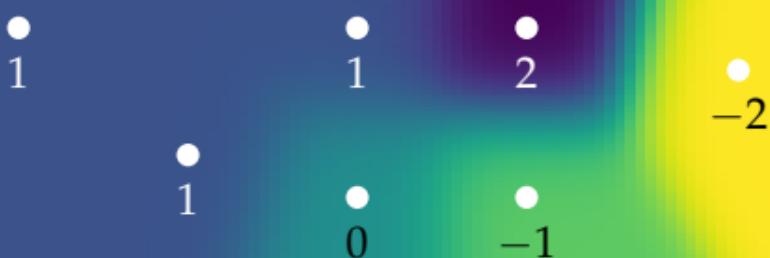
$$d(\mathbf{s}, \mathbf{s}') = \|\mathbf{s} - \mathbf{s}'\|_1$$

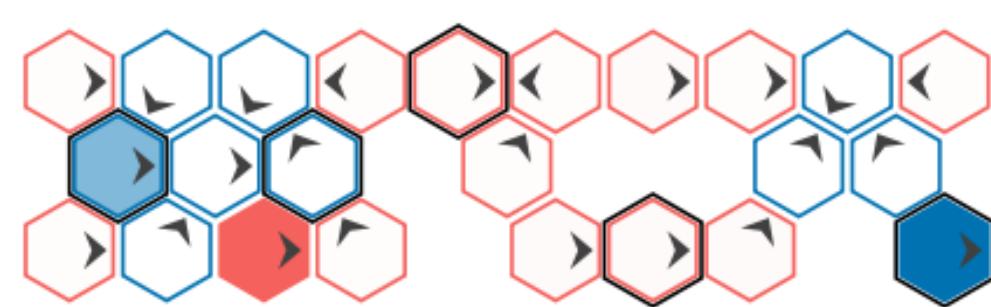


$$d(\mathbf{s}, \mathbf{s}') = \|\mathbf{s} - \mathbf{s}'\|_2^2$$

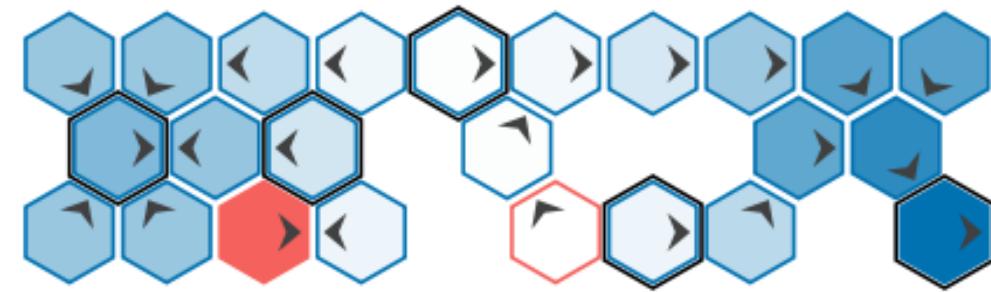


$$d(\mathbf{s}, \mathbf{s}') = \exp(-\|\mathbf{s} - \mathbf{s}'\|_2^2)$$

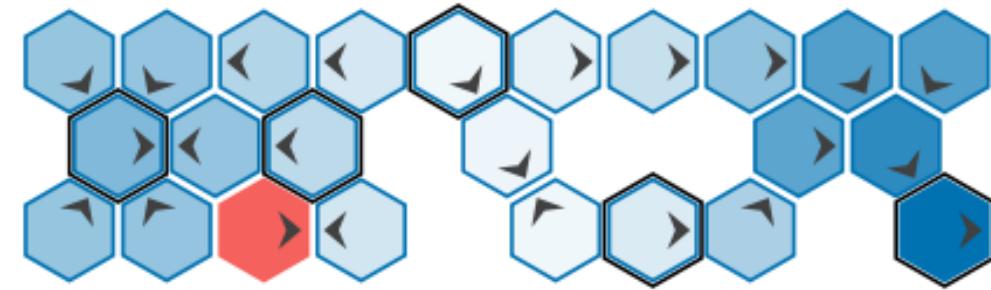




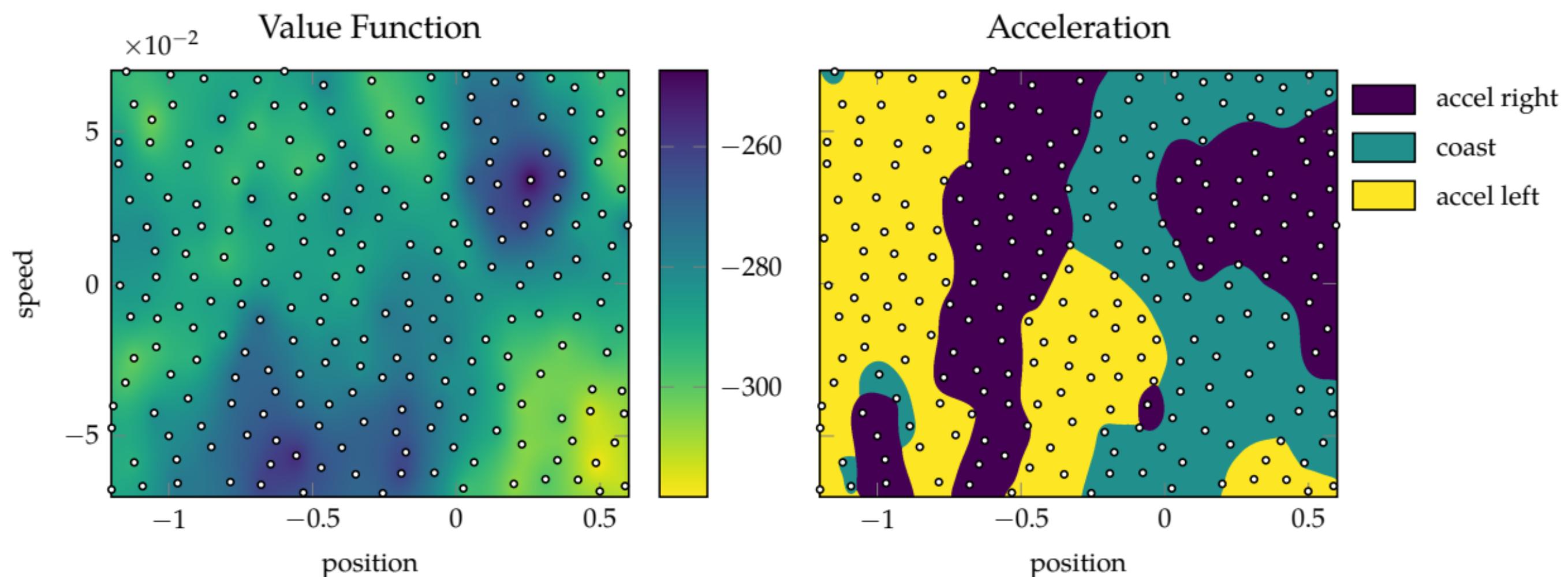
Initial value function $U^{(1)}$

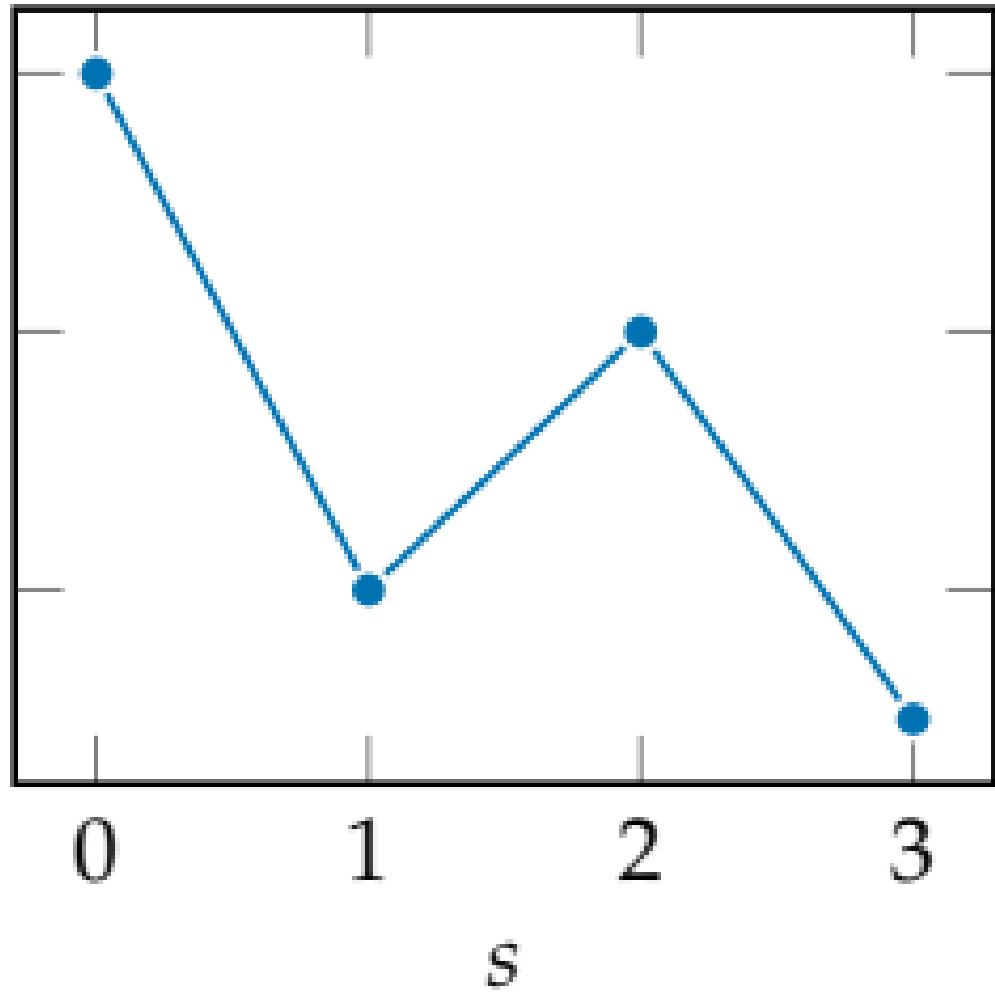


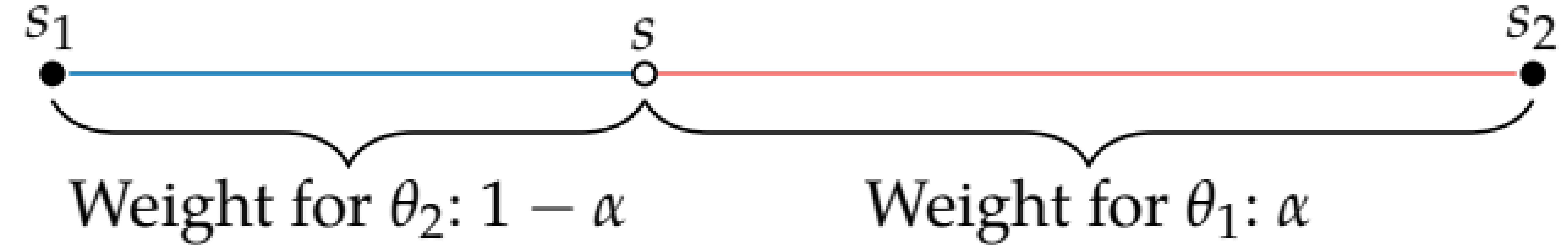
Iteration 2

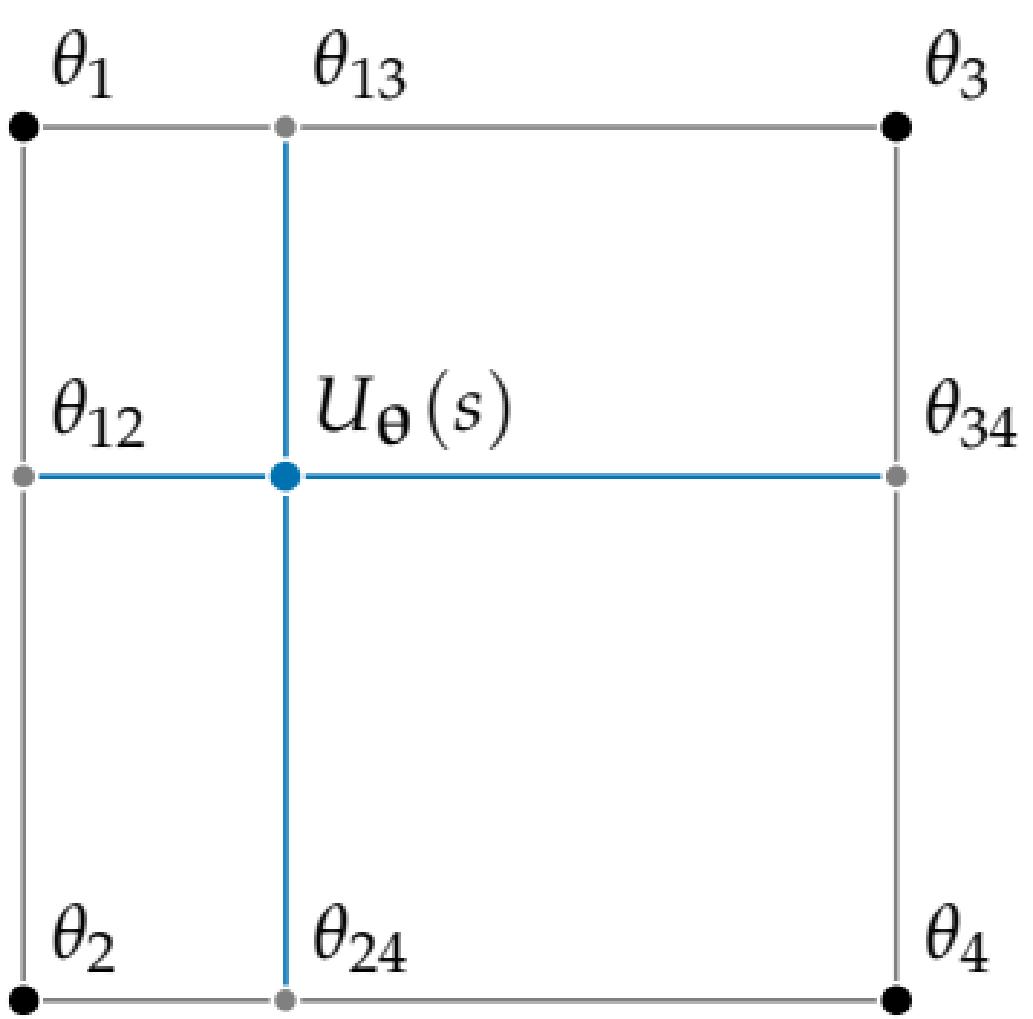


Iteration 3



$U(s)$ c_2 2 1 0 1 2 3 s 





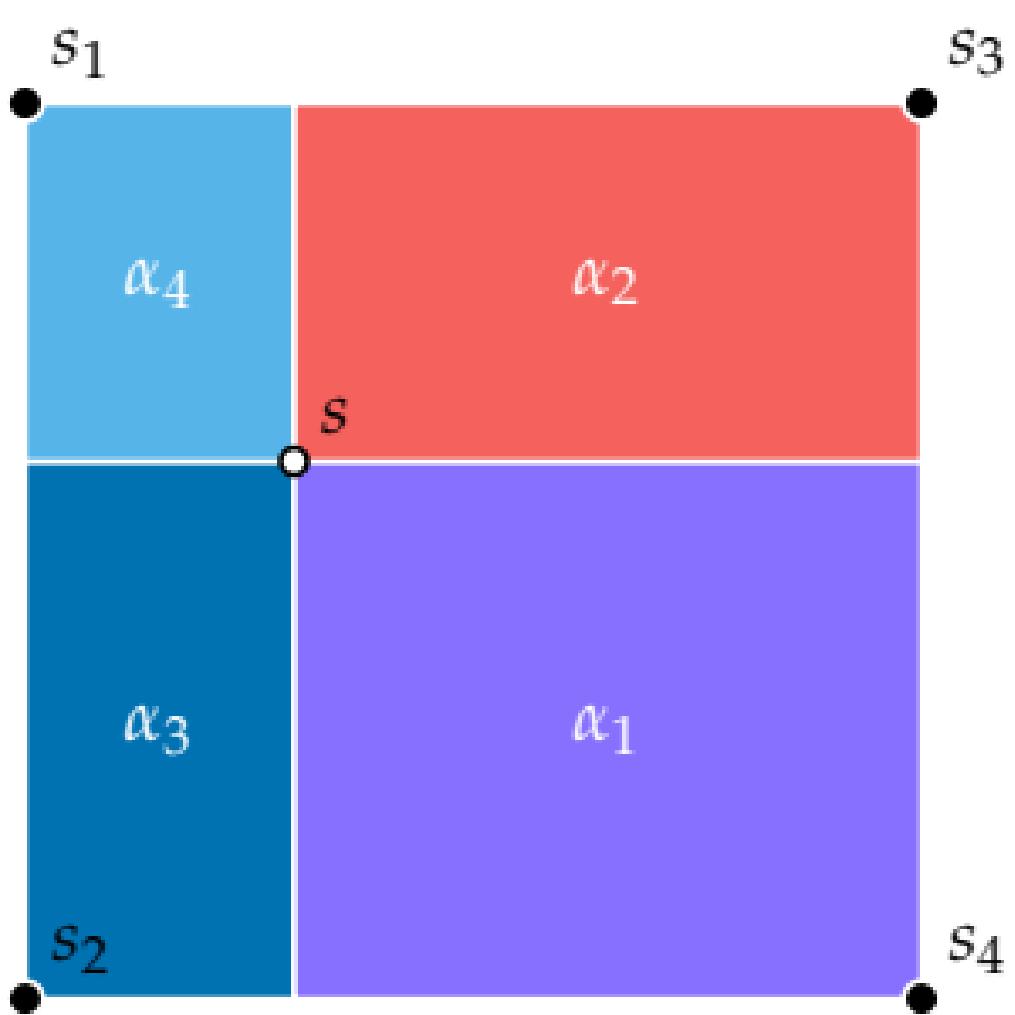
$\theta_{12} = \text{1D interpolation between } \theta_1 \text{ and } \theta_2 \text{ along the vertical axis}$

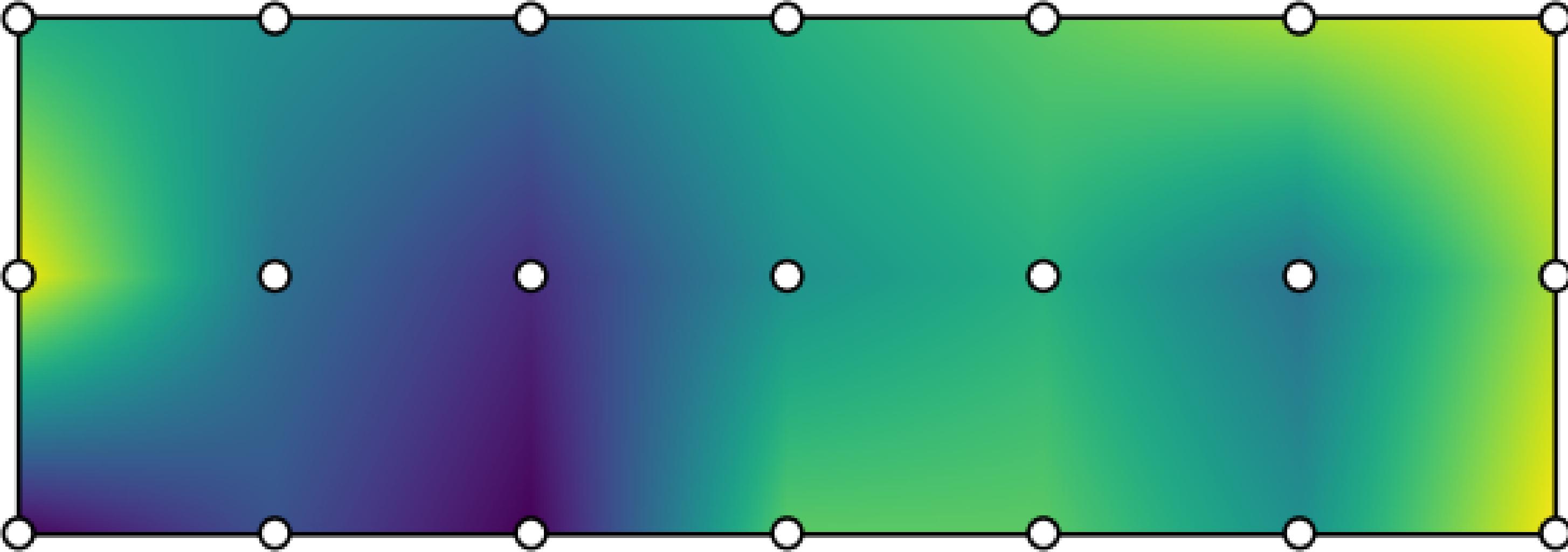
$\theta_{24} = \text{1D interpolation between } \theta_2 \text{ and } \theta_4 \text{ along the horizontal axis}$

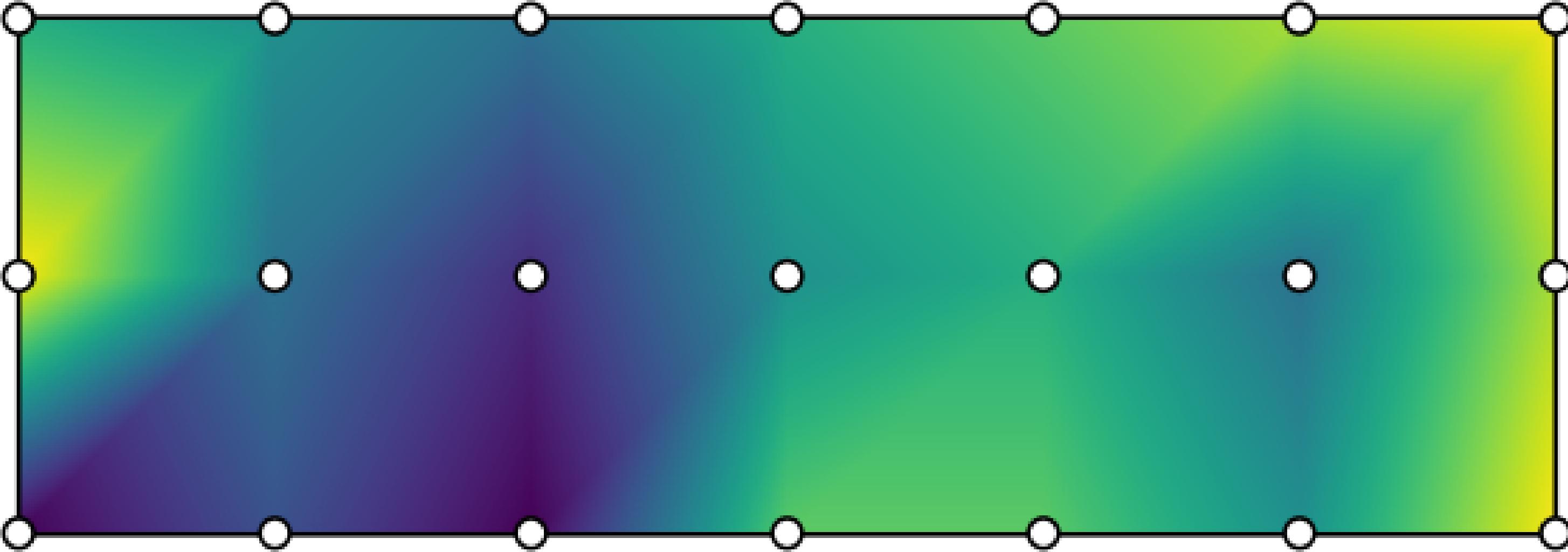
$\theta_{13} = \text{1D interpolation between } \theta_1 \text{ and } \theta_3 \text{ along the horizontal axis}$

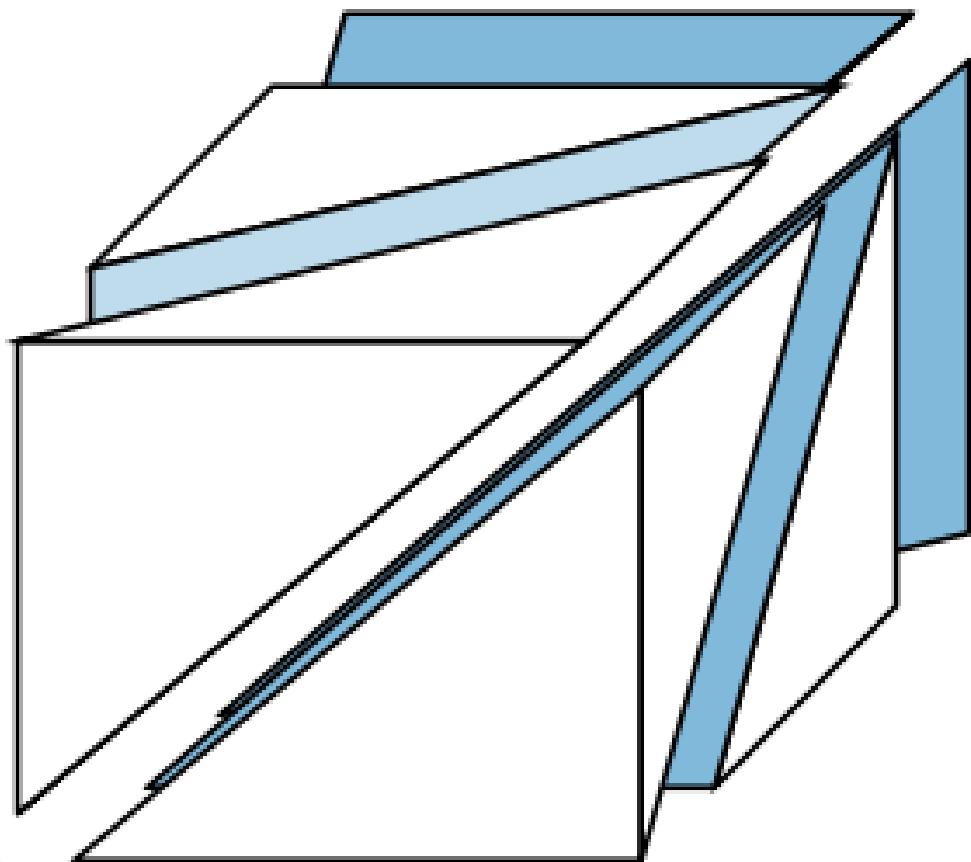
$\theta_{34} = \text{1D interpolation between } \theta_3 \text{ and } \theta_4 \text{ along the vertical axis}$

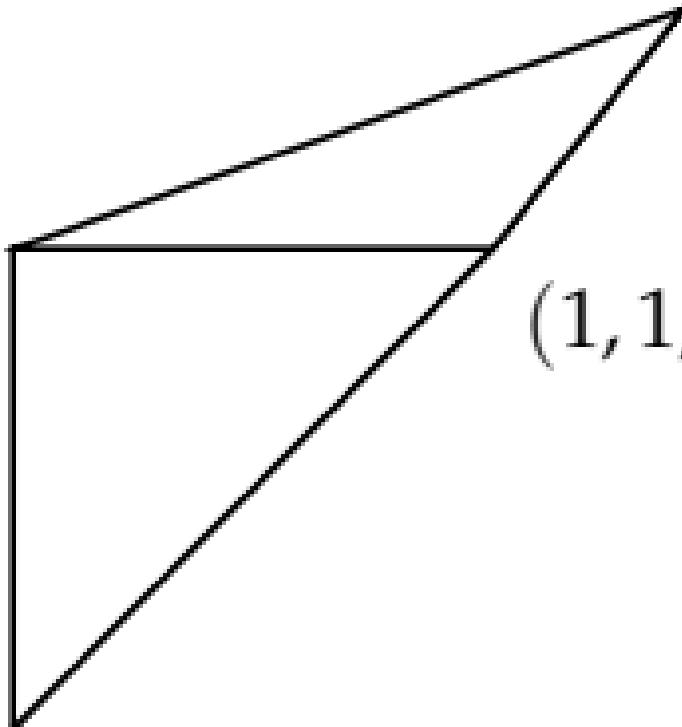
$$U_\theta(s) = \begin{cases} \text{1D interpolation between } \theta_{12} \text{ and } \theta_{34} \text{ along the horizontal axis} \\ \quad \text{or} \\ \text{1D interpolation between } \theta_{13} \text{ and } \theta_{24} \text{ along the vertical axis} \end{cases}$$

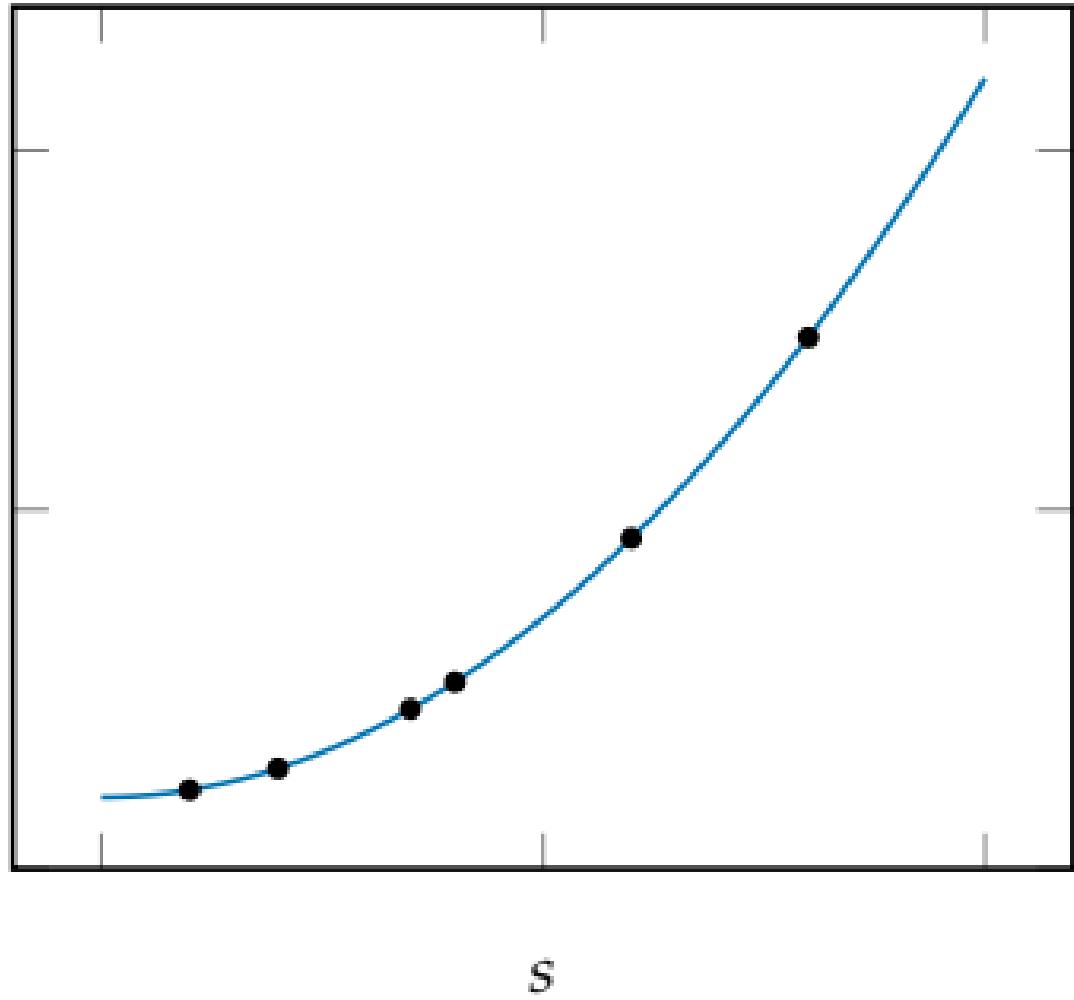
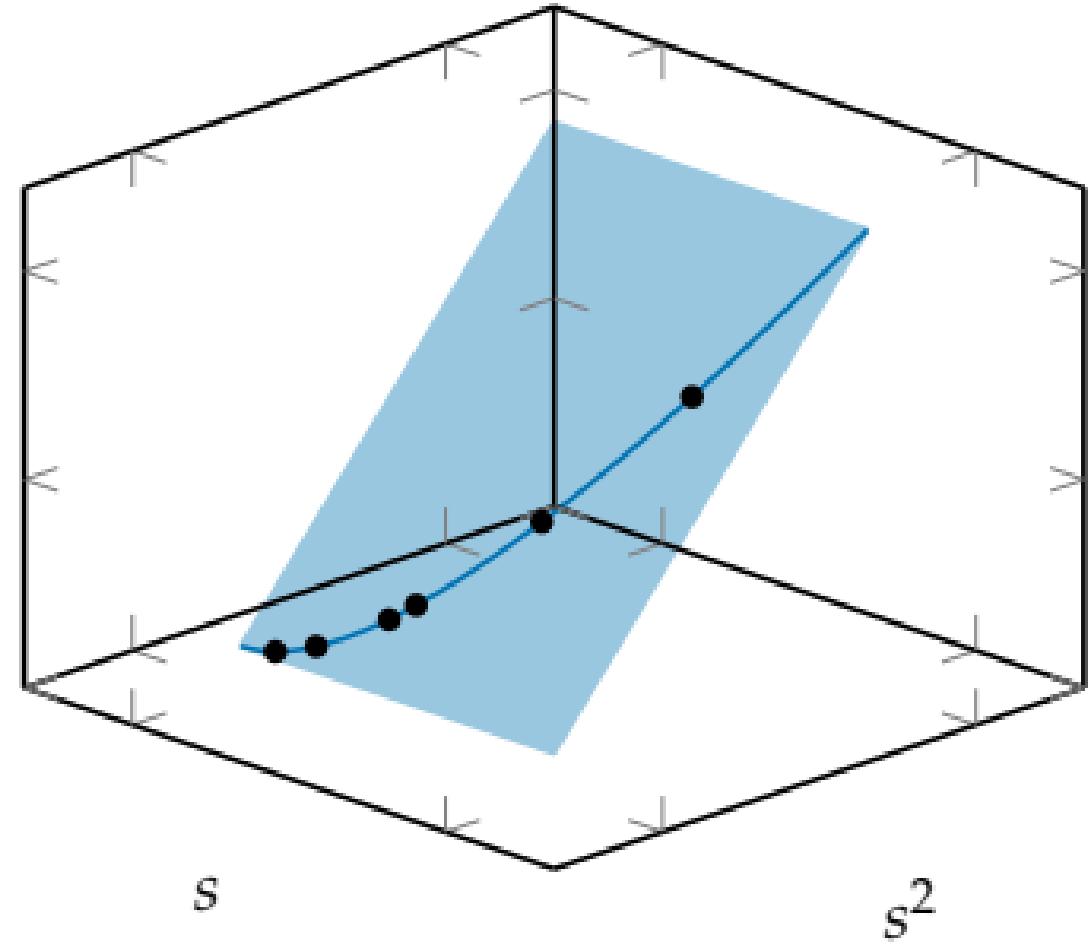






$(1, 1, 1)$ $(0, 0, 0)$ 

$(1, 1, 1)$ $(0, 1, 0)$ $(1, 1, 0)$ $(0, 0, 0)$ 

$U_\Theta(s)$  $U_\Theta(s)$ 

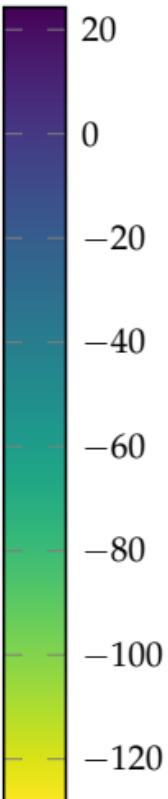
$U_\theta(s)$ $\times 10^{-2}$

speed

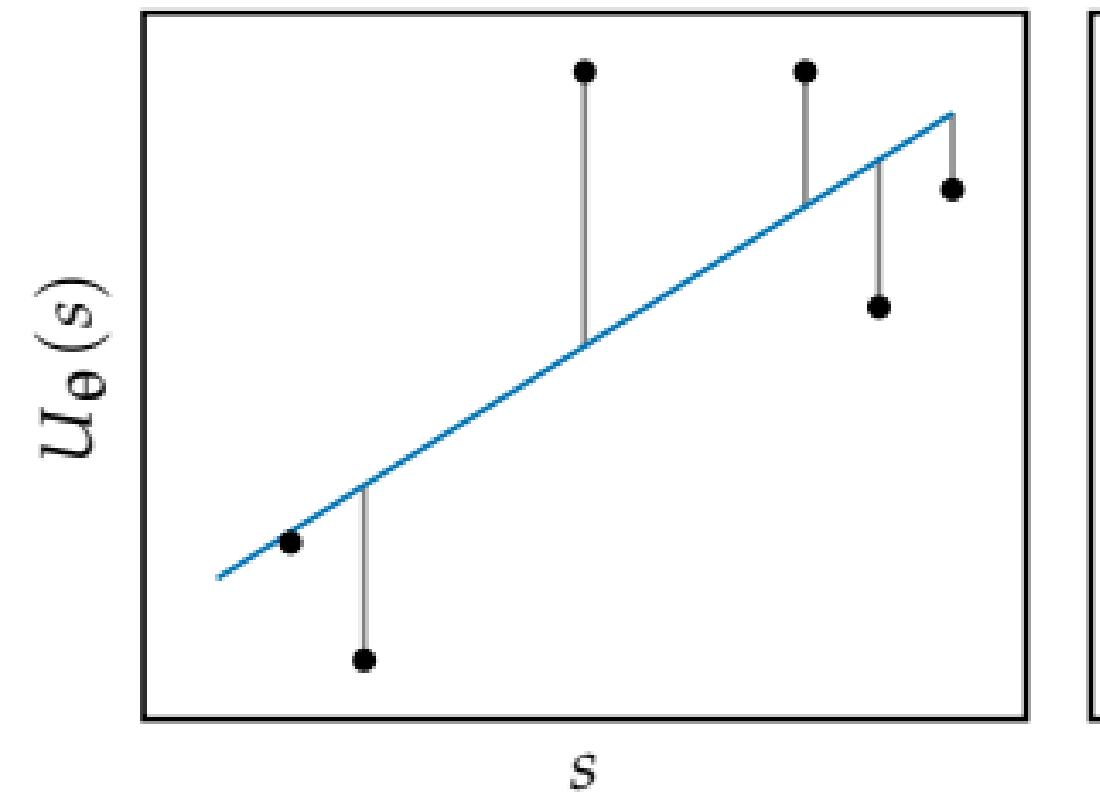
6
4
2
0
-2
-4
-6

-1.2 -1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6

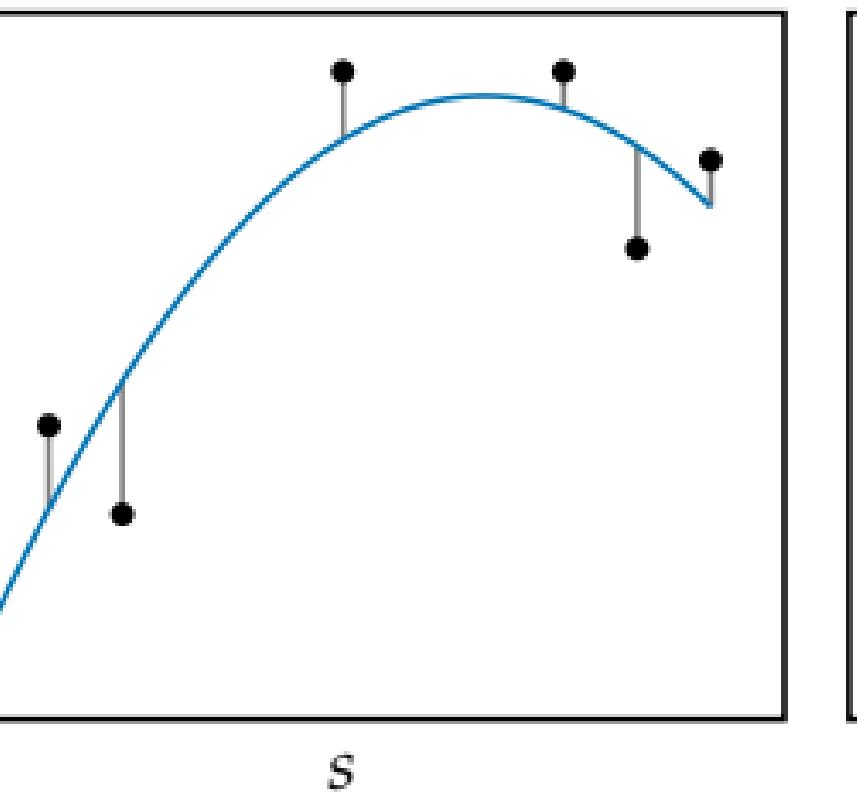
position



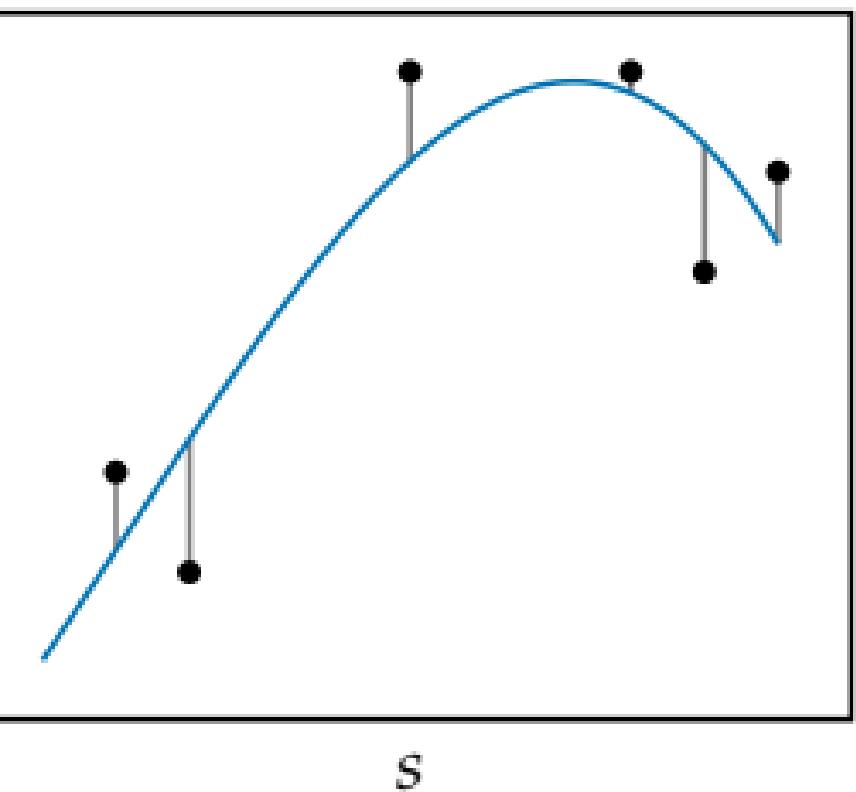
linear



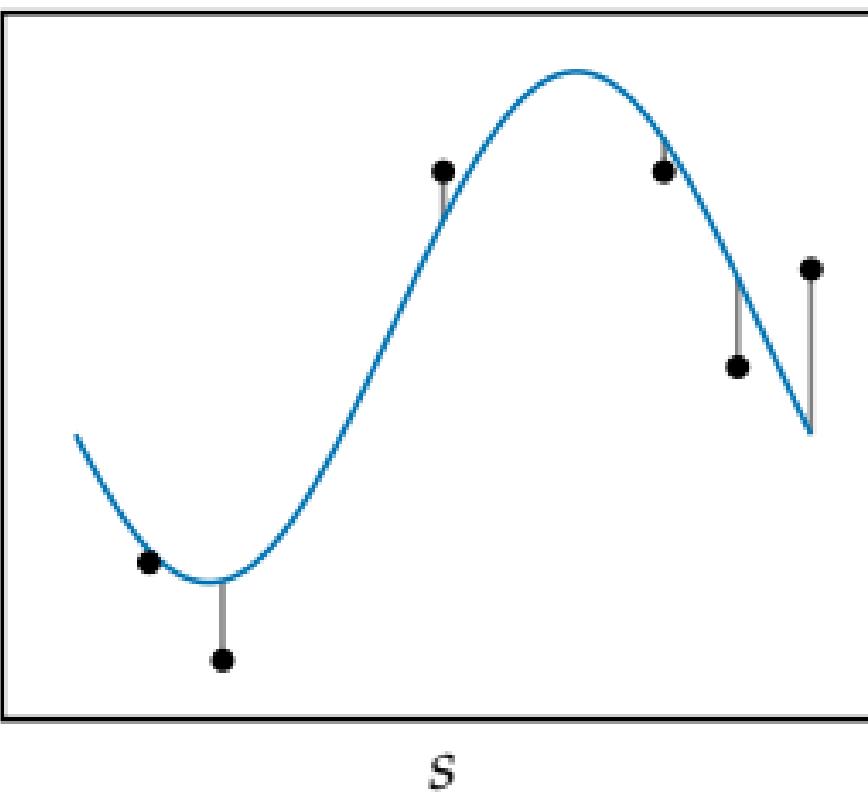
quadratic

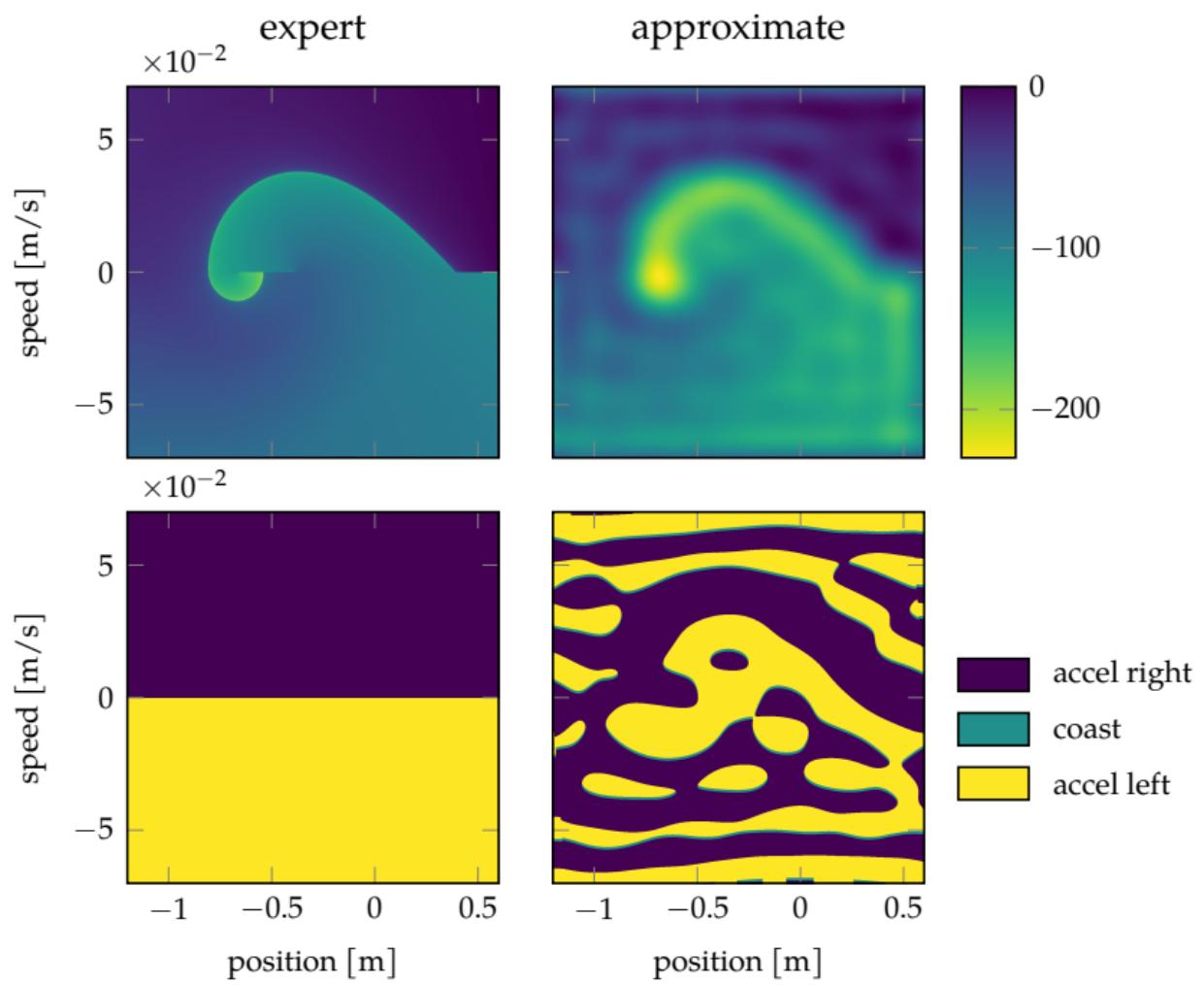


cubic

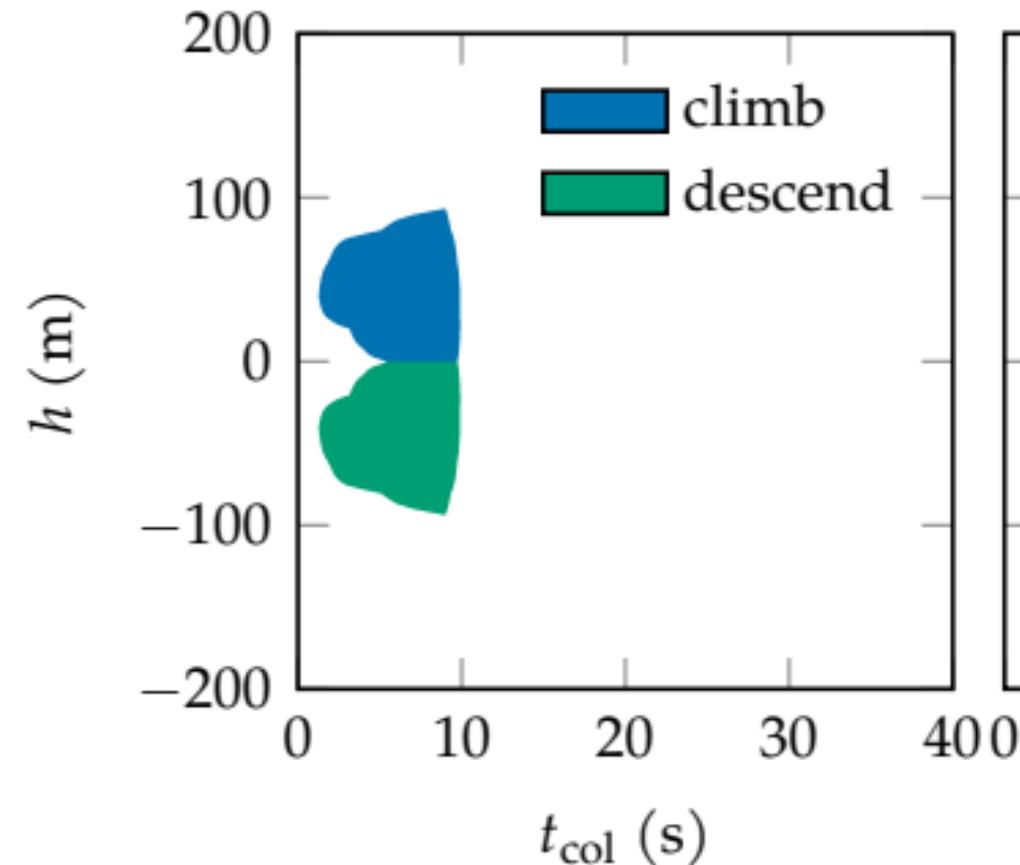


sinusoidal

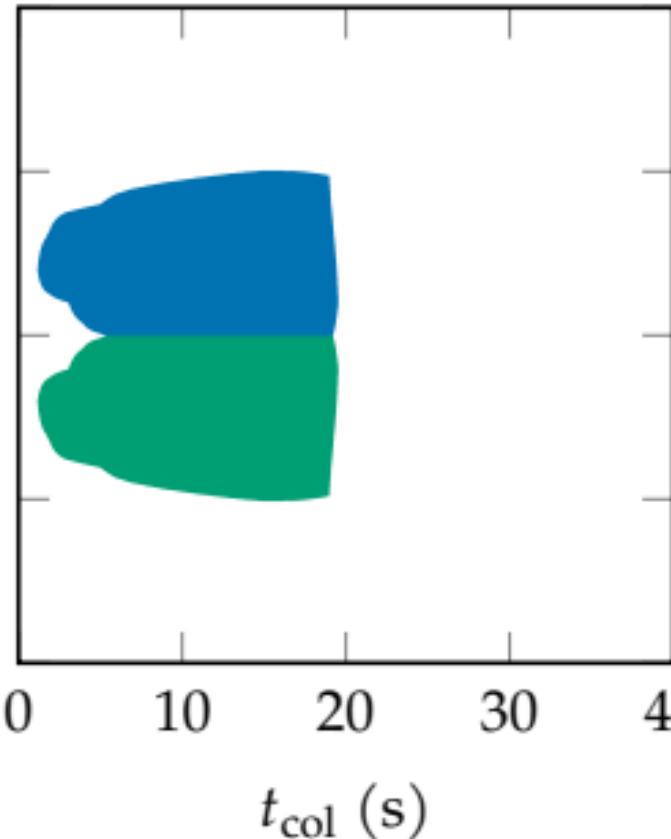




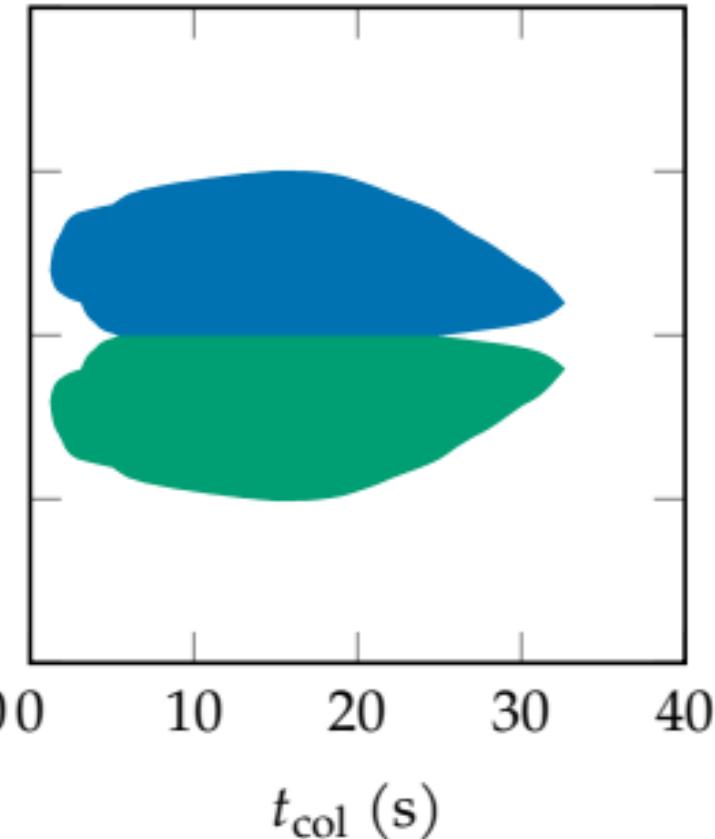
Horizon 10

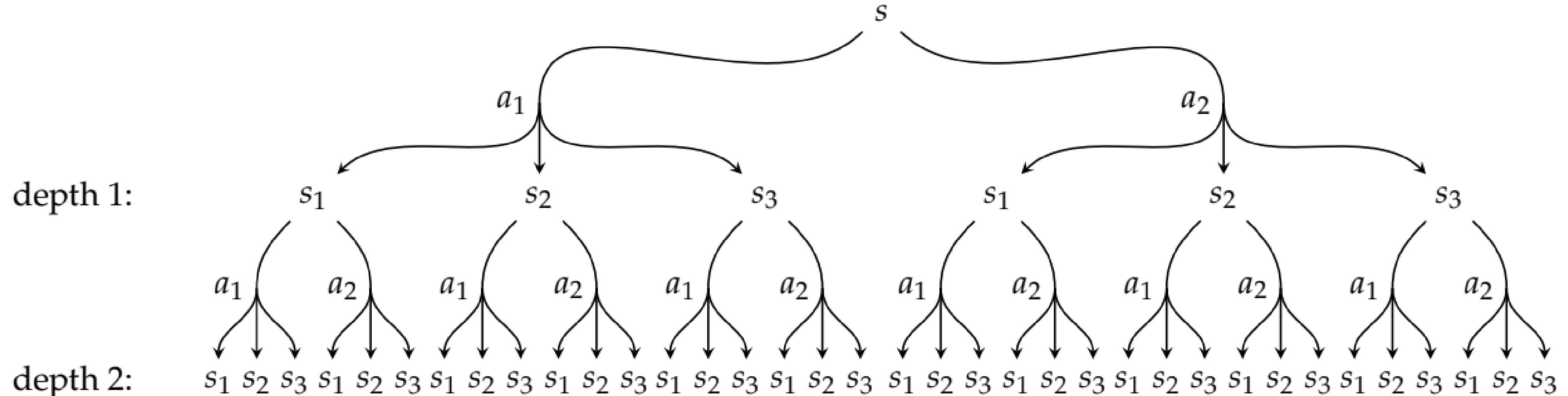


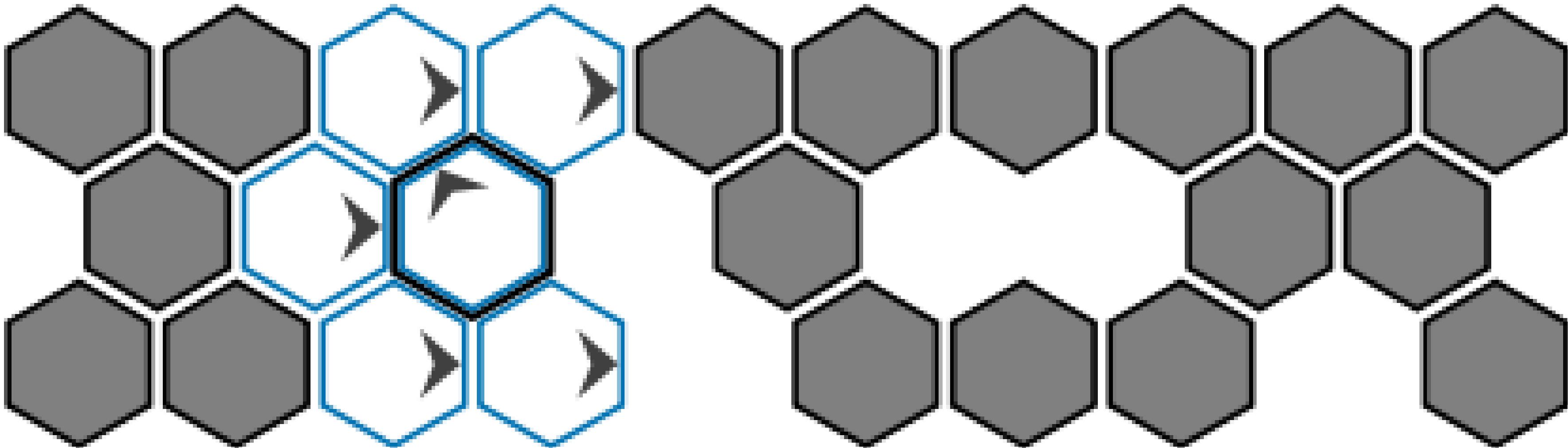
Horizon 20

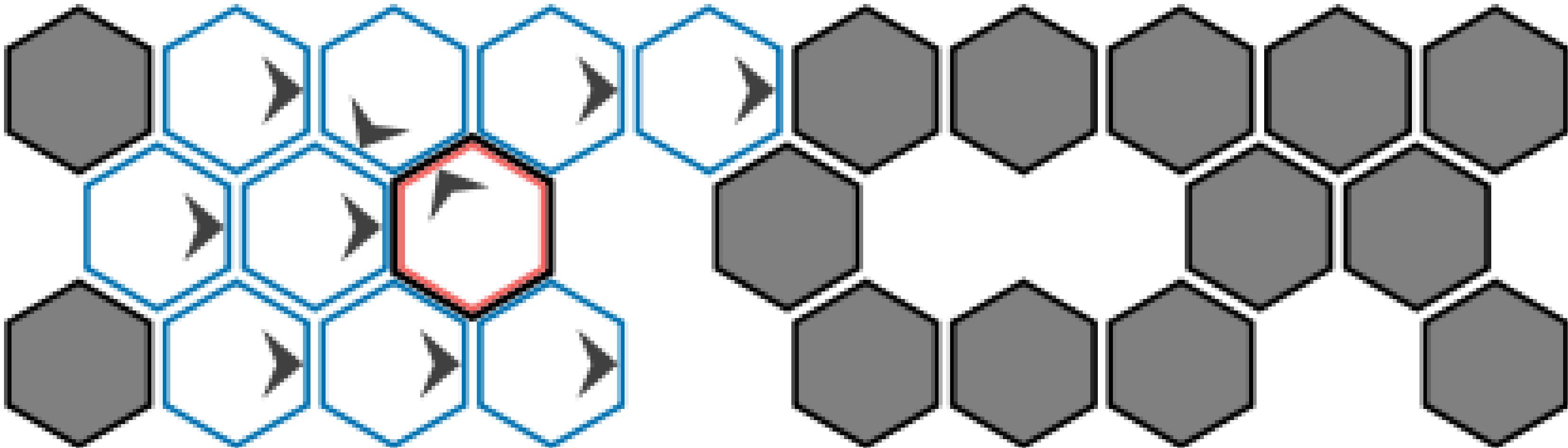


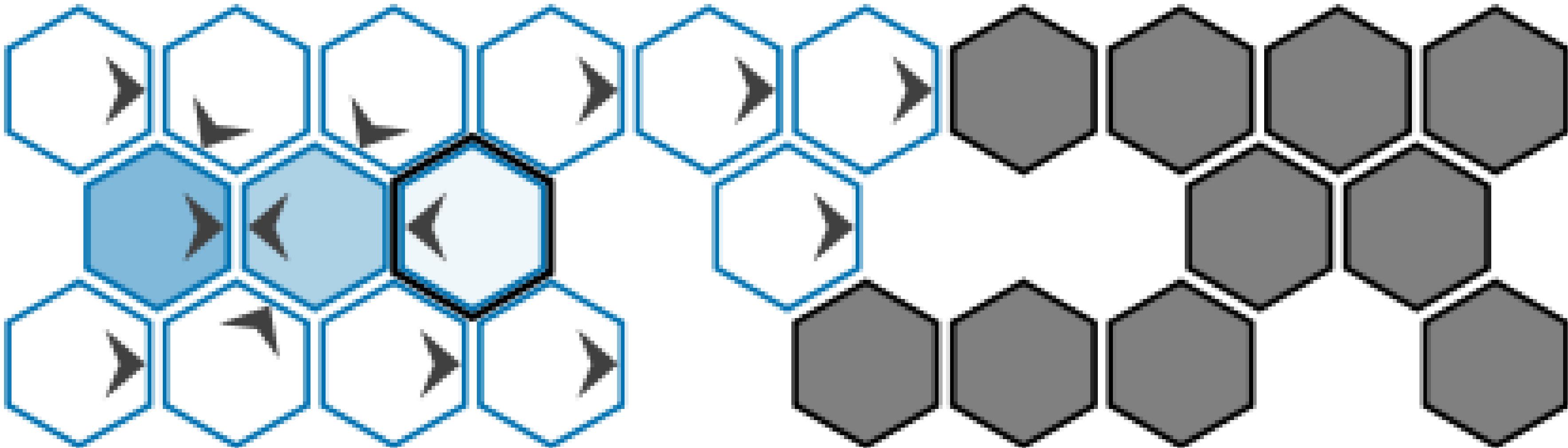
Horizon 40

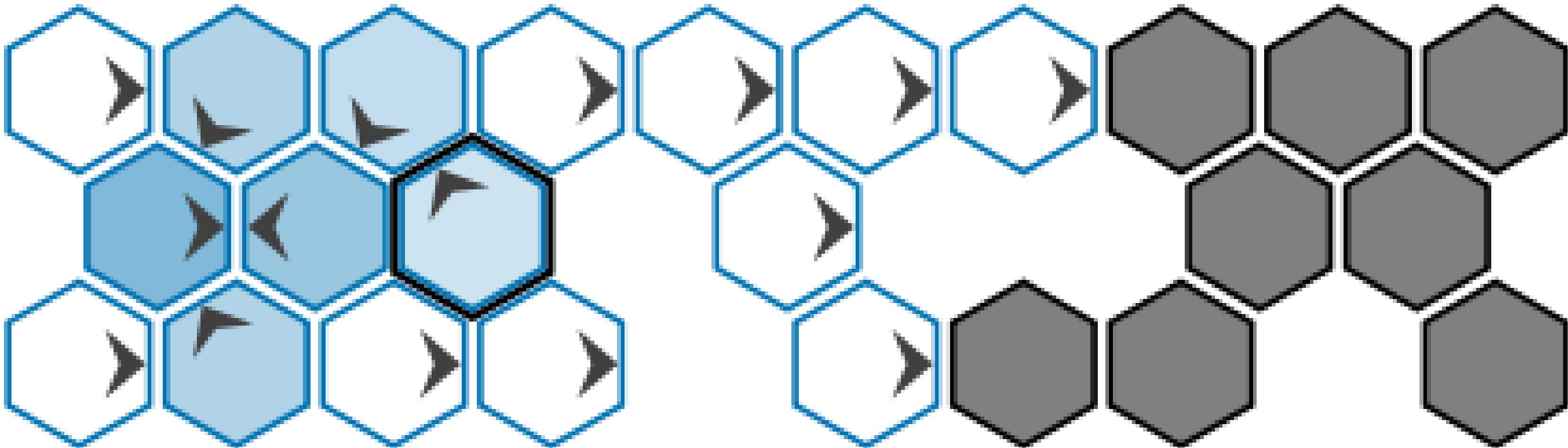


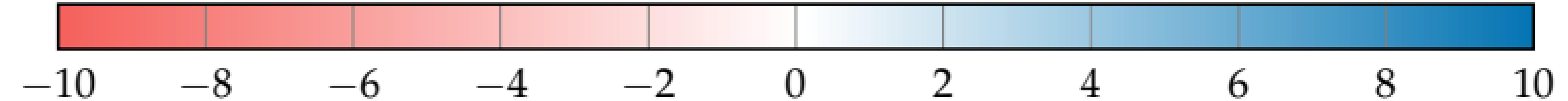


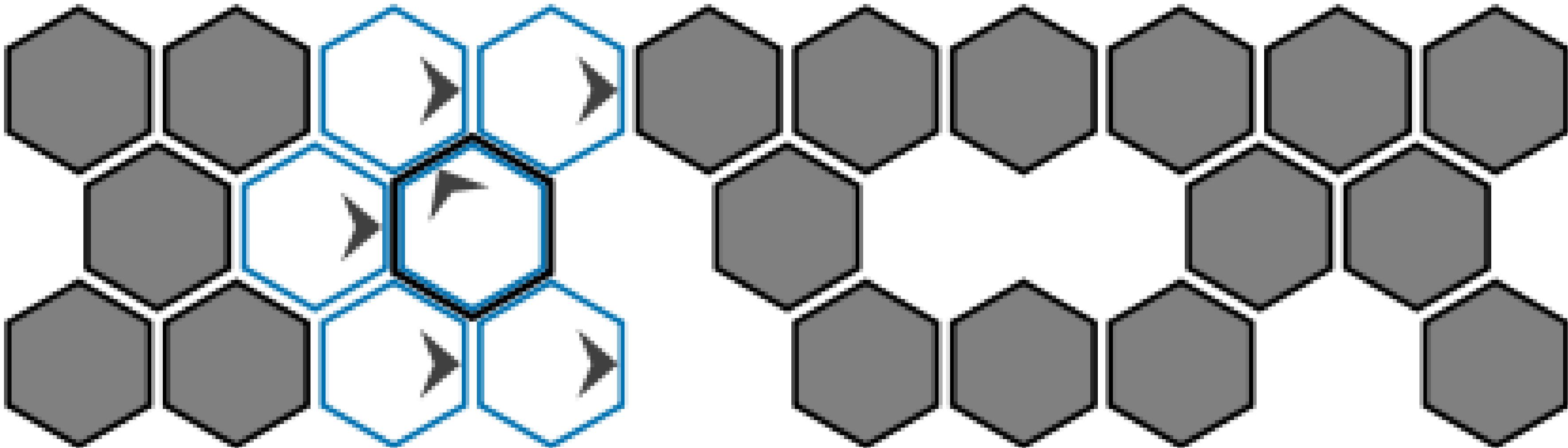


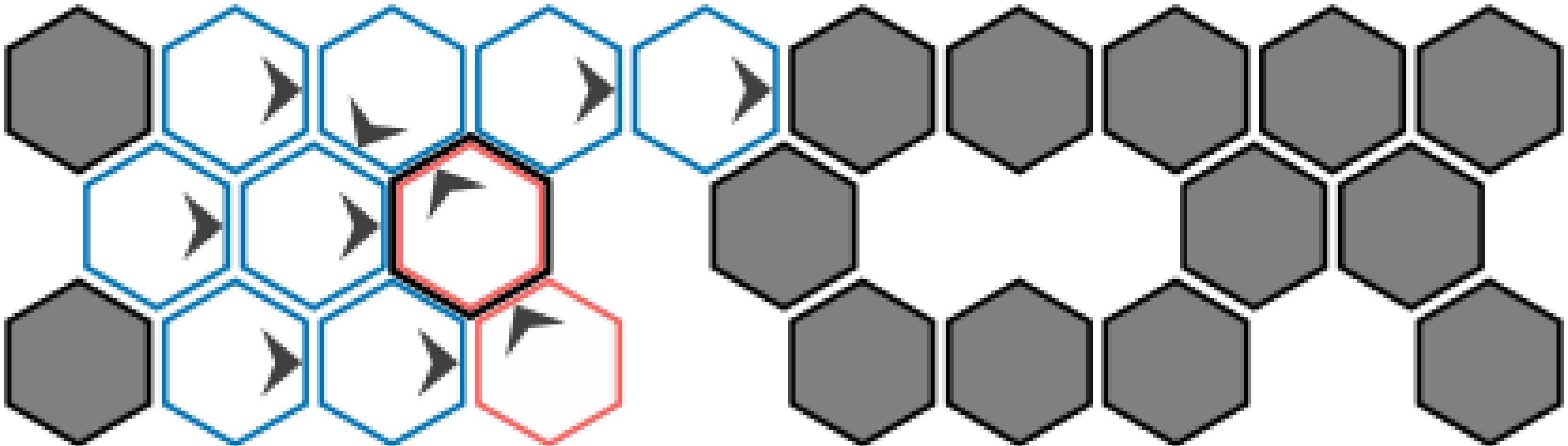


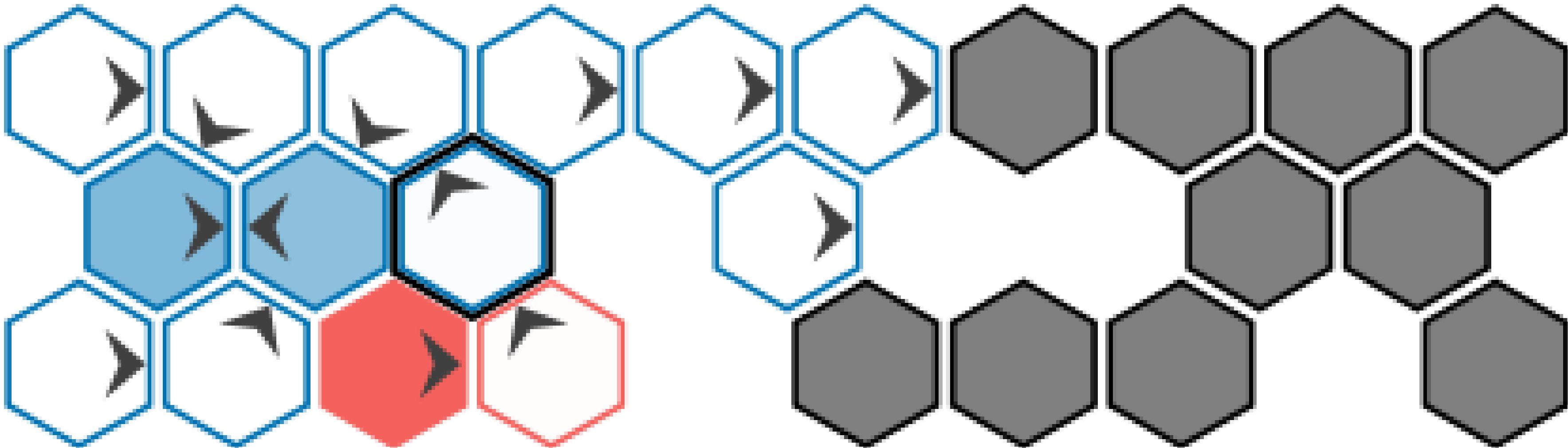


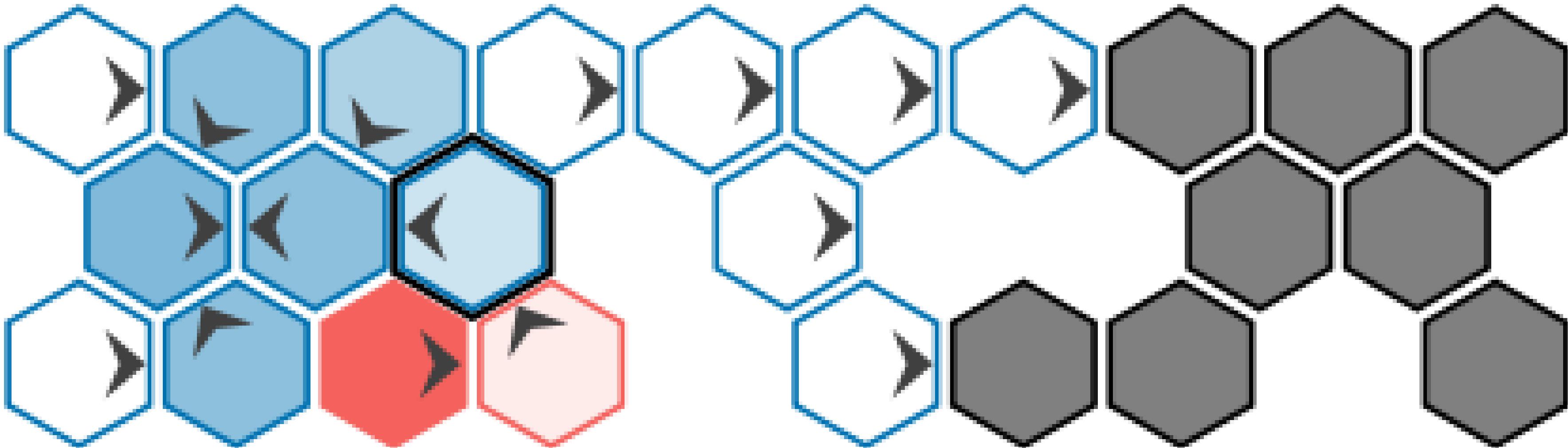


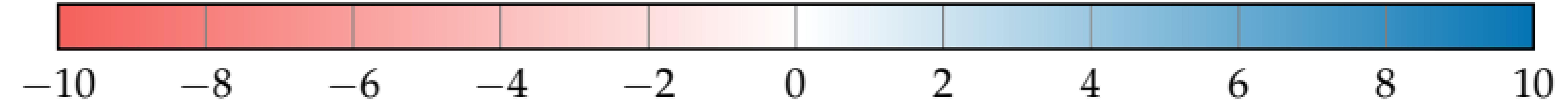


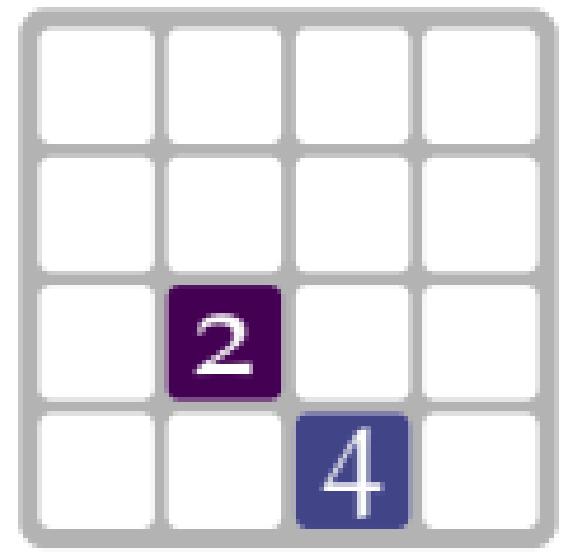












left

$$N = 0 \quad Q = 0$$

down

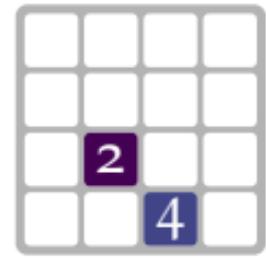
$$N = 0 \quad Q = 0$$

right

$$N = 0 \quad Q = 0$$

up

$$N = 0 \quad Q = 0$$

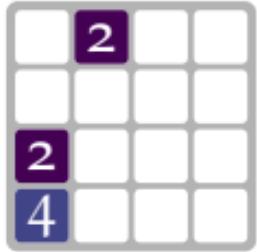


left
 $N = 1$
 $Q = 72$

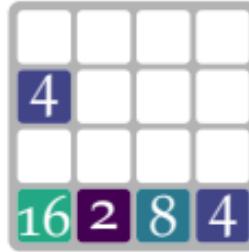
down
 $N = 0$
 $Q = 0$

right
 $N = 0$
 $Q = 0$

up
 $N = 0$
 $Q = 0$



$$U(s) = 72$$

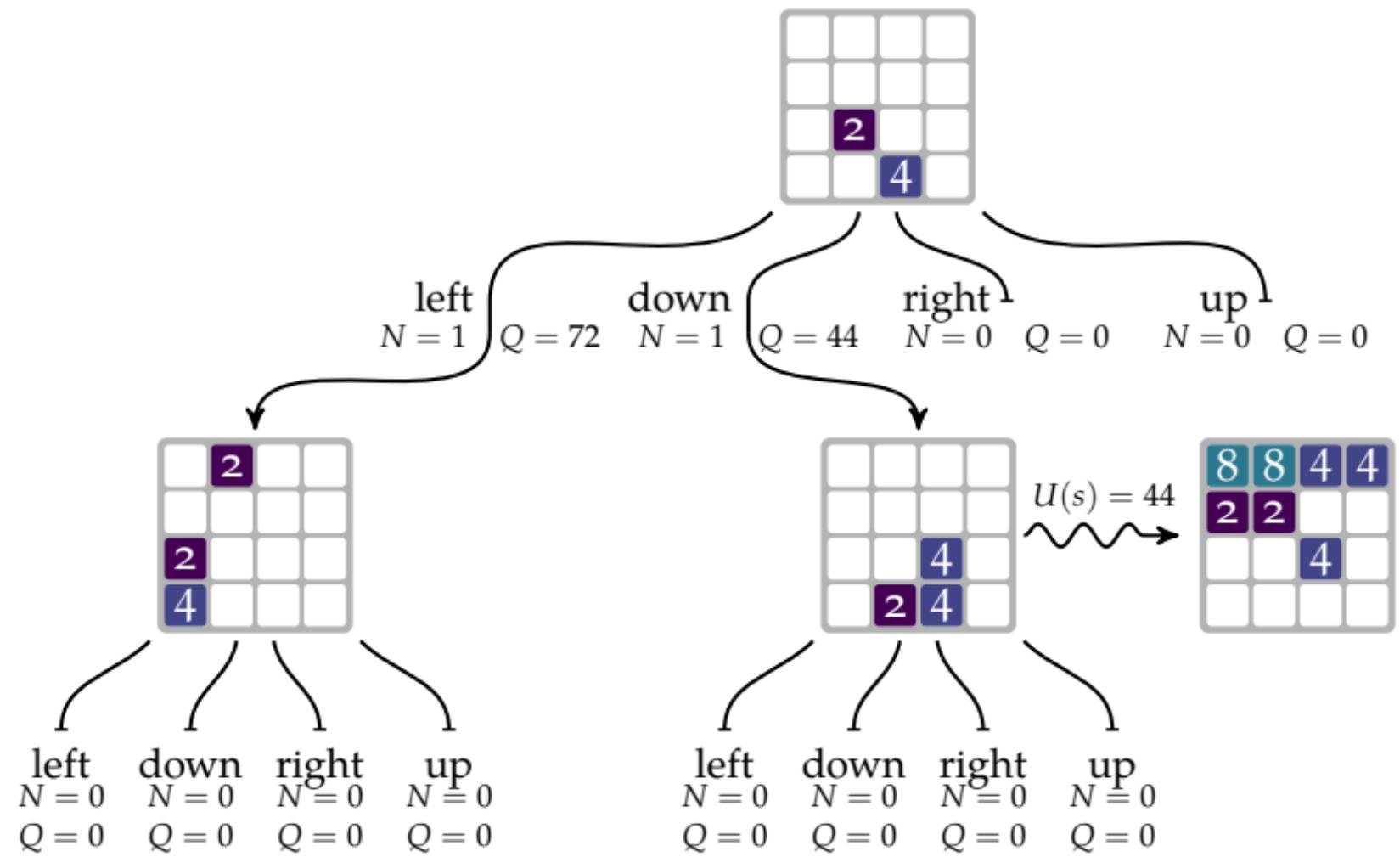


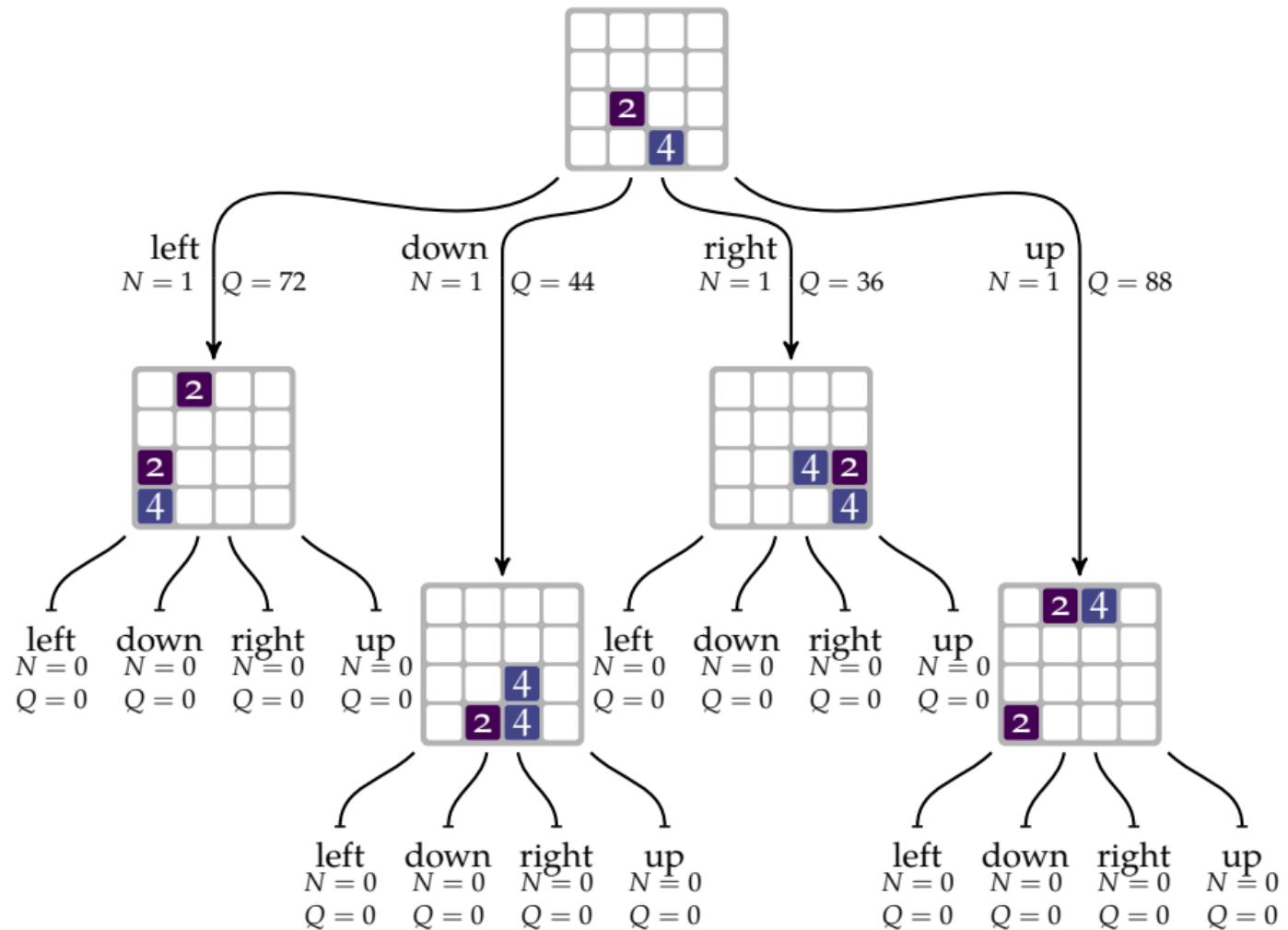
left
 $N = 0$
 $Q = 0$

down
 $N = 0$
 $Q = 0$

right
 $N = 0$
 $Q = 0$

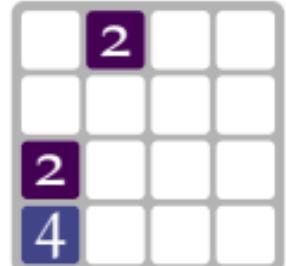
up
 $N = 0$
 $Q = 0$







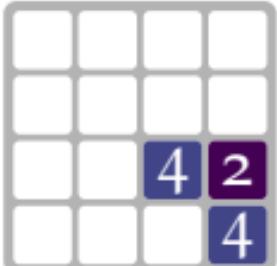
left
 $N = 1$ $Q = 72$



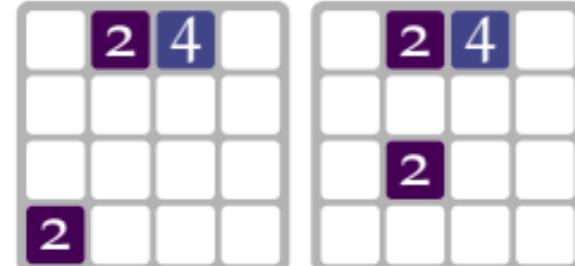
down
 $N = 1$ $Q = 44$

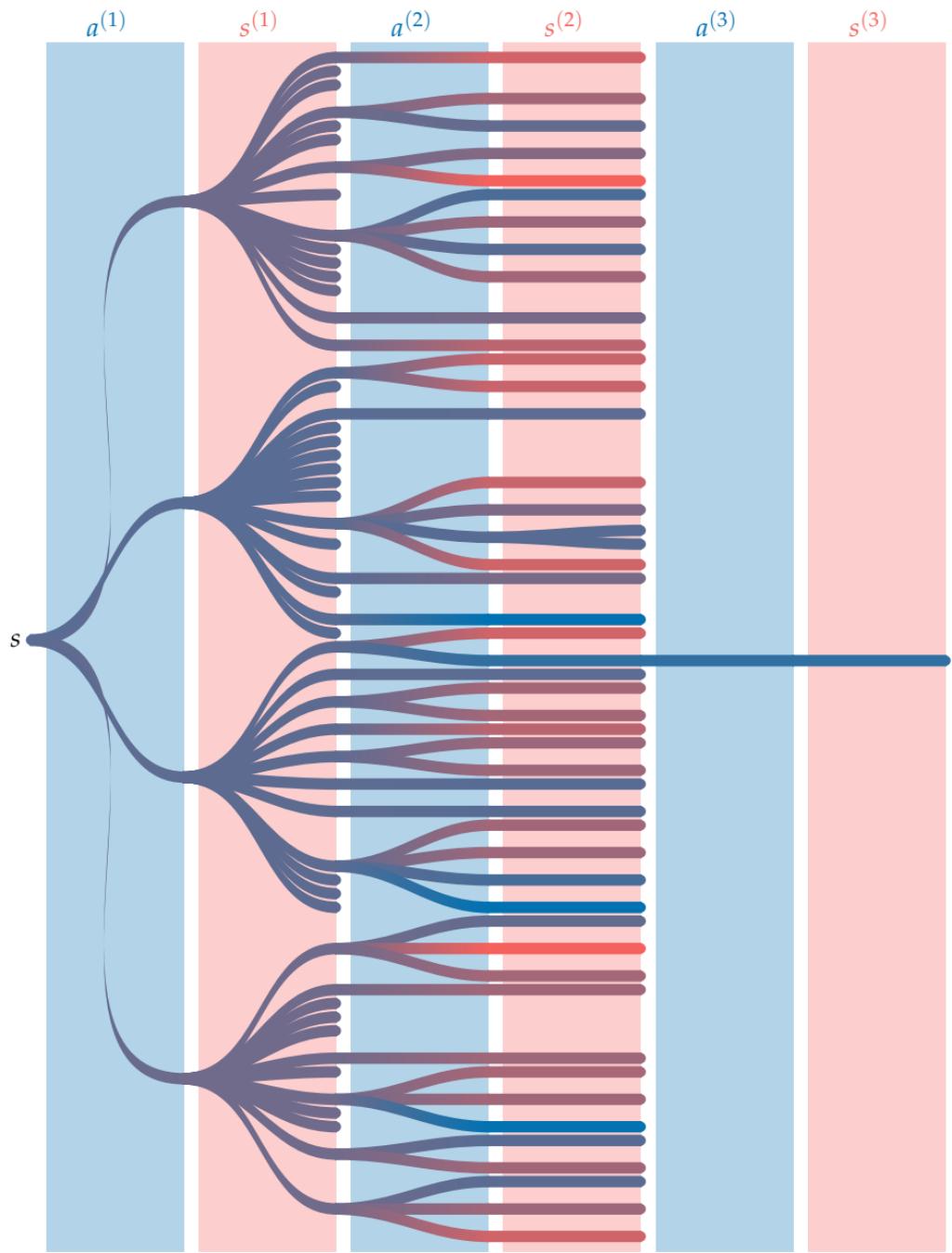


right
 $N = 1$ $Q = 36$



up
 $N = 2$ $Q = 66$



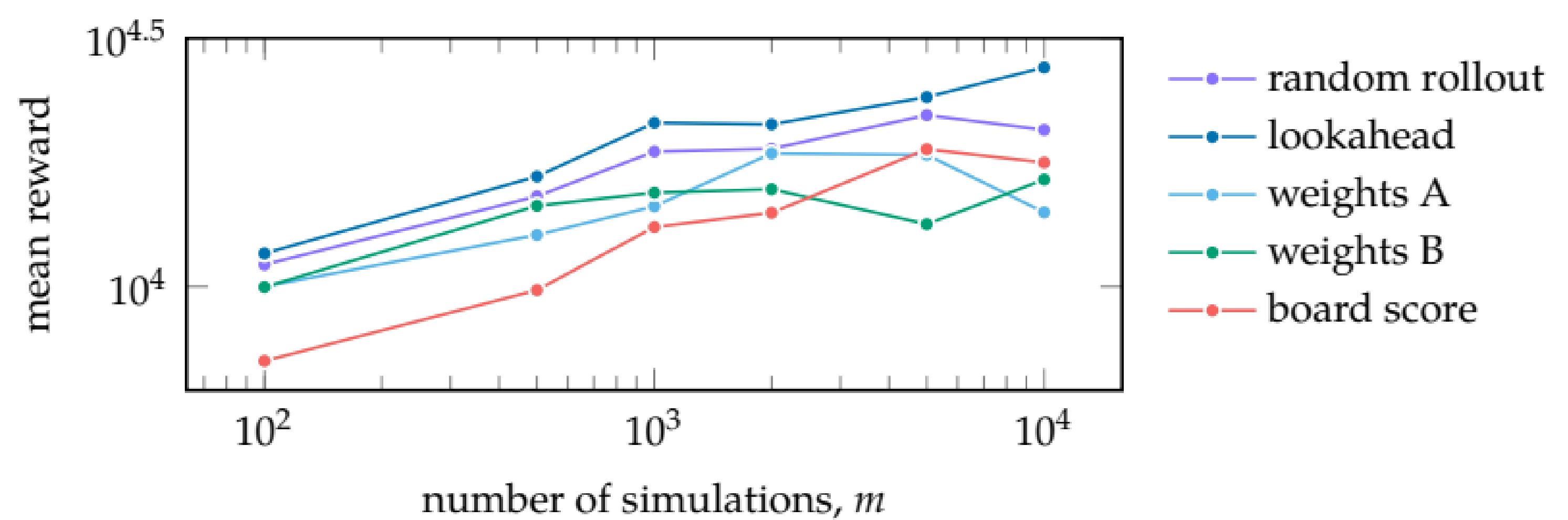


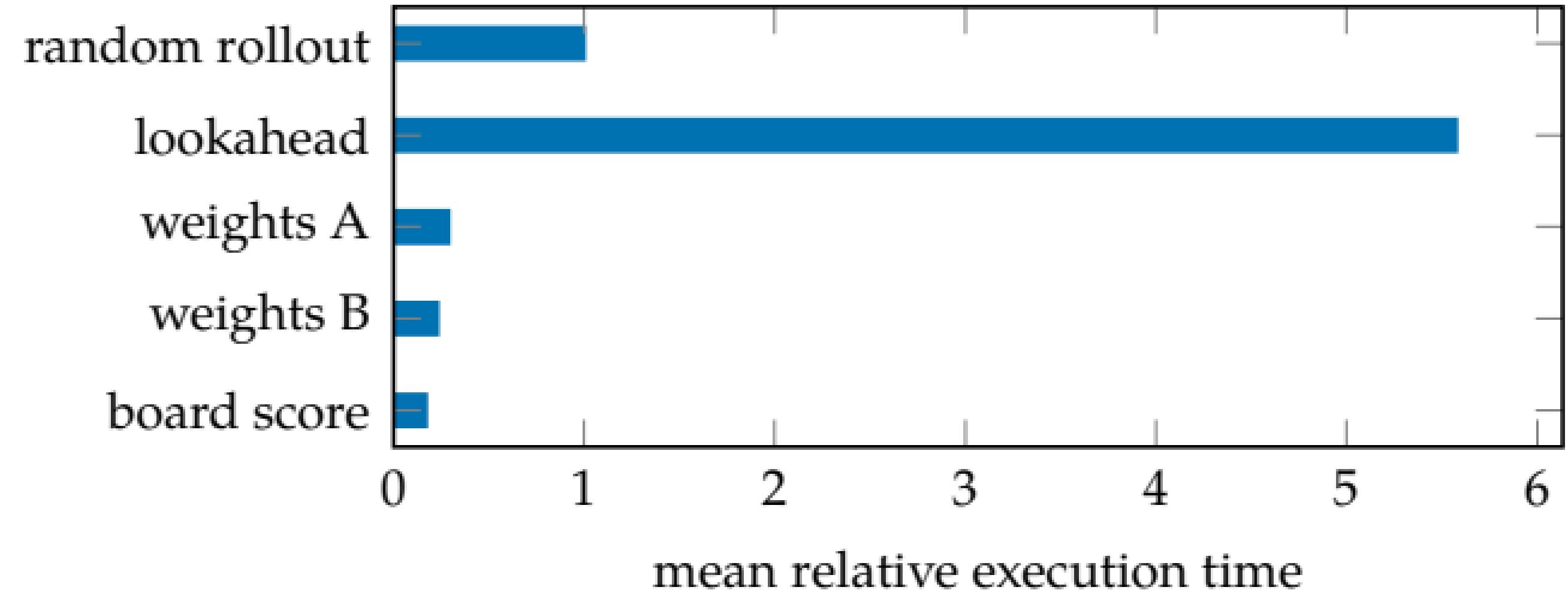
1	2	3	4
2	3	4	5
3	4	5	6
4	5	6	7

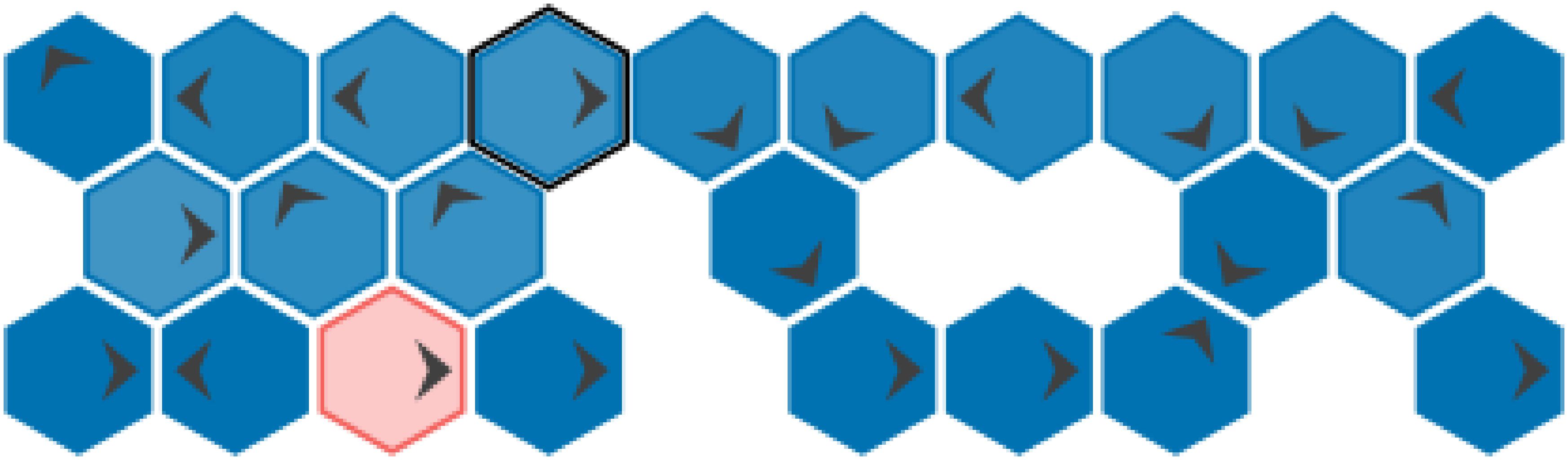
heuristic A weights

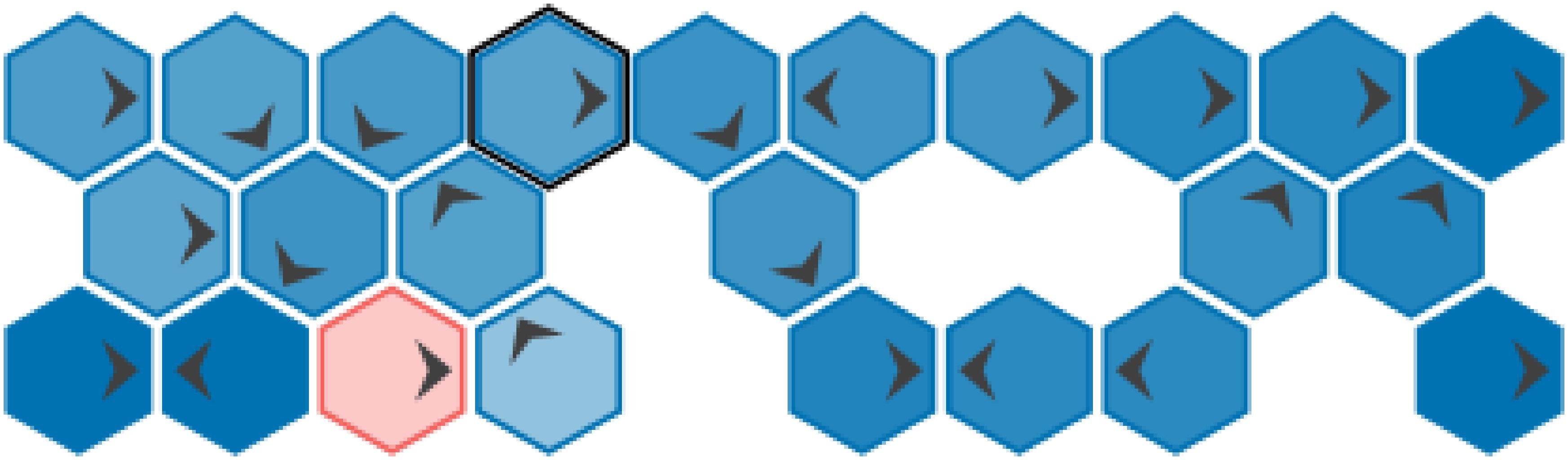
0	1	2	3
7	6	5	4
8	9	10	11
15	14	13	12

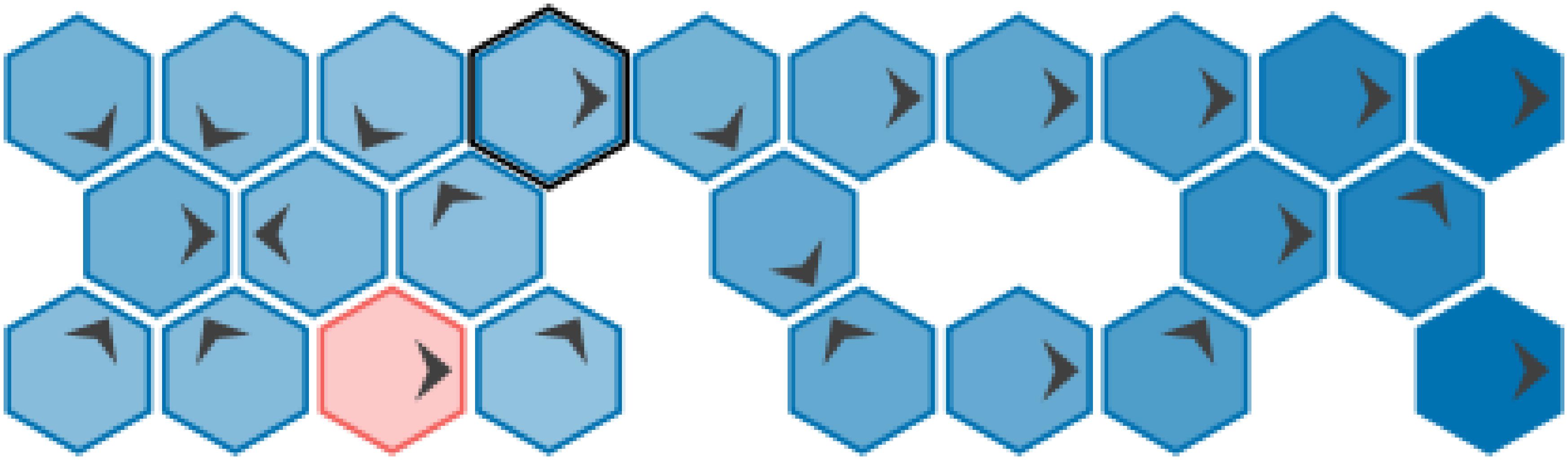
heuristic B weights

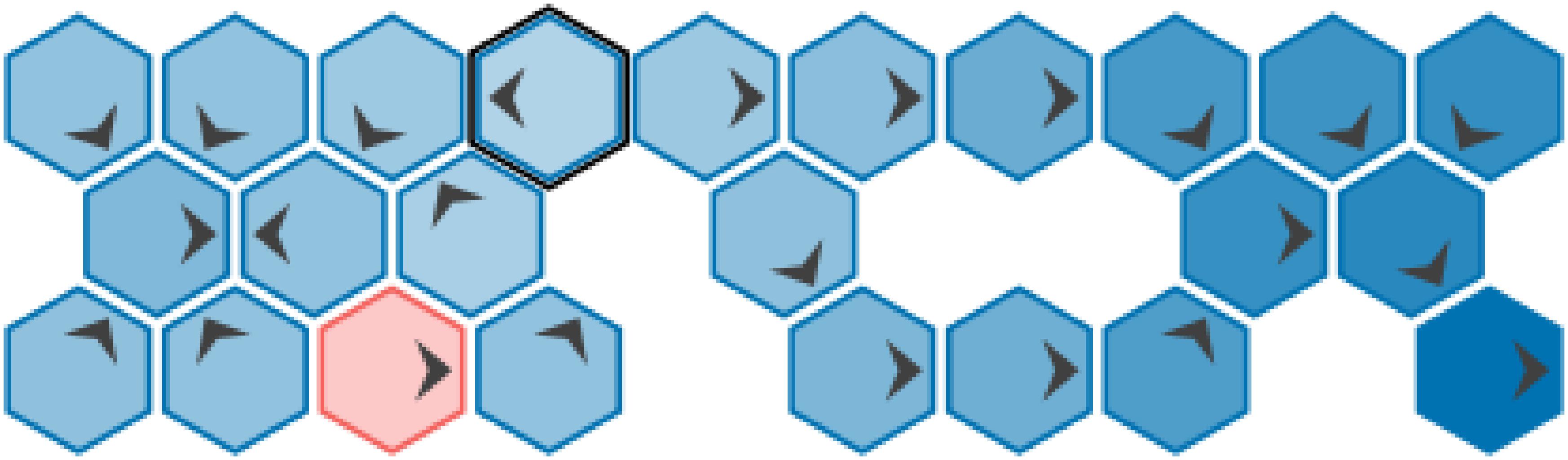


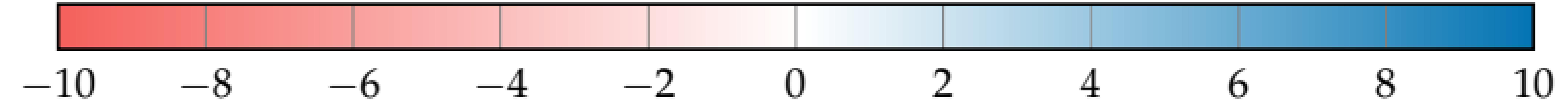


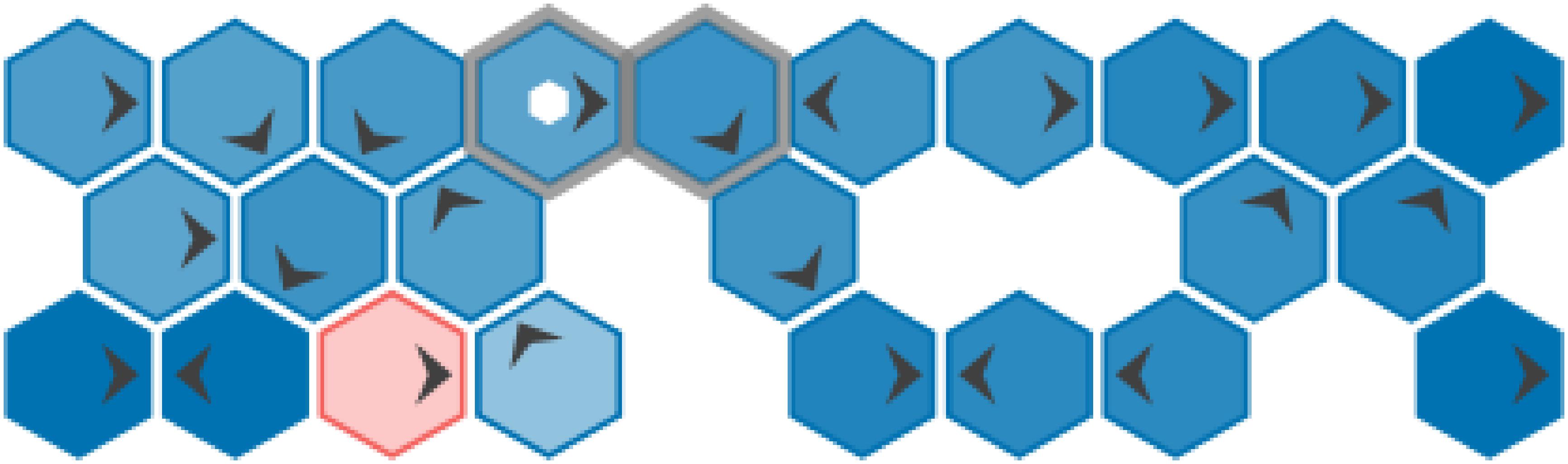


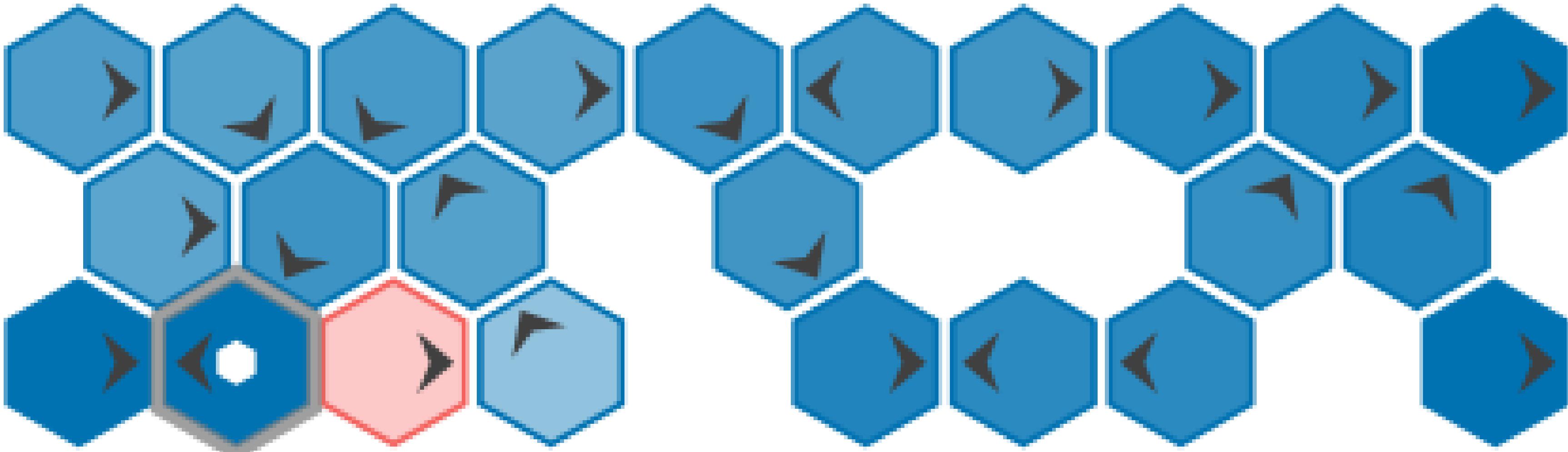


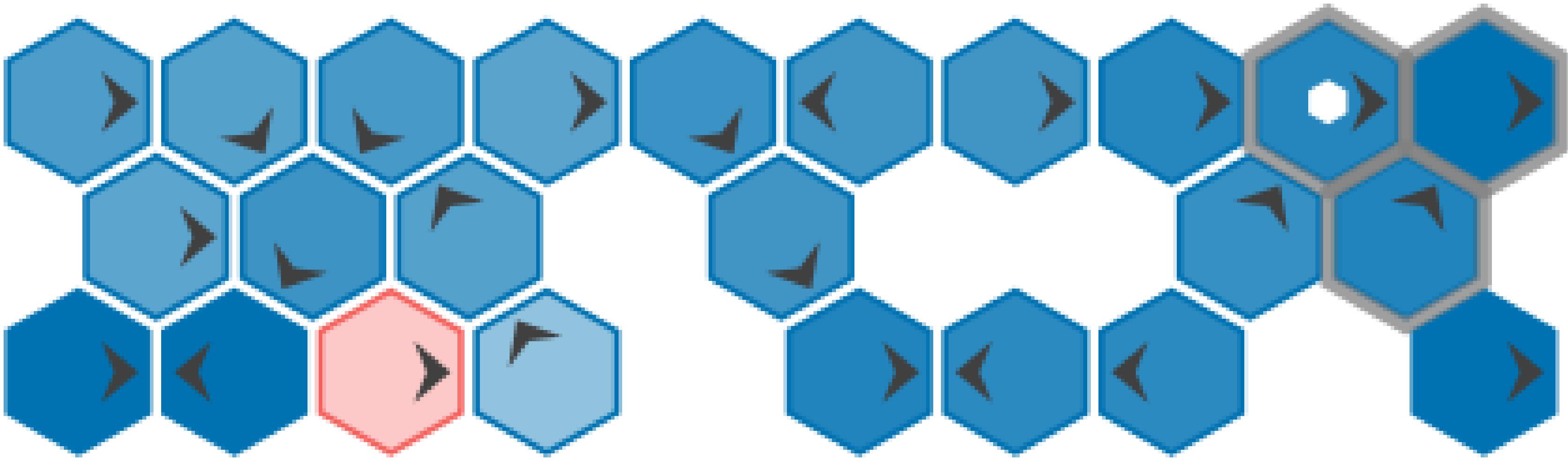


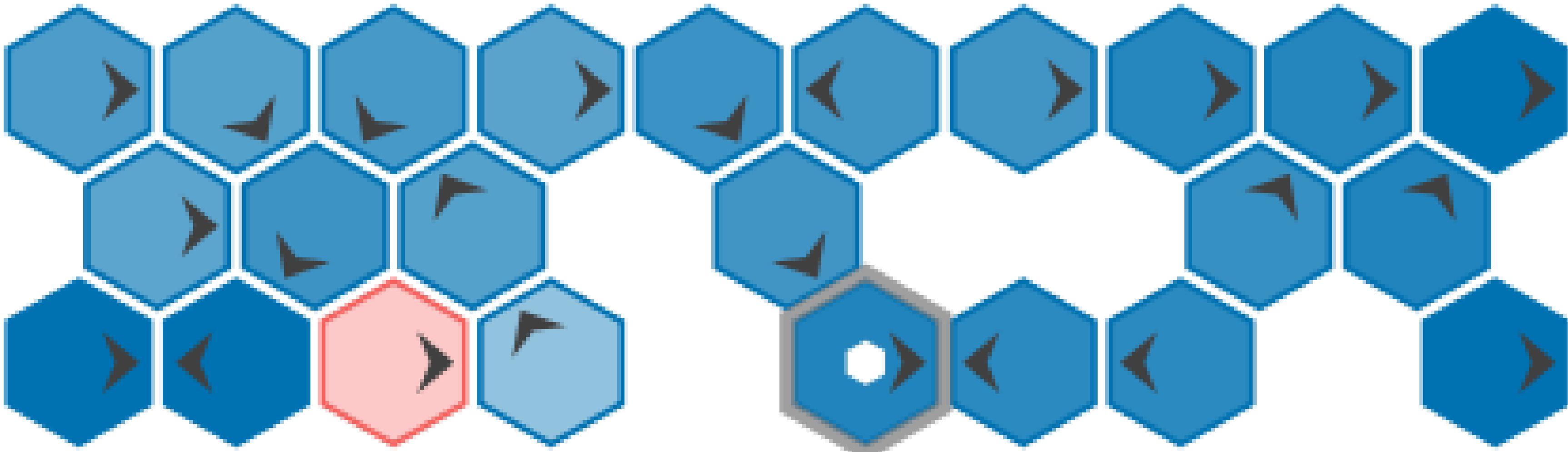


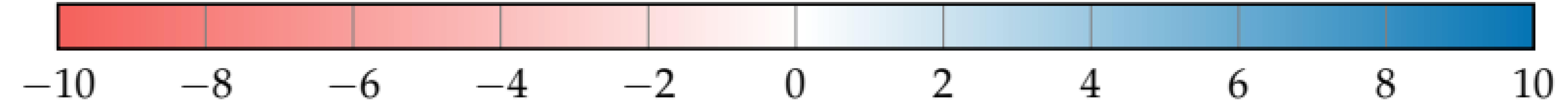




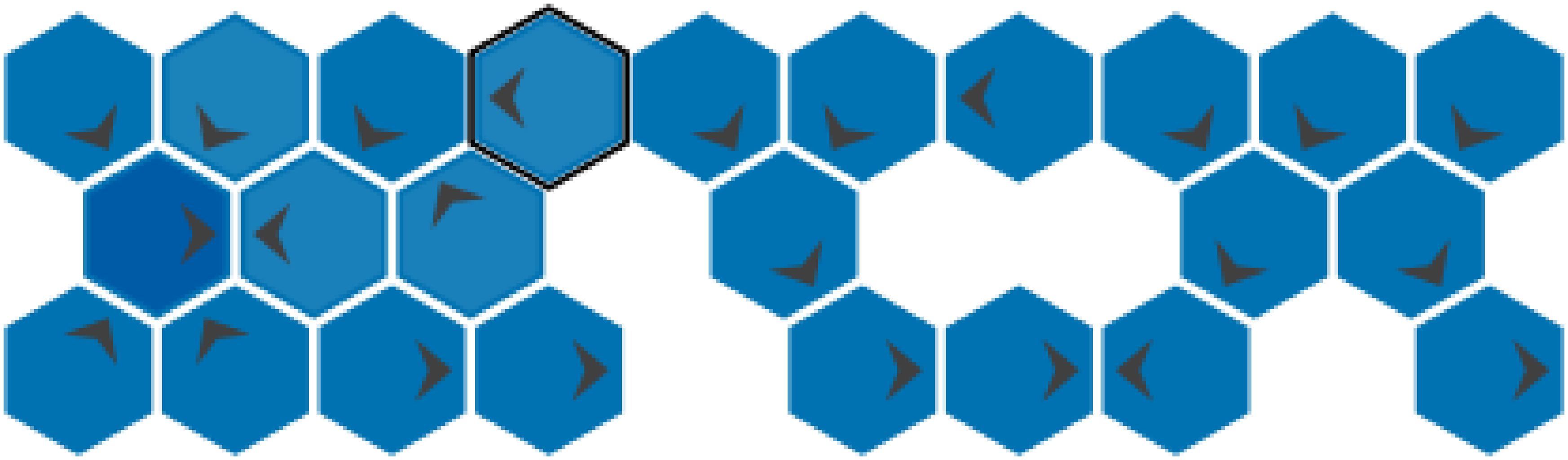


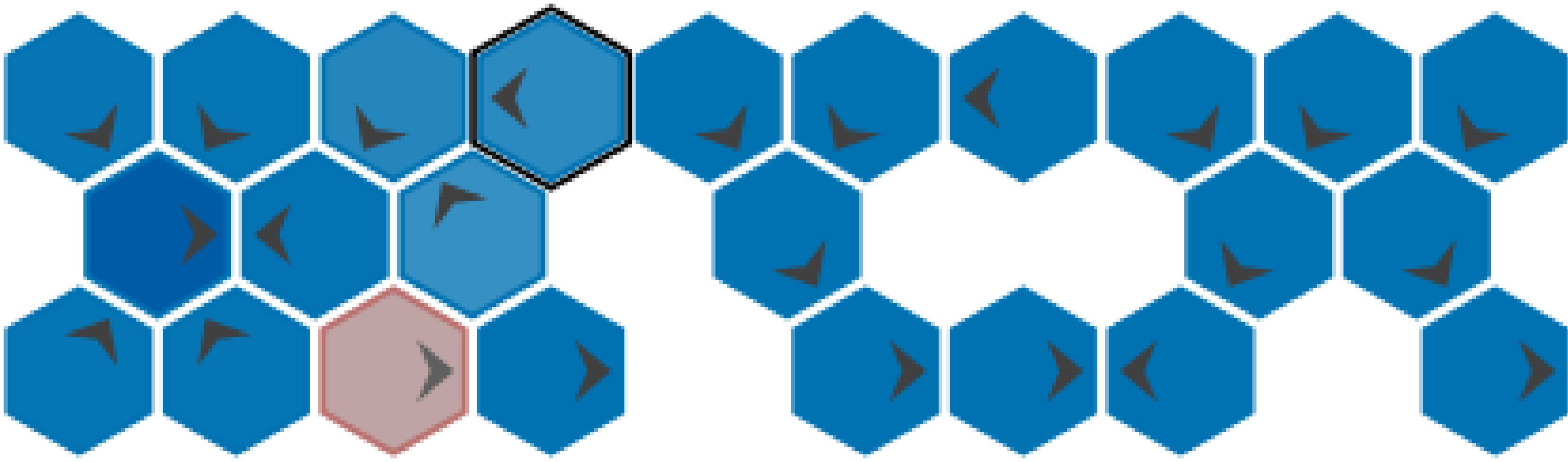


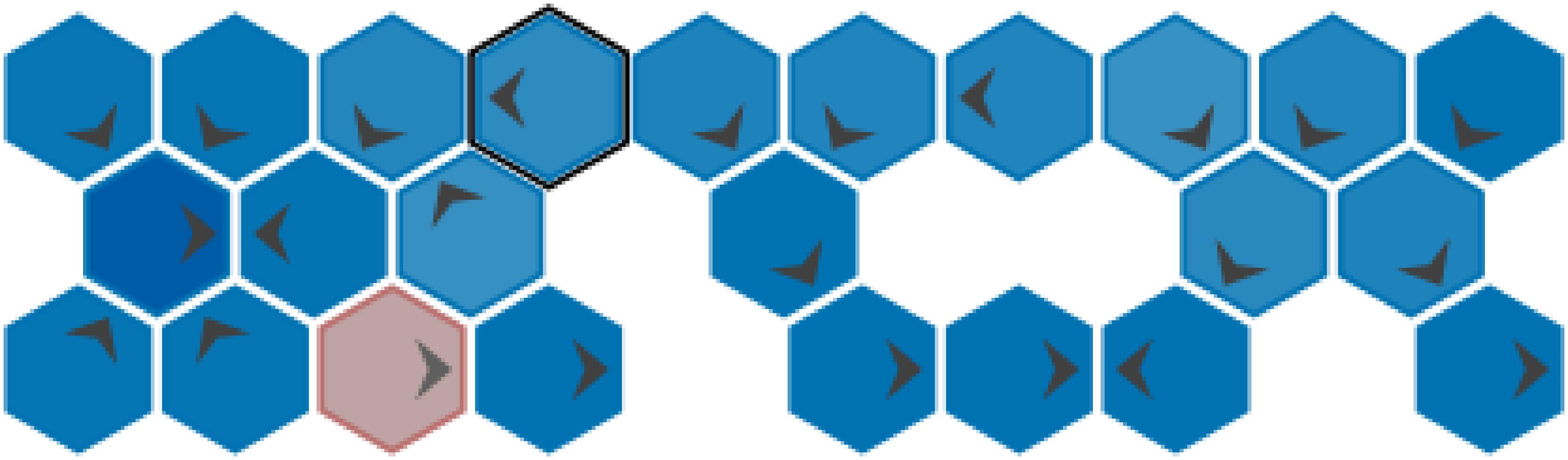


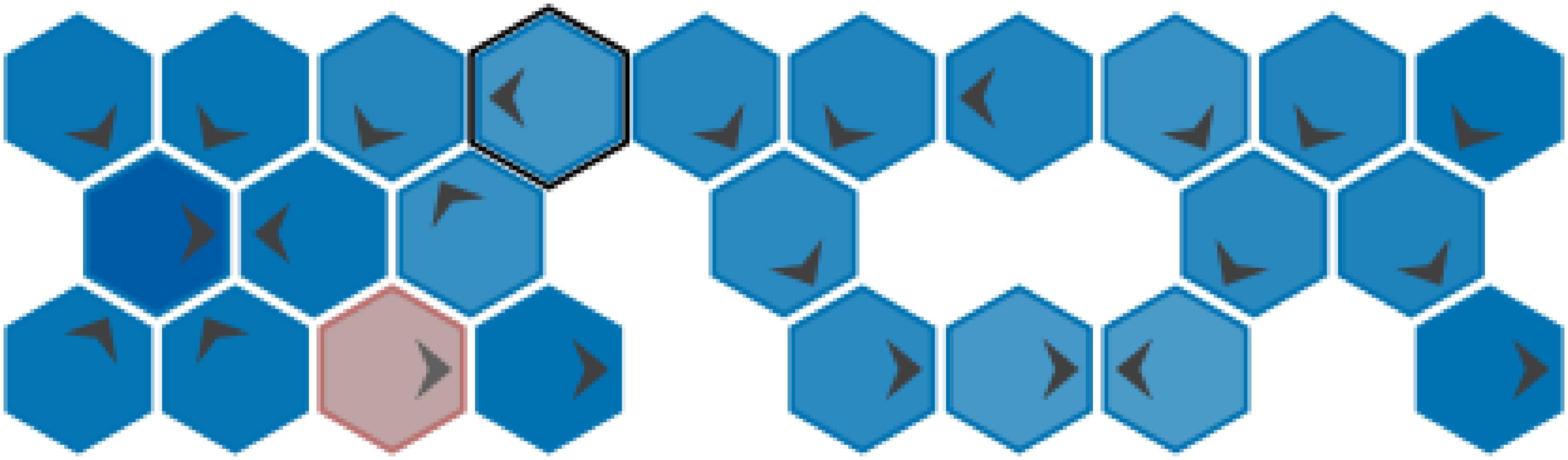


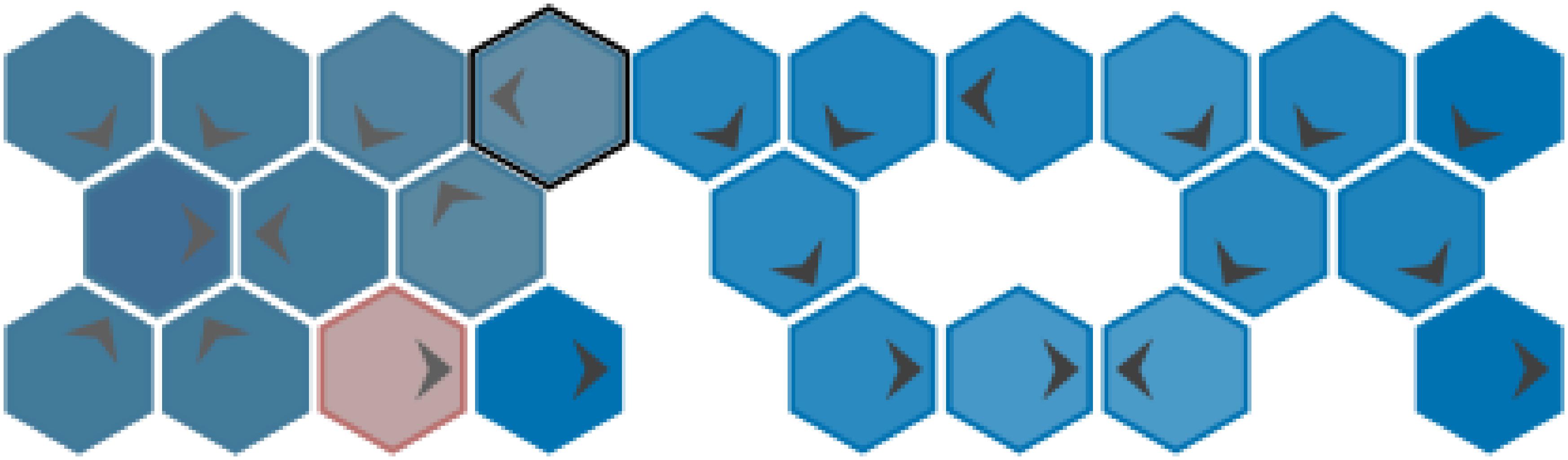


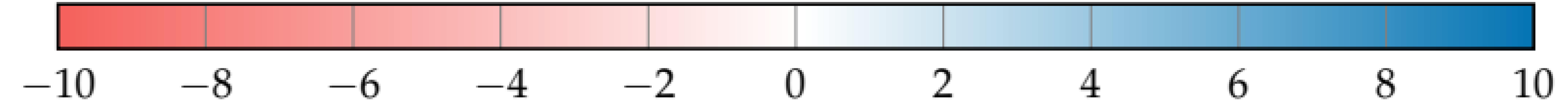


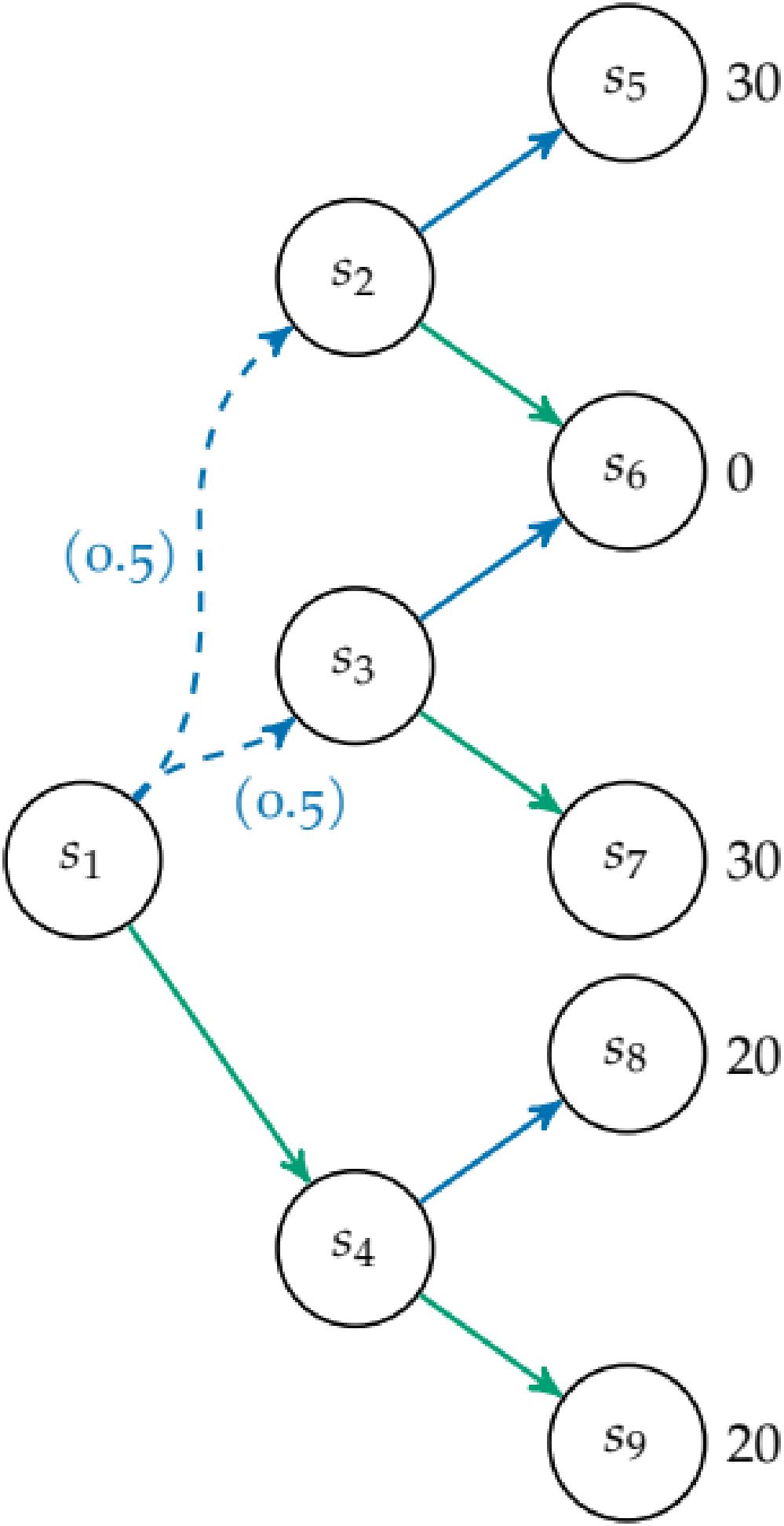


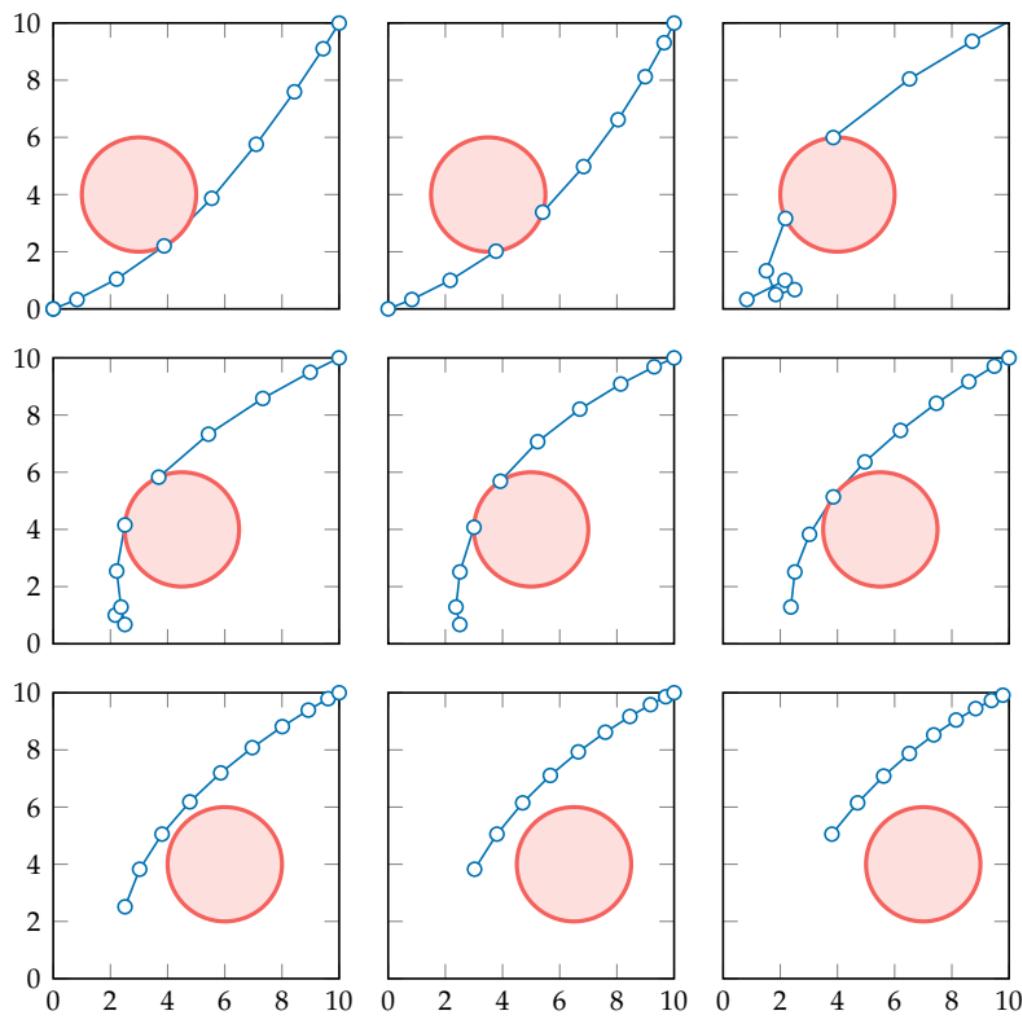


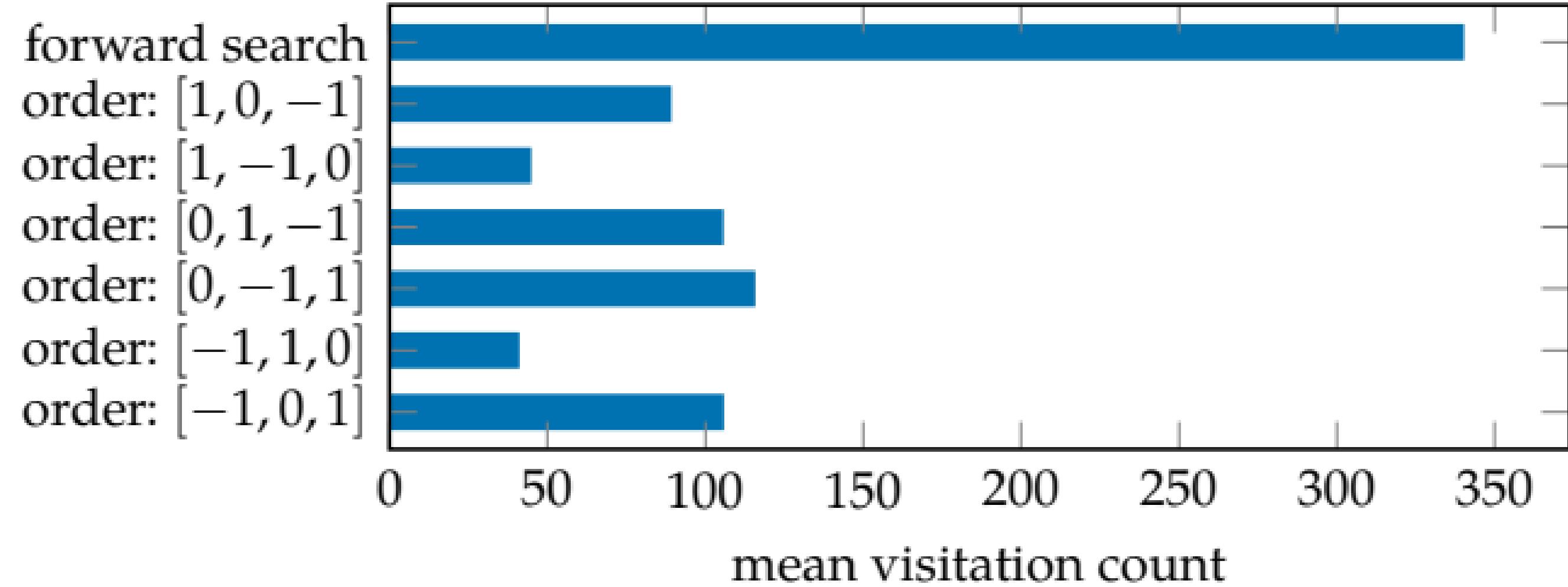


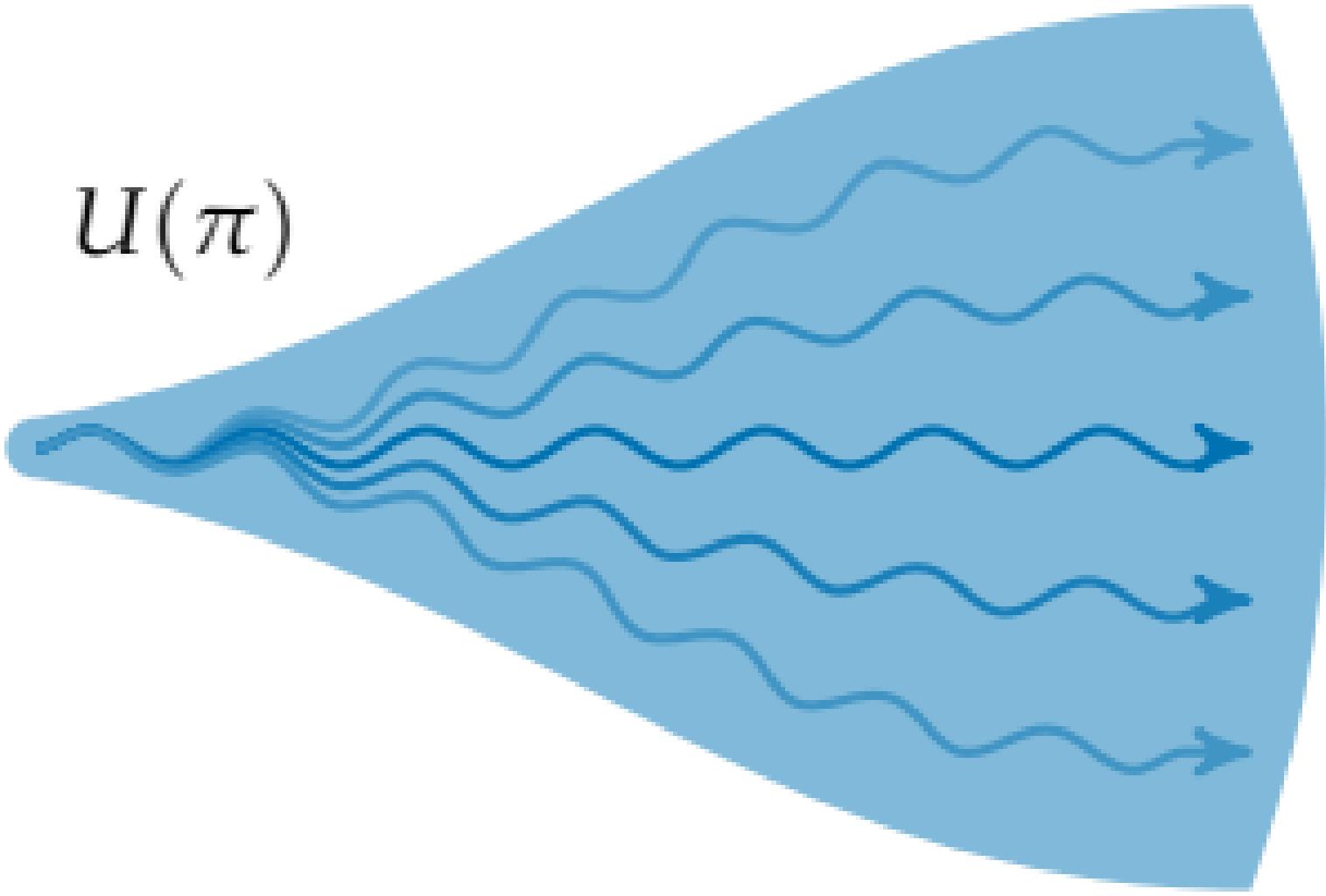


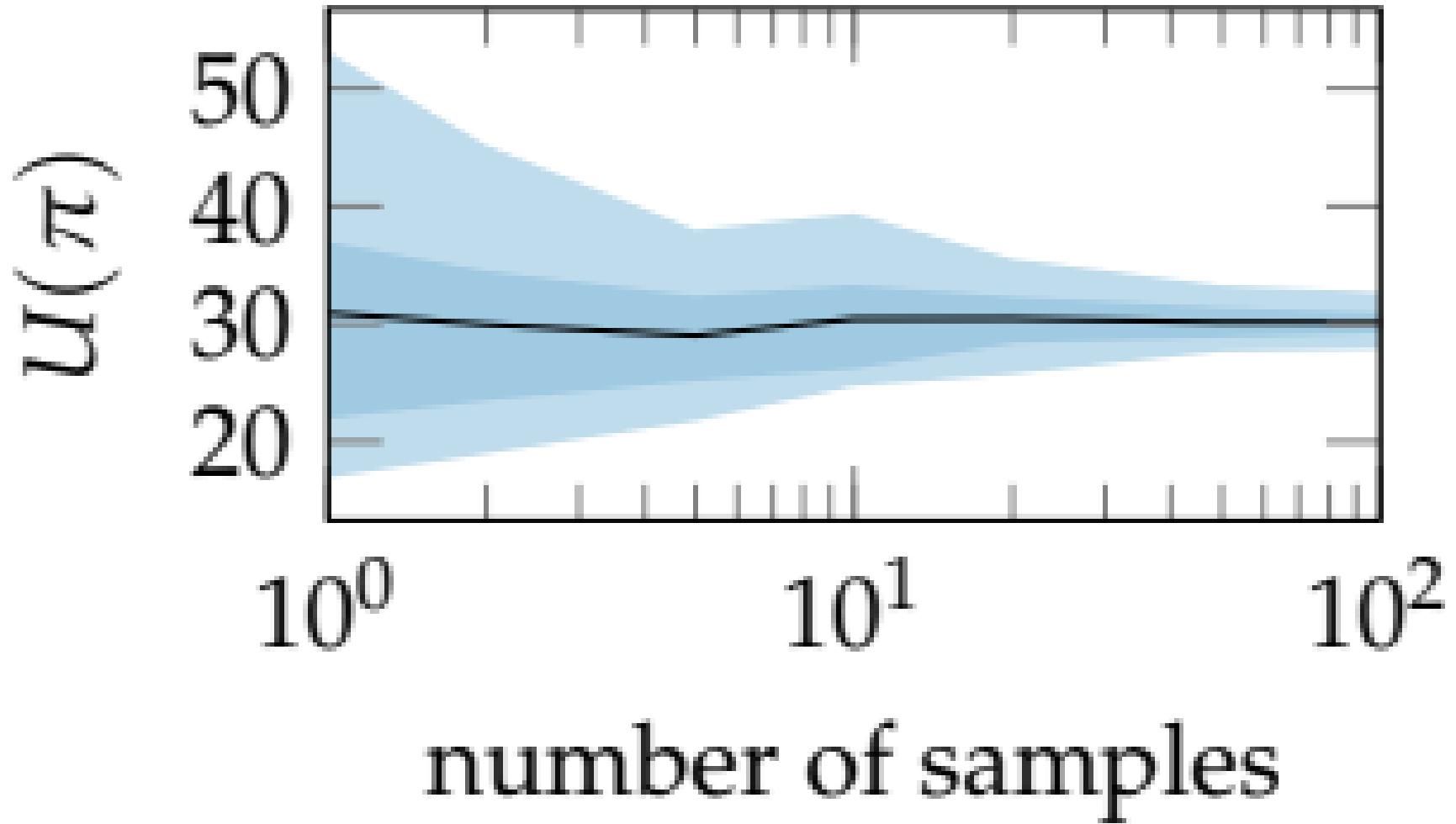


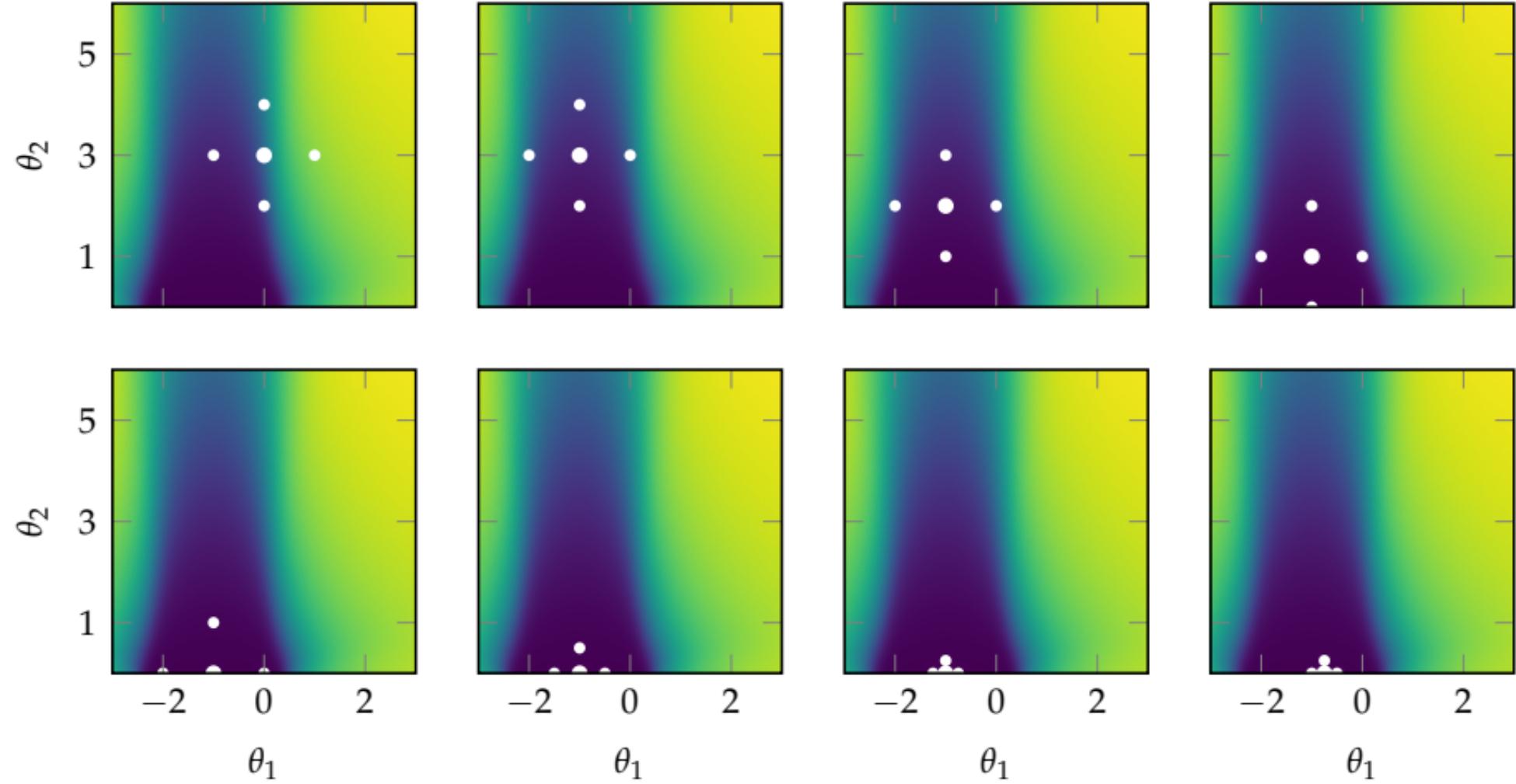


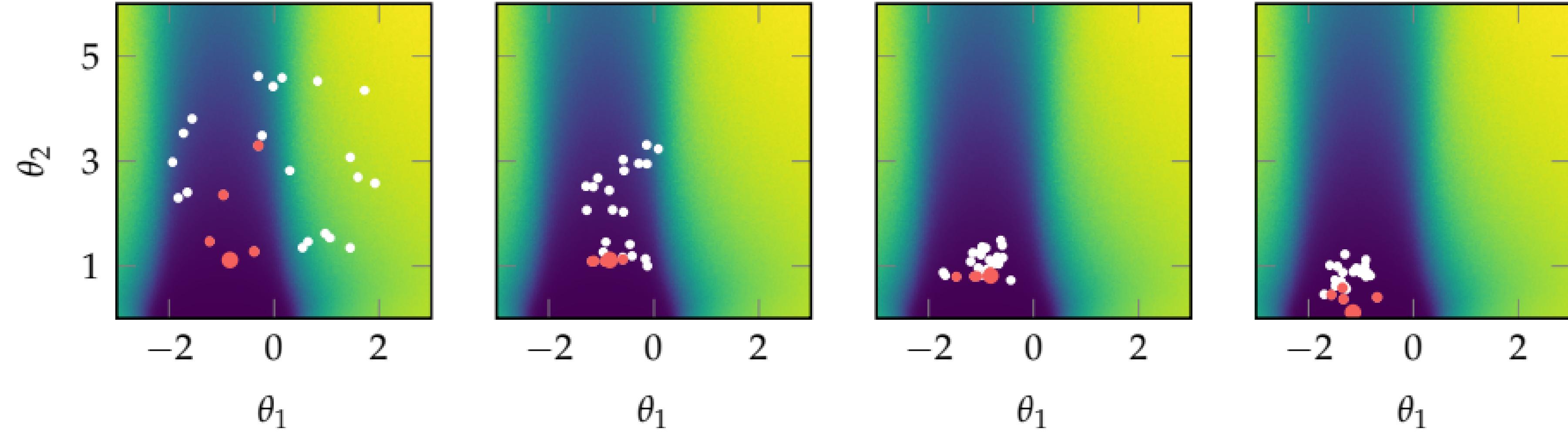


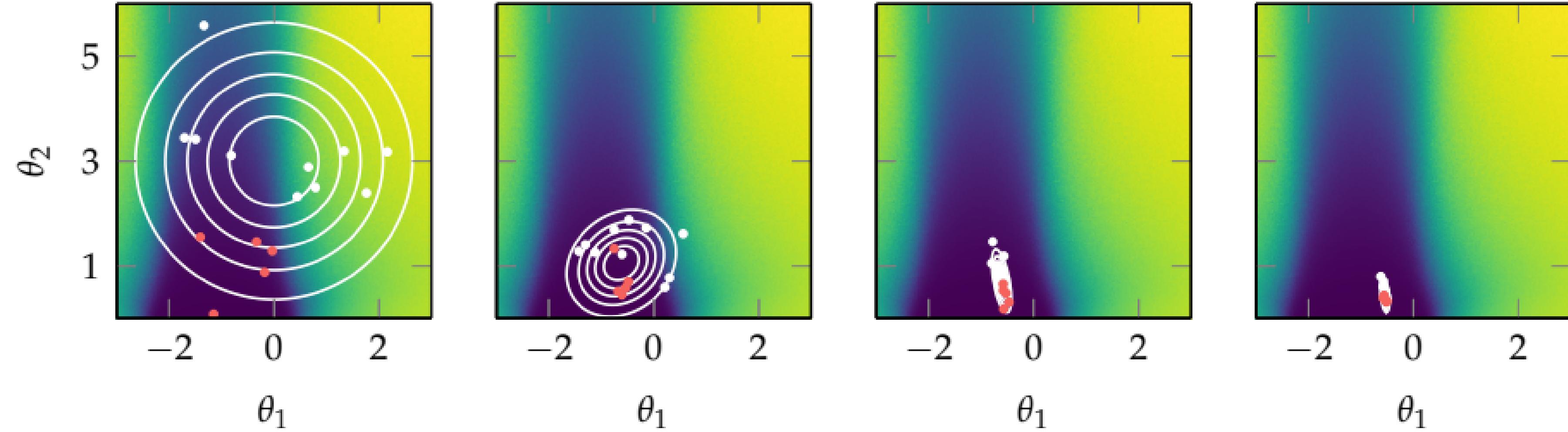


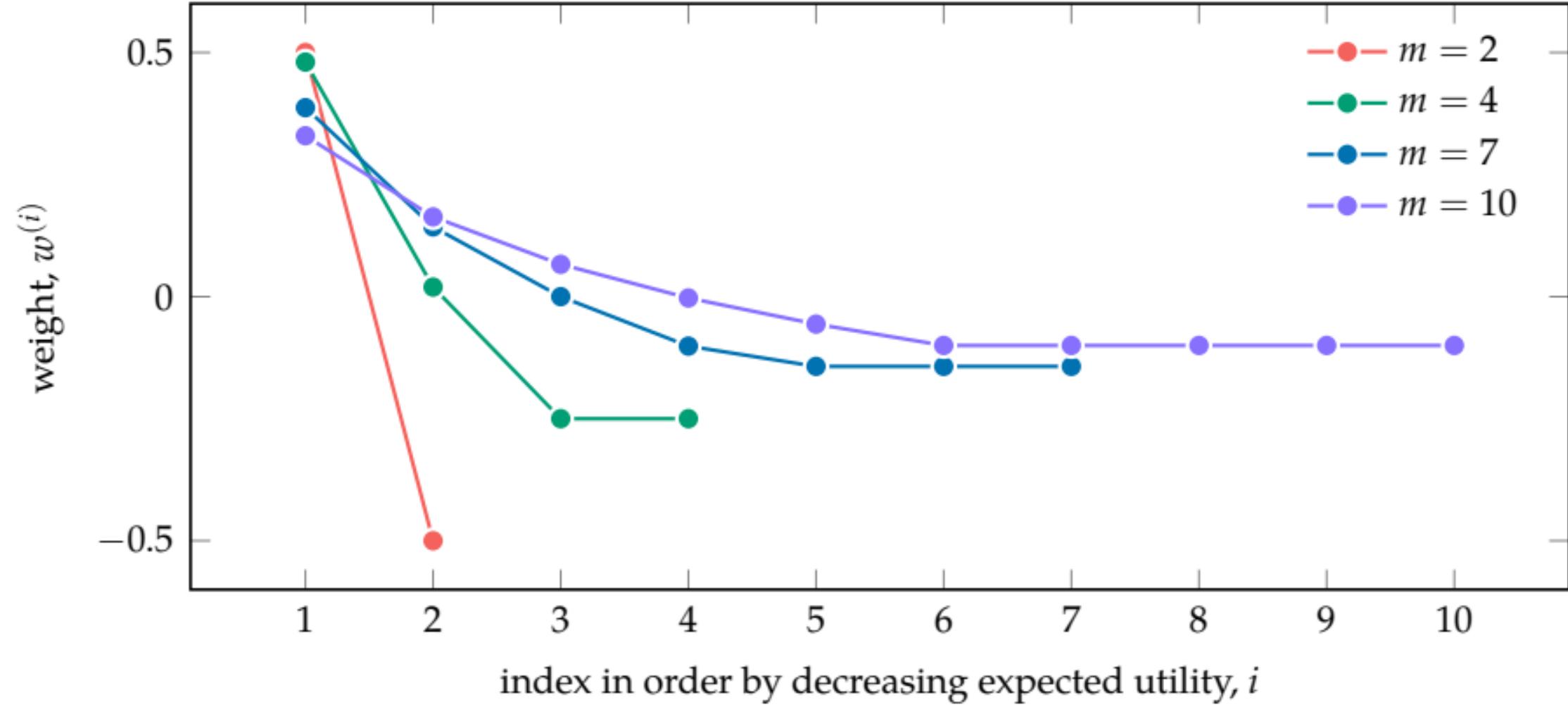
$U(\pi)$ 

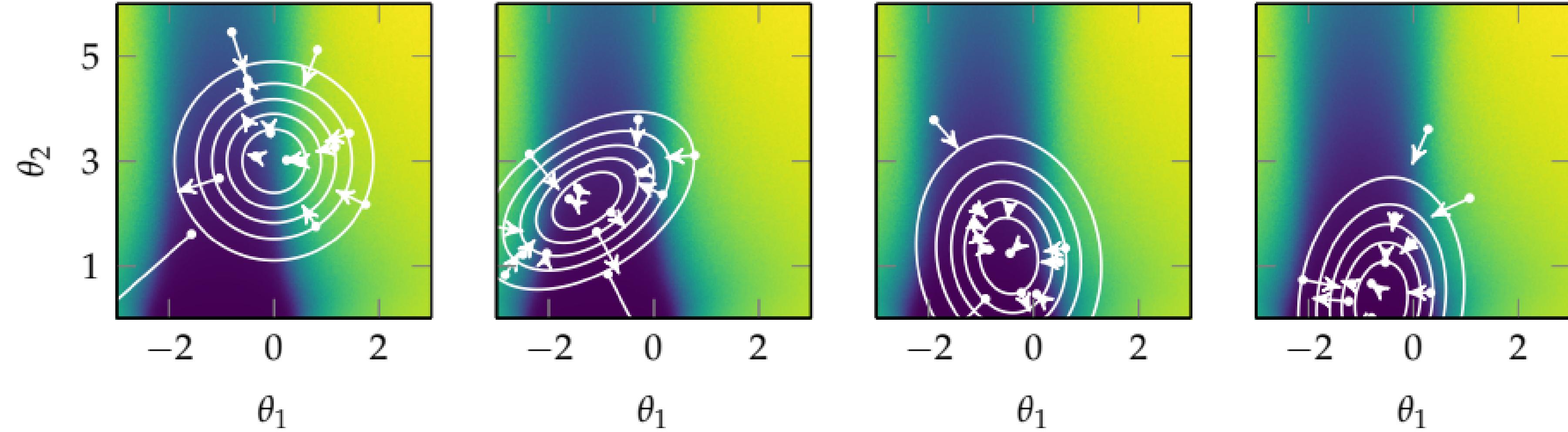


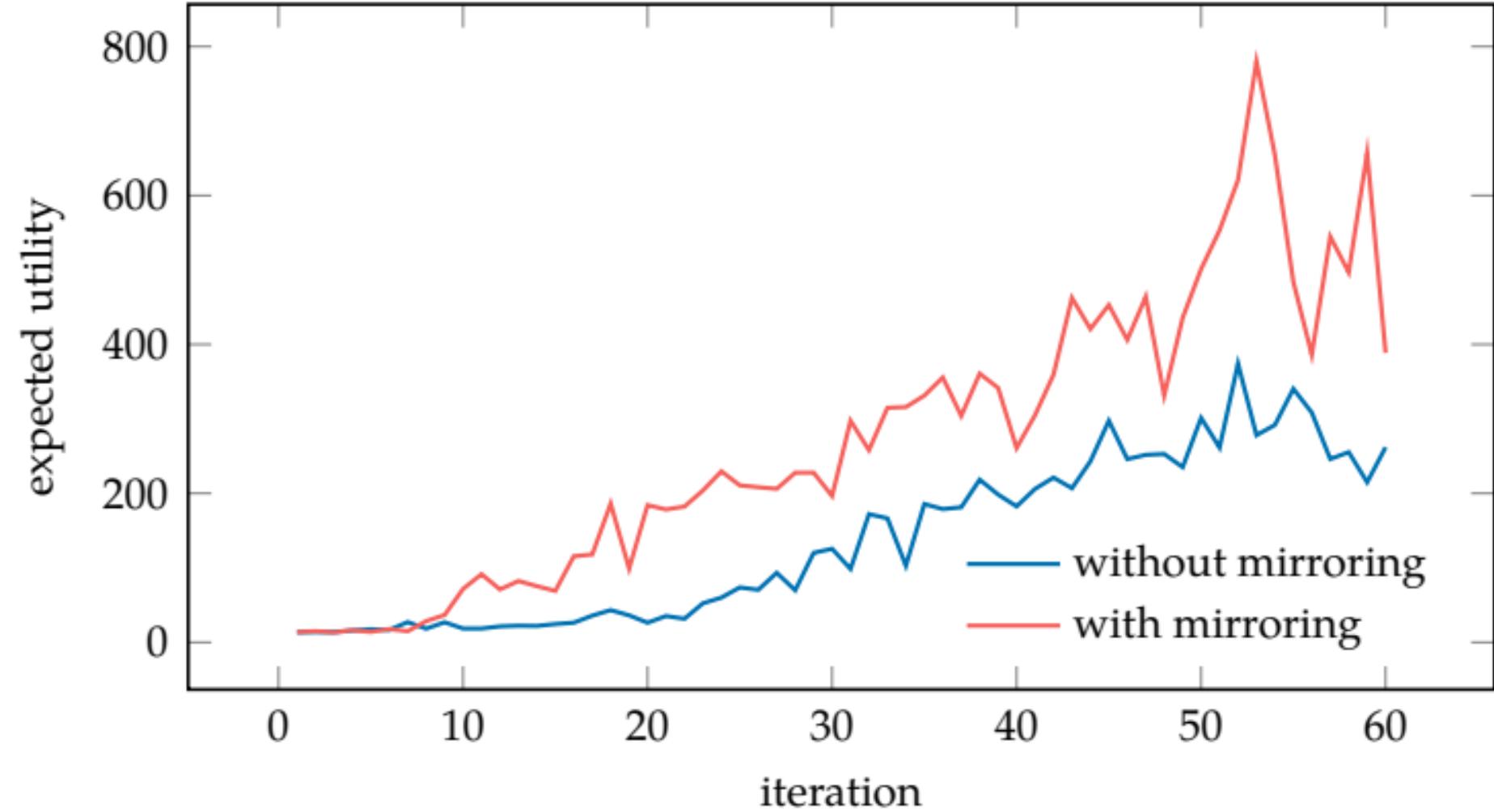


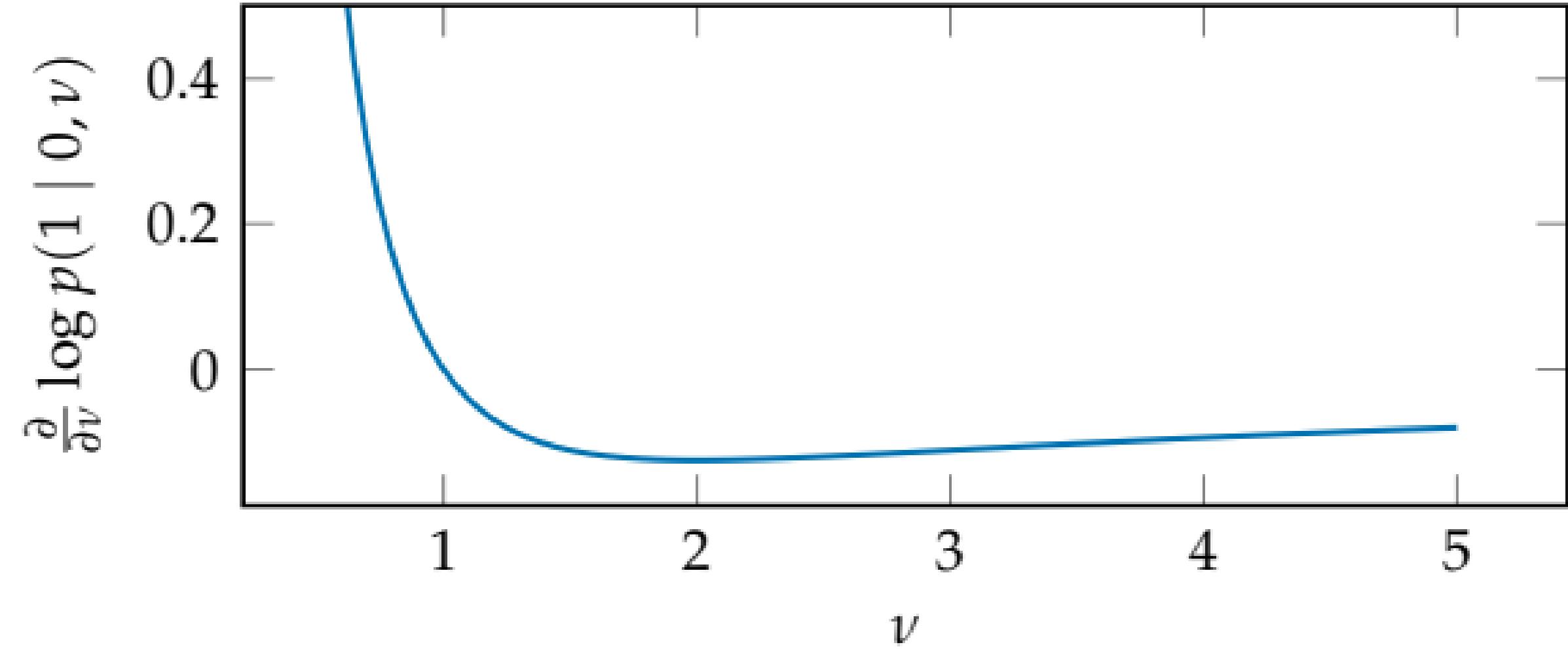


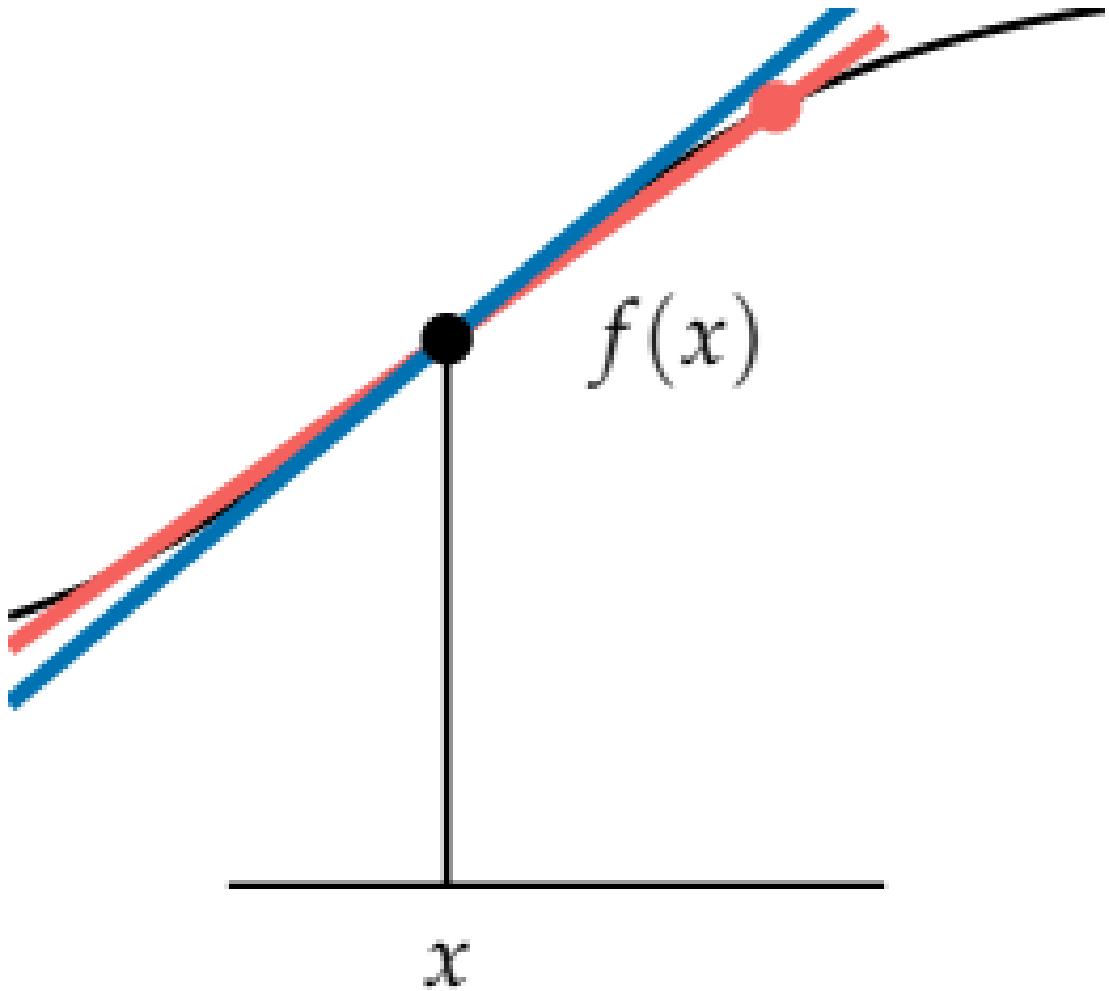










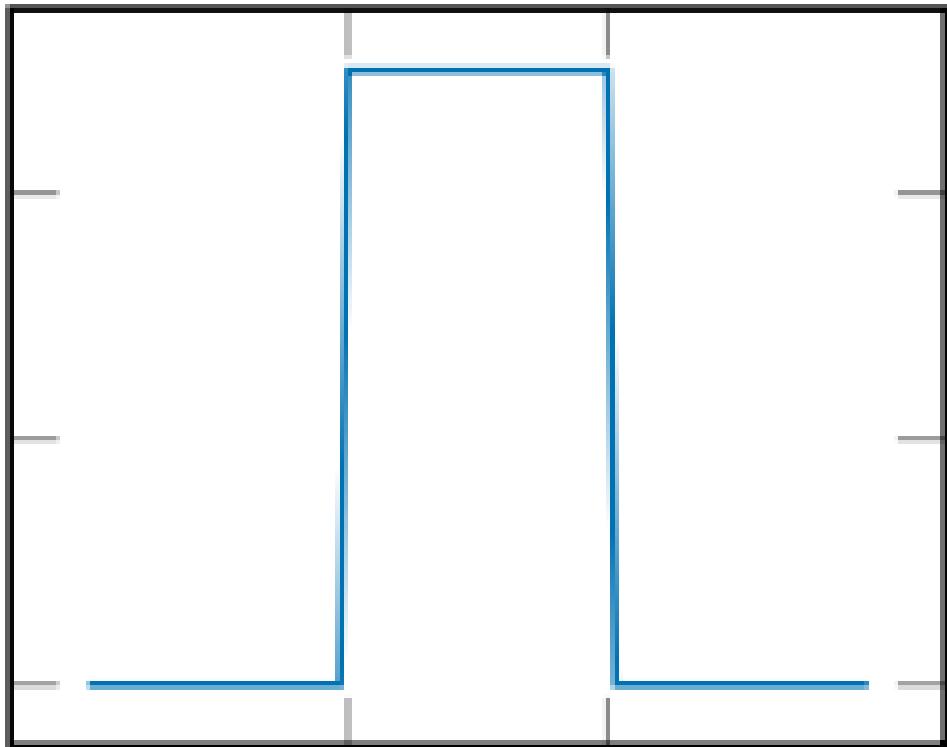


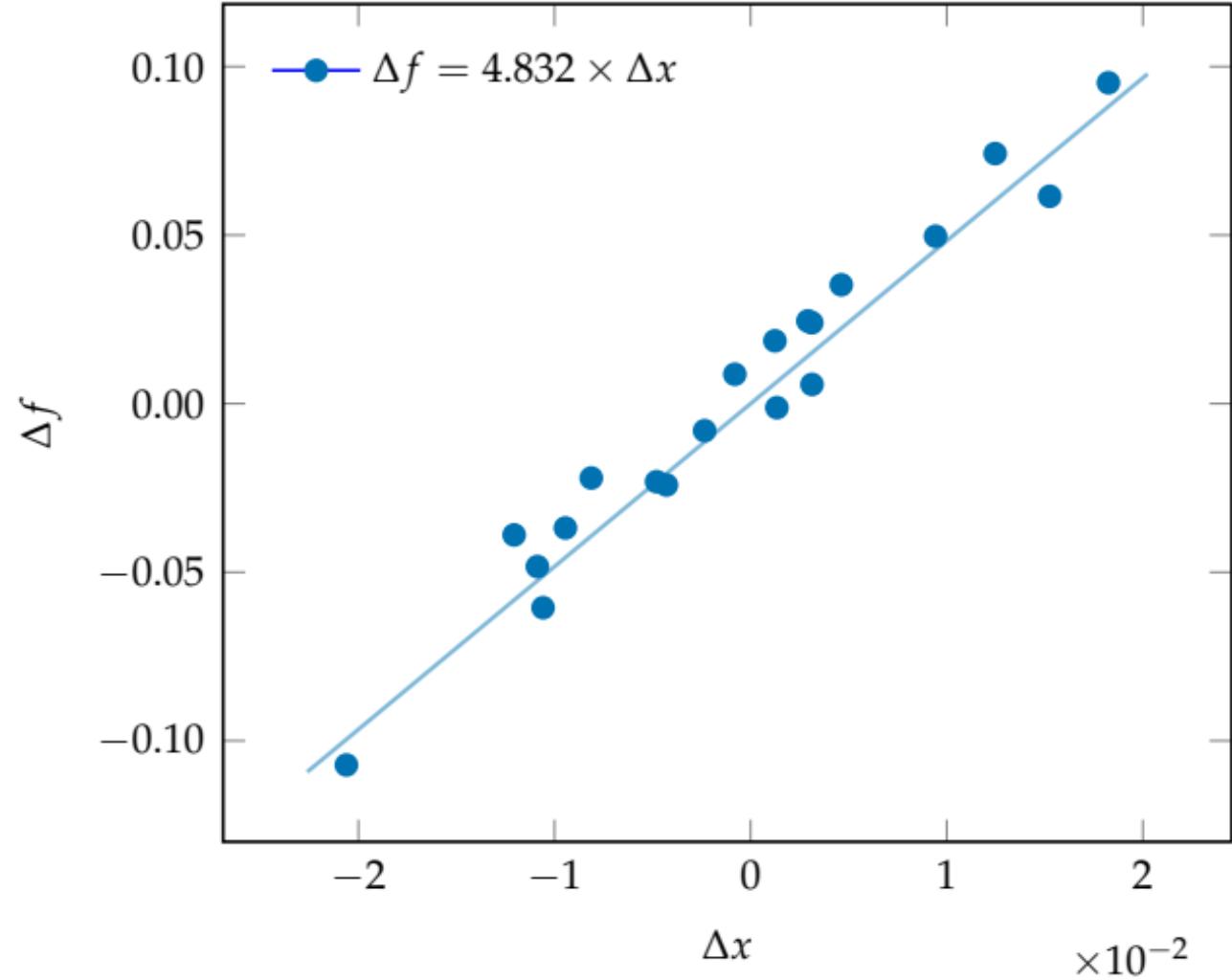
$\pi(a | s)$

0.4

0.2

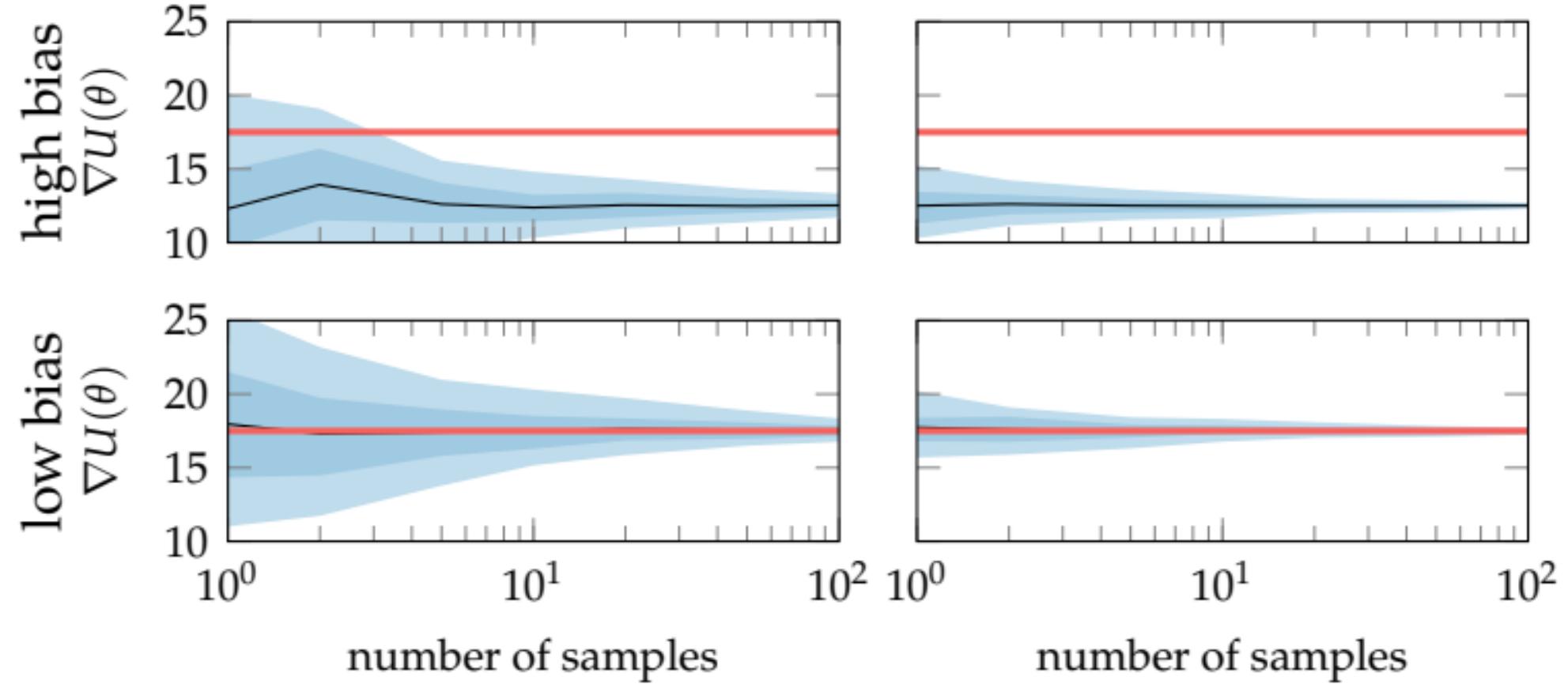
0

 $\theta_1 = \theta'_1 \quad \theta_2 = 100\theta'_2$ a 

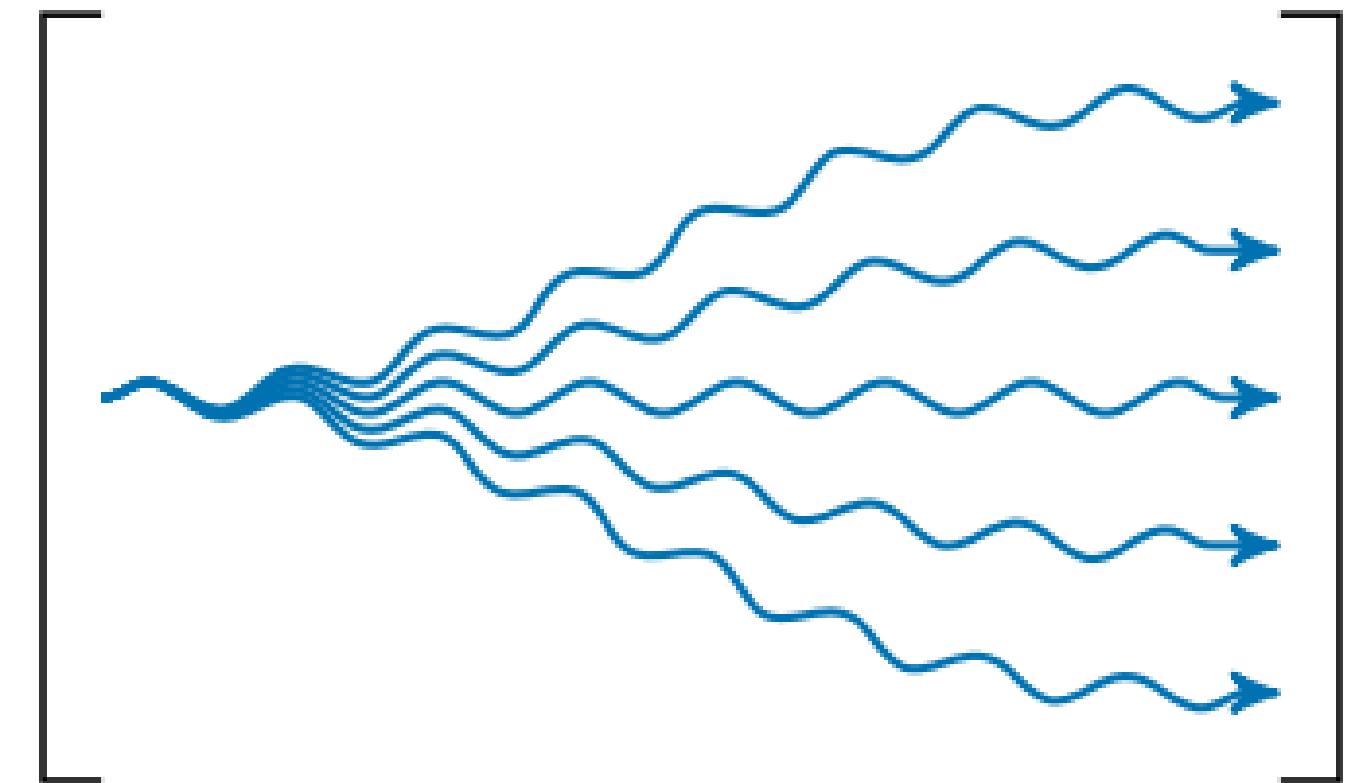


high variance

low variance

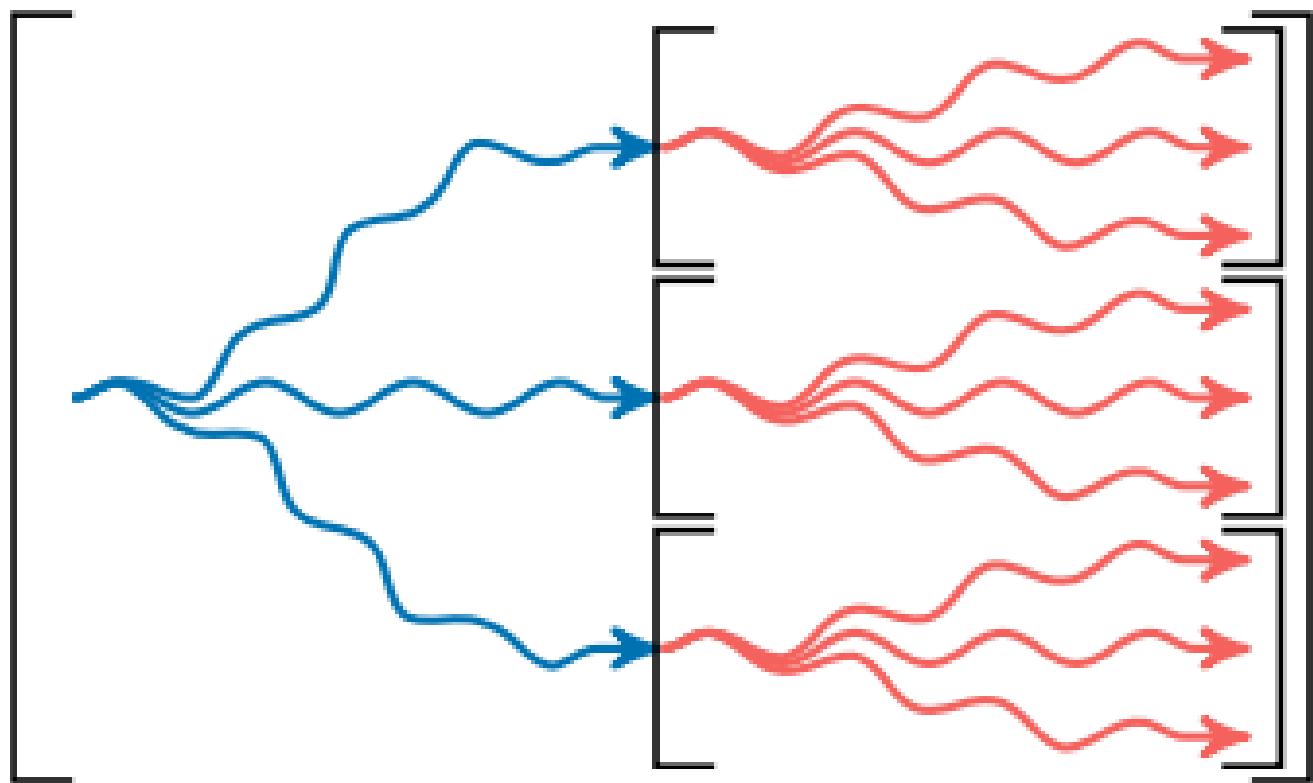


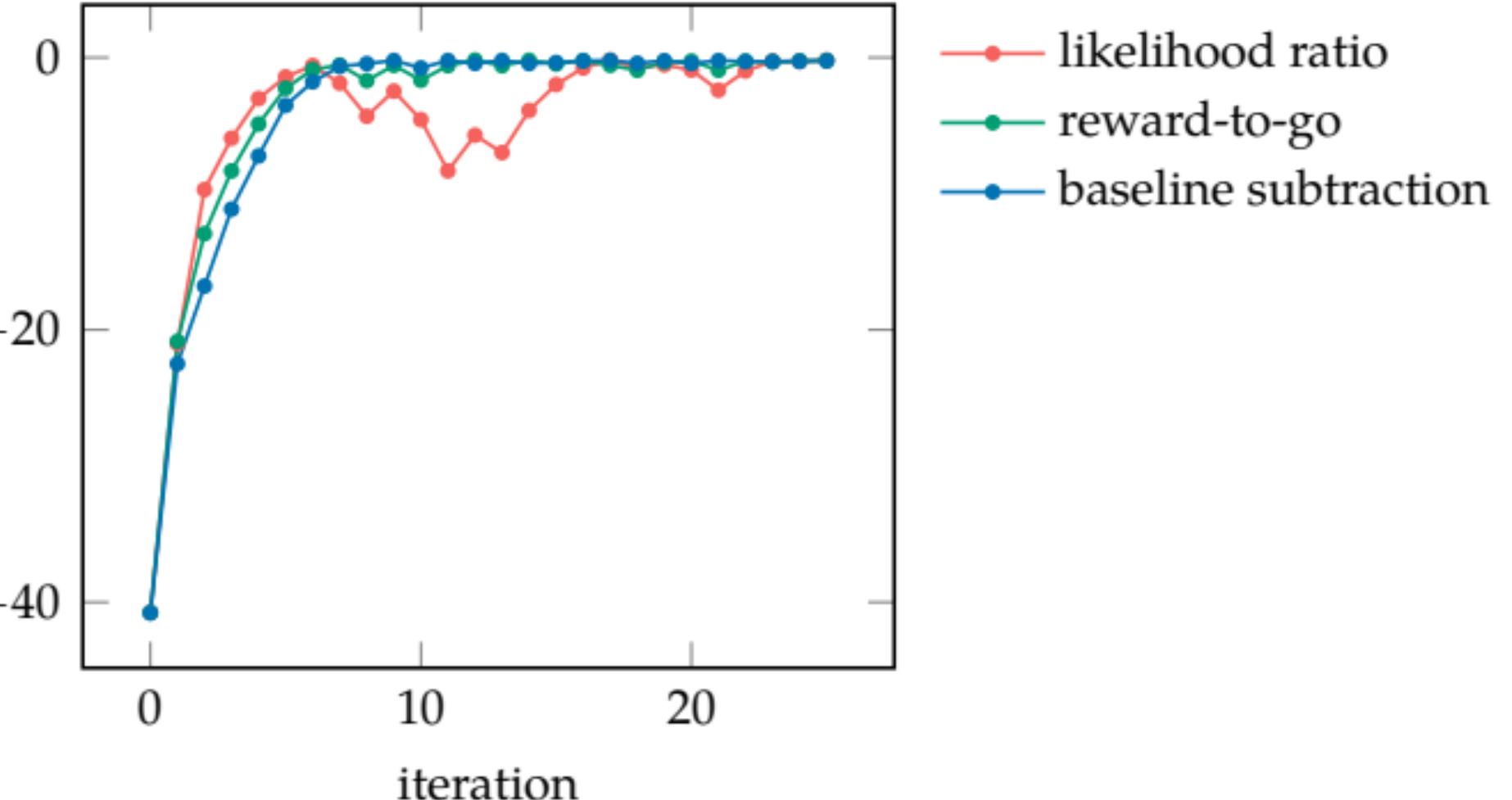
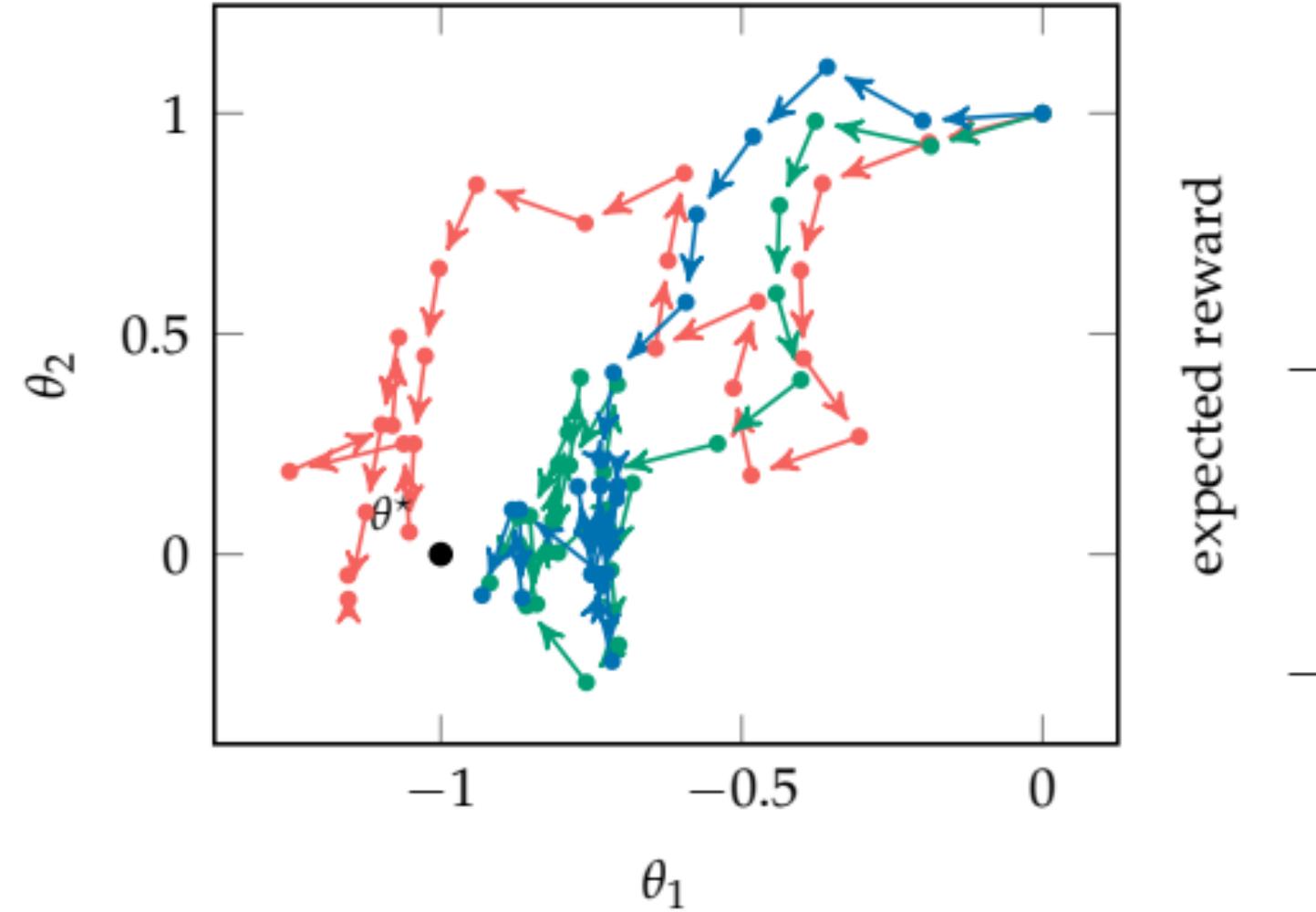
$$\mathbb{E}_\tau[f(\tau)]$$

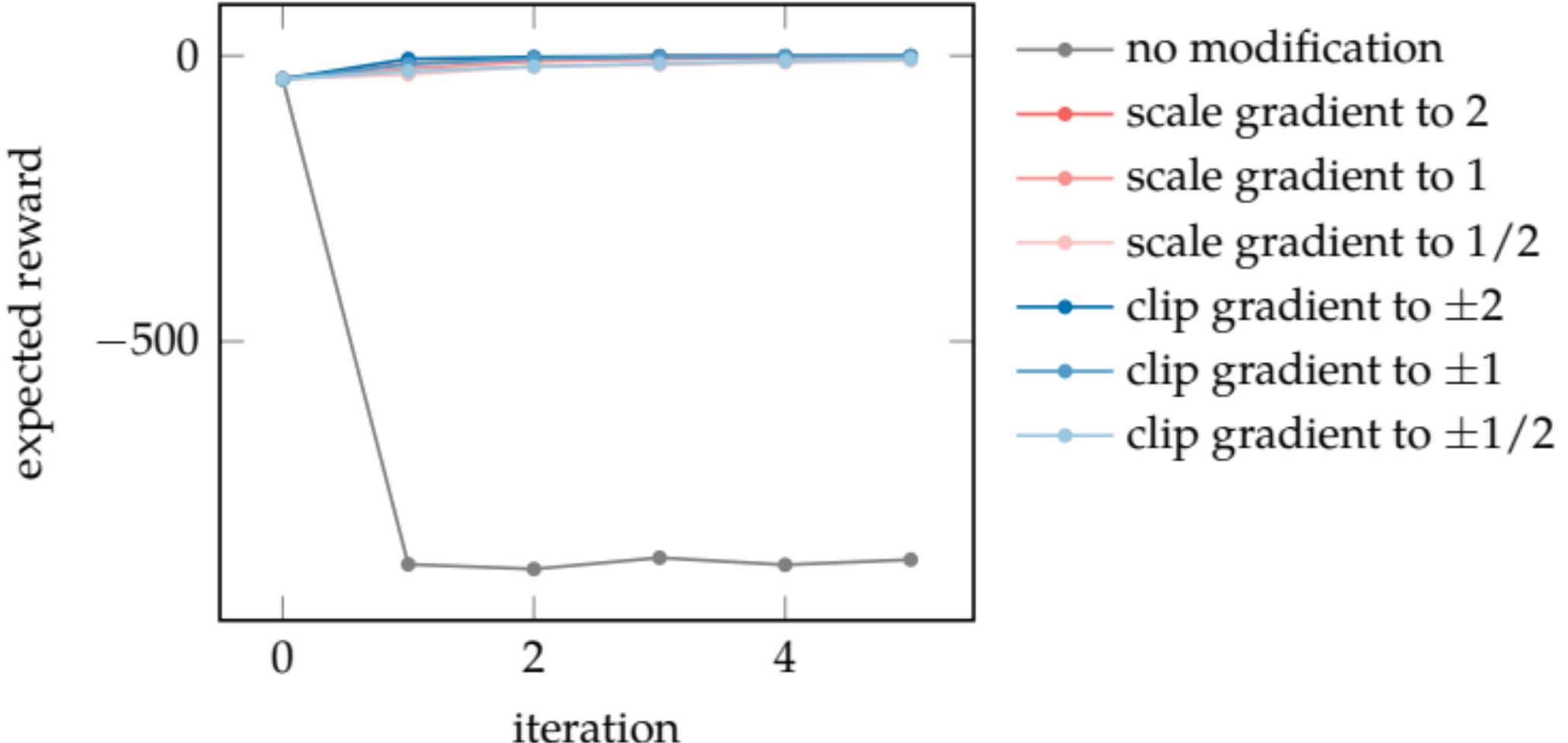
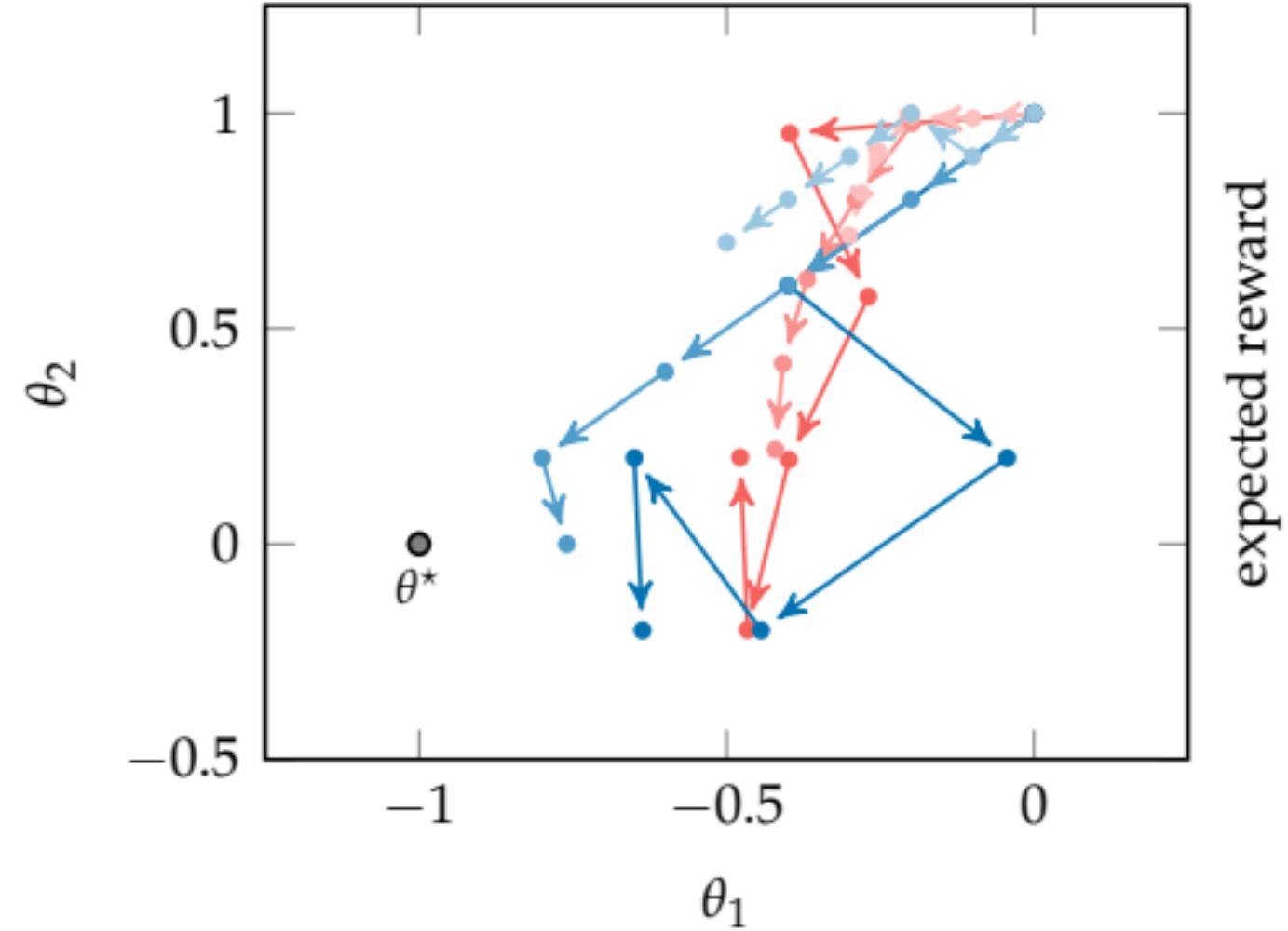


=

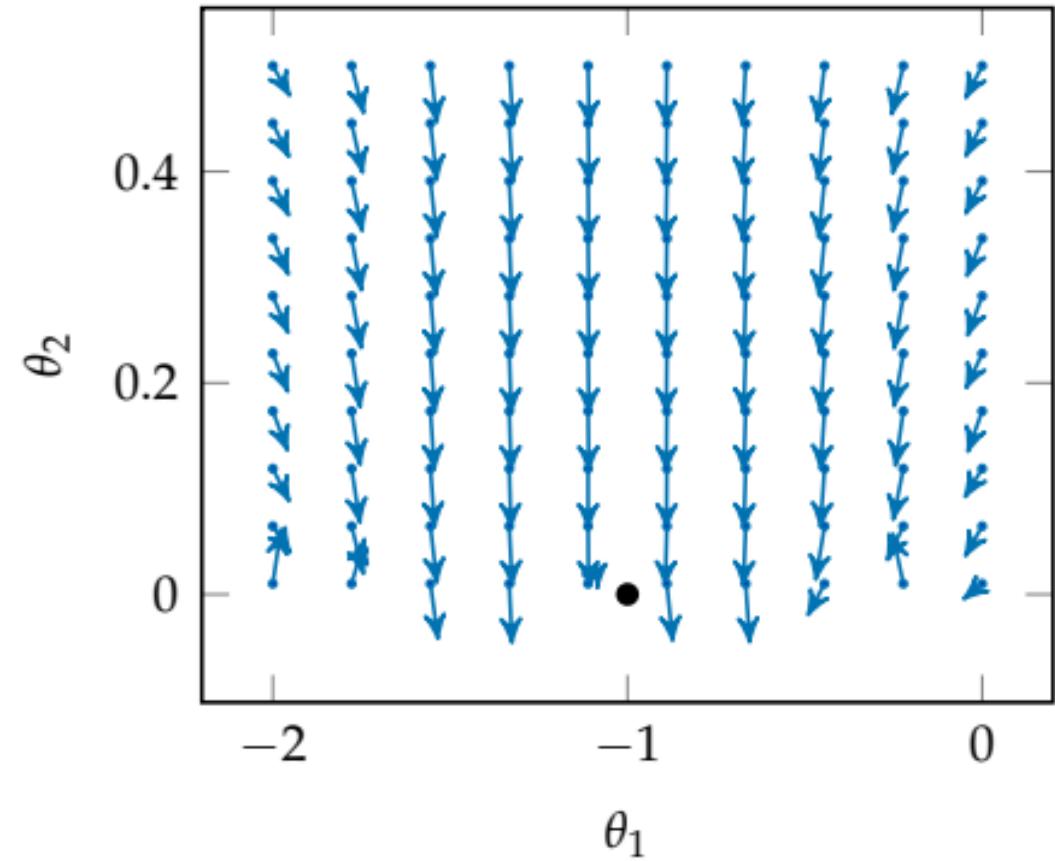
$$\mathbb{E}_{\tau_{1:k}} \left[\mathbb{E}_{\tau_{k+1:d}} [f(\tau)] \right]$$



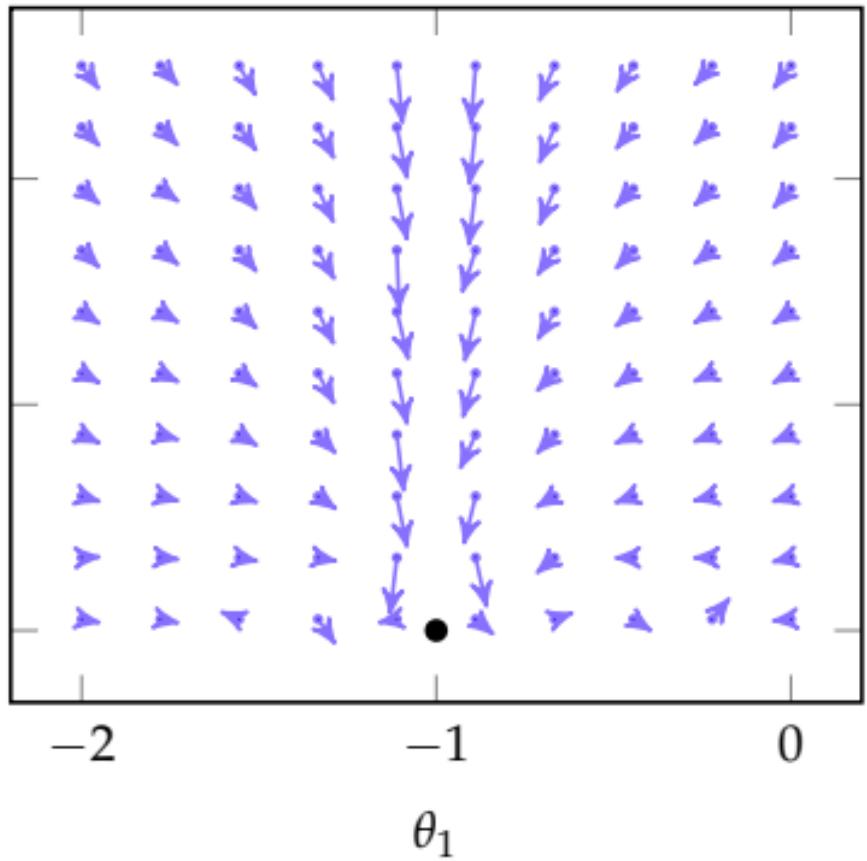




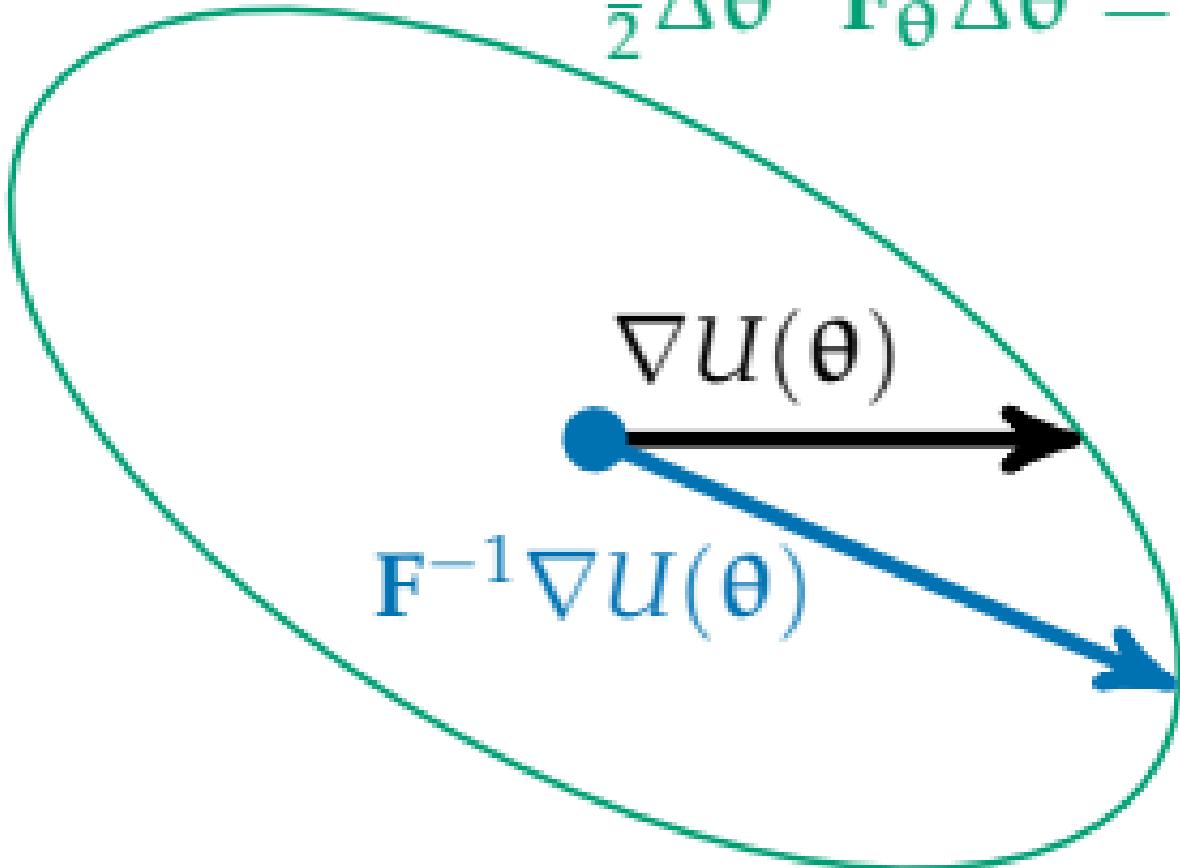
true gradient



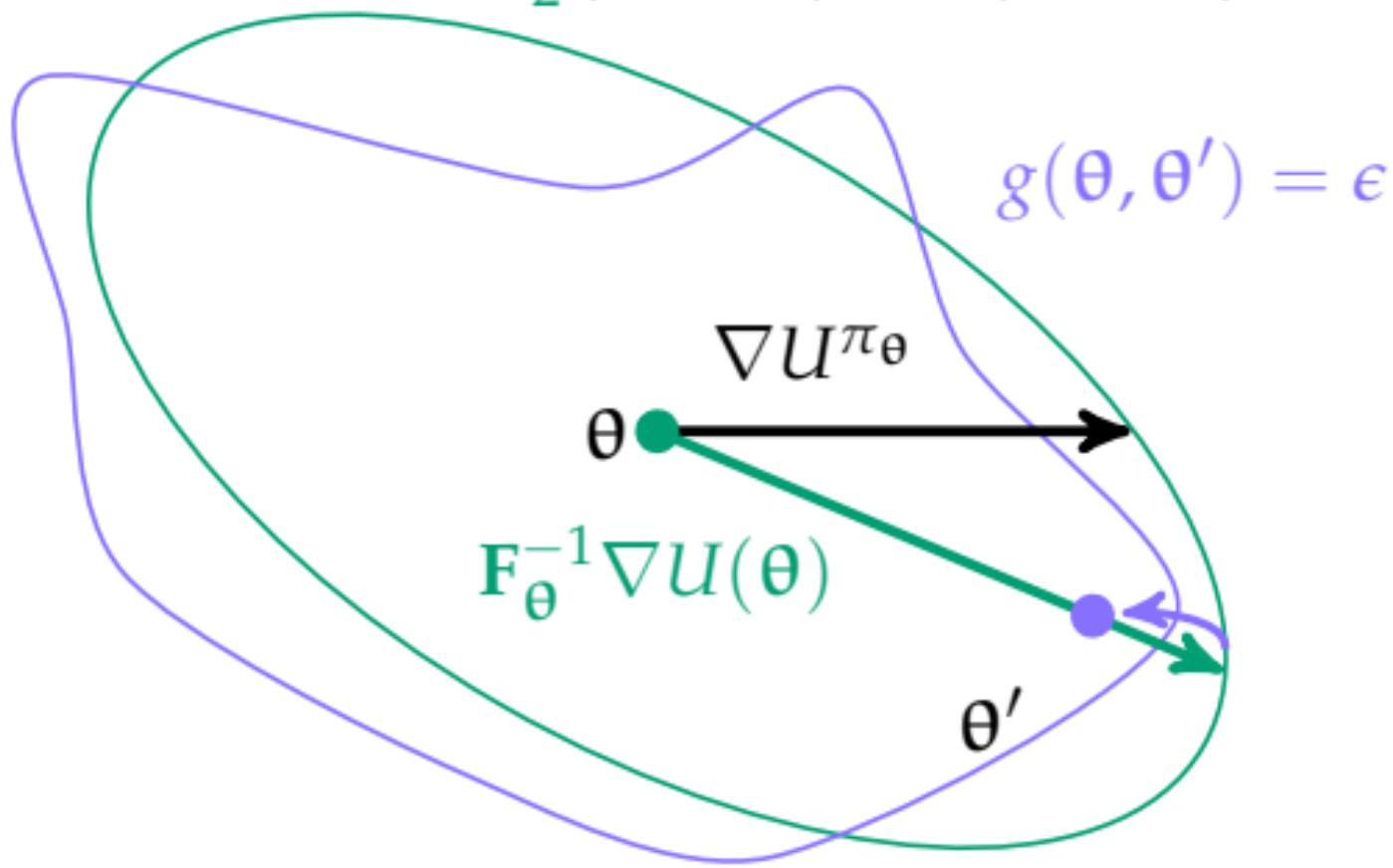
natural gradient

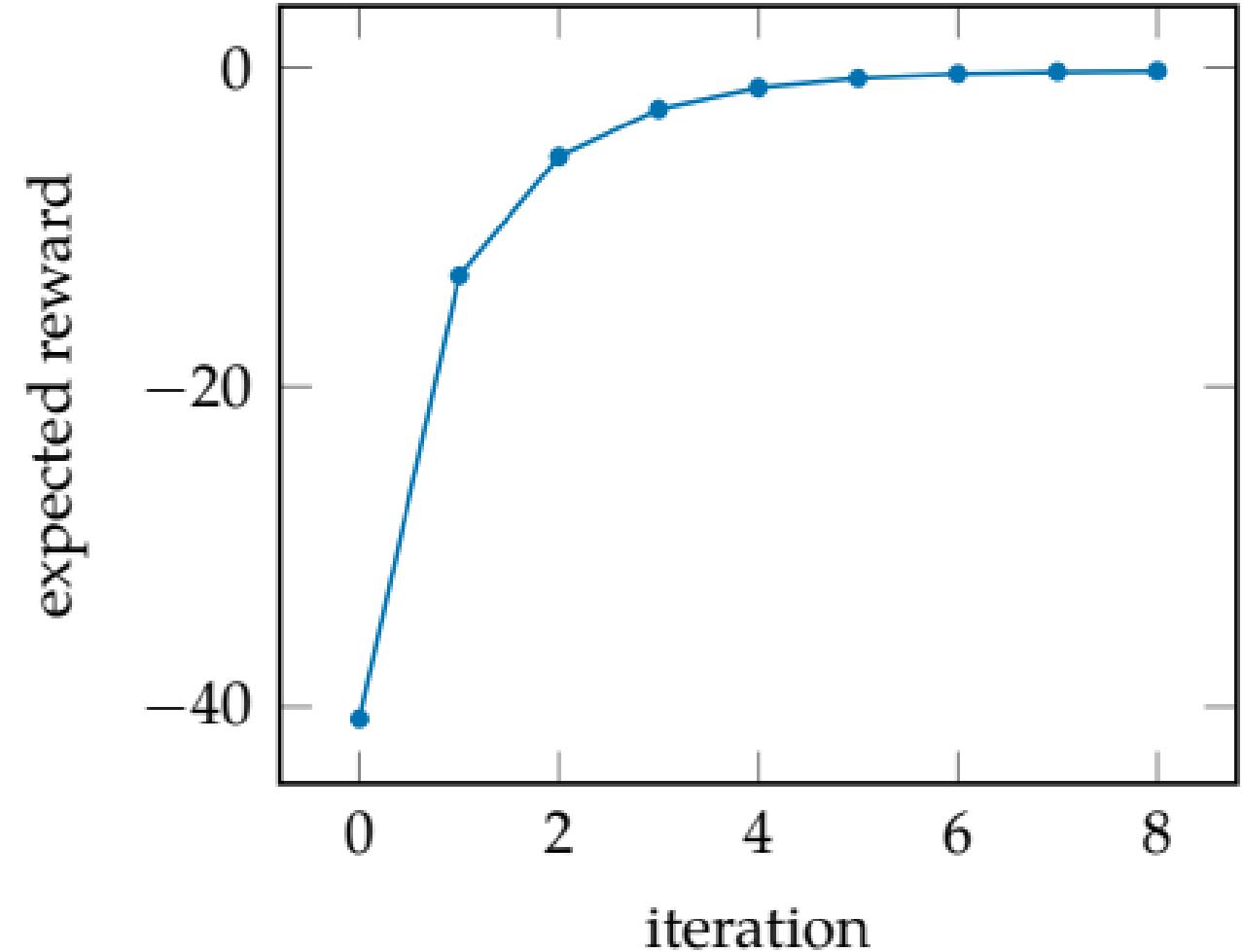
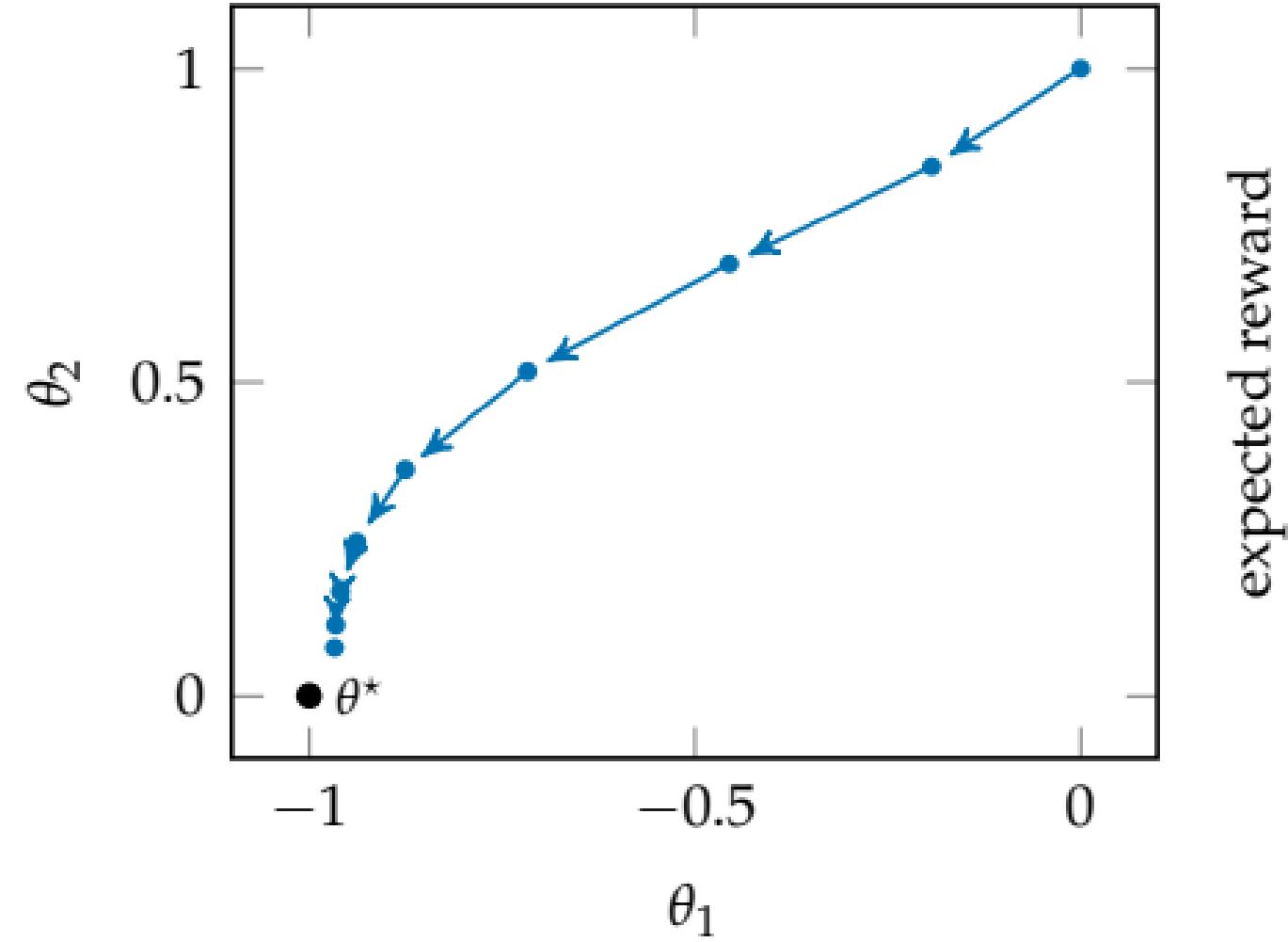


$$\frac{1}{2} \Delta\theta^\top F_\theta \Delta\theta = \epsilon$$



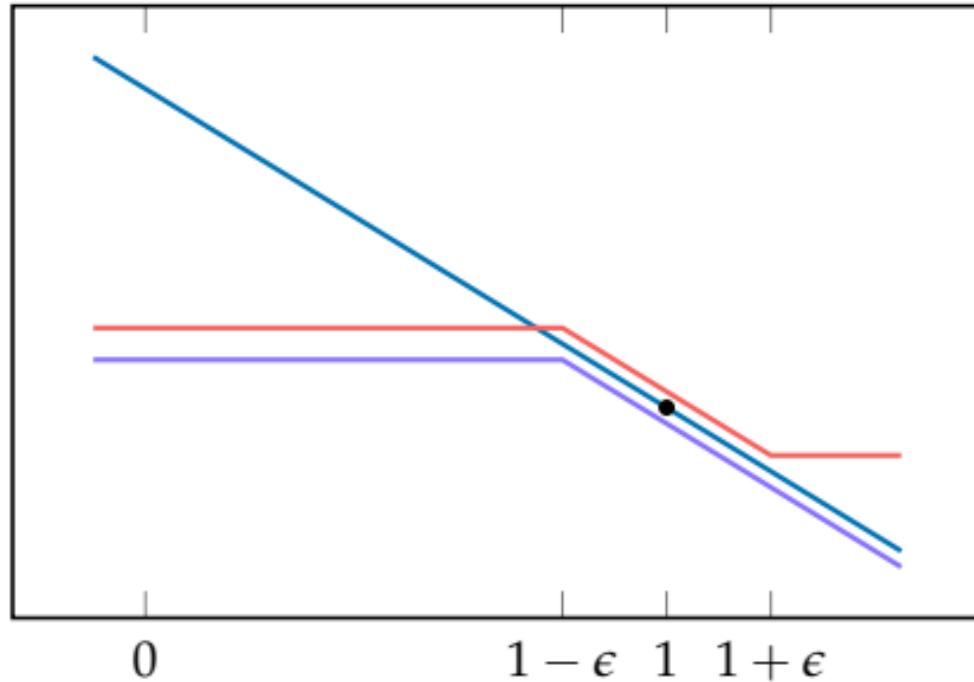
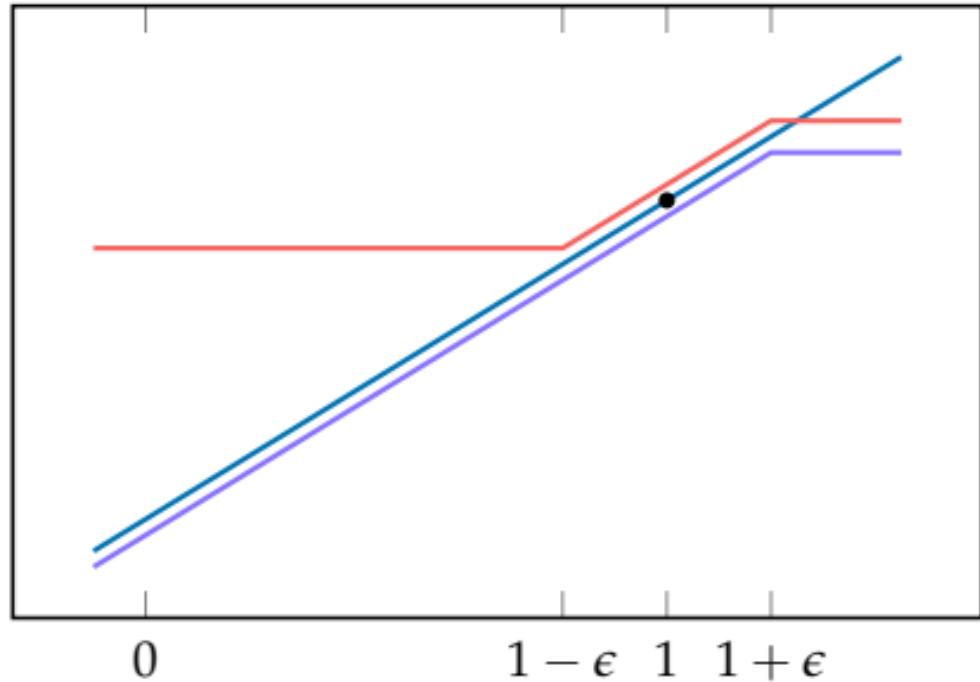
$$\frac{1}{2}(\theta' - \theta)^\top \mathbf{F}_\theta (\theta' - \theta) = \epsilon$$



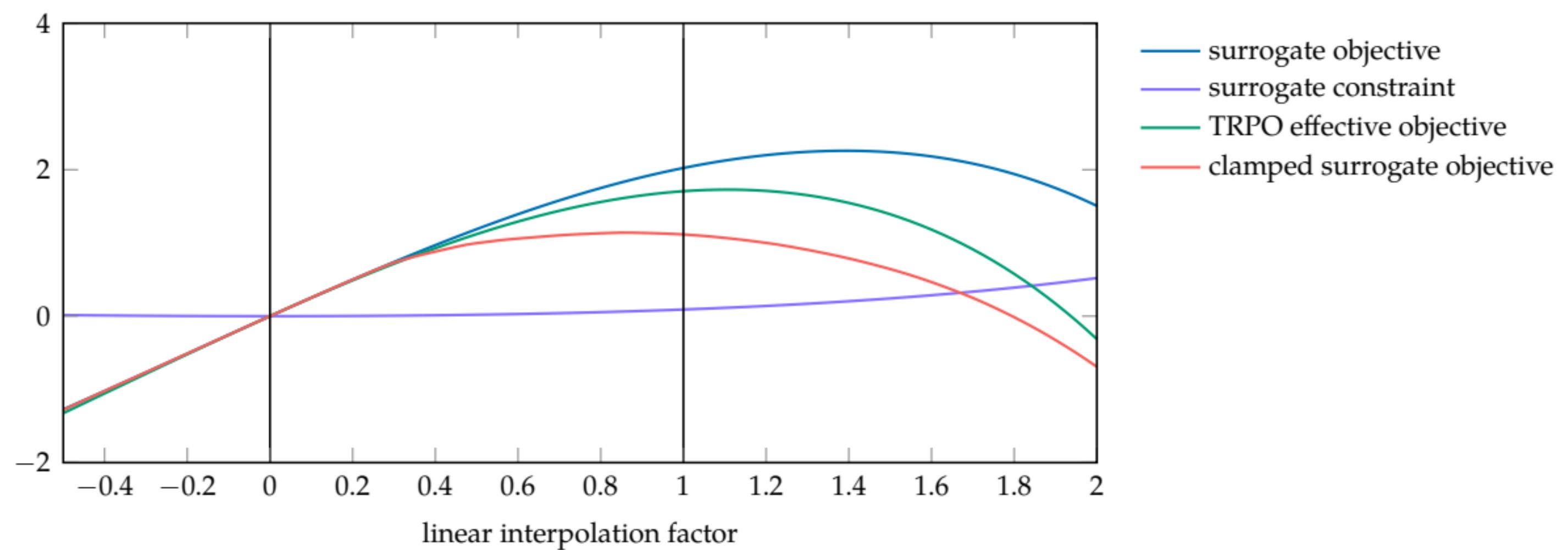


$A > 0$ $A < 0$

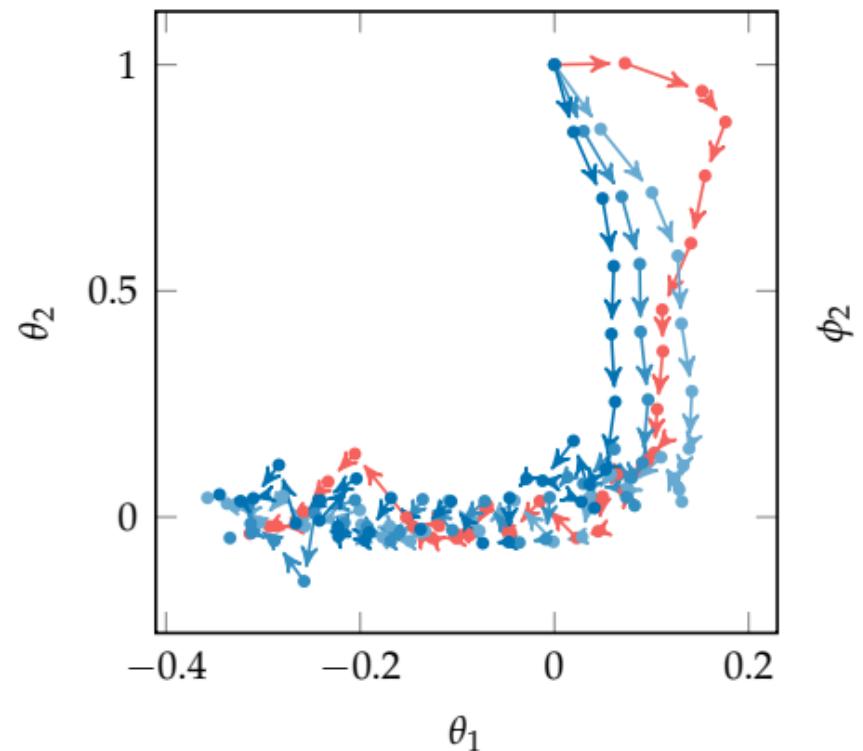
objective function



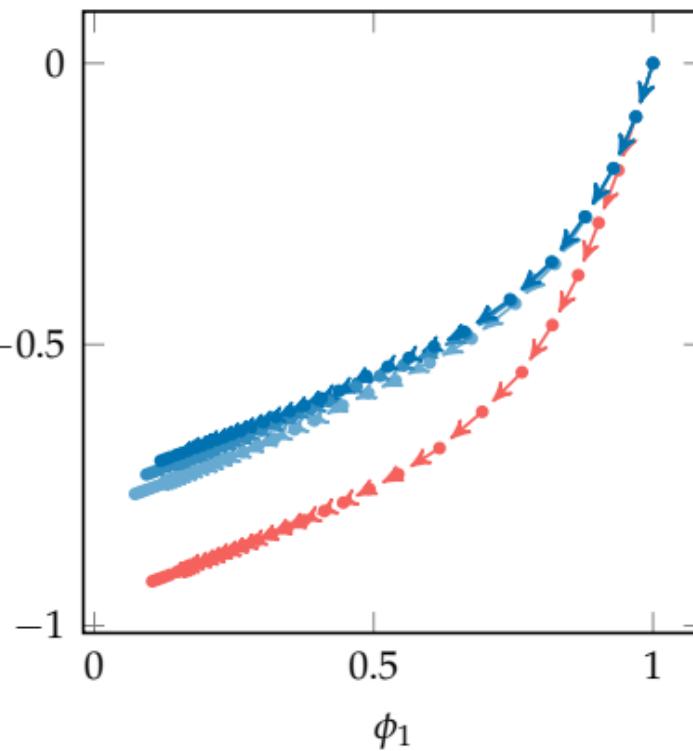
— original objective — clamped objective — lower-bound objective



policy parameterization



value function parameterization

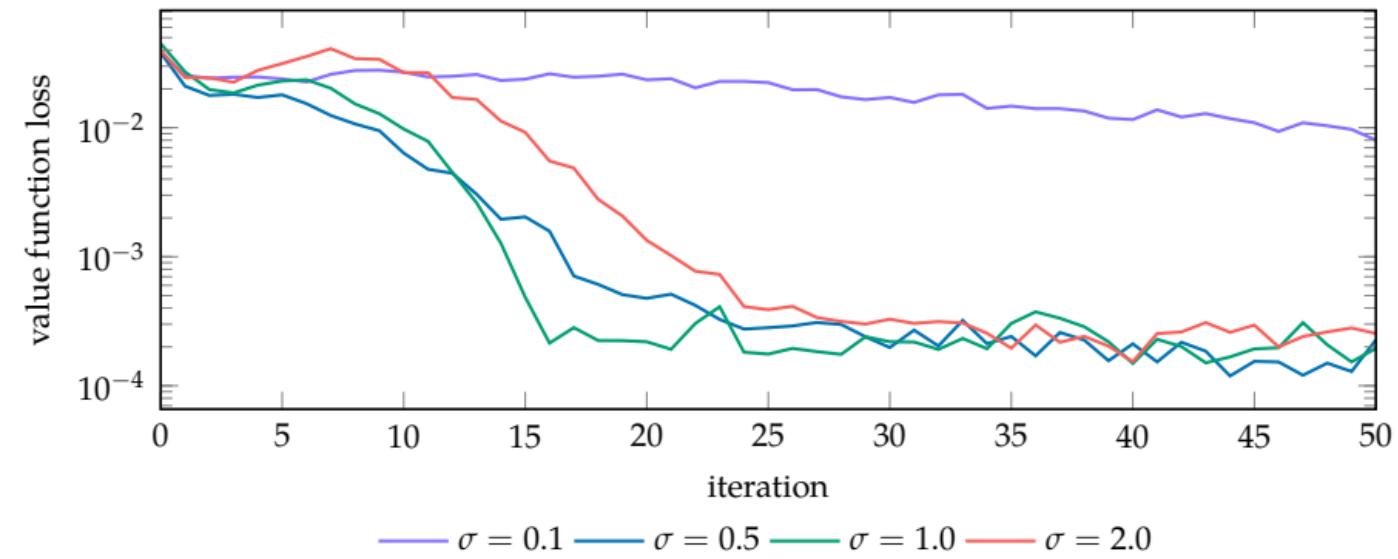
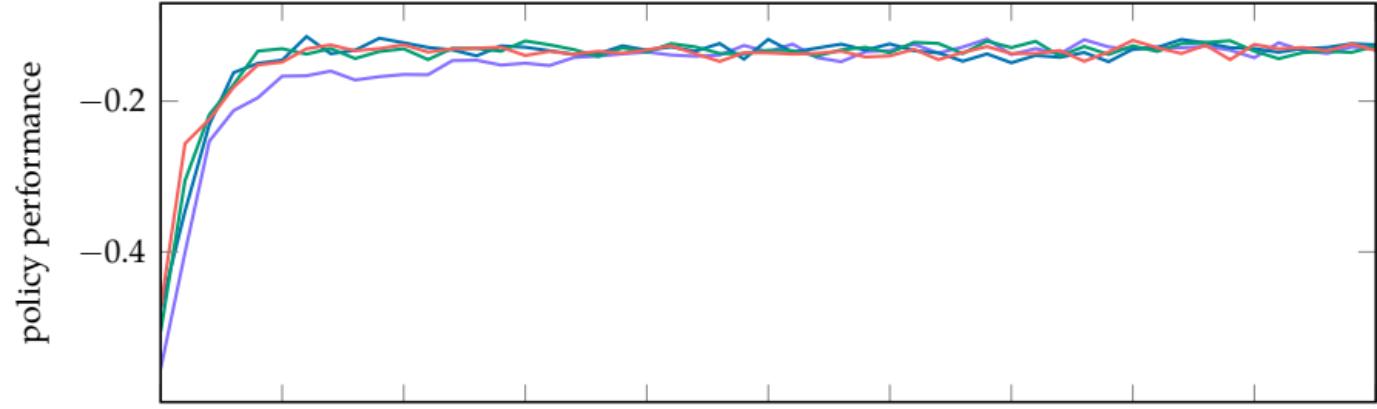


—●— actor-critic

—●— generalized advantage estimation, $\lambda = 0.7$

—●— generalized advantage estimation, $\lambda = 0.5$

—●— generalized advantage estimation, $\lambda = 0.9$



$\times 10^{-3}$

$p(\text{miss distance})$

2

1

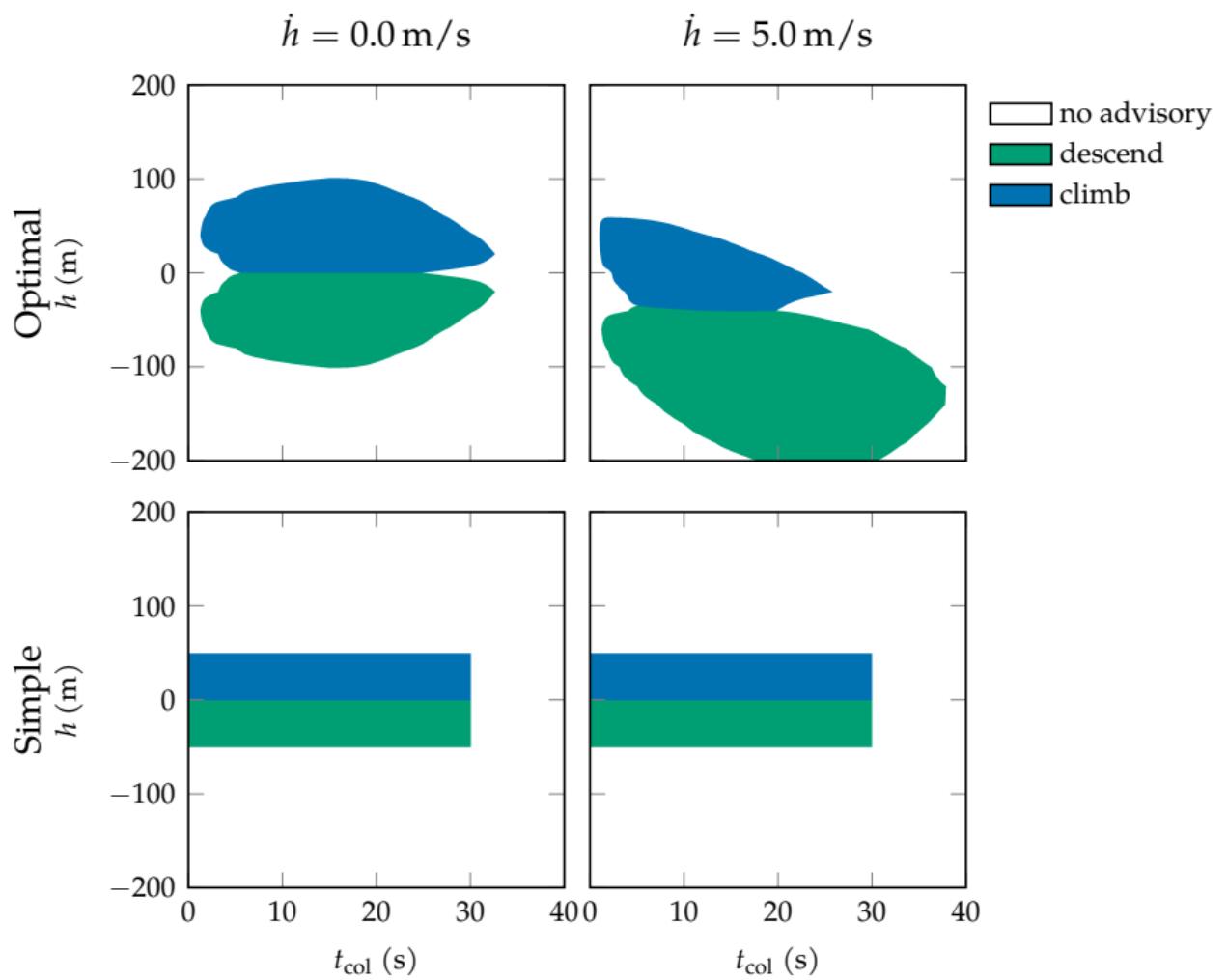
0

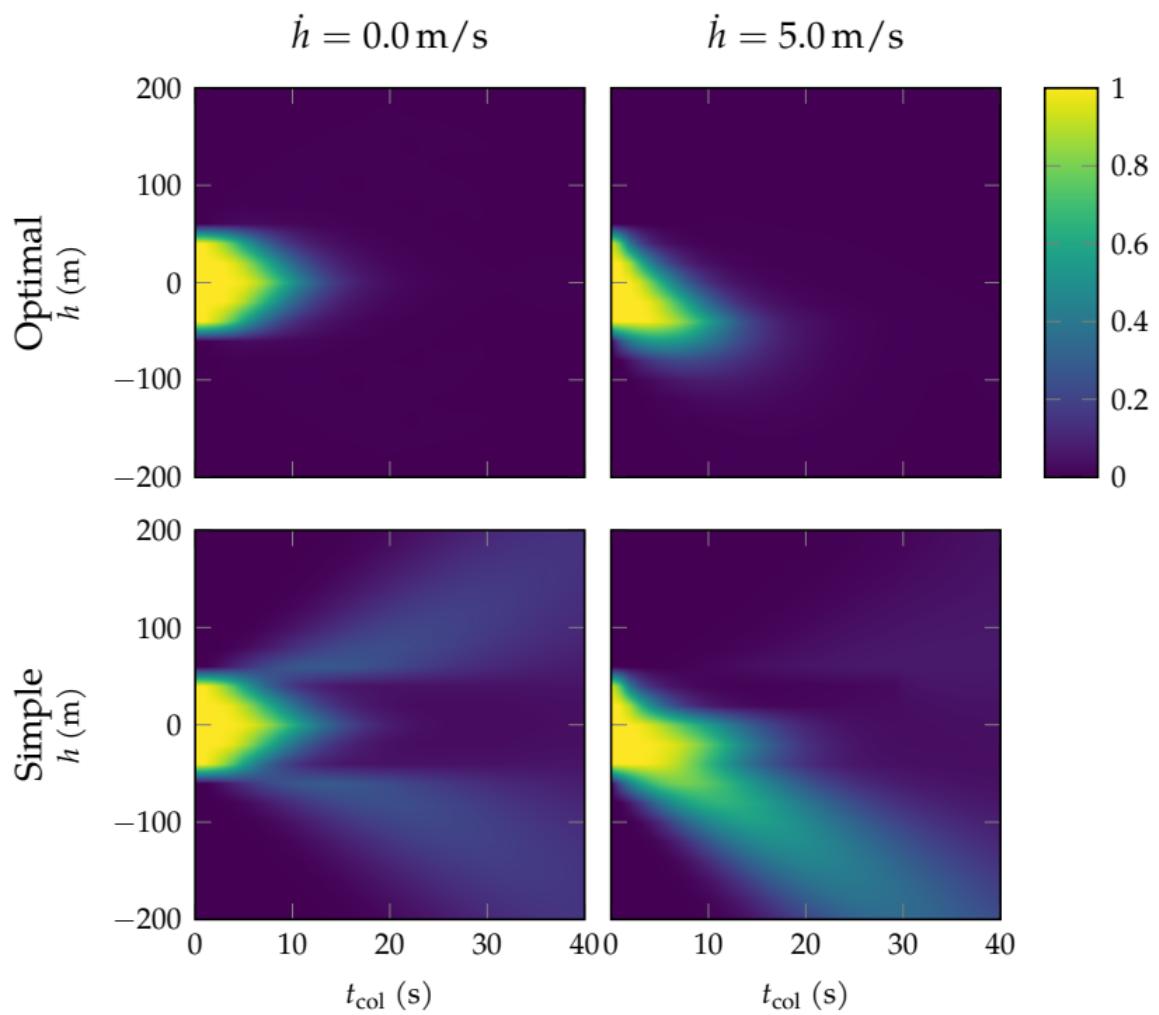
0

500

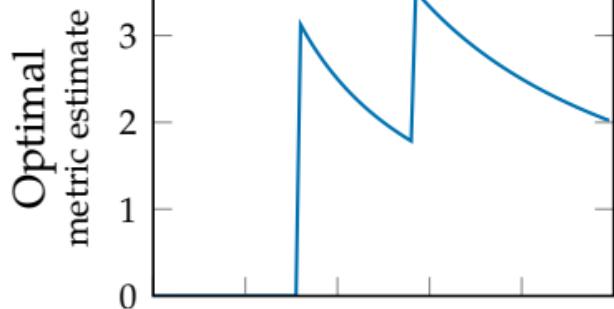
1,000

miss distance

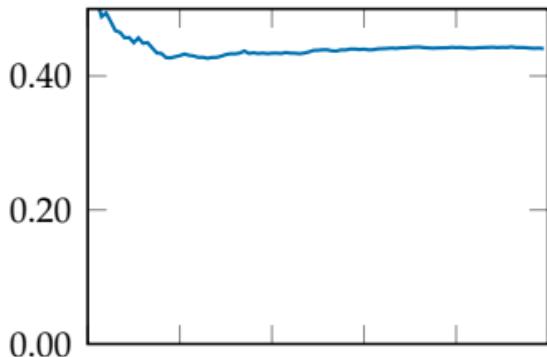




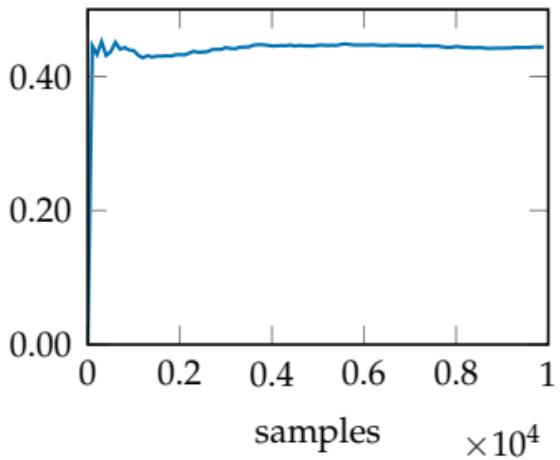
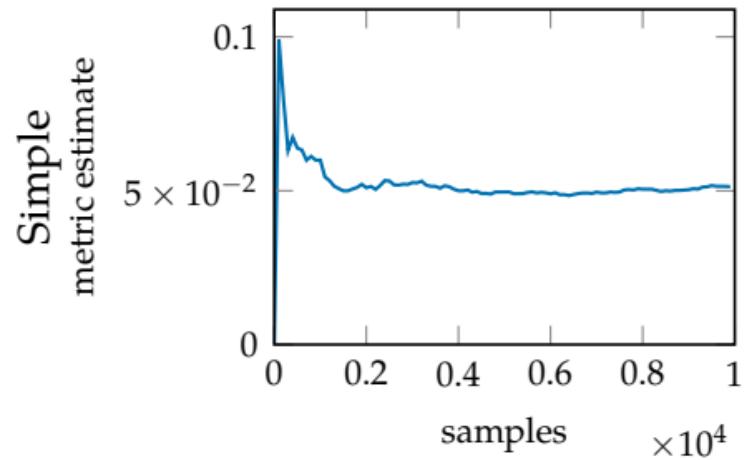
$\times 10^{-4}$ Collision

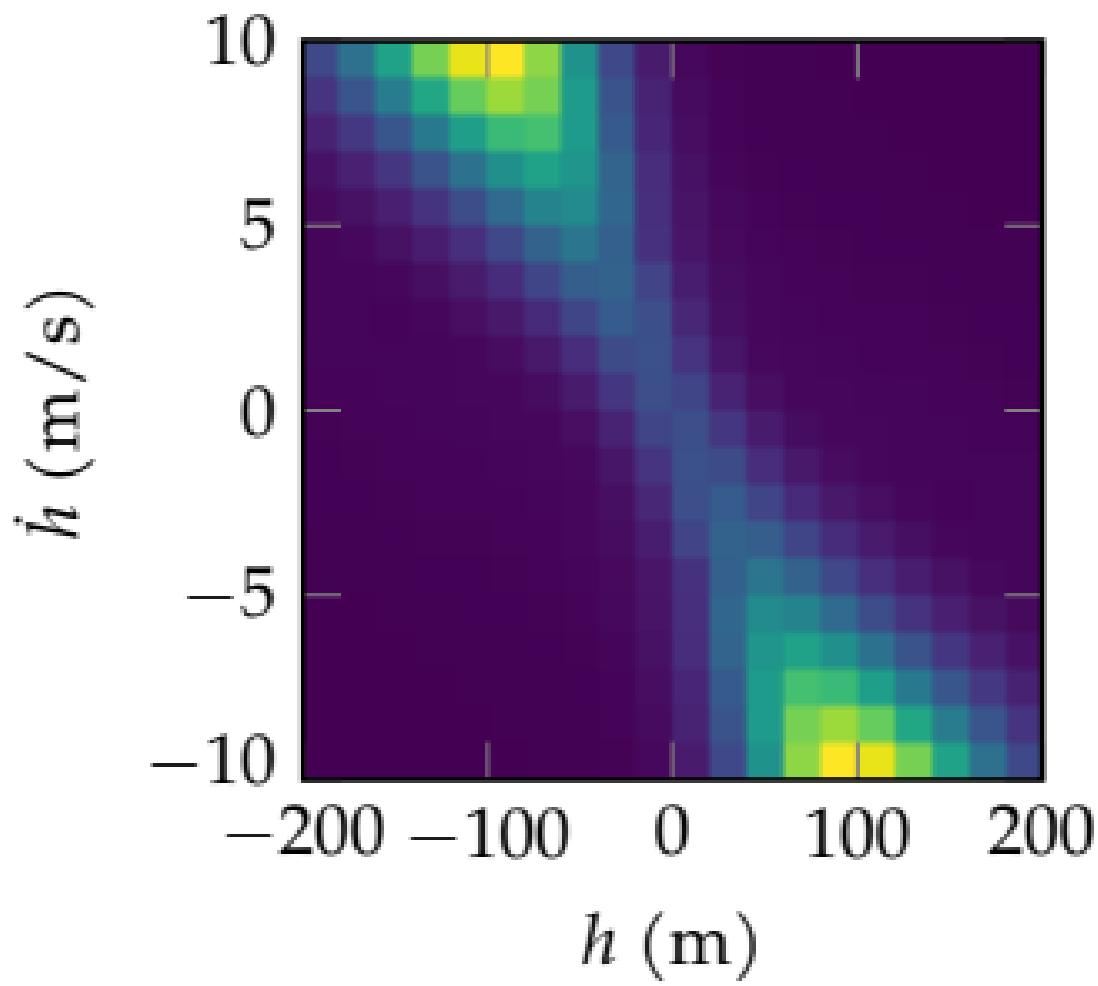


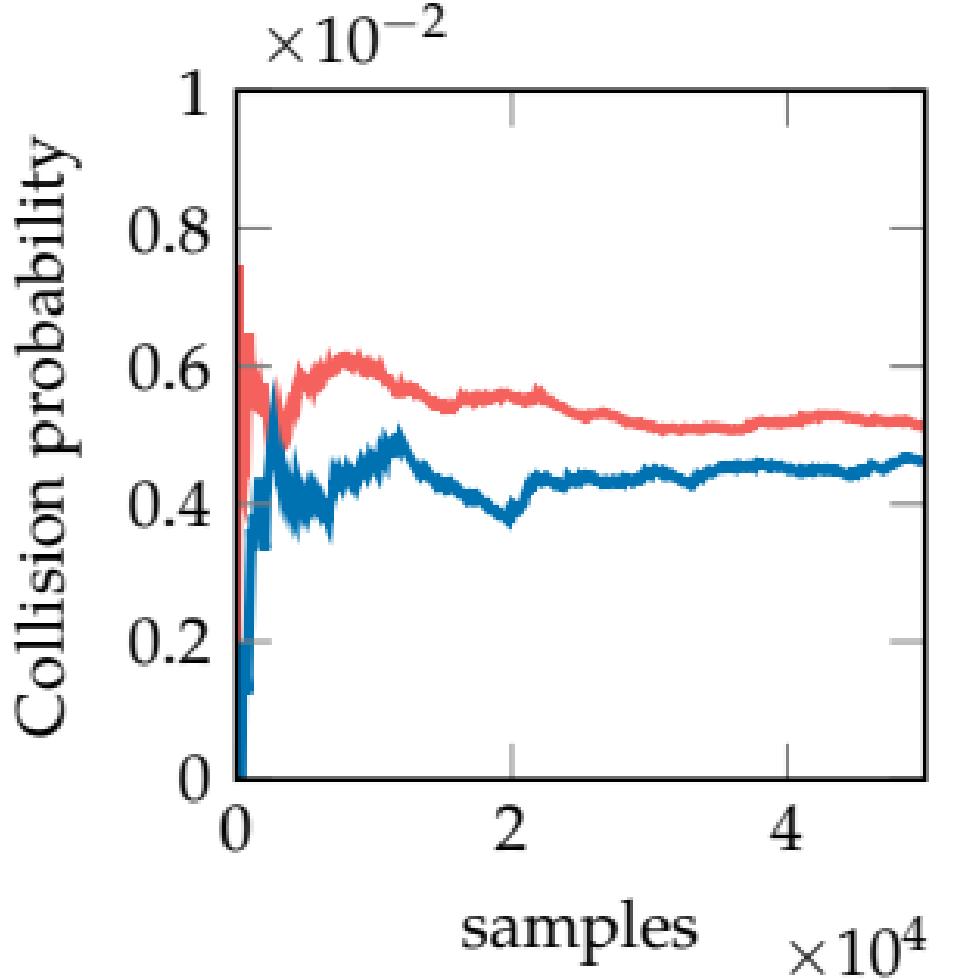
Advisory



Simple metric estimate

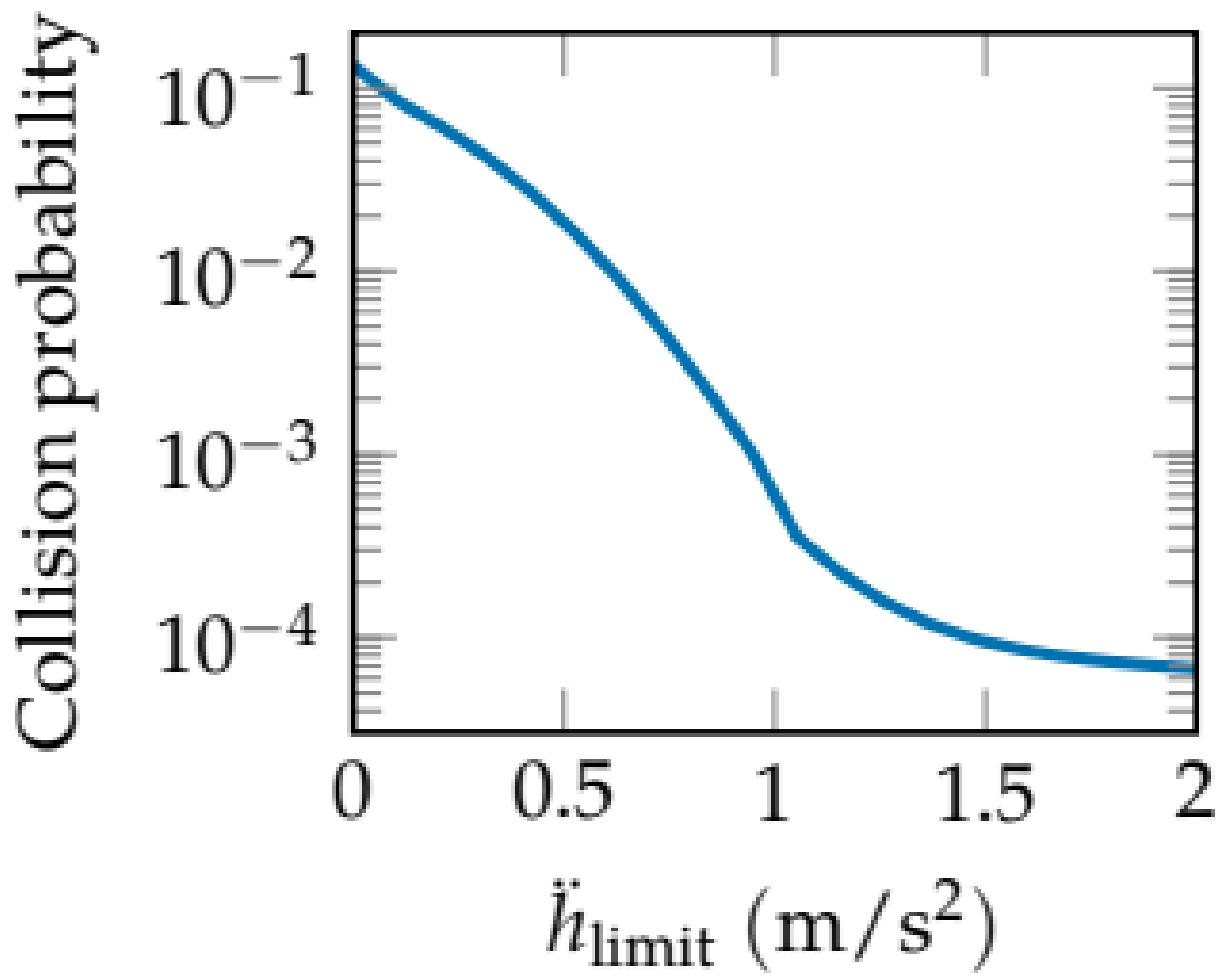


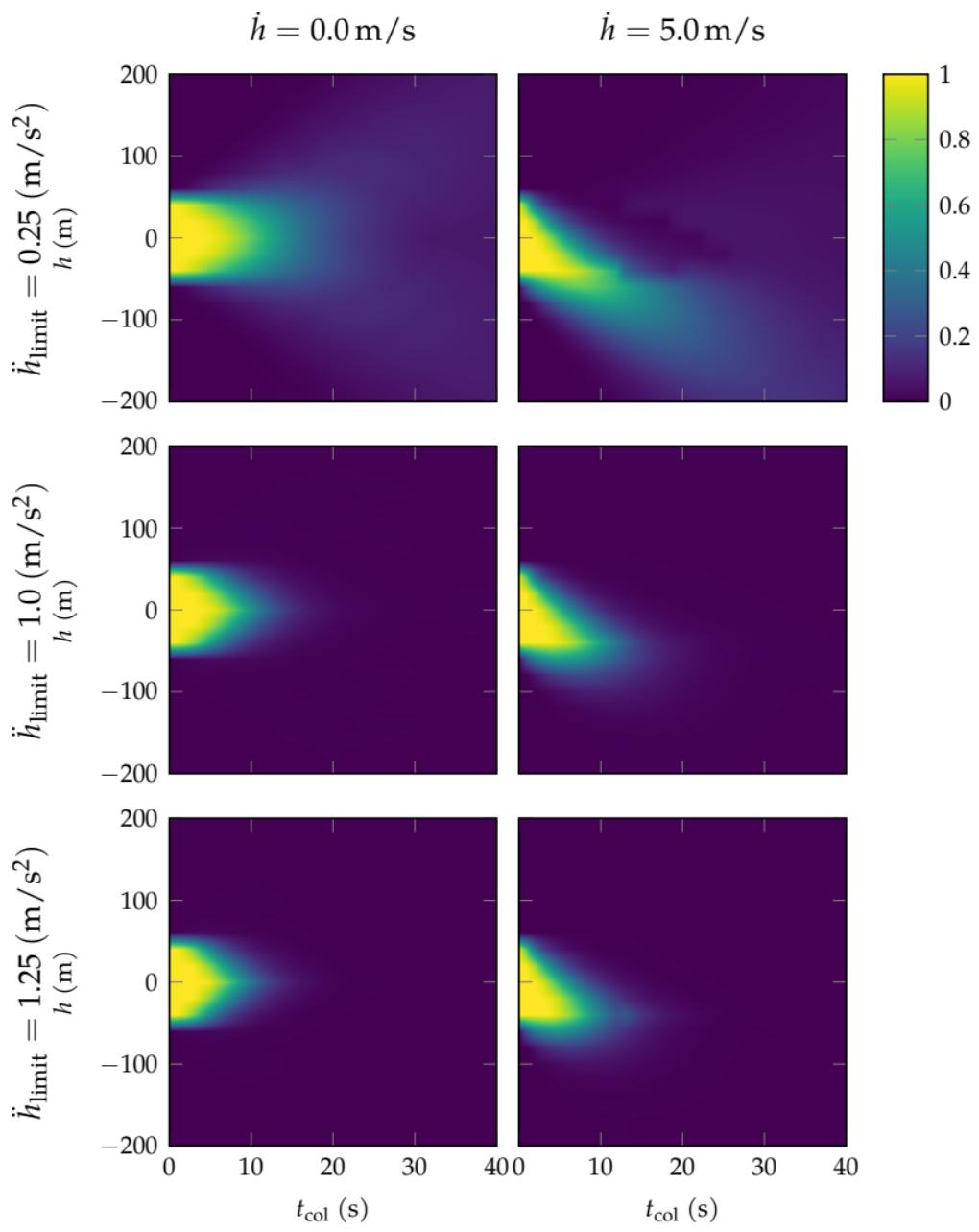


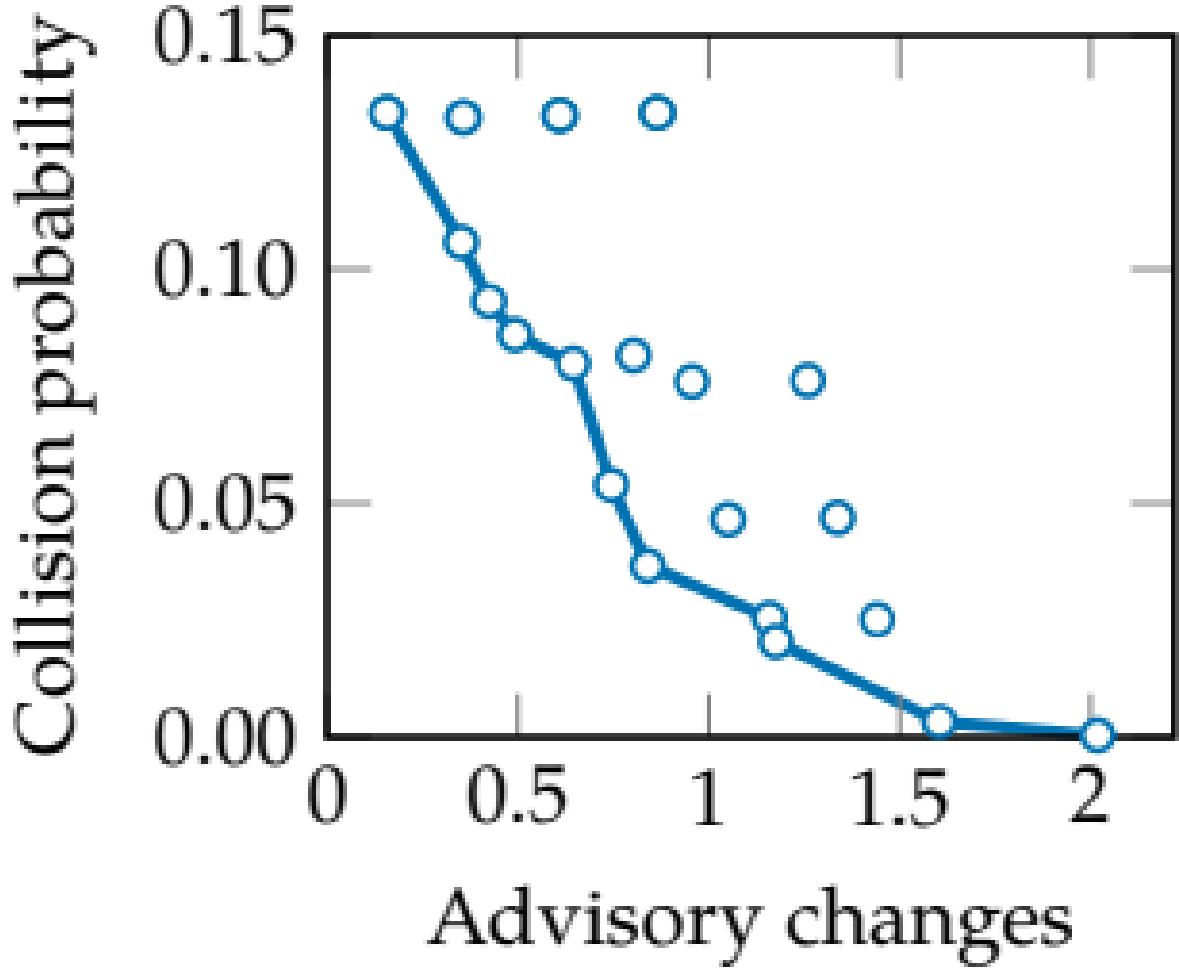


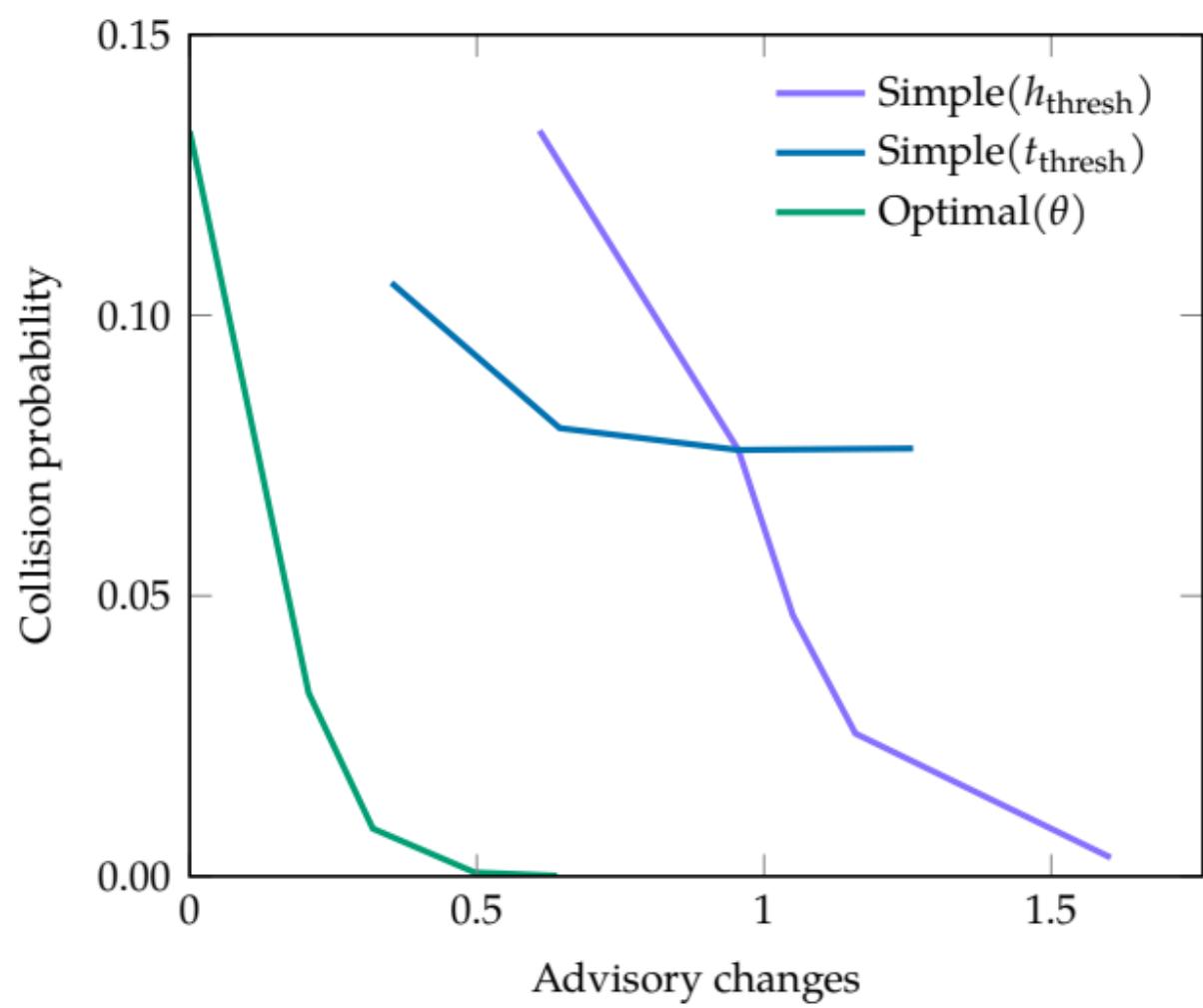
— importance sampling

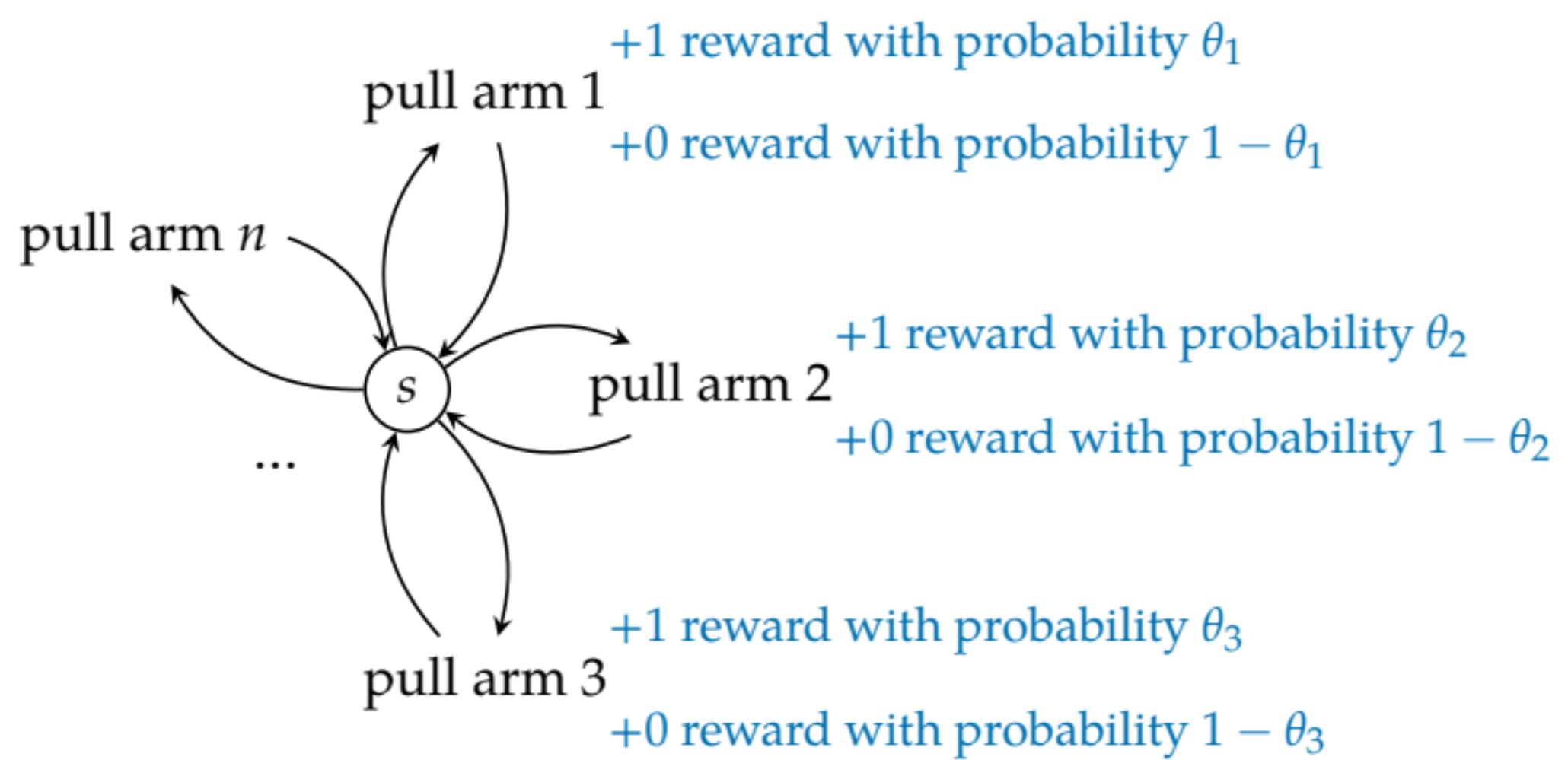
— direct sampling

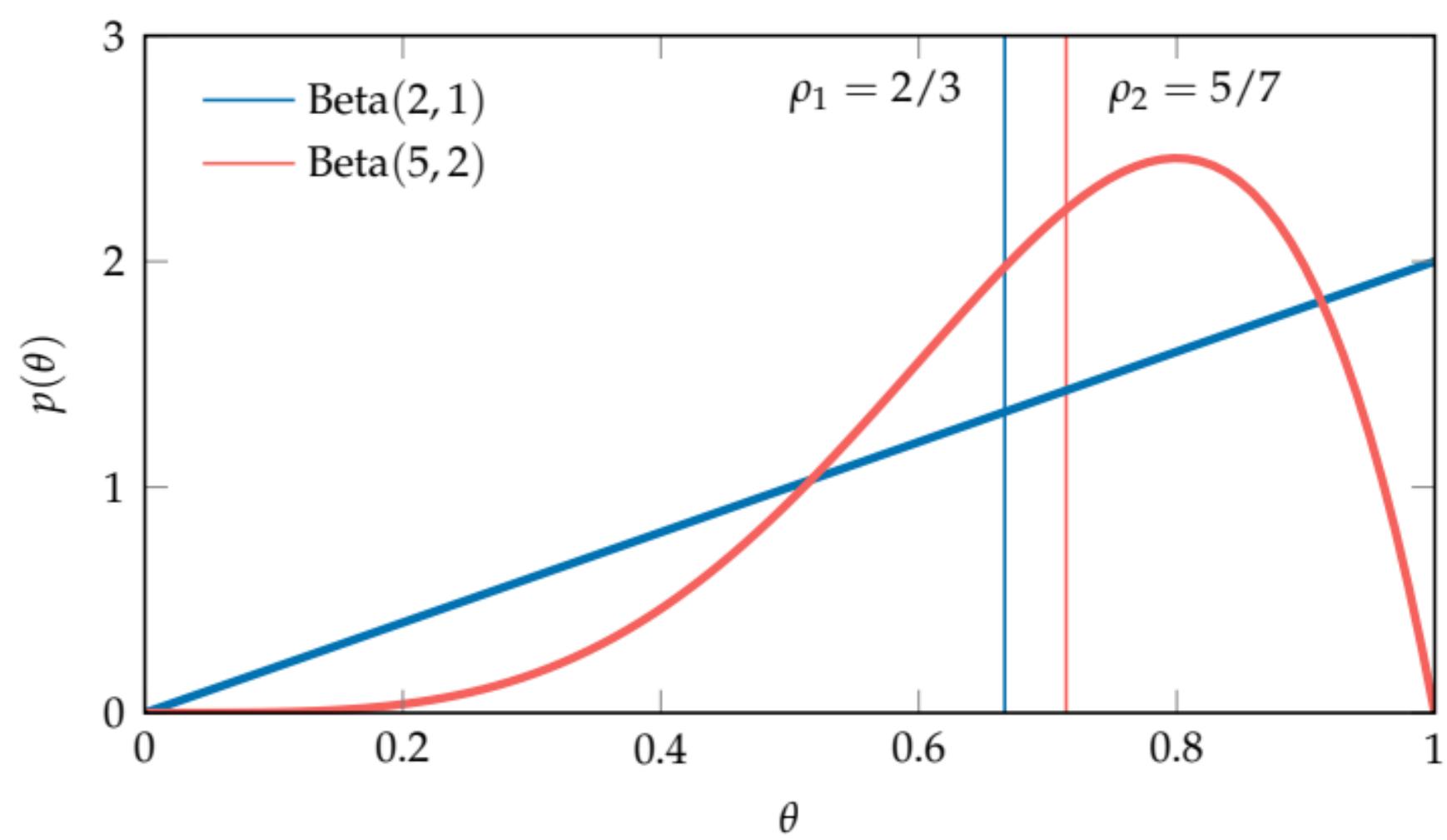


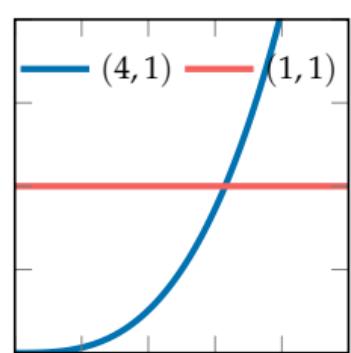
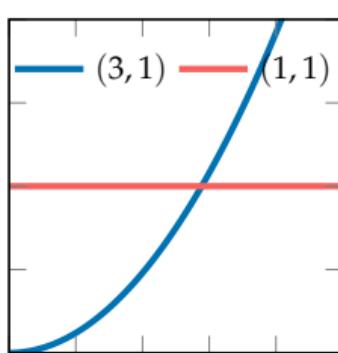
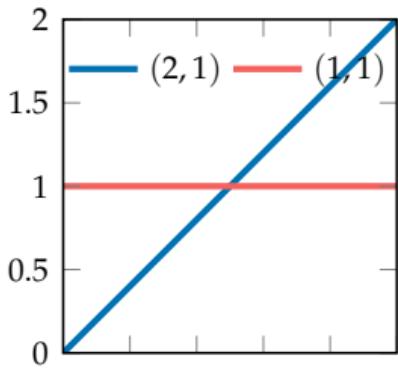
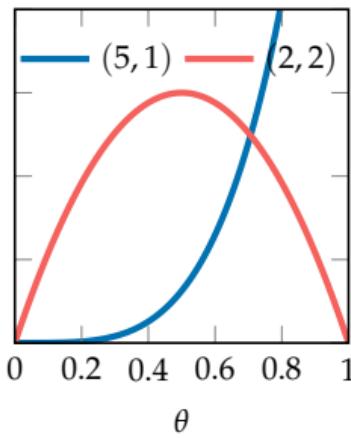
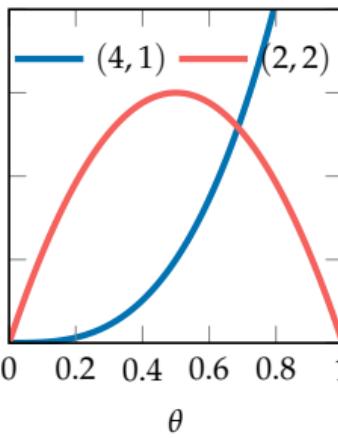
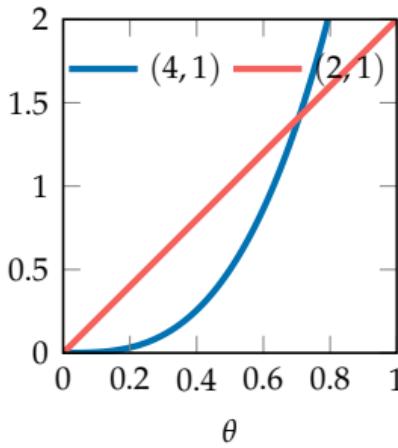


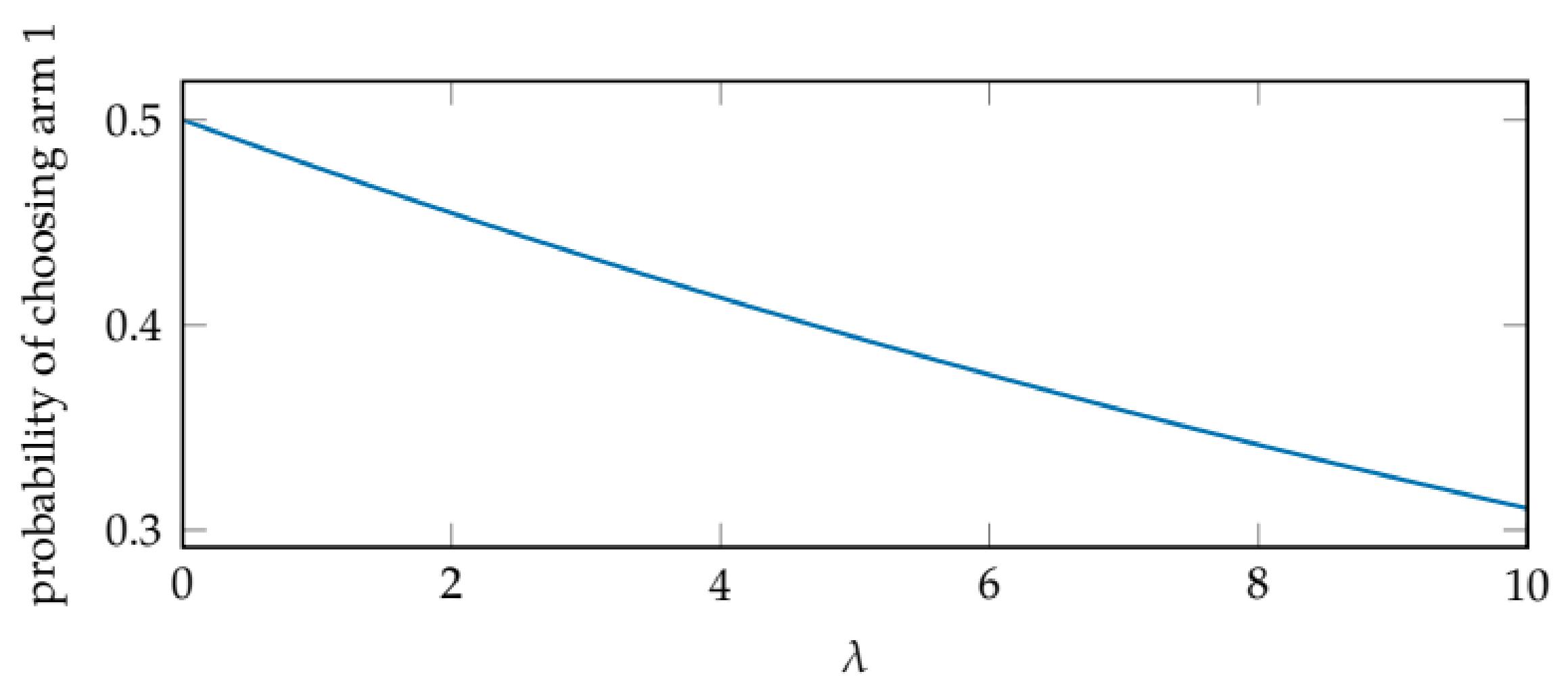


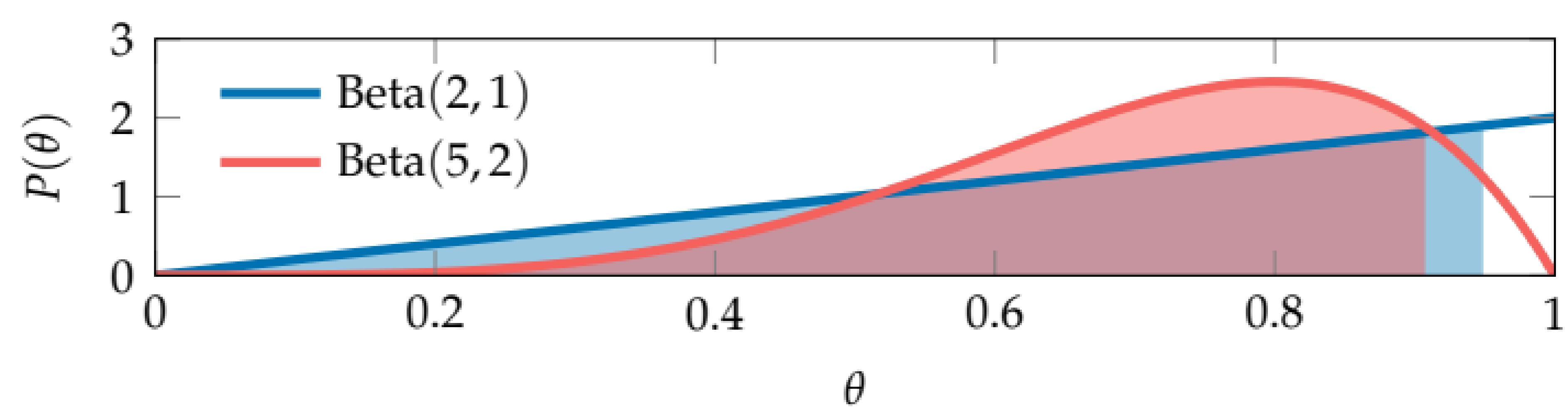






$t = 1, a = 1, r = 1$ $t = 2, a = 1, r = 1$ $t = 3, a = 1, r = 1$ $p(\theta)$  $t = 4, a = 2, r = 1$ $t = 5, a = 2, r = 0$ $t = 6, a = 1, r = 1$ $p(\theta)$ 





$[0, 0, 0, 0], U^* = 1.083$

pull 1

pull 2

 $[1, 0, 0, 0], U^* = 2/3$ $[0, 1, 0, 0], U^* = 1/2$ $[0, 0, 1, 0], U^* = 2/3$ $[0, 0, 0, 1], U^* = 1/2$

pull 1

pull 2

pull 1

pull 2

pull 1

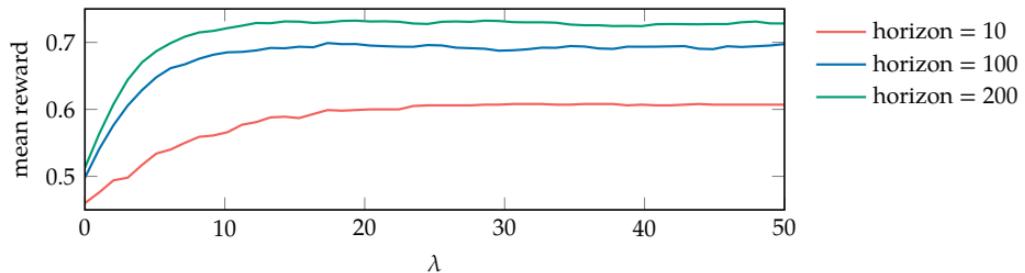
pull 2

pull 1

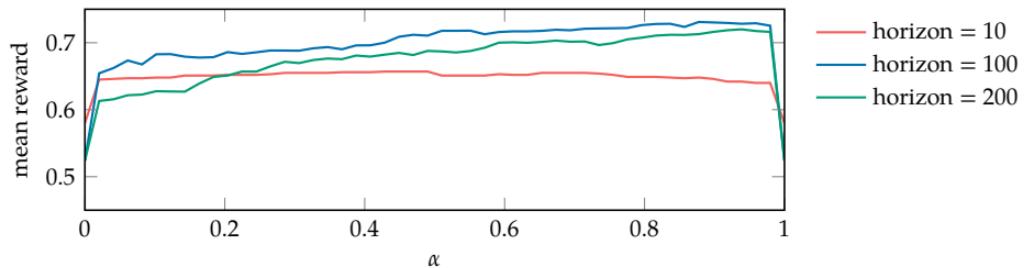
pull 2

 $[2, 0, 0, 0]$ $[1, 1, 0, 0]$ $[1, 0, 0, 1]$ $[0, 2, 0, 0]$ $[0, 1, 0, 1]$ $[1, 0, 1, 0]$ $[0, 1, 1, 0]$ $[0, 0, 2, 0]$ $[0, 0, 1, 1]$ $[0, 0, 0, 2]$

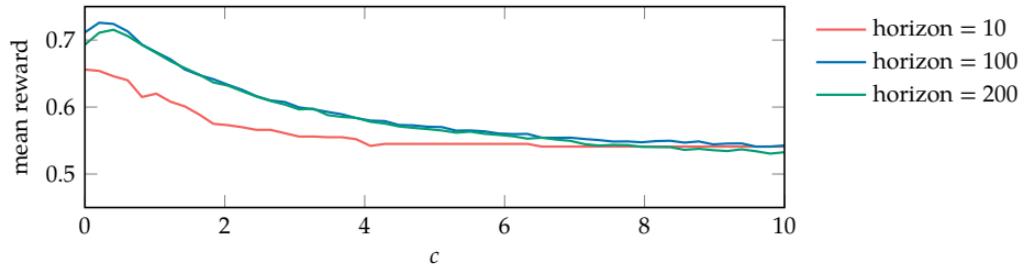
softmax exploration with constant precision

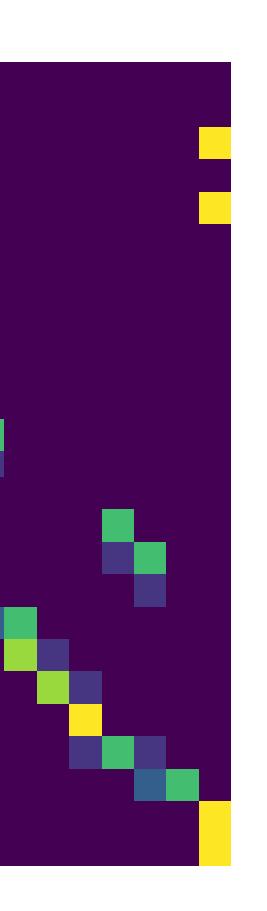
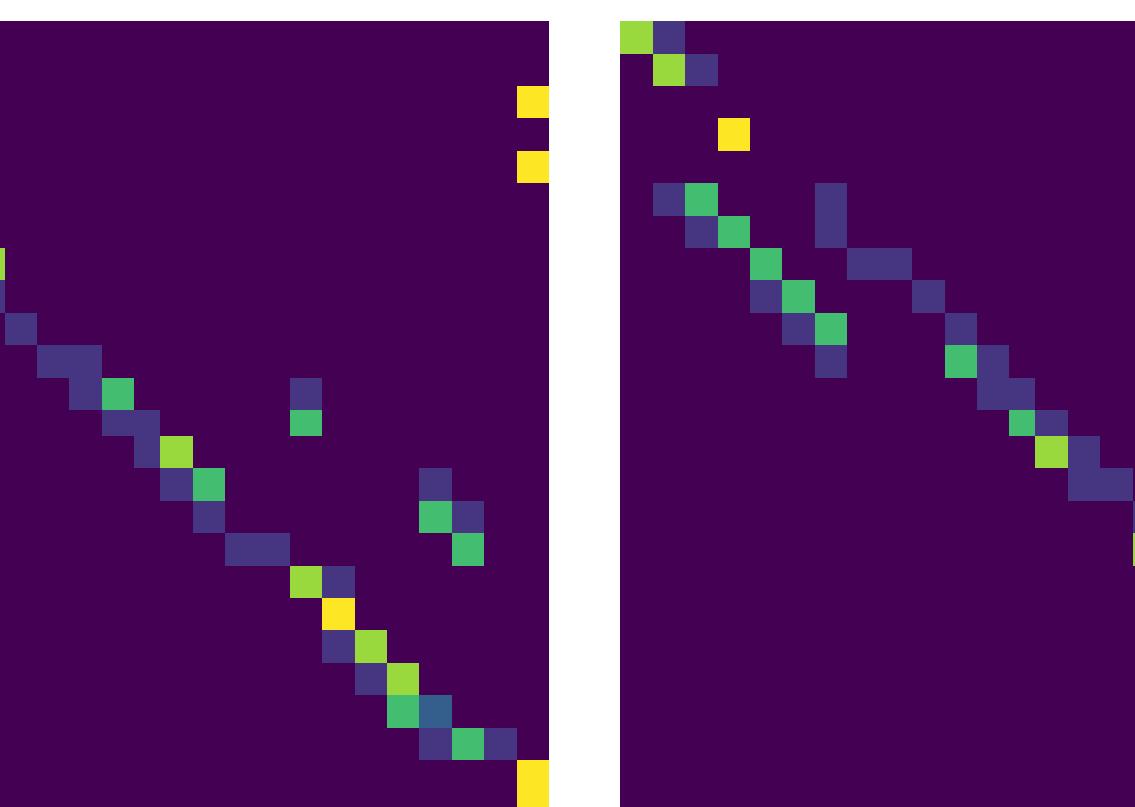
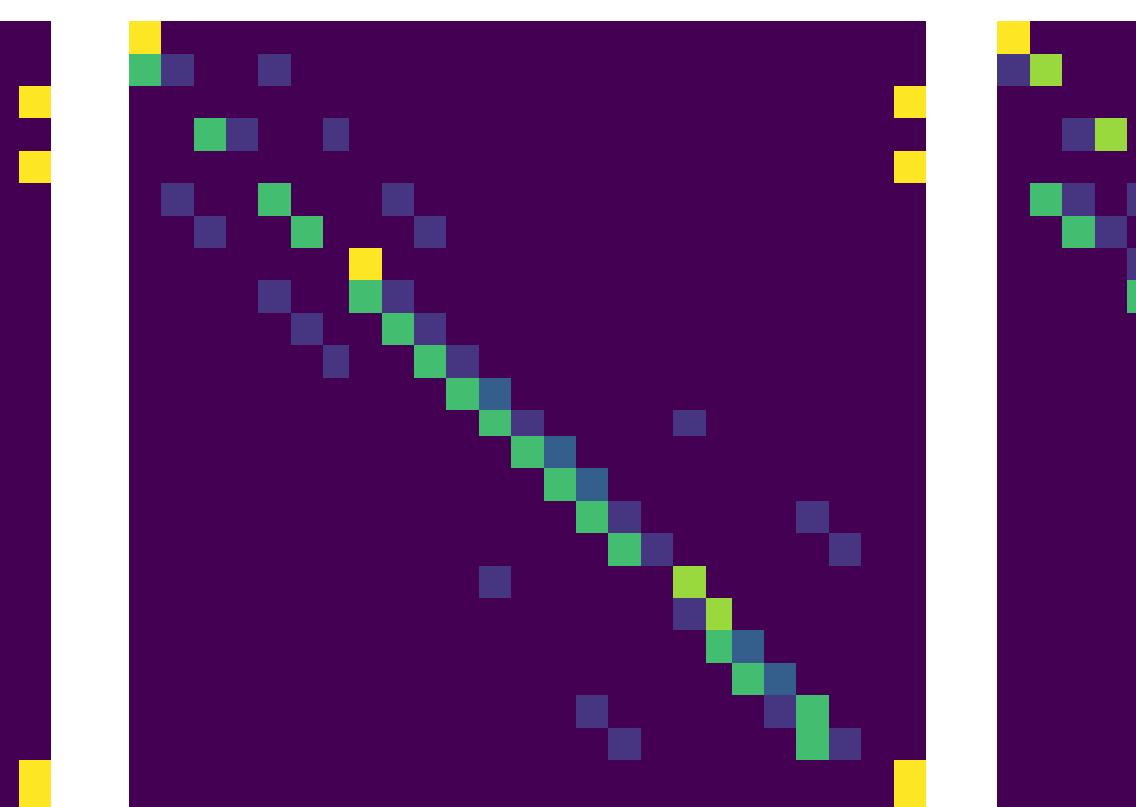
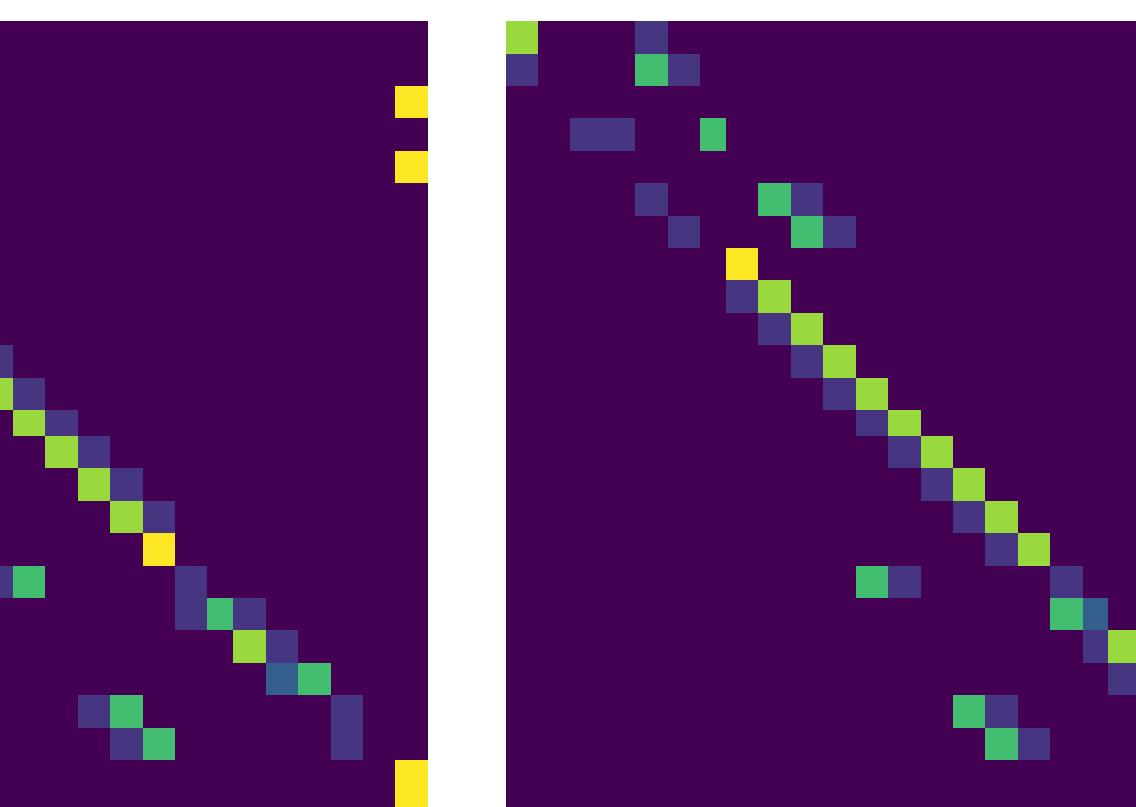
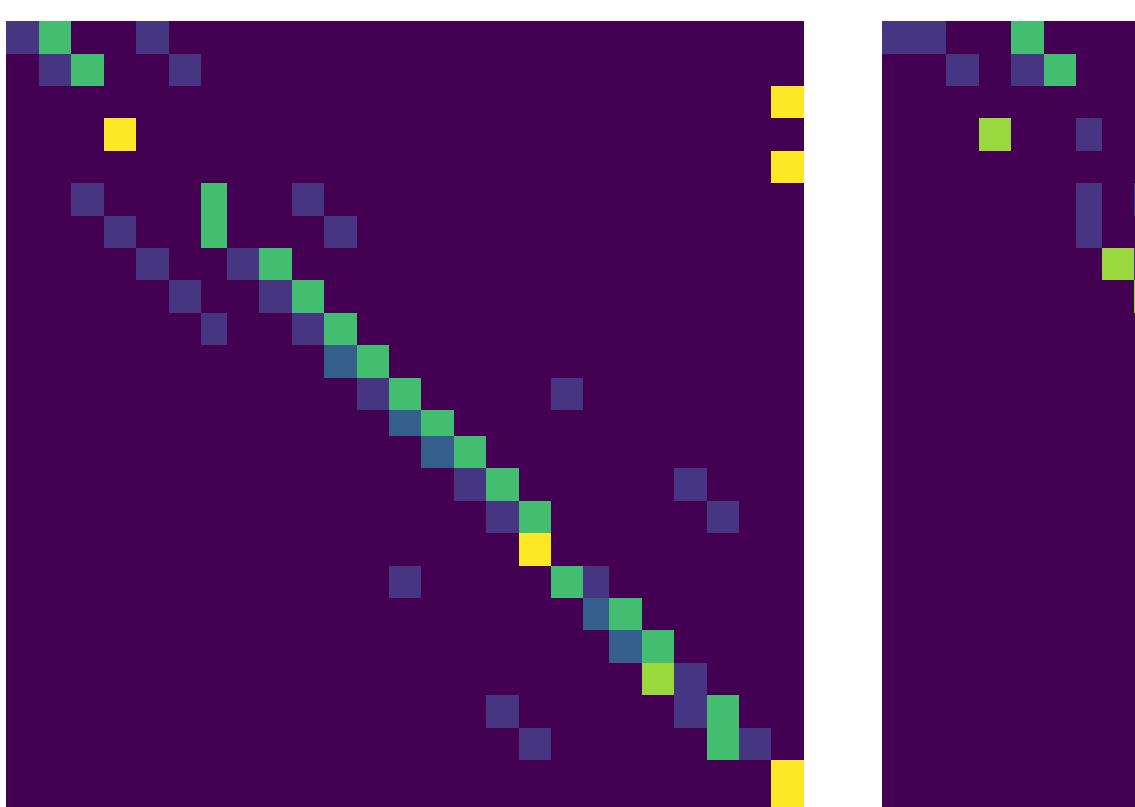


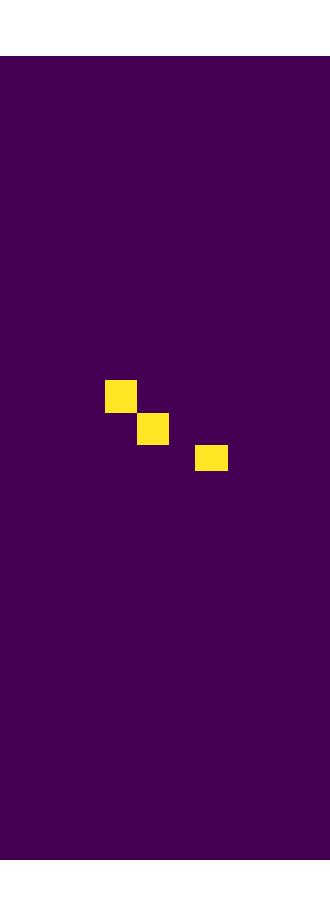
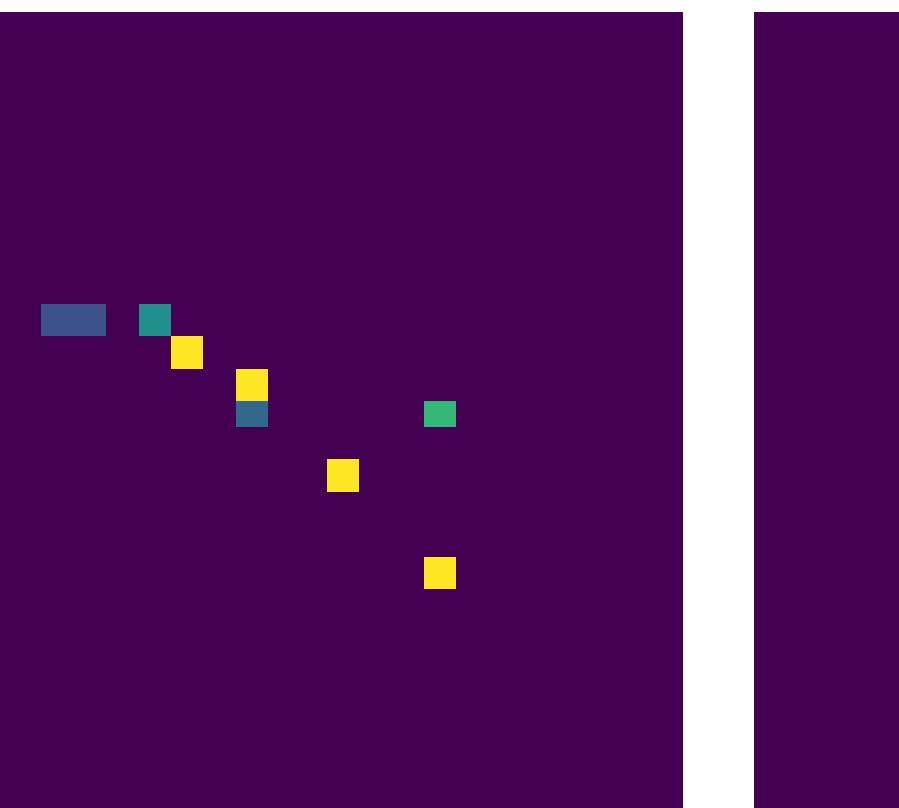
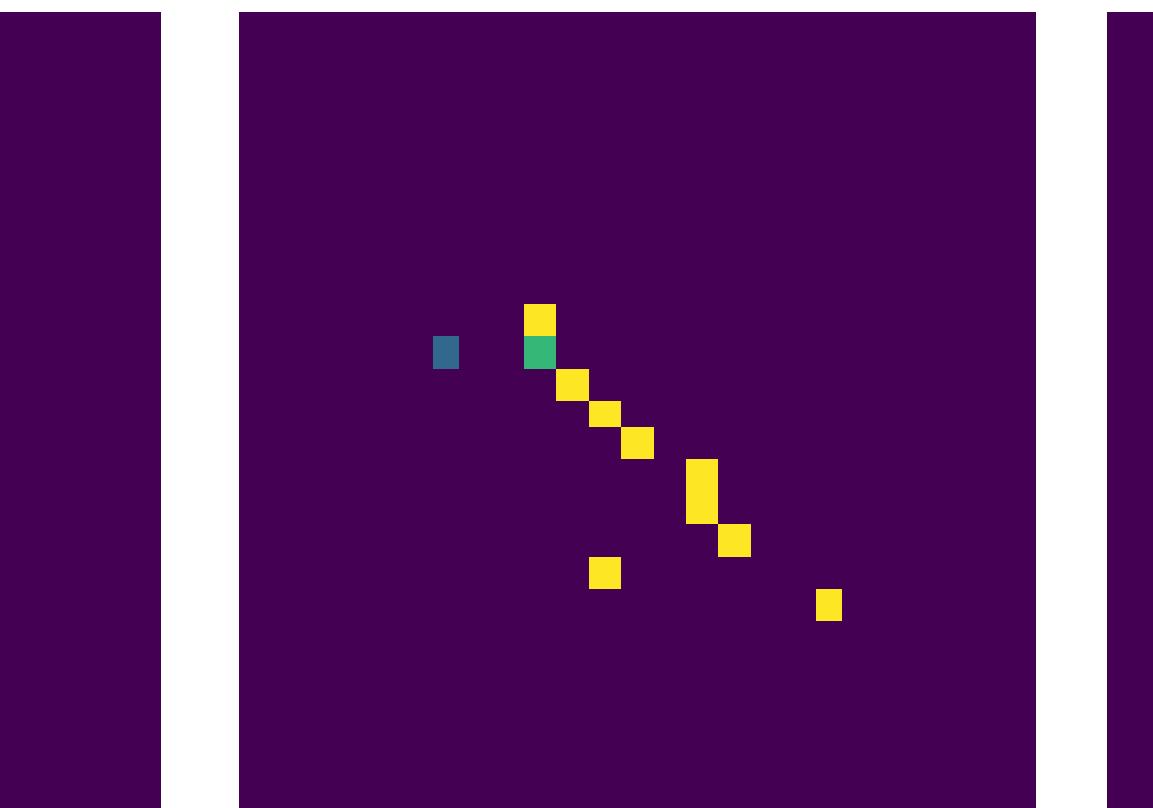
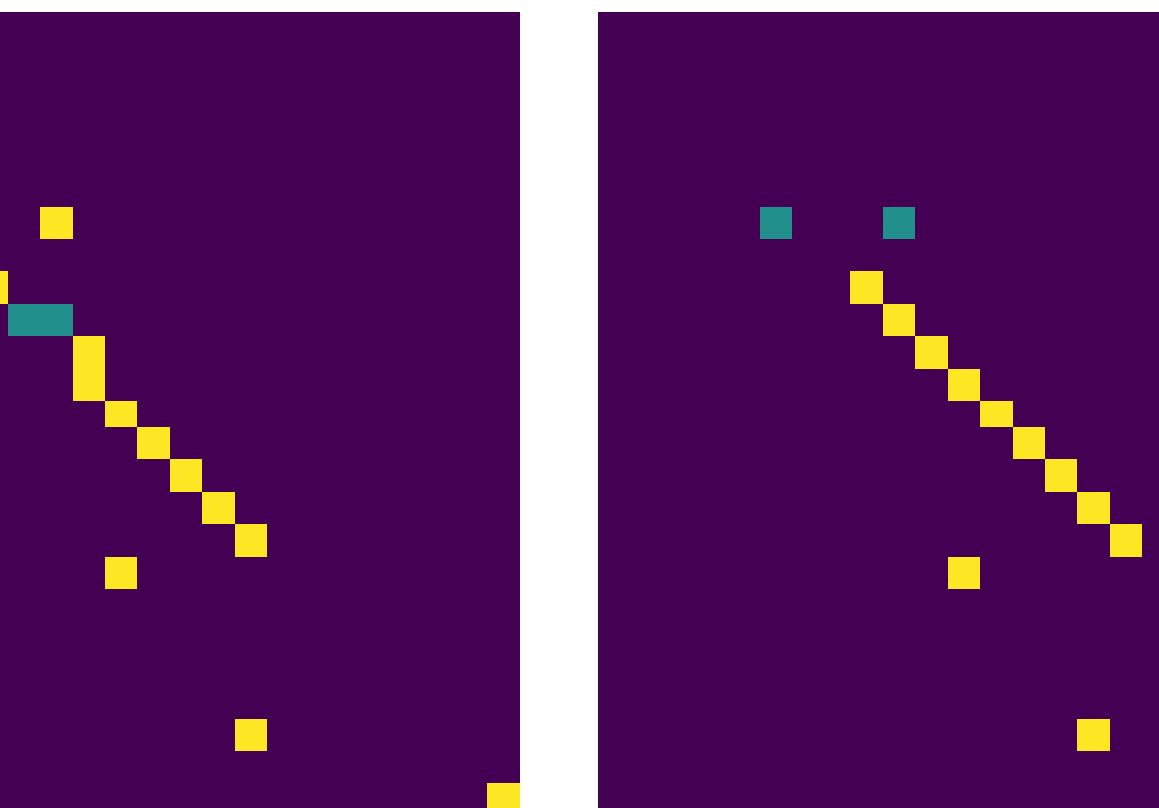
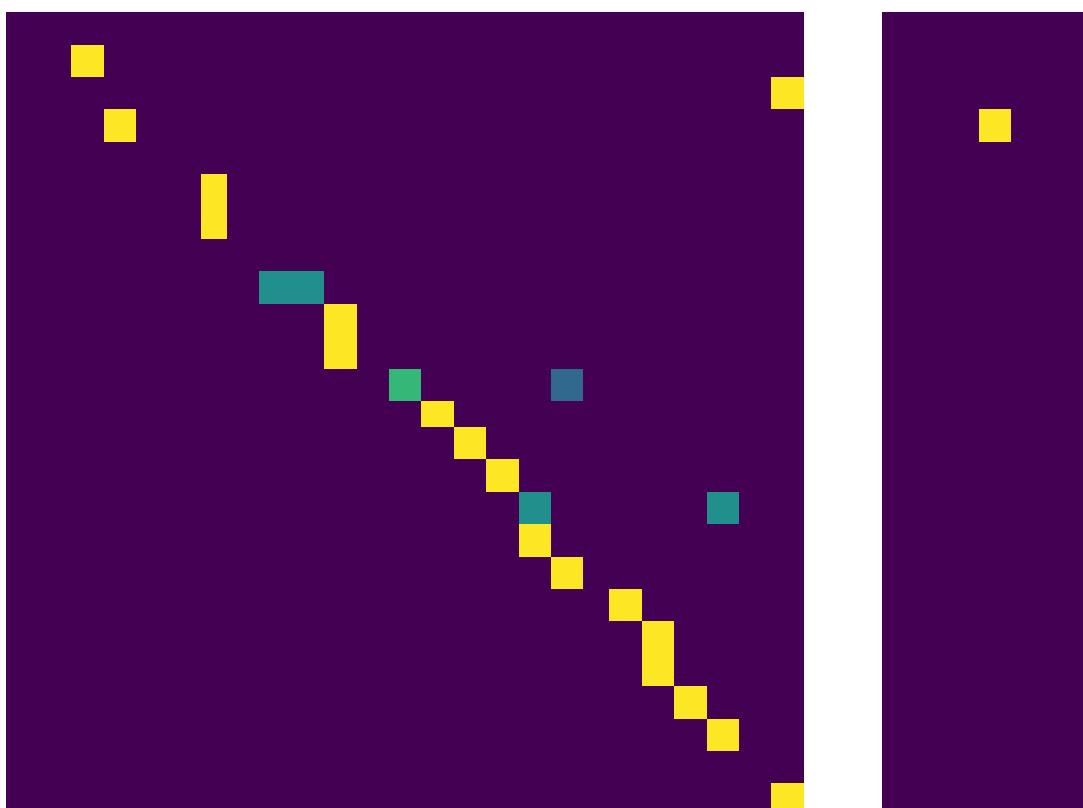
quantile exploration

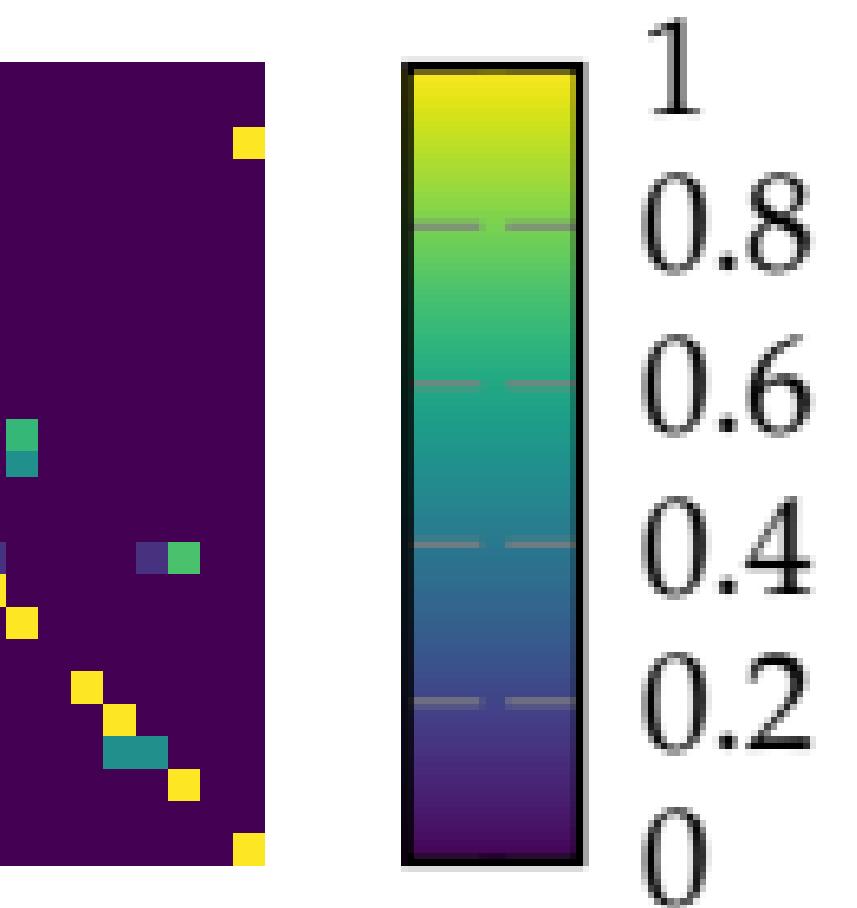
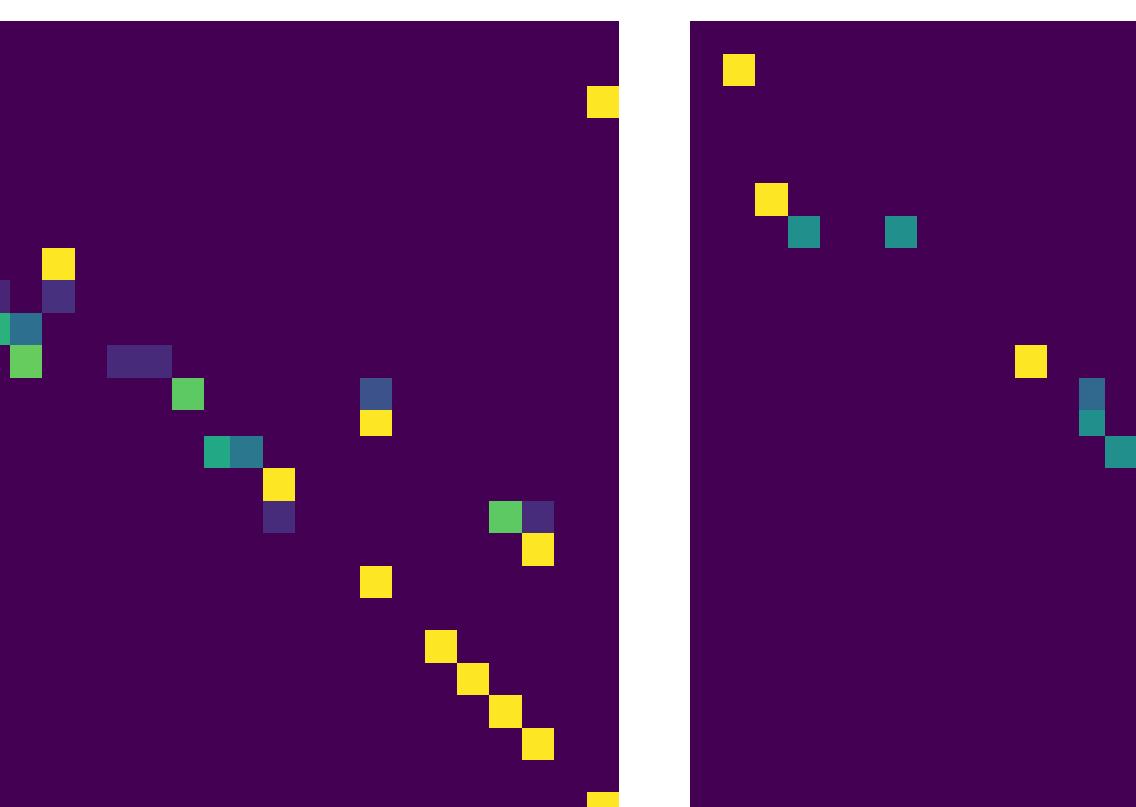
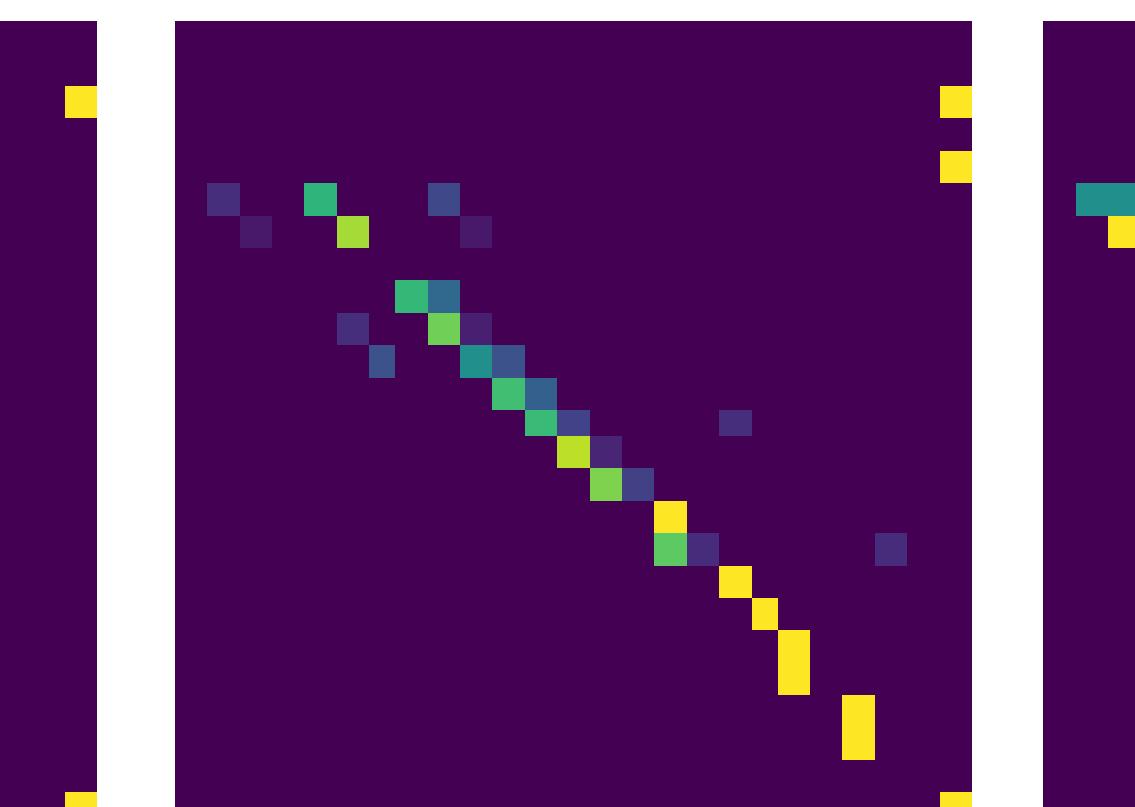
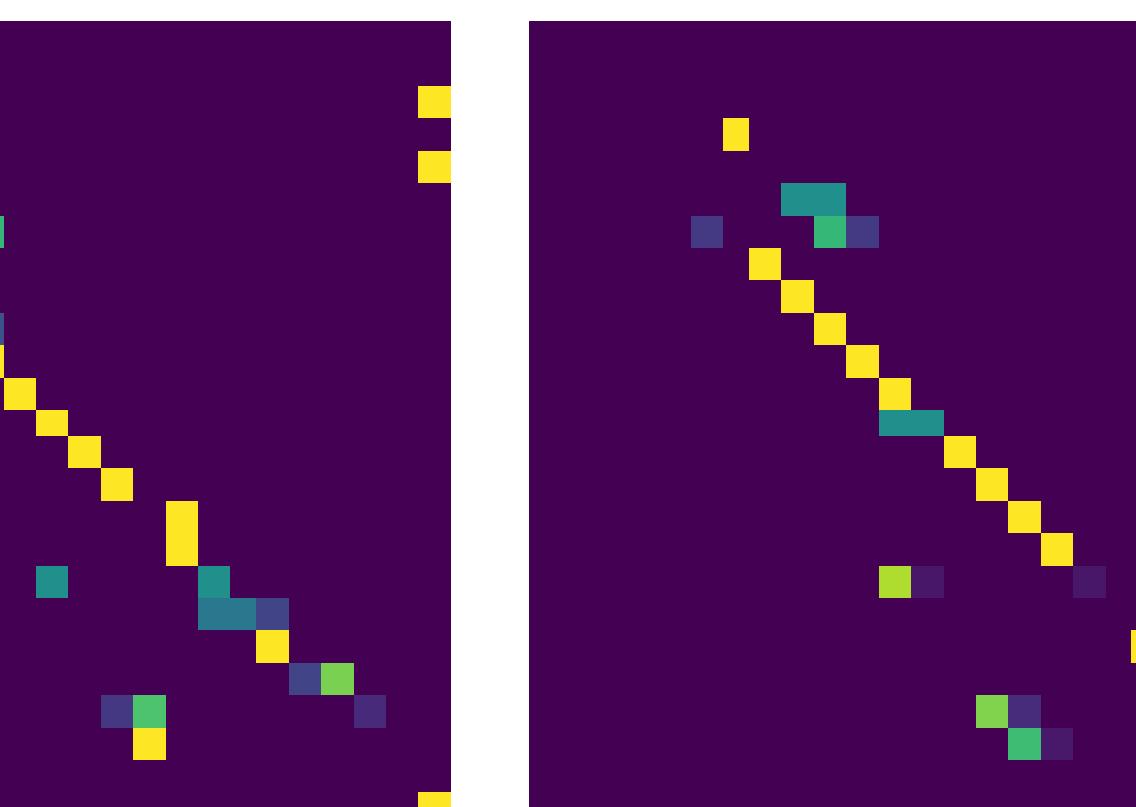
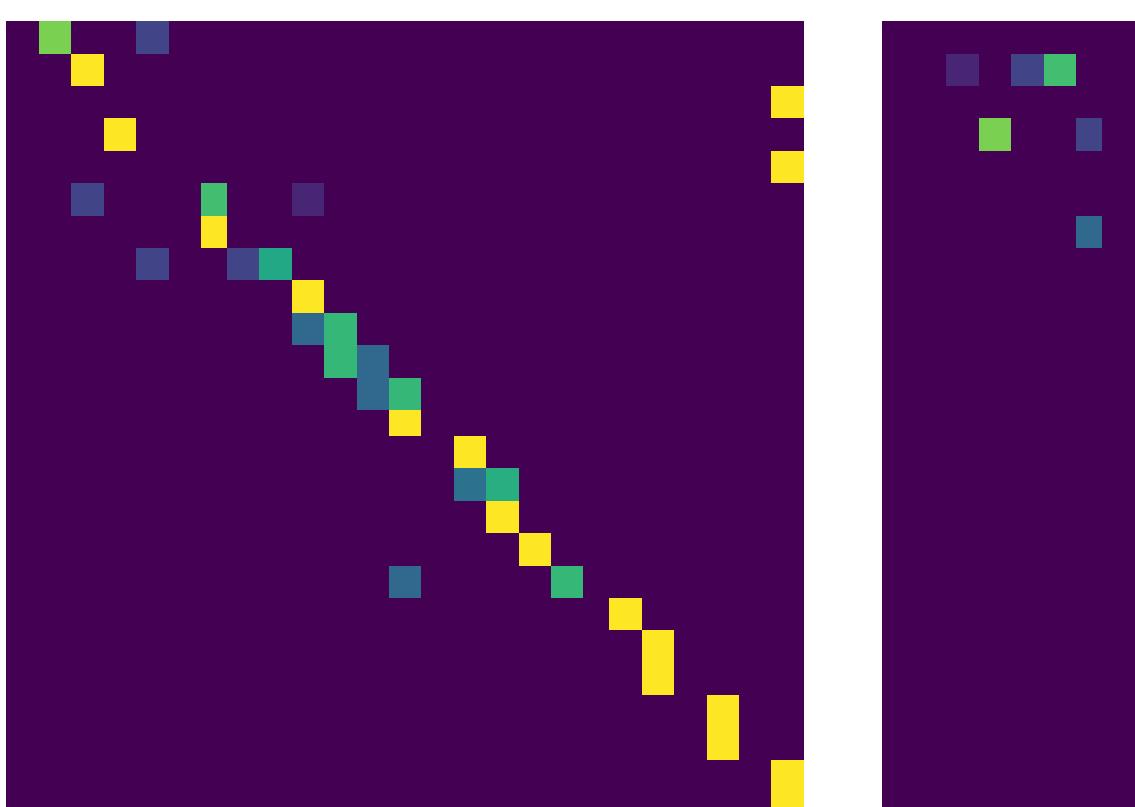


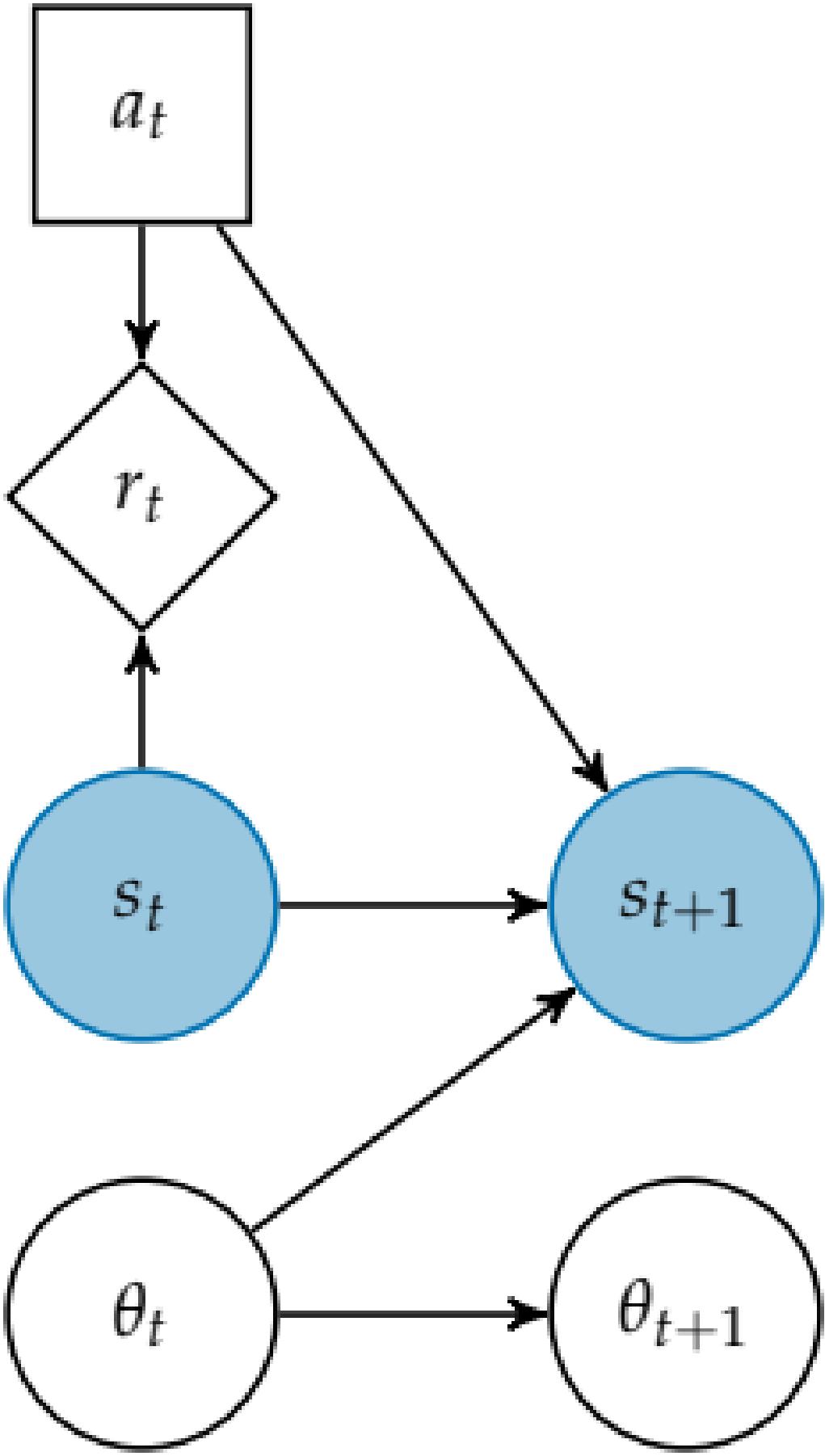
UCB1

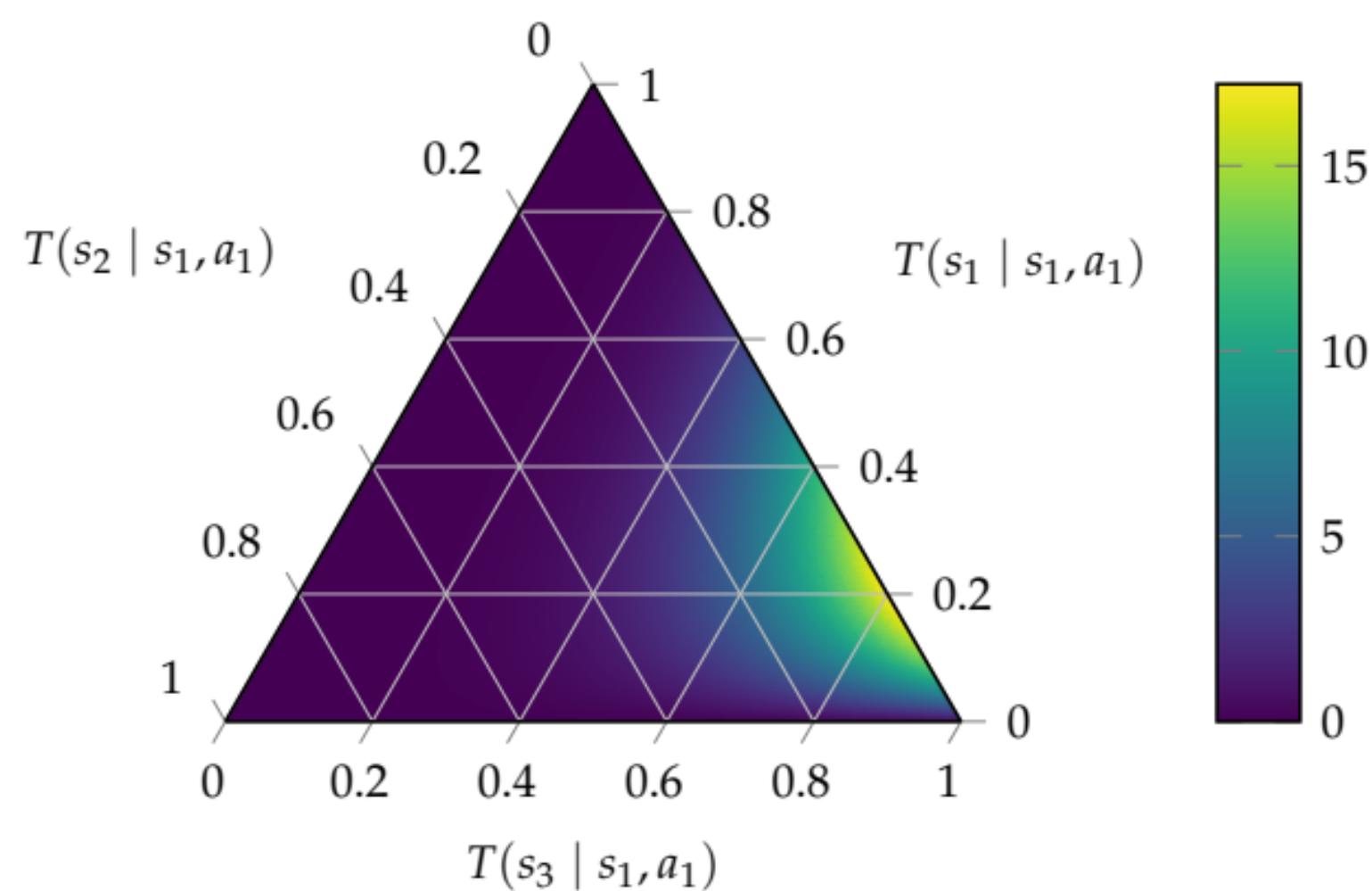


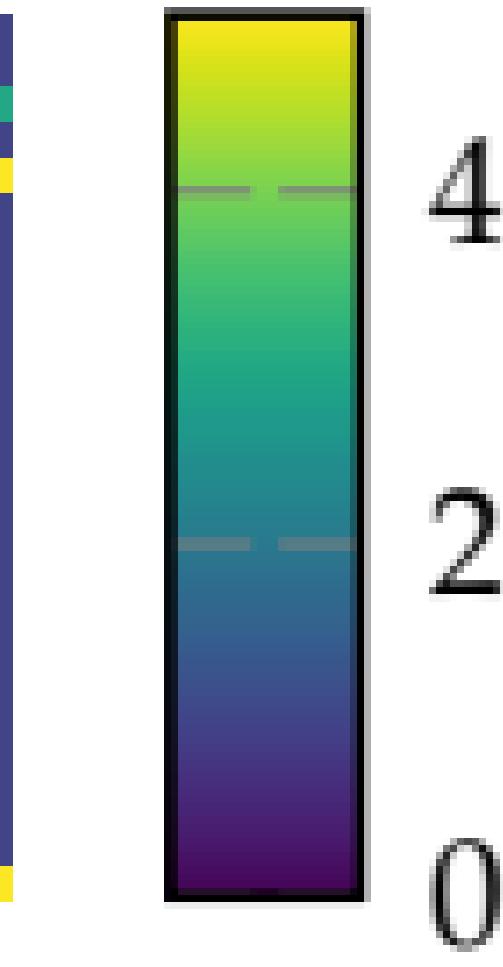
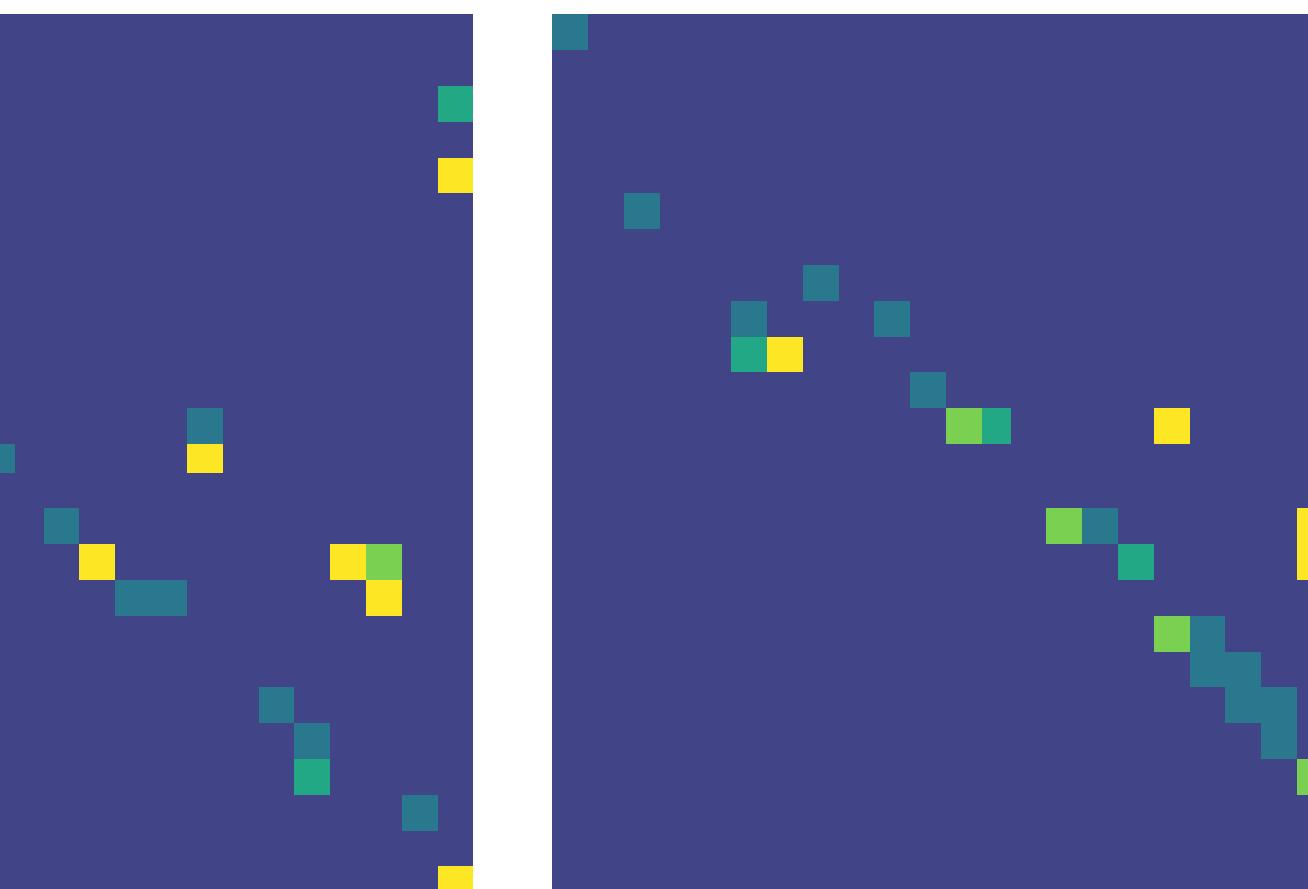
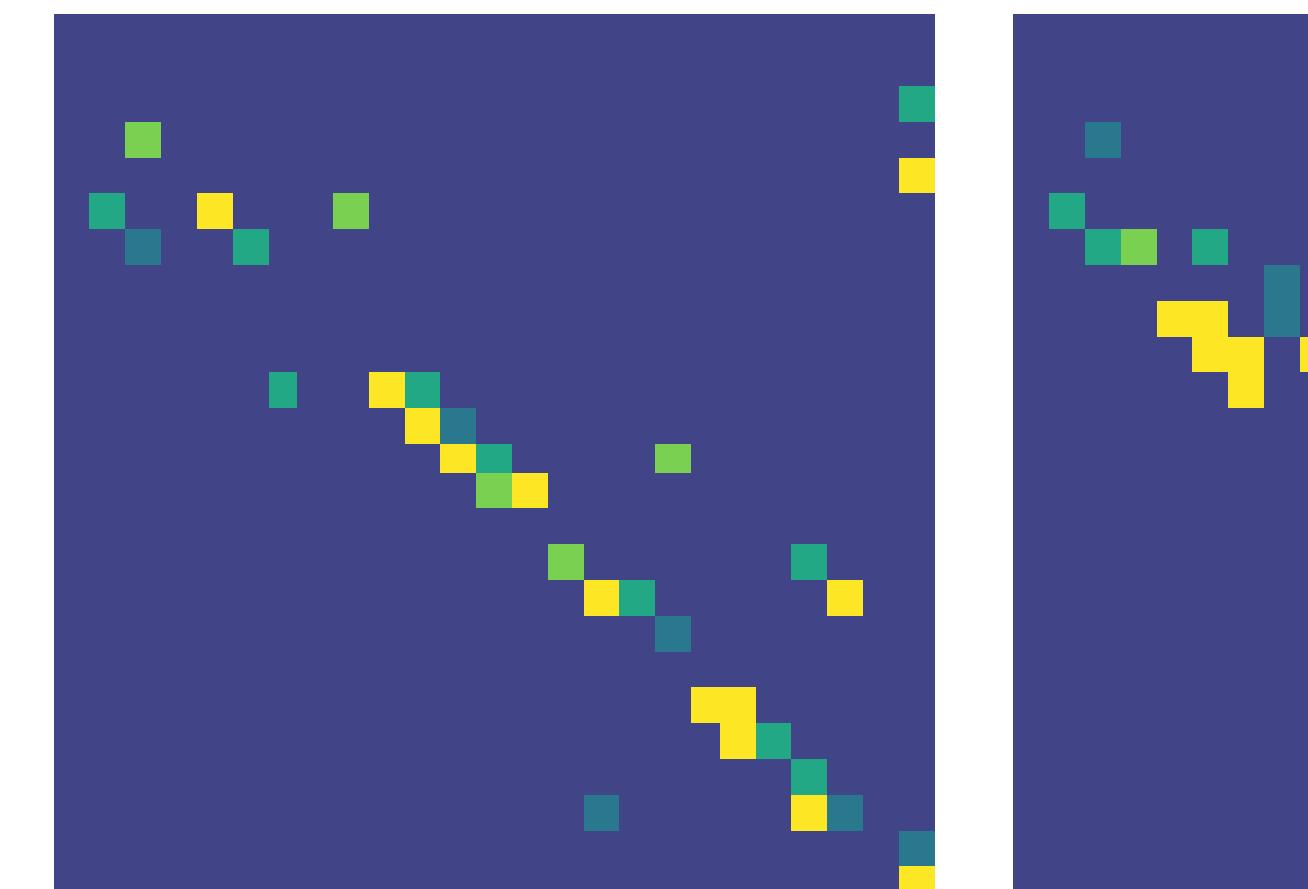
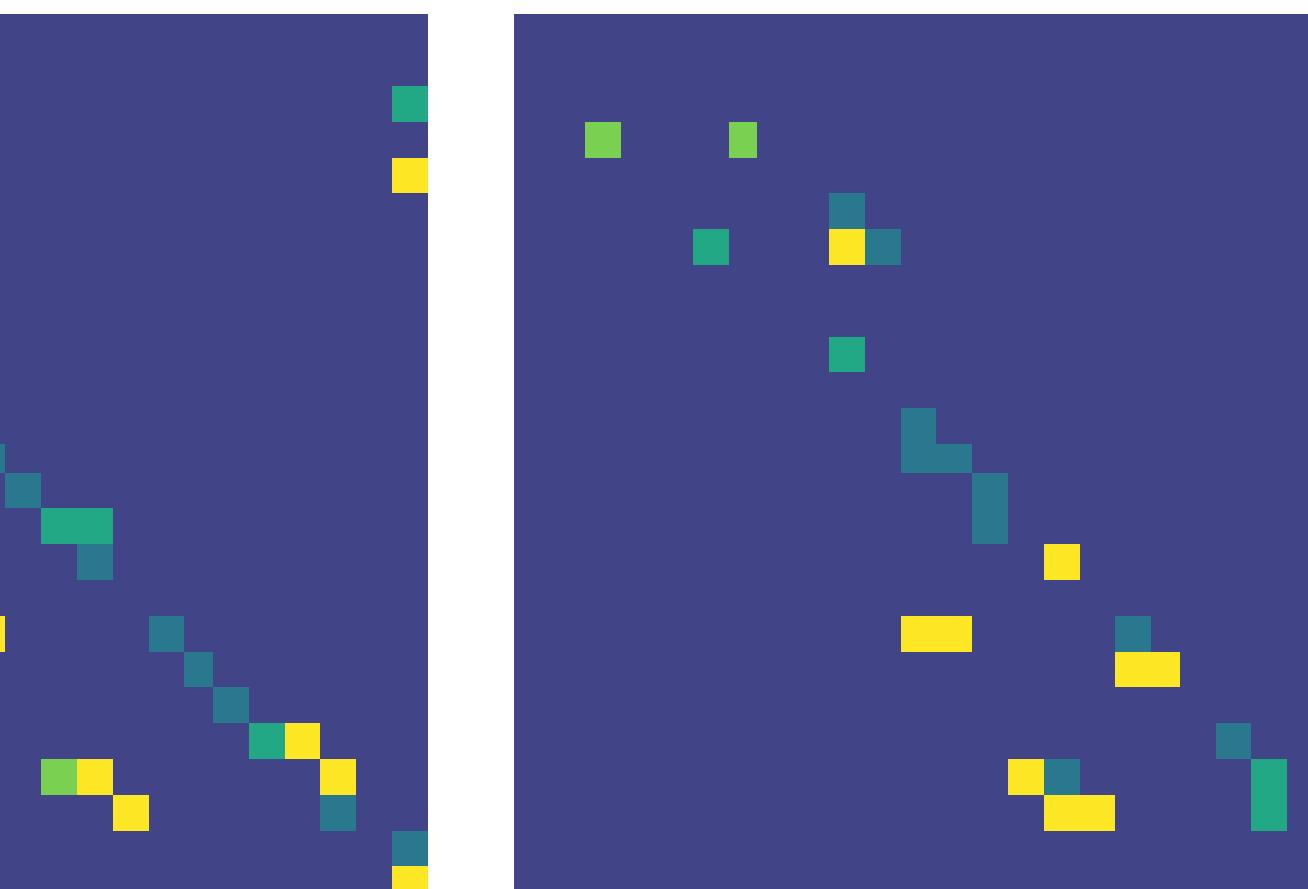
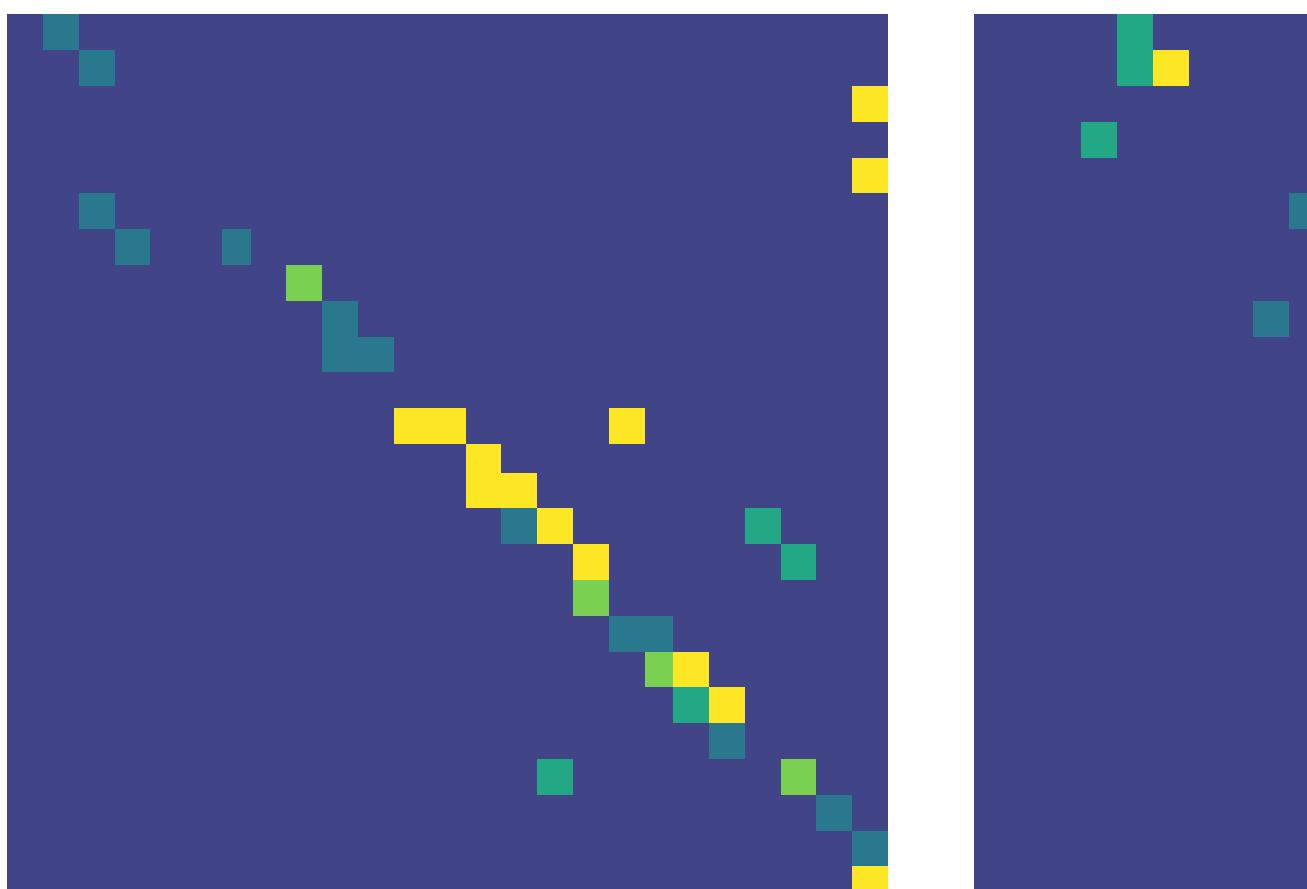


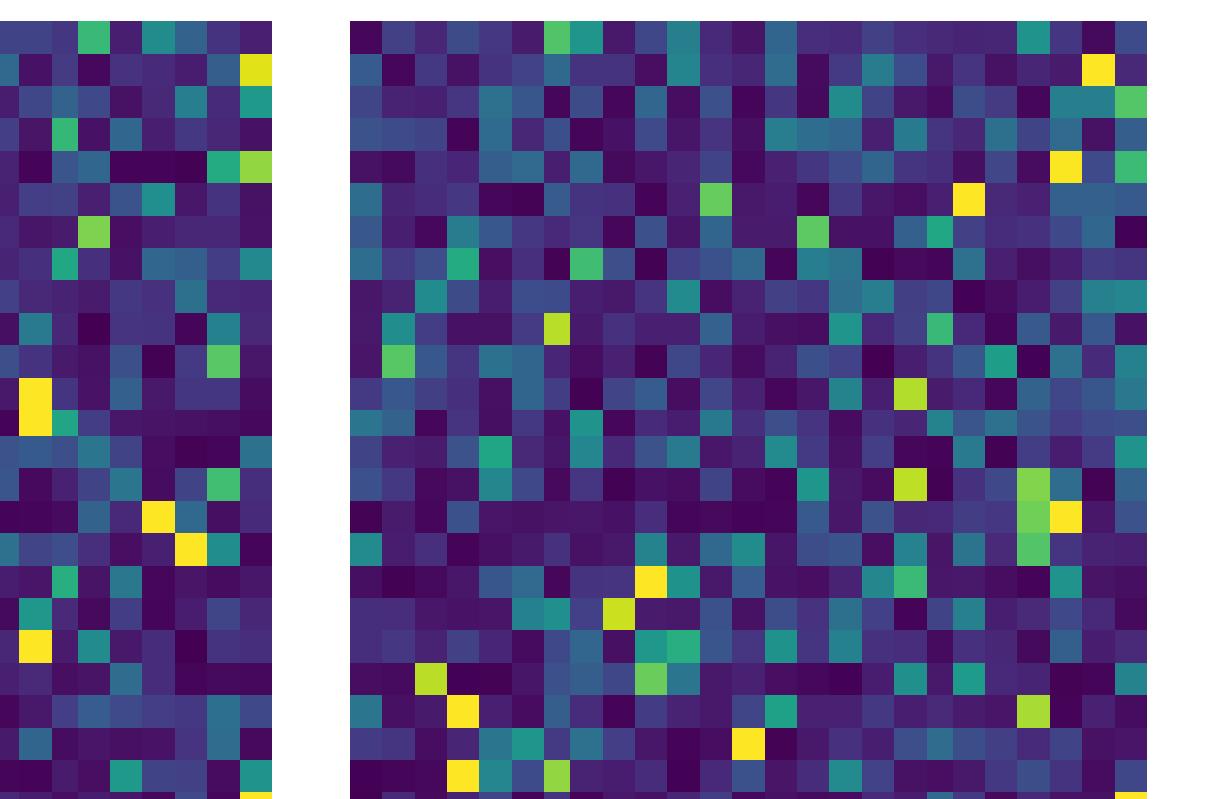
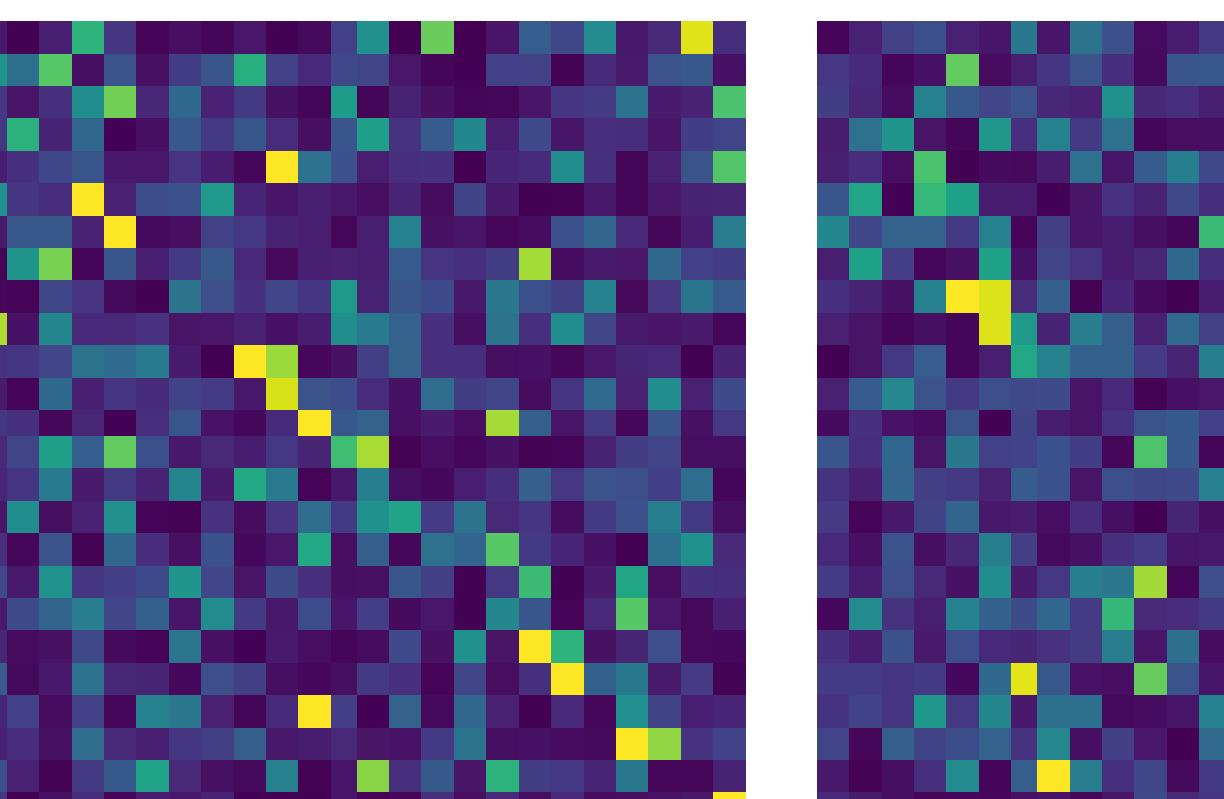
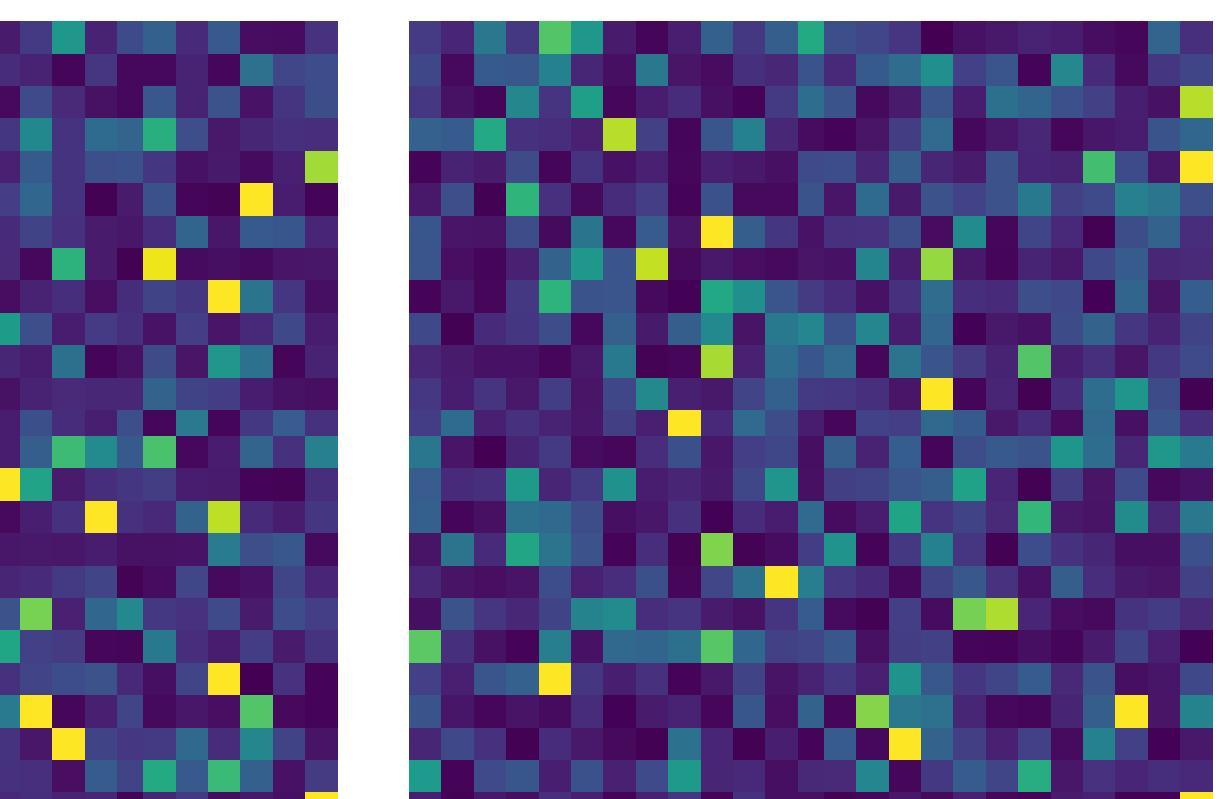
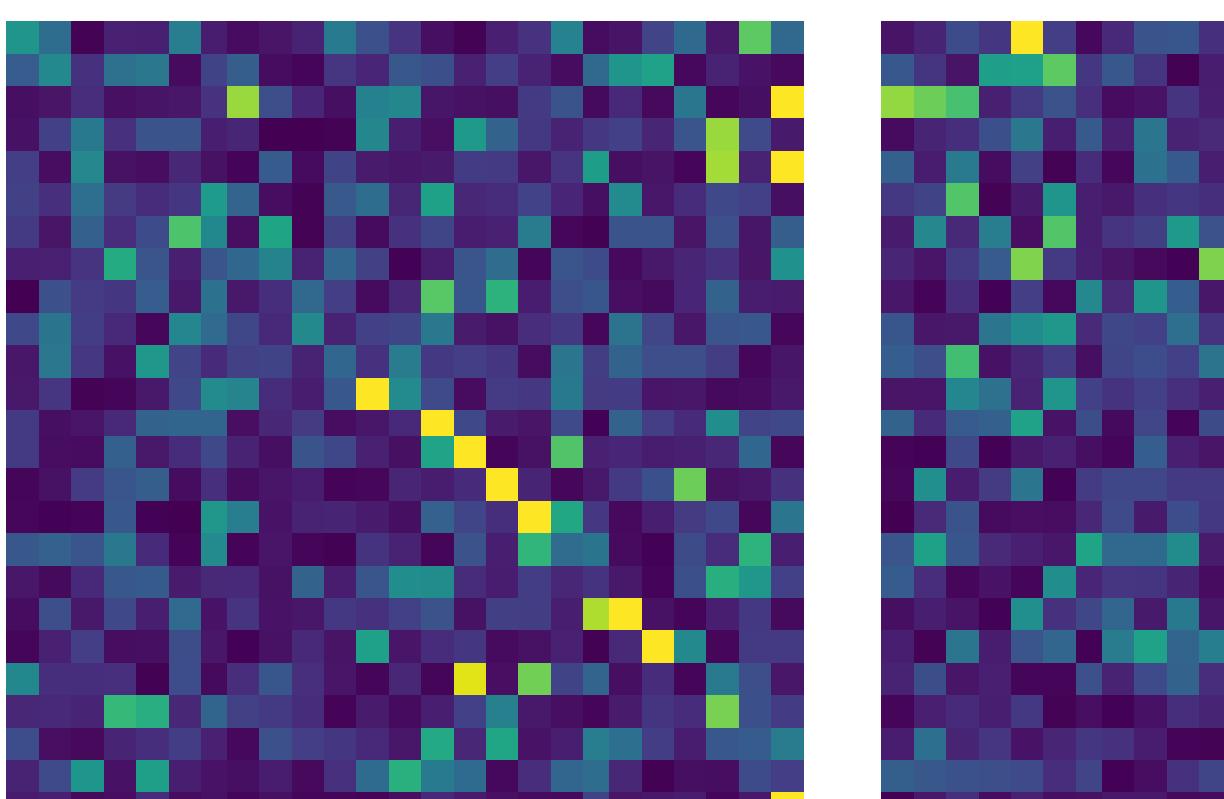


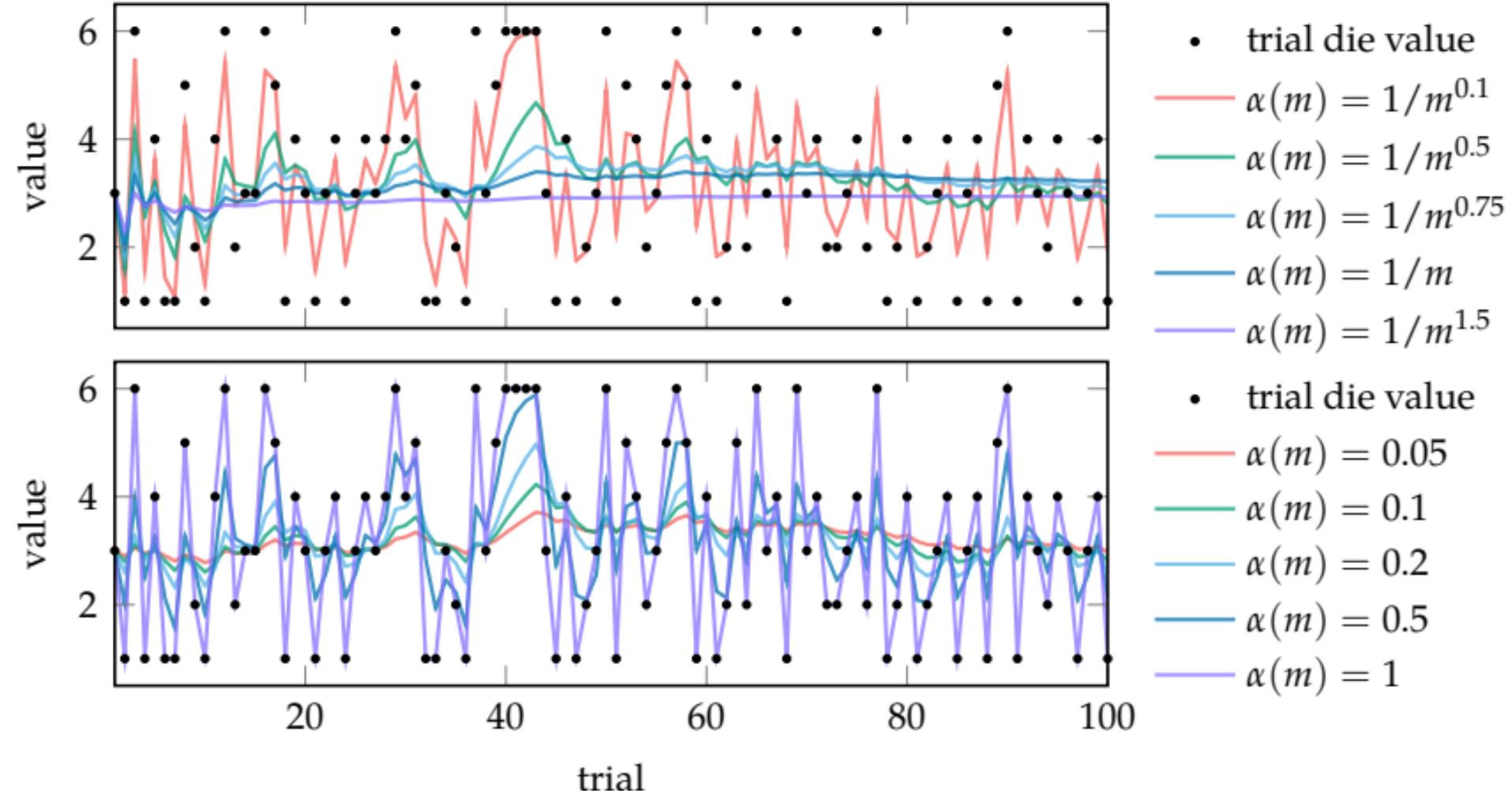


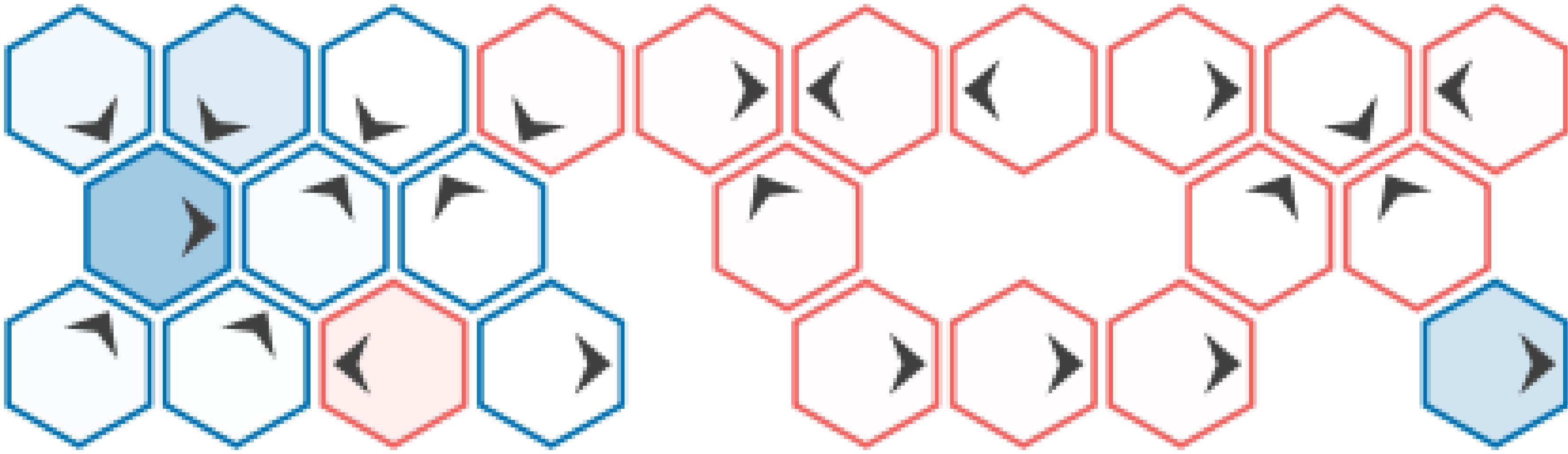


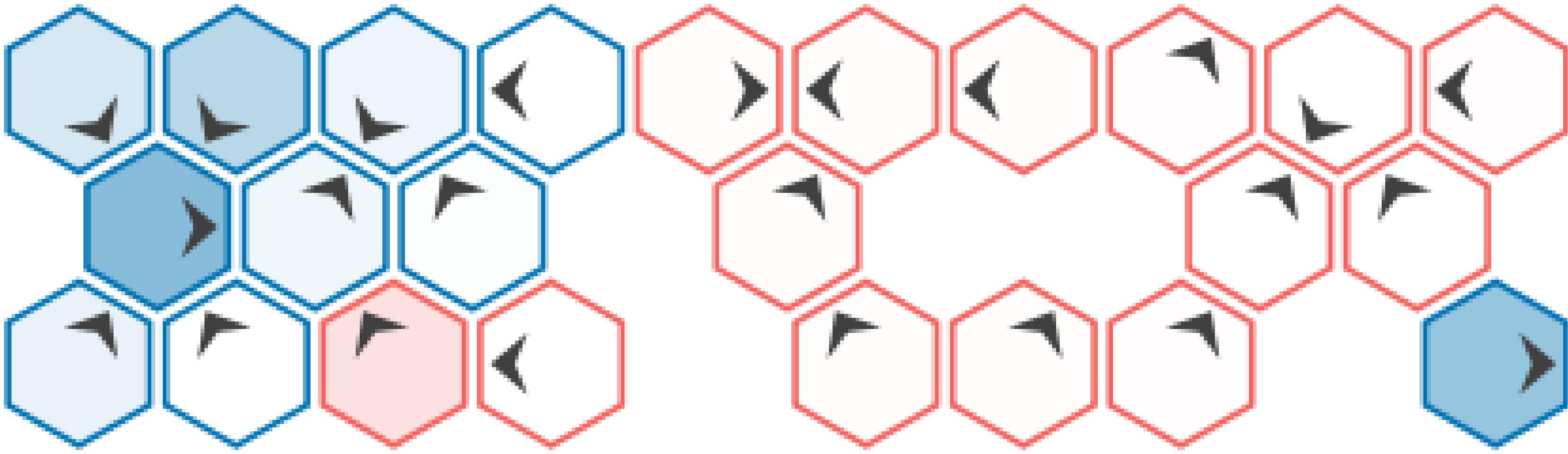


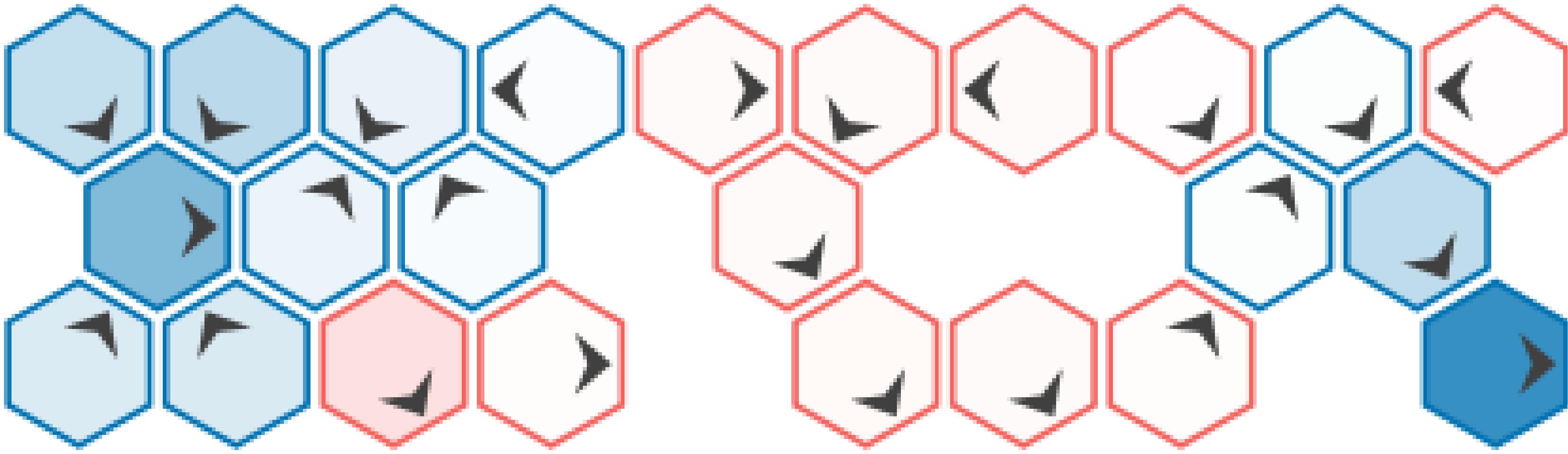


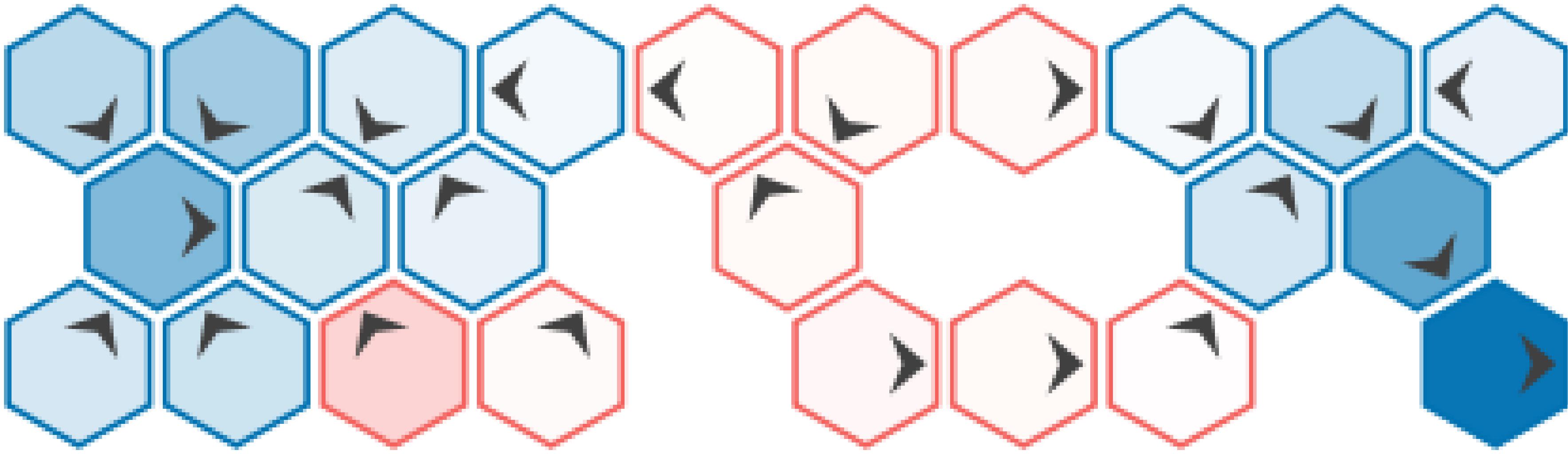


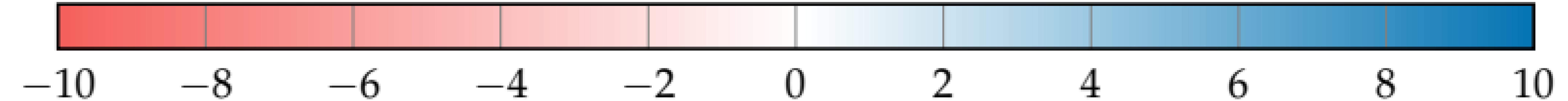


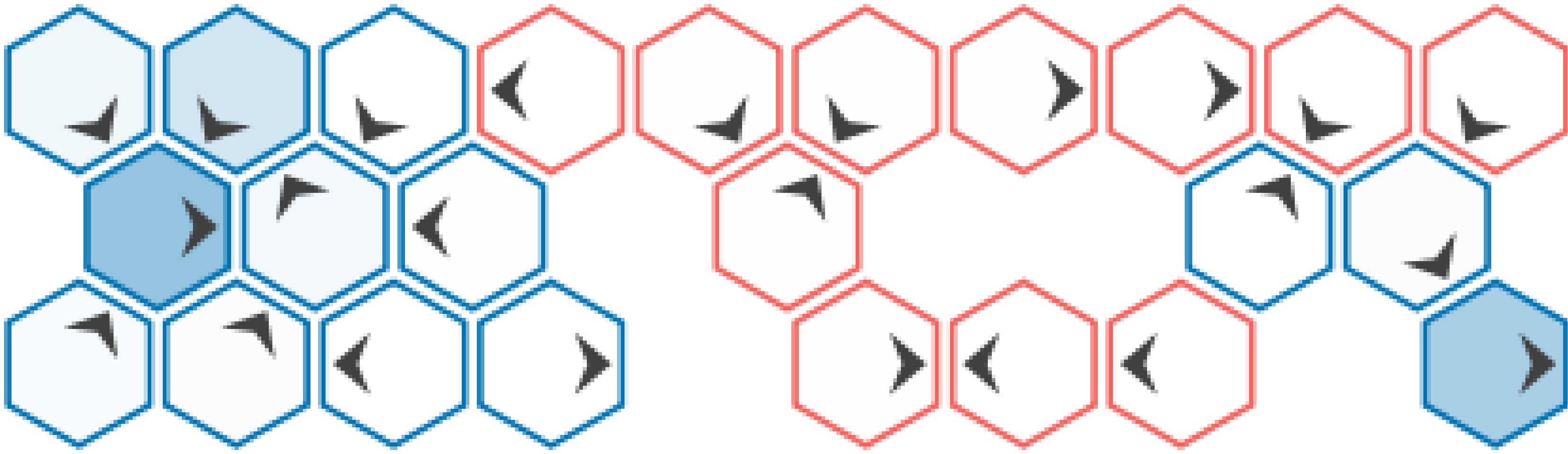


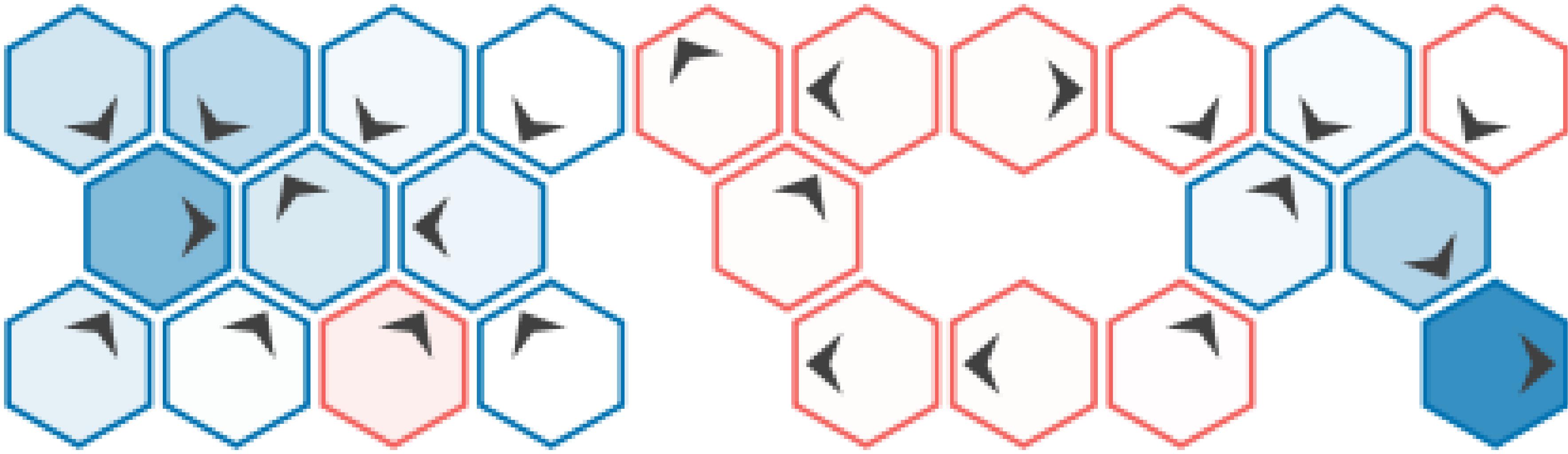


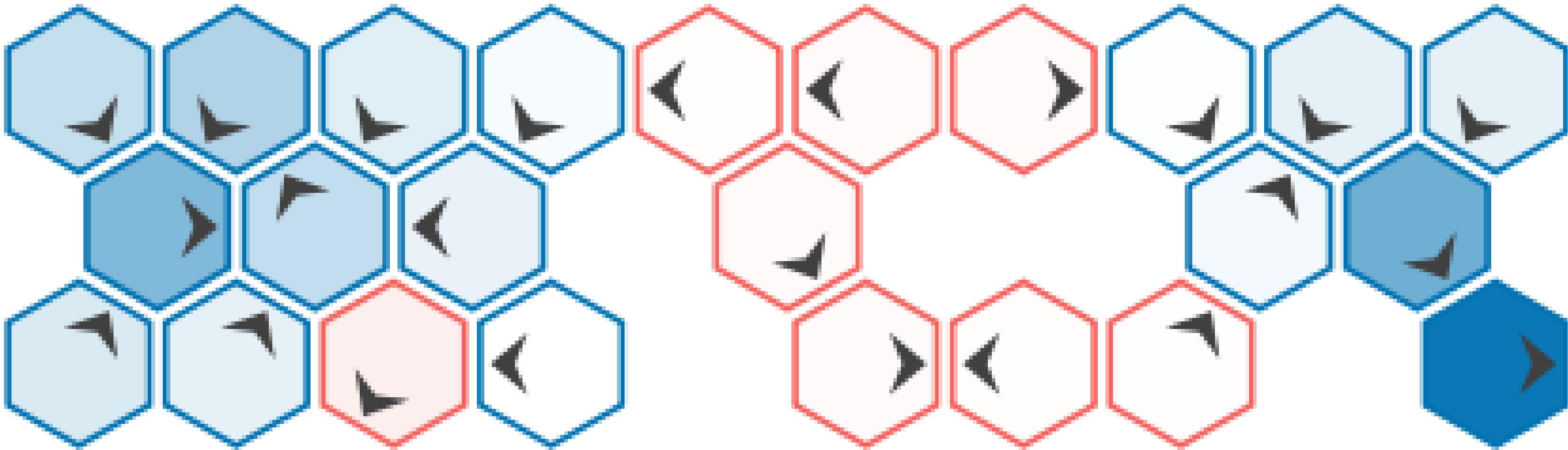


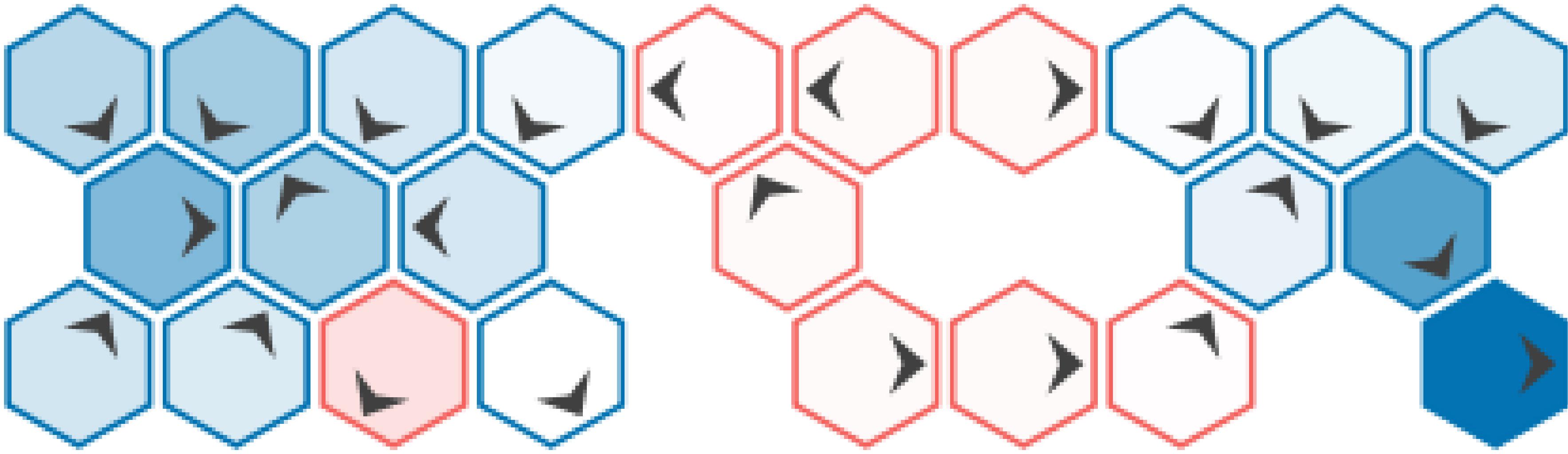


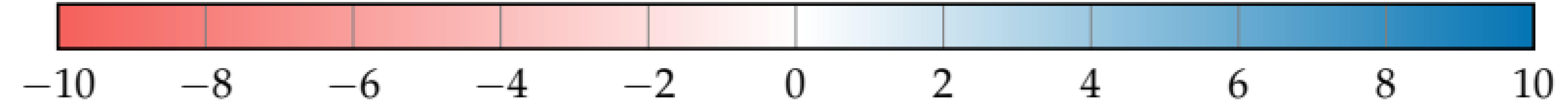


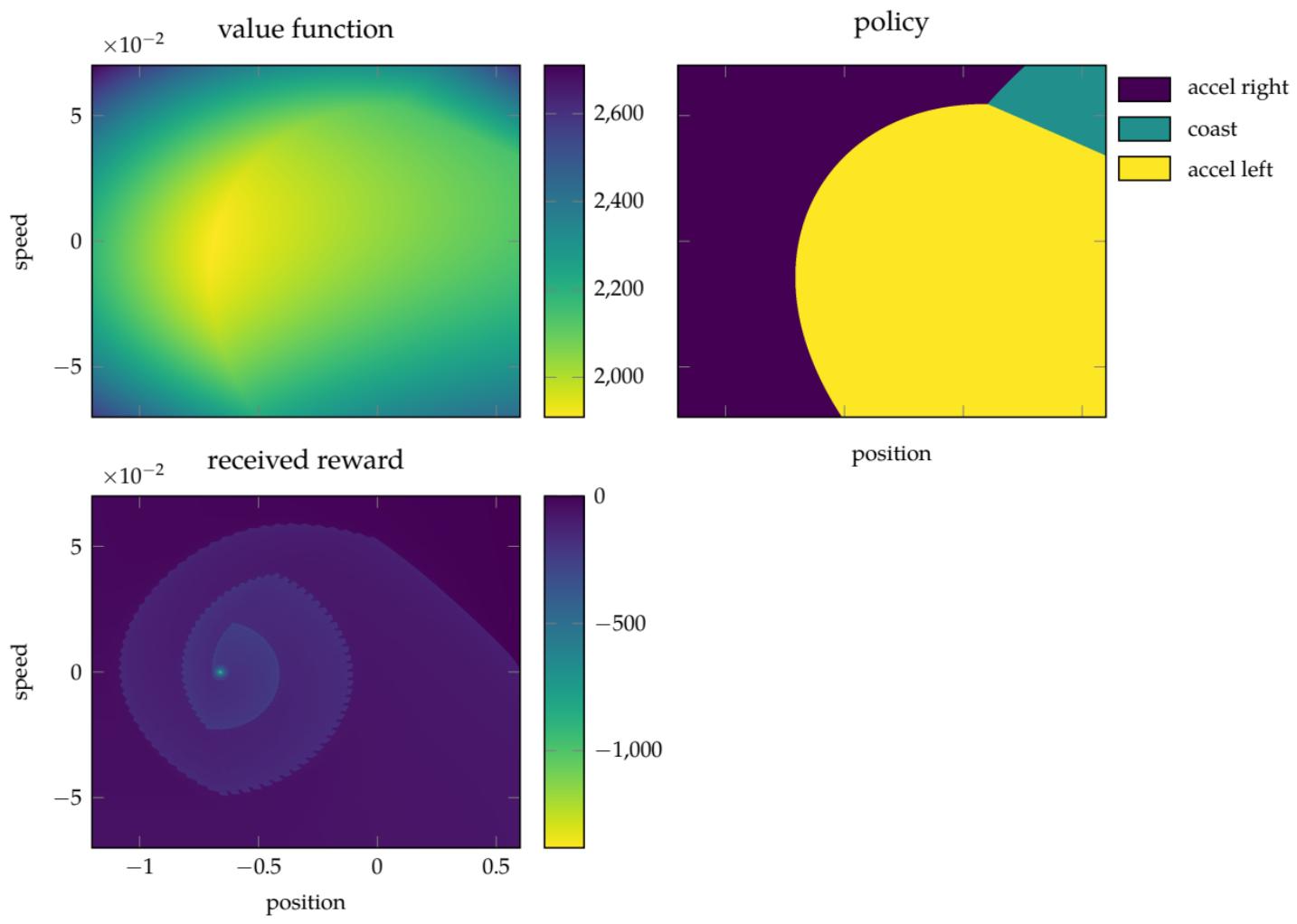


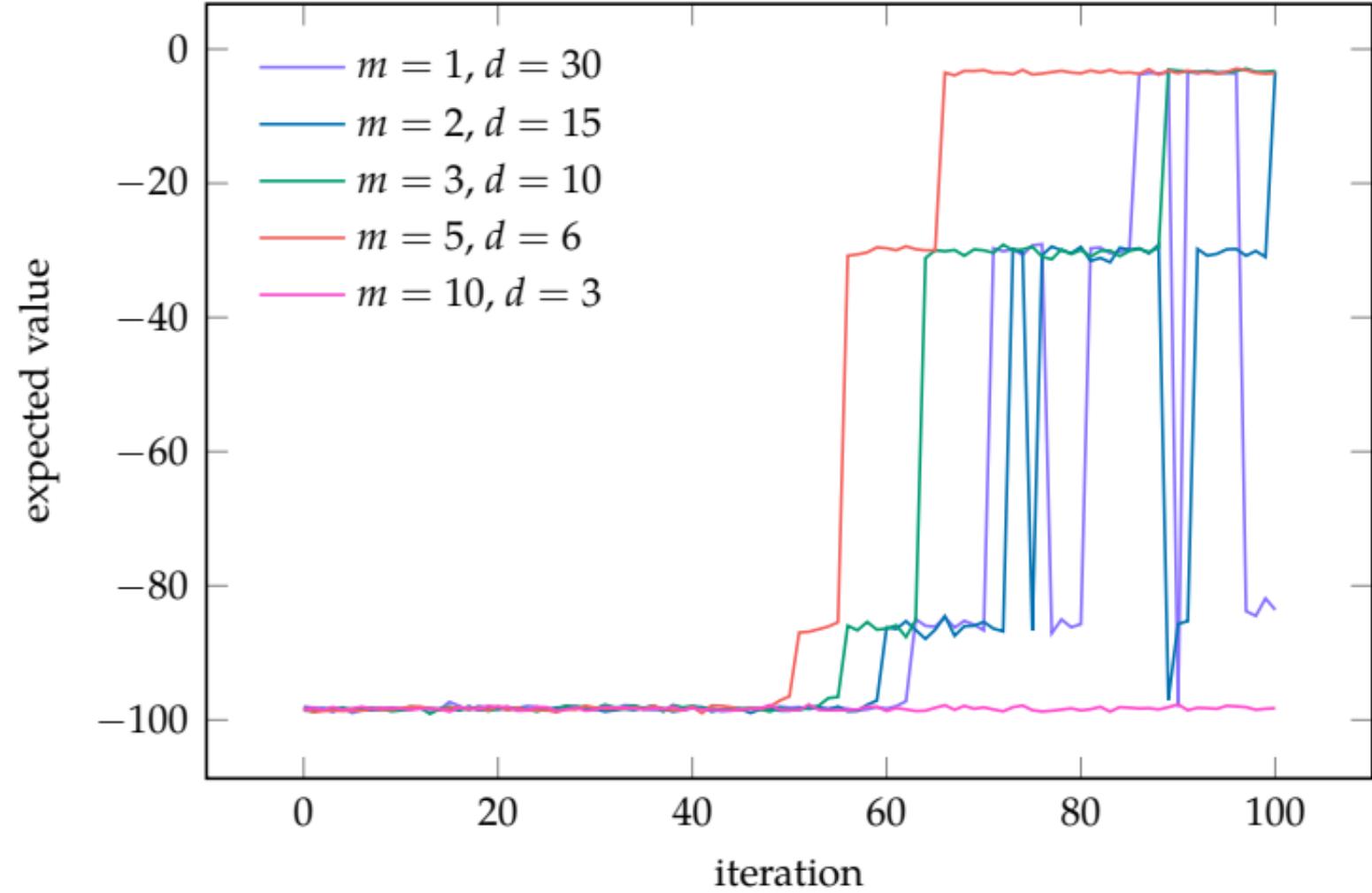






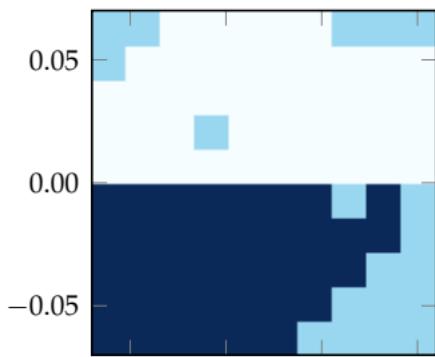






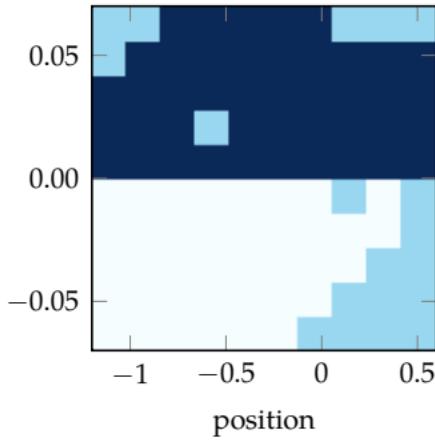
$P(a = -1 \mid s)$ $P(a = 0 \mid s)$

speed

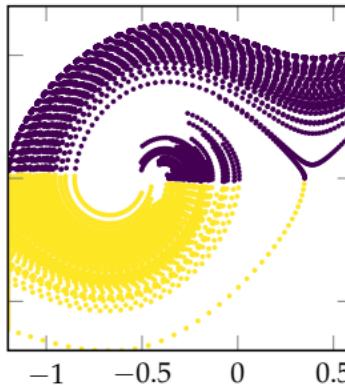
 $P(a = 1 \mid s)$

expert demonstrations

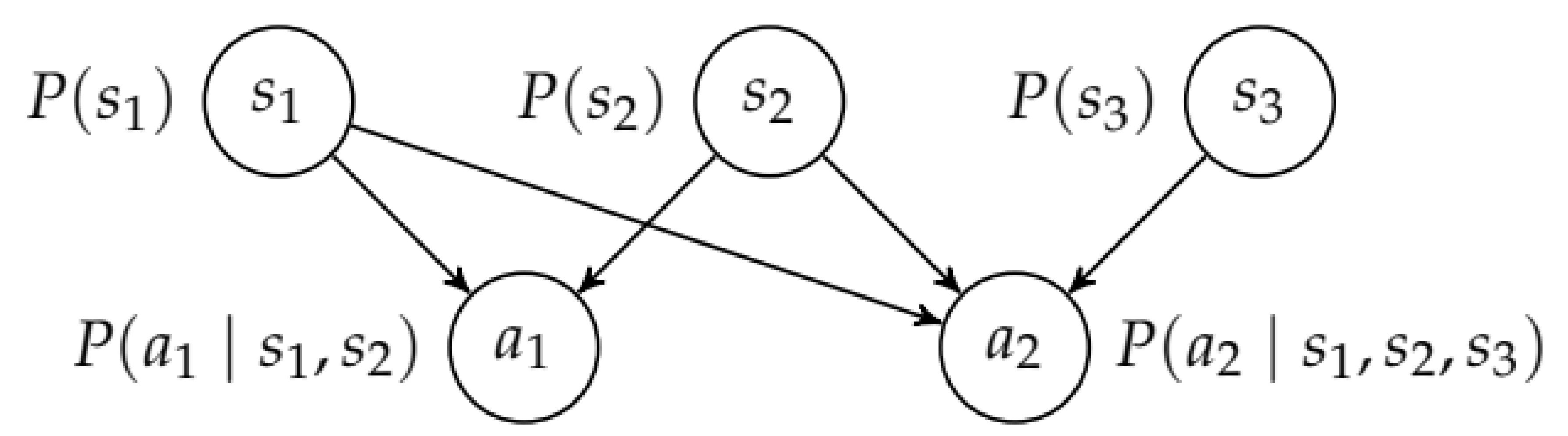
speed

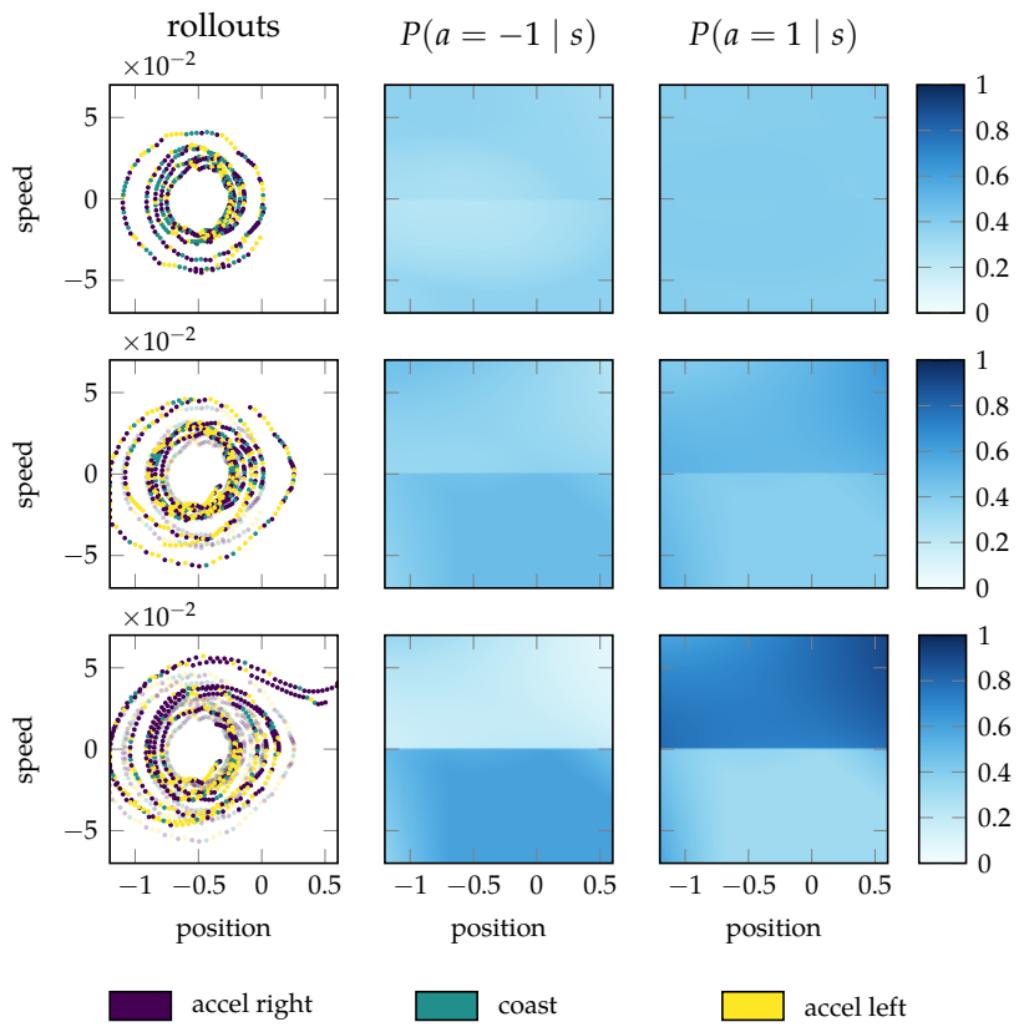


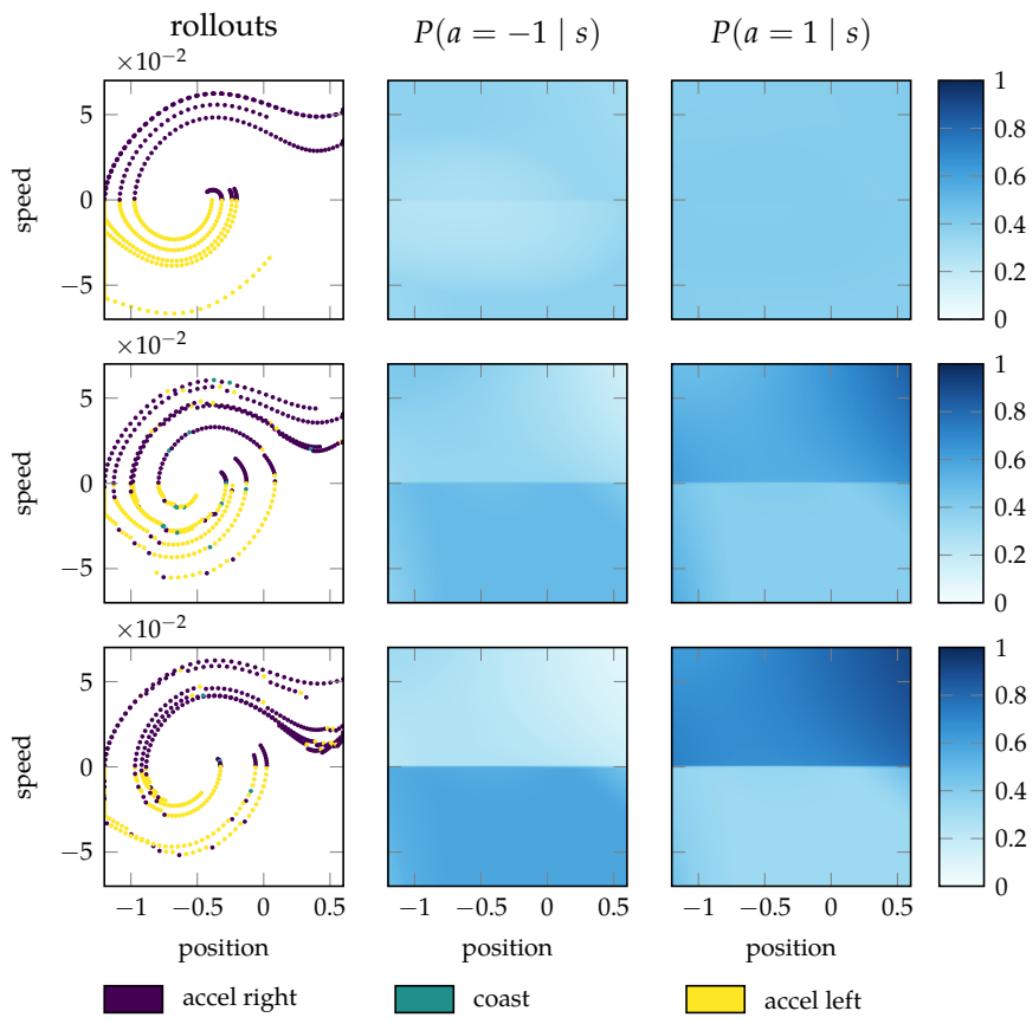
position

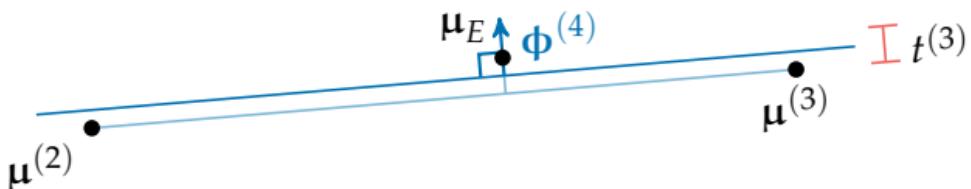
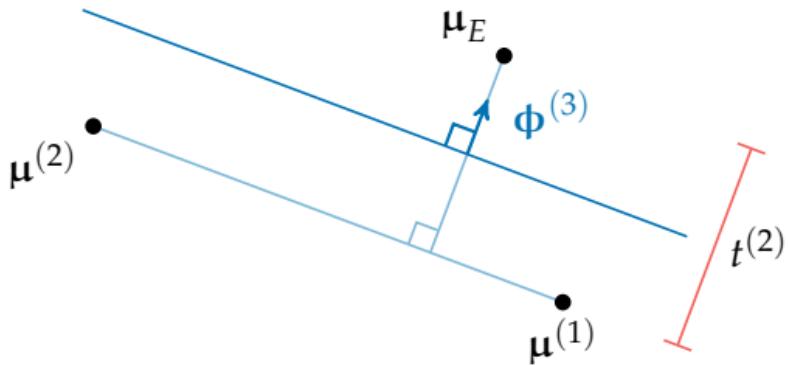
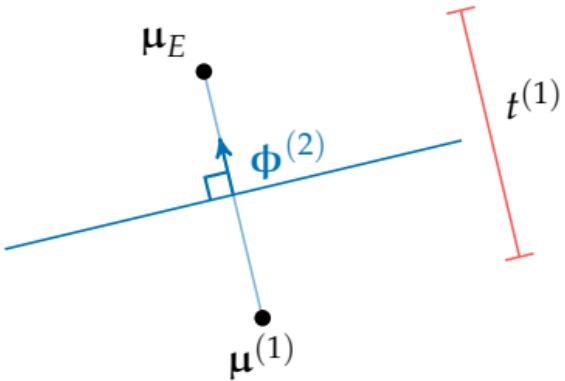


- accel right
- coast
- accel left

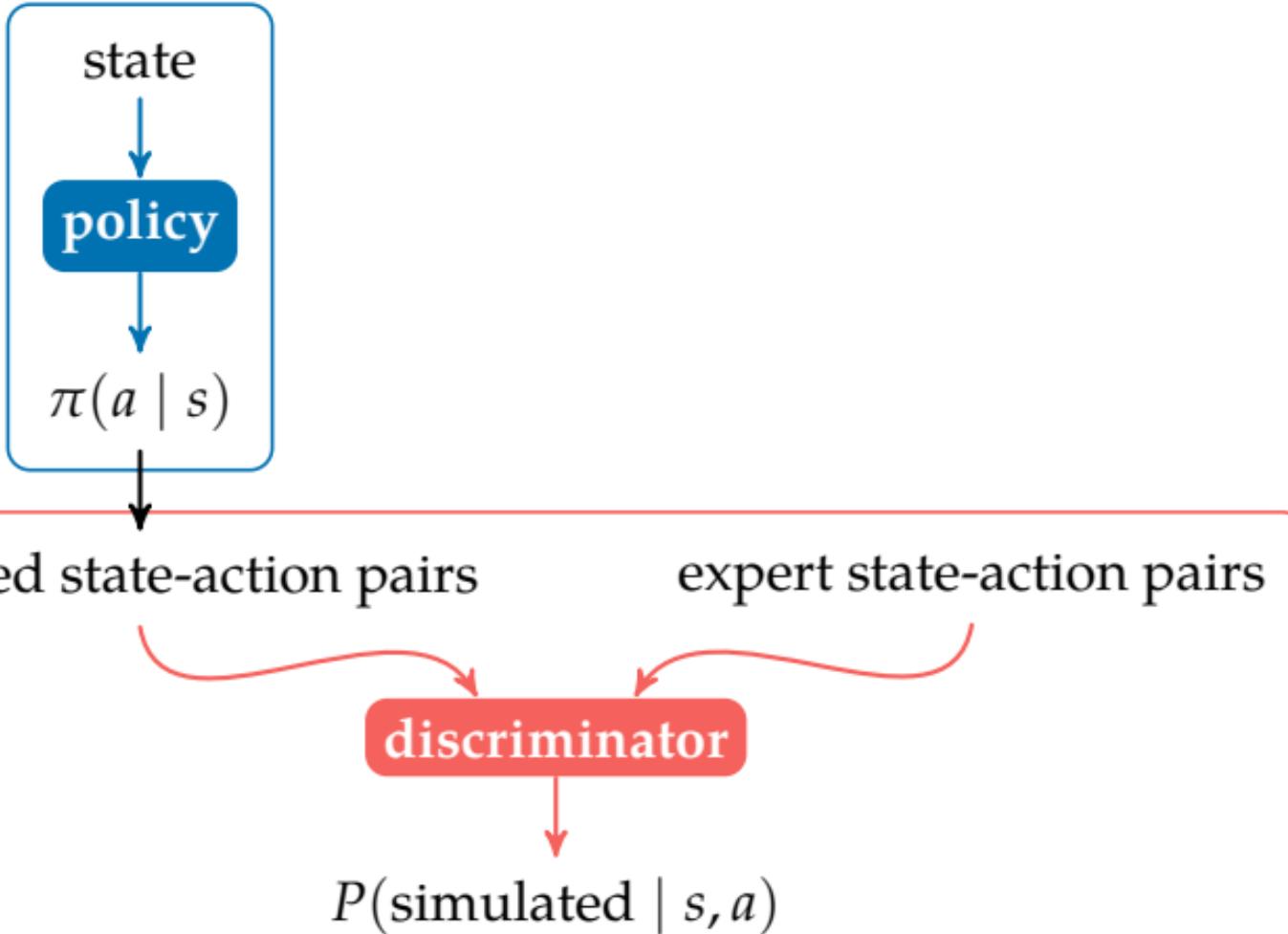






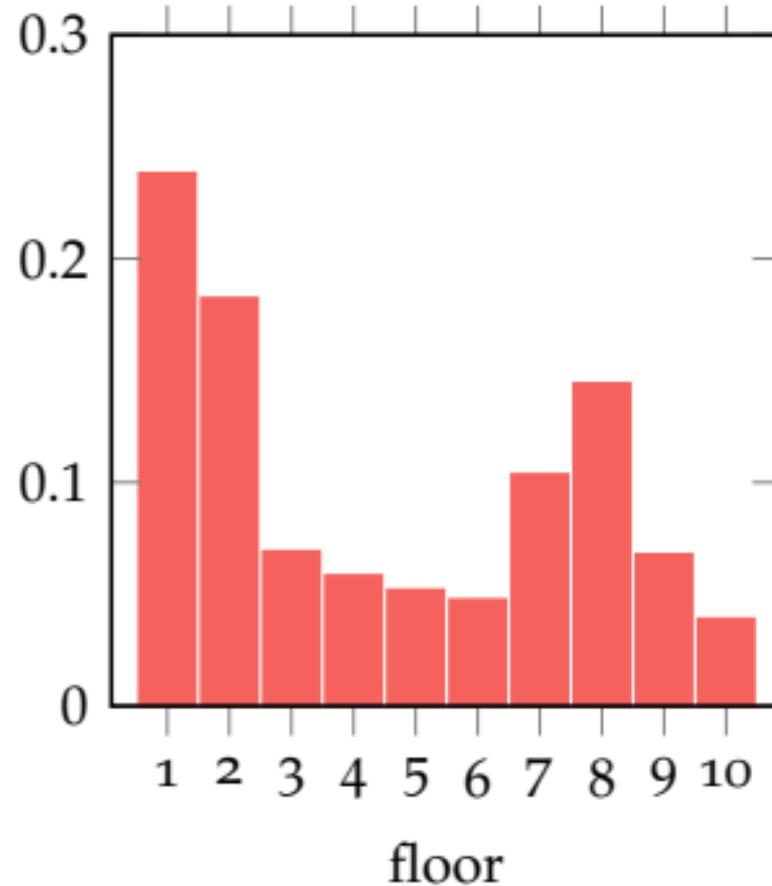


$\mu^{(1)}$

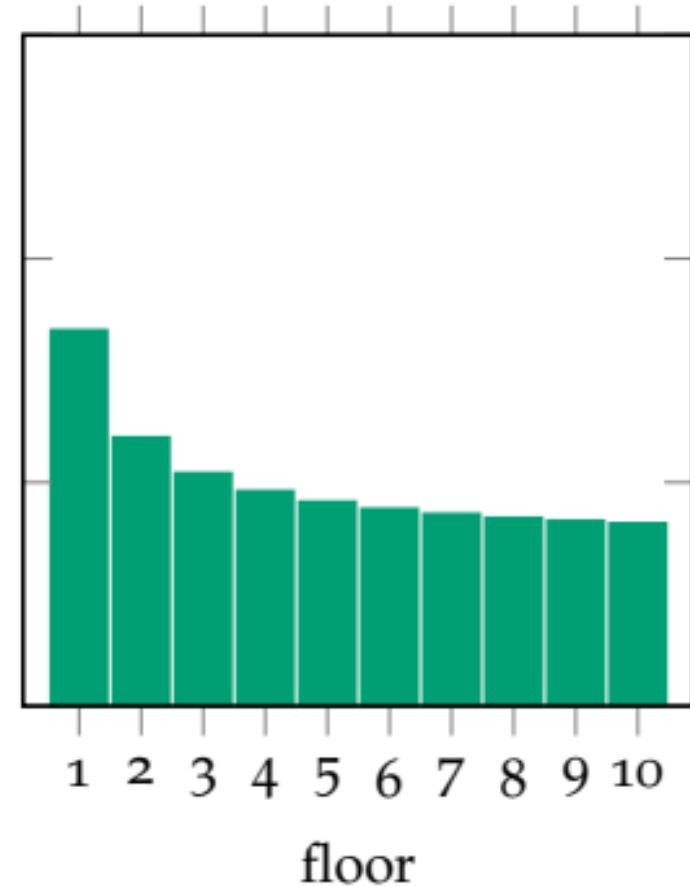


policy A

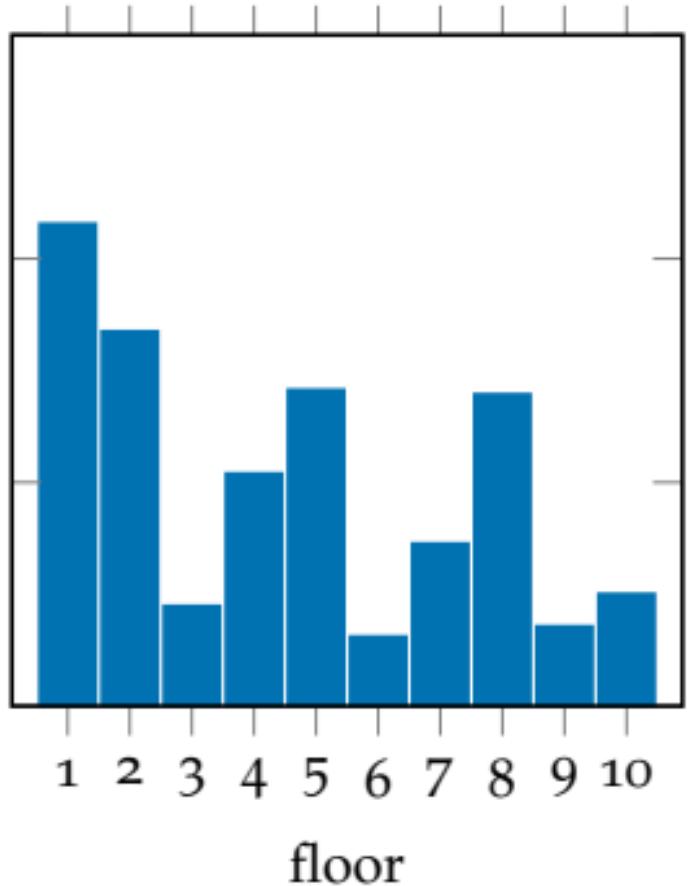
relative duration

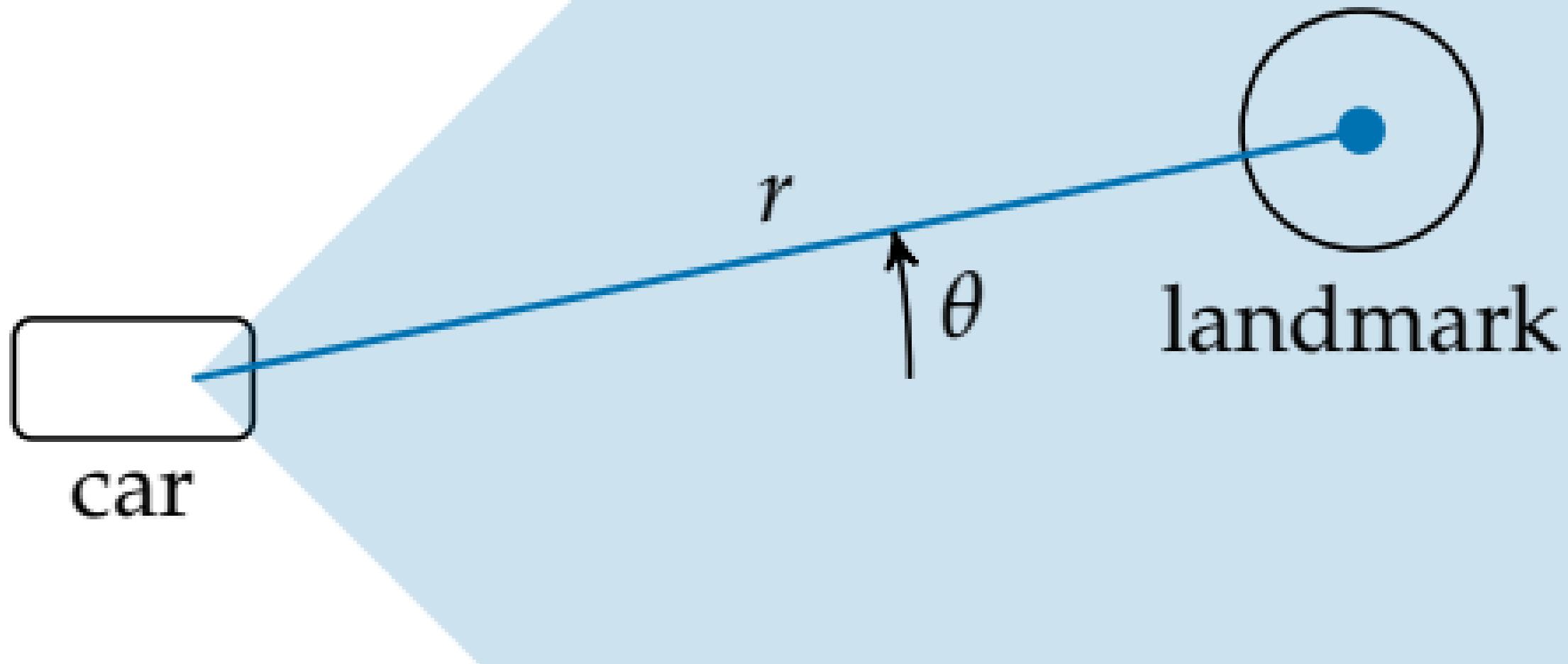


policy B

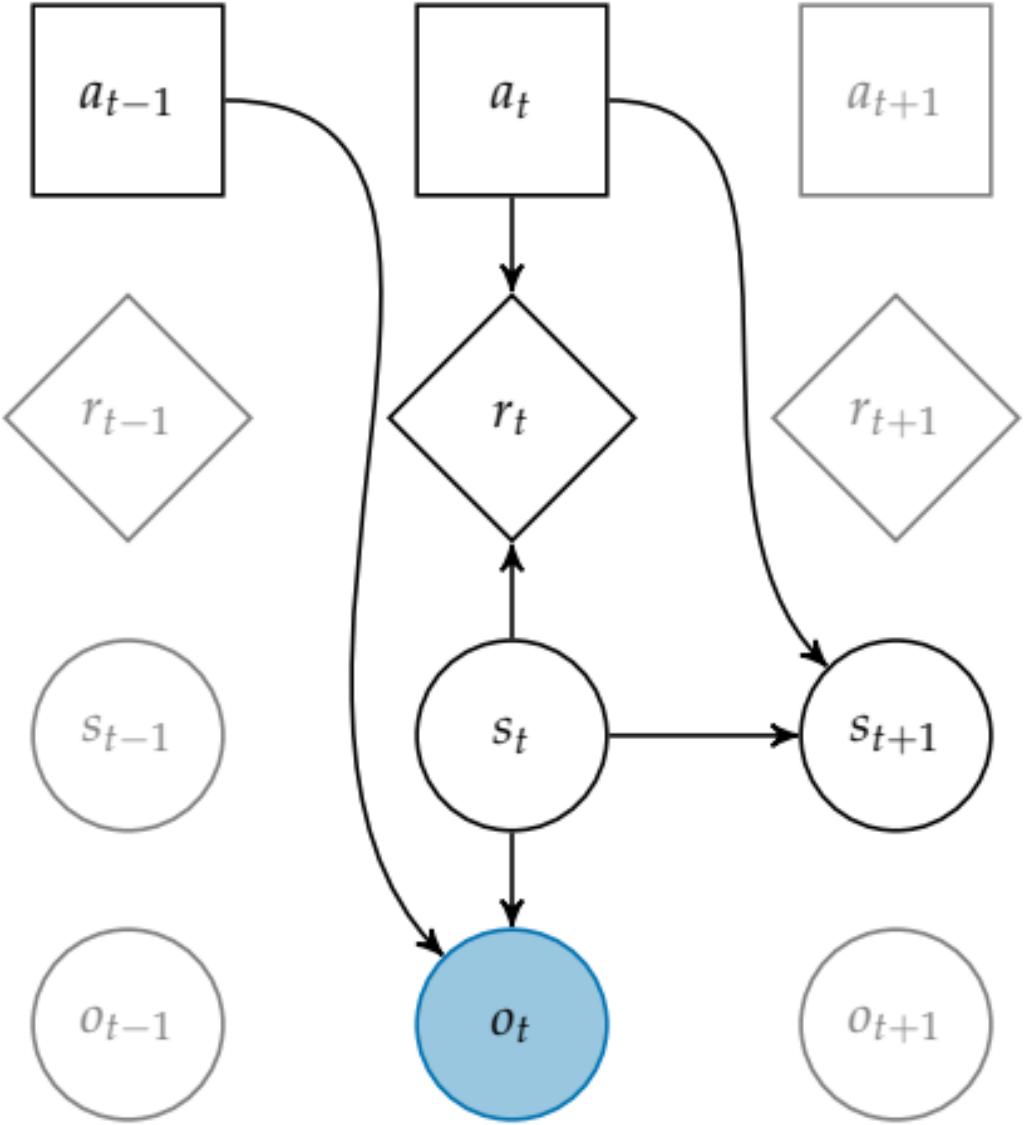


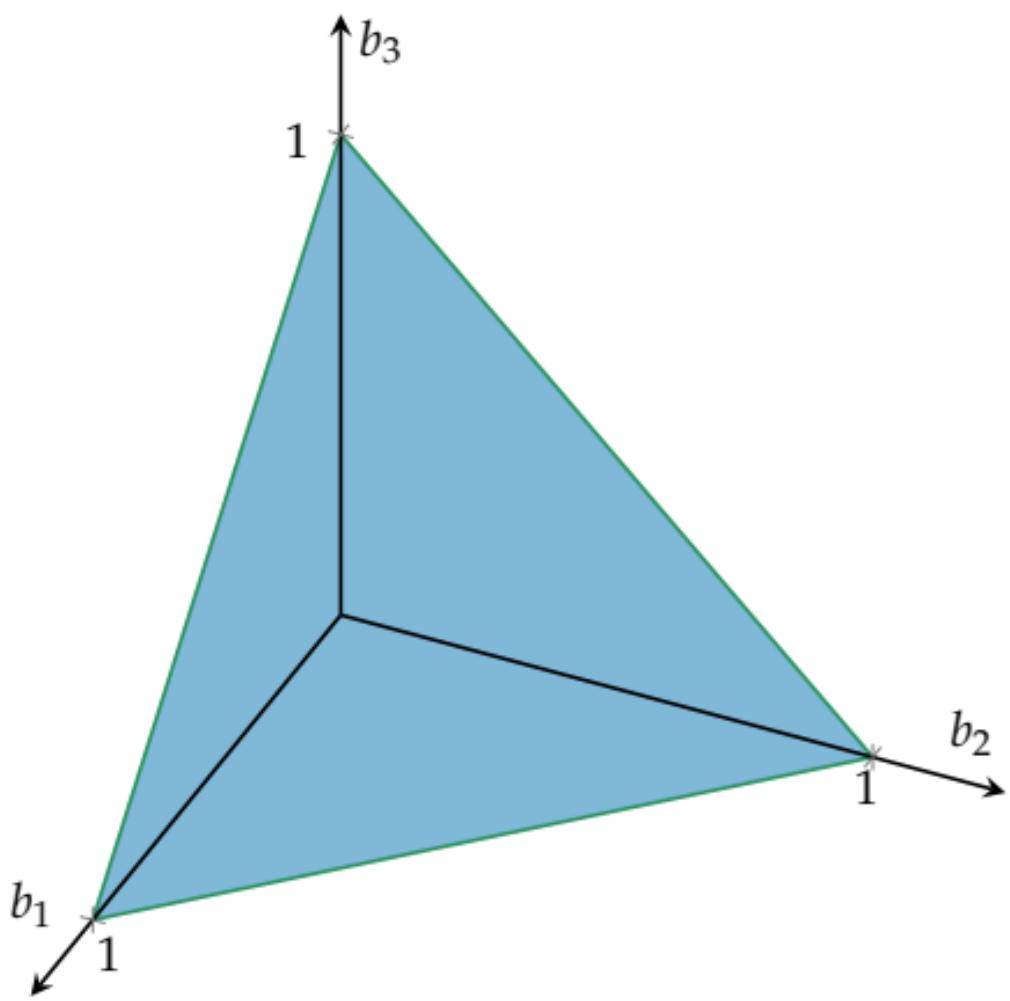
policy C





landmark



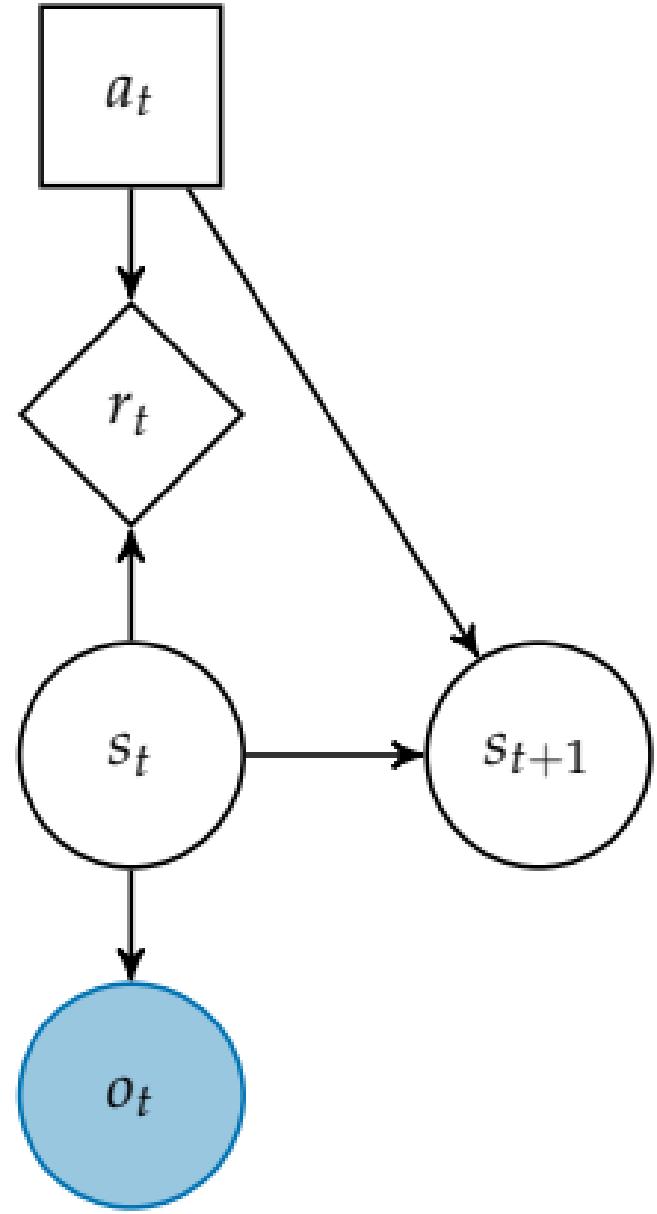


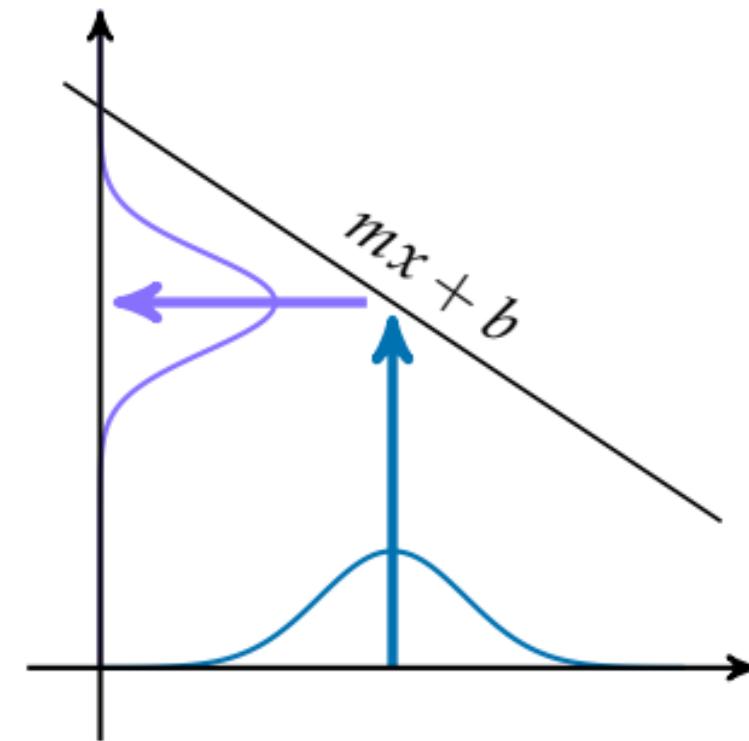
feed / sing
/ ignore

reward

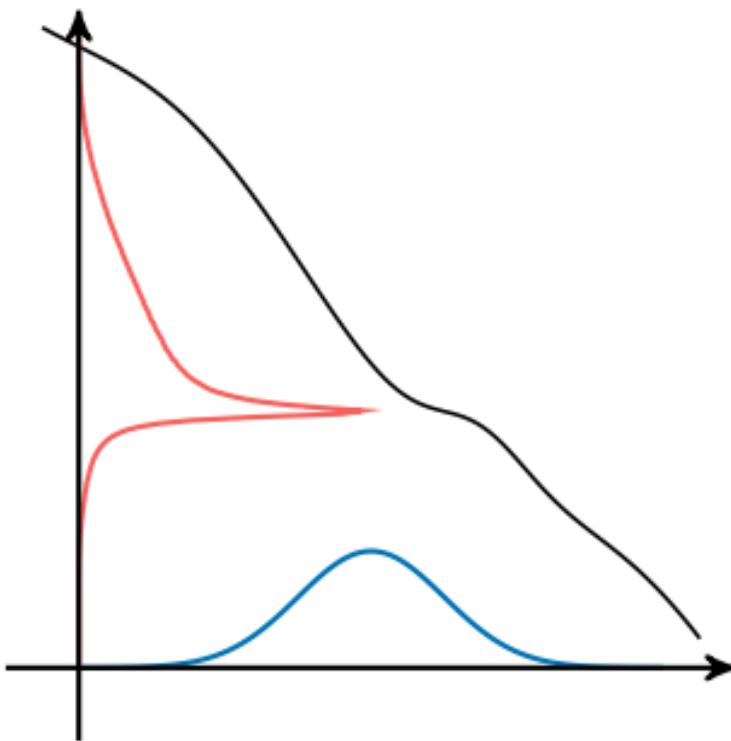
hungry

crying

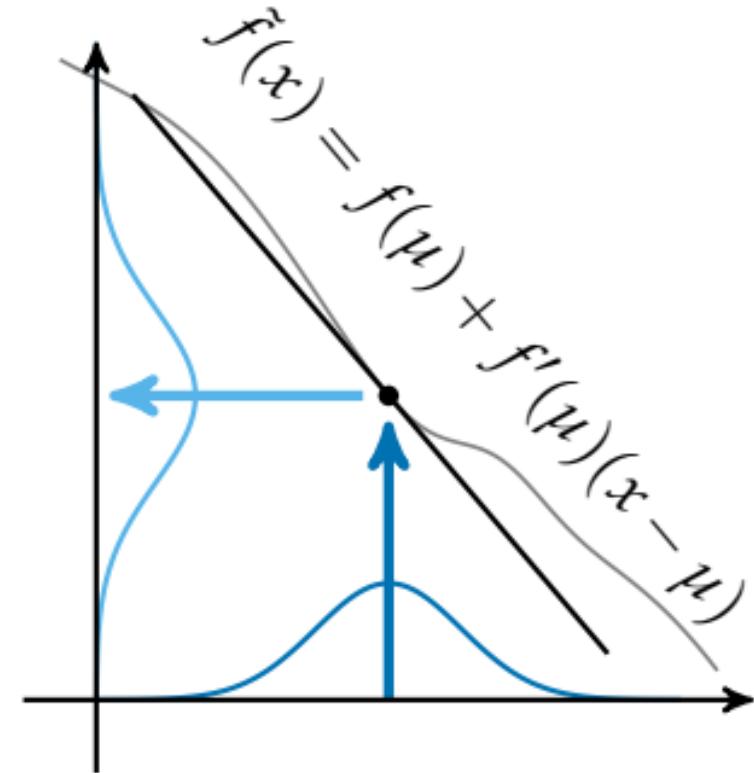




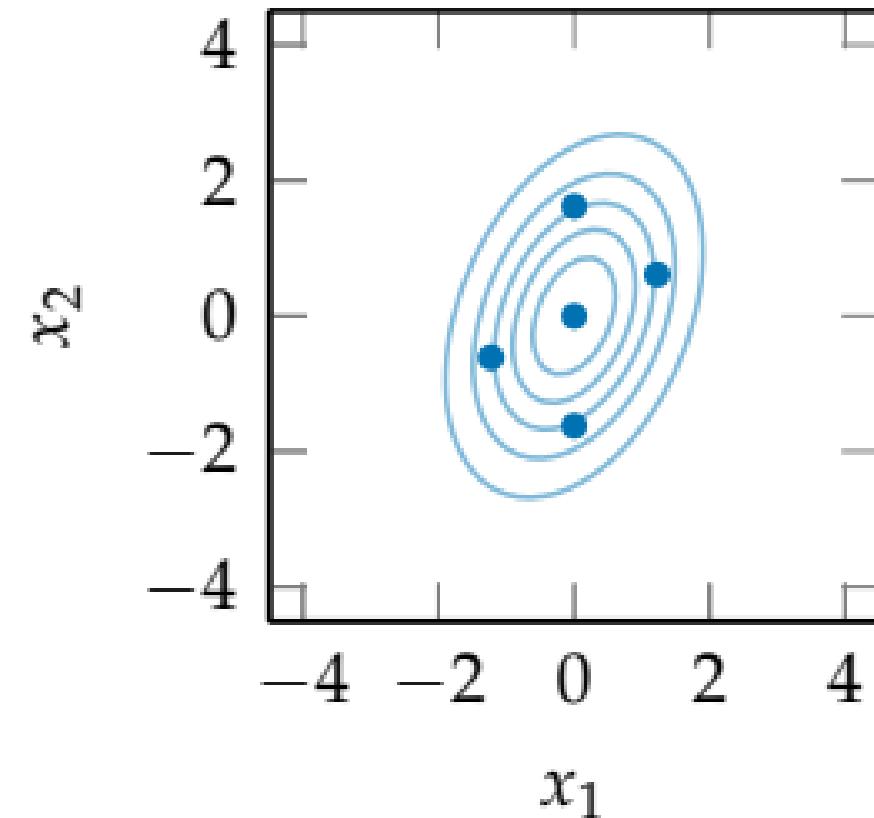
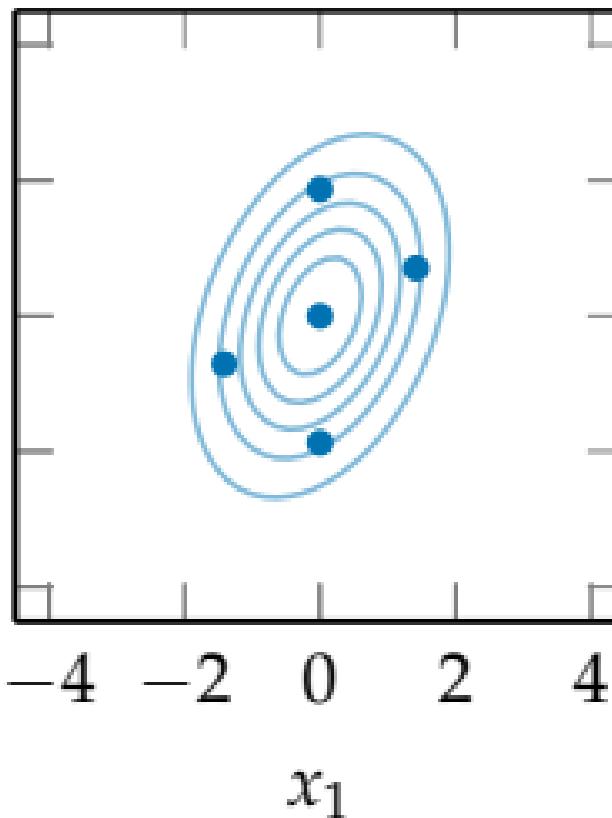
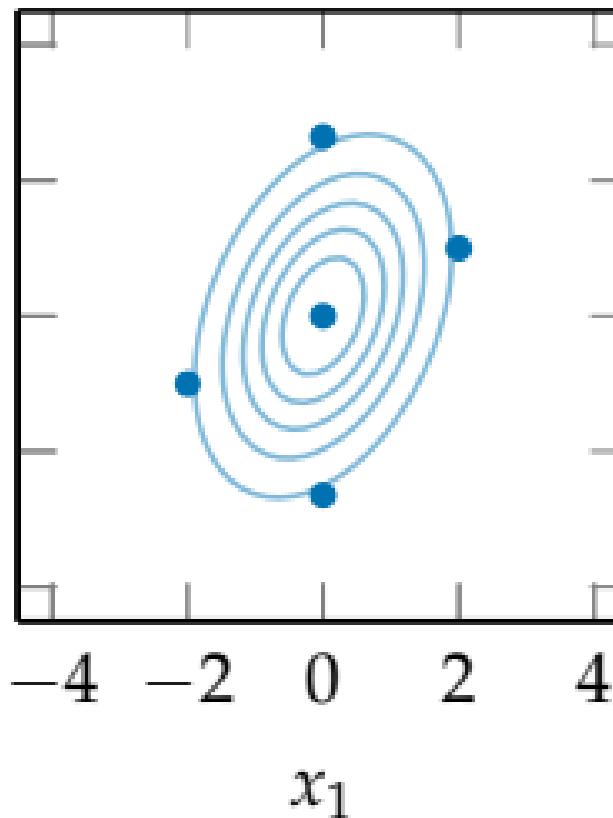
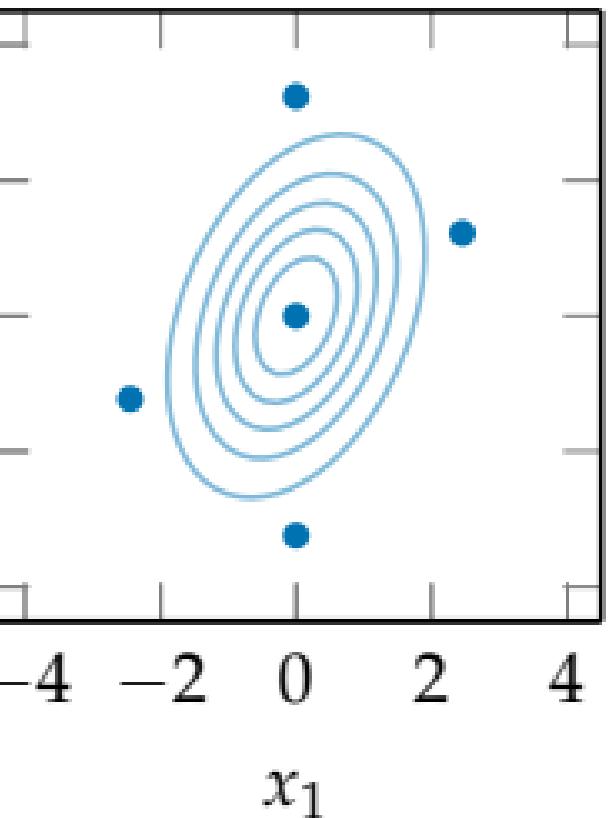
linear dynamics

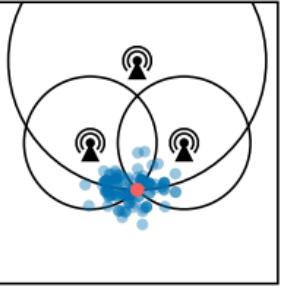
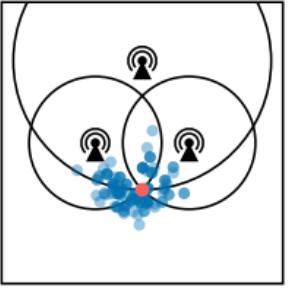
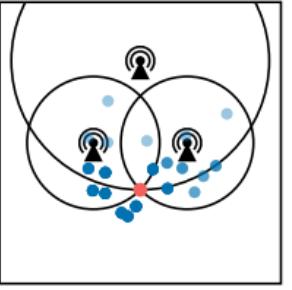
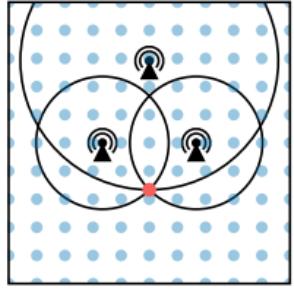
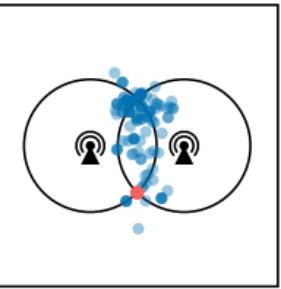
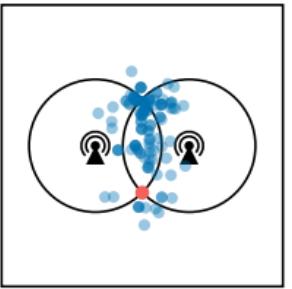
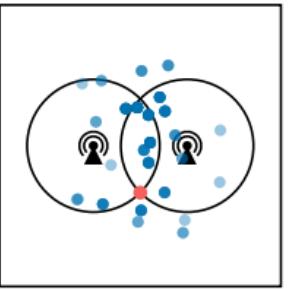
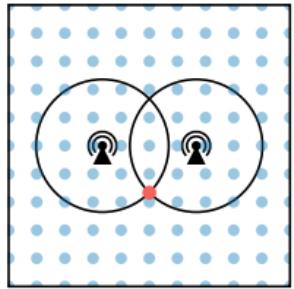
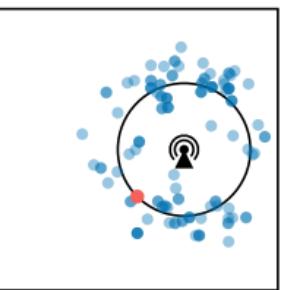
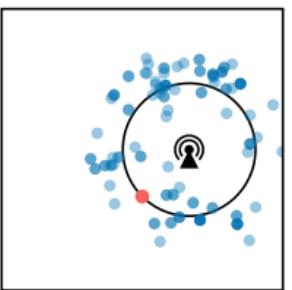
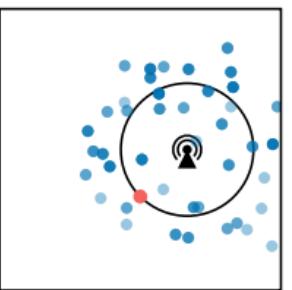
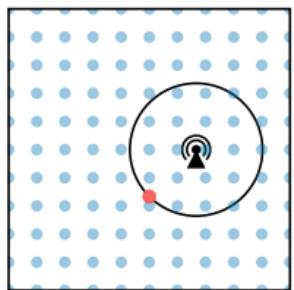


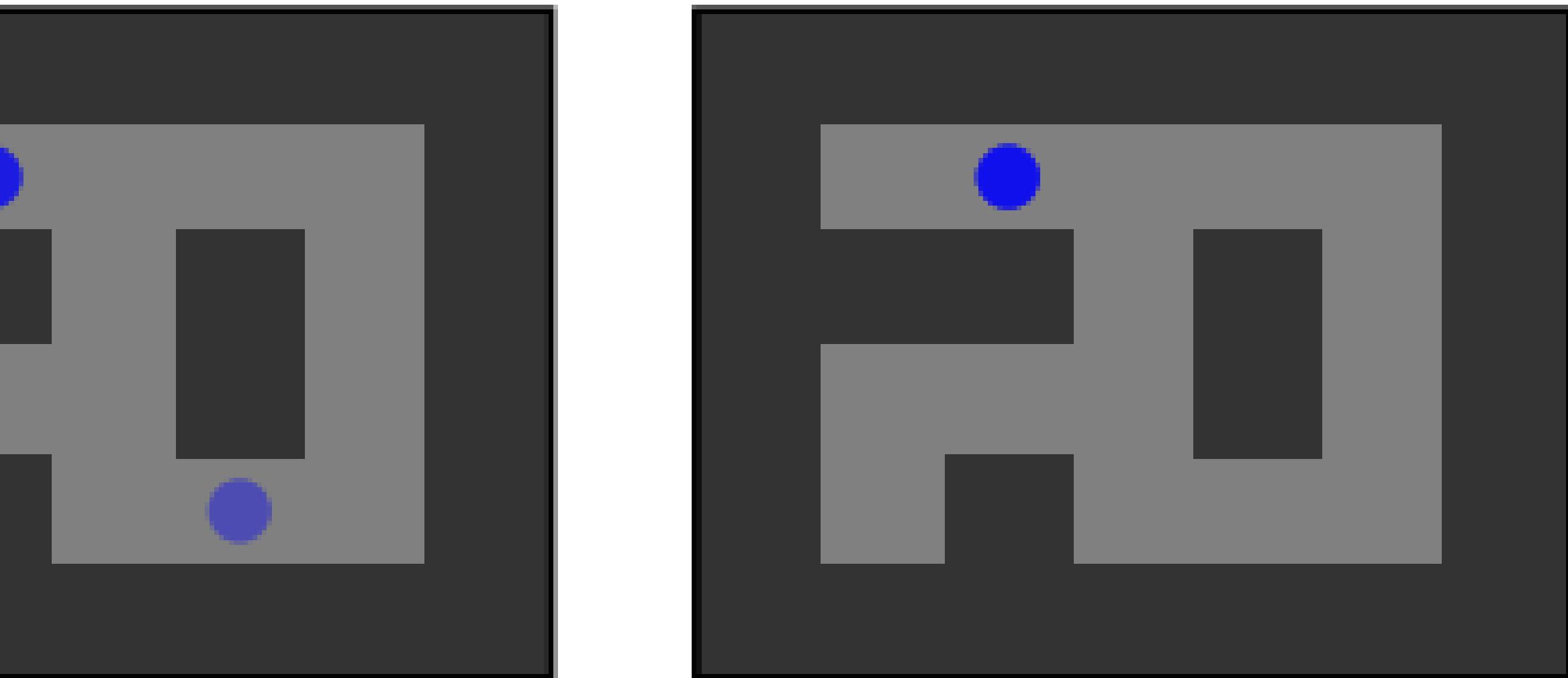
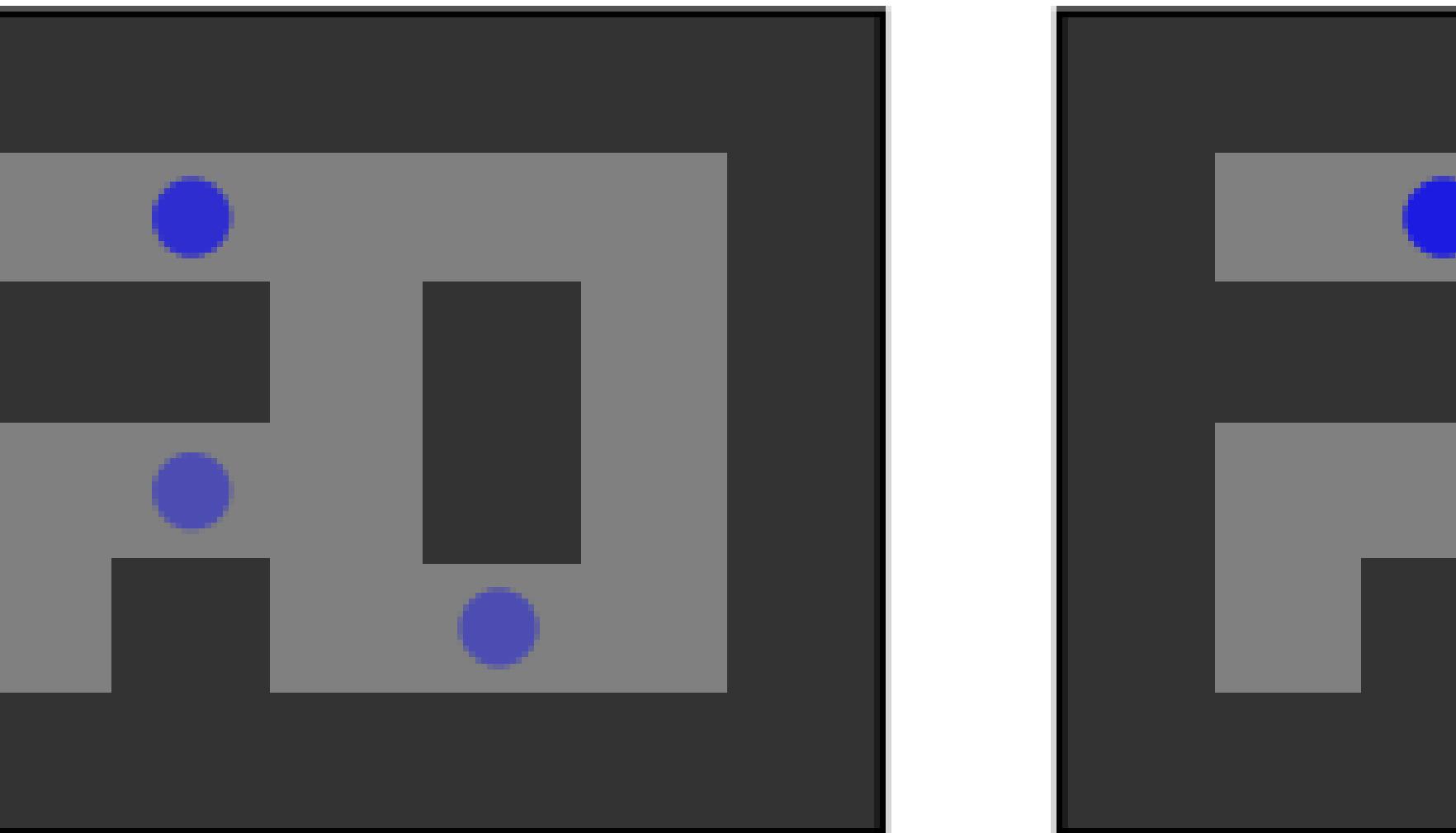
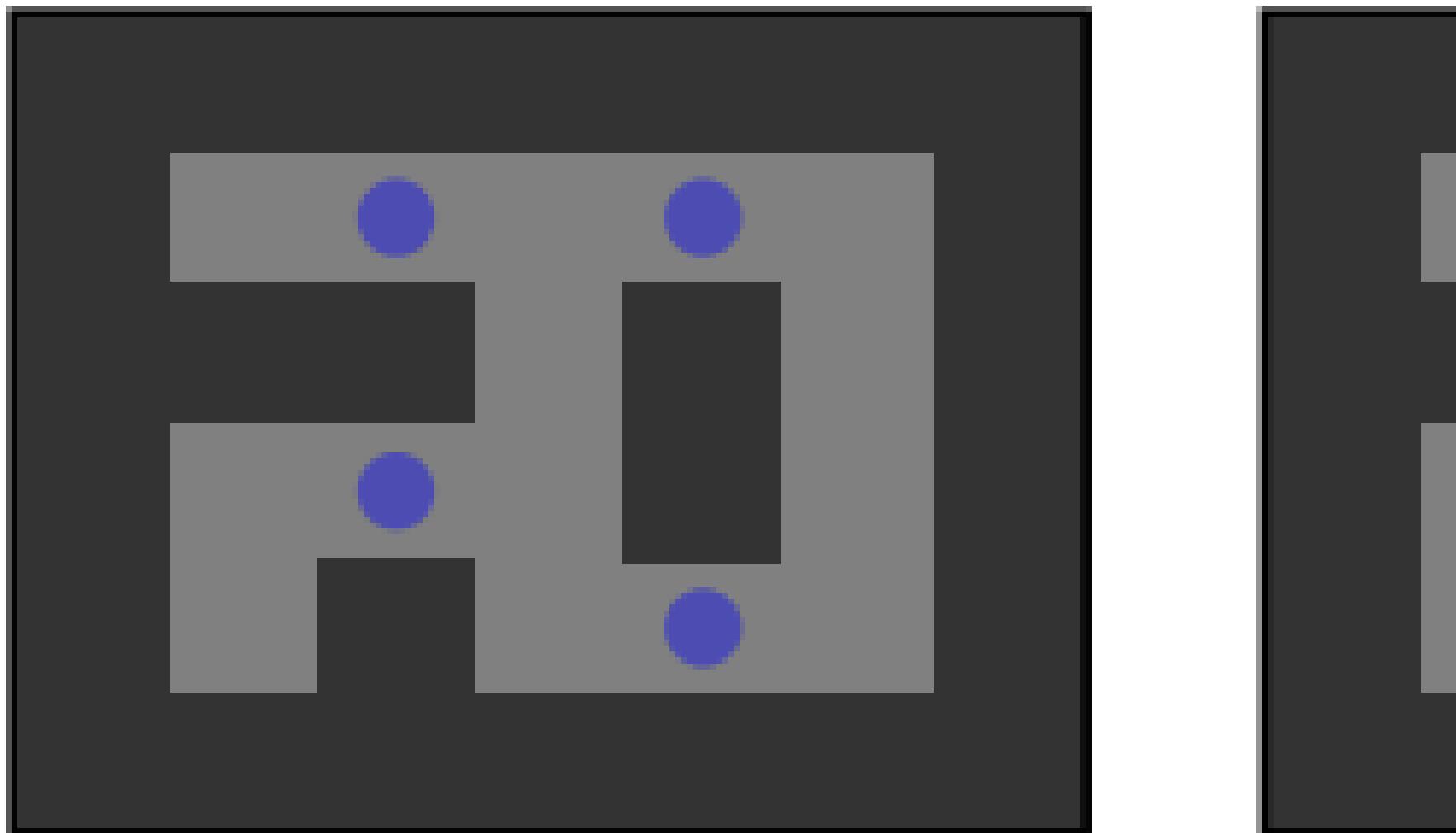
nonlinear dynamics

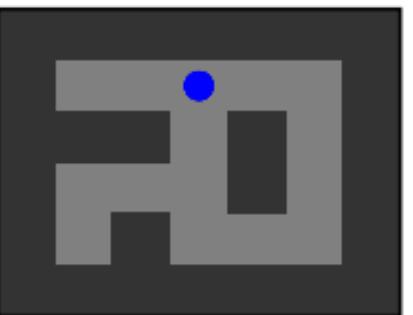
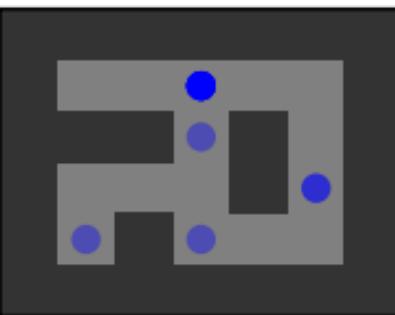
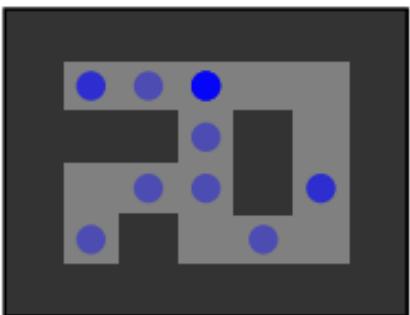
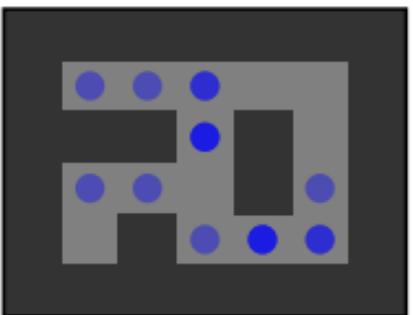
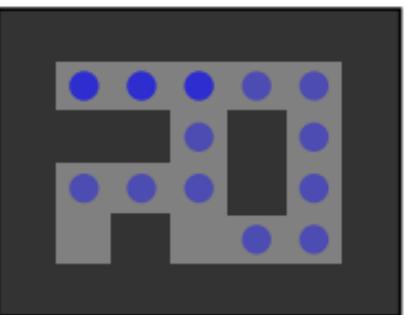
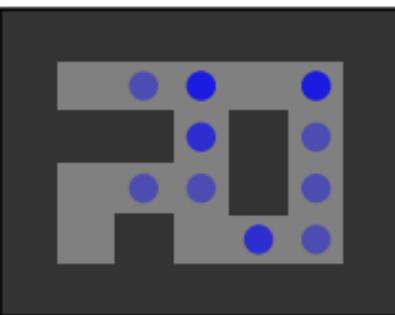


linear approximation

$\lambda = -0.5$  $\lambda = 0.0$  $\lambda = 2.0$  $\lambda = 4.0$ 

$t = 1$ $t = 2$ $t = 3$ $t = 4$  x x x x y y y



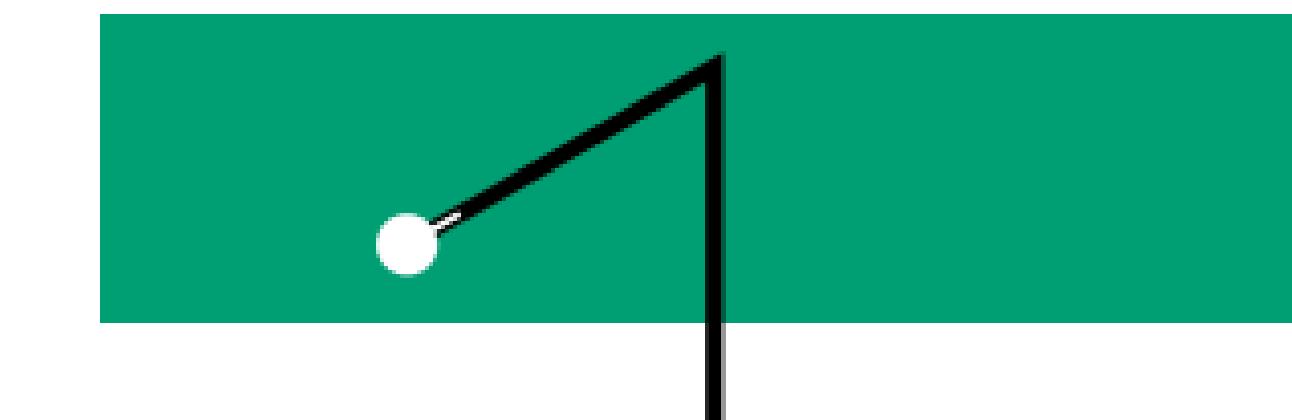
$w_{slow} = 1.0$
 $w_{fast} = 1.0$  $w_{slow} = 0.99$
 $w_{fast} = 0.7$  $w_{slow} = 0.98$
 $w_{fast} = 0.49$  $w_{slow} = 0.97$
 $w_{fast} = 0.34$  $w_{slow} = 0.96$
 $w_{fast} = 0.24$  $w_{slow} = 0.95$
 $w_{fast} = 0.17$  $w_{slow} = 0.94$
 $w_{fast} = 0.12$  $w_{slow} = 0.93$
 $w_{fast} = 0.1$ 



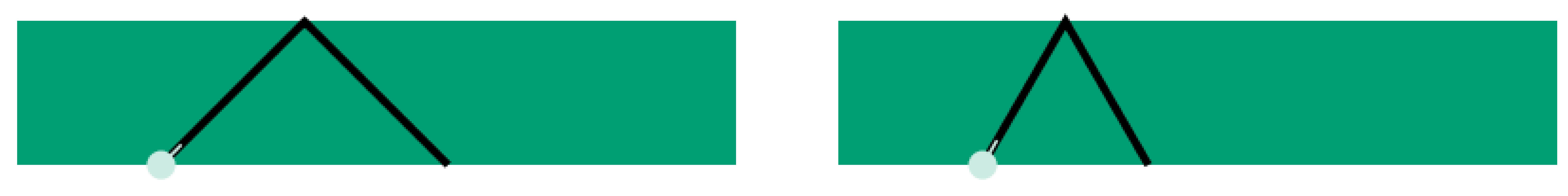
A straight path of length $2w$
and a diagonal path of length $\sqrt{2}w$

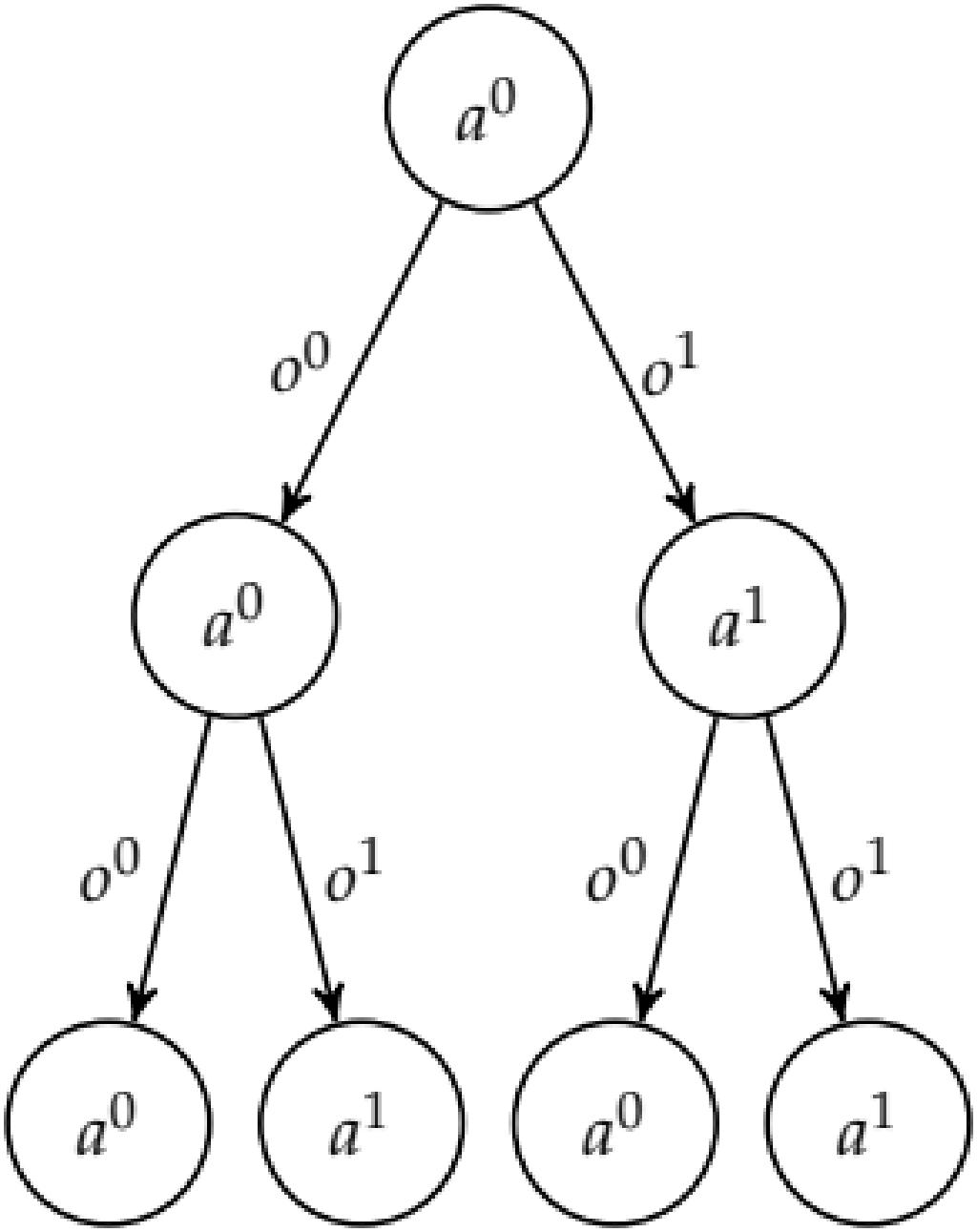


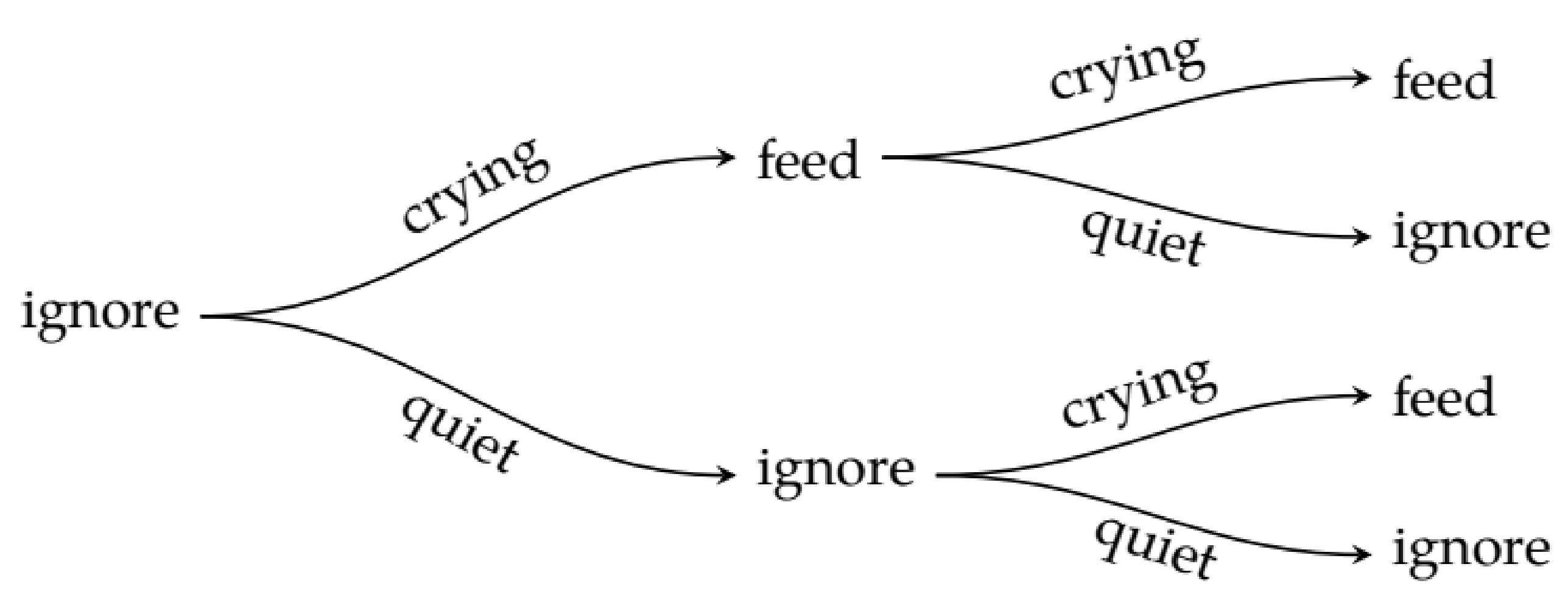
Two perpendicular segments,
each of length $\sqrt{2}w$

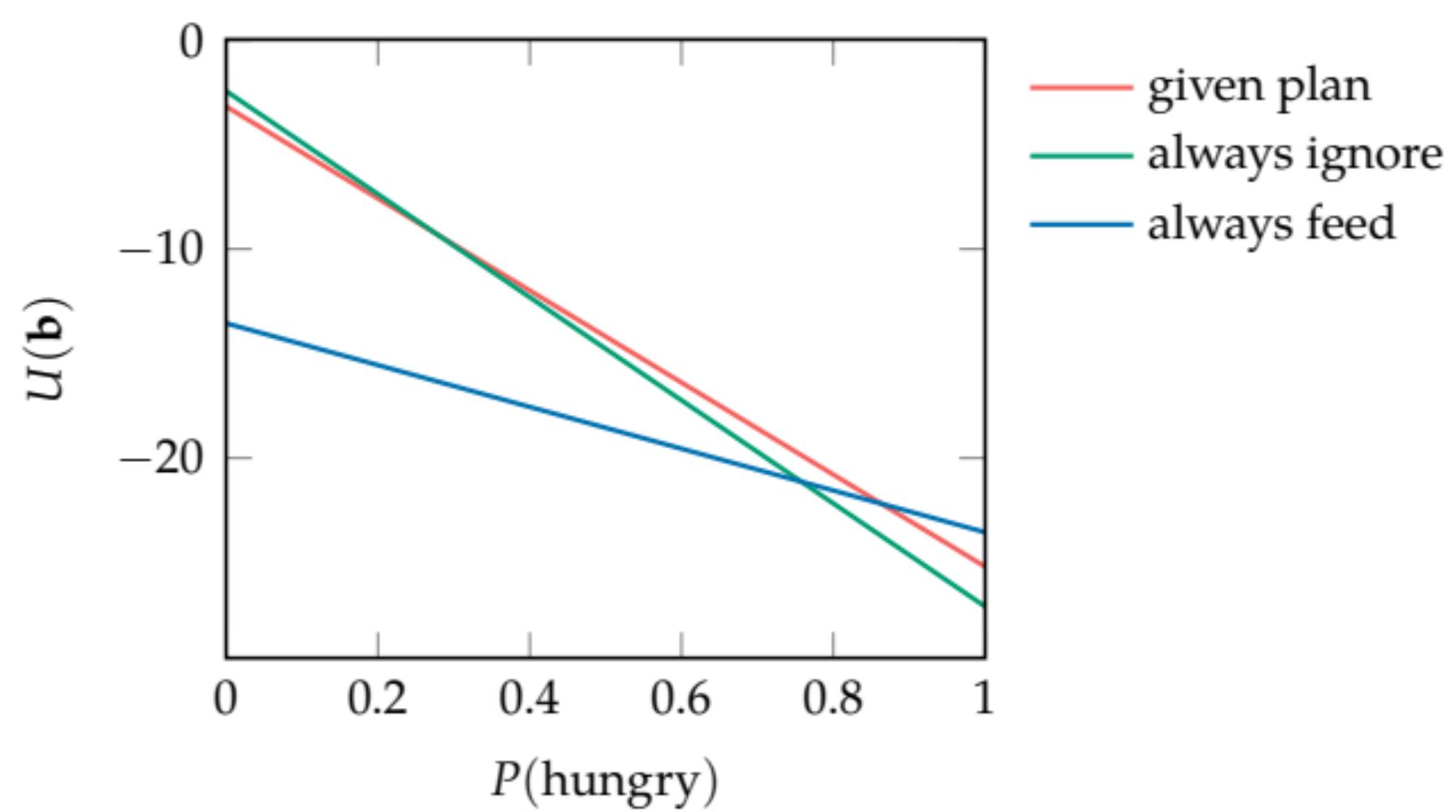


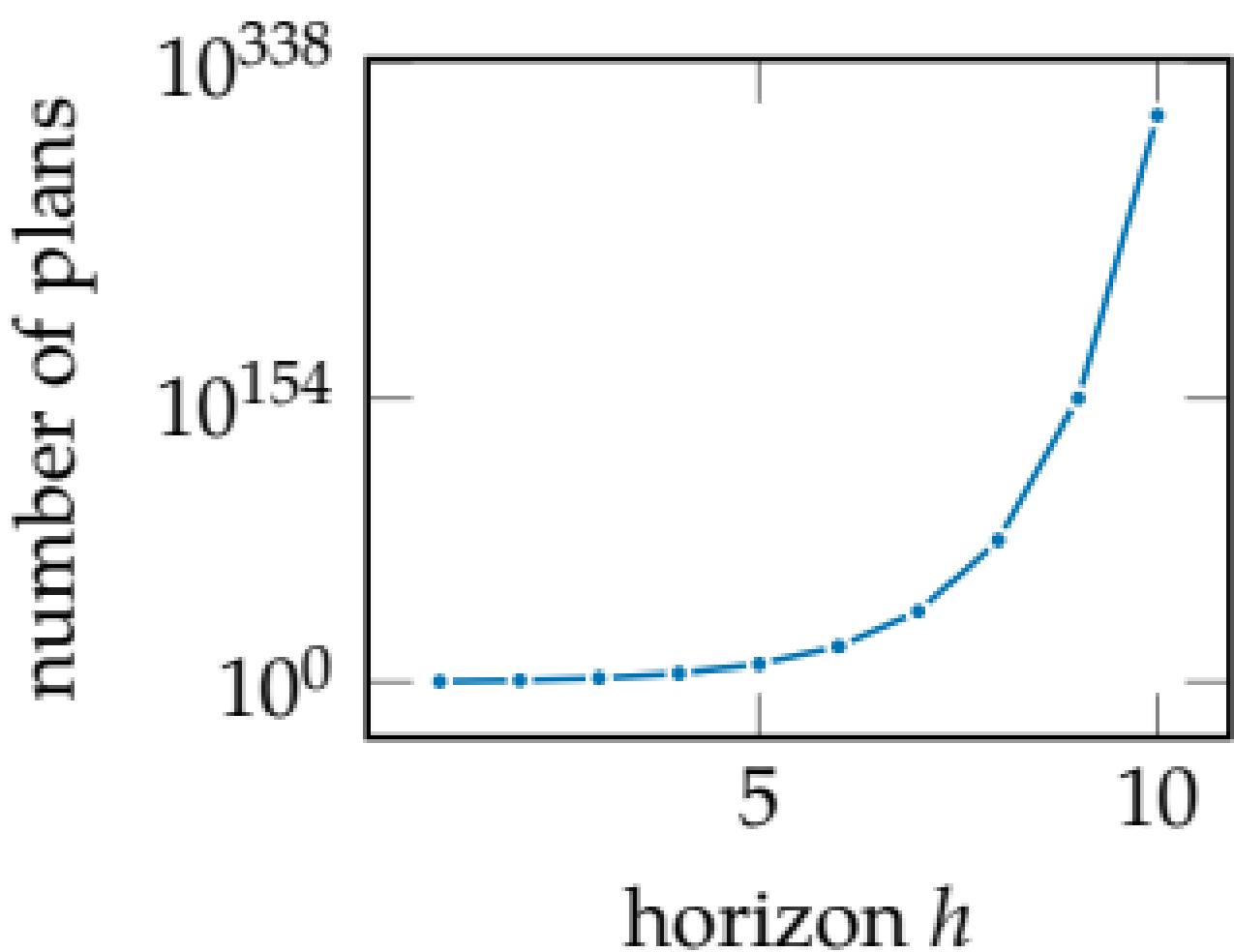
Two legs of an equilateral triangle, each of length $\frac{2\sqrt{3}}{3}w$











$$R(b, \text{feed}) = -10$$

$$\gamma P(\text{crying} \mid b, \text{feed}) U(\text{Update}(b, \text{feed}, \text{crying})) = -0.18$$

$$\gamma P(\text{quiet} \mid b, \text{feed}) U(\text{Update}(b, \text{feed}, \text{quiet})) = -1.62$$

$$\rightarrow Q(b, \text{feed}) = -11.8$$

$$R(b, \text{ignore}) = -5$$

$$\gamma P(\text{crying} \mid b, \text{ignore}) U(\text{Update}(b, \text{ignore}, \text{crying})) = -6.09$$

$$\gamma P(\text{quiet} \mid b, \text{ignore}) U(\text{Update}(b, \text{ignore}, \text{quiet})) = -2.81$$

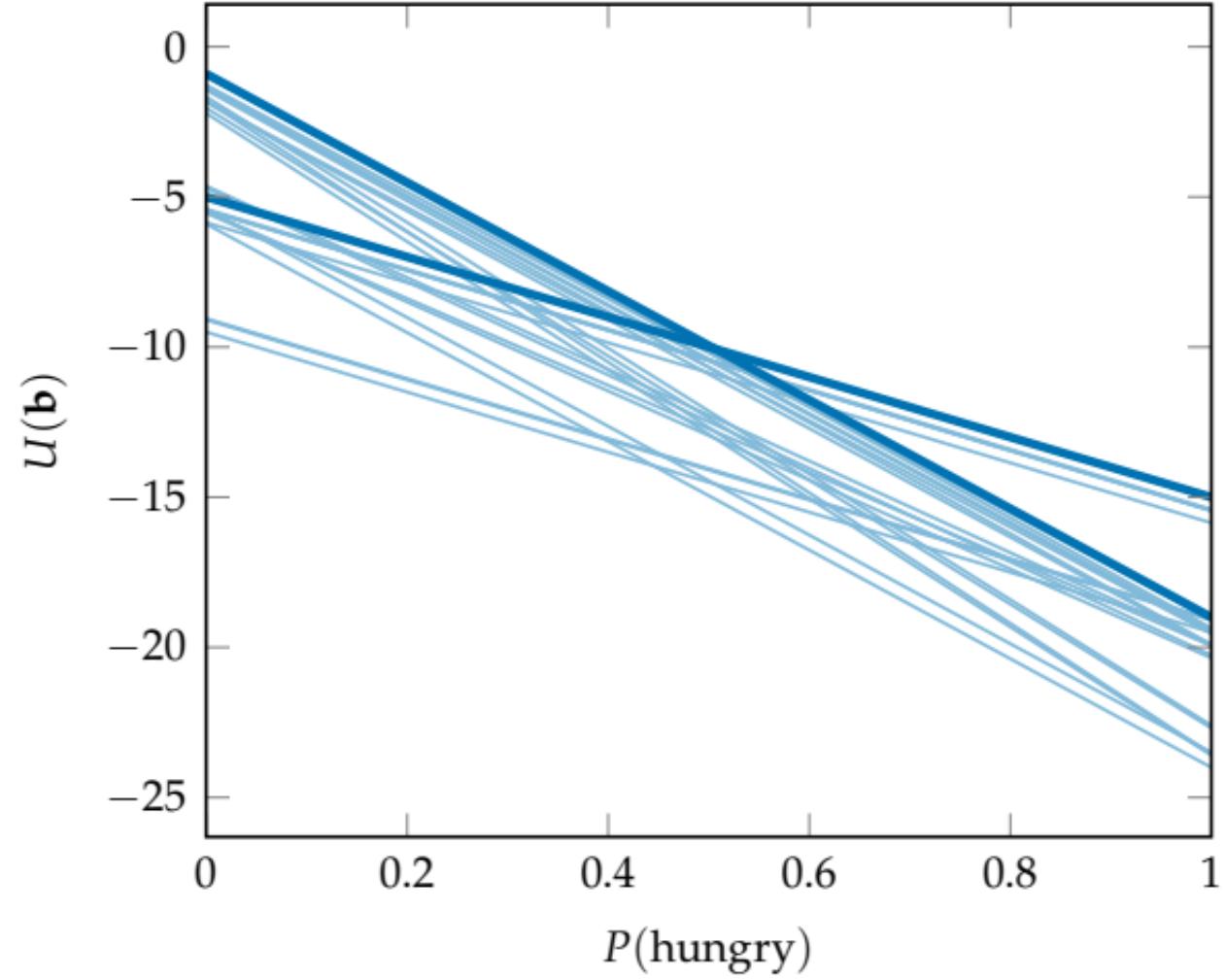
$$\rightarrow Q(b, \text{ignore}) = -13.9$$

$$R(b, \text{sing}) = -5.5$$

$$\gamma P(\text{crying} \mid b, \text{sing}) U(\text{Update}(b, \text{sing}, \text{crying})) = -6.68$$

$$\gamma P(\text{quiet} \mid b, \text{sing}) U(\text{Update}(b, \text{sing}, \text{quiet})) = -1.85$$

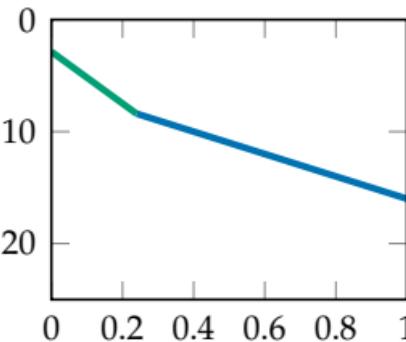
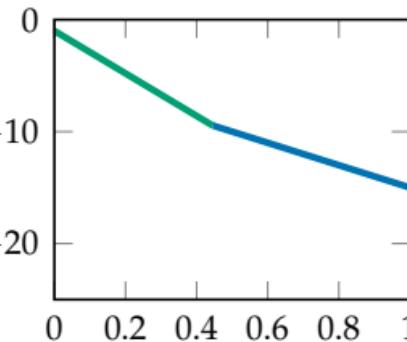
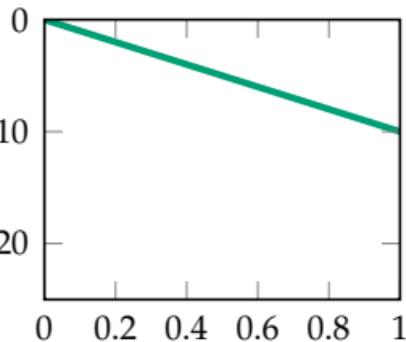
$$\rightarrow Q(b, \text{sing}) = -14.0$$



1-step plans

2-step plans

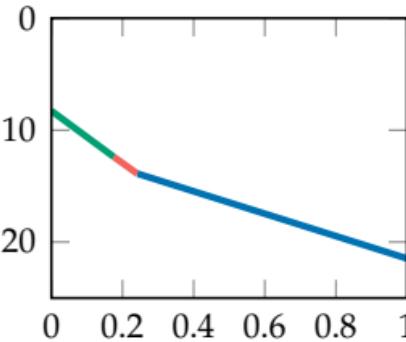
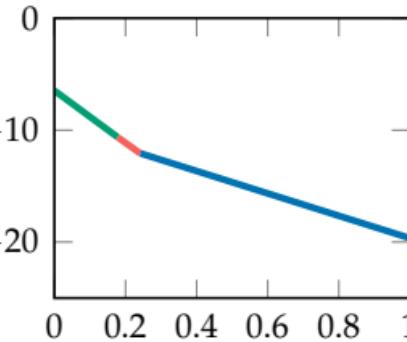
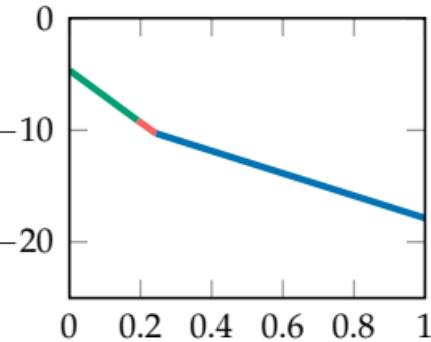
3-step plans

 $U(\mathbf{p})$ 

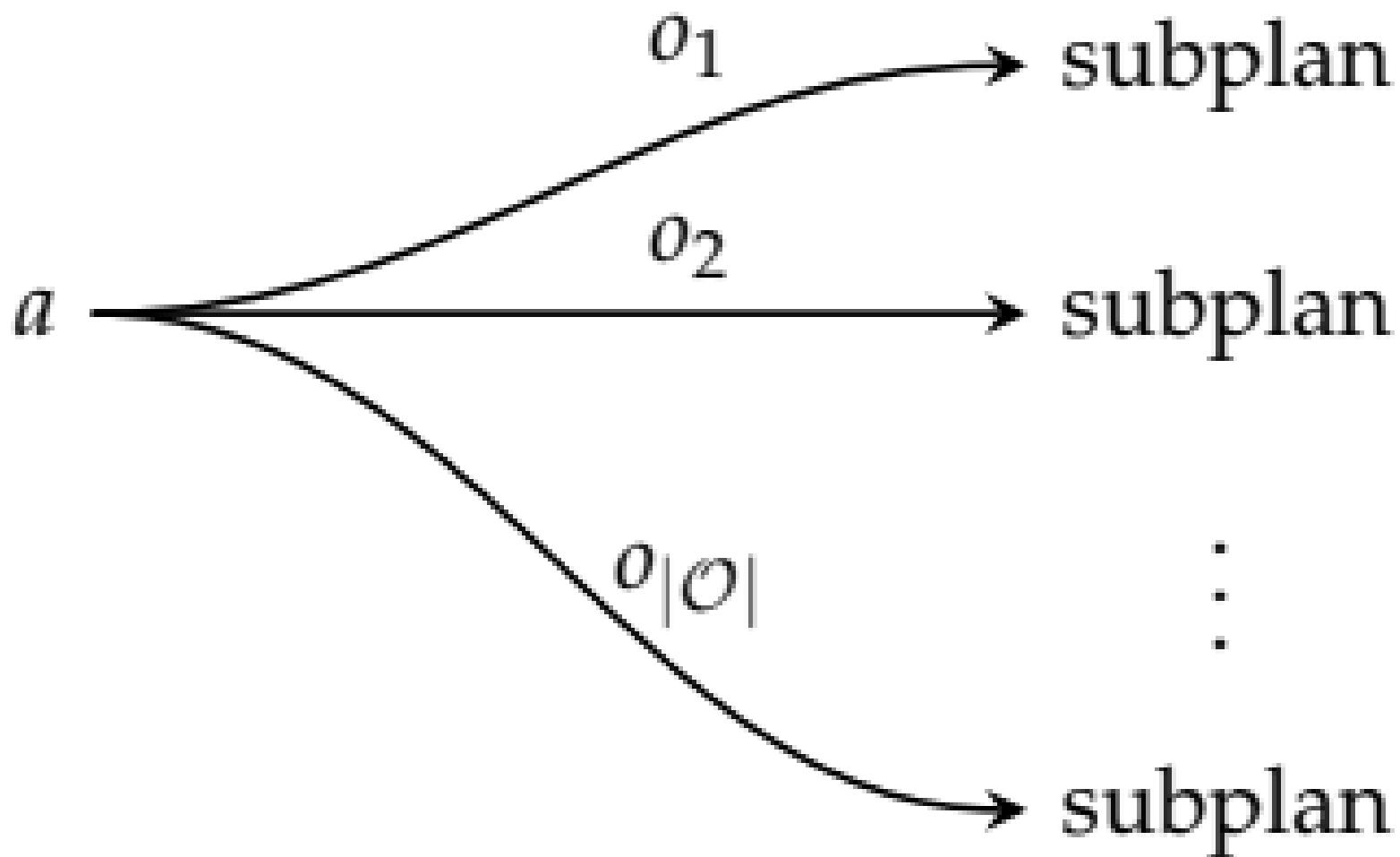
4-step plans

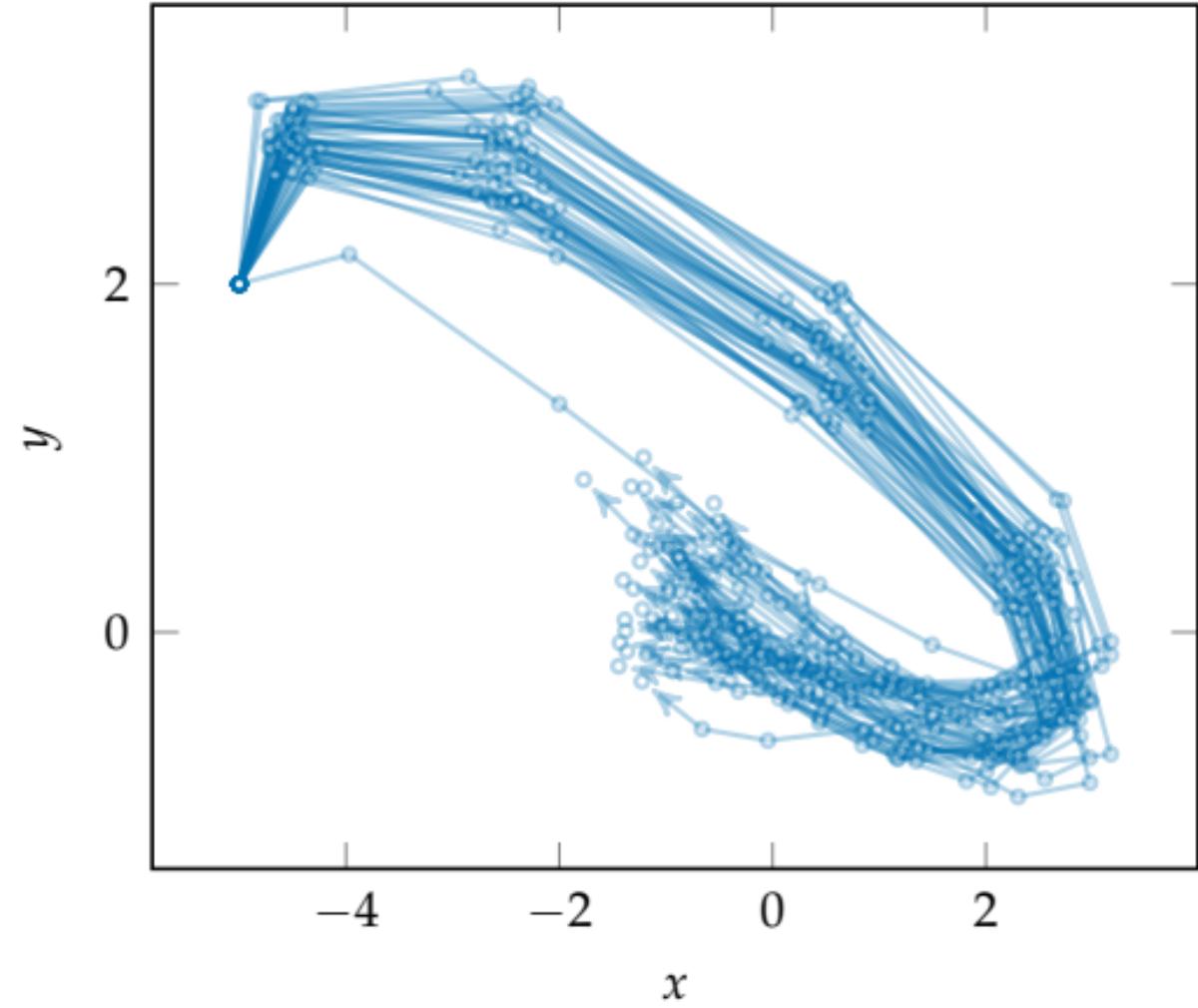
5-step plans

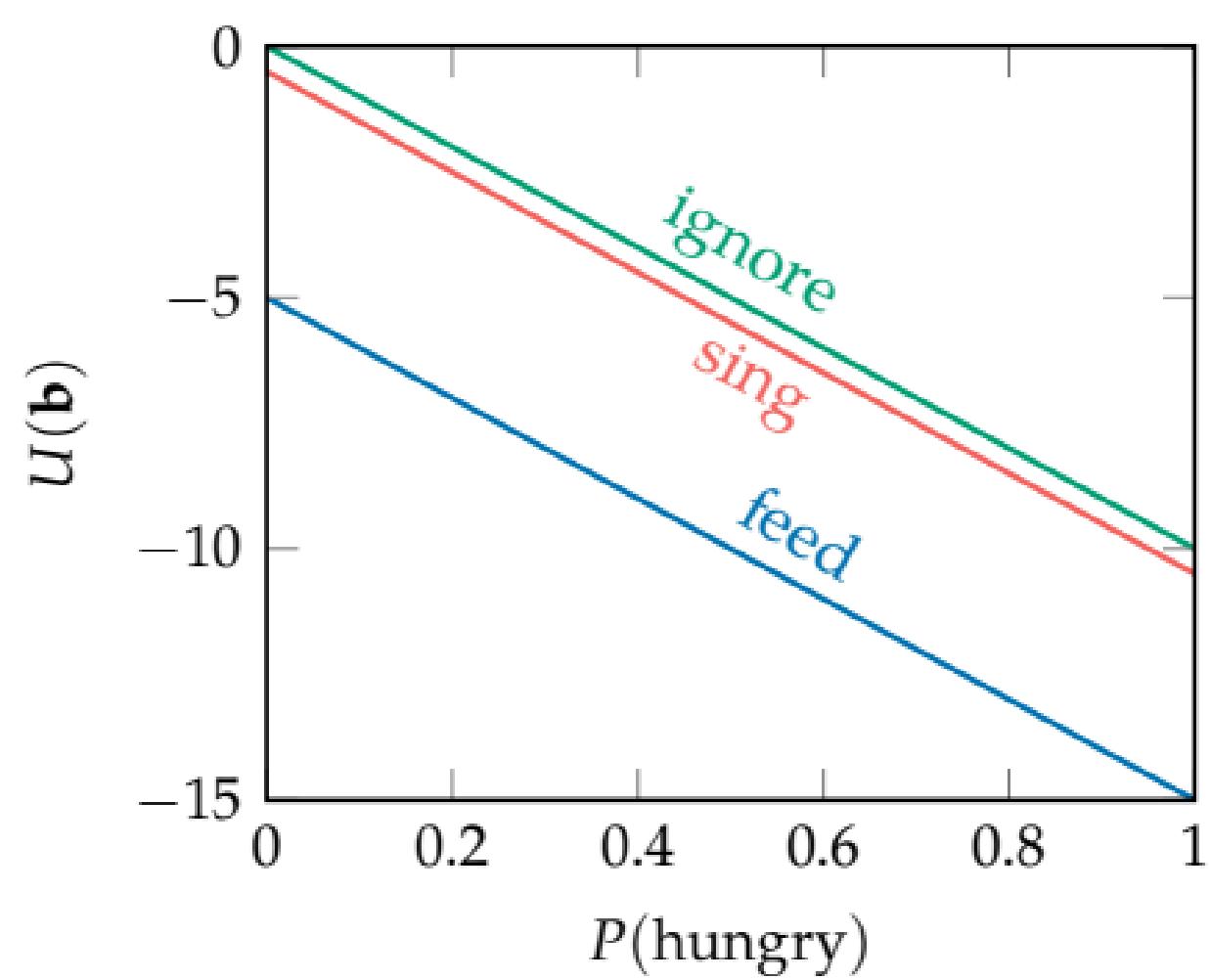
6-step plans

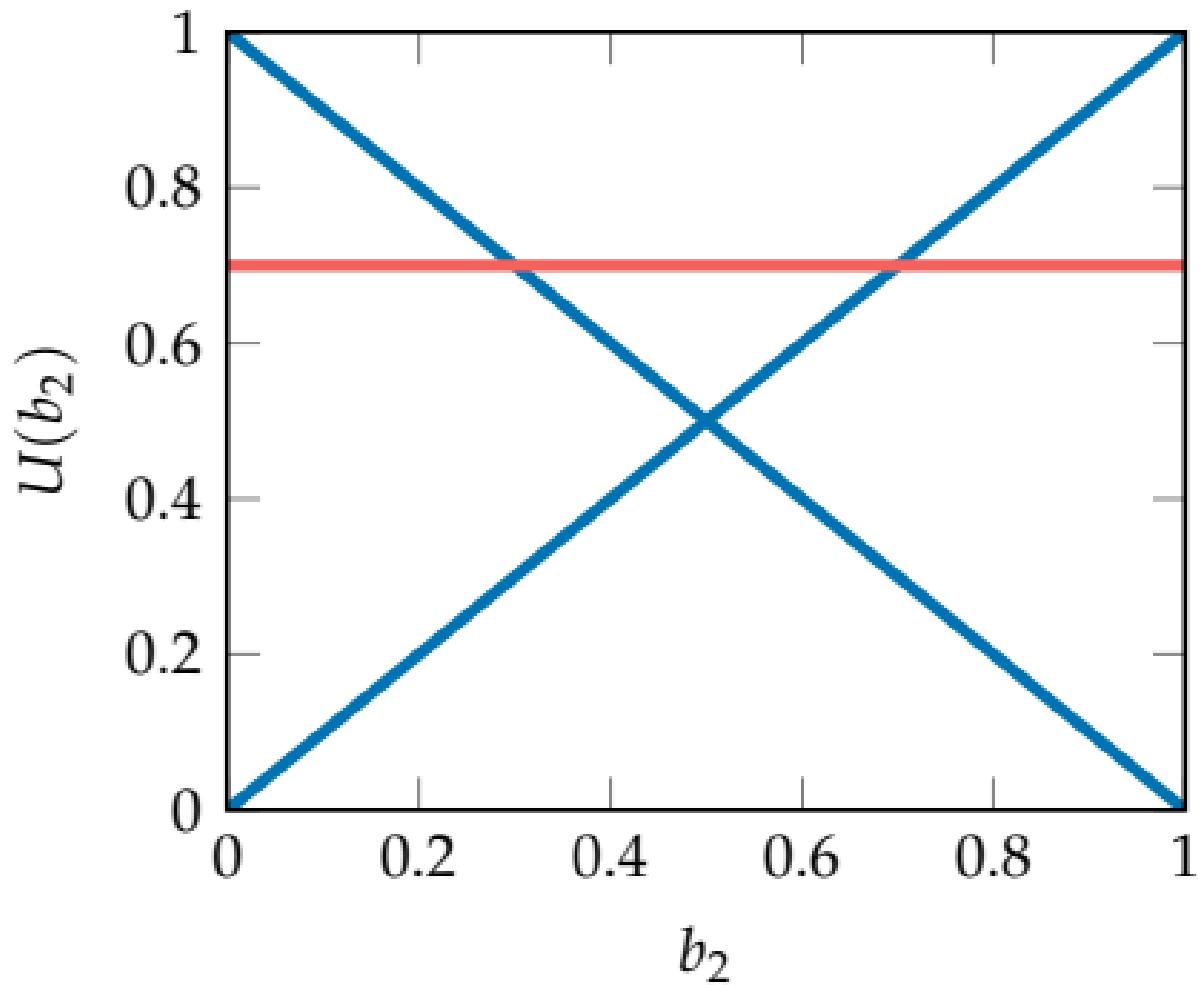
 $U(\mathbf{p})$  $P(\text{hungry})$ $P(\text{hungry})$ $P(\text{hungry})$

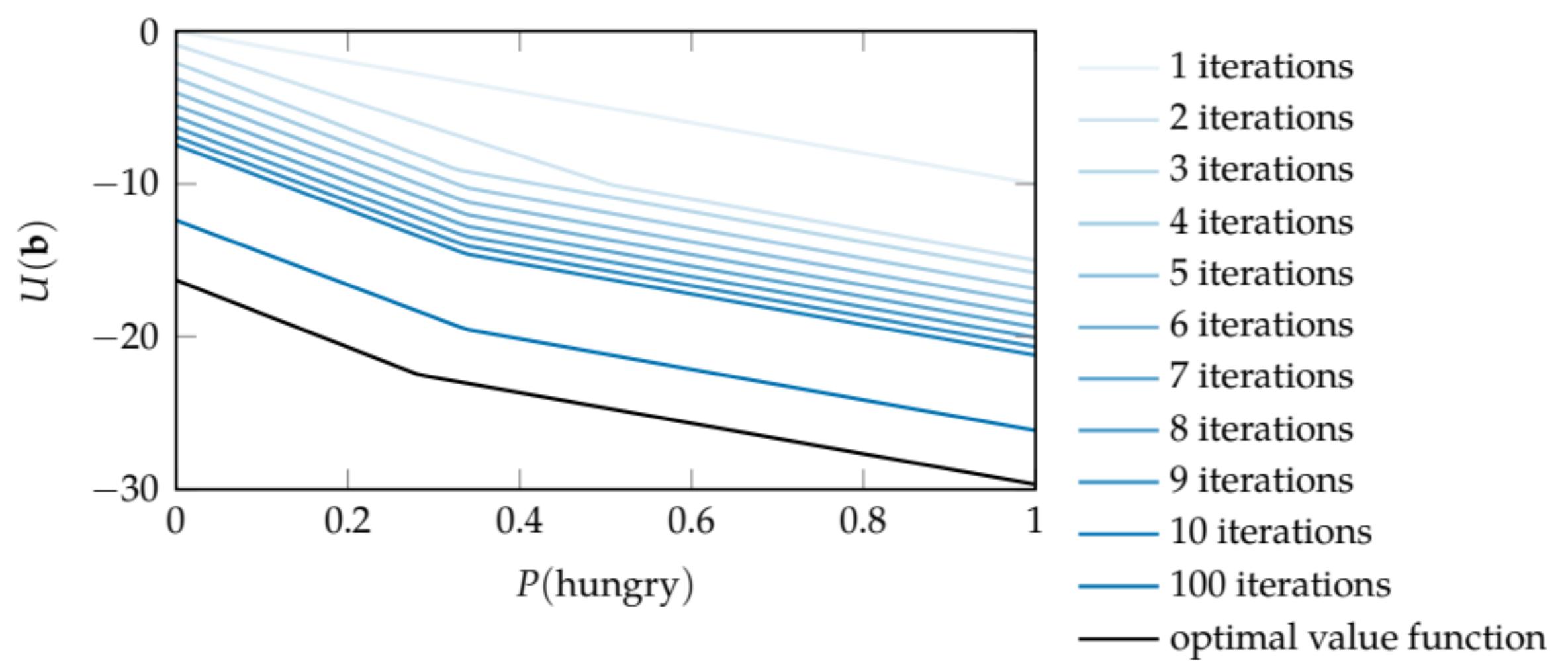
— ignore — sing — feed

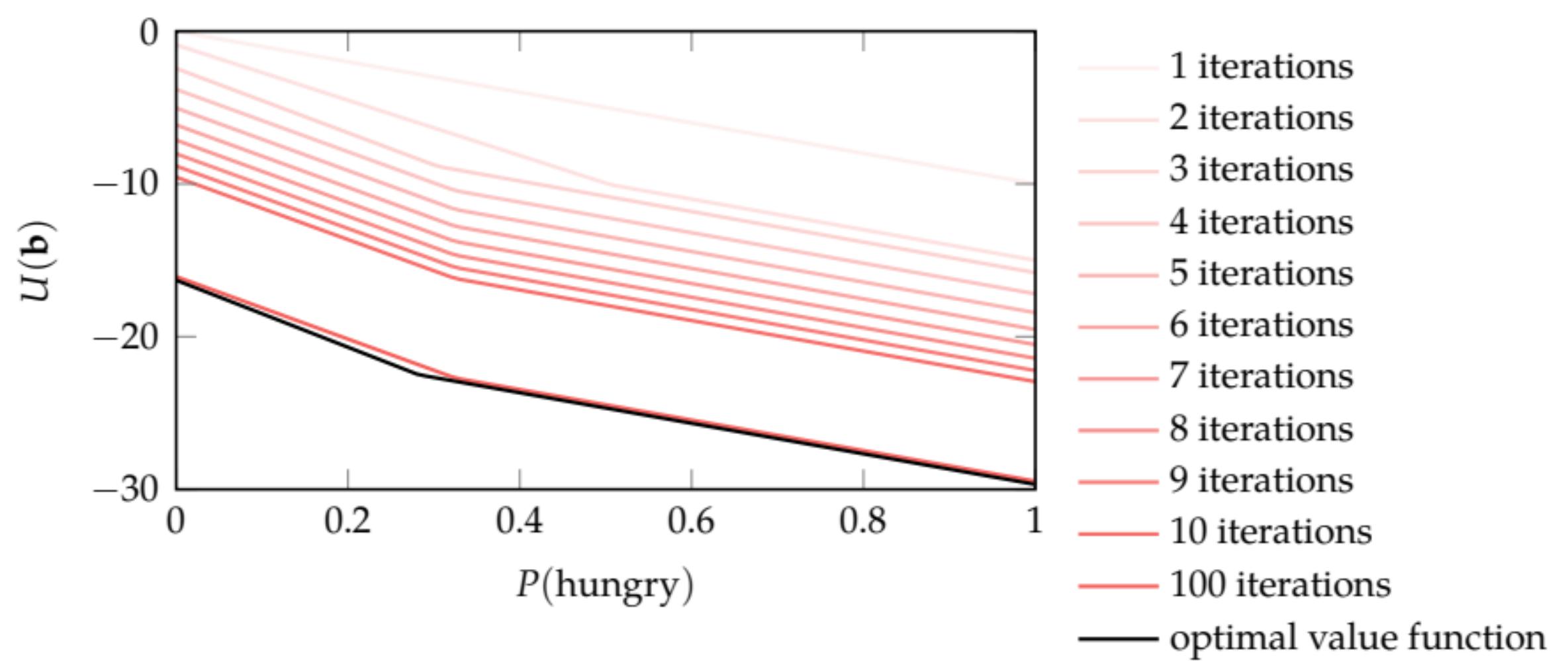


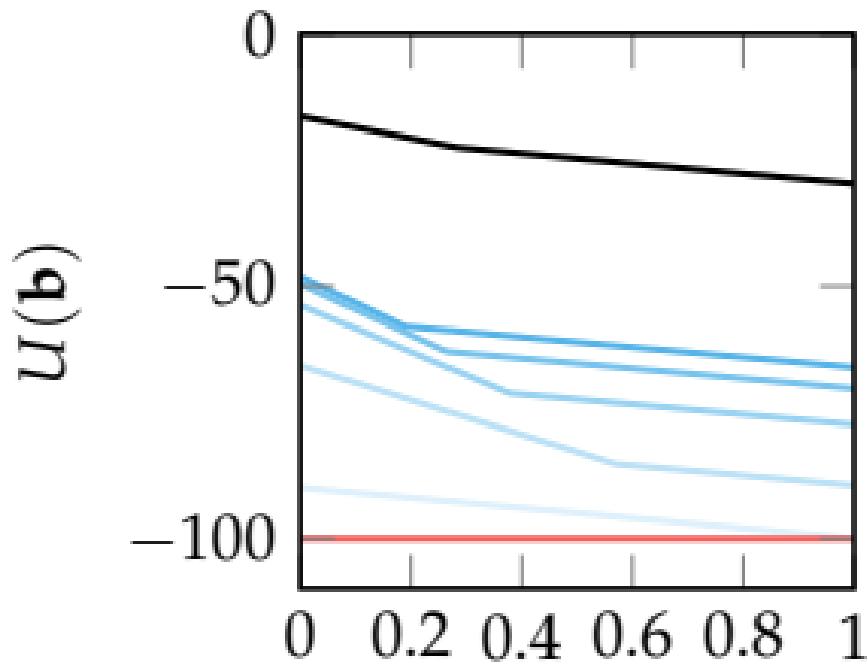






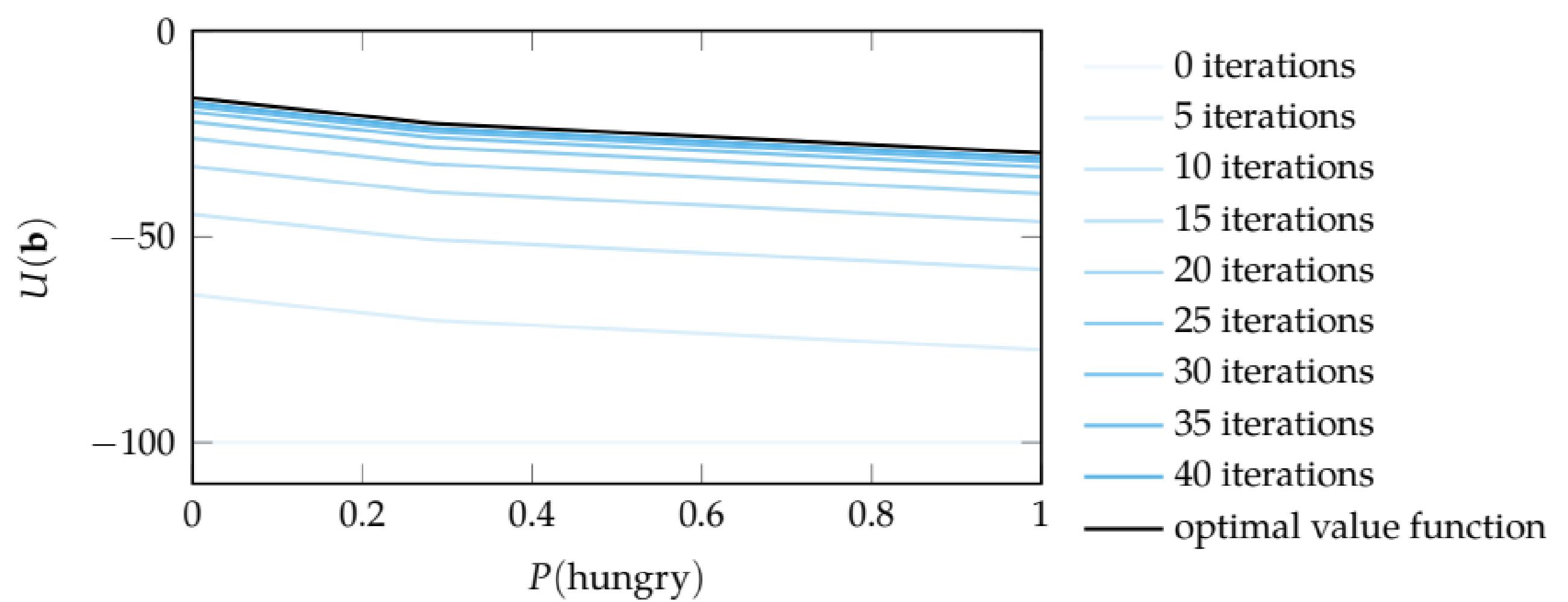


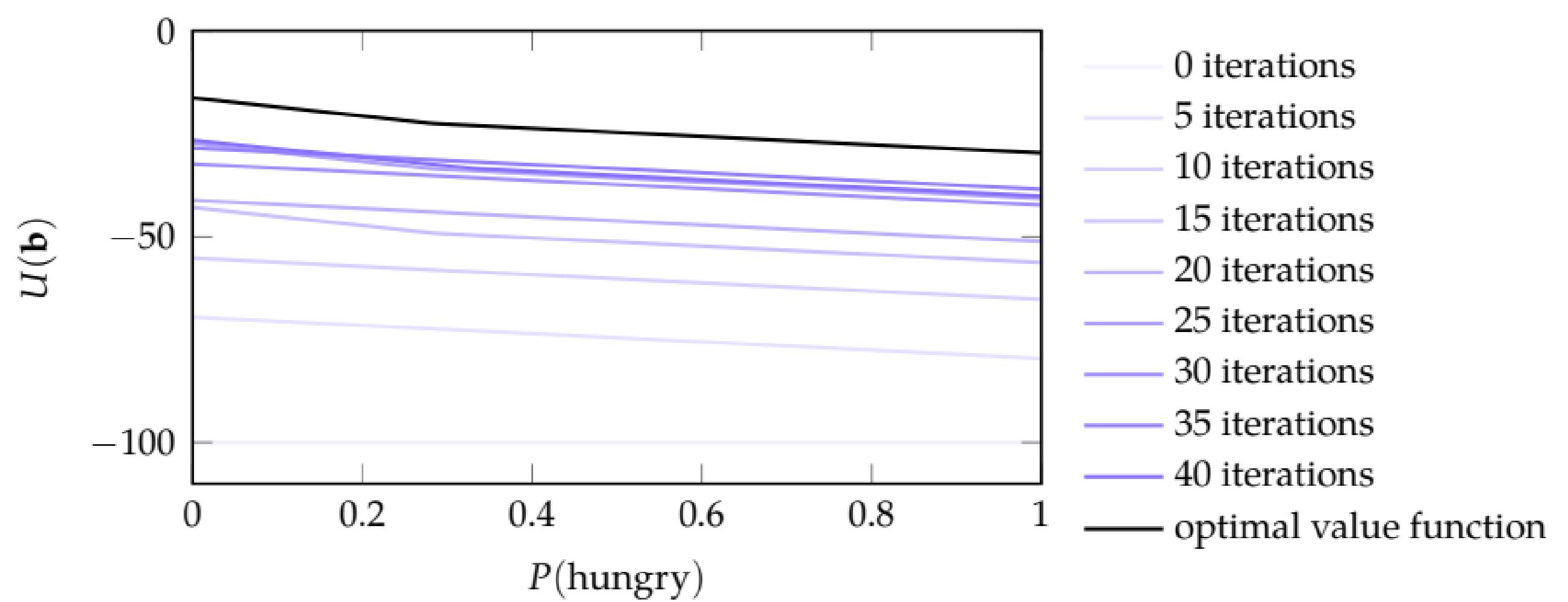




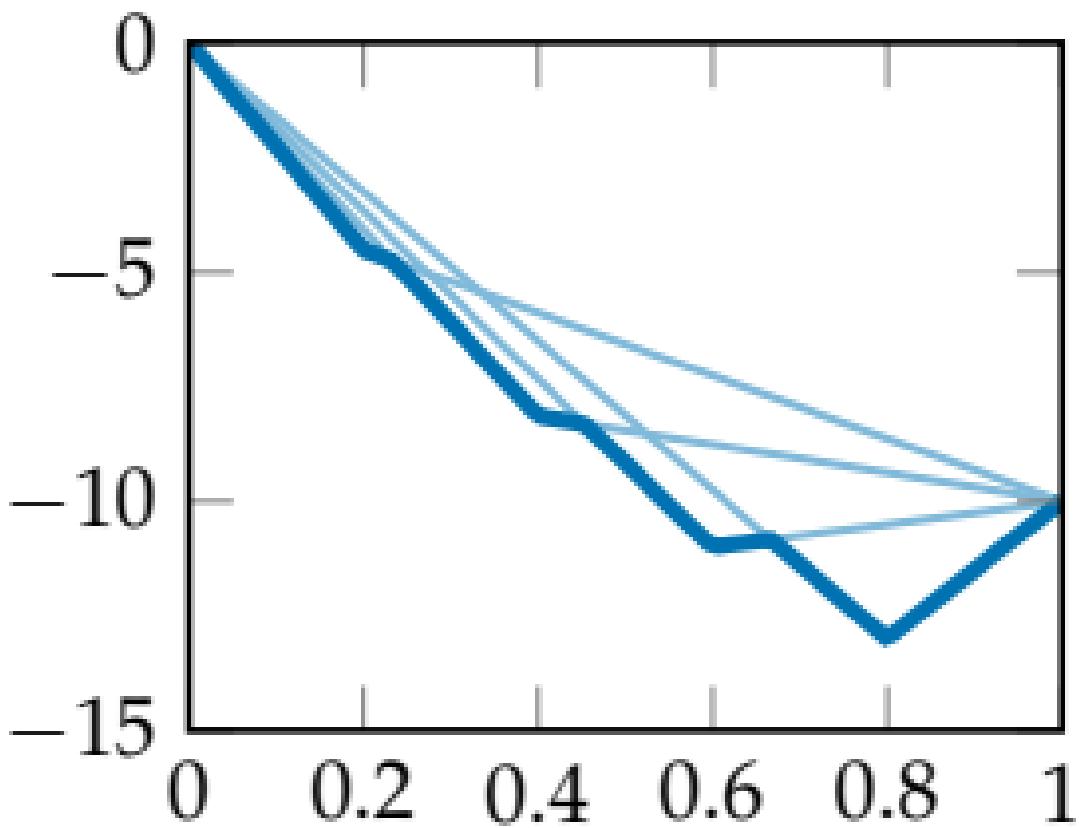
$P(\text{hungry})$

- blind 1 — blind 5
- blind 10 — blind 15
- blind 20 — optimal
- BAWS



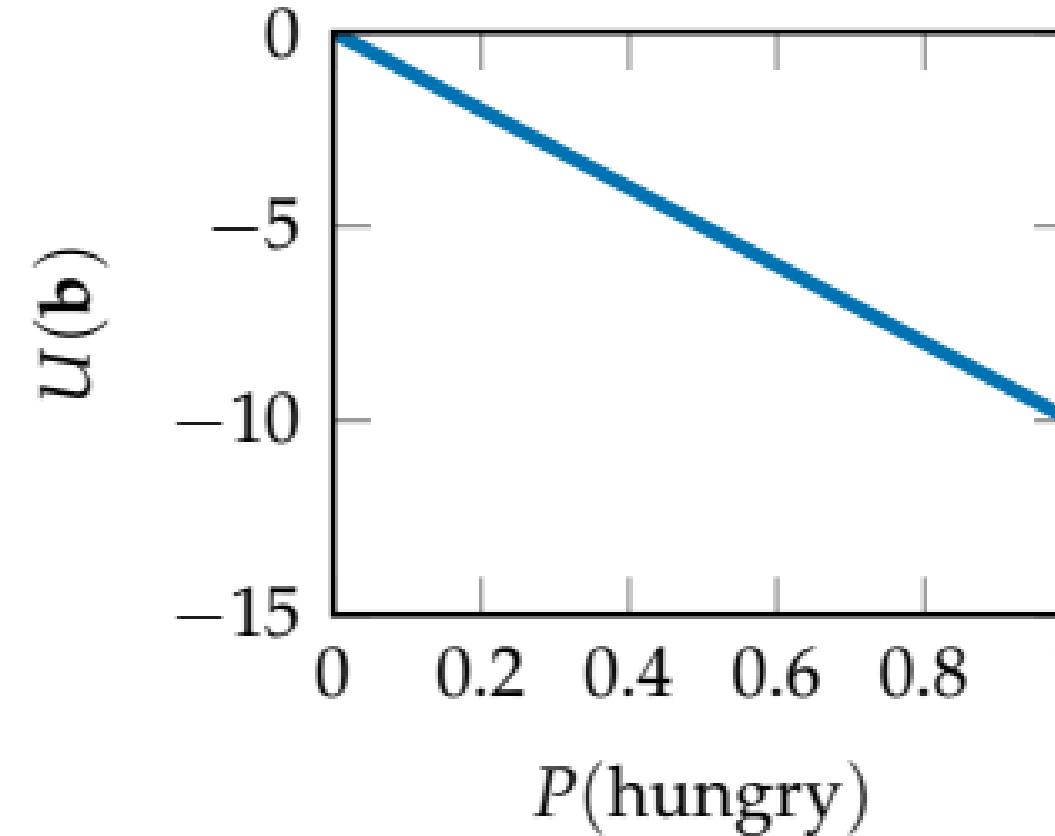


$U(\mathbf{b})$

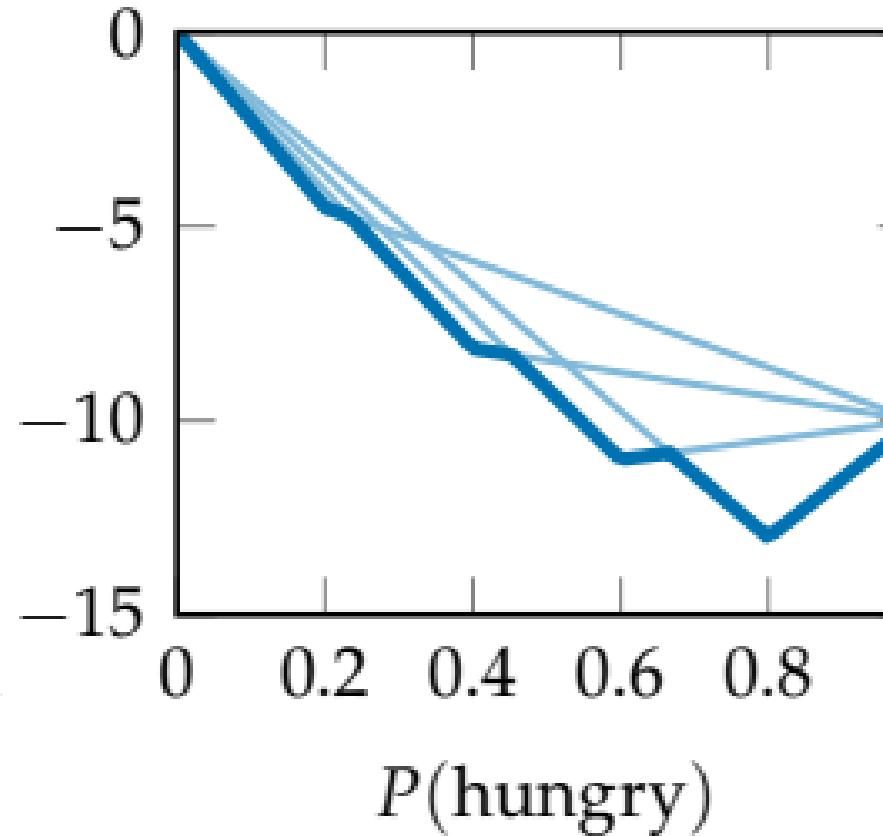


$P(\text{hungry})$

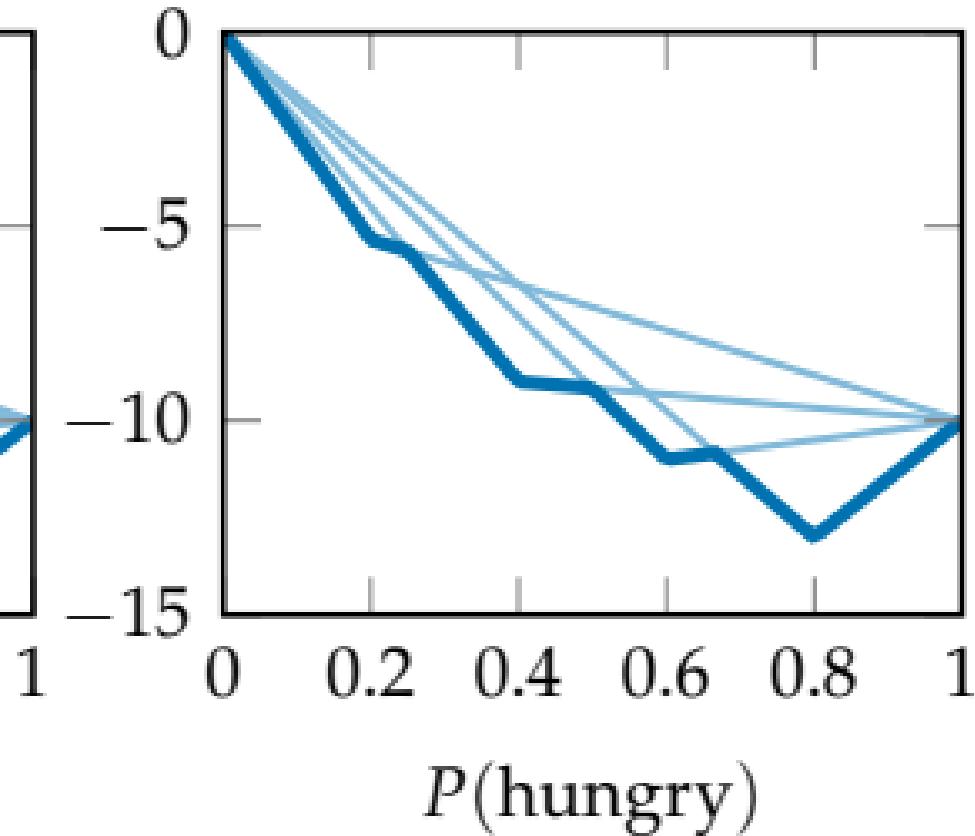
iteration 1 bound

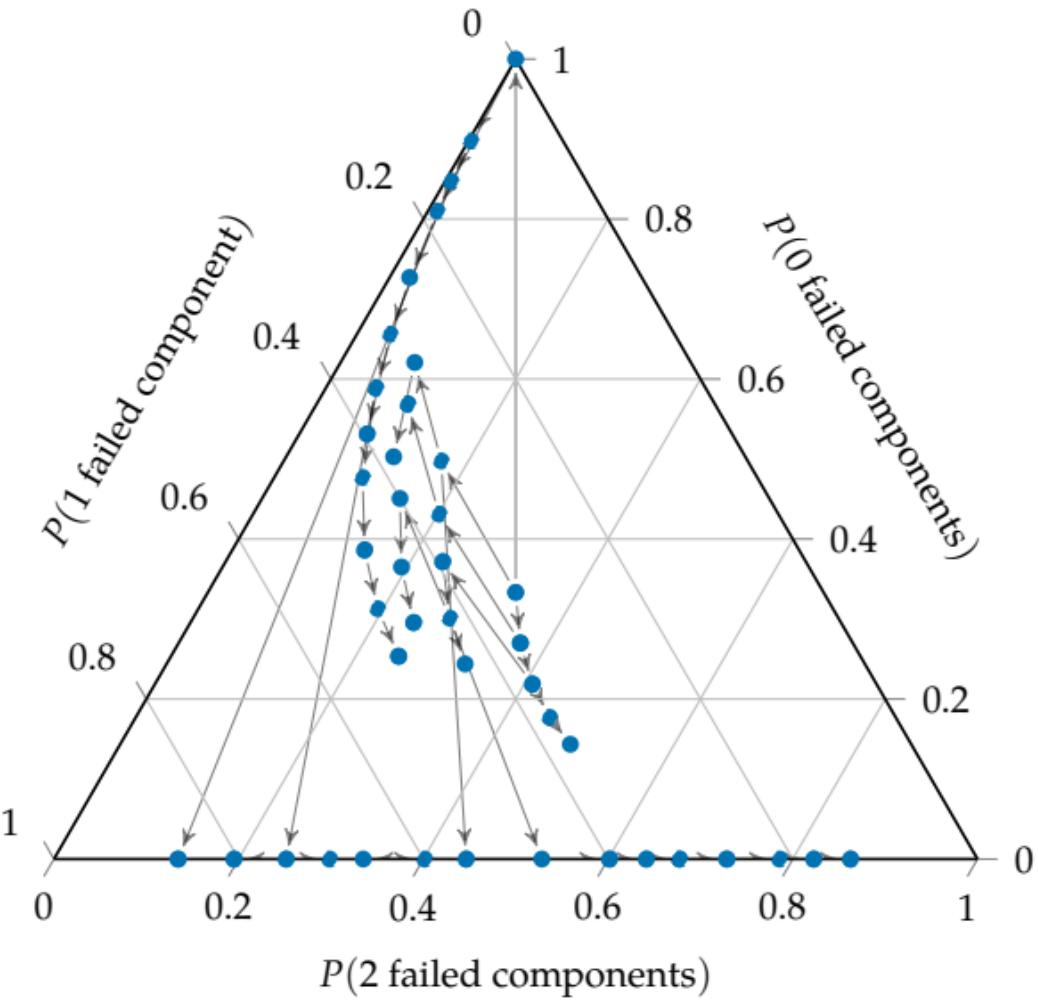


iteration 2 bound

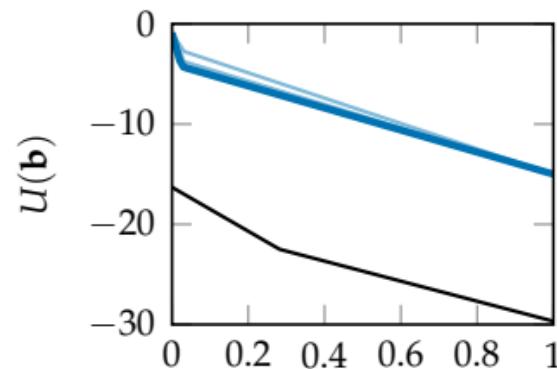


iteration 3 bound

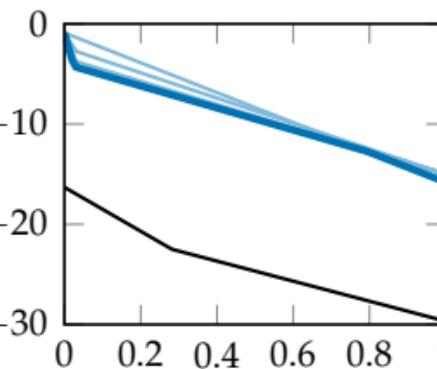




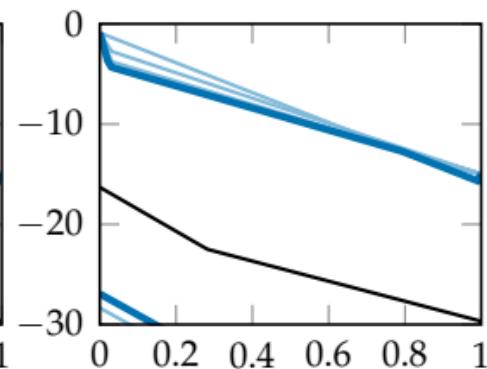
iteration 1



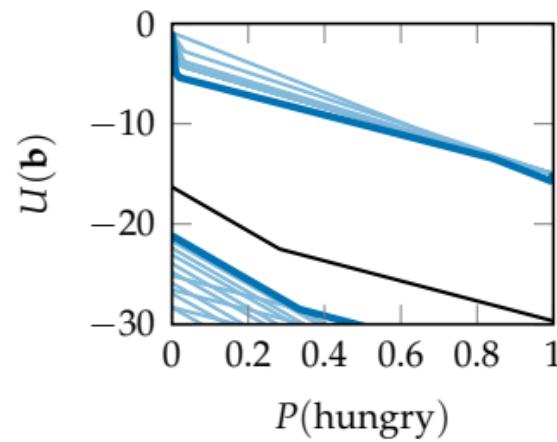
iteration 2



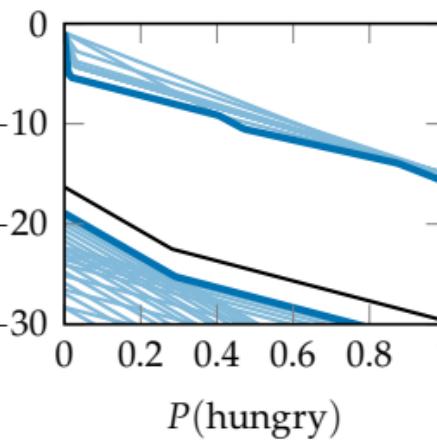
iteration 3



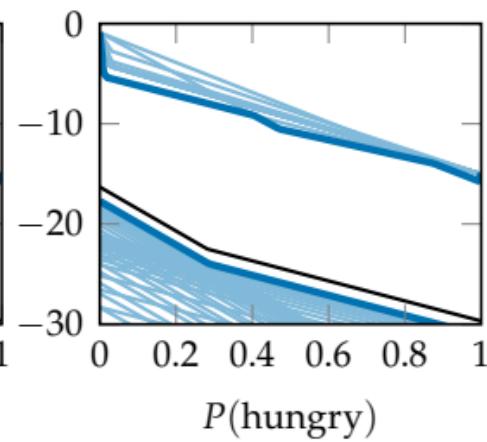
iteration 4

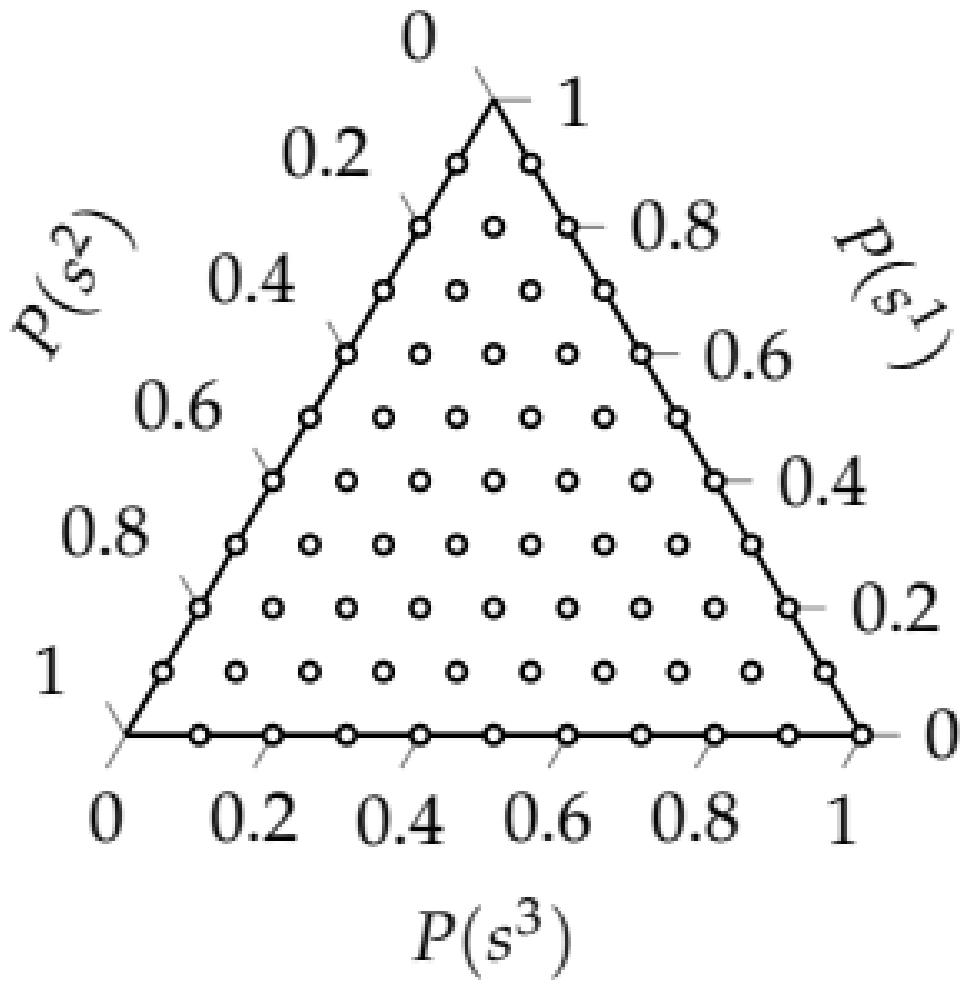


iteration 5

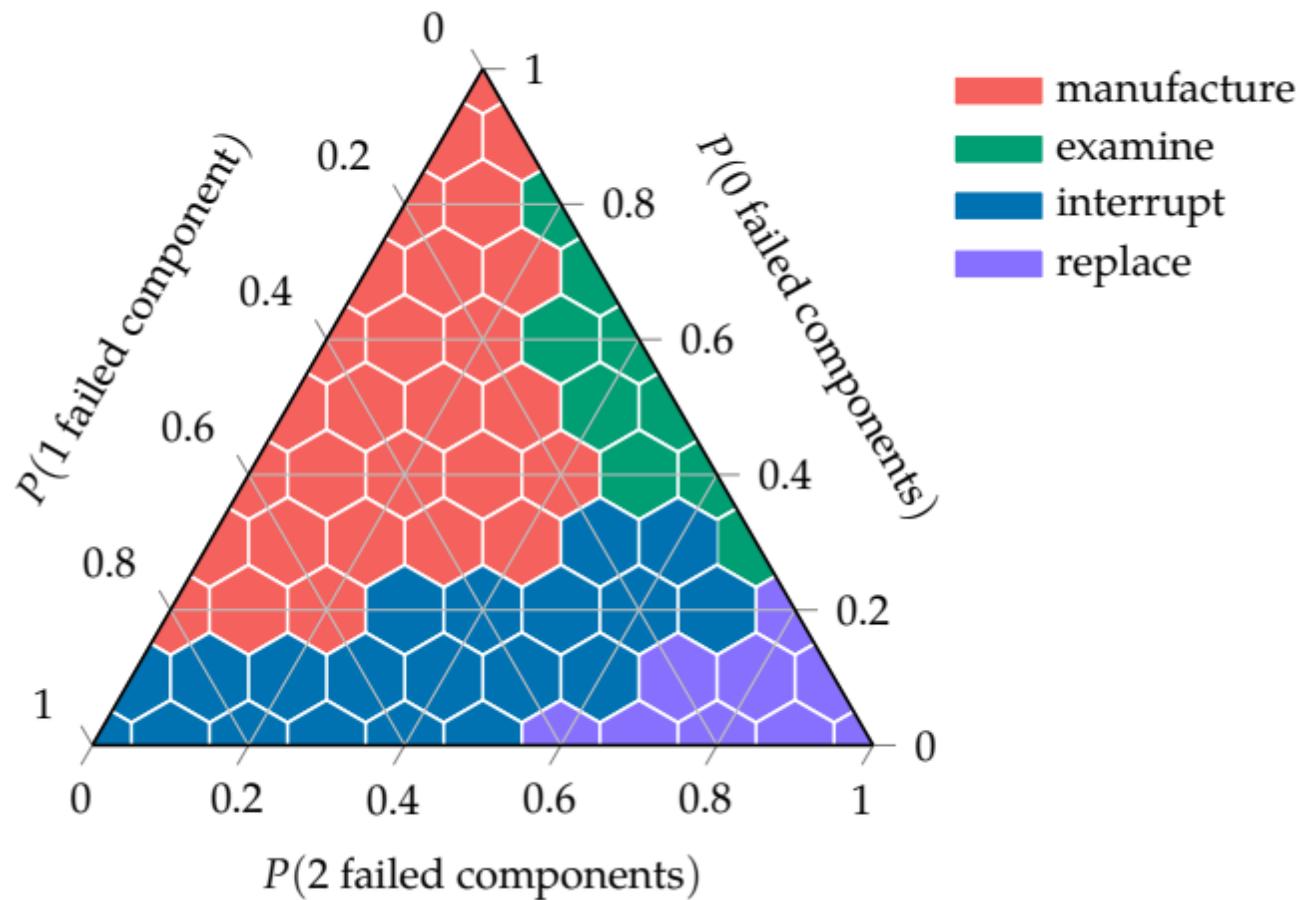


iteration 6

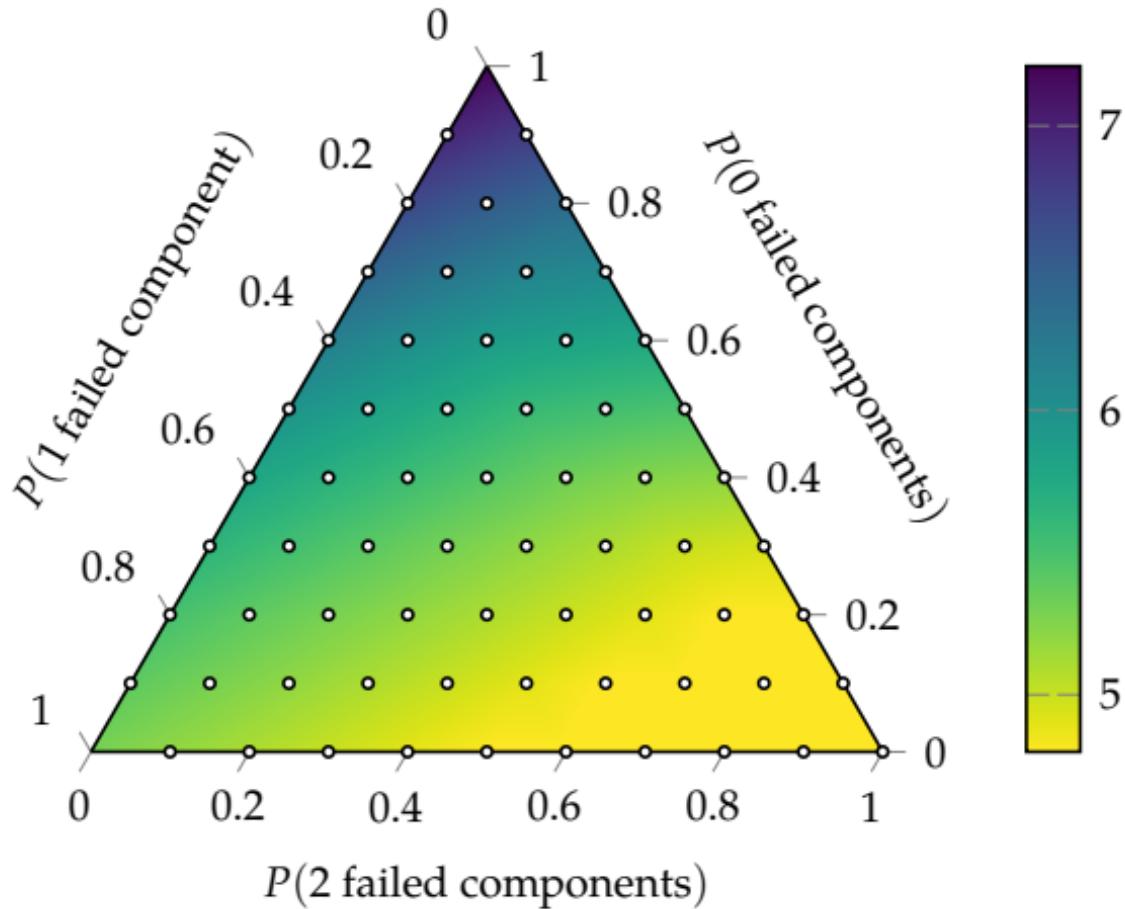




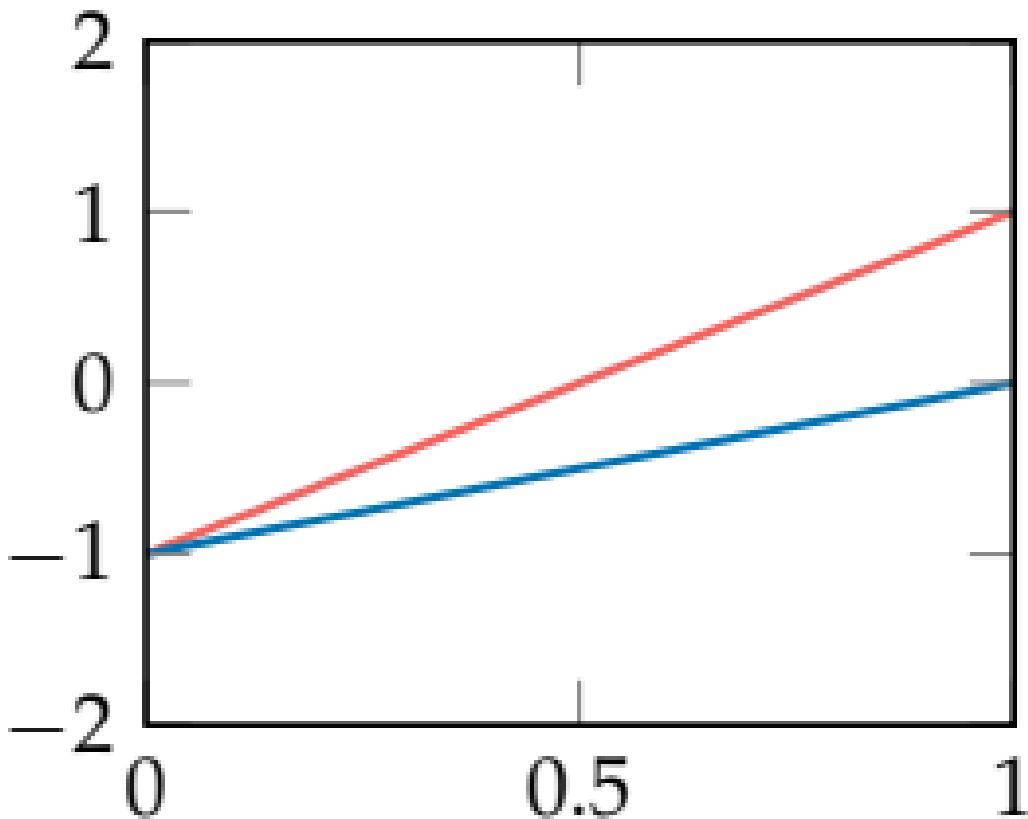
Policy



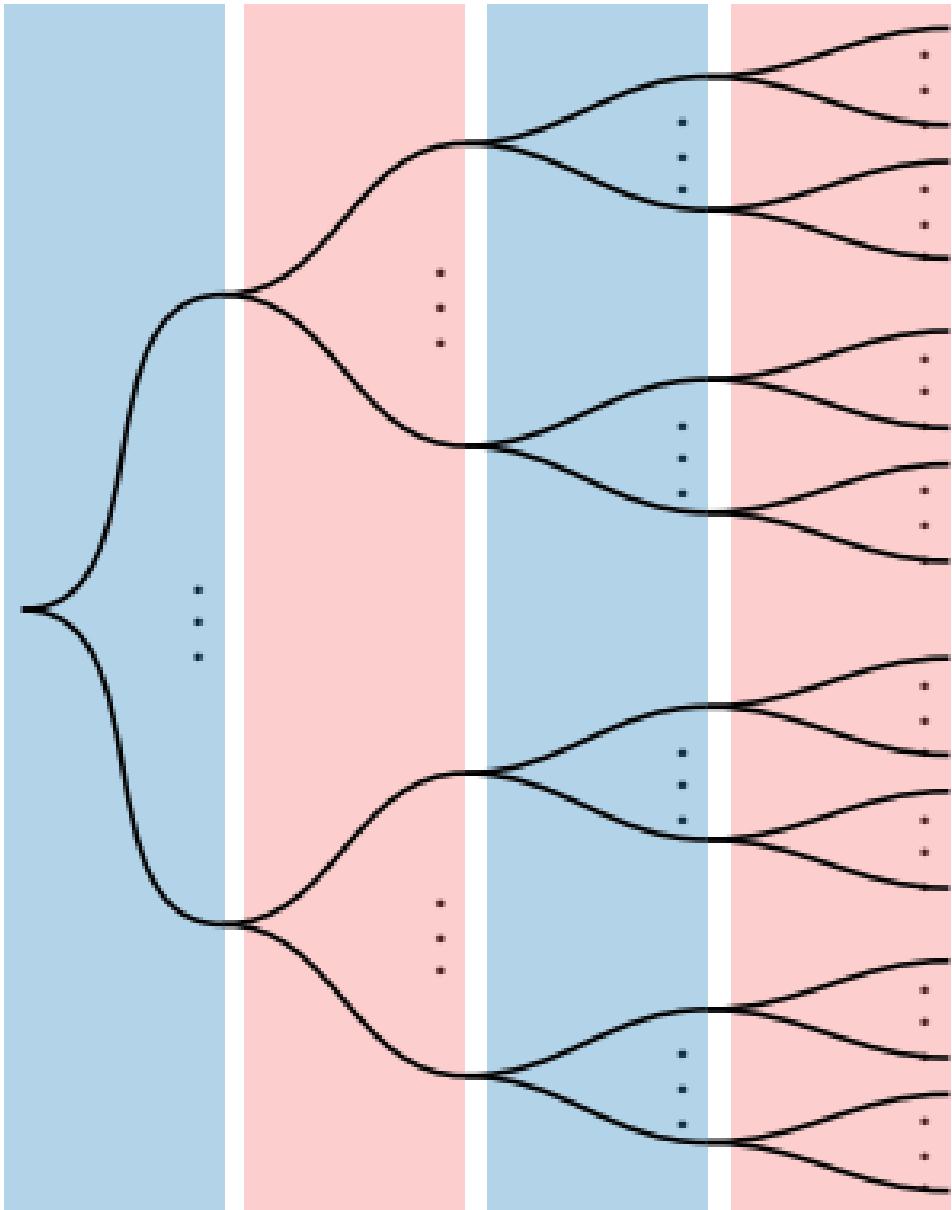
Value Function

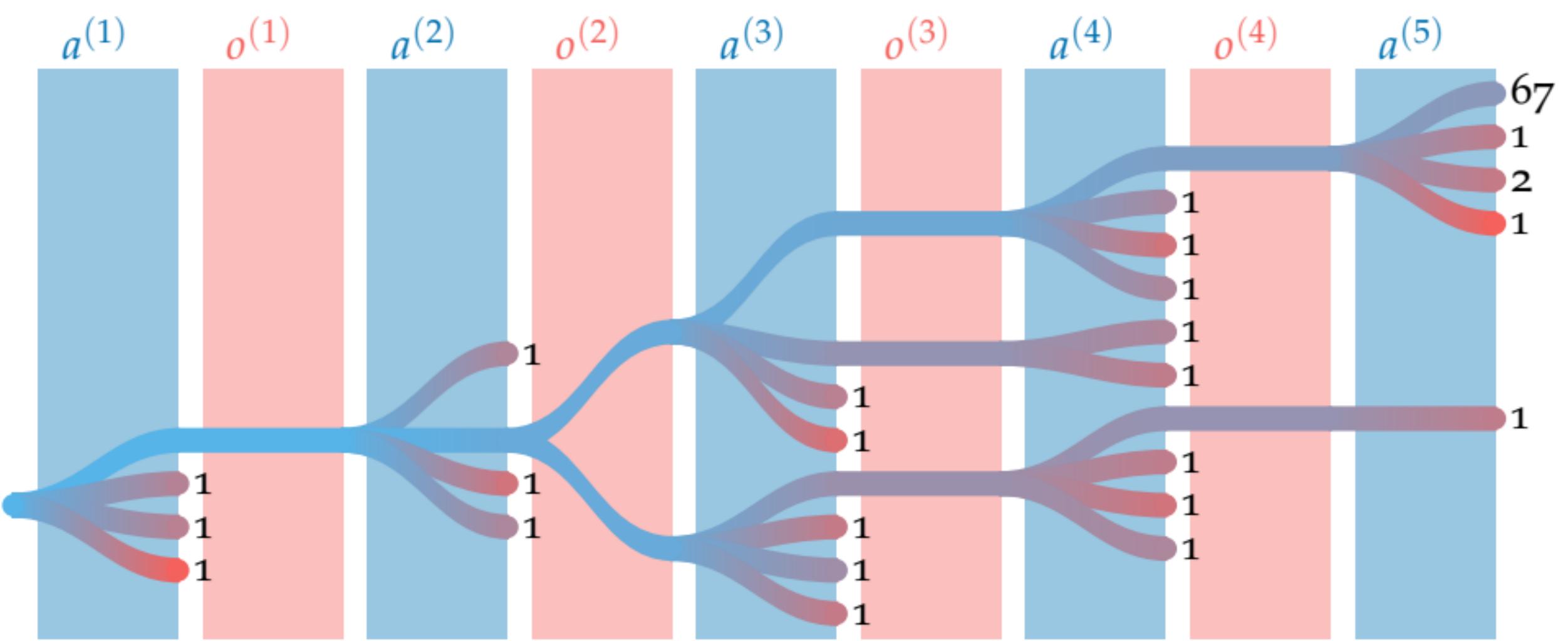


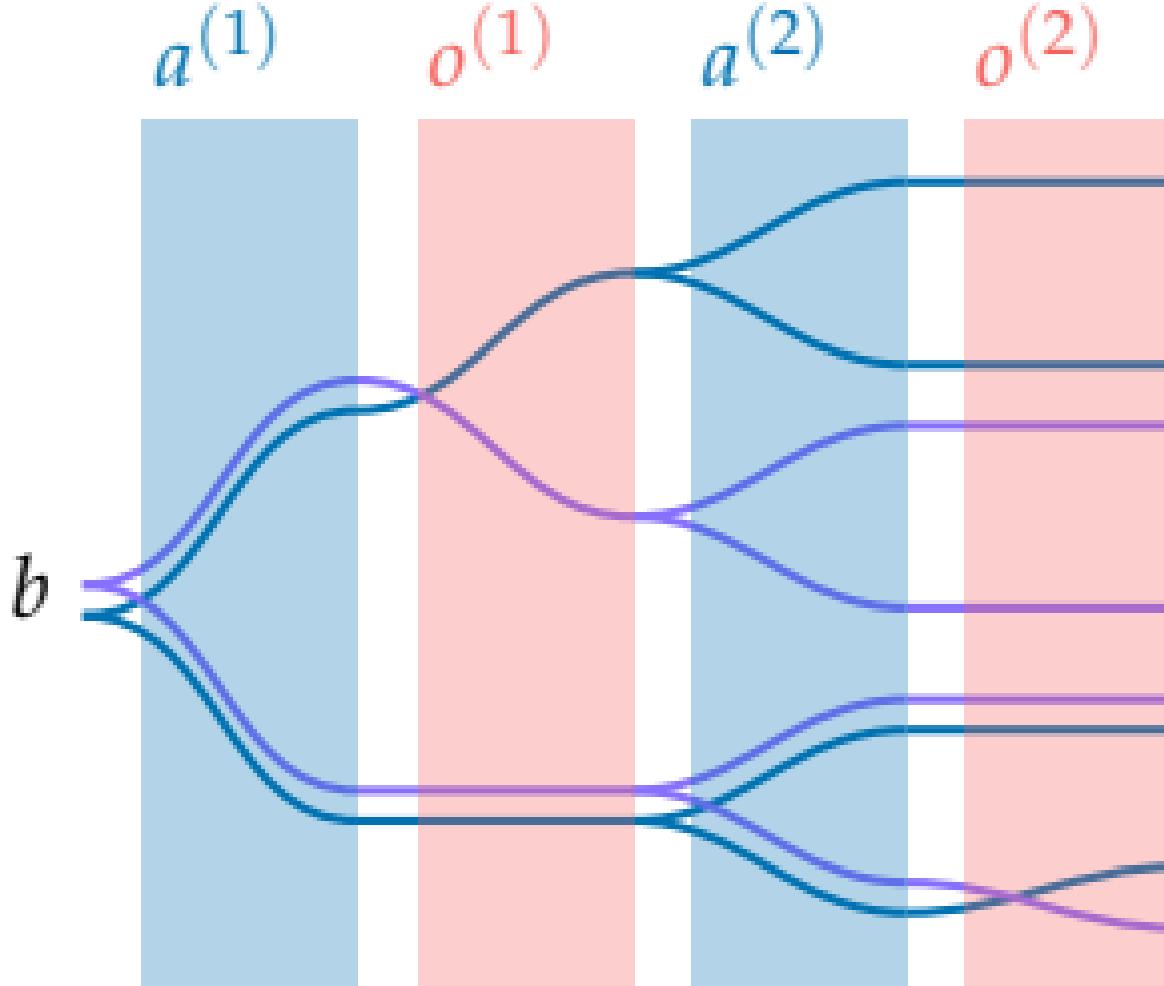
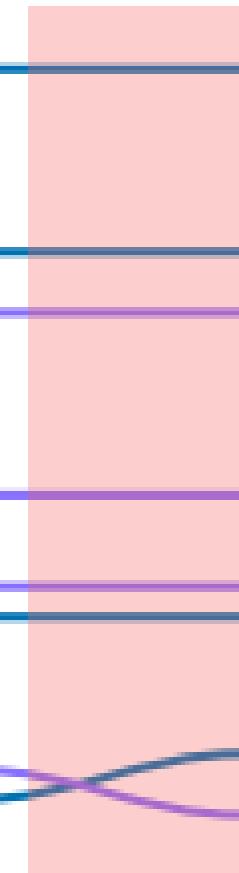
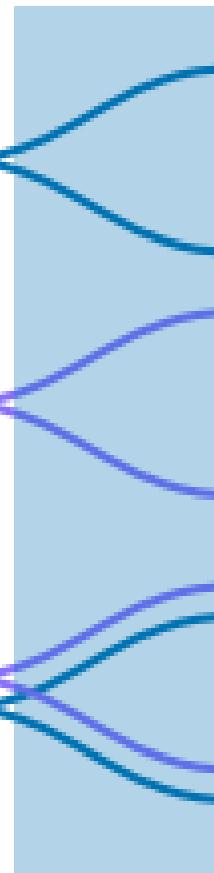
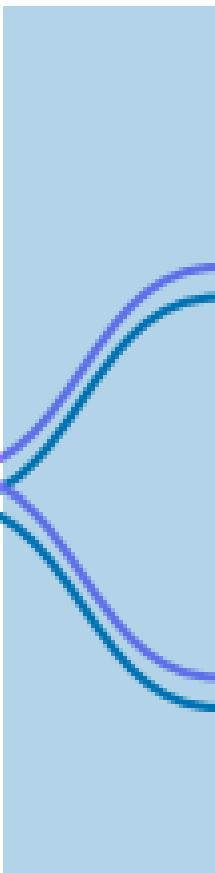
$U(\mathbf{b})$



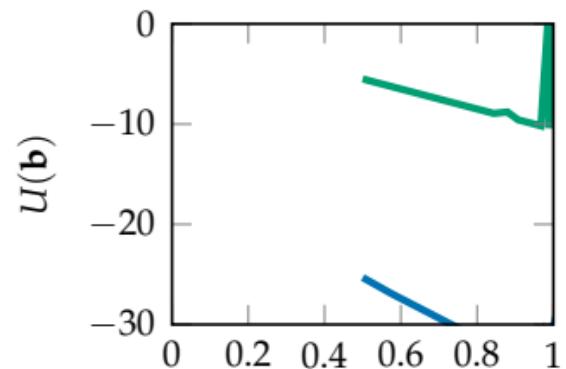
$P(s^1)$

$a^{(1)}$ $o^{(1)}$ $a^{(2)}$ $o^{(2)}$ b 

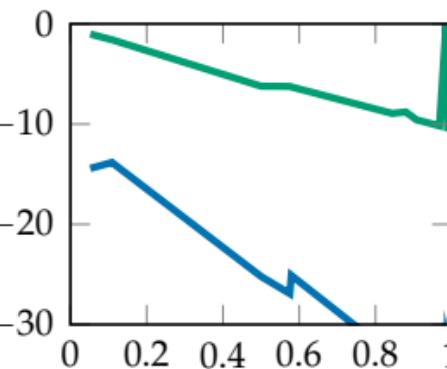


$a^{(1)}$ $o^{(1)}$ $a^{(2)}$ $o^{(2)}$ b 

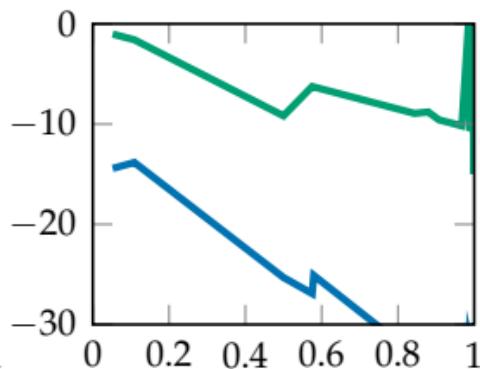
iteration 1



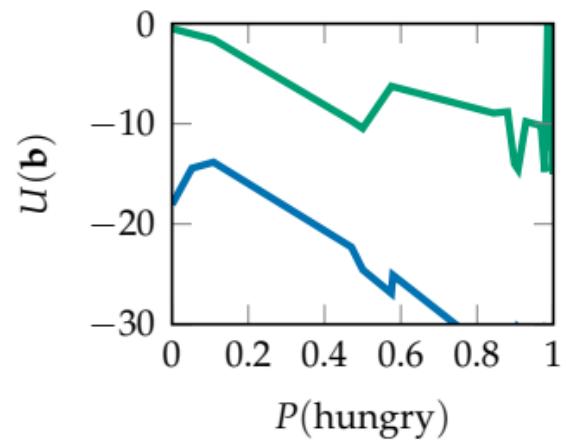
iteration 2



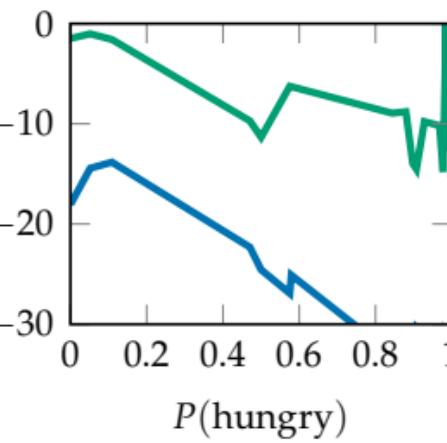
iteration 3



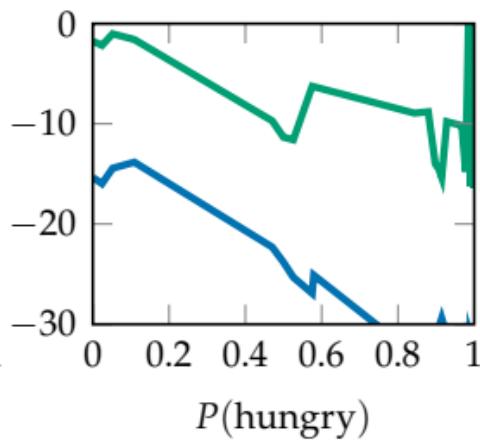
iteration 4

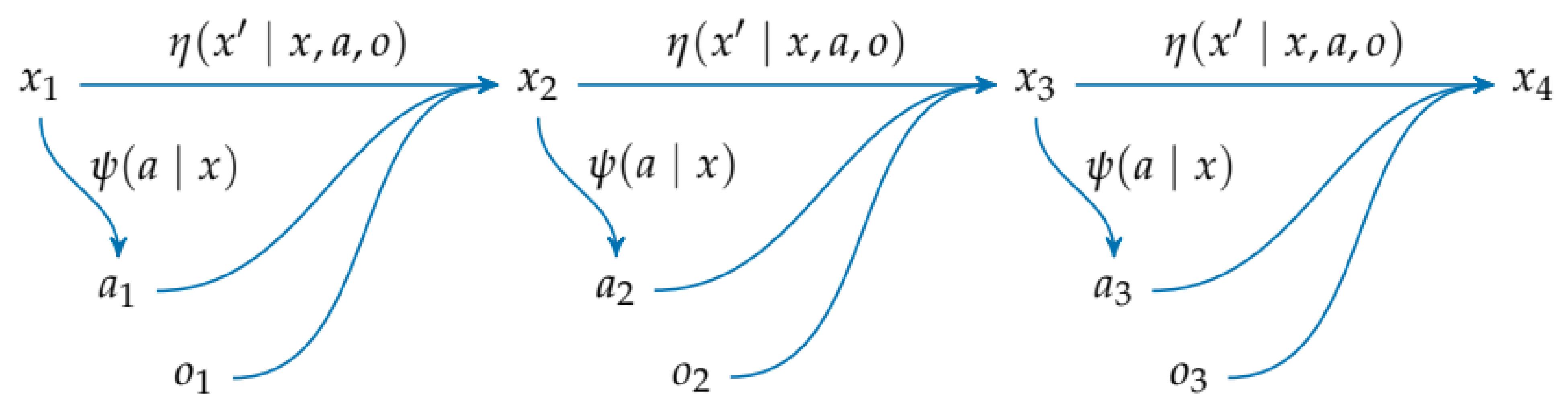


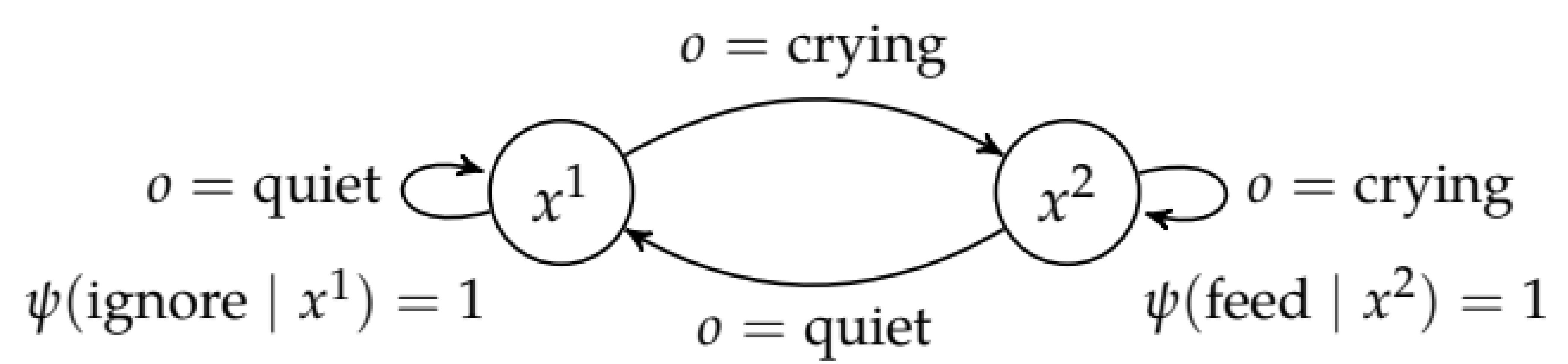
iteration 5

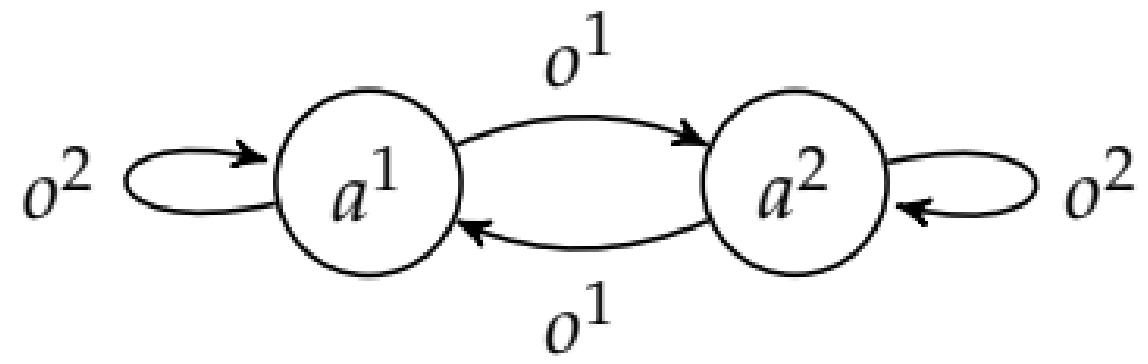
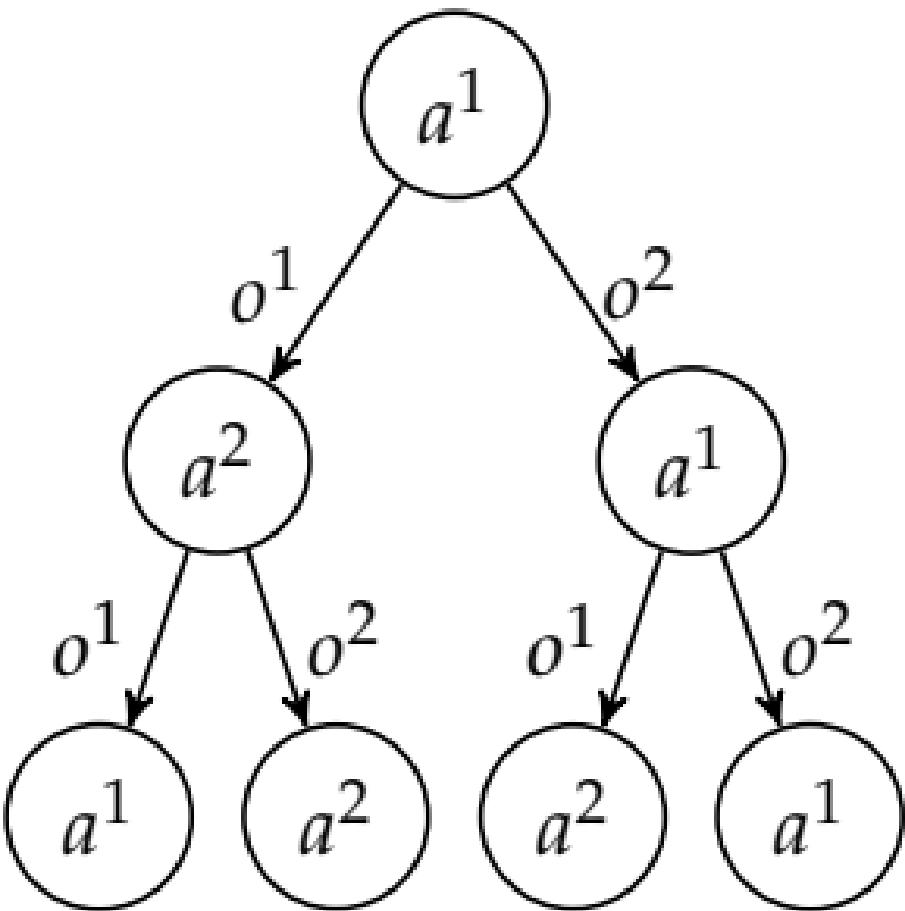


iteration 6







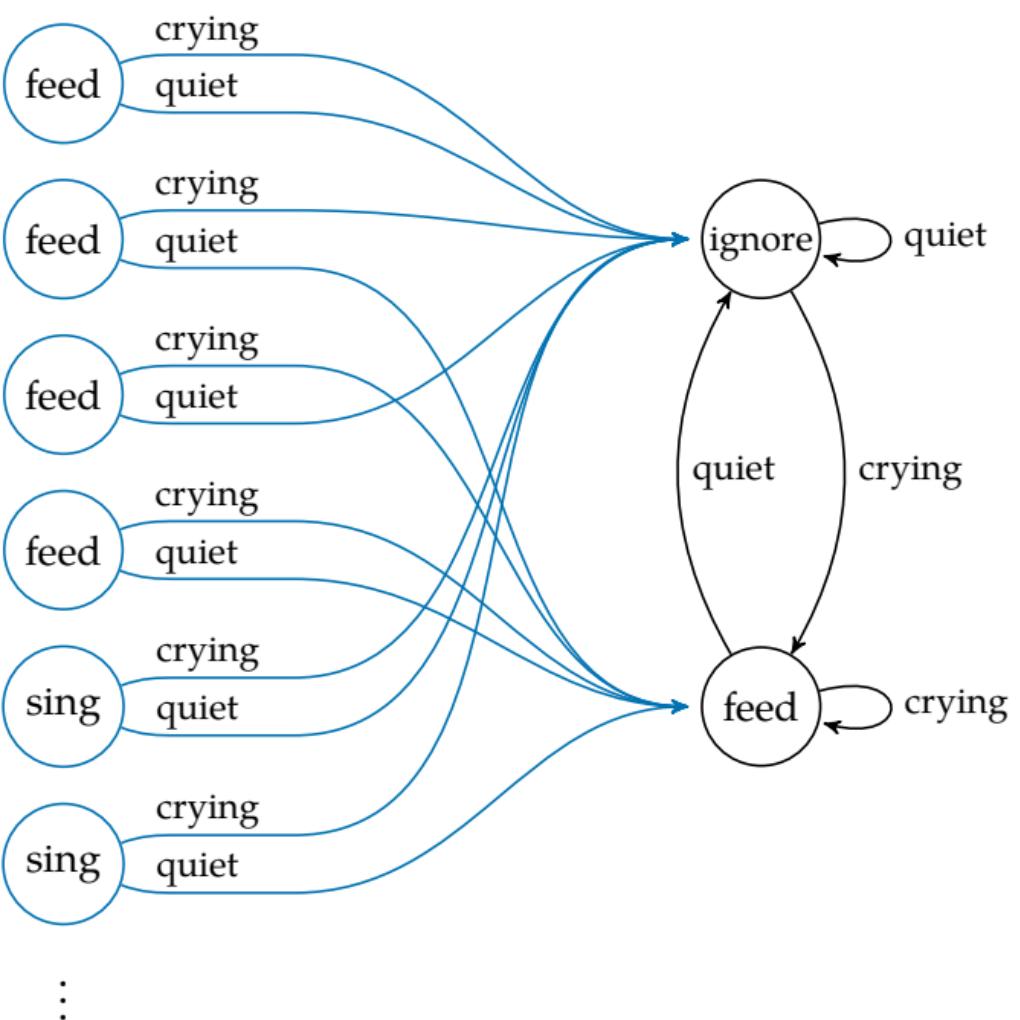


o^1 = quiet

o^2 = crying

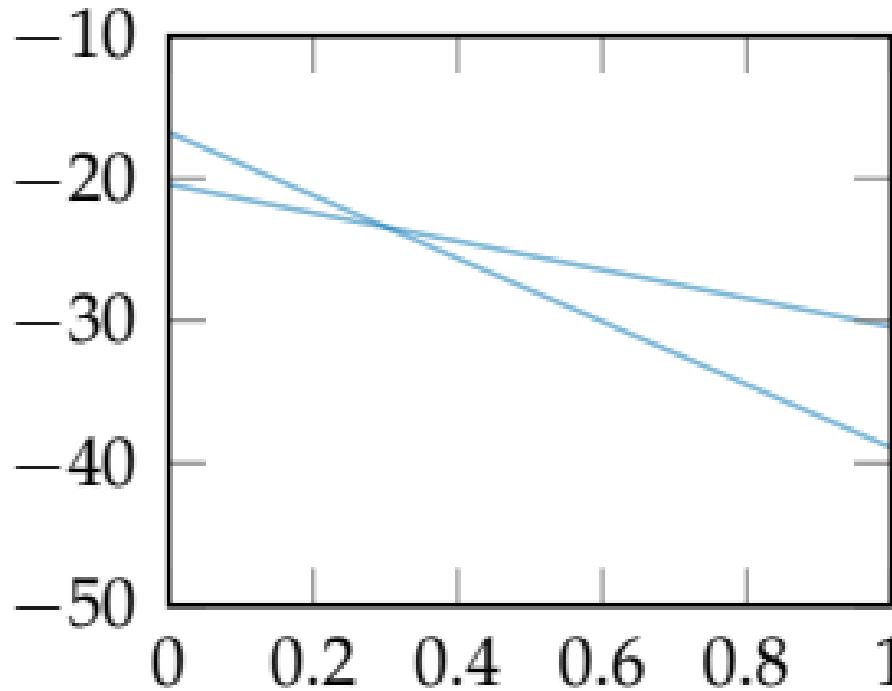
a^1 = ignore

a^2 = feed

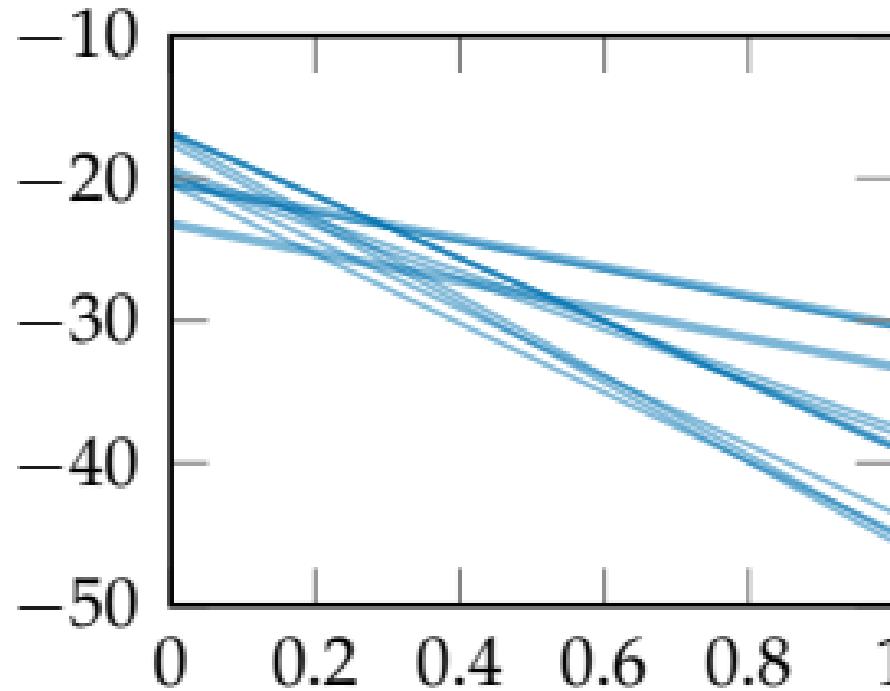


policy evaluation 1

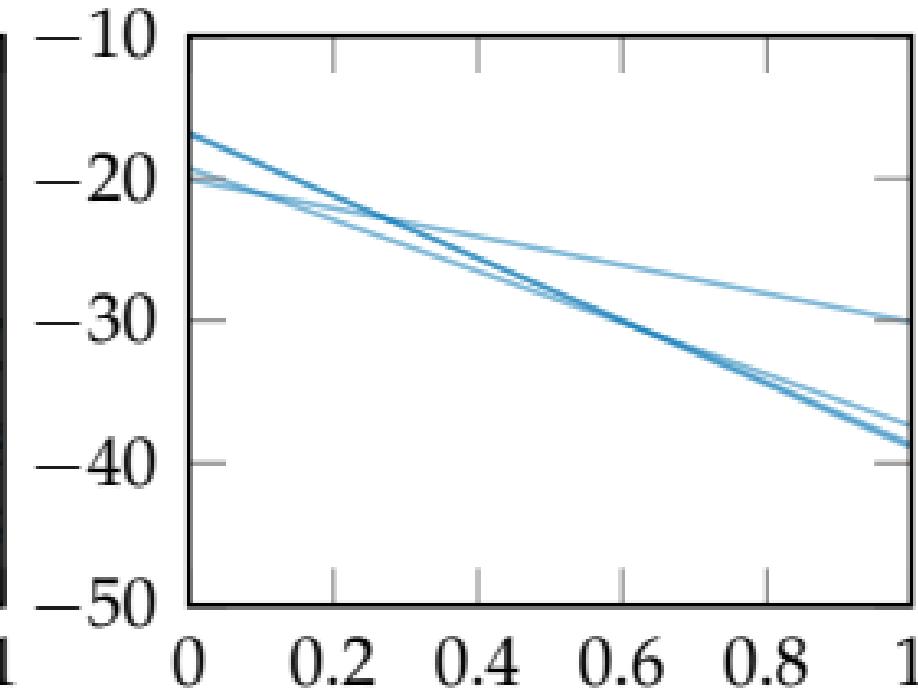
$U(x, s)$



policy improvement 1



pruning 1



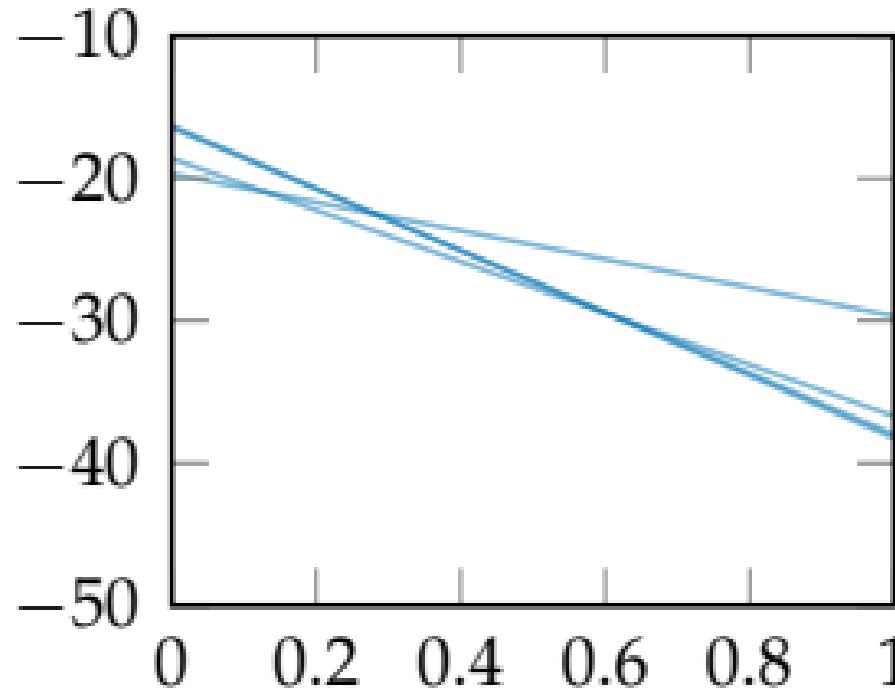
$P(\text{hungry})$

$P(\text{hungry})$

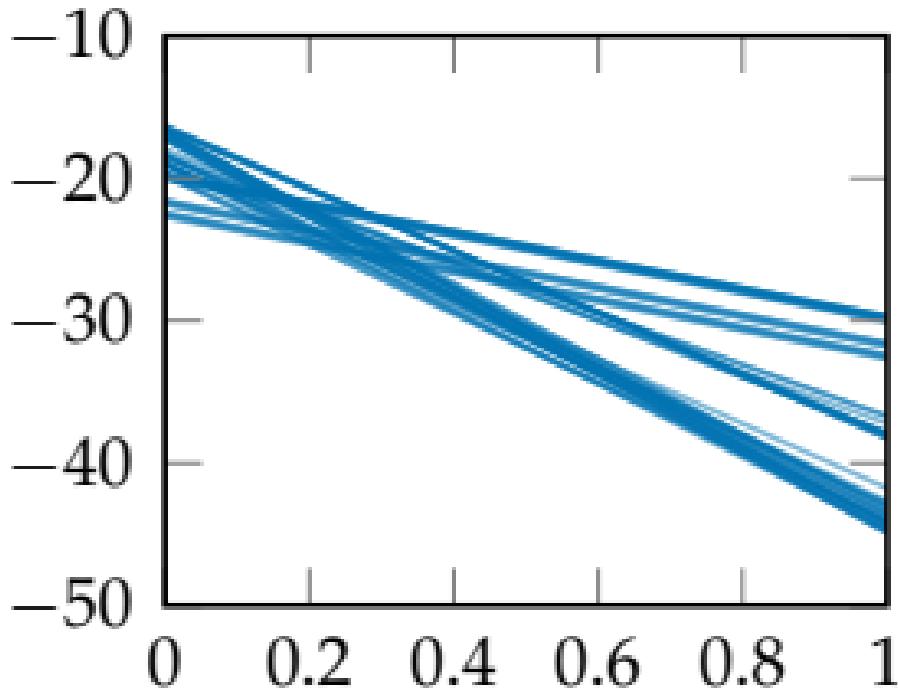
$P(\text{hungry})$

policy evaluation 2

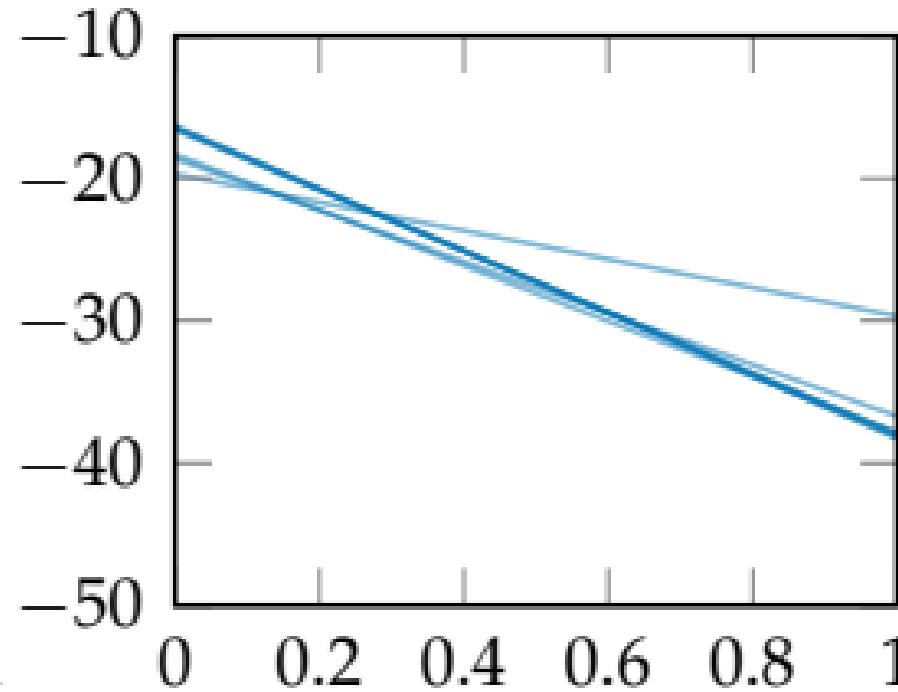
$U(x, s)$



policy improvement 2



pruning 2

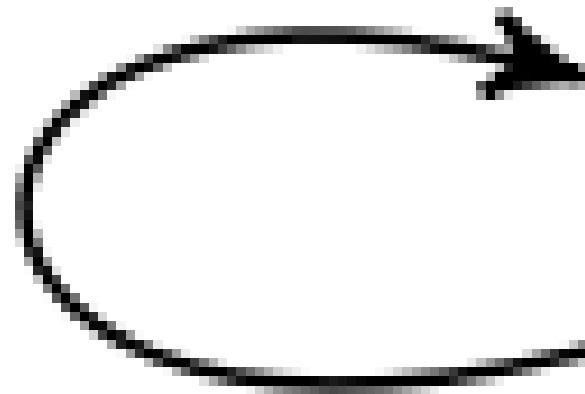


$P(\text{hungry})$

$P(\text{hungry})$

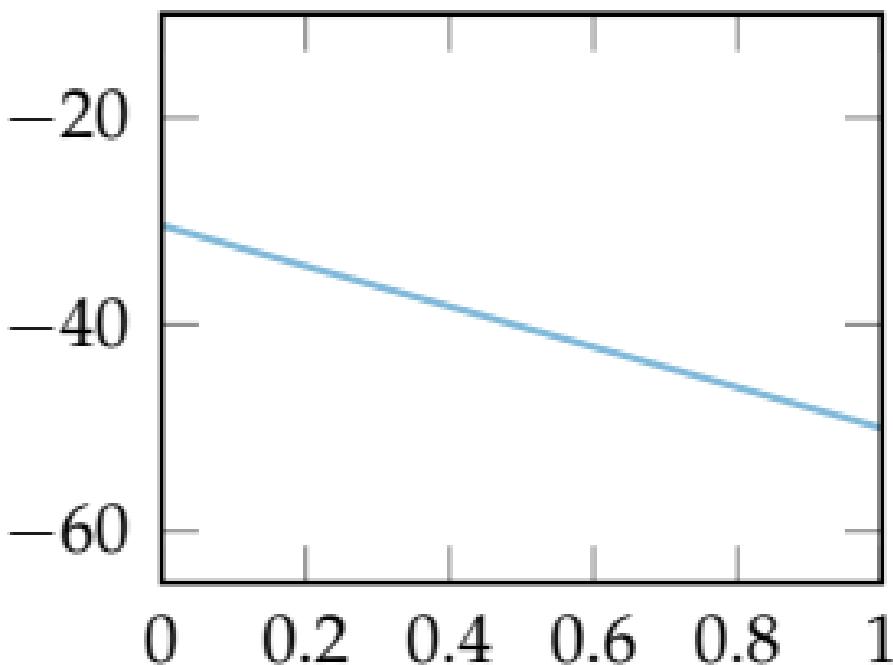
$P(\text{hungry})$

crying, quiet

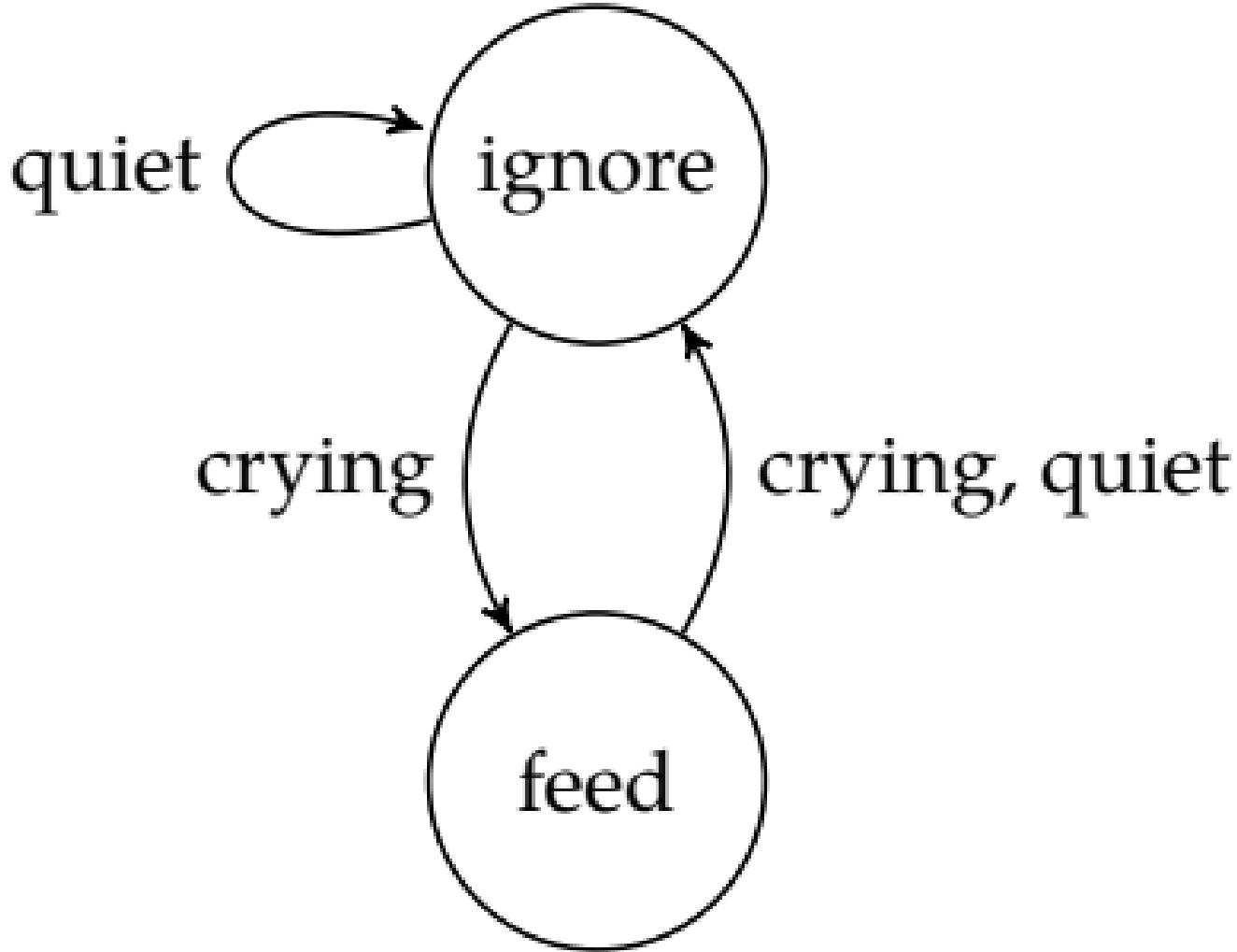


ignore

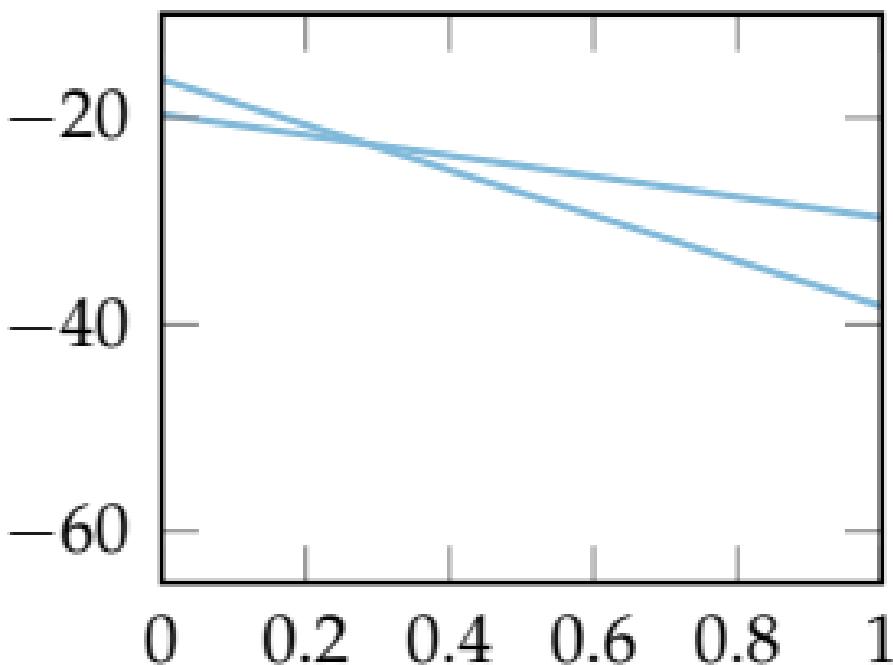
controller utility ($k = 1$)



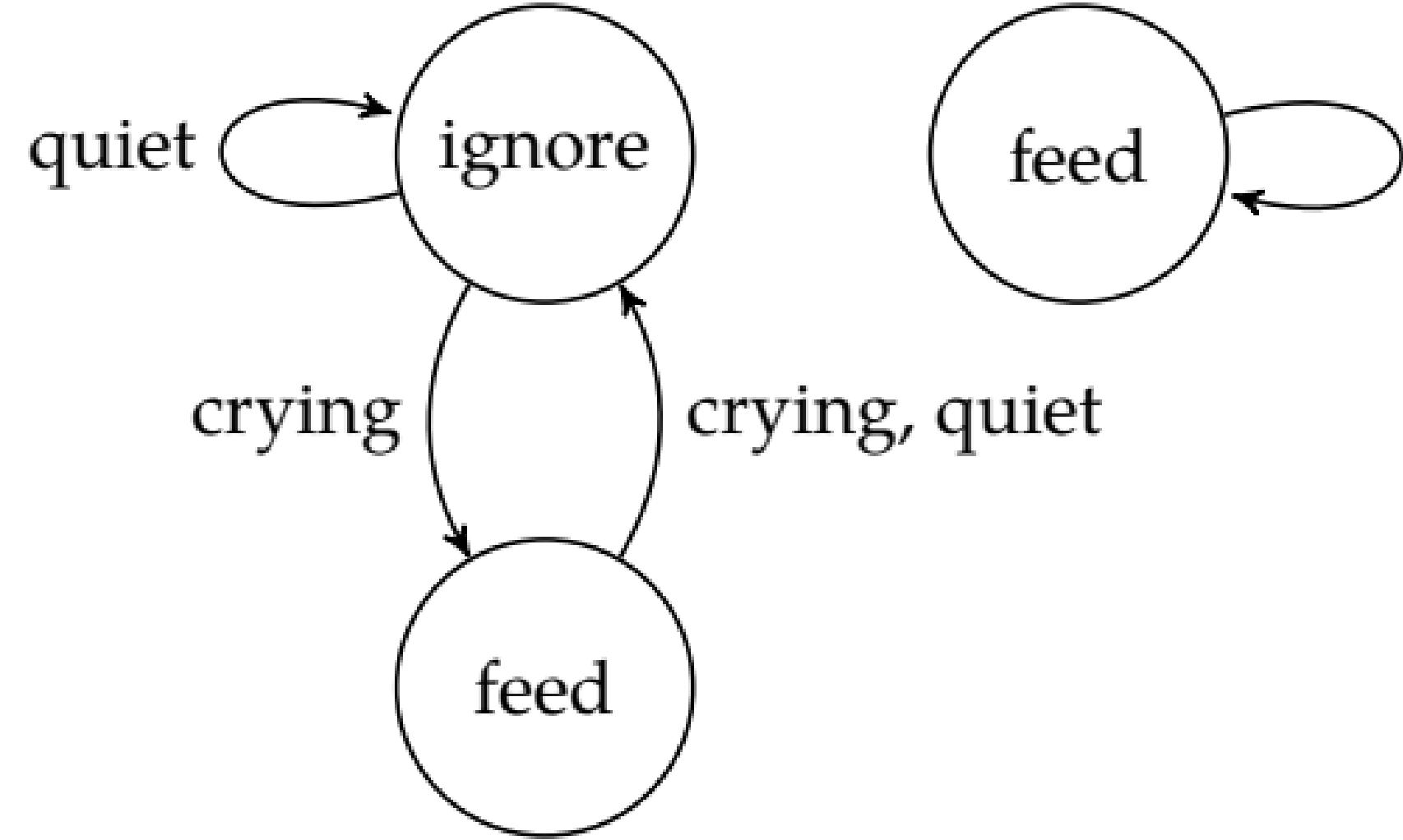
$P(\text{hungry})$



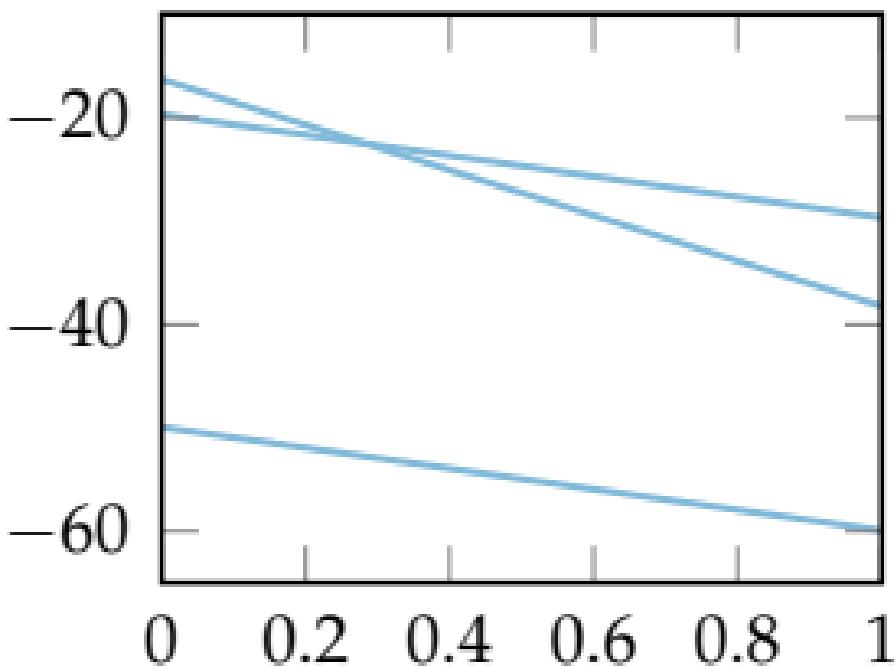
controller utility ($k = 2$)



$P(\text{hungry})$



controller utility ($k = 3$)

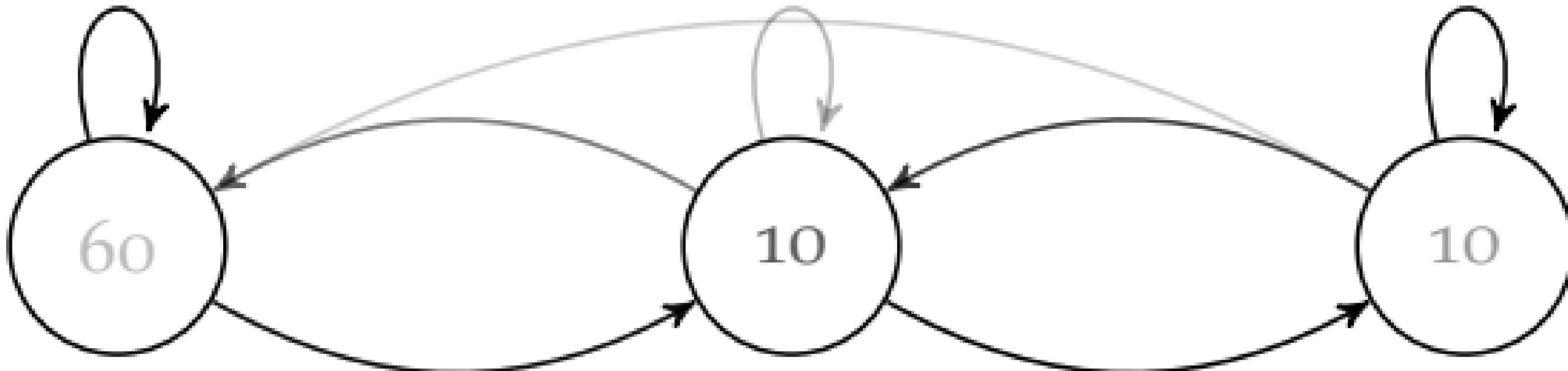


$P(\text{hungry})$

catch, drop

drop

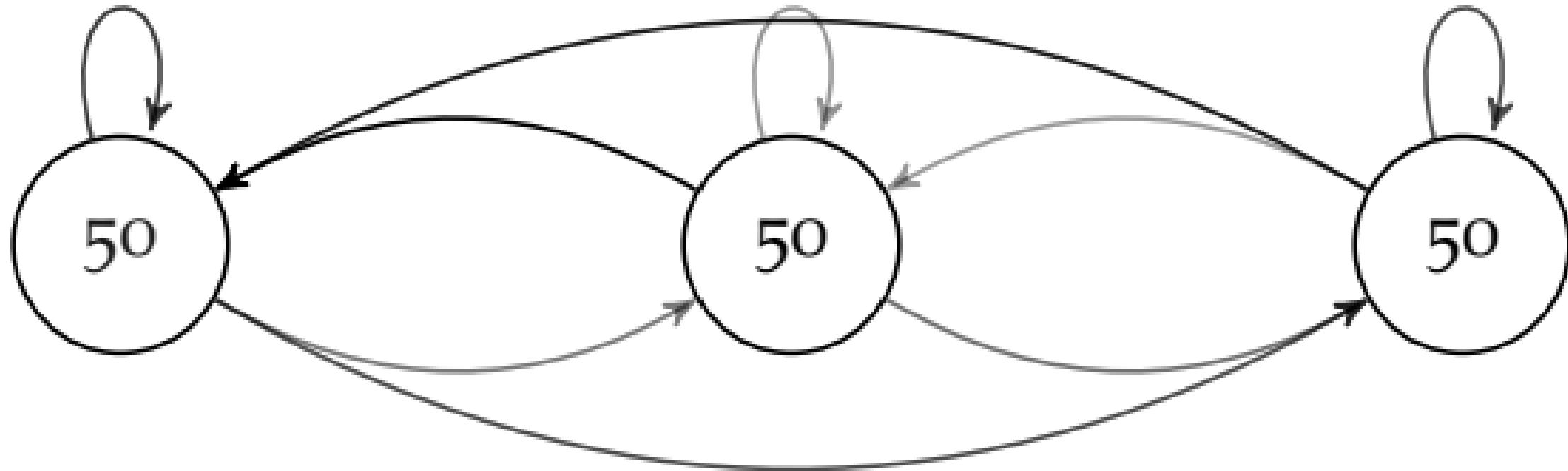
catch, drop



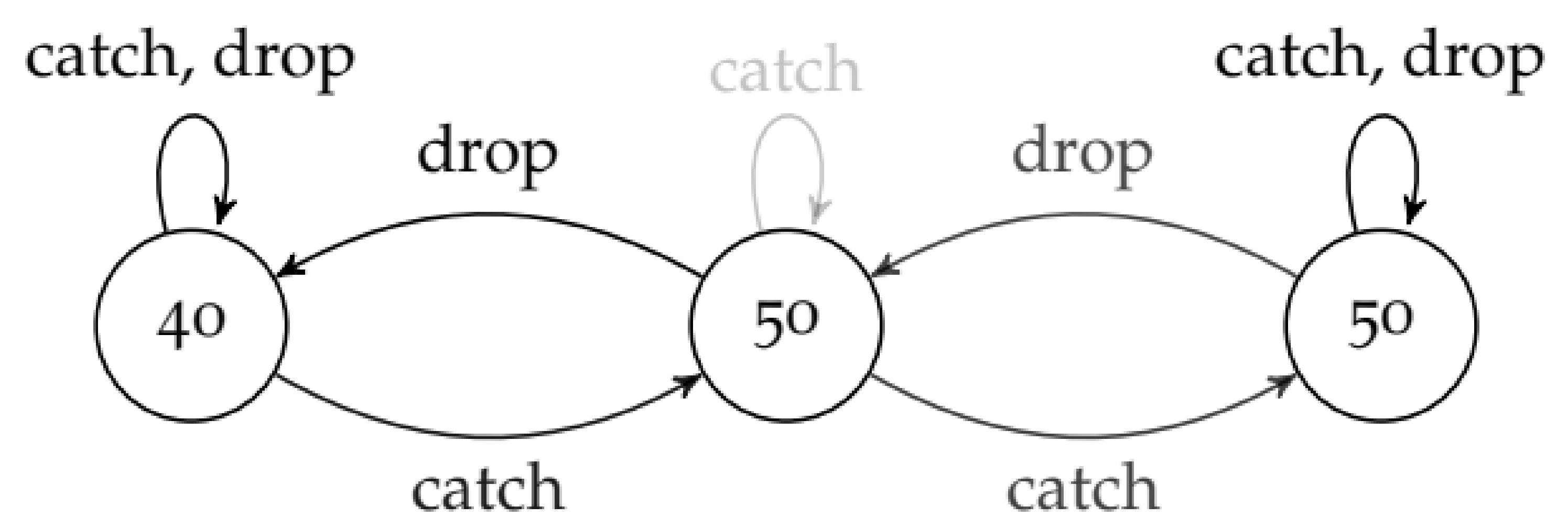
catch, drop

catch, drop

catch, drop



catch, drop



agent 2

cooperate defect

agent 1

defect

-1, -1

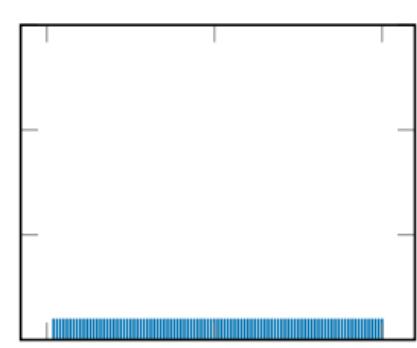
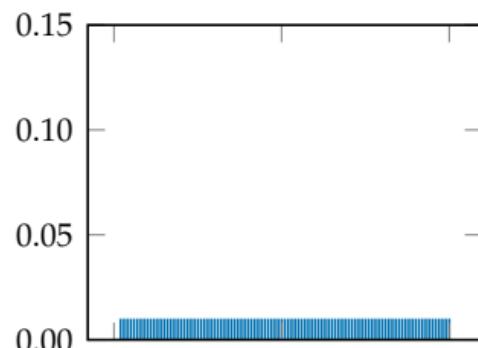
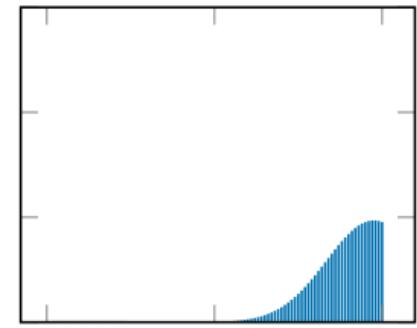
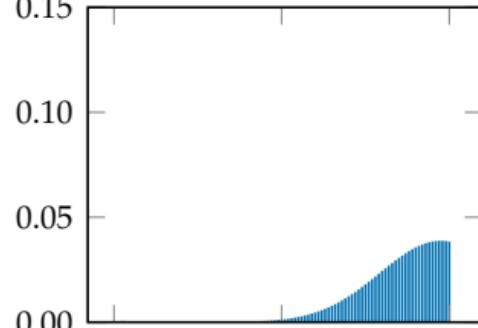
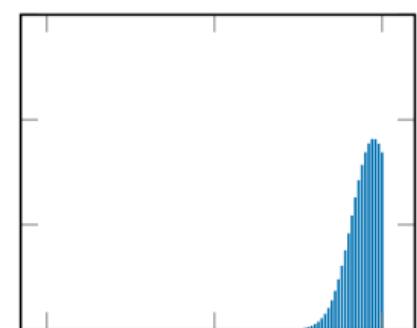
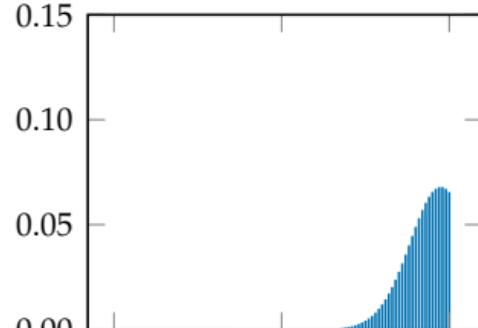
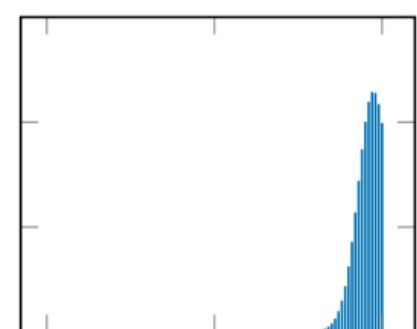
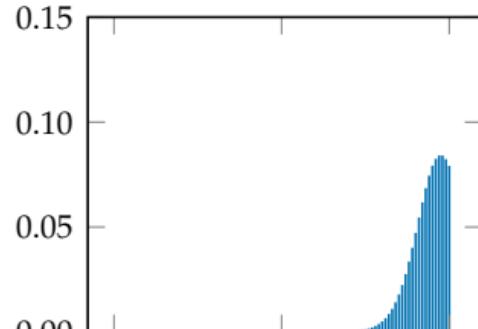
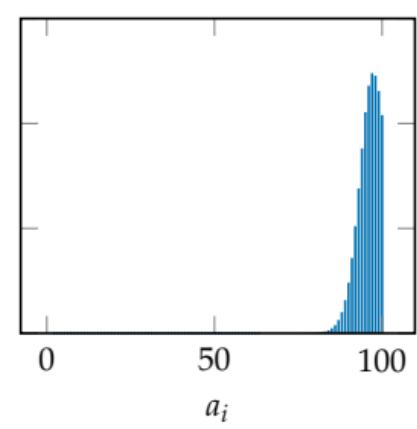
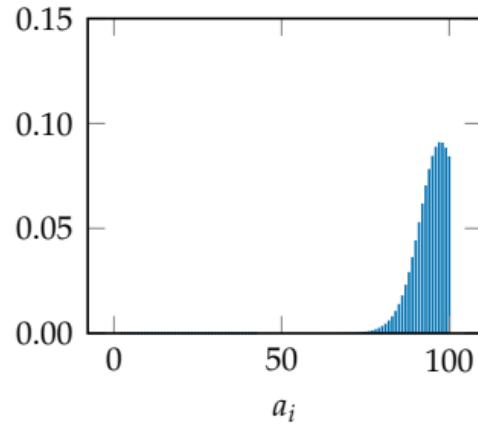
-4, 0

0, -4

-3, -3

		agent 2		
		rock	paper	scissors
agent 1	rock	0, 0	-1, 1	1, -1
	paper	1, -1	0, 0	-1, 1
	scissors	-1, 1	1, -1	0, 0

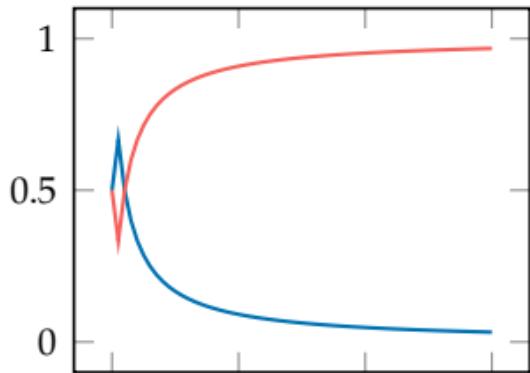
		agent 2		
		rock	paper	scissors
agent 1	rock	0, 0	-1, 1	1, -1
	paper	1, -1	0, 0	-1, 1
	scissors	-1, 1	1, -1	0, 0

$\lambda = 0.3$ $\lambda = 0.5$ $k = 0$
 $P(a_i)$  $k = 1$
 $P(a_i)$  $k = 2$
 $P(a_i)$  $k = 3$
 $P(a_i)$  $k = 4$
 $P(a_i)$ 

opponent model

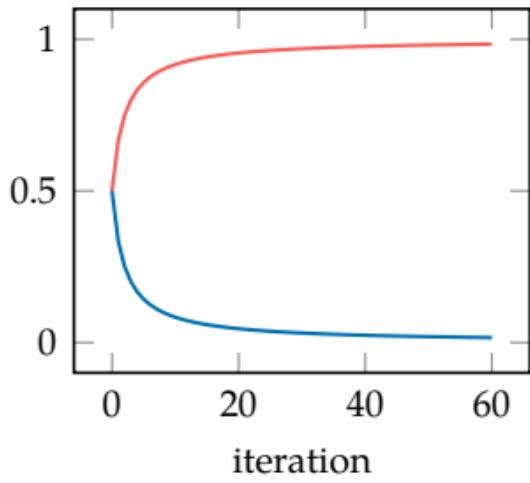
policy

agent 1
 $P(\text{action})$



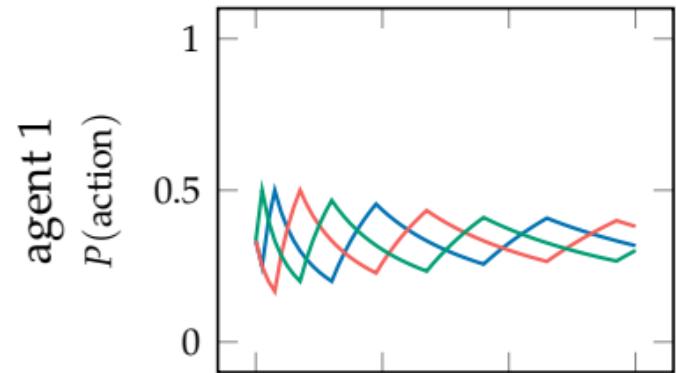
cooperate
defect

agent 2
 $P(\text{action})$

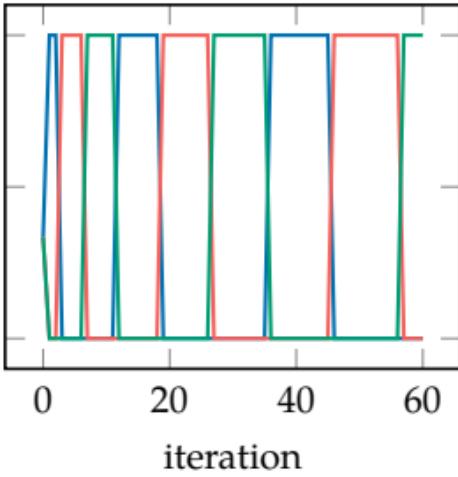
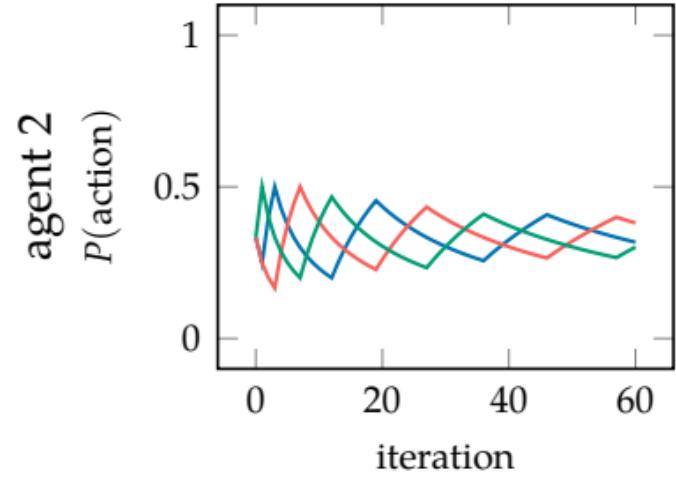
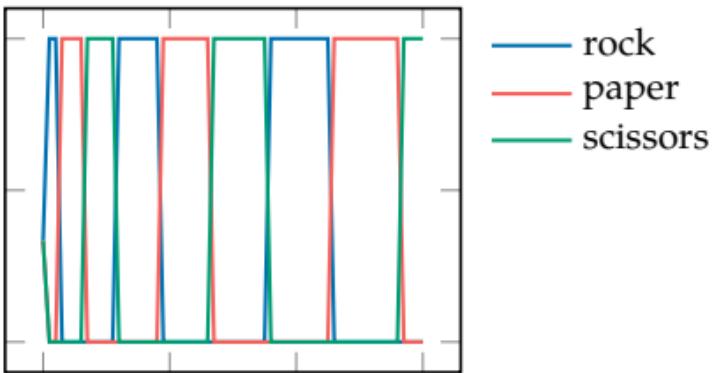


iteration

opponent model

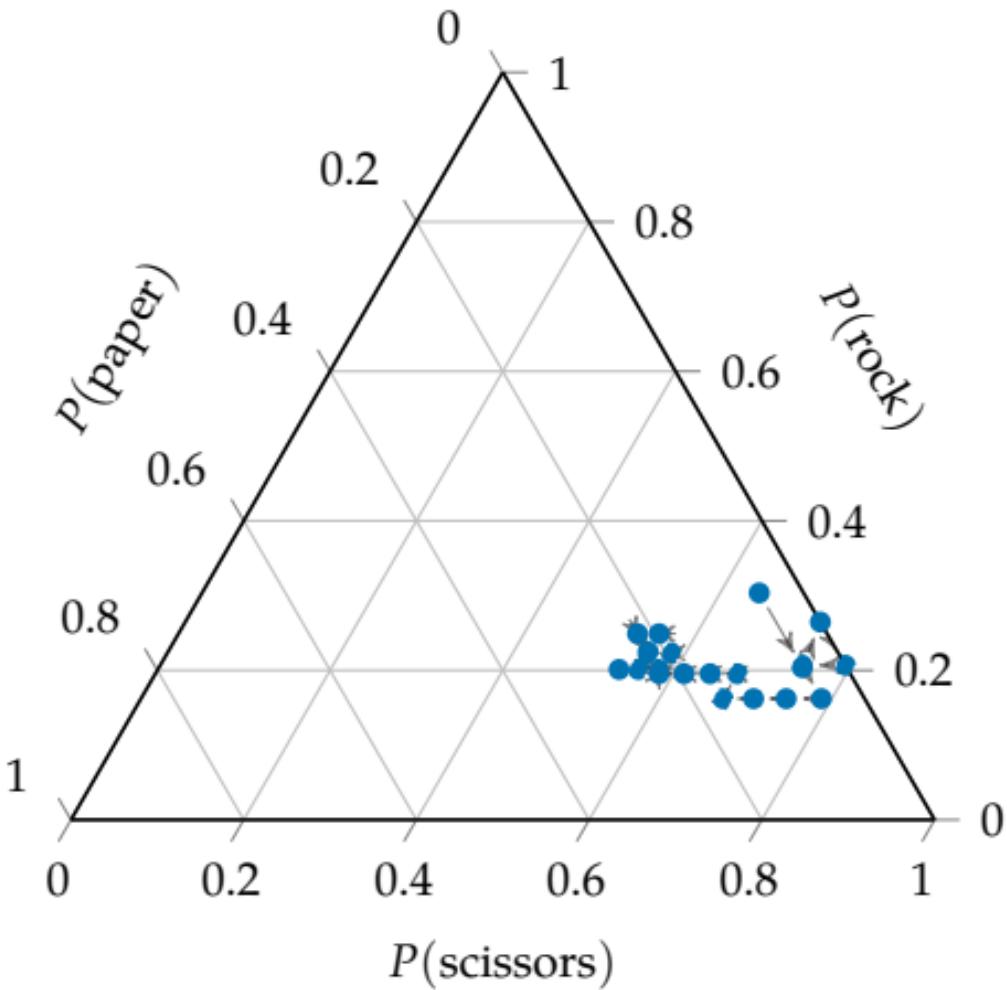


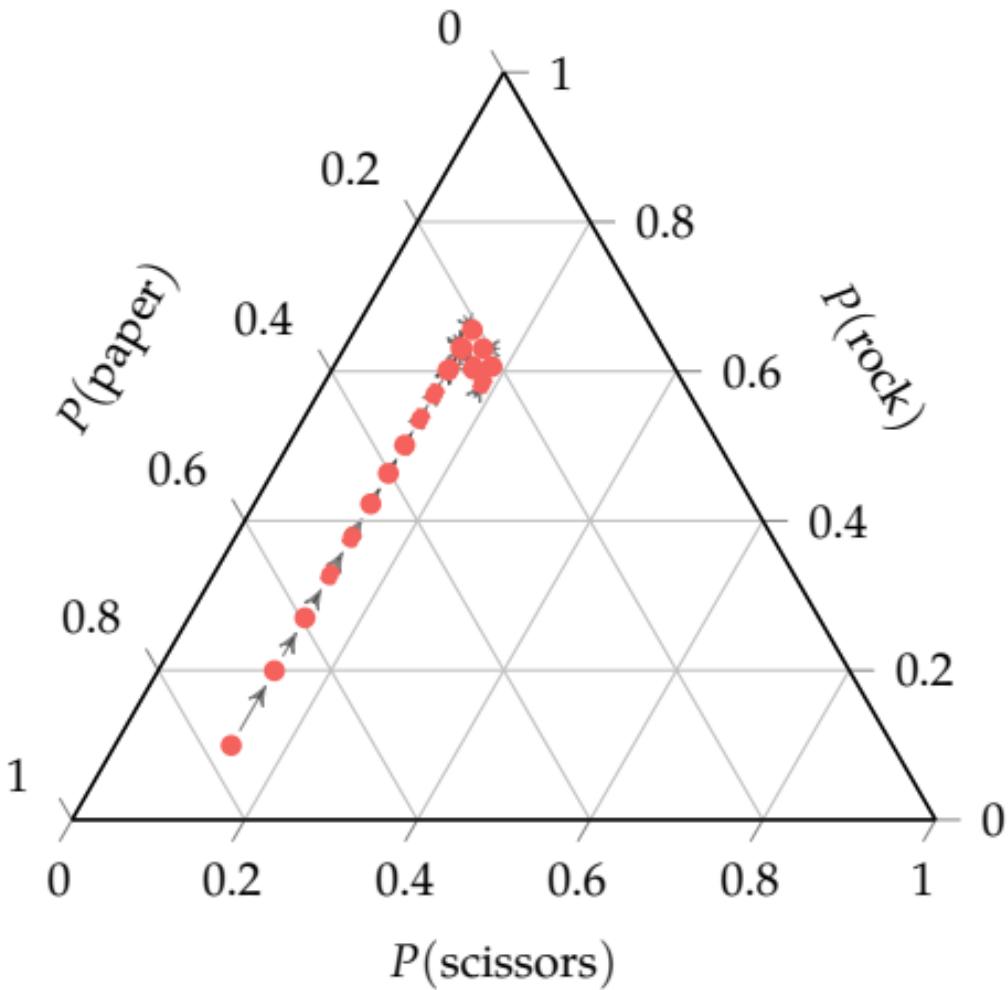
policy



iteration

iteration





		agent 2	
		climb	descend
		climb	descend
agent 1	climb	-5, -5	-1, 0
	descend	0, -1	-4, -4

agent 2

dinner

movie

agent 1

dinner

movie

2, 1

0, 0

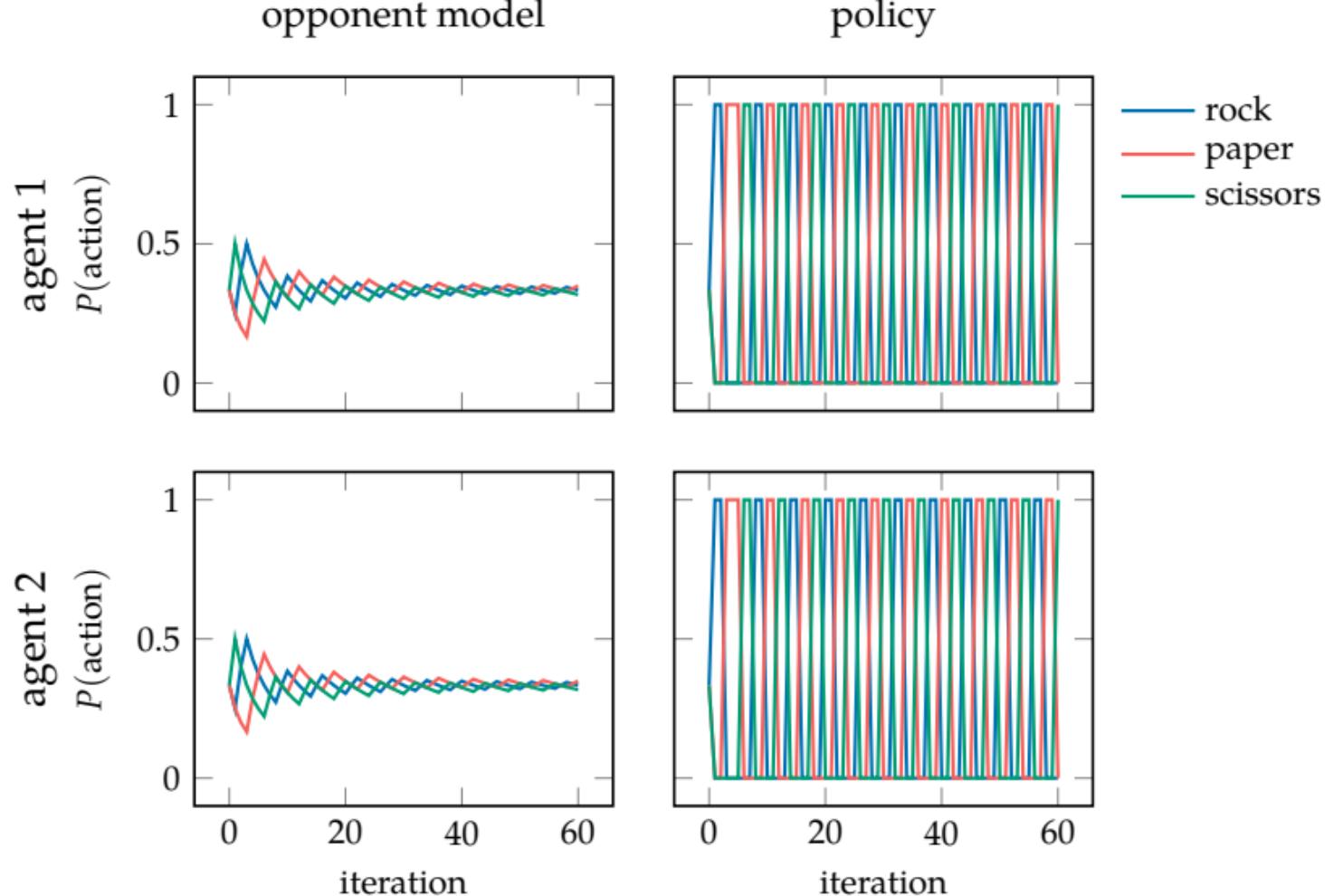
0, 0

1, 2

Iteration	Agent 1's Action	Agent 2's Action	Rewards
1	paper	rock	1.0, -1.0
2	paper	scissors	-1.0, 1.0
3	rock	scissors	1.0, -1.0
4	rock	paper	-1.0, 1.0
5	scissors	paper	1.0, -1.0
6	scissors	rock	-1.0, 1.0
7	paper	rock	1.0, -1.0
8	paper	scissors	-1.0, 1.0
9	rock	scissors	1.0, -1.0
10	rock	paper	-1.0, 1.0

		agent 2		
		rock	paper	scissors
agent 1	rock	0, 0	0, 1	1, 0
	paper	1, 0	0, 0	0, 1
	scissors	0, 1	1, 0	0, 0

opponent model

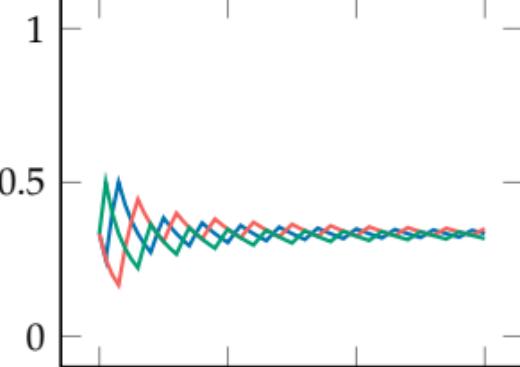


policy

iteration

agent 1

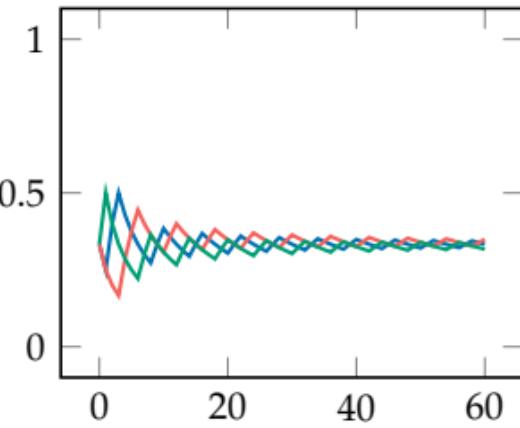
$P(\text{action})$

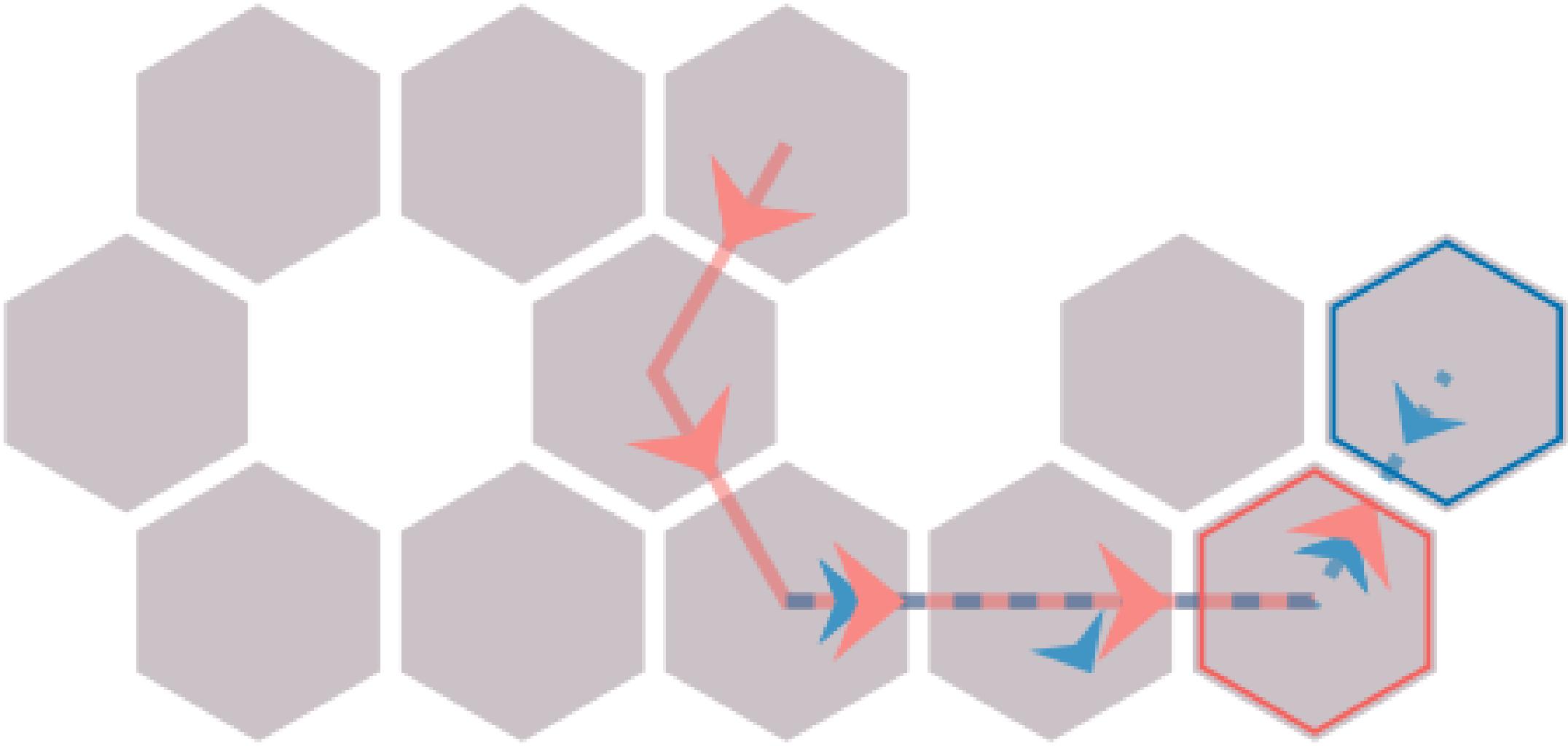


iteration

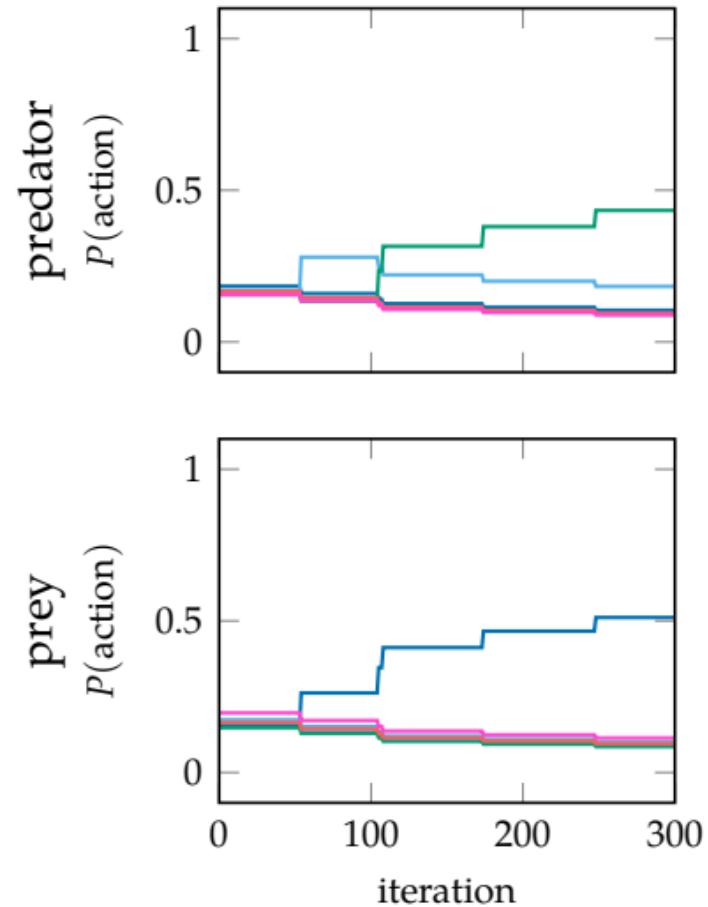
agent 2

$P(\text{action})$

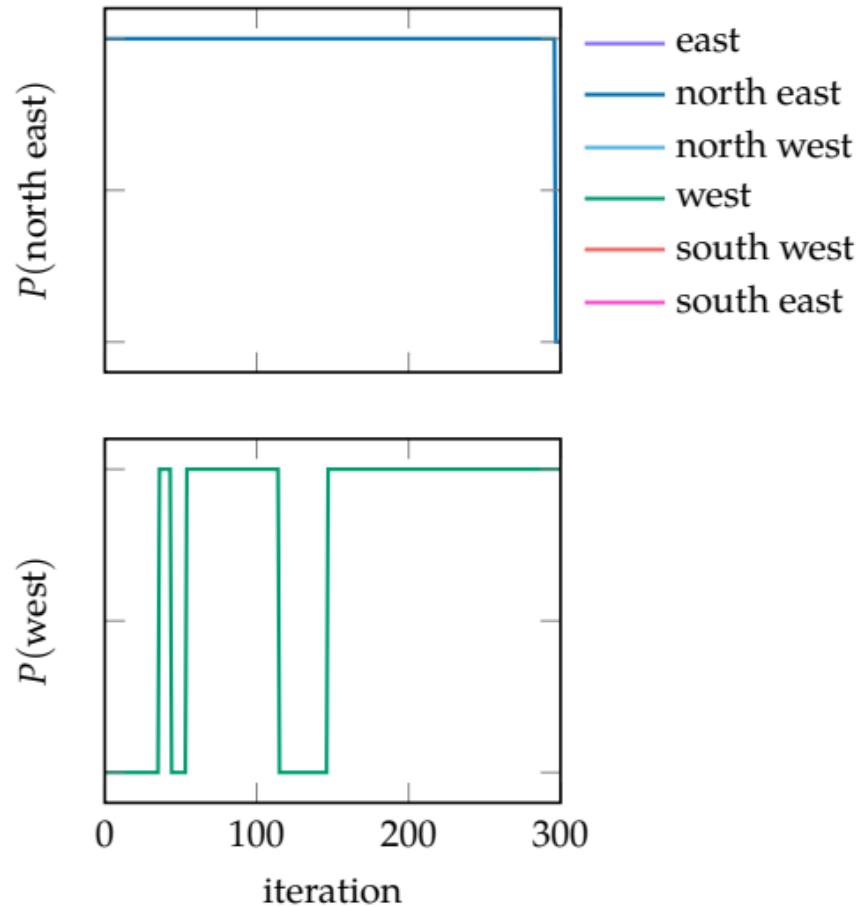




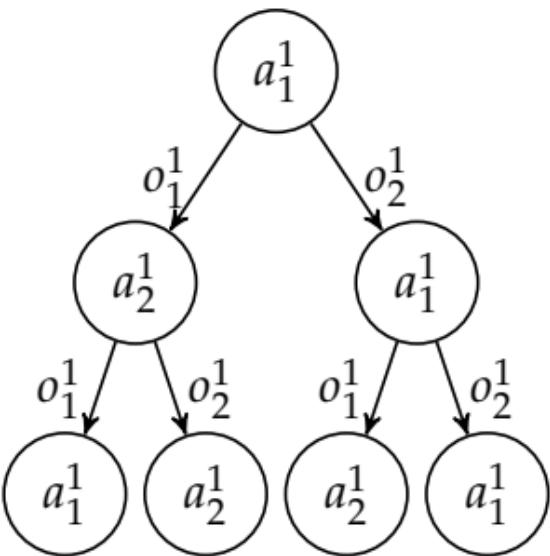
opponent model



policy



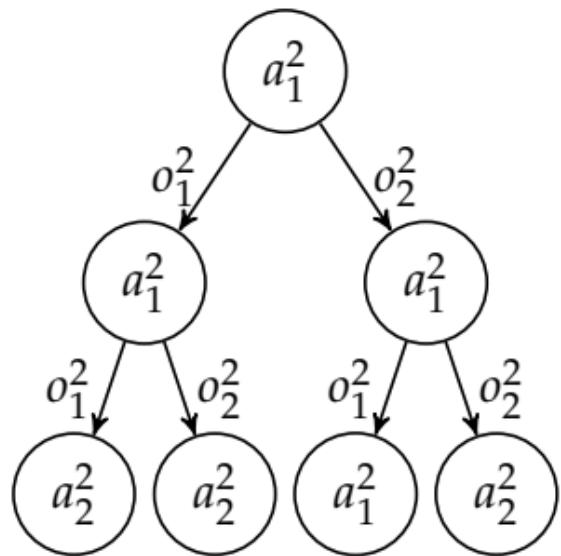
agent 1's policy π^1



o_1^1 = quiet a_1^1 = ignore

o_2^1 = crying a_2^1 = feed

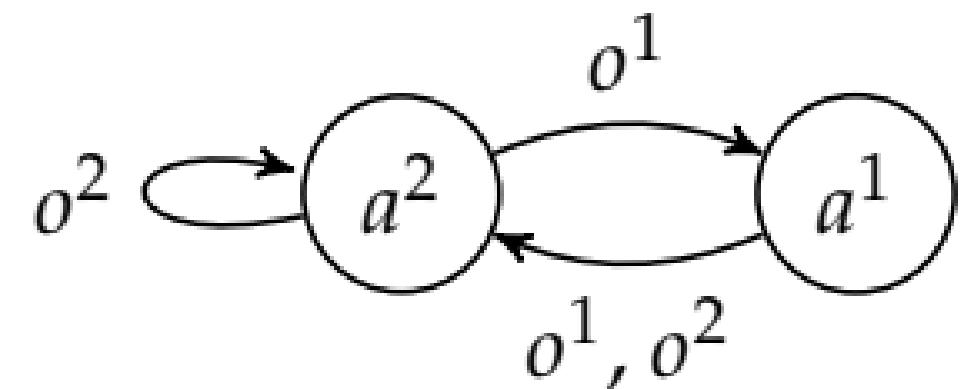
agent 2's policy π^2



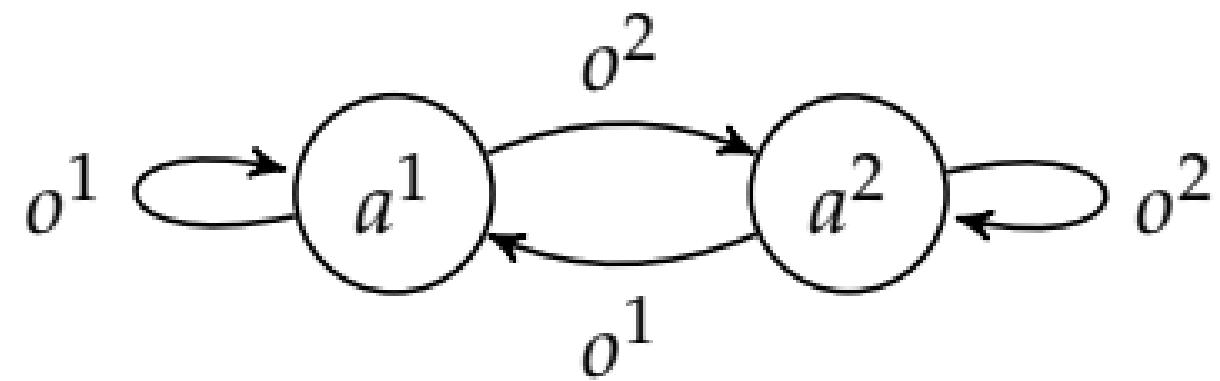
o_1^2 = quiet a_1^2 = ignor

o_2^2 = crying a_2^2 = feed

agent 1's policy π^1



agent 2's policy π^2



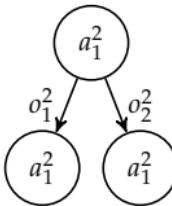
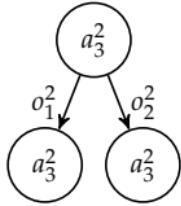
$o^1 = \text{quiet}$ $a^1 = \text{ignore}$

$o^1 = \text{hungry}$ $a^1 = \text{feed}$

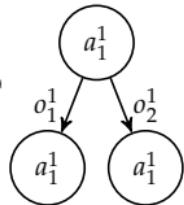
agent 2

π_1^2

π_c^2

 \dots 

π_1^1



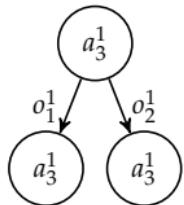
$$U^{\pi_1^1, \pi_1^2, 1}(b), \\ U^{\pi_1^1, \pi_1^2, 2}(b)$$

$$U^{\pi_1^1, \pi_c^2, 1}(b), \\ U^{\pi_1^1, \pi_c^2, 2}(b)$$

 \vdots \vdots \dots \vdots

agent 1

π_r^1

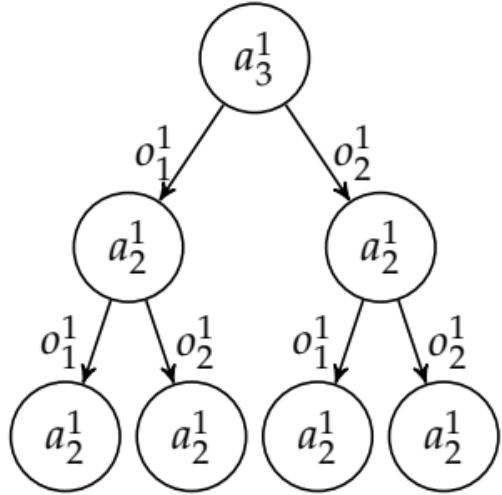
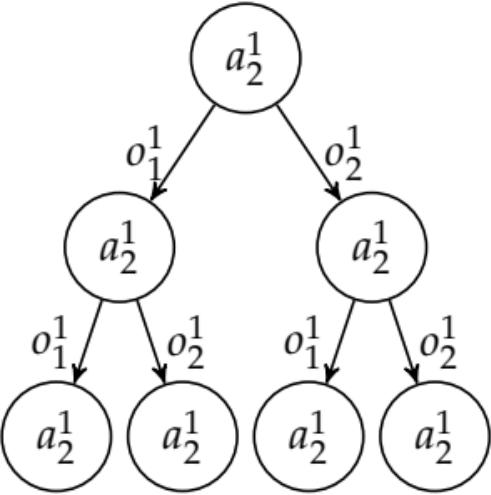
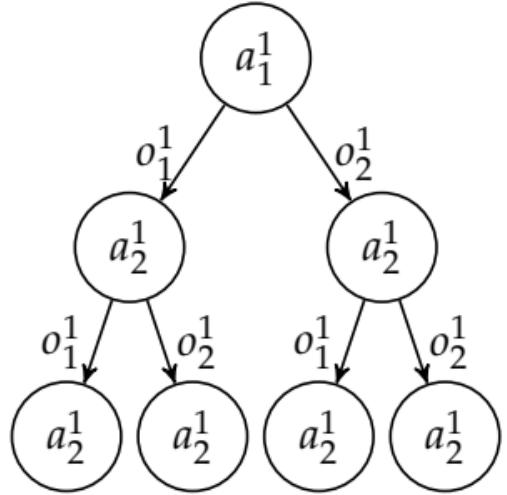


$$U^{\pi_r^1, \pi_1^2, 1}(b), \\ U^{\pi_r^1, \pi_1^2, 2}(b)$$

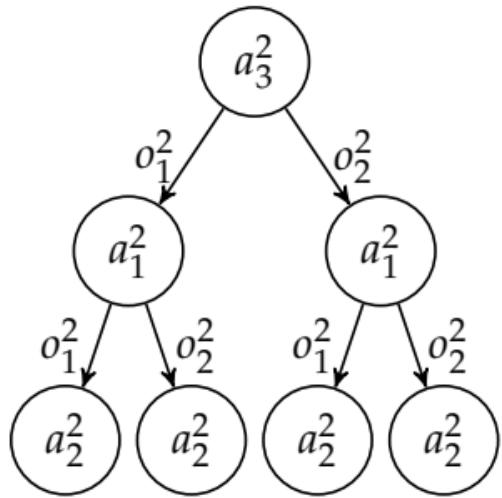
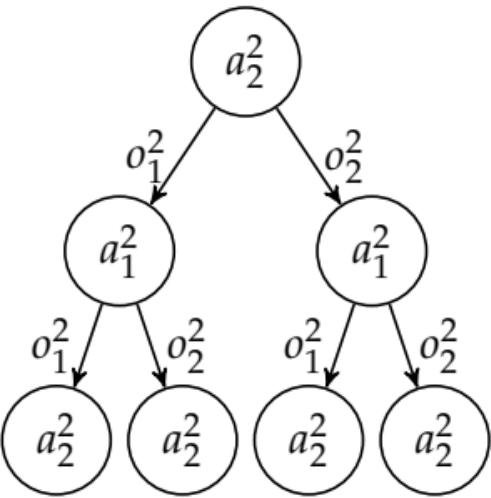
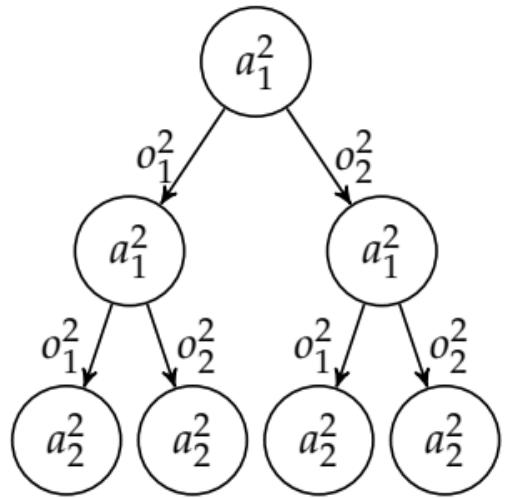
$$U^{\pi_r^1, \pi_c^2, 1}(b), \\ U^{\pi_r^1, \pi_c^2, 2}(b)$$

 \dots

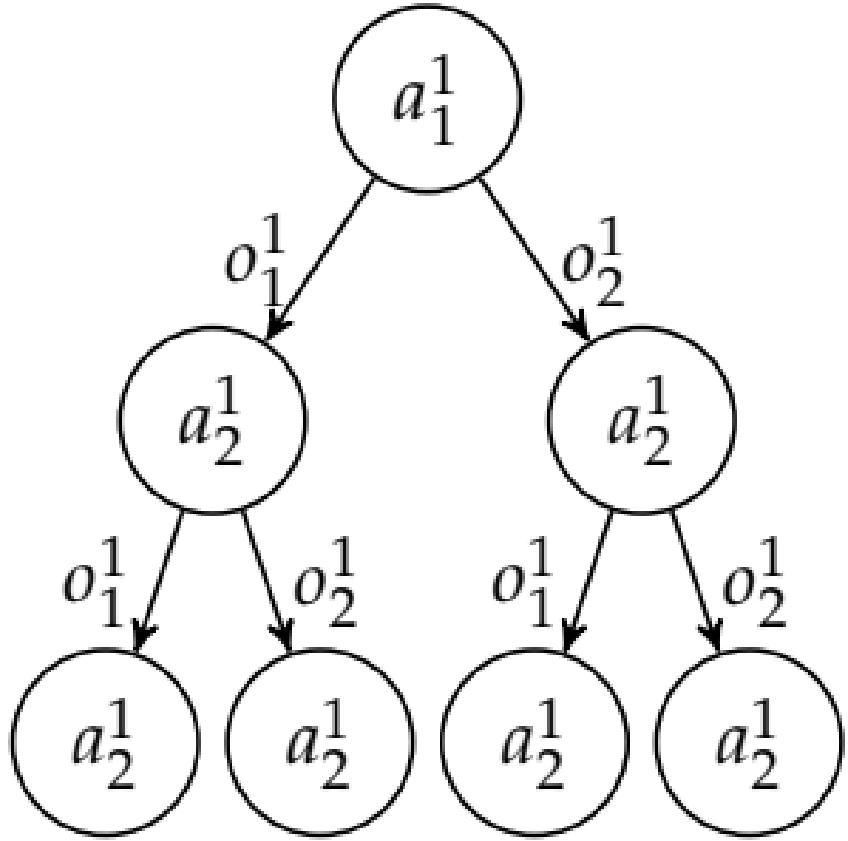
agent 1 policies



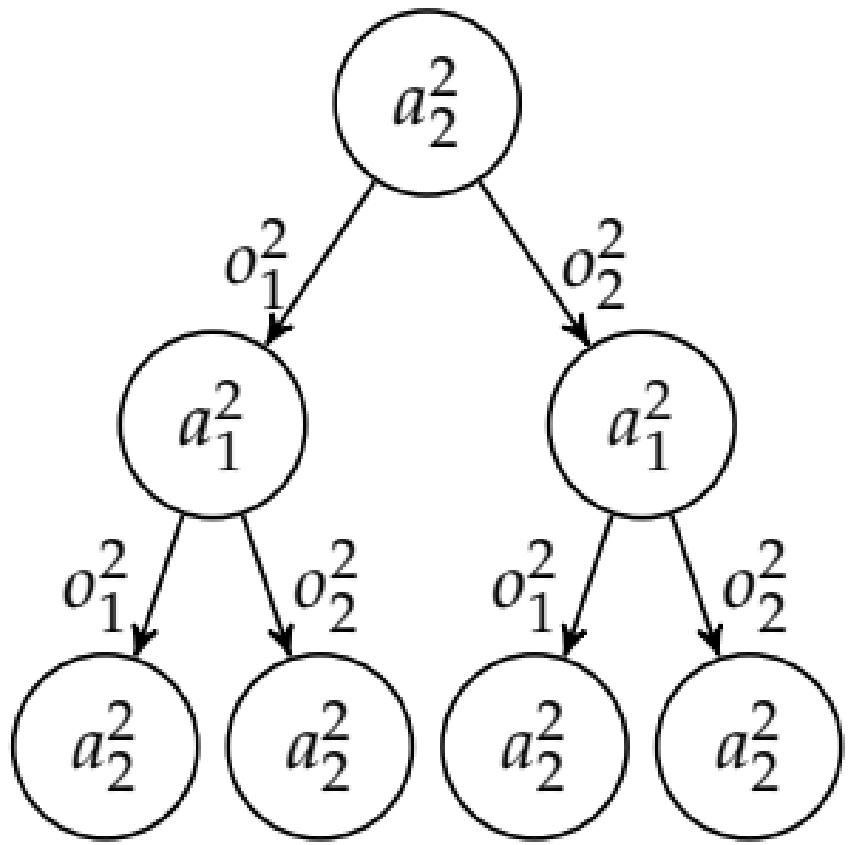
agent 2 policies

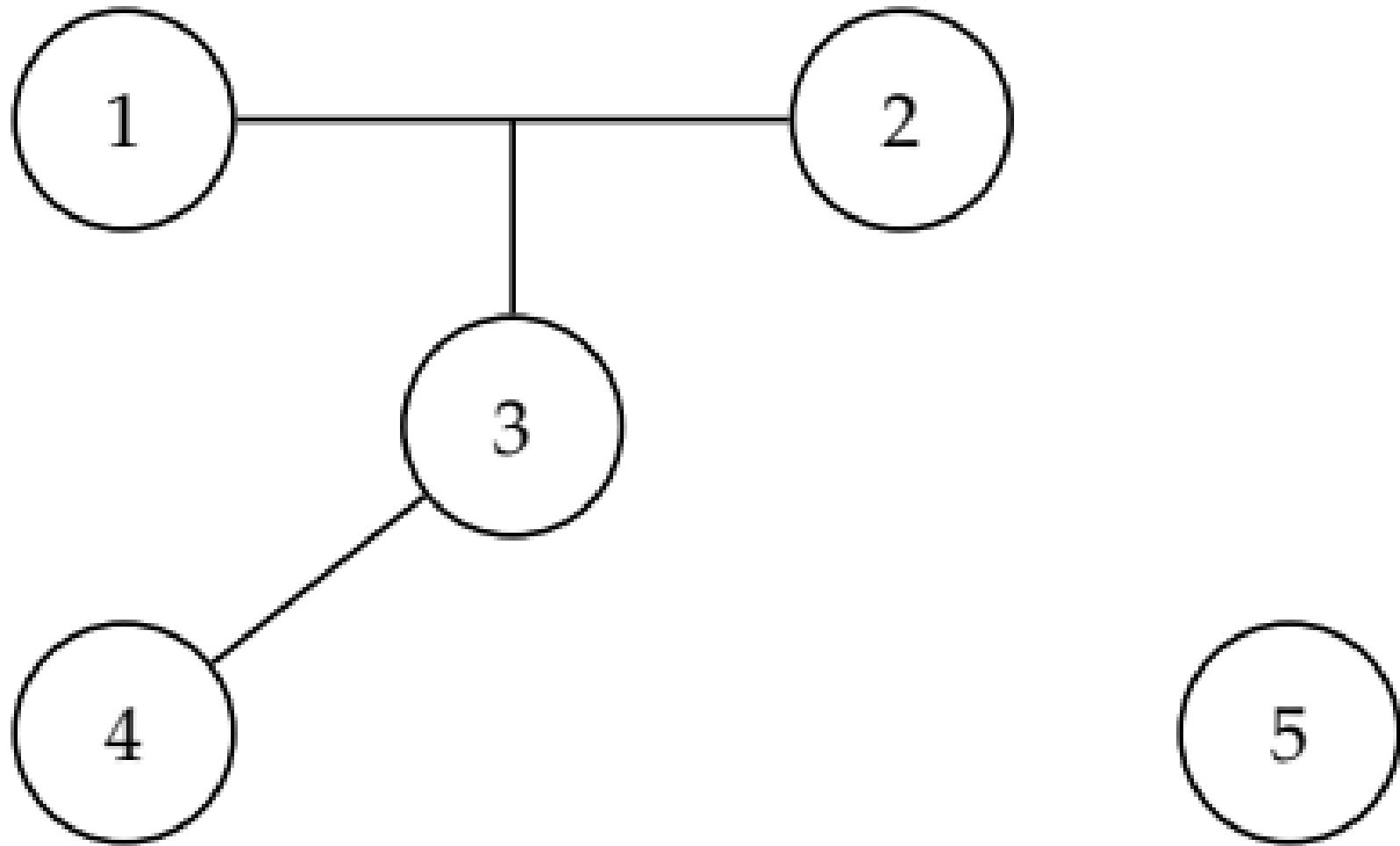


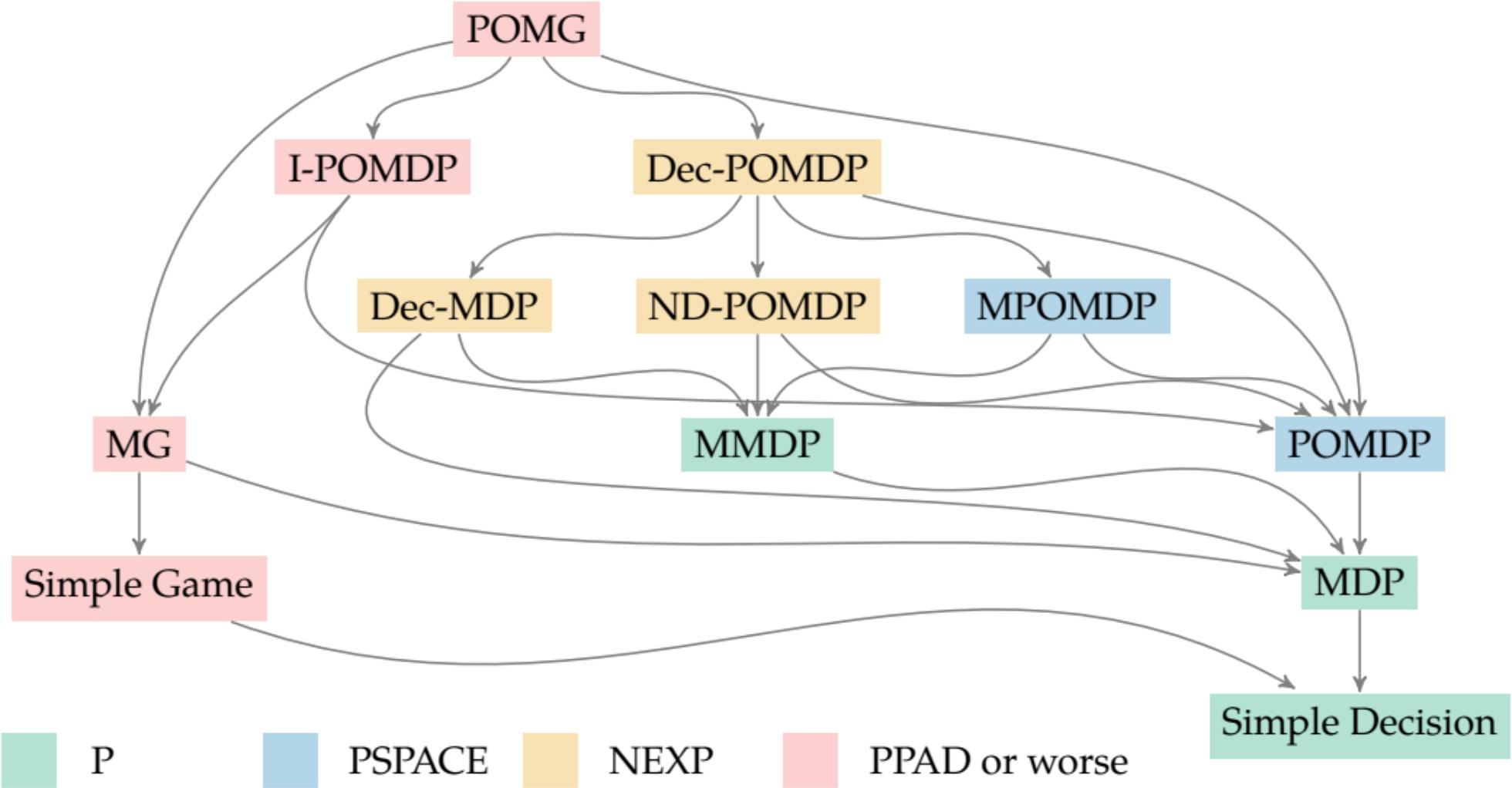
agent 1 policies



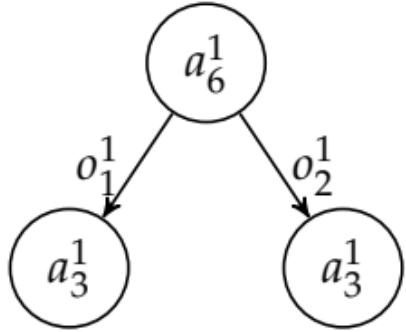
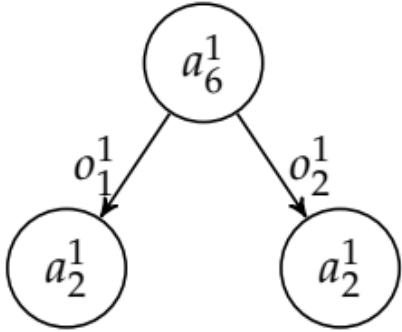
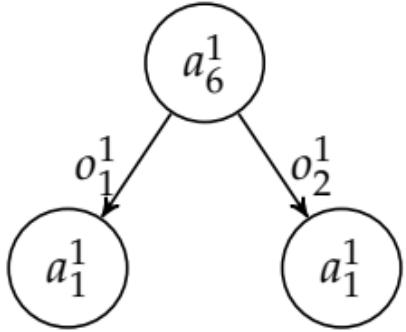
agent 2 policies



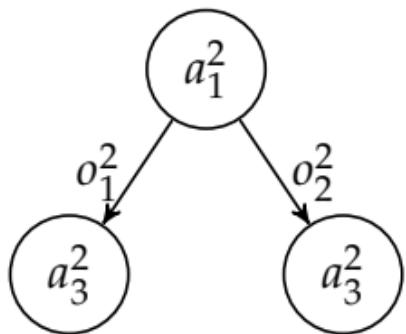
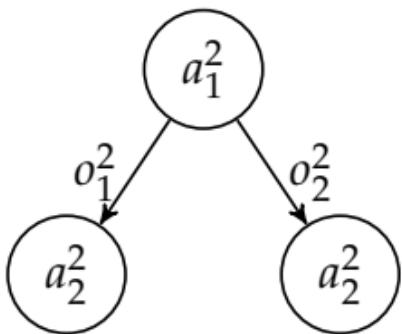
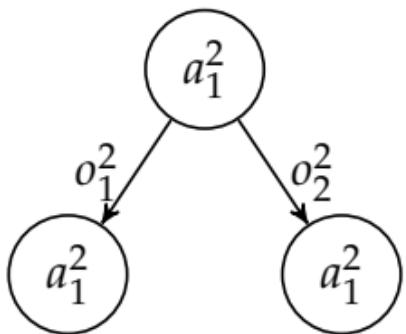




agent 1 policies

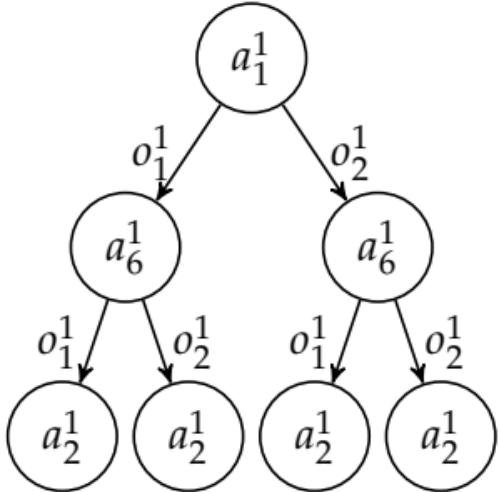
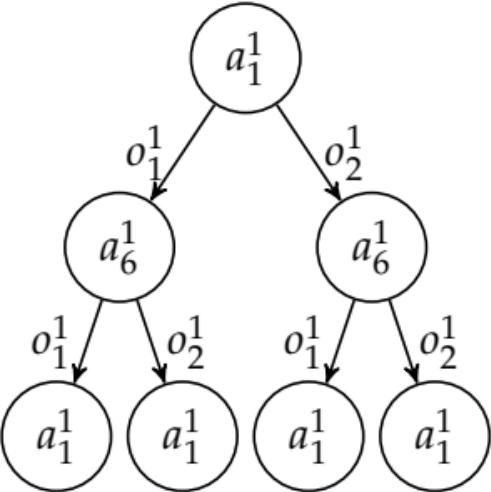
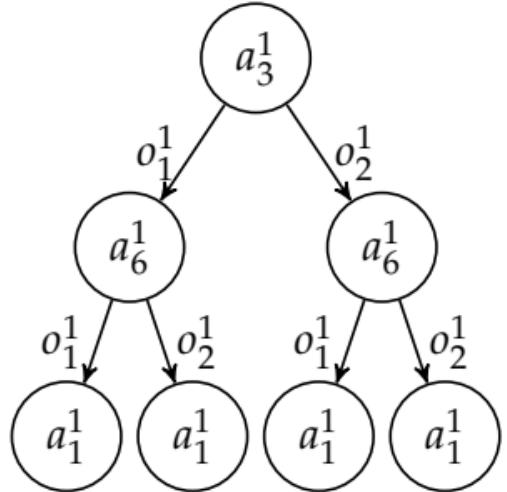


agent 2 policies

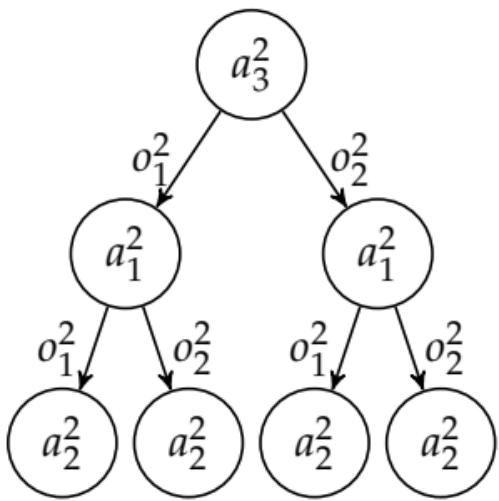
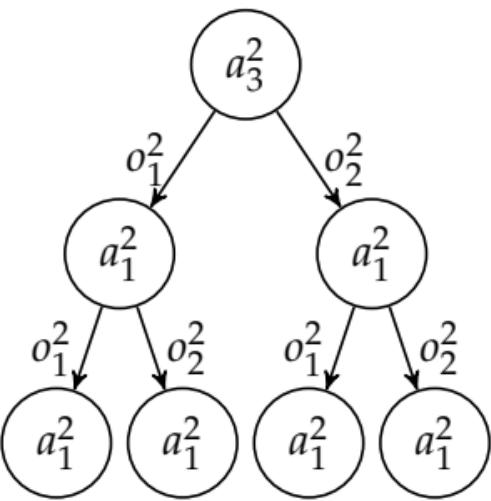
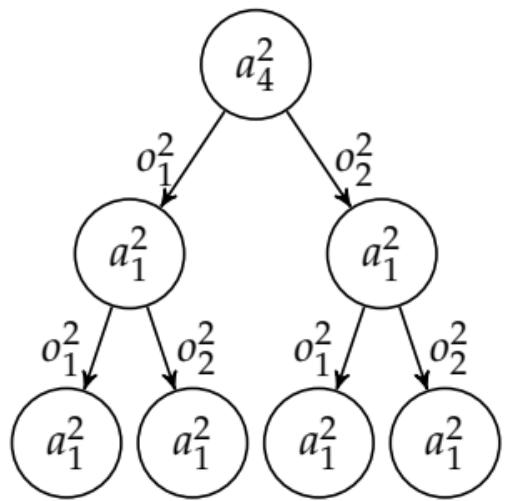


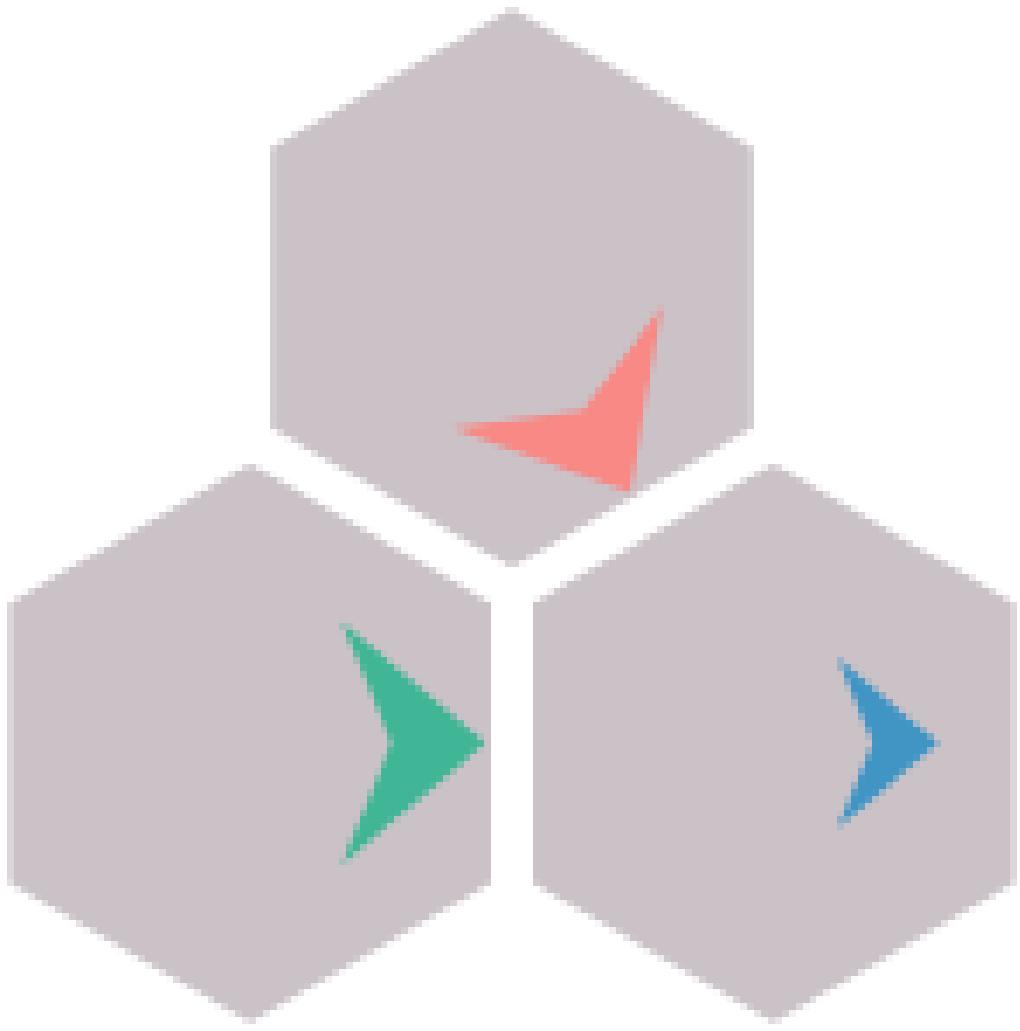
$b_3 = [0.0, 0.03, 0.01, 0.0, 0.03, 0.01, 0.0$
 $.0, 0.08, 0.03, 0.0, 0.0$
 $0, 0.01, 0.08, 0.34, 0.03$
 $.01, 0.0, 0.01, 0.0]$

agent 1 policies



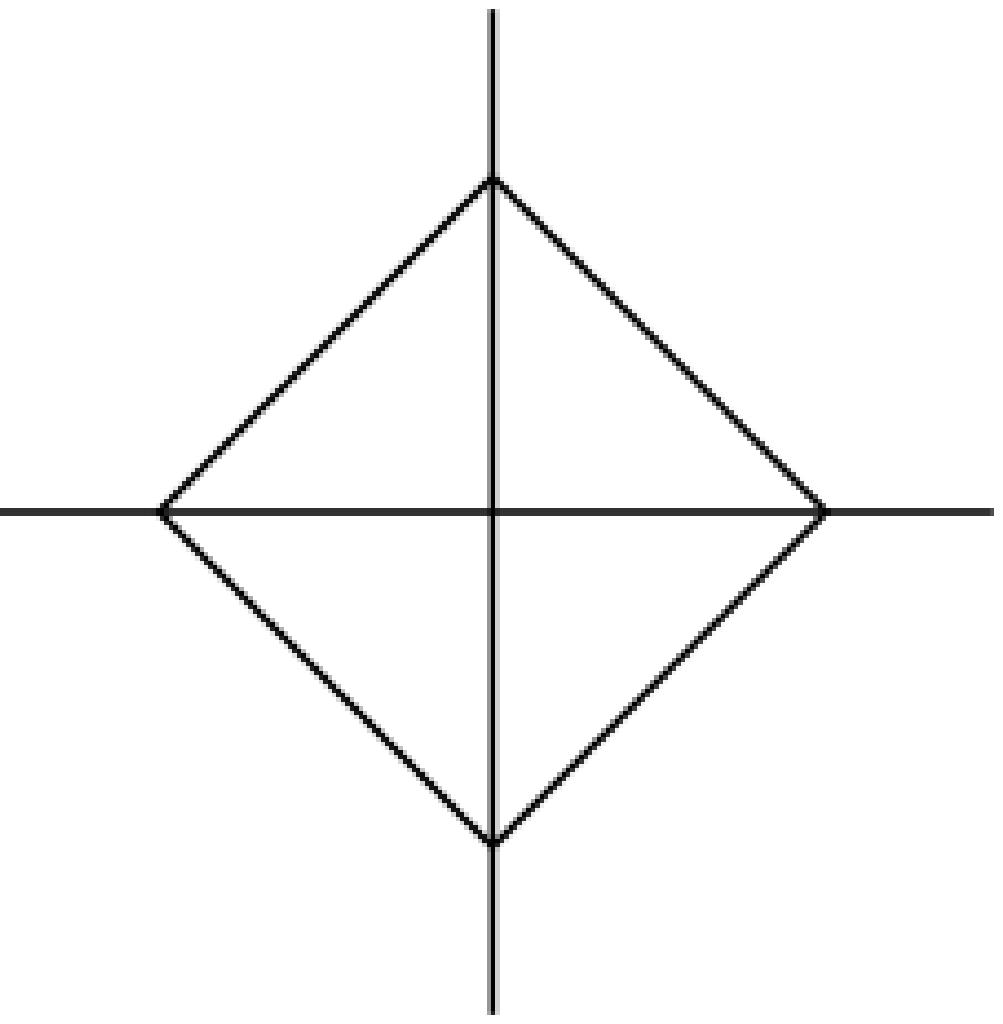
agent 2 policies





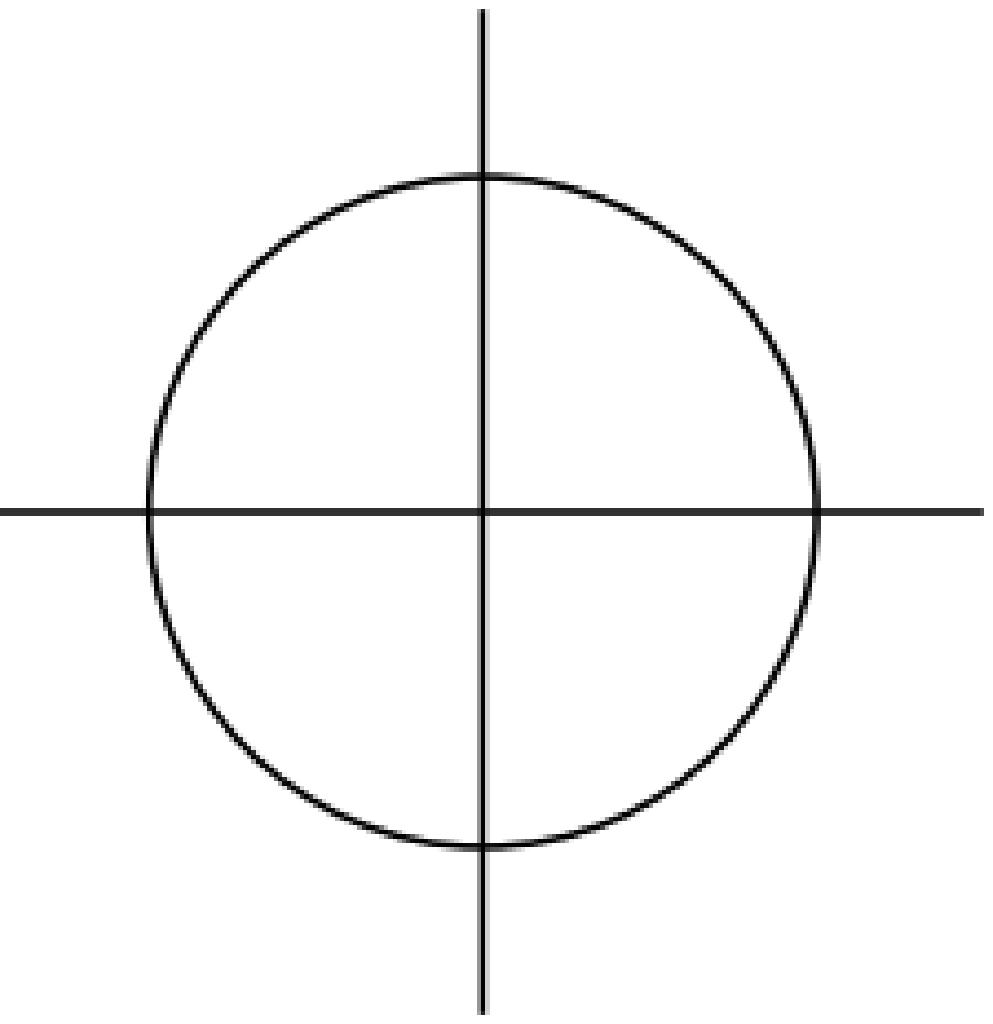
$$L_1: \|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

This metric is often referred to as the *taxicab norm*.



$$L_2: \|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

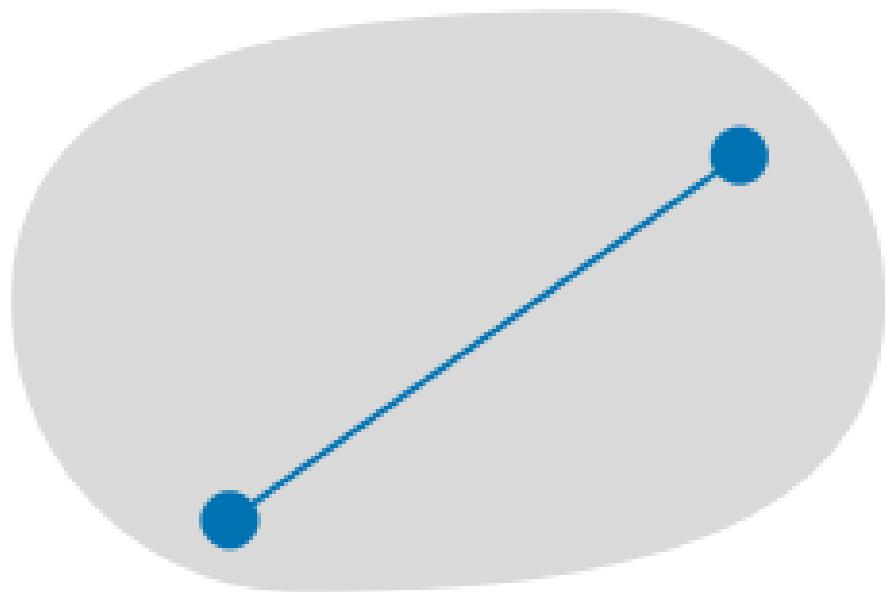
This metric is often referred to as the *Euclidean norm*.



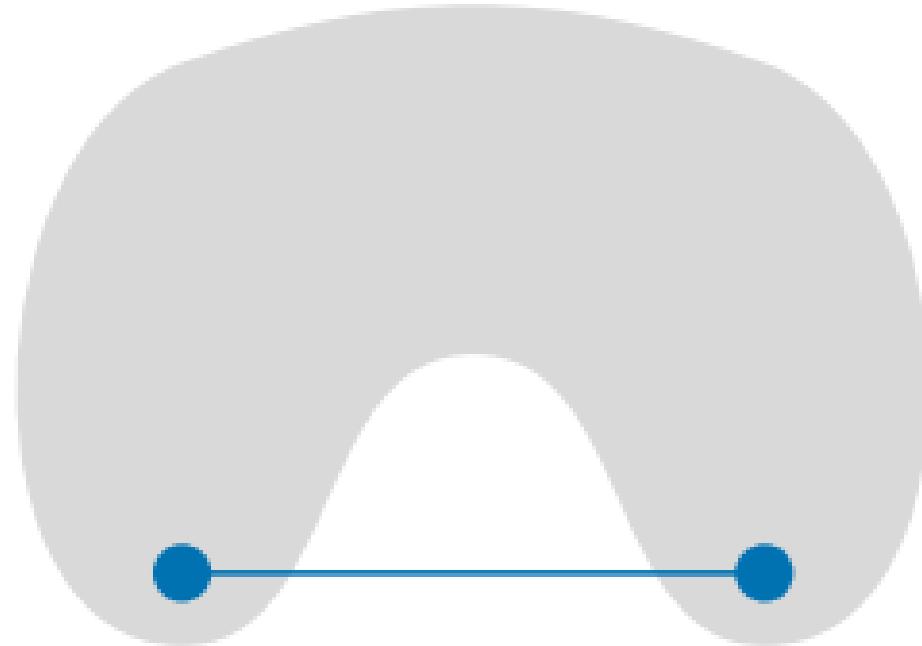
$$L_{\infty}: \|\mathbf{x}\|_{\infty} = \max(|x_1|, |x_2|, \dots, |x_n|)$$

This metric is often referred to as the *max norm*, *Chebyshev norm*, or *chessboard norm*.

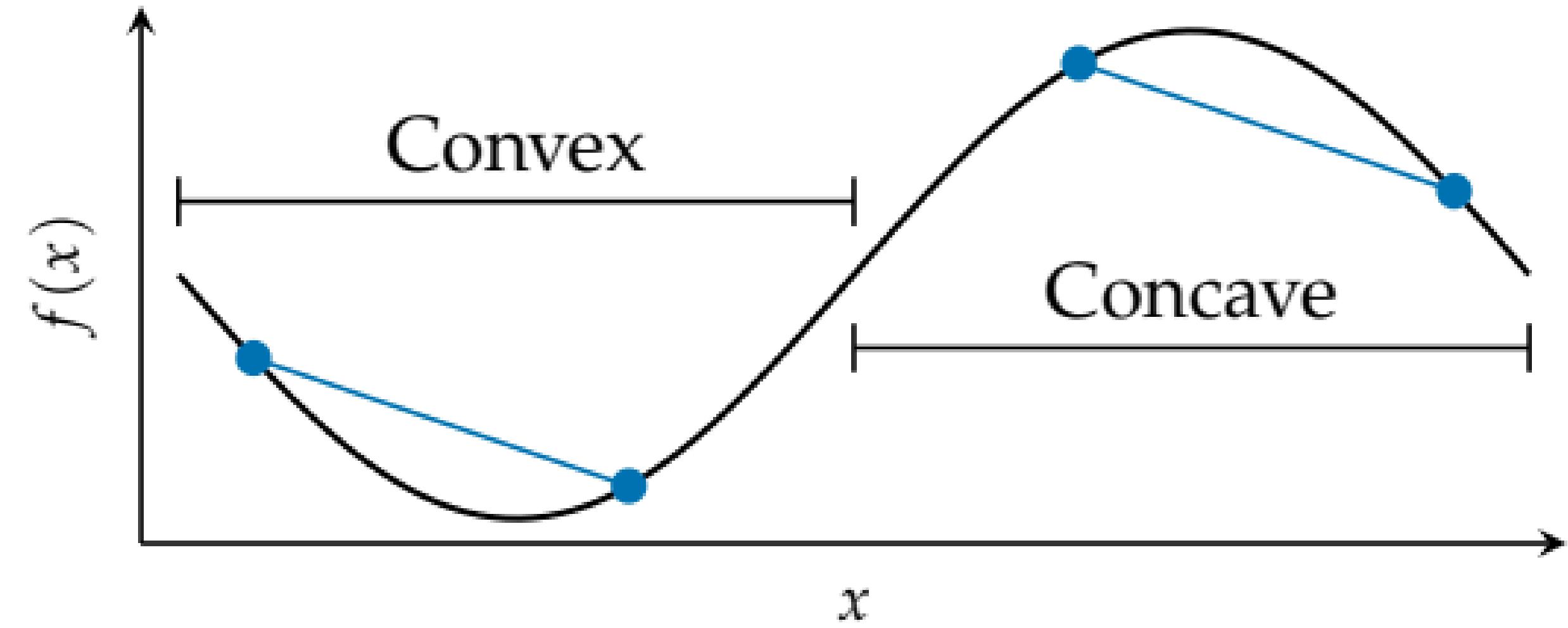
The latter name comes from the minimum number of moves that a king needs to move between two squares in chess.

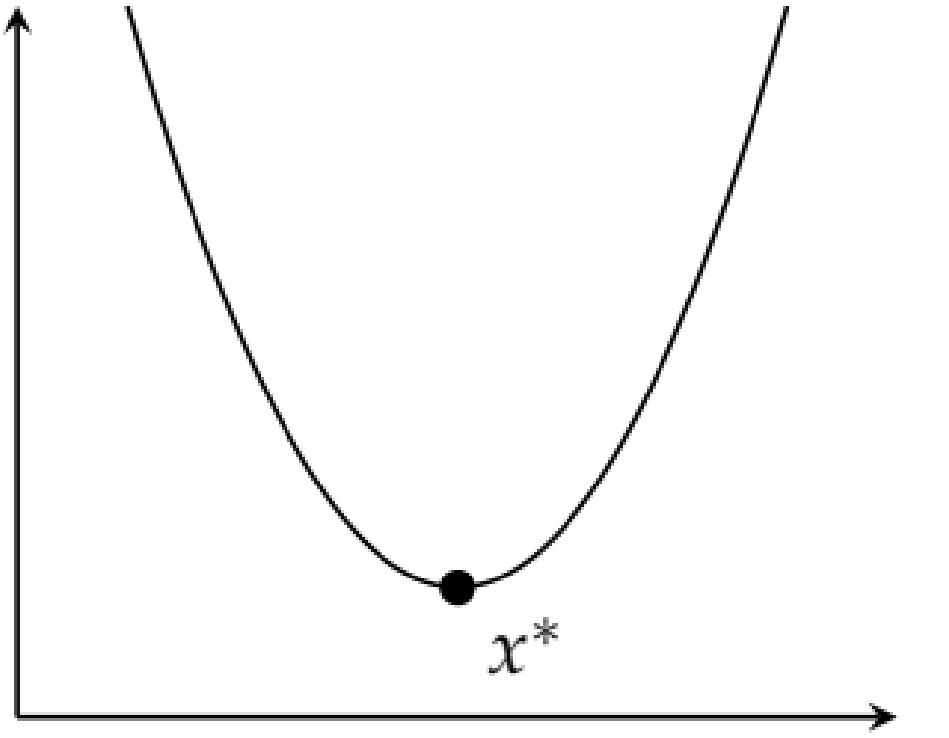


a convex set

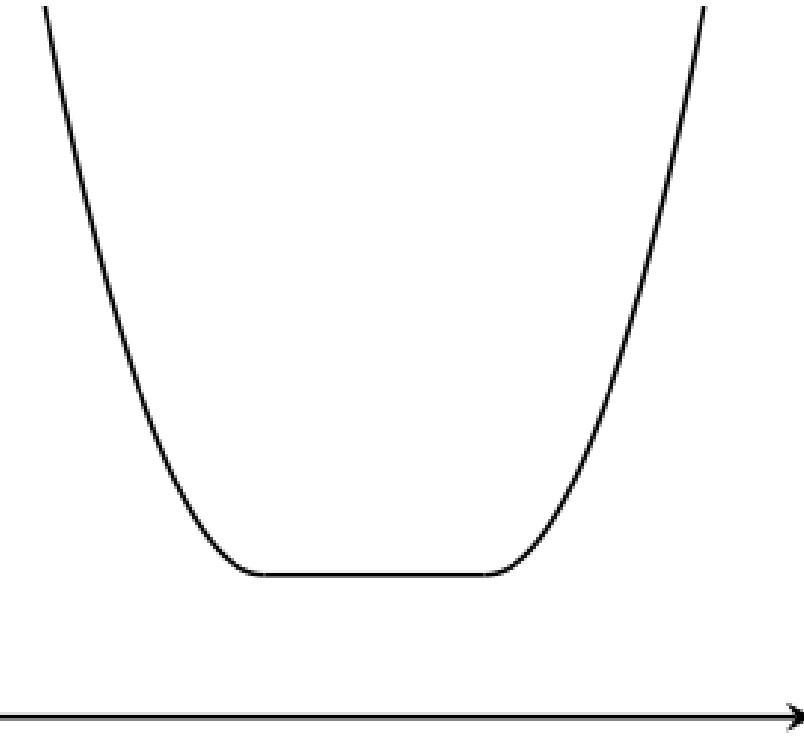


a nonconvex set

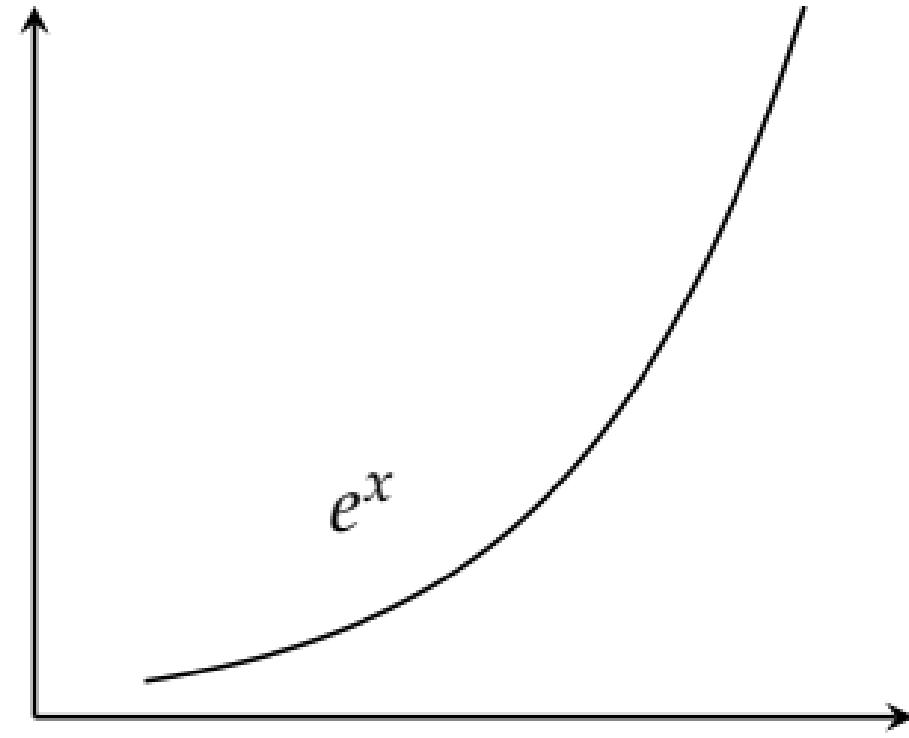




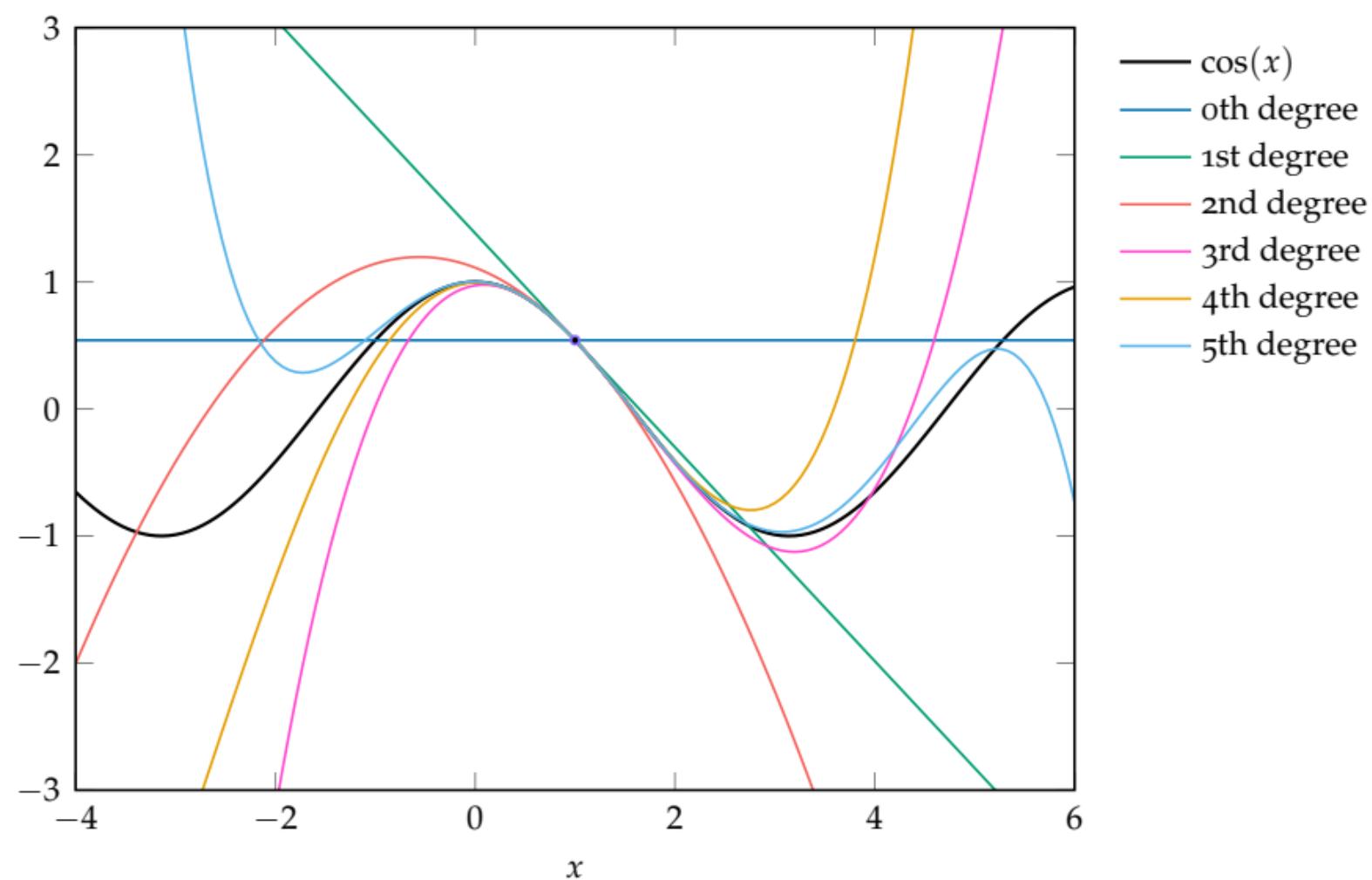
strictly convex function with
one global minimum

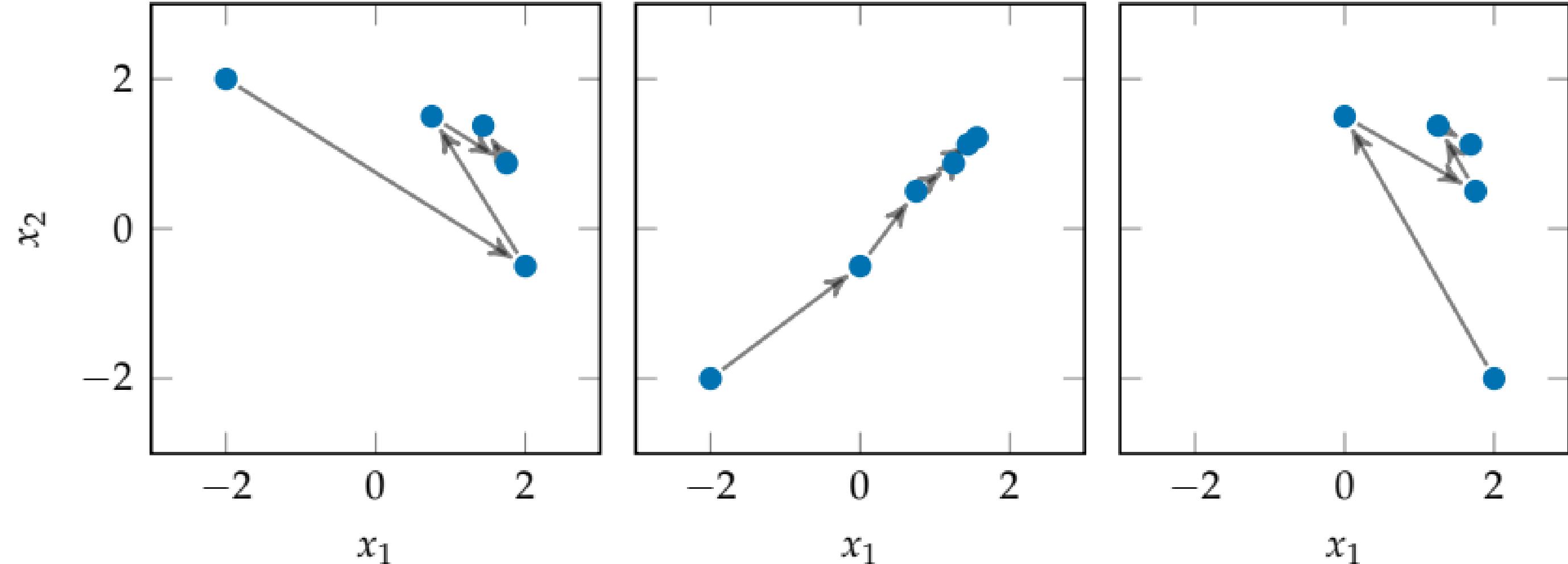


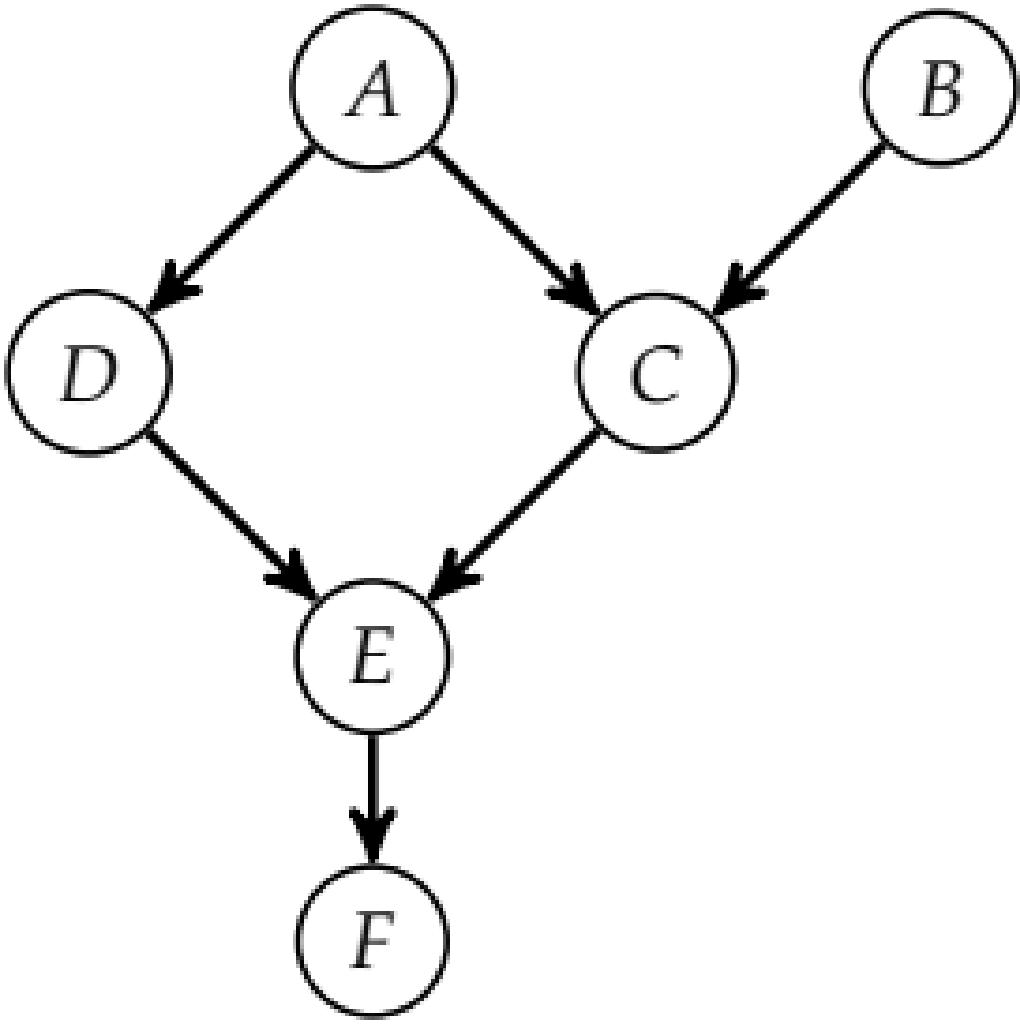
convex function without a
unique global minimum

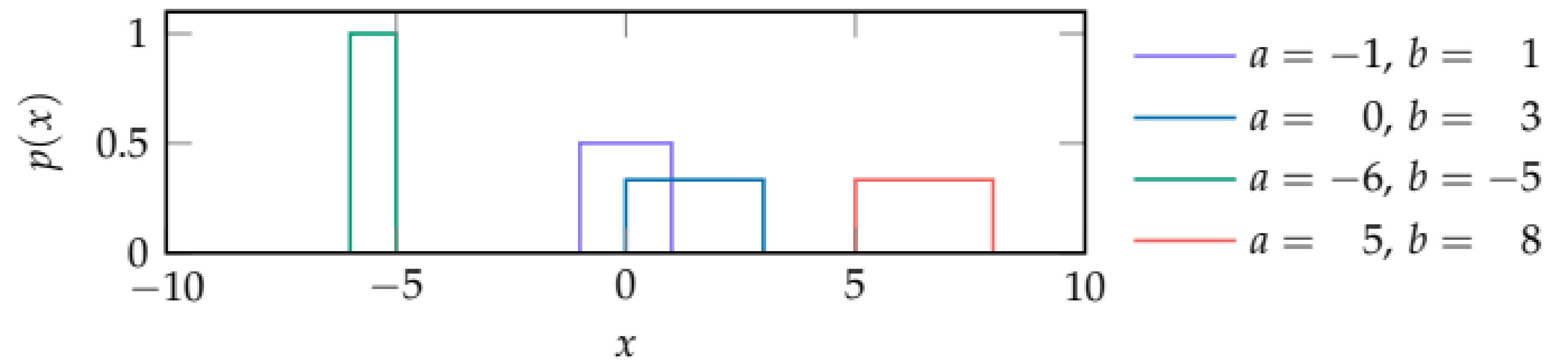


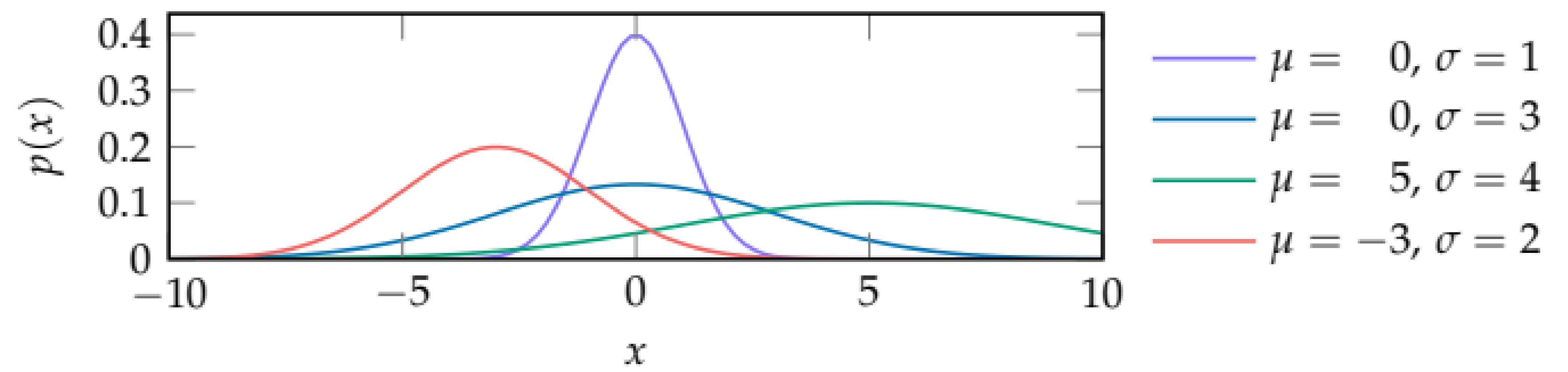
strictly convex function
without a global minimum

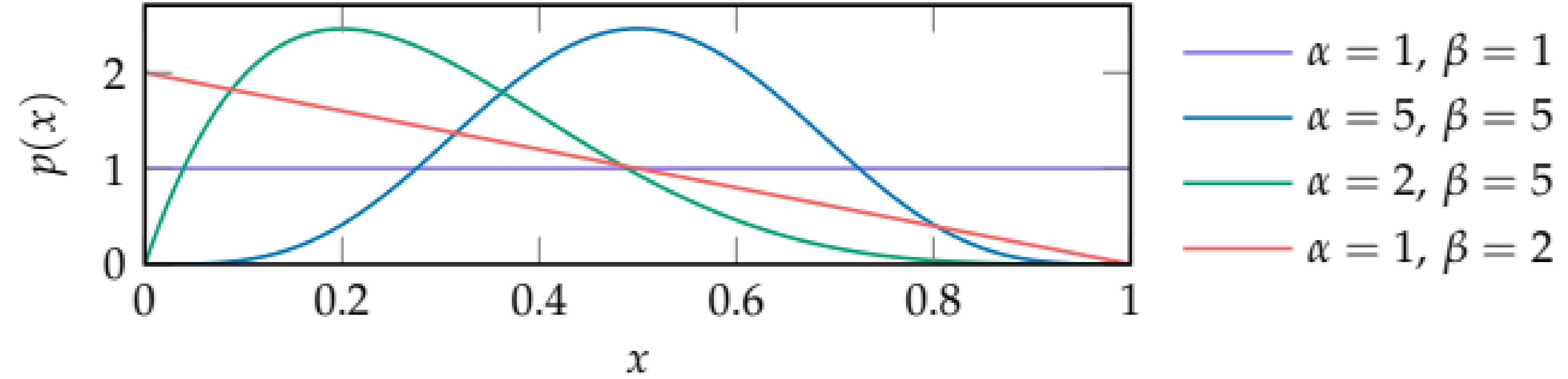








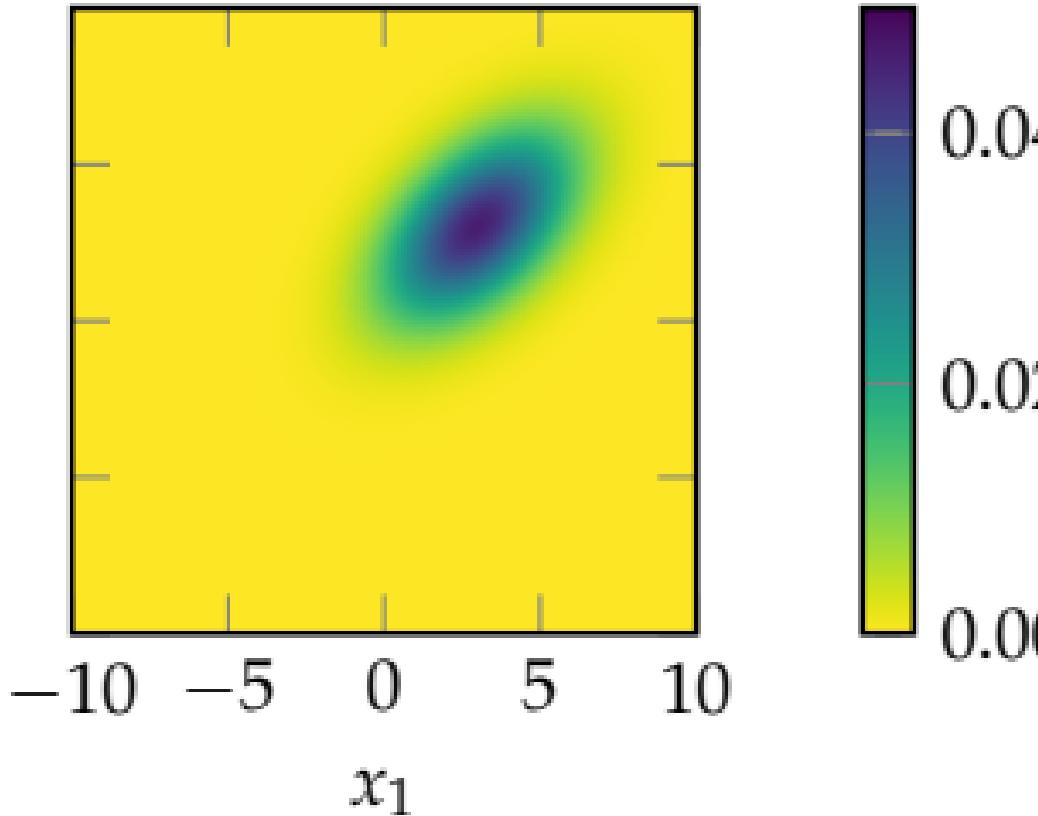
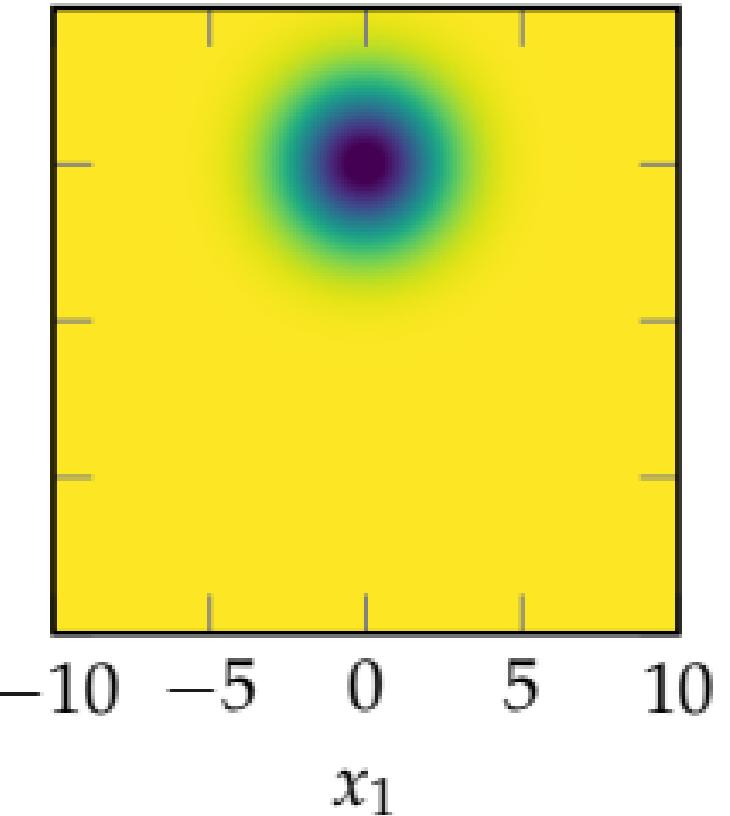
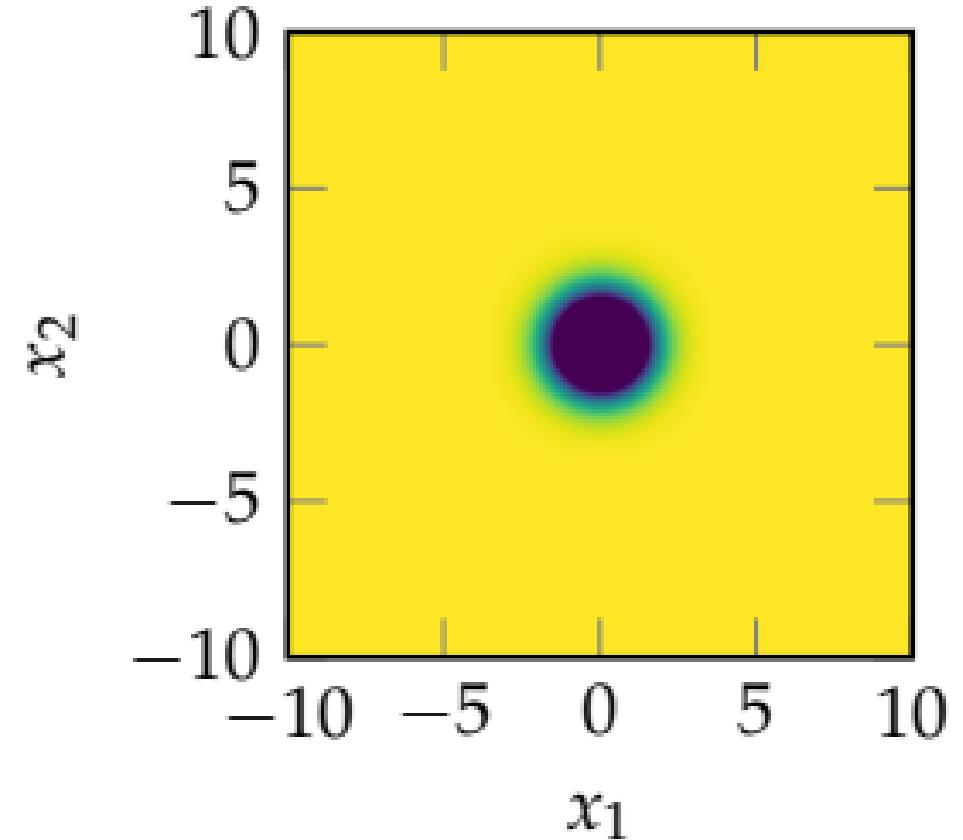




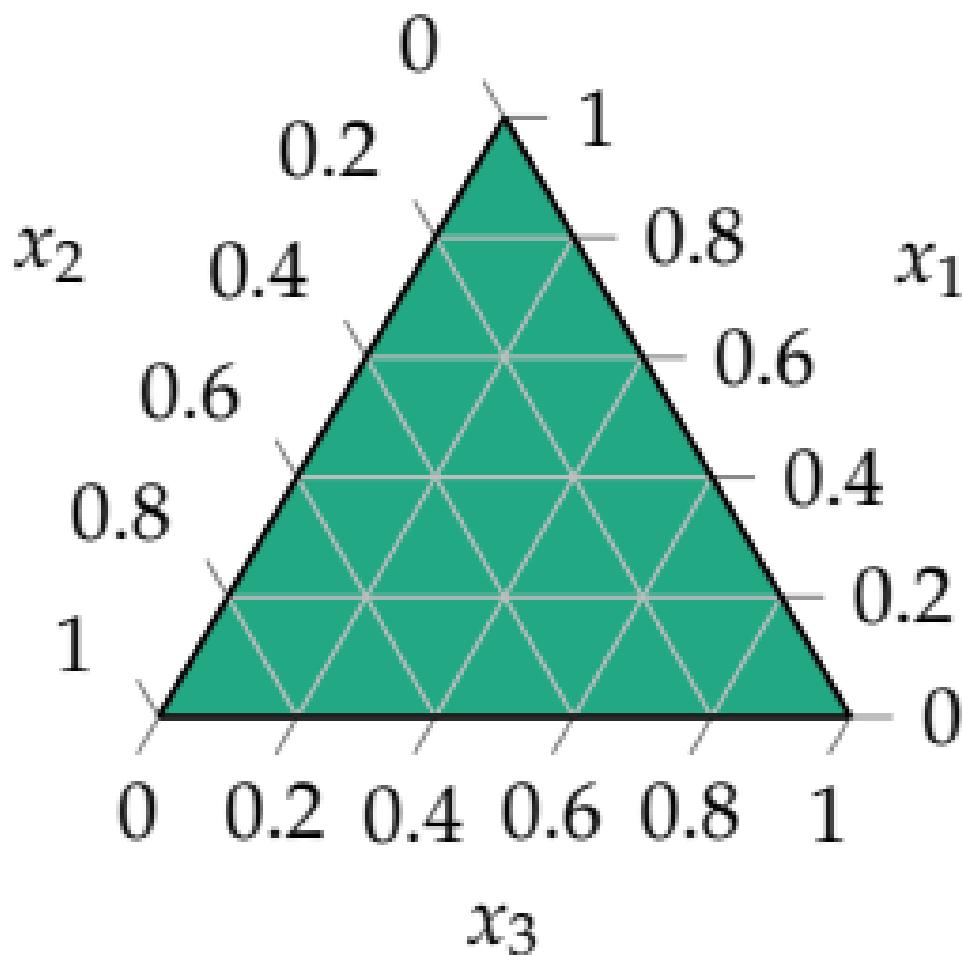
$$\mu = [0, 0], \Sigma = [1 \ 0; \ 0 \ 1]$$

$$\mu = [0, 5], \Sigma = [3 \ 0; \ 0 \ 3]$$

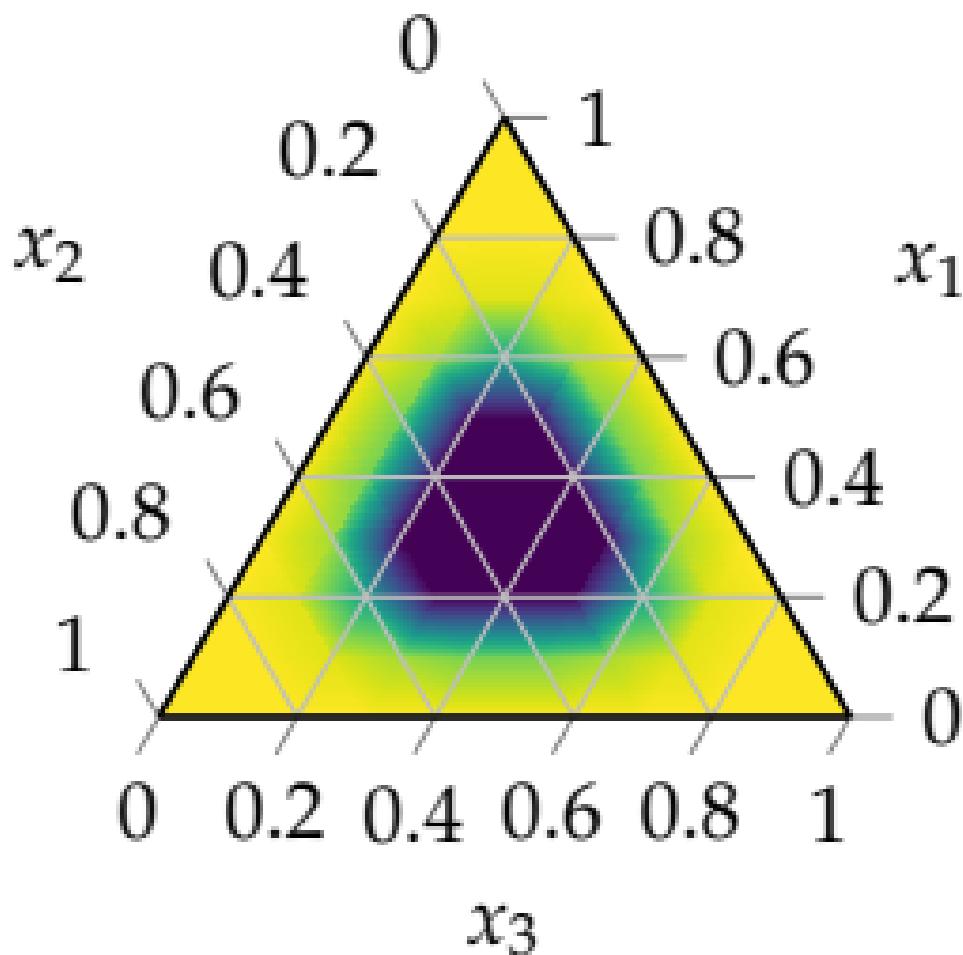
$$\mu = [3, 3], \Sigma = [4 \ 2; \ 2 \ 4]$$



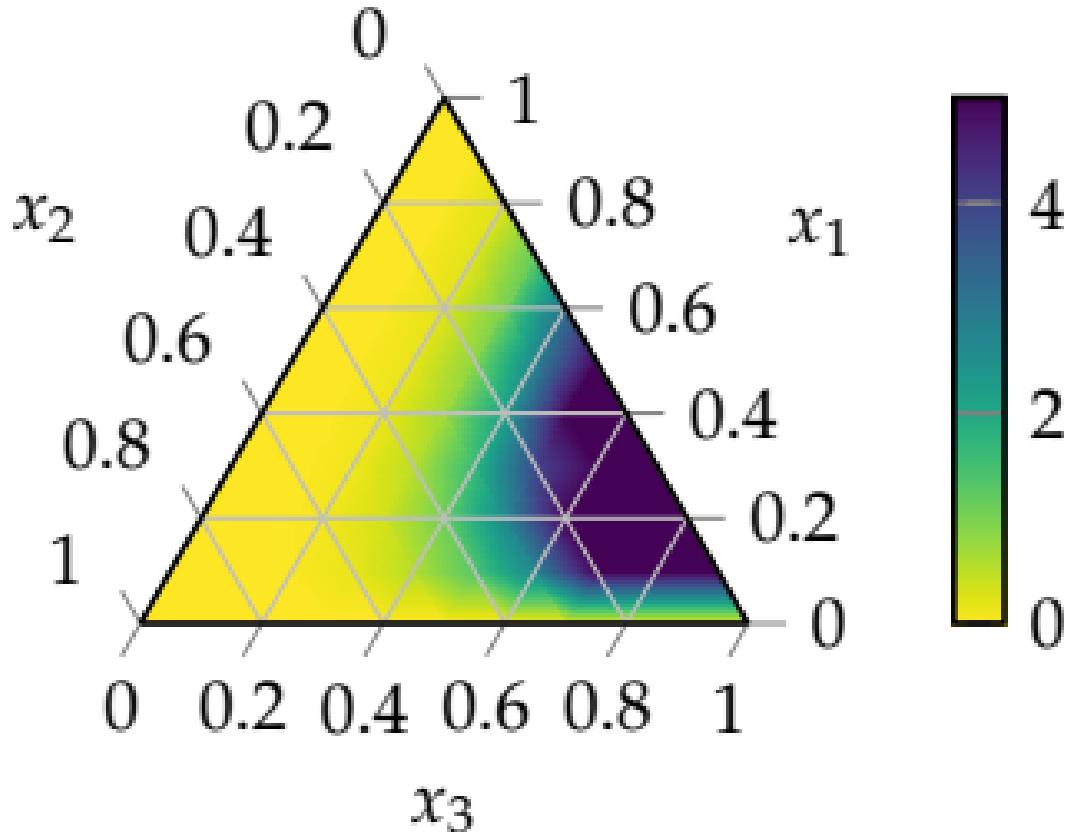
$$\alpha = [1, 1, 1]$$

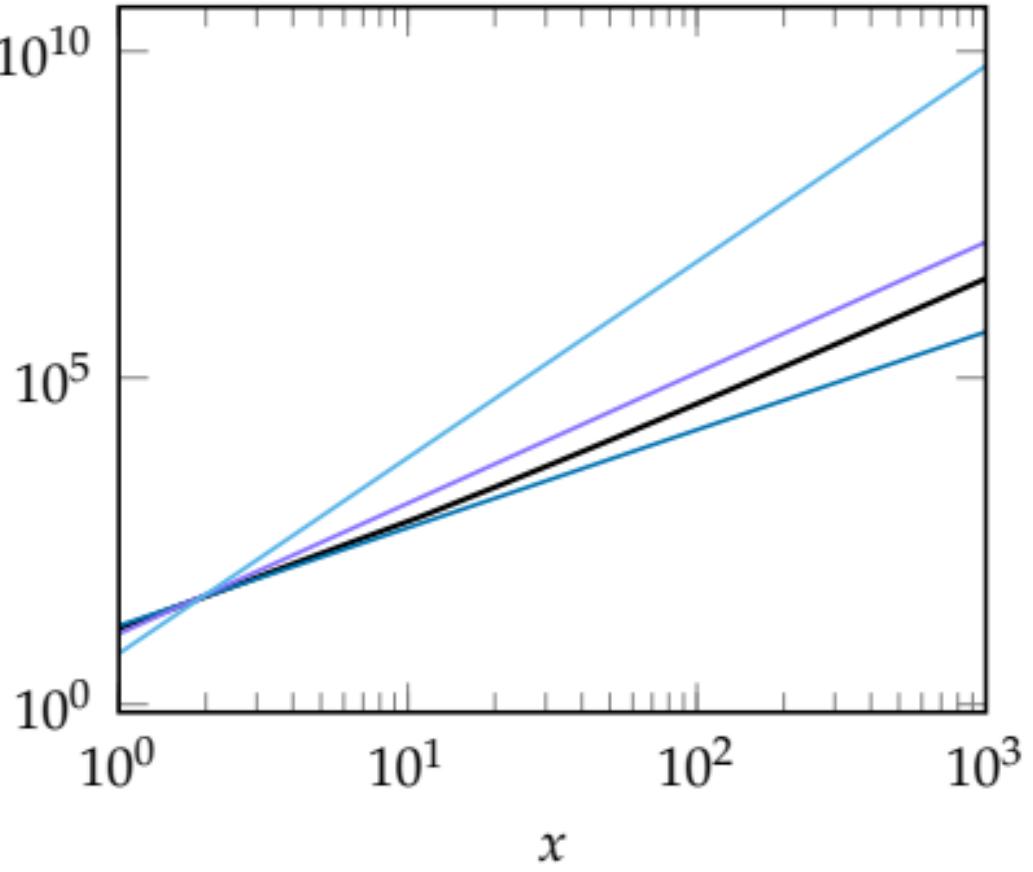
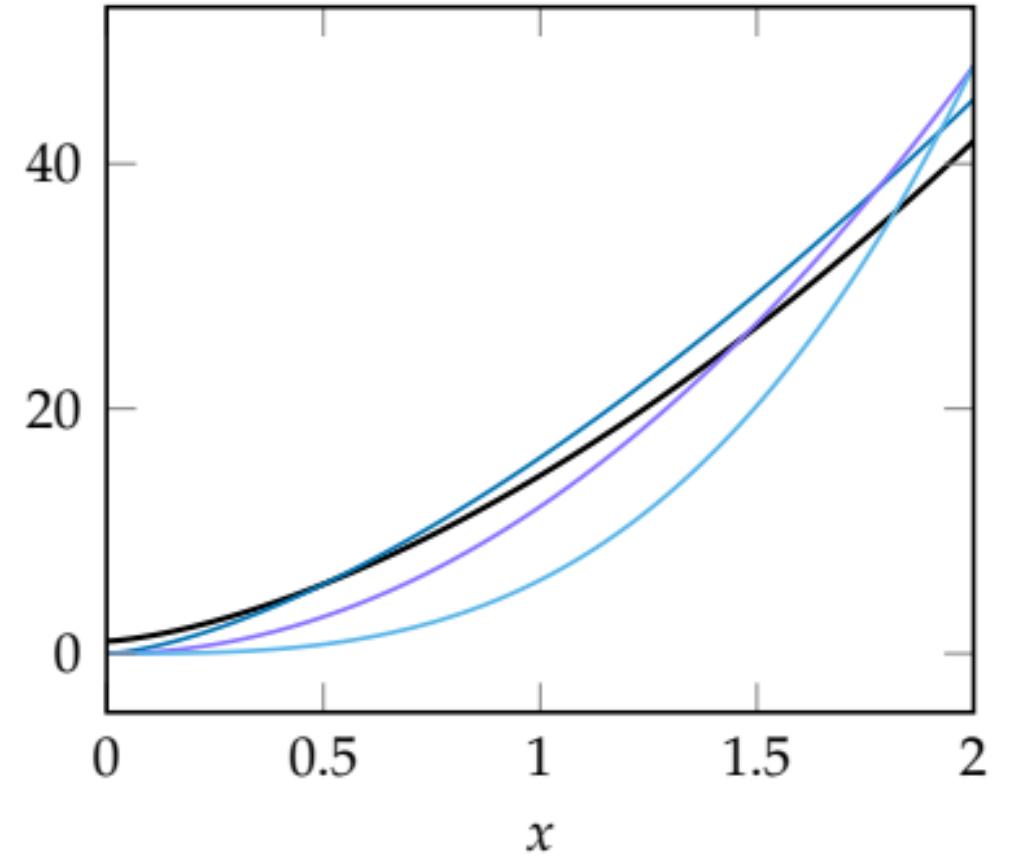


$$\alpha = [5, 5, 5]$$



$$\alpha = [2, 1, 5]$$





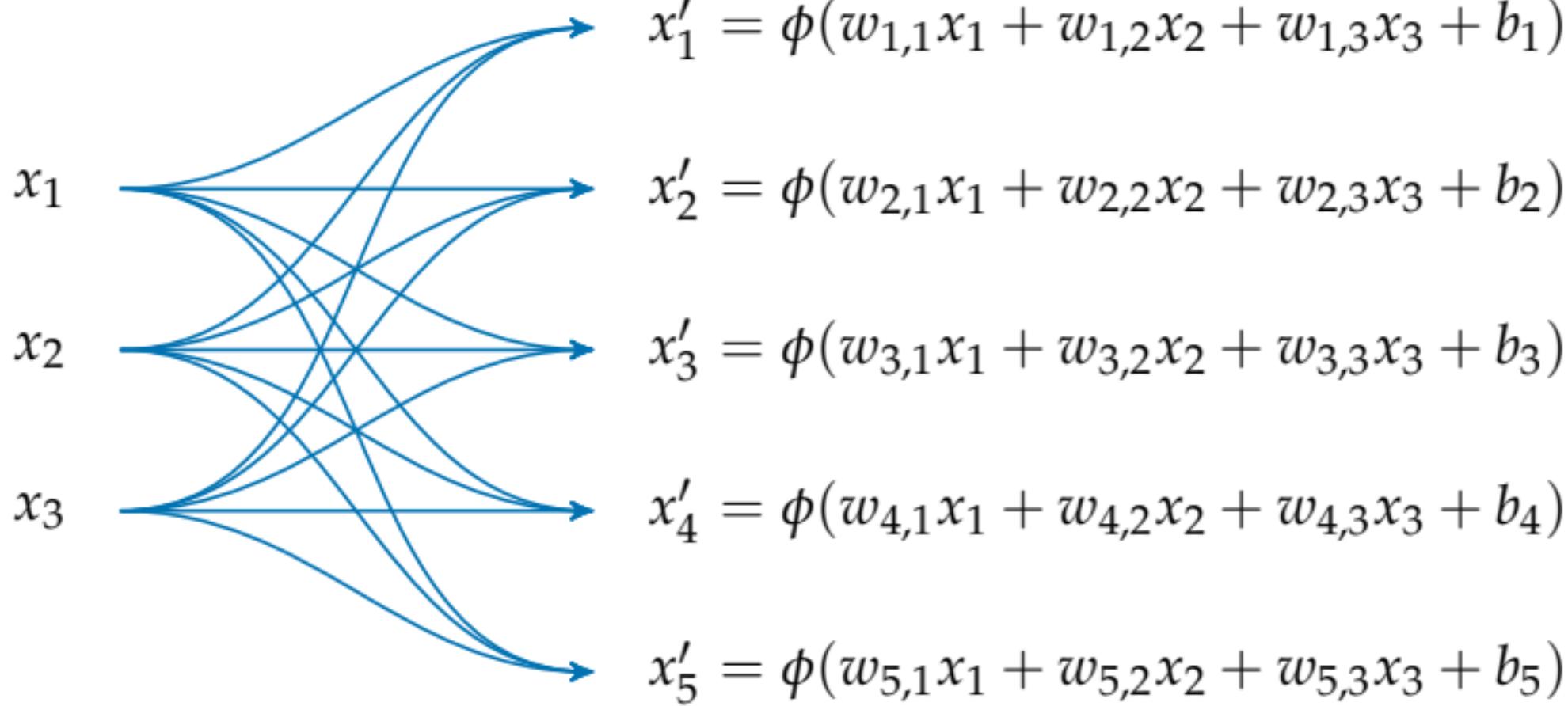
— $f(x)$ — $16x^{3/2}$ — $12x^2$ — $6x^3$

NP-hard

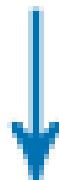
NP-complete

NP

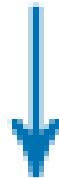
P



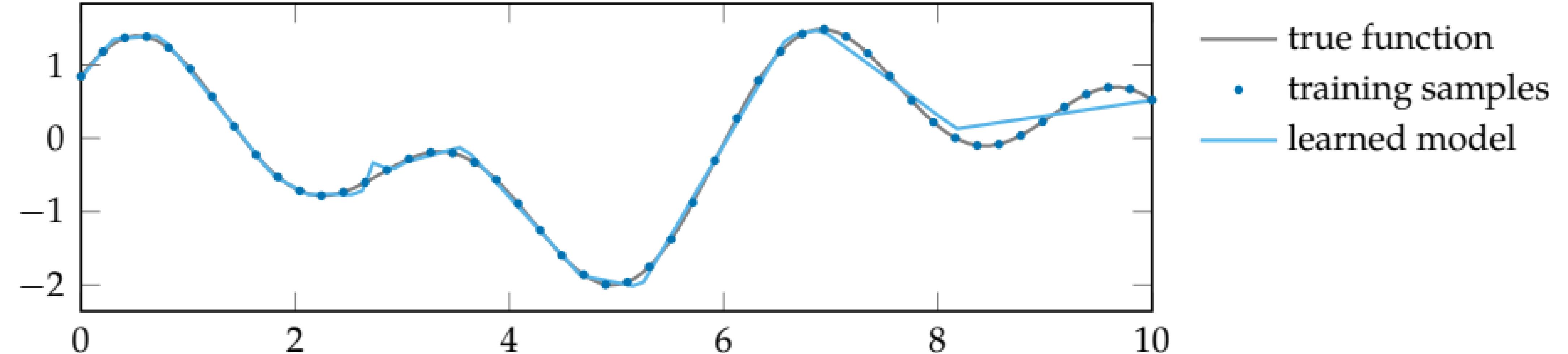
$$\mathbf{x} \in \mathbb{R}^3$$

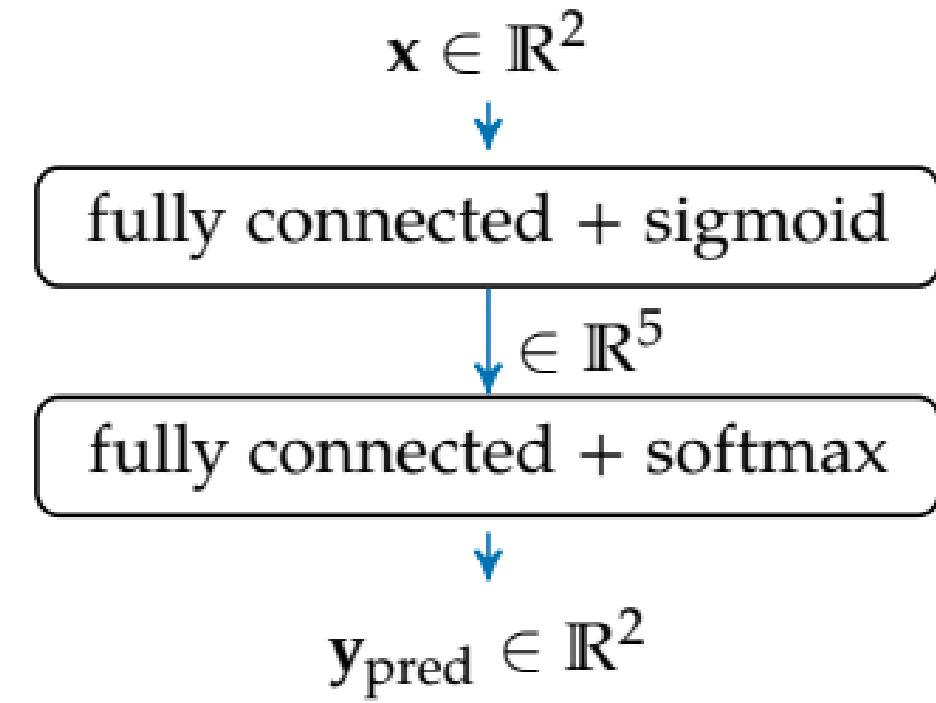
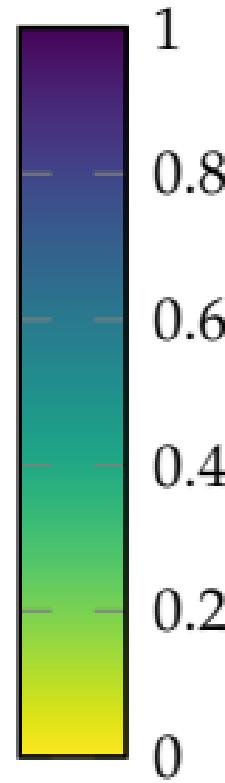
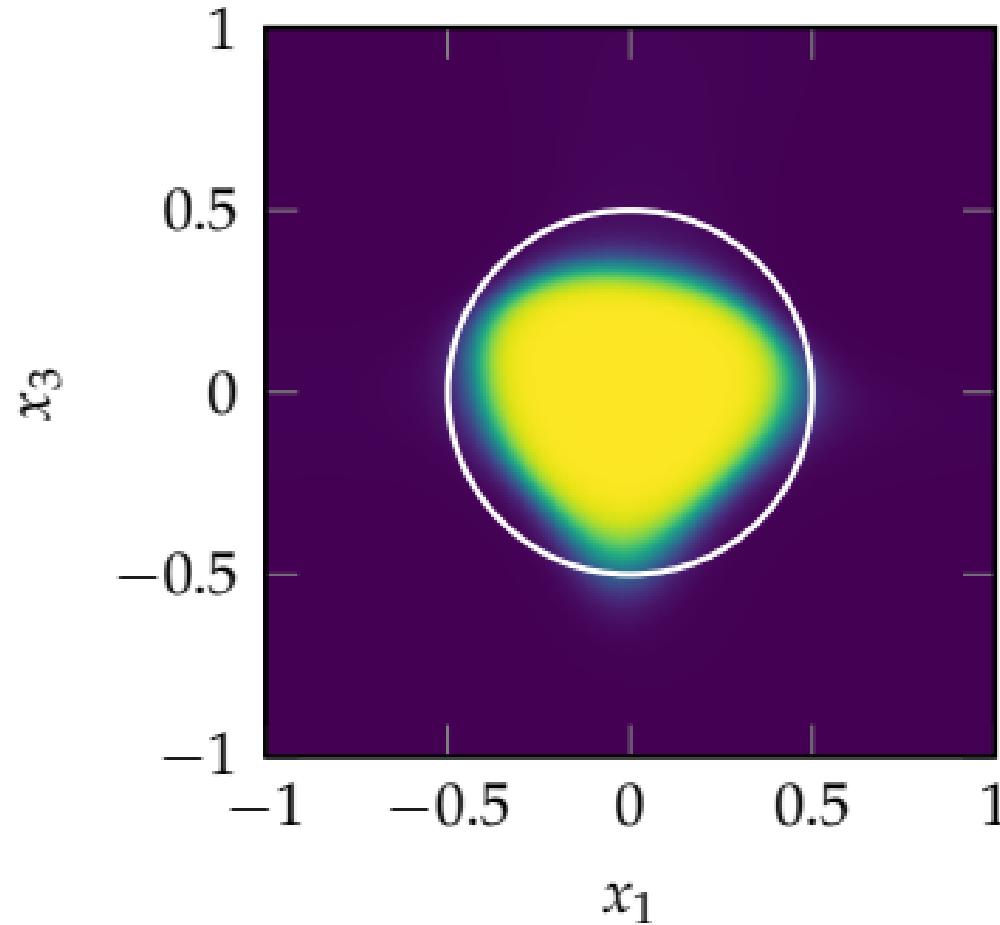


fully connected + ϕ



$$\mathbf{x}' \in \mathbb{R}^5$$





sigmoid

$$1/(1 + \exp(-x))$$

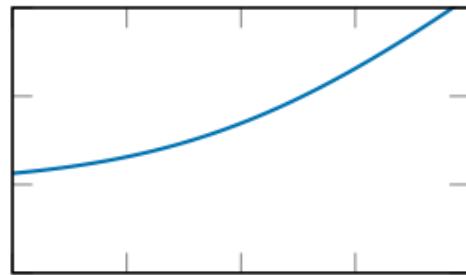
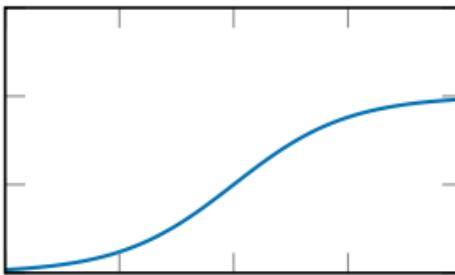
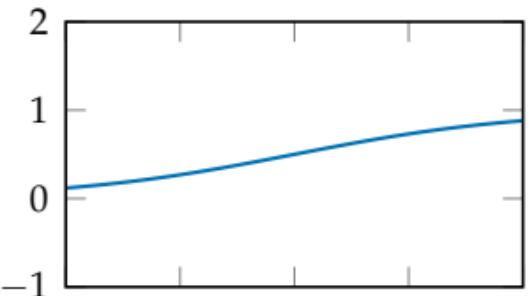
$\phi(x)$

tanh

$$\tanh(x)$$

softplus

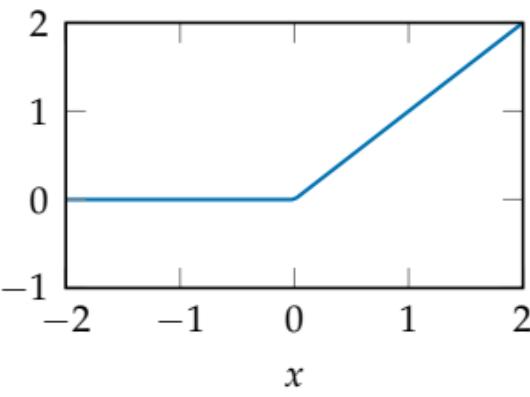
$$\log(1 + \exp(x))$$



relu

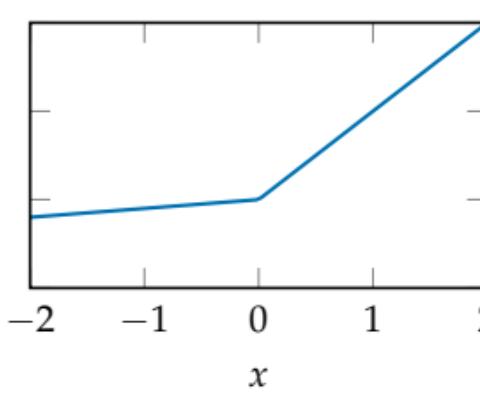
$$\max(0, x)$$

$\phi(x)$



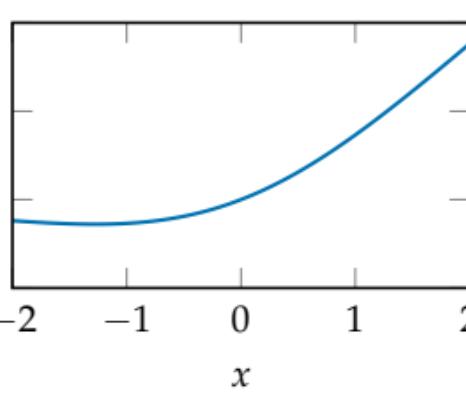
leaky relu

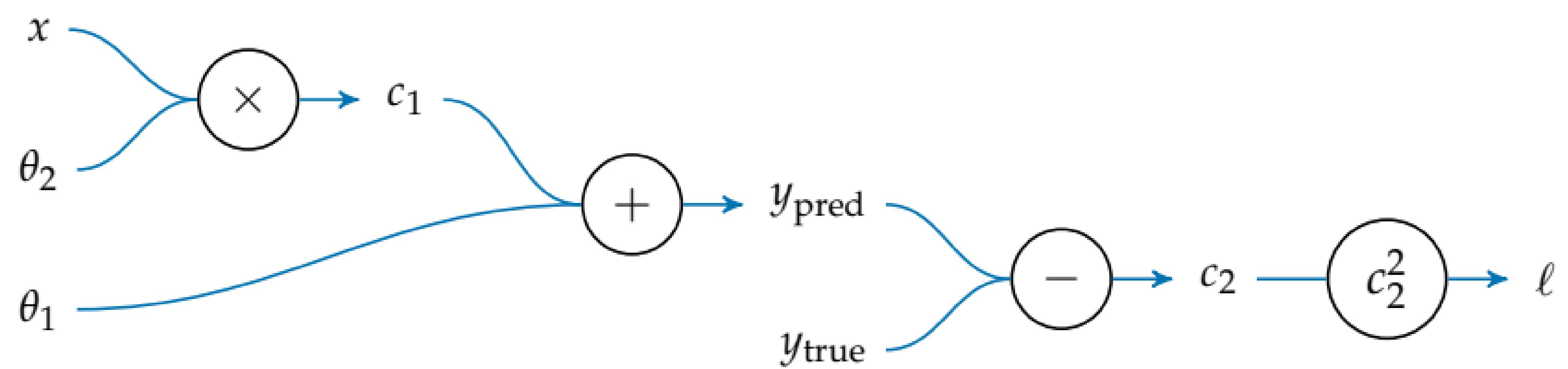
$$\max(\alpha x, x)$$

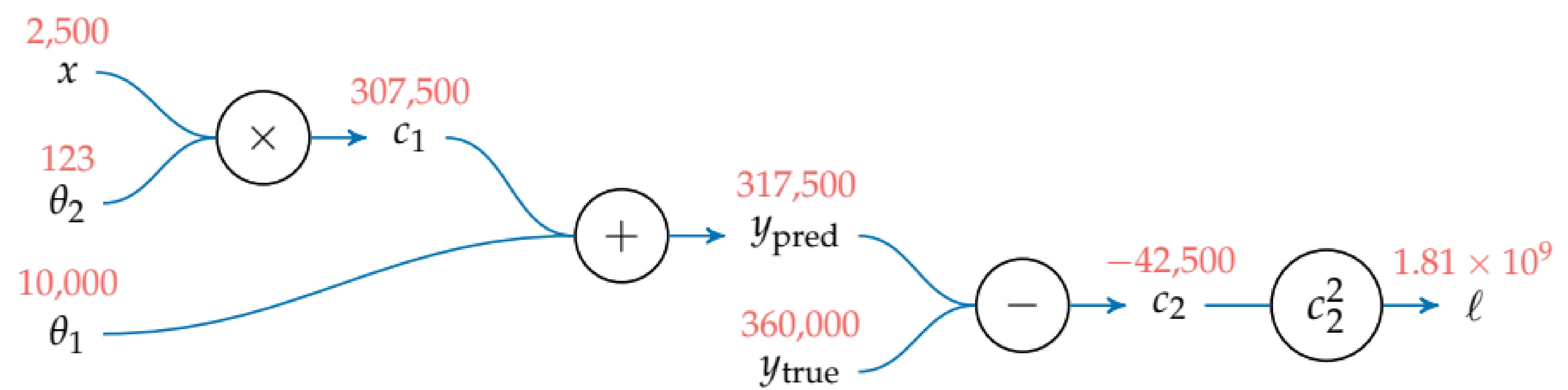


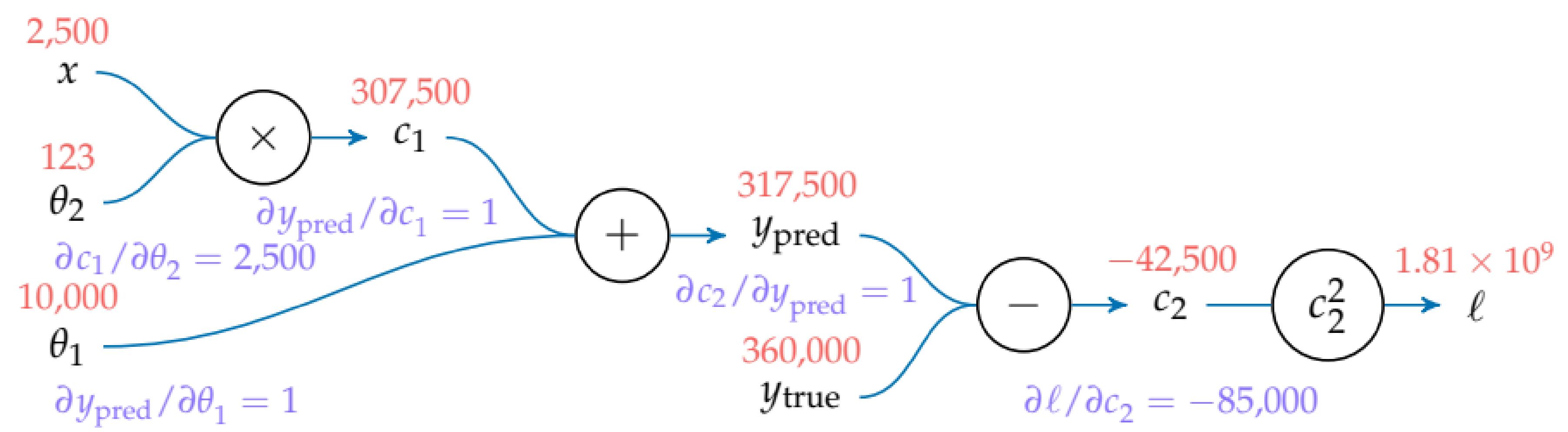
swish

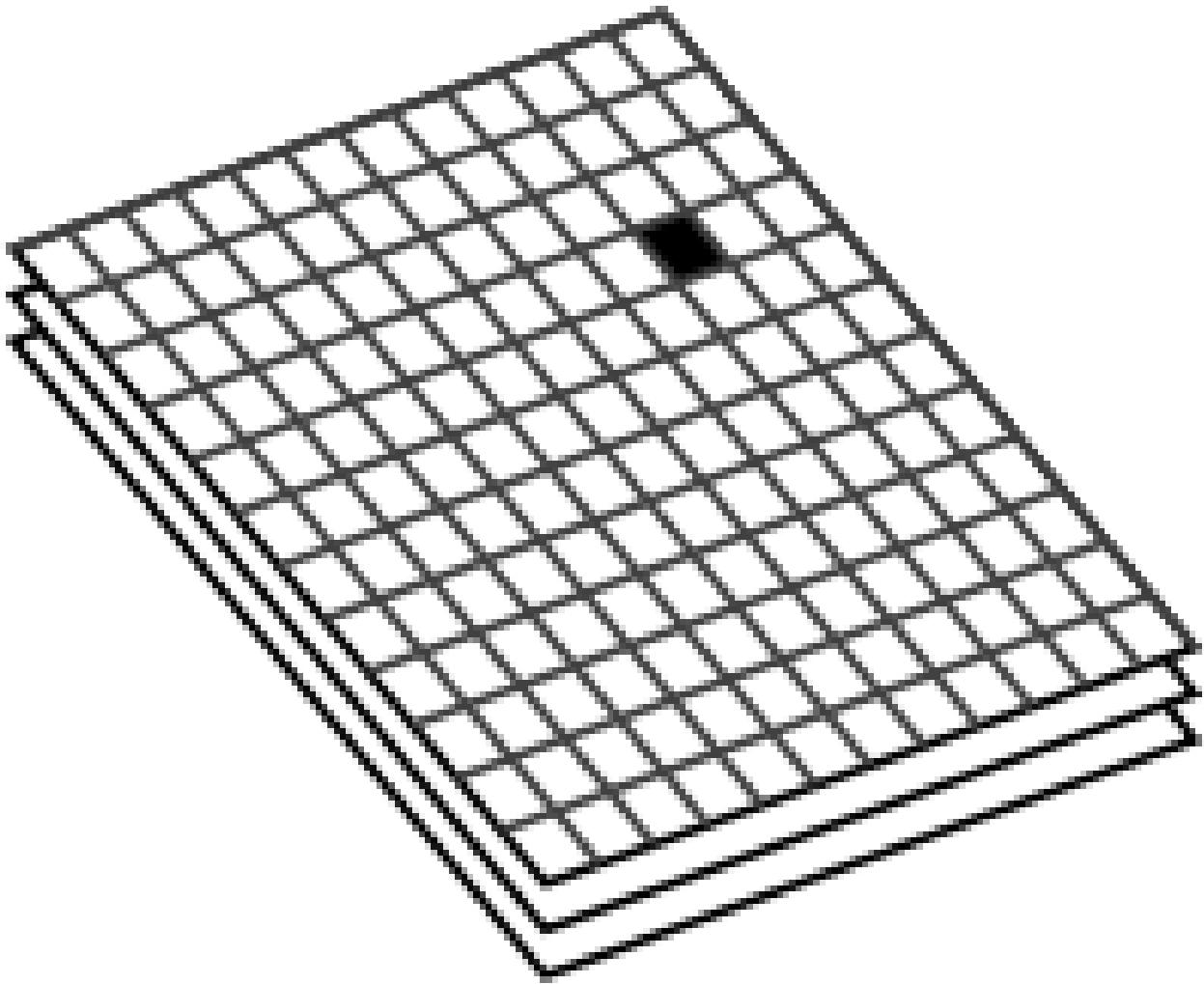
$$x \operatorname{sigmoid}(x)$$



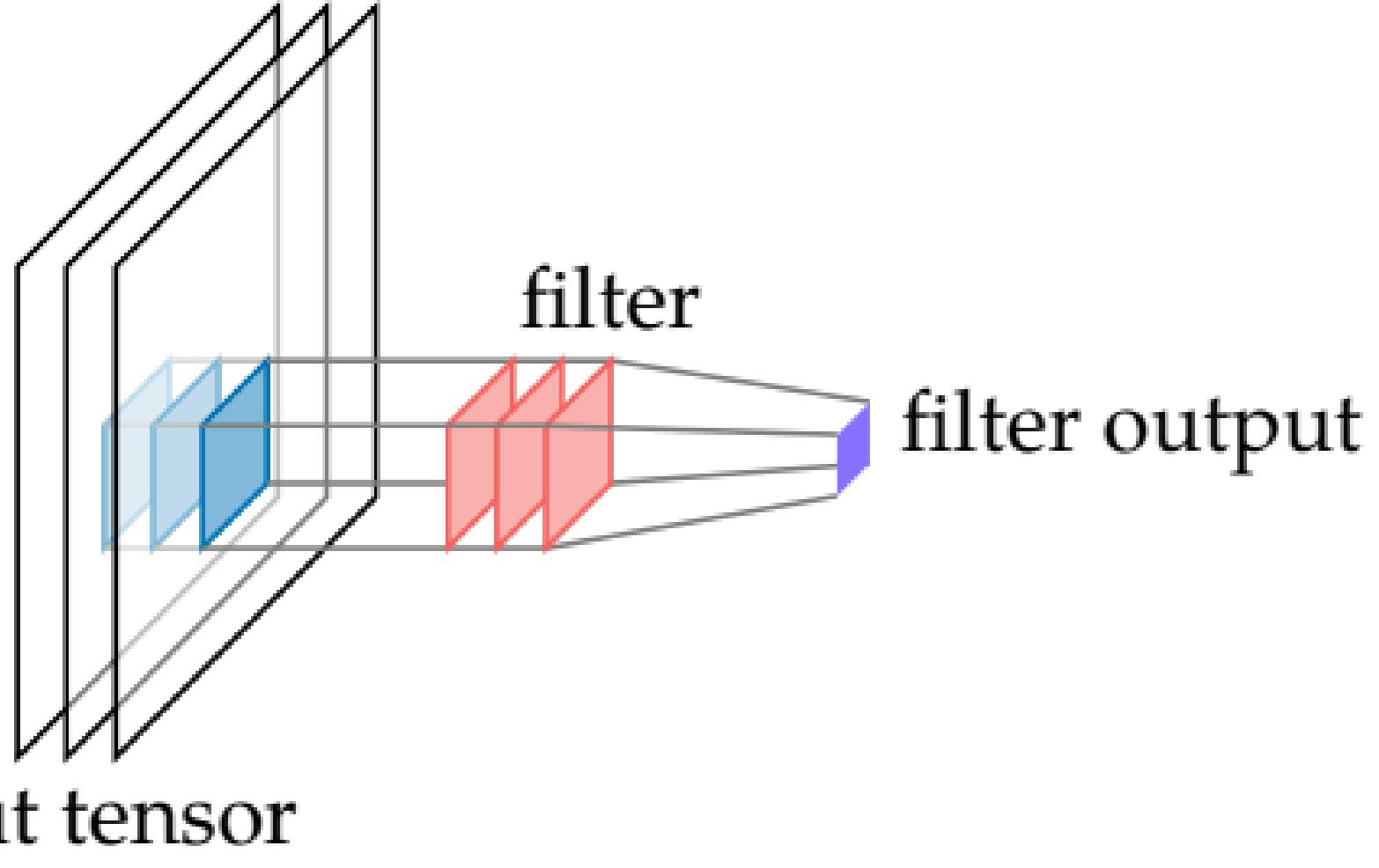




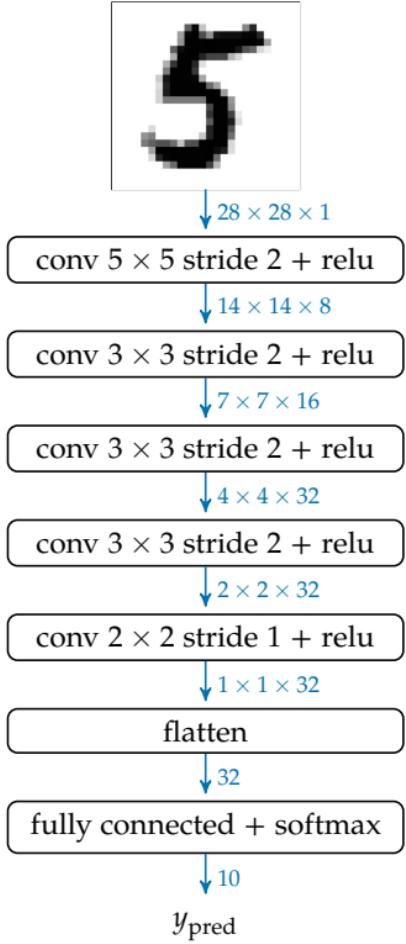




receptive field



The MNIST data set contains handwritten digits in the form of 28×28 monochromatic images. It is often used to test image classification networks. To the right, we have a sample convolutional neural network that takes an MNIST image as input and produces a categorical probability distribution over the 10 possible digits. Convolutional layers are used to efficiently extract features. The model shrinks in the first two dimensions and expands in the third dimension (the number of features) as the network depth increases. Eventually reaching a first and second dimension of 1 ensures that information from across the entire image can affect every feature. The flatten operation takes the $1 \times 1 \times 32$ input and flattens it into a 32-component output. Such operations are common when transitioning between convolutional and fully connected layers. This model has 19,722 parameters. The parameters can be tuned to maximize the likelihood of the training data.



$\{x_1, x_2, x_3, \dots\}$  y

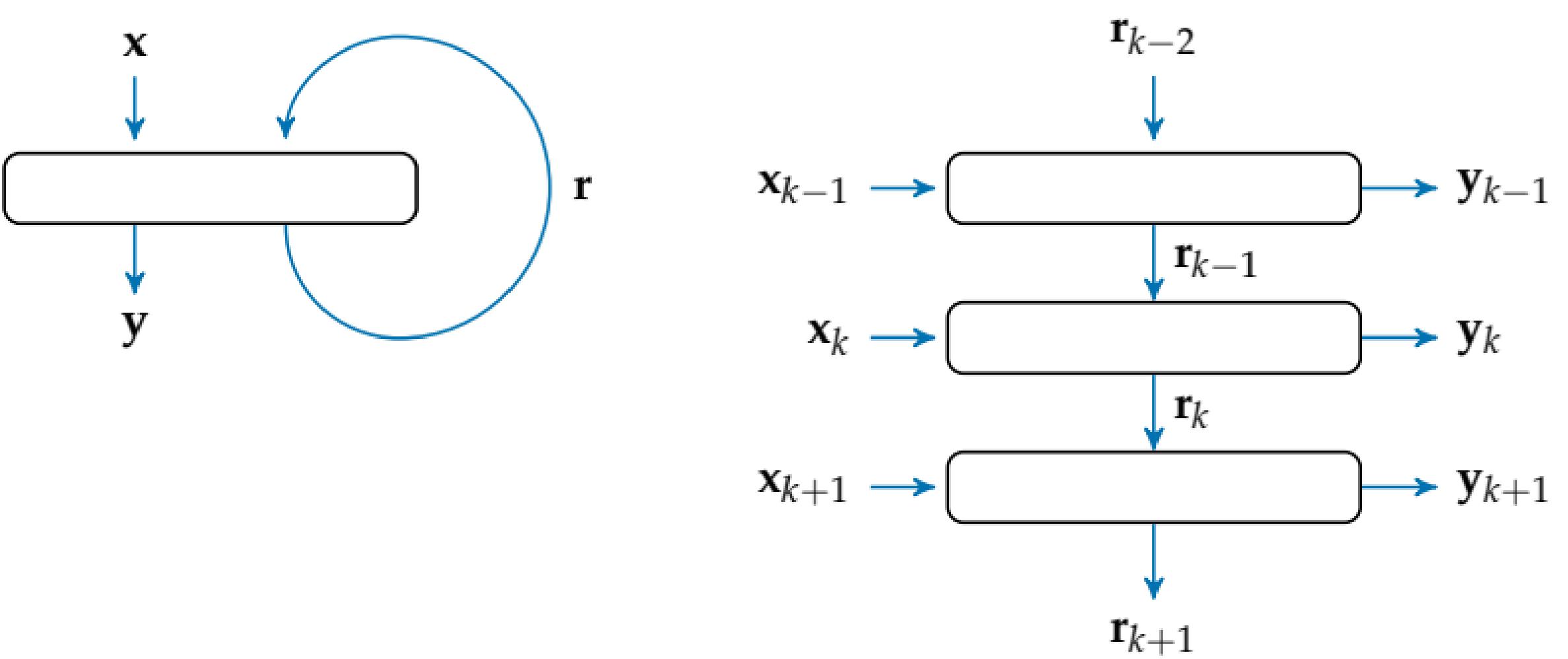
many-to-one

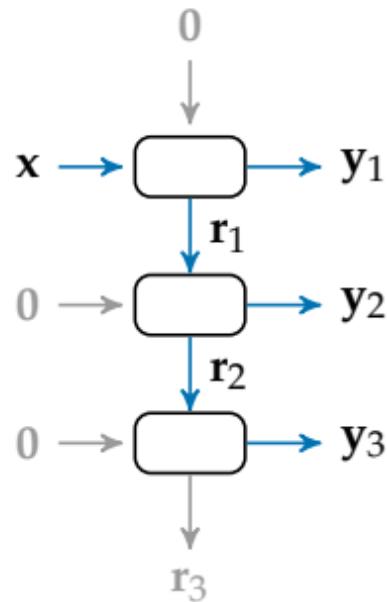
 x  $\{y_1, y_2, y_3, \dots\}$

one-to-many

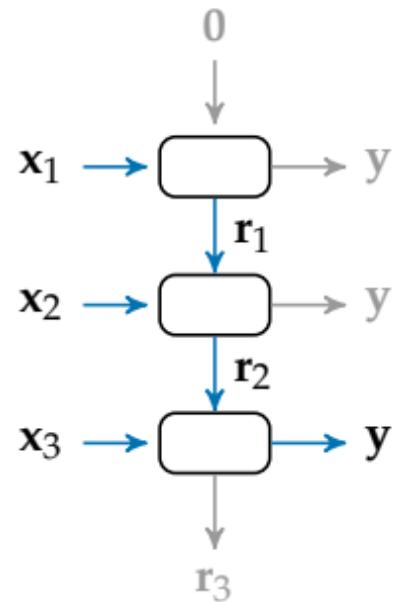
 $\{x_1, x_2, x_3, \dots\}$  $\{y_1, y_2, y_3, \dots\}$

many-to-many

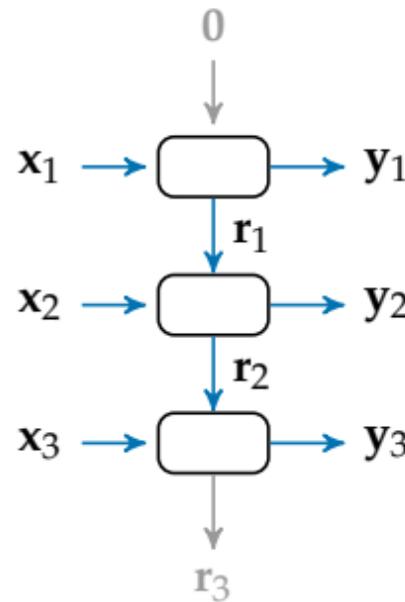




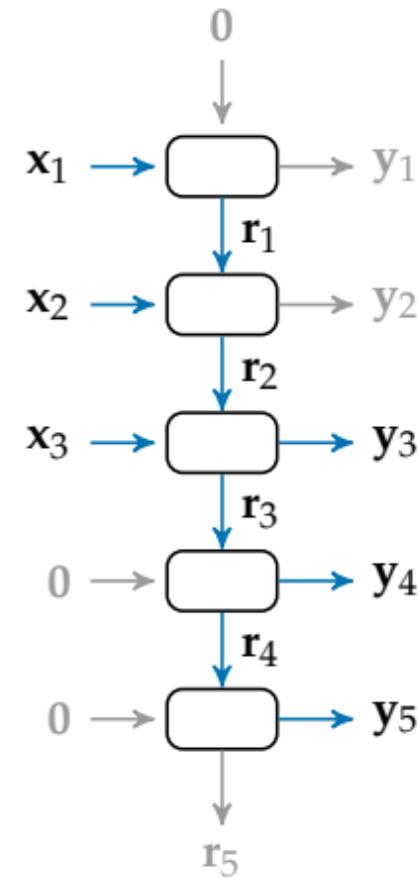
one-to-many



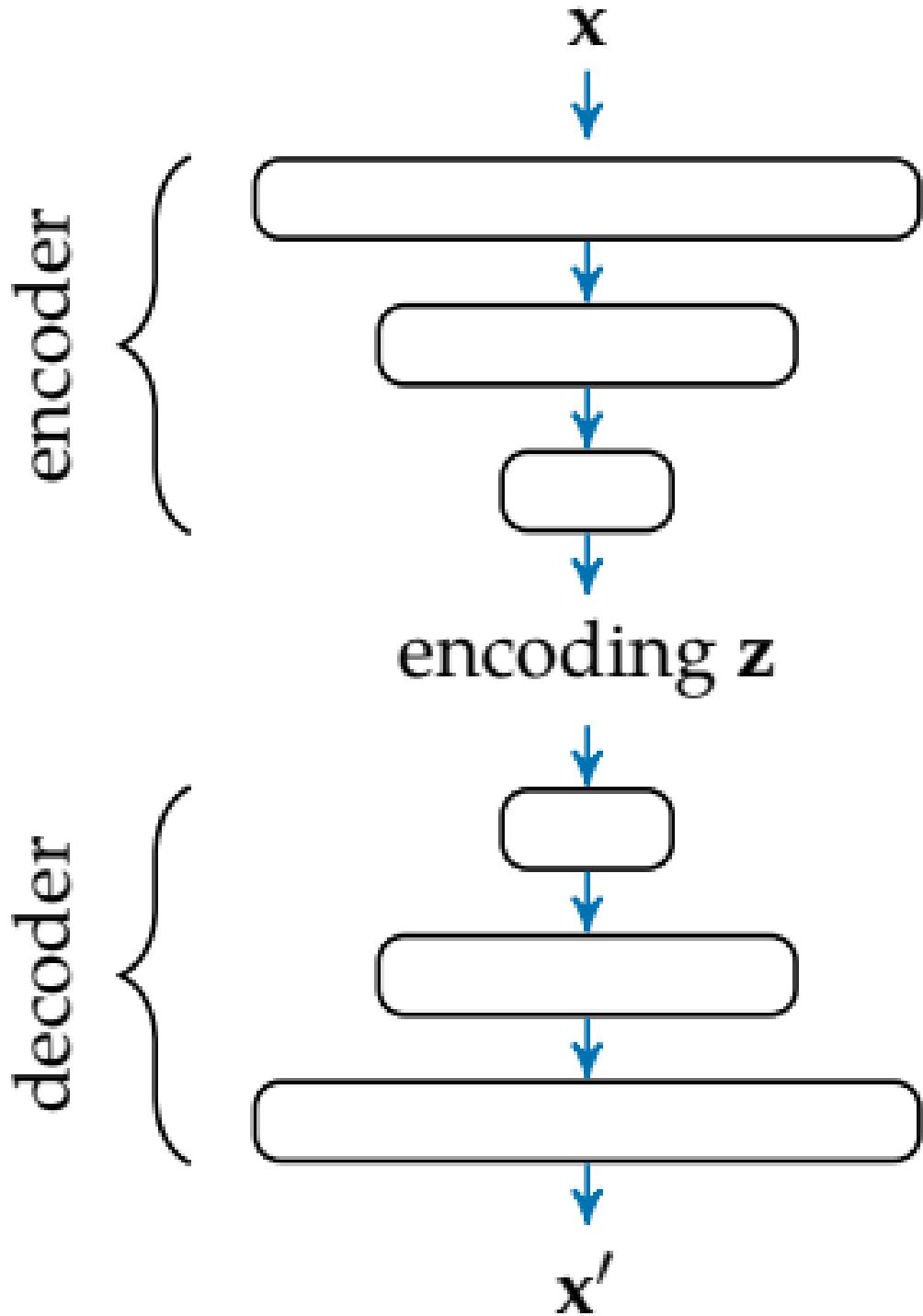
many-to-one

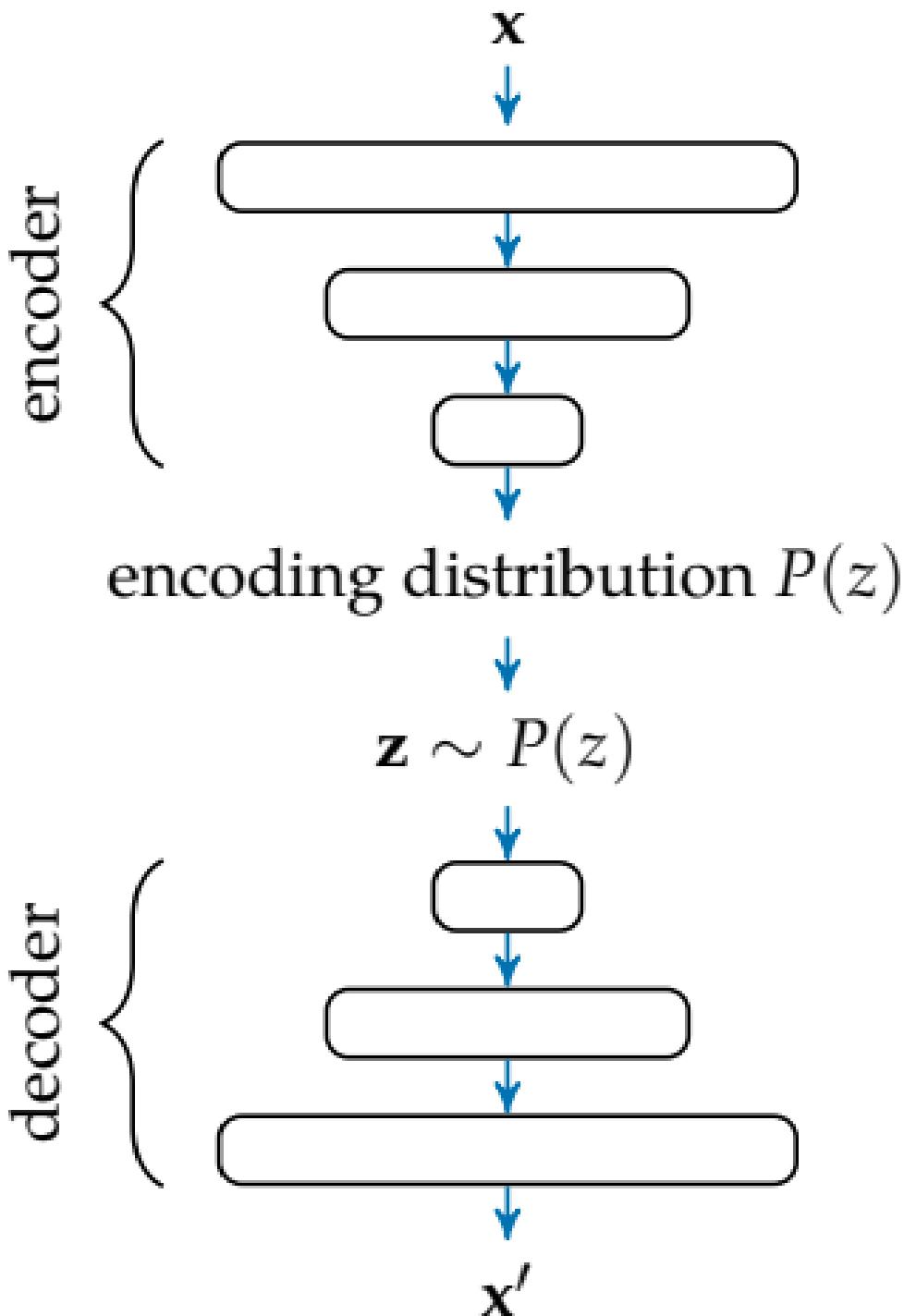


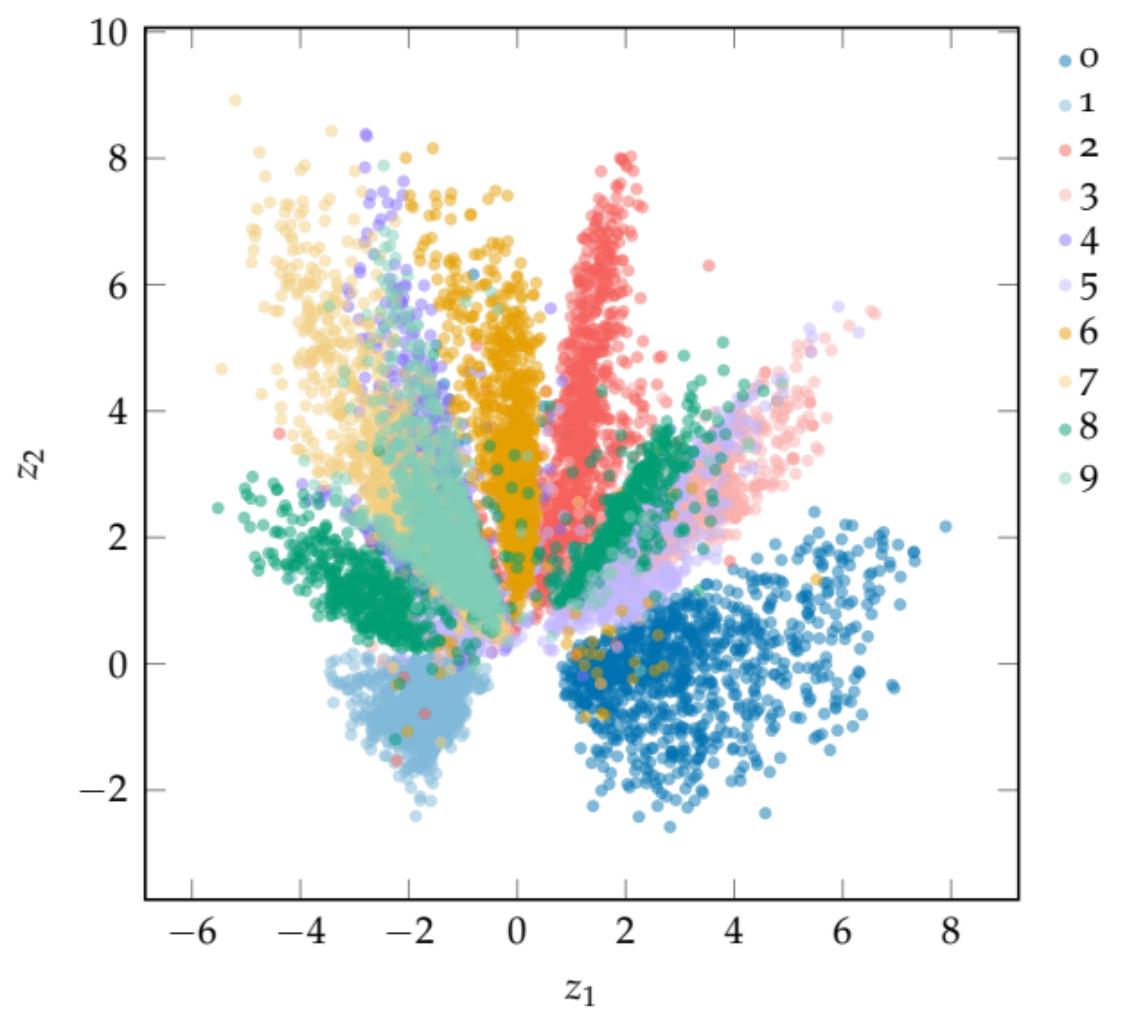
many-to-many

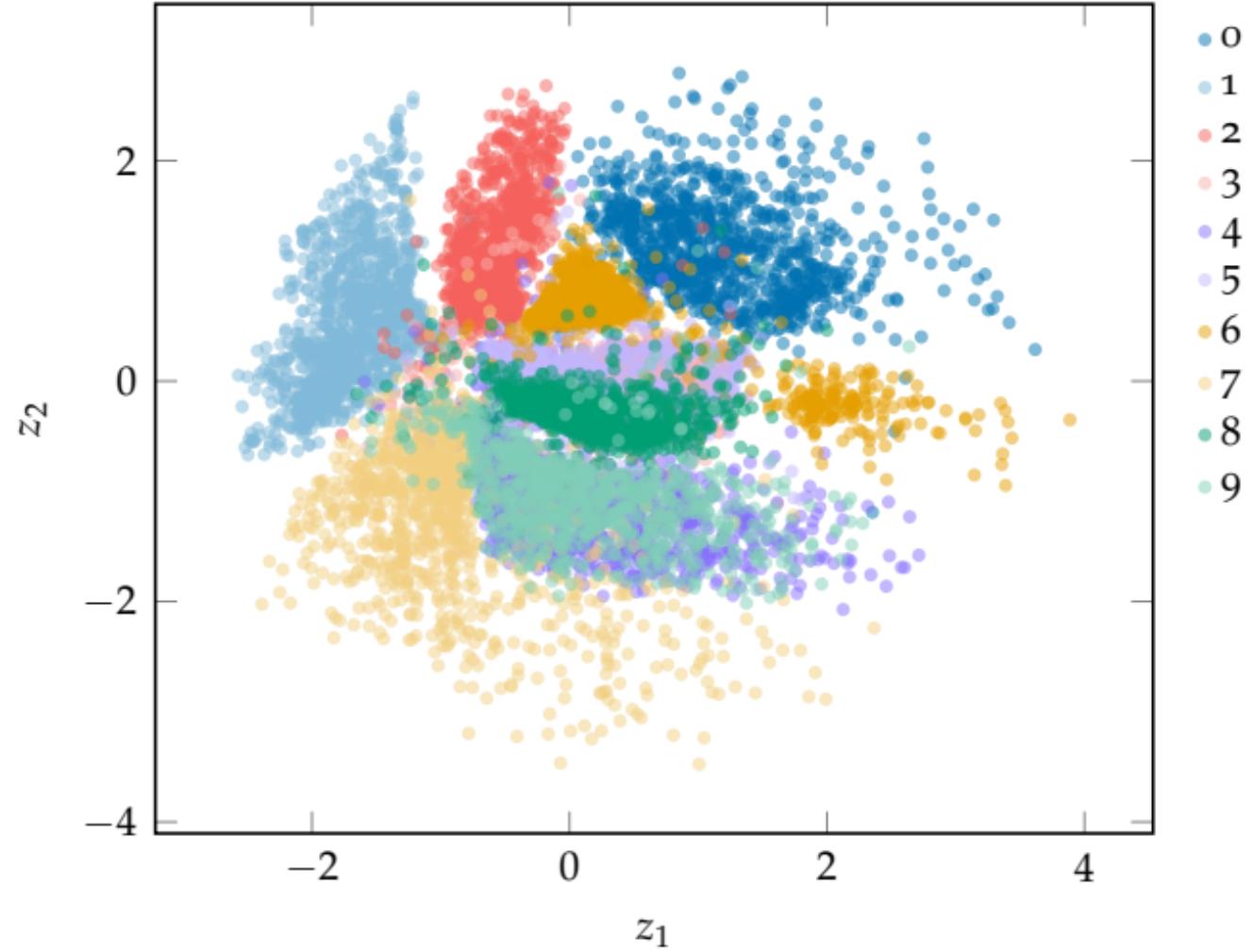


many-to-many

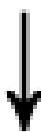








x



primary network

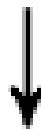


y

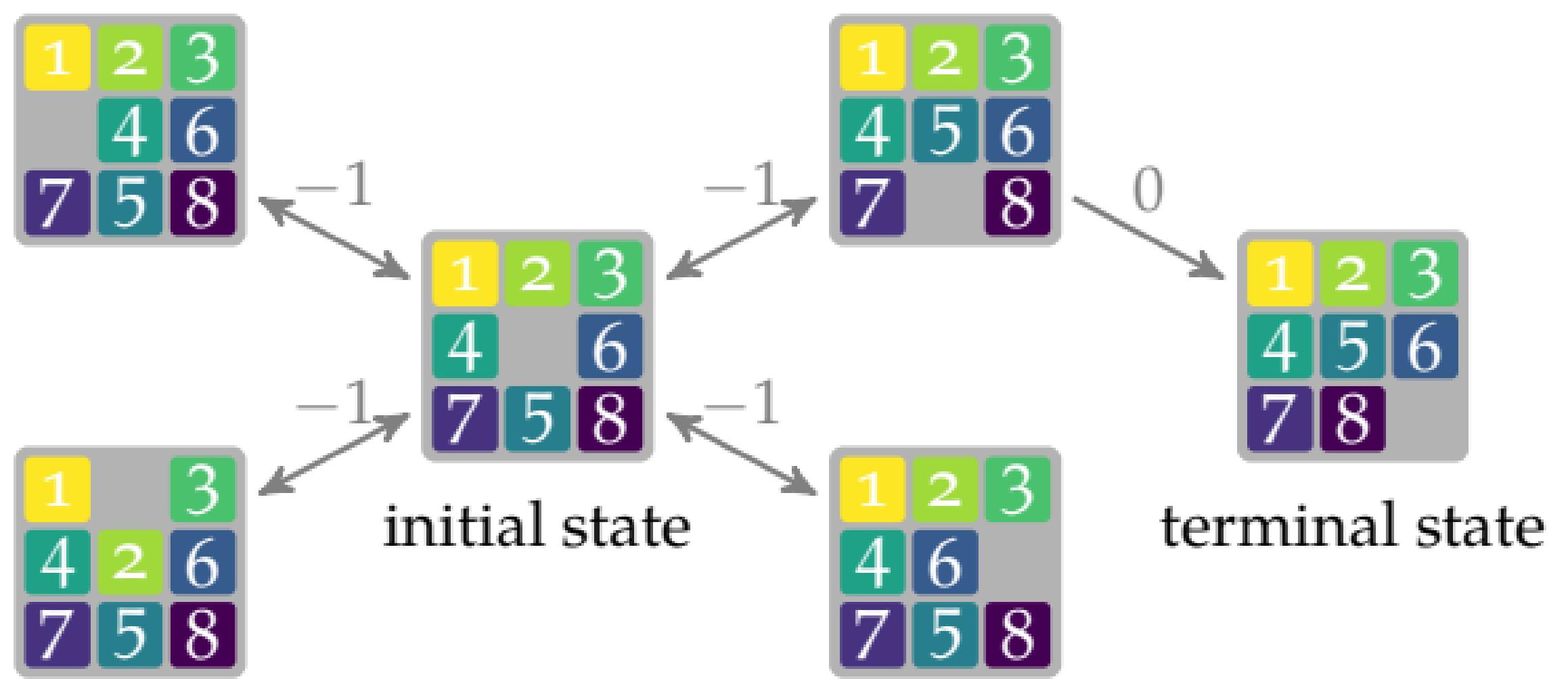
y_{true}



discriminator



P(true)



1	2	3
4		6
7	5	8

-1

-1

-1

-1

1	2	3
4	5	6
7		8

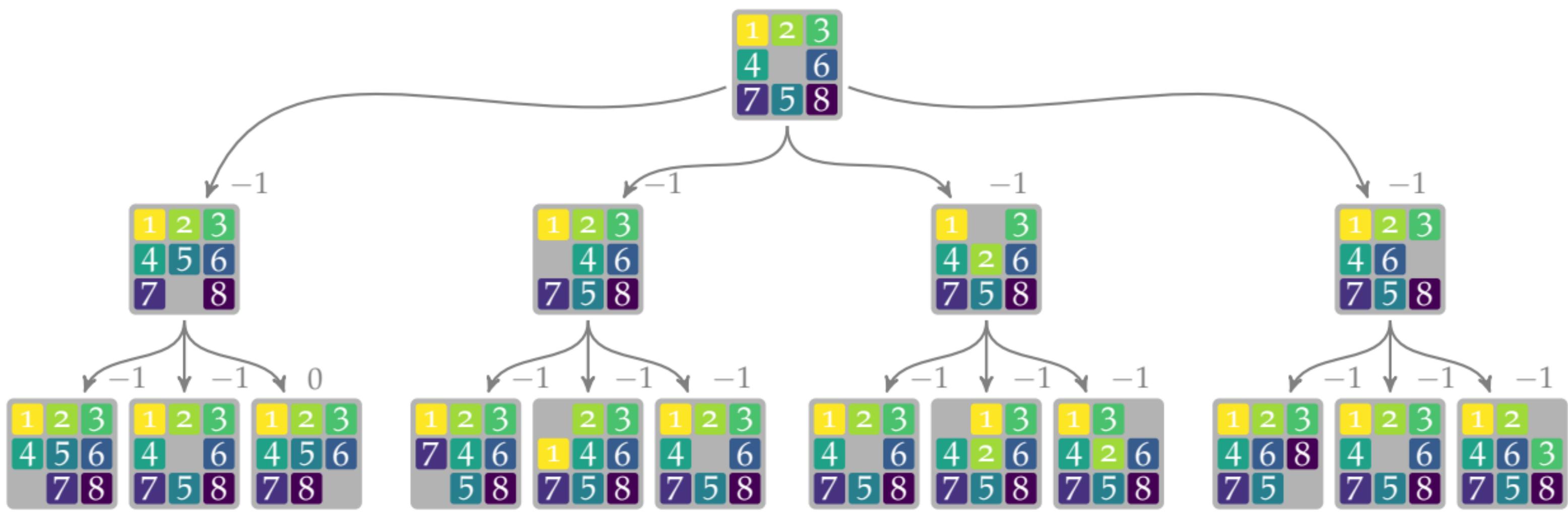
1	2	3
	4	6
7	5	8

1		3
4	2	6
7	5	8

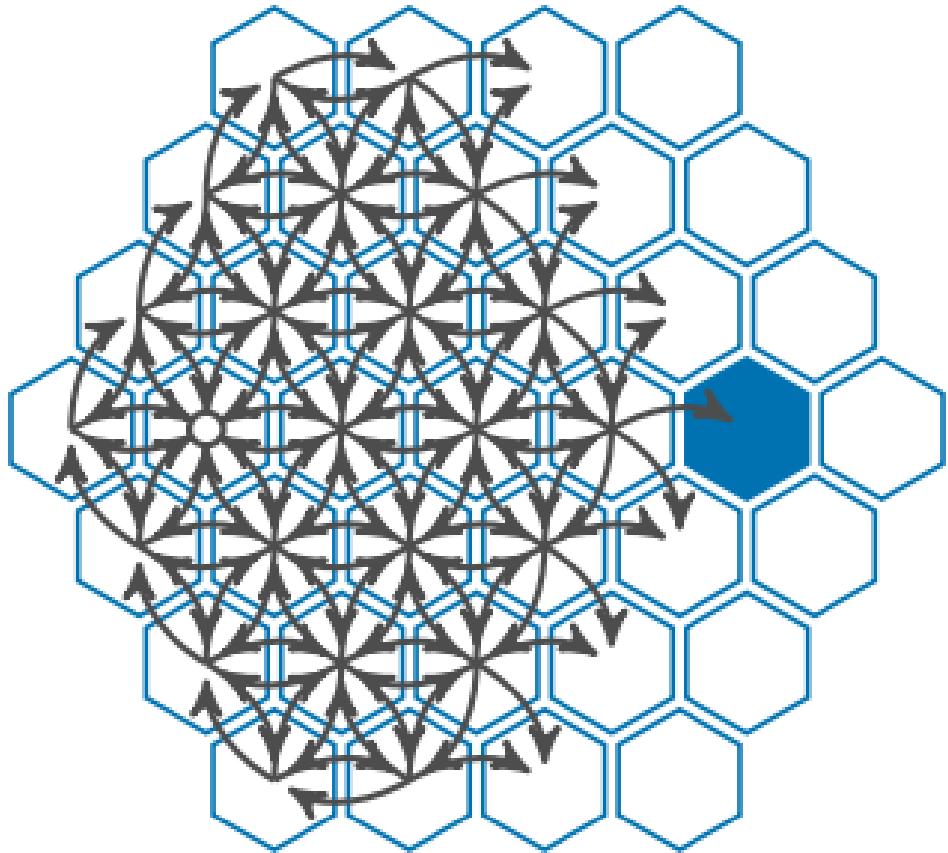
1	2	3
4	6	
7	5	8

↓ 0

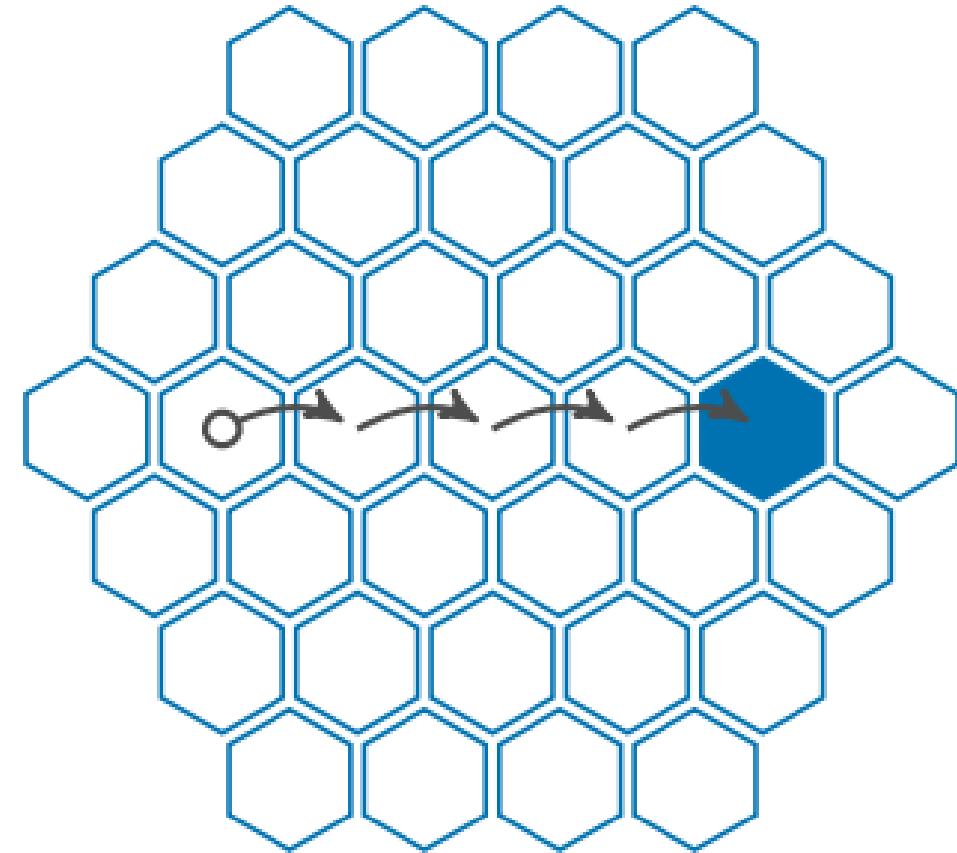
1	2	3
4	5	6
7	8	

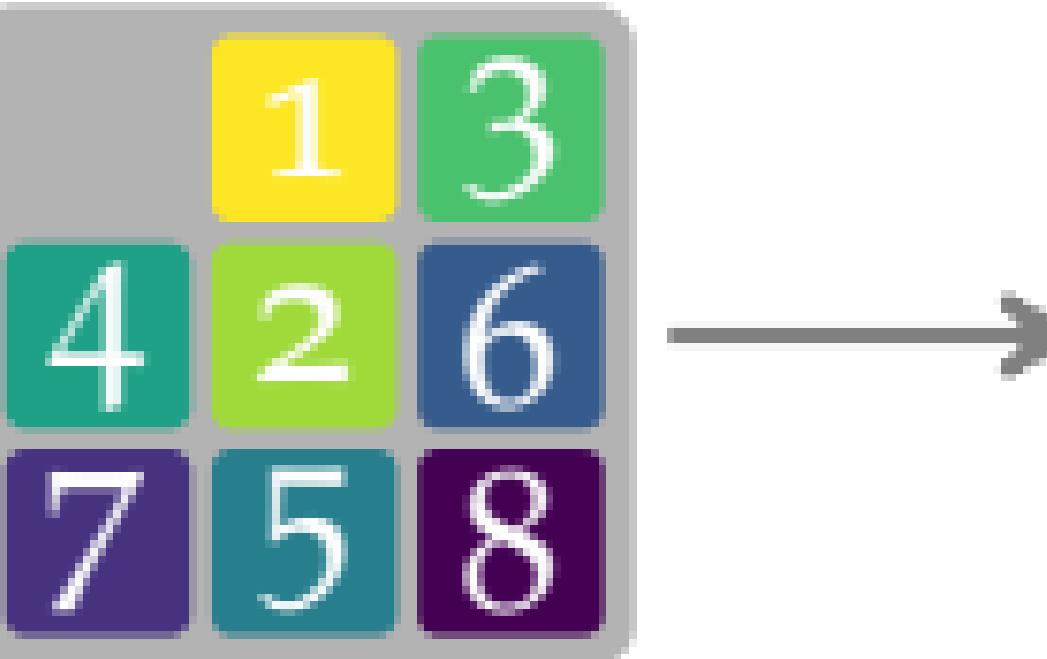


forward search

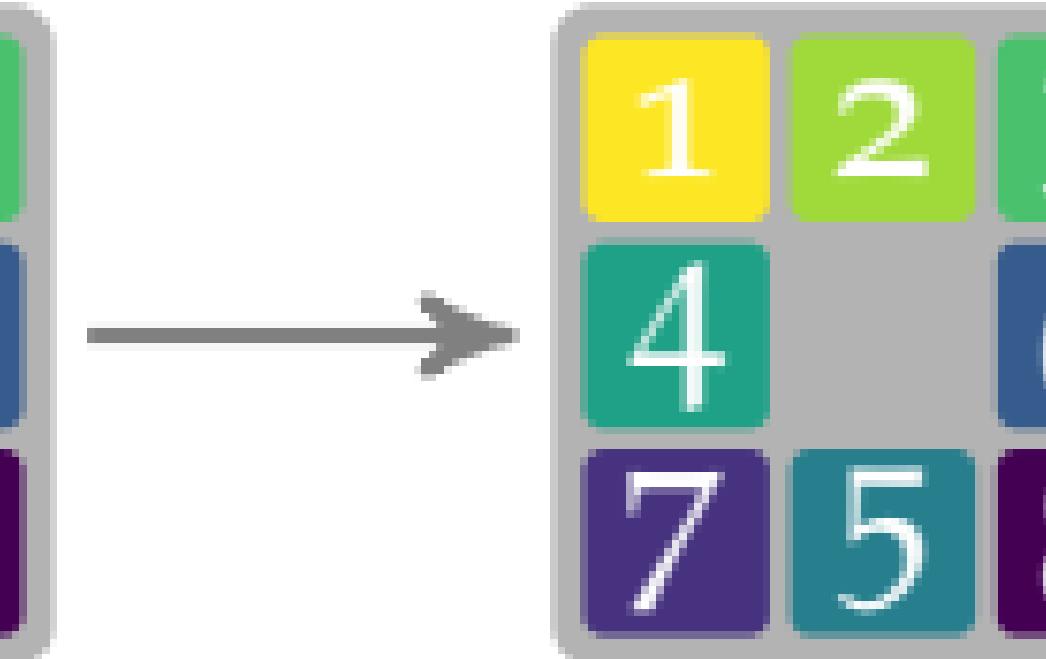


branch and bound

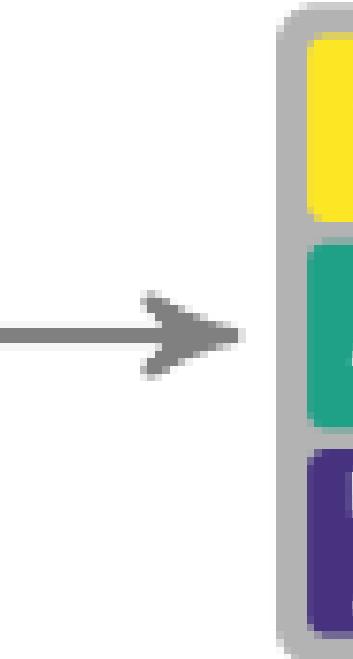
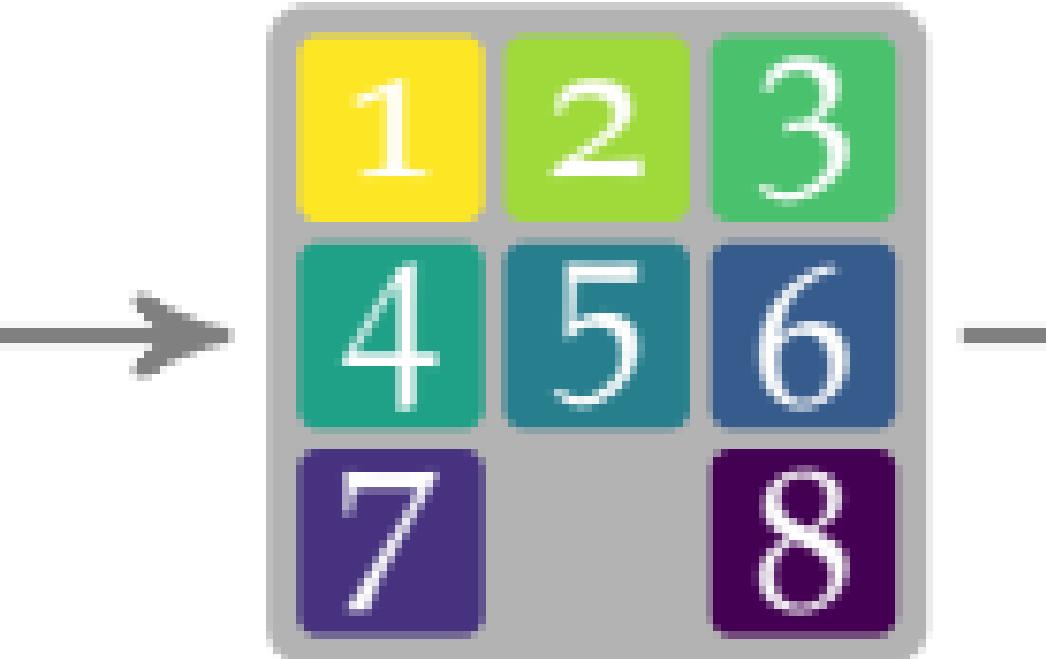




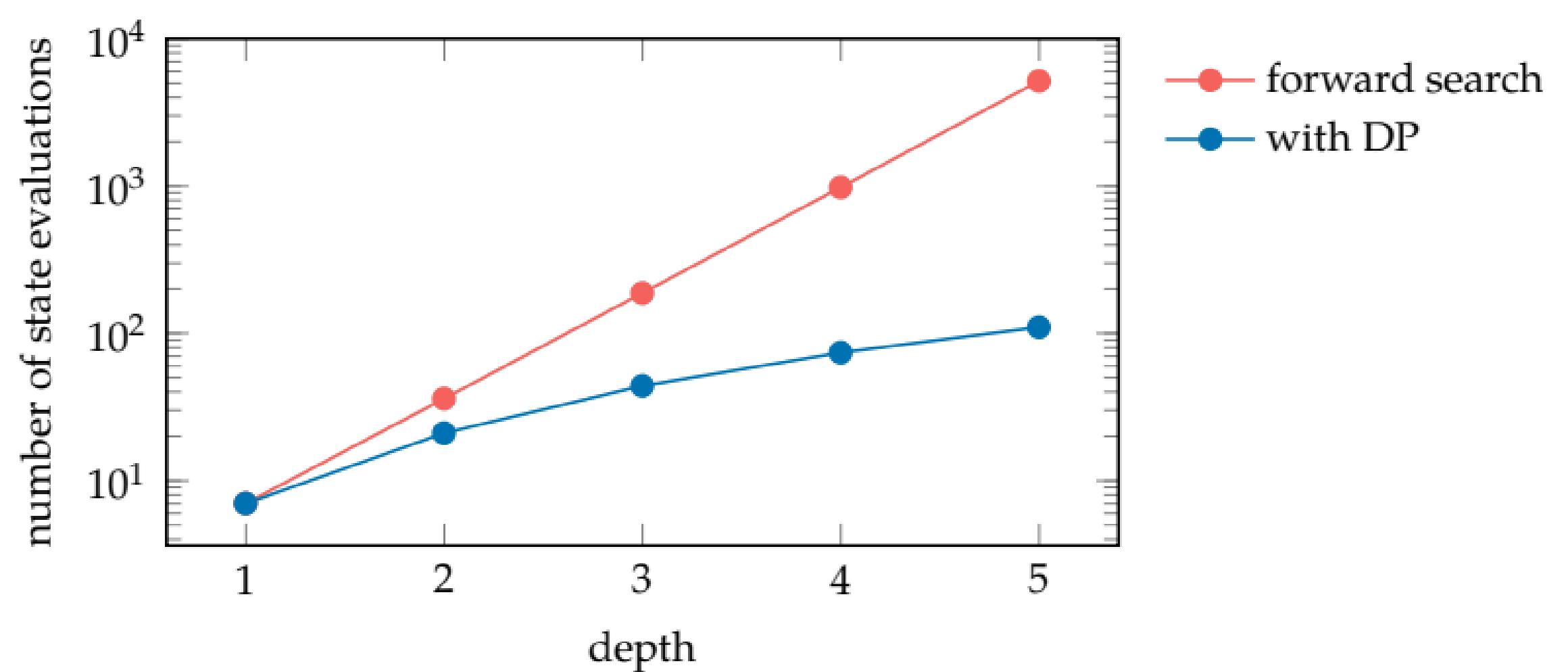
initial state



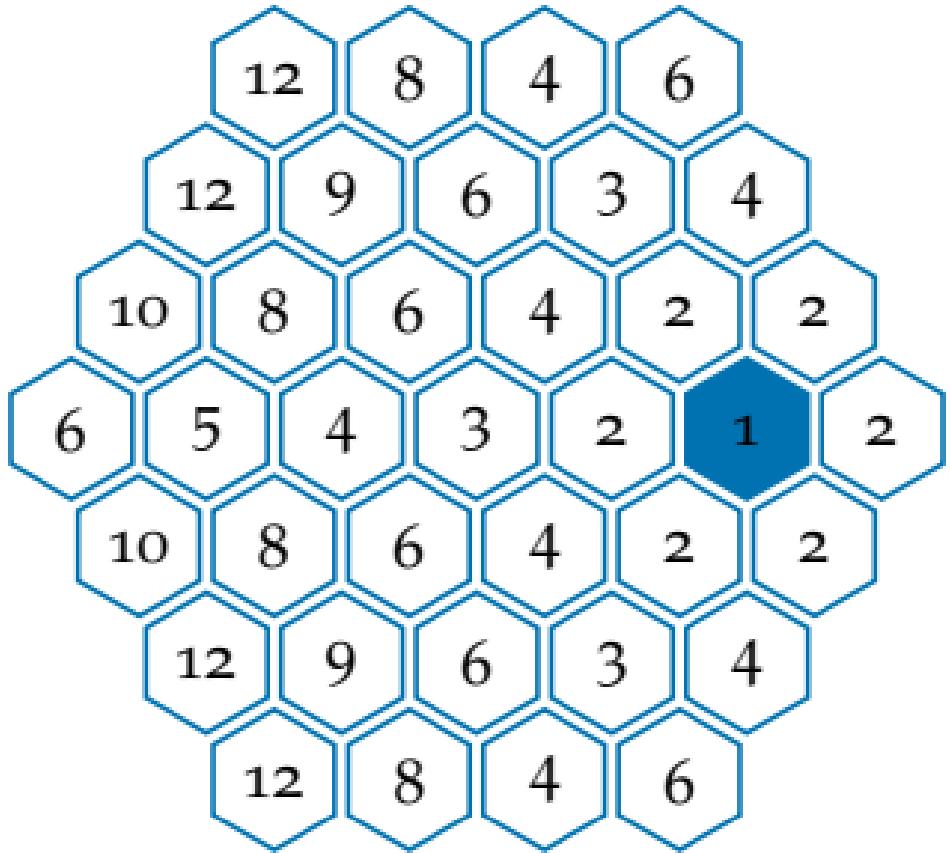
intermediate state



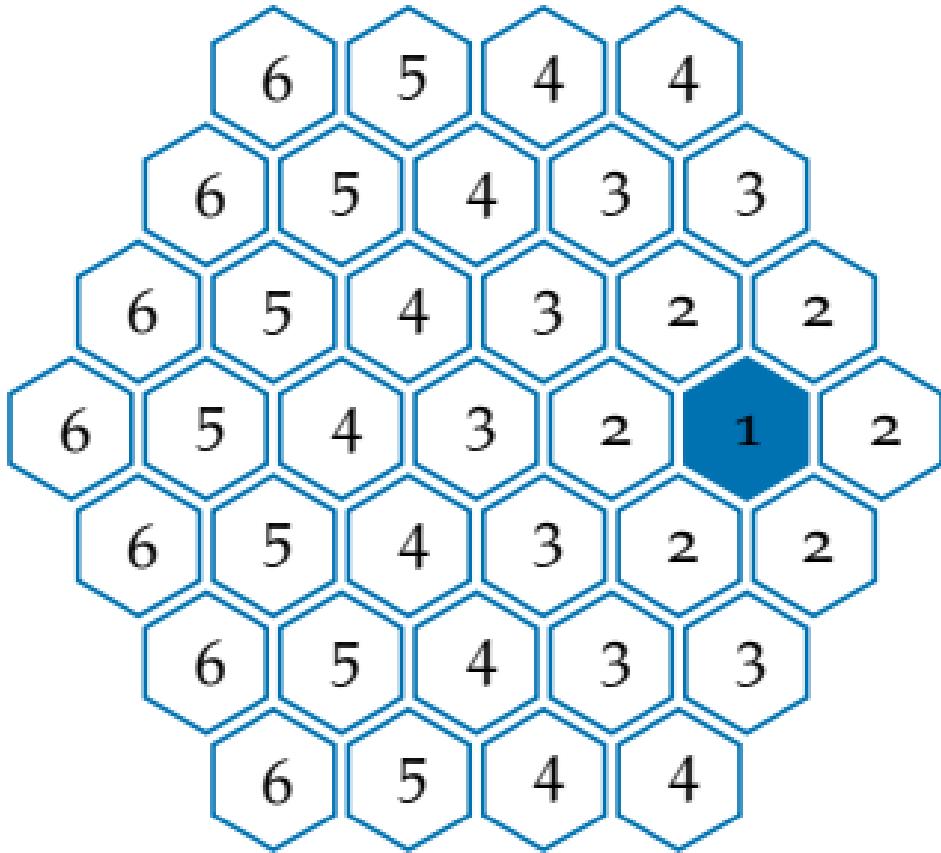
terminal state

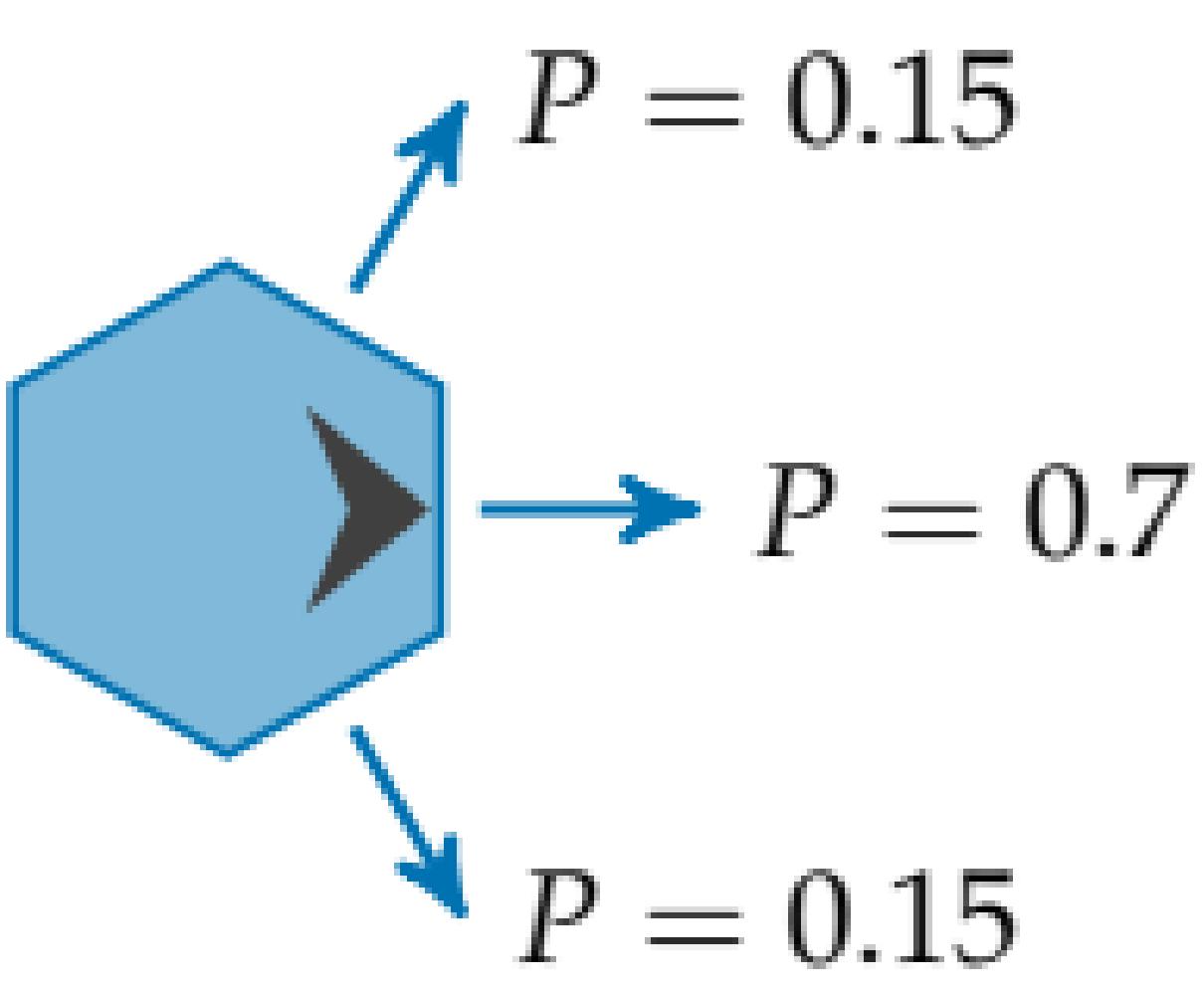


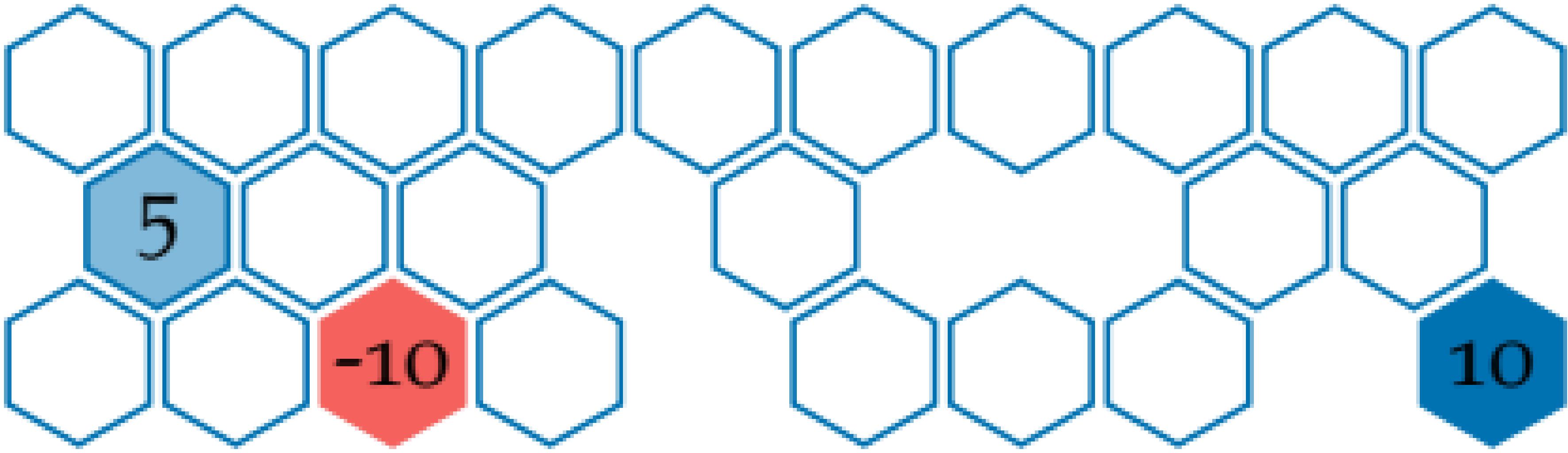
branch and bound

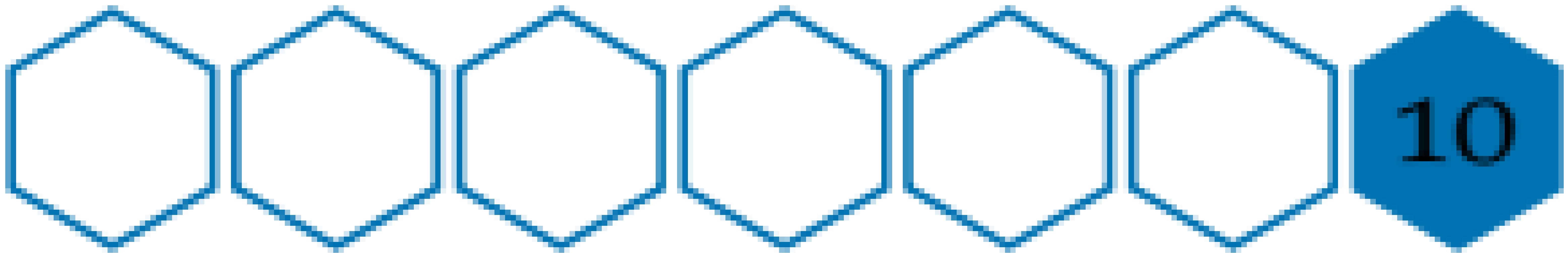


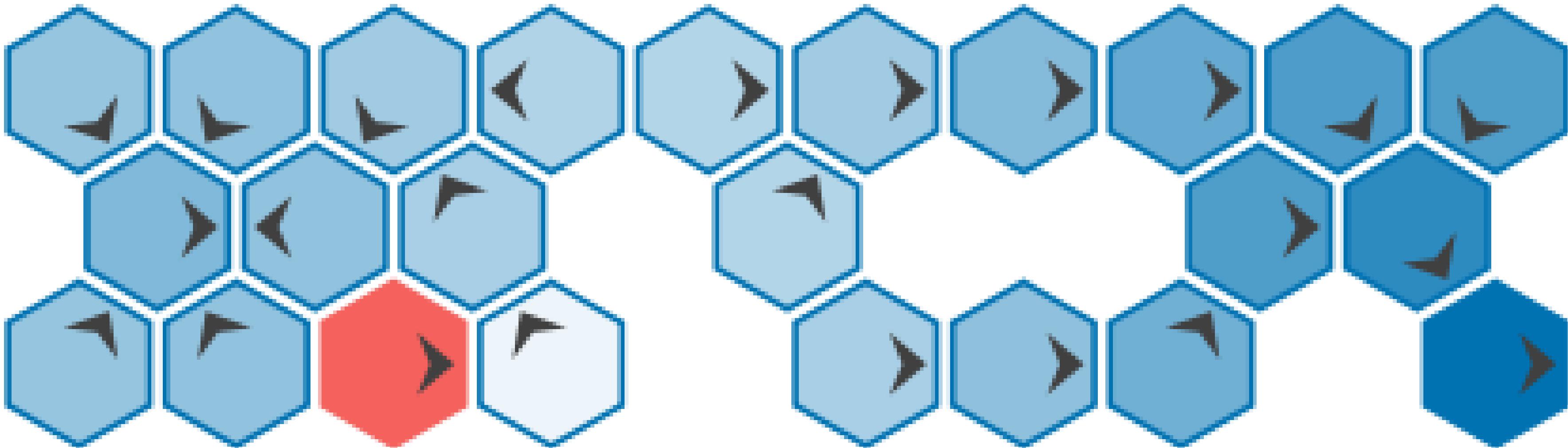
heuristic search

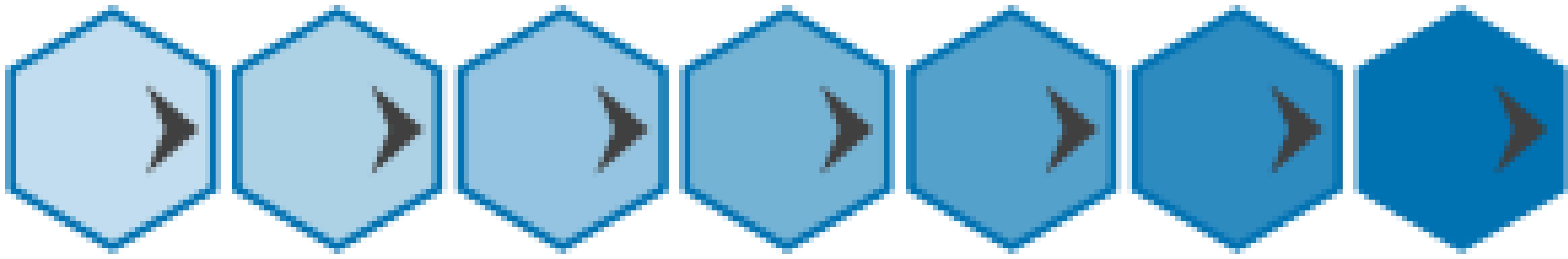


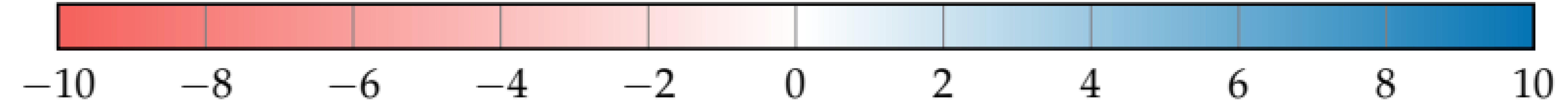






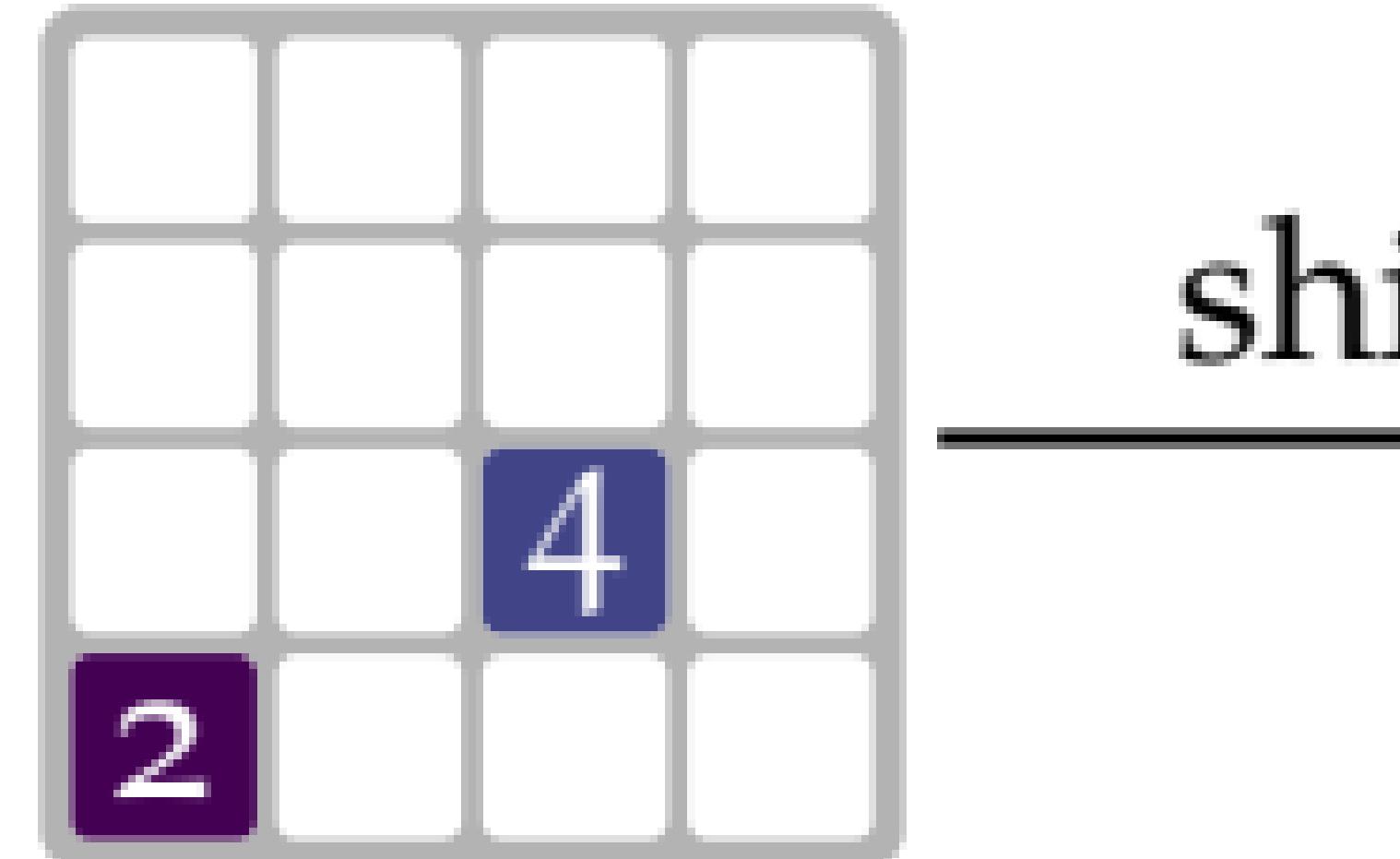




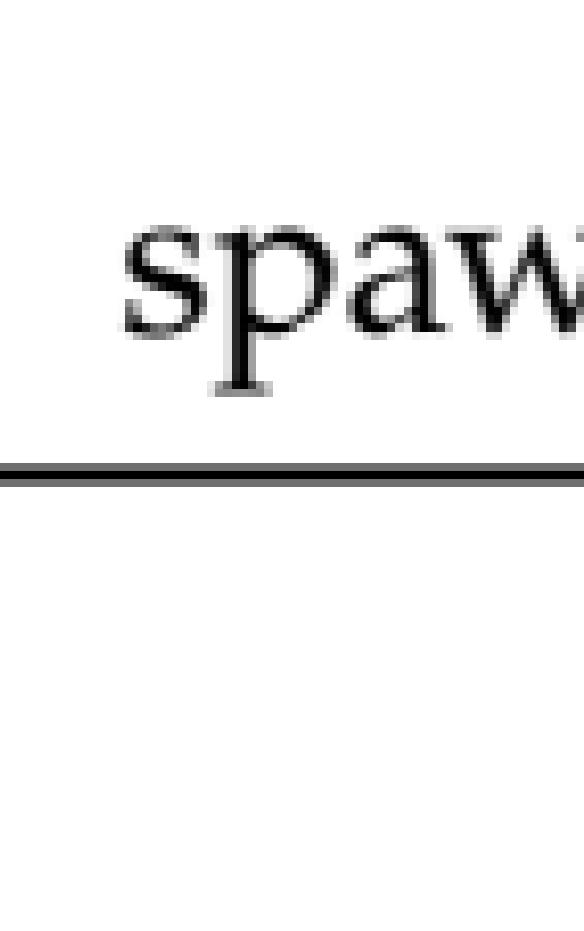


2

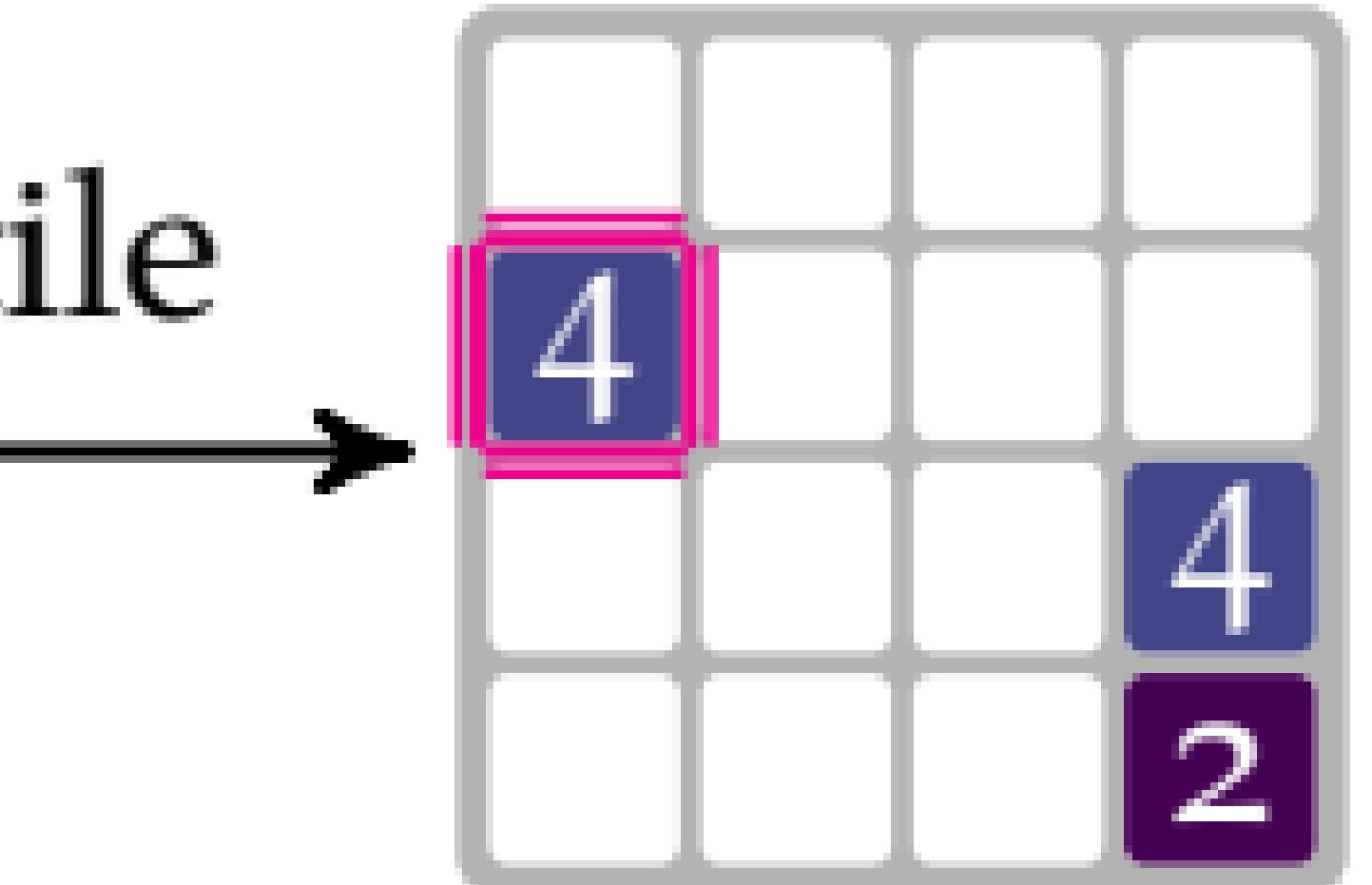
4

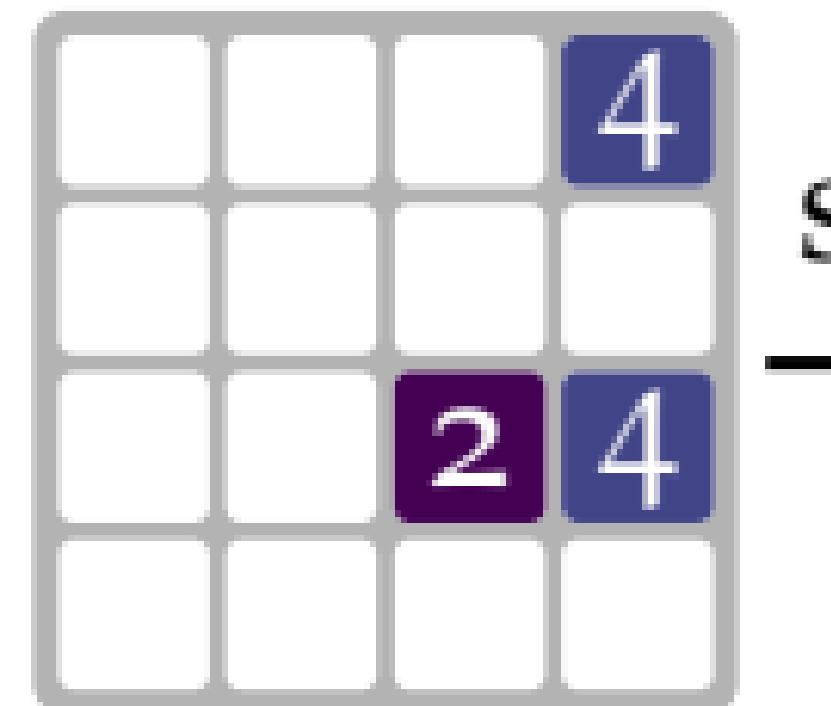


shift tiles

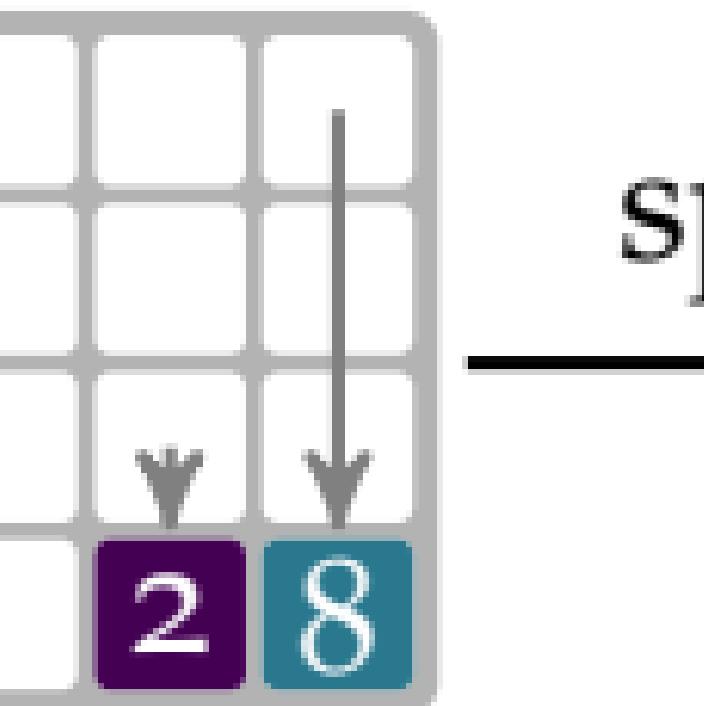


spawn tile



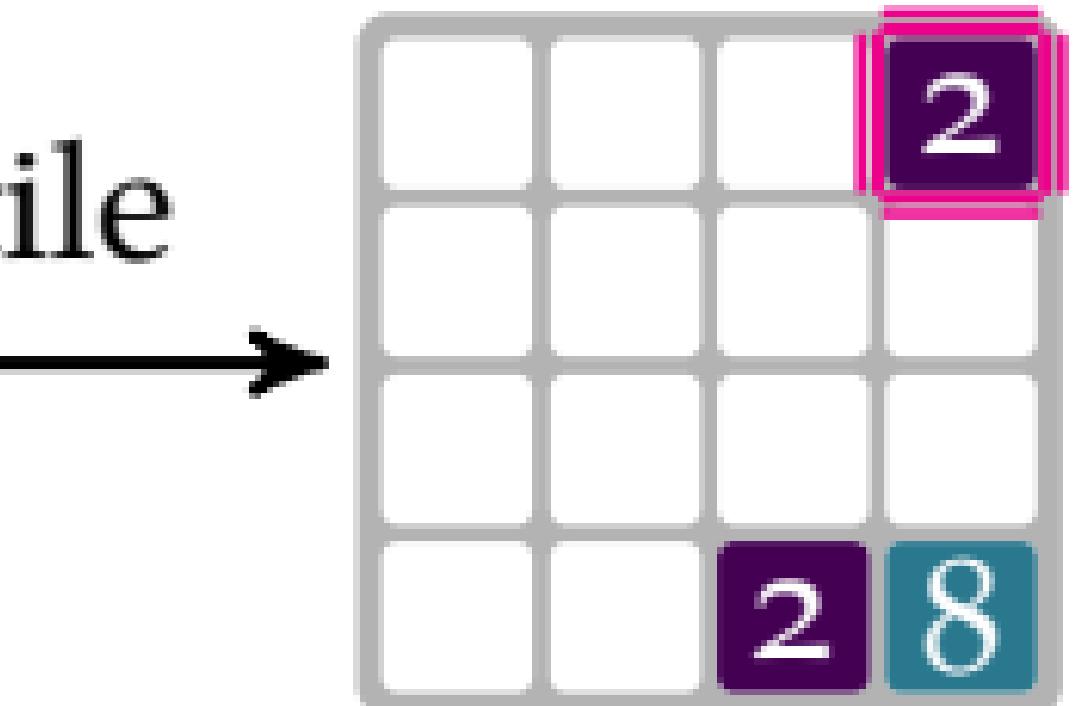


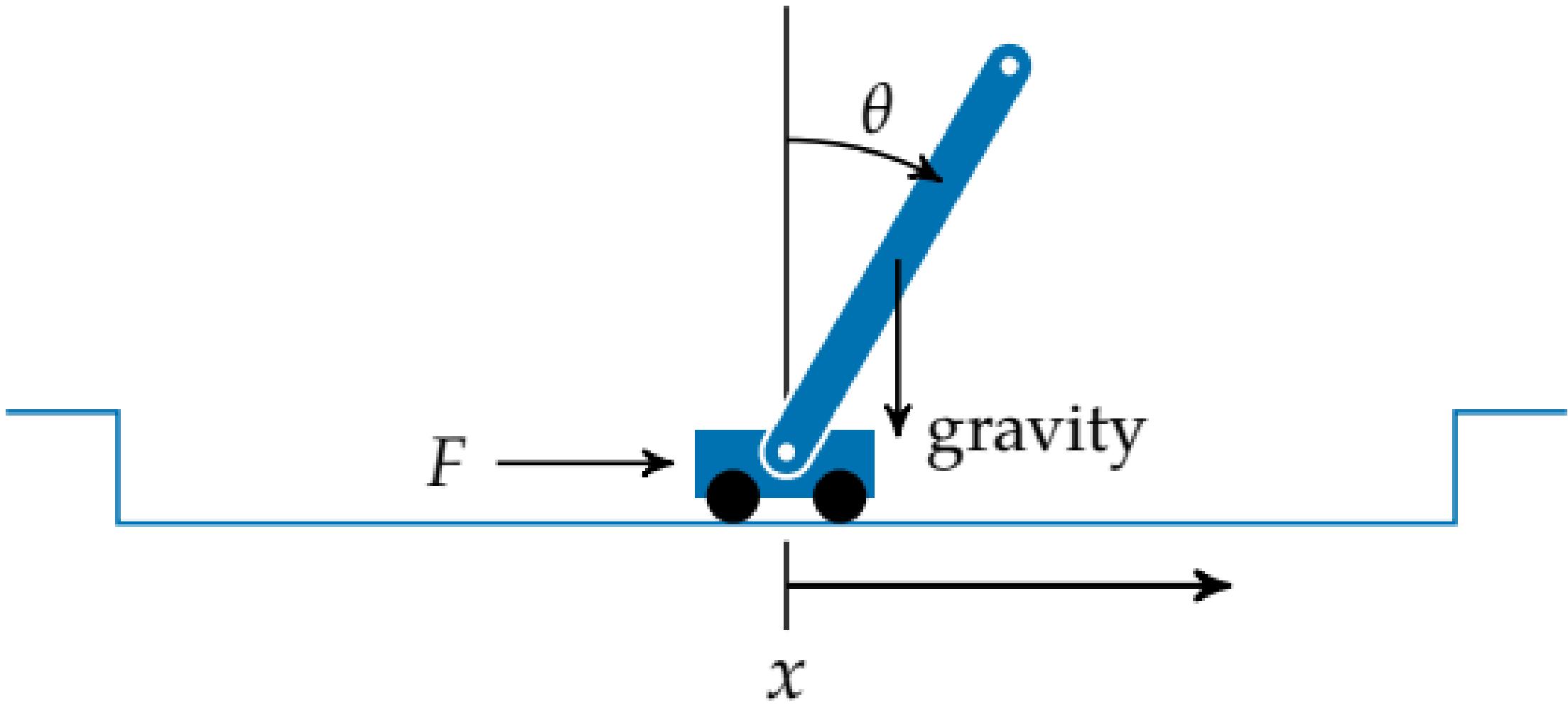
shift & merge



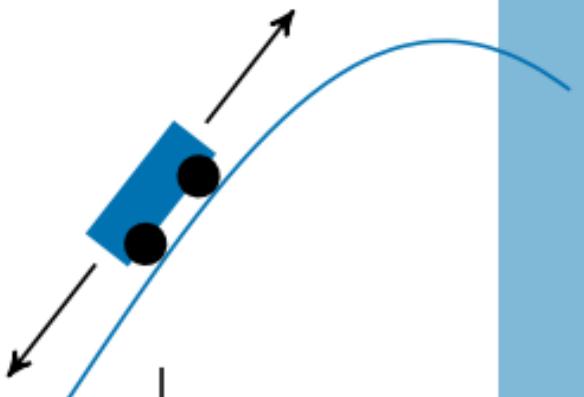
+8 reward

spawn tile



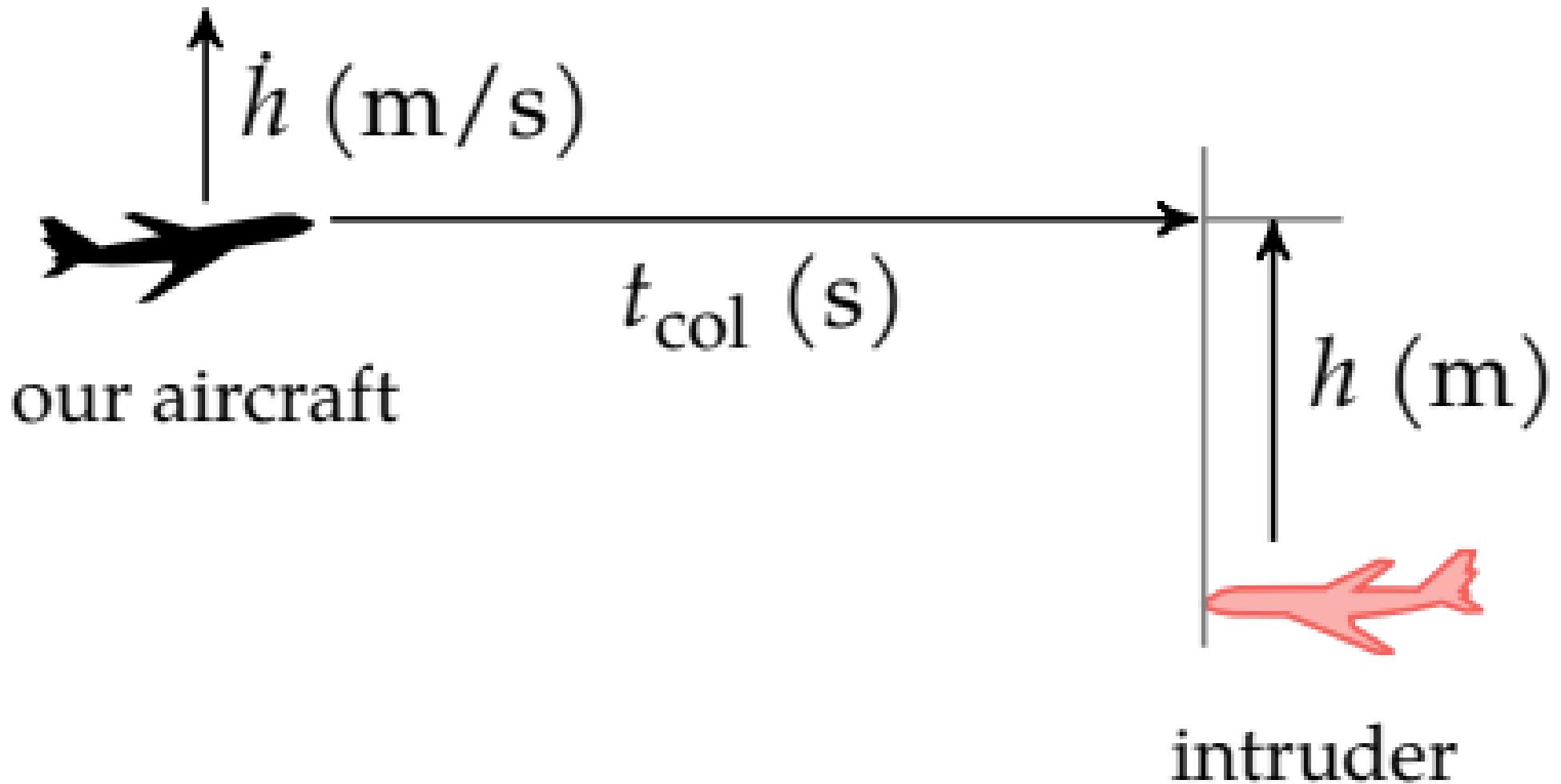


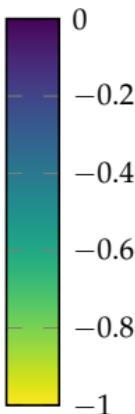
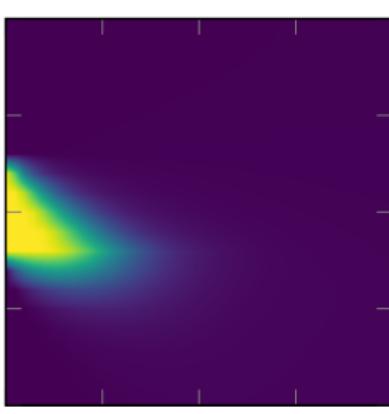
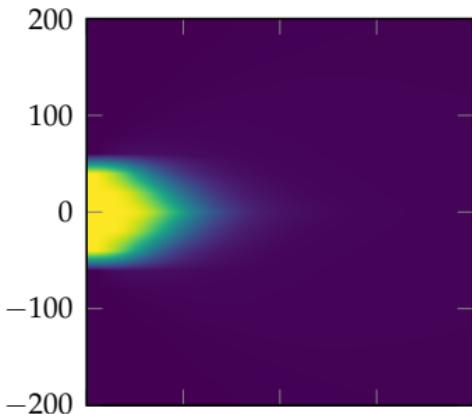
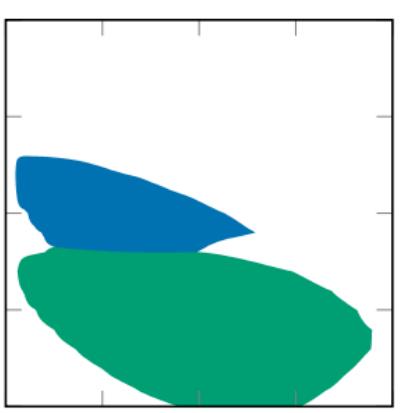
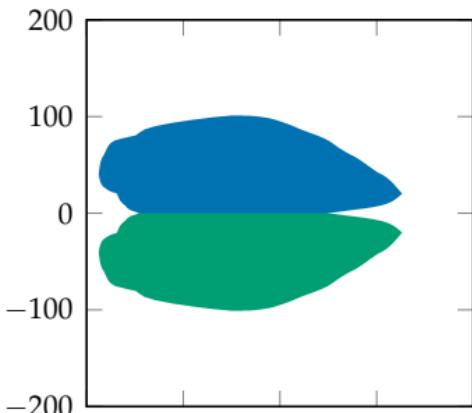
goal



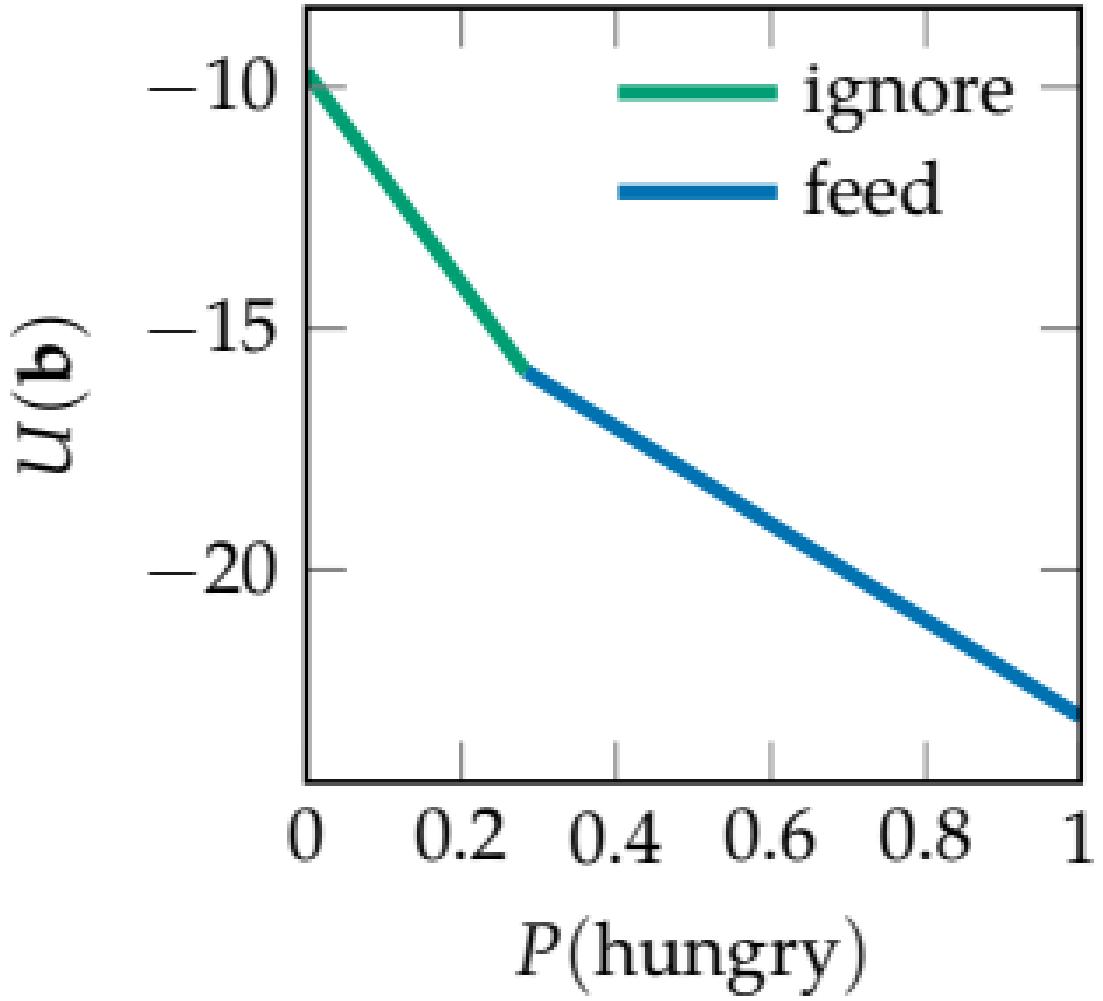
gravity

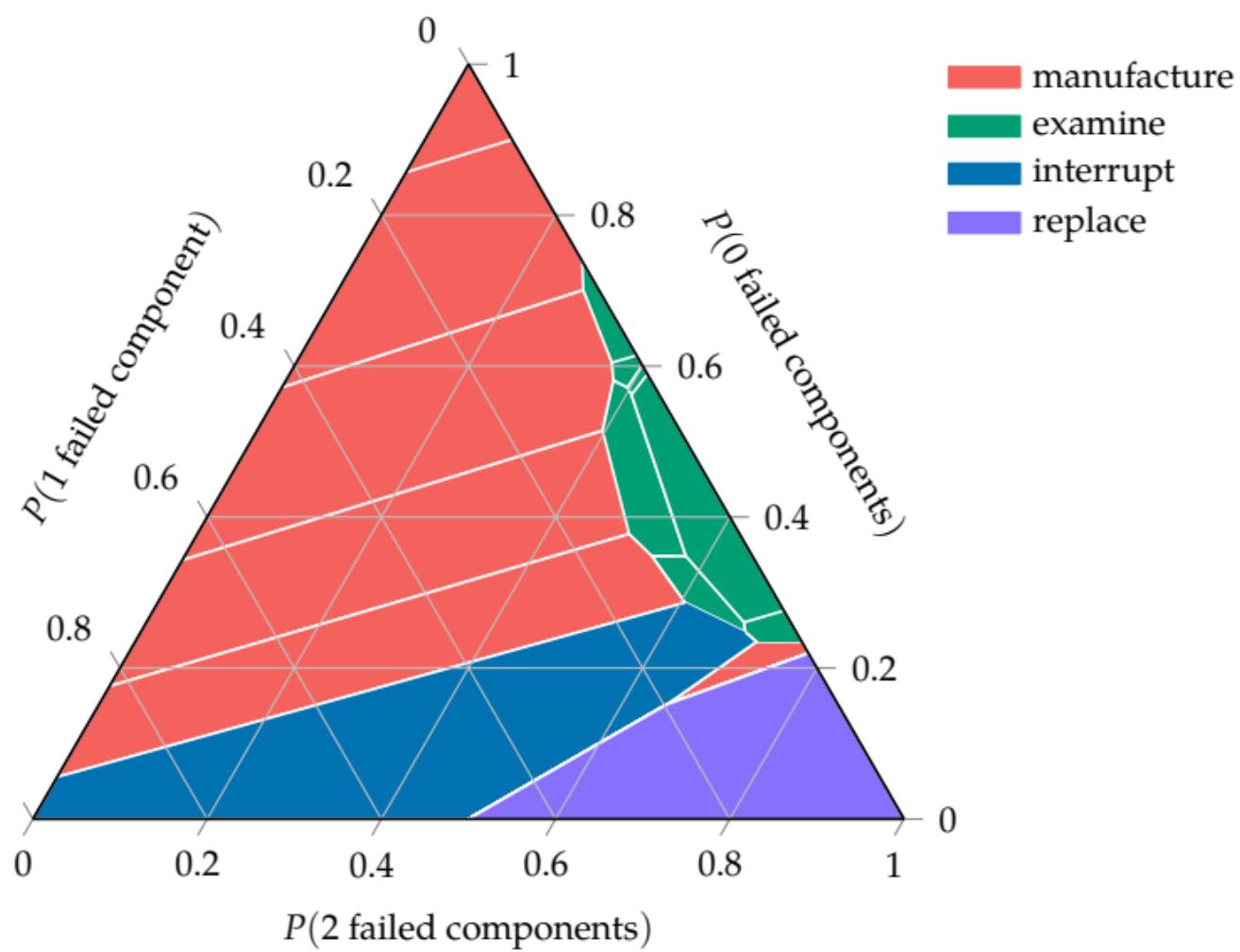
x



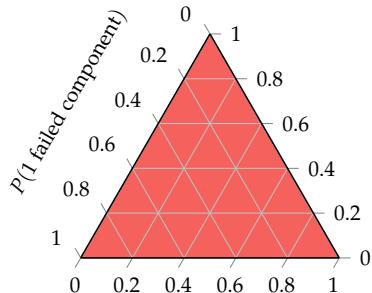
$\dot{h} = 0.0(\text{m/s})$ $\dot{h} = 5.0(\text{m/s})$ $h(\text{m})$  $h(\text{m})$  $t_{\text{col}}(\text{s})$ $t_{\text{col}}(\text{s})$

- no advisory
- descend
- climb

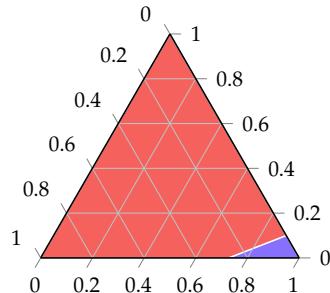




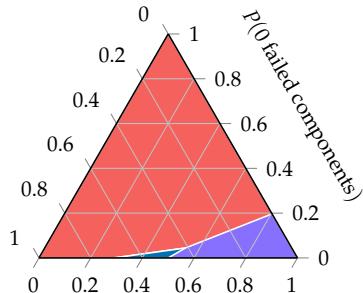
5-step plan



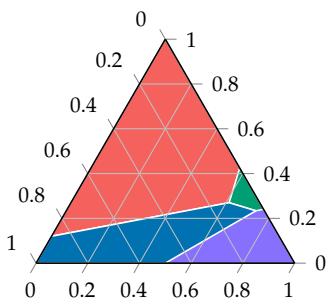
6-step plan



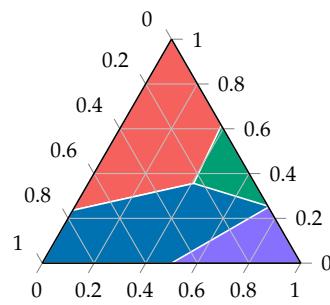
7-step plan



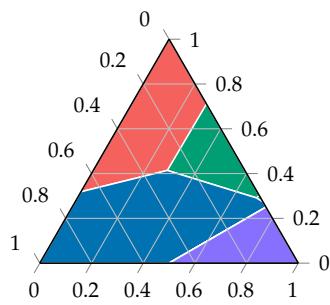
8-step plan



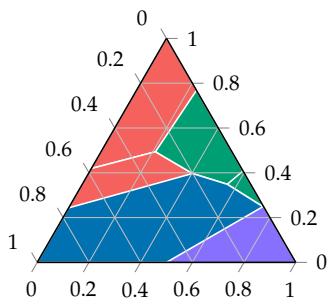
9-step plan



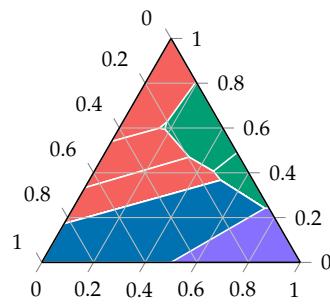
10-step plan



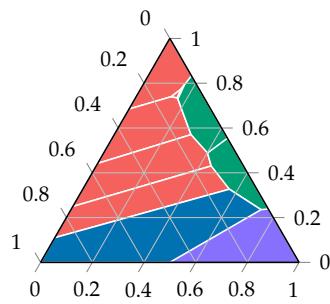
11-step plan



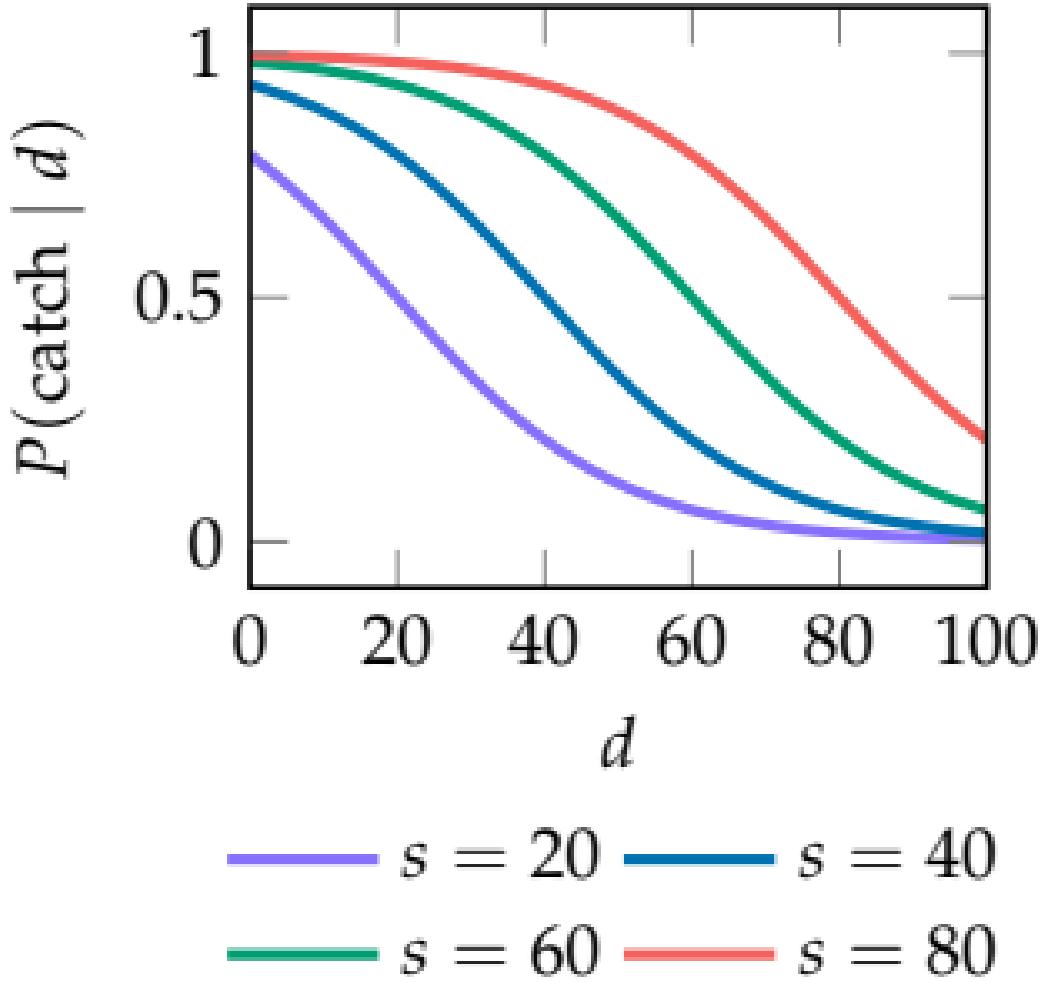
12-step plan



13-step plan

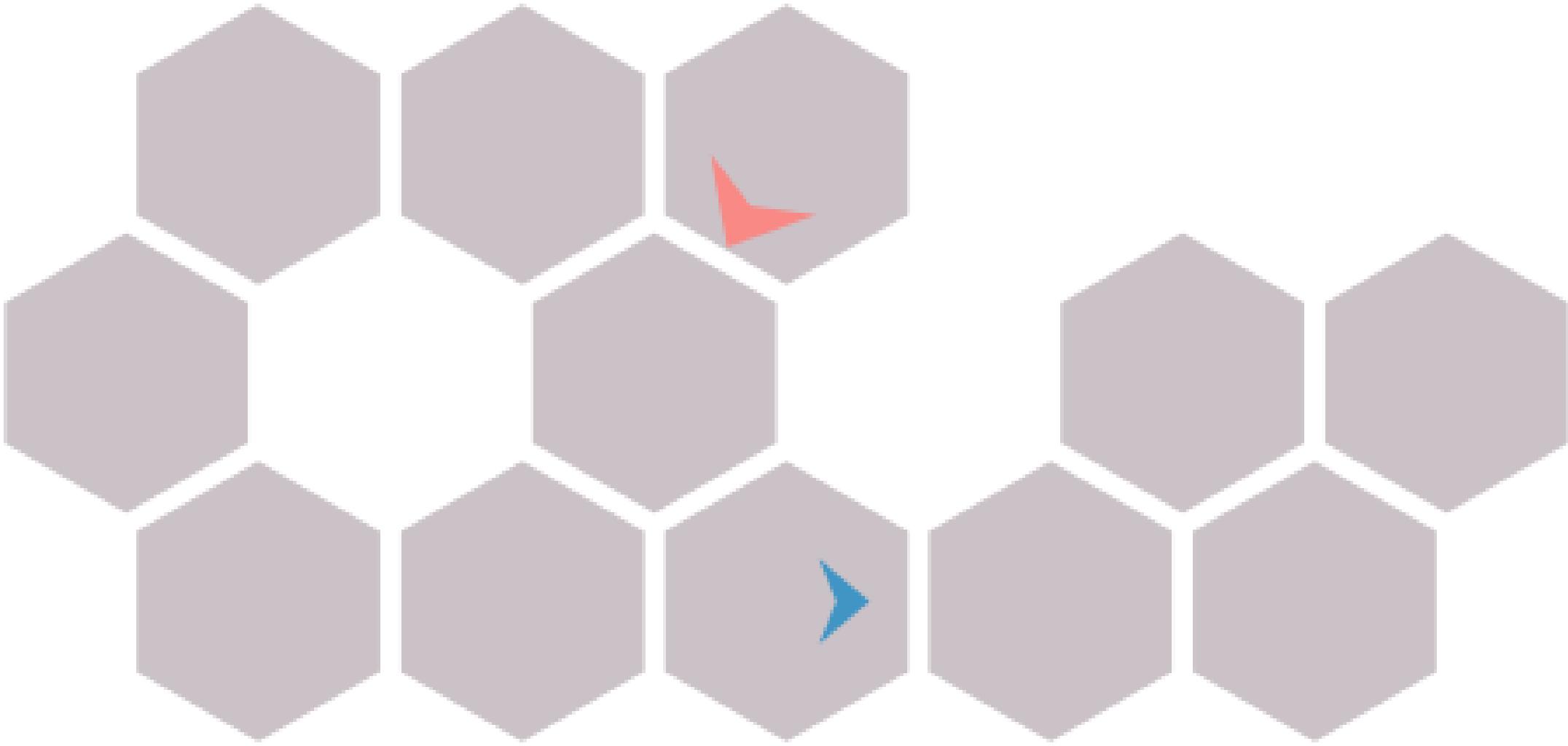
 $P(2 \text{ failed components})$

- manufacture
- examine
- interrupt
- replace



	agent 2	
	cooperate	defect
agent 1	cooperate	$-1, -1$
	defect	$-4, 0$
	cooperate	$0, -4$
	defect	$-3, -3$

		agent 2	
		rock	paper
agent 1	rock	0, 0	-1, 1
	paper	1, -1	0, 0
	scissors	-1, 1	1, -1



Any

