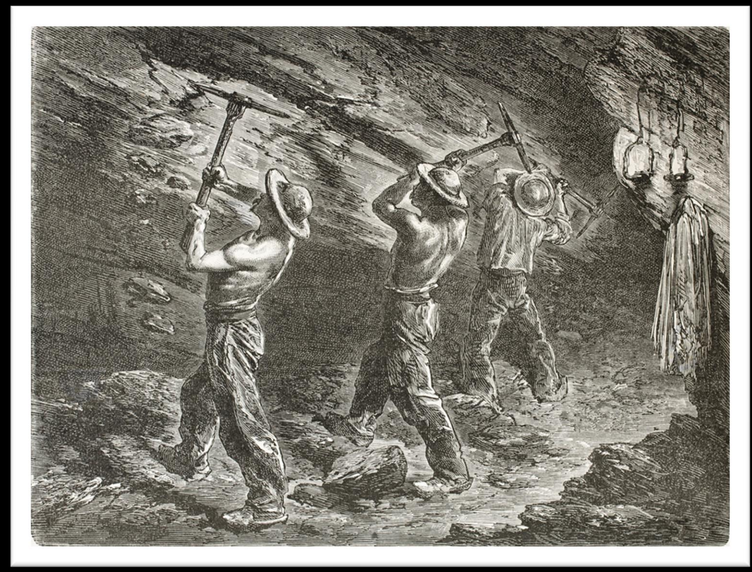
Coordinated Resource Extraction

Deep Q-Networks for centralized cooperative multi-agent reinforcement learning

Emil G. Stålvinge, Anders Köhler & Julián R. Marrades

The State

The Labor

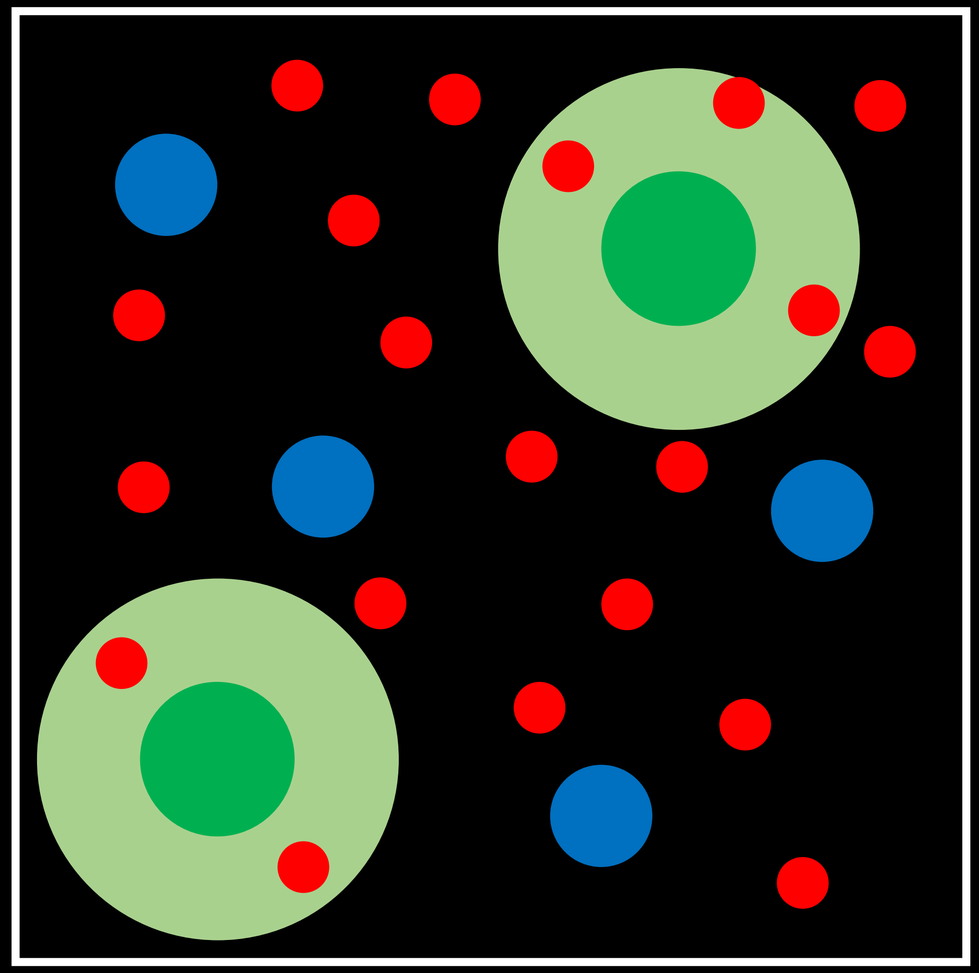


The Riot

How can we model this?

The game

$p = 1$



Resource

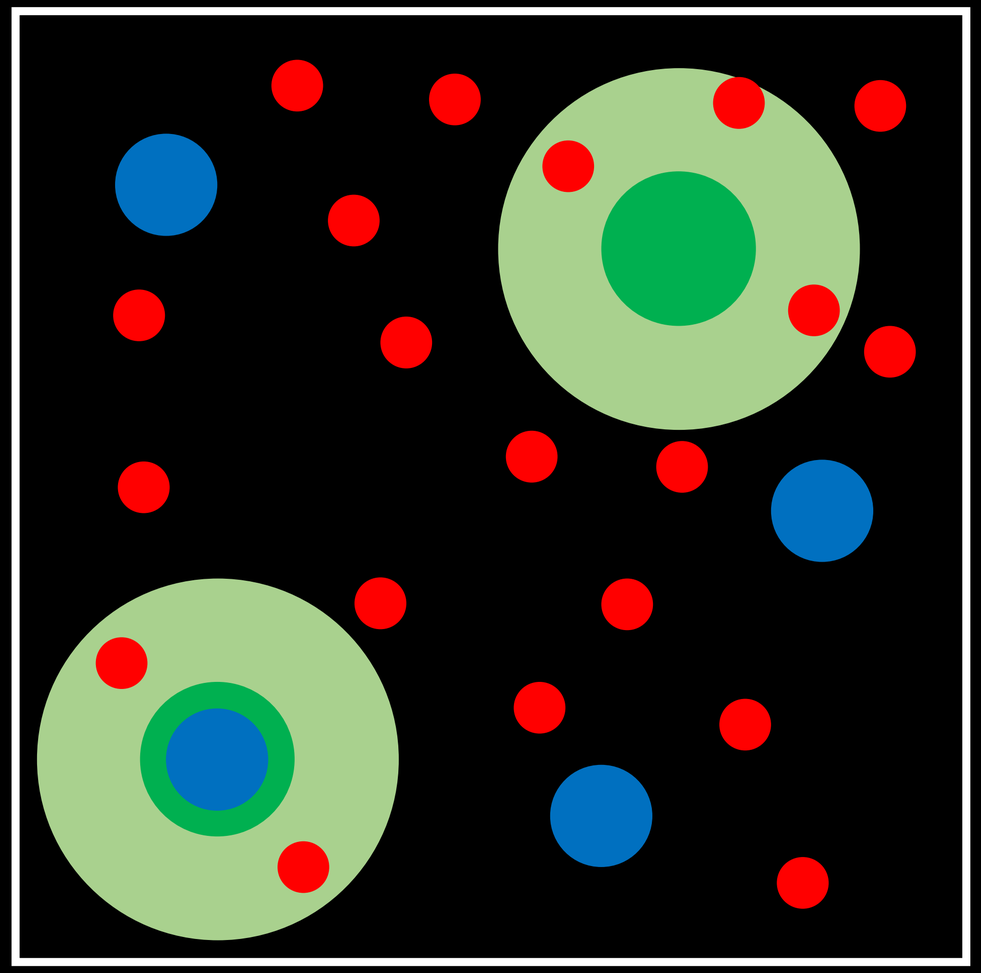
Circle of influence

Enforcer

Civilian

The game

$p = 1$



Resource

Circle of influence

Enforcer

Civilian

Single-Agent RL

Multi-Agent RL

SARL

Given

- A, S
- $\delta: S \times A \to S$
- $\mathcal{R}: S \times A \to \mathbb{R}$

Find π so that

$$V^{\pi}(s_t) \coloneqq \sum_{i=1}^{\infty} \gamma^i \cdot r_{t+i}$$

is maximized

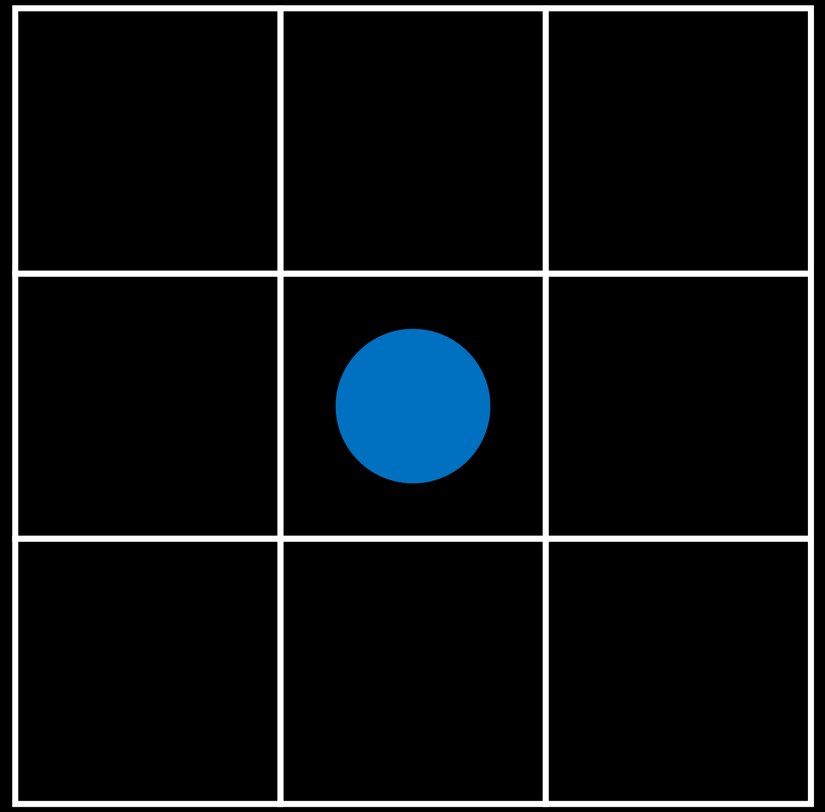
MARL

Let m be the number of agents

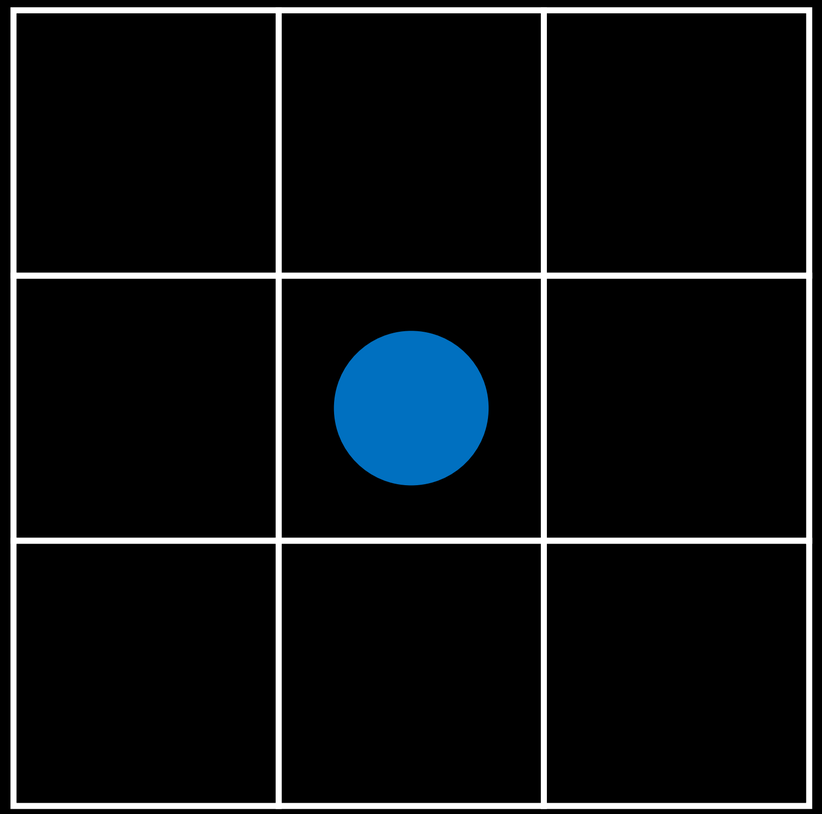
- S grows exponentially with m
- $A \coloneqq A_1 \times A_2 \times \cdots \times A_n$
- $\delta: S \times A_1 \times A_2 \times \cdots \times A_n \to S$
- $\mathcal{R}: S \times A_1 \times A_2 \times \cdots \times A_n \to \mathbb{R}$ or $\mathbb{R}^m$

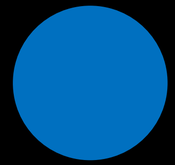
Train with DQN!

MARL to SARL



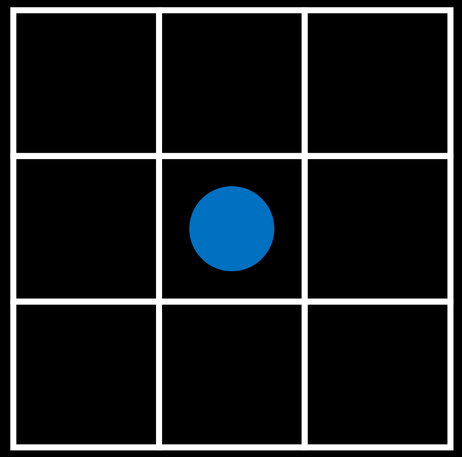
MARL to SARL

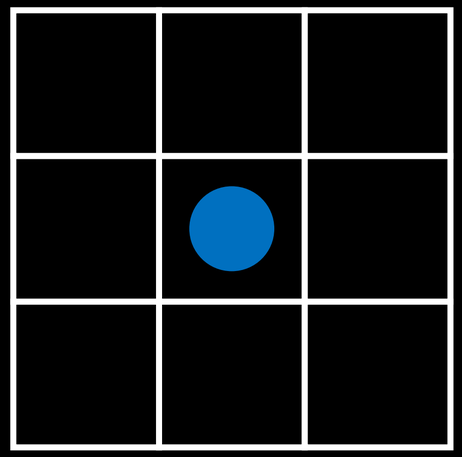


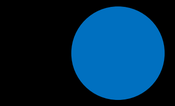


$$A = [0, 1, 2, 3, 4]$$
$$size(A) = 5$$

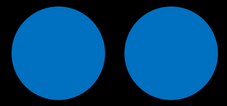
MARL to SARL





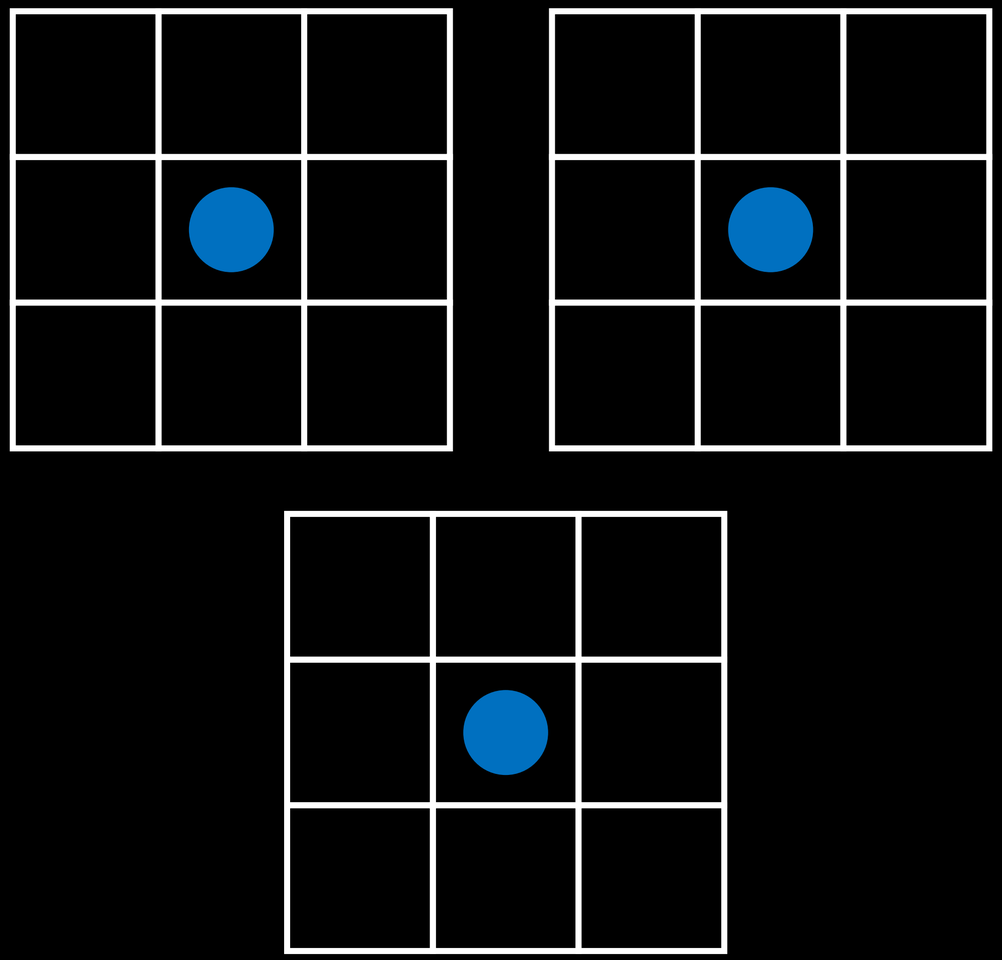


$$A = [0, 1, 2, 3, 4]$$
$$size(A) = 5$$

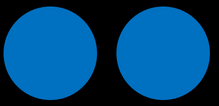


$$A = [(0, 0), (0, 1), (0, 2), \dots]$$
$$size(A) = 5^2 = 25$$

MARL to SARL



$$A = [0, 1, 2, 3, 4]$$
$$size(A) = 5$$

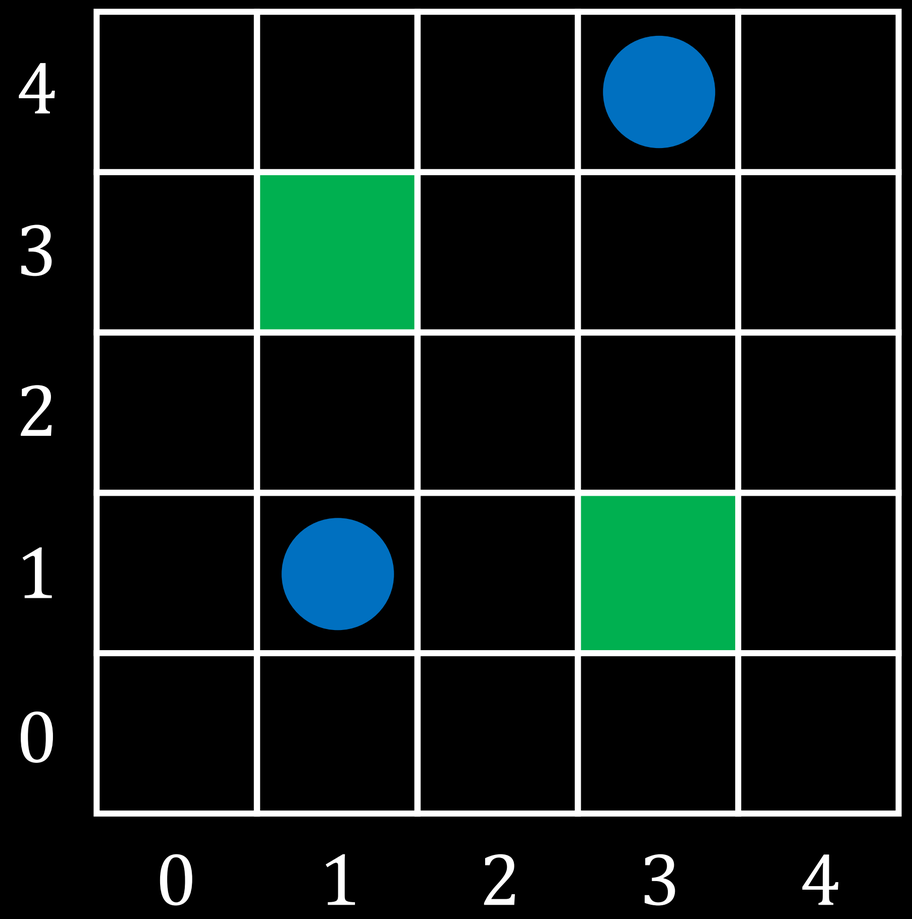


$$A = [(0, 0), (0, 1), (0, 2), \dots]$$
$$size(A) = 5^2 = 25$$

$$A = [(0, 0, 0), (0, 0, 1), \dots]$$
$$size(A) = 5^3 = 125$$

v1.0 & v1.5

$g = 5$



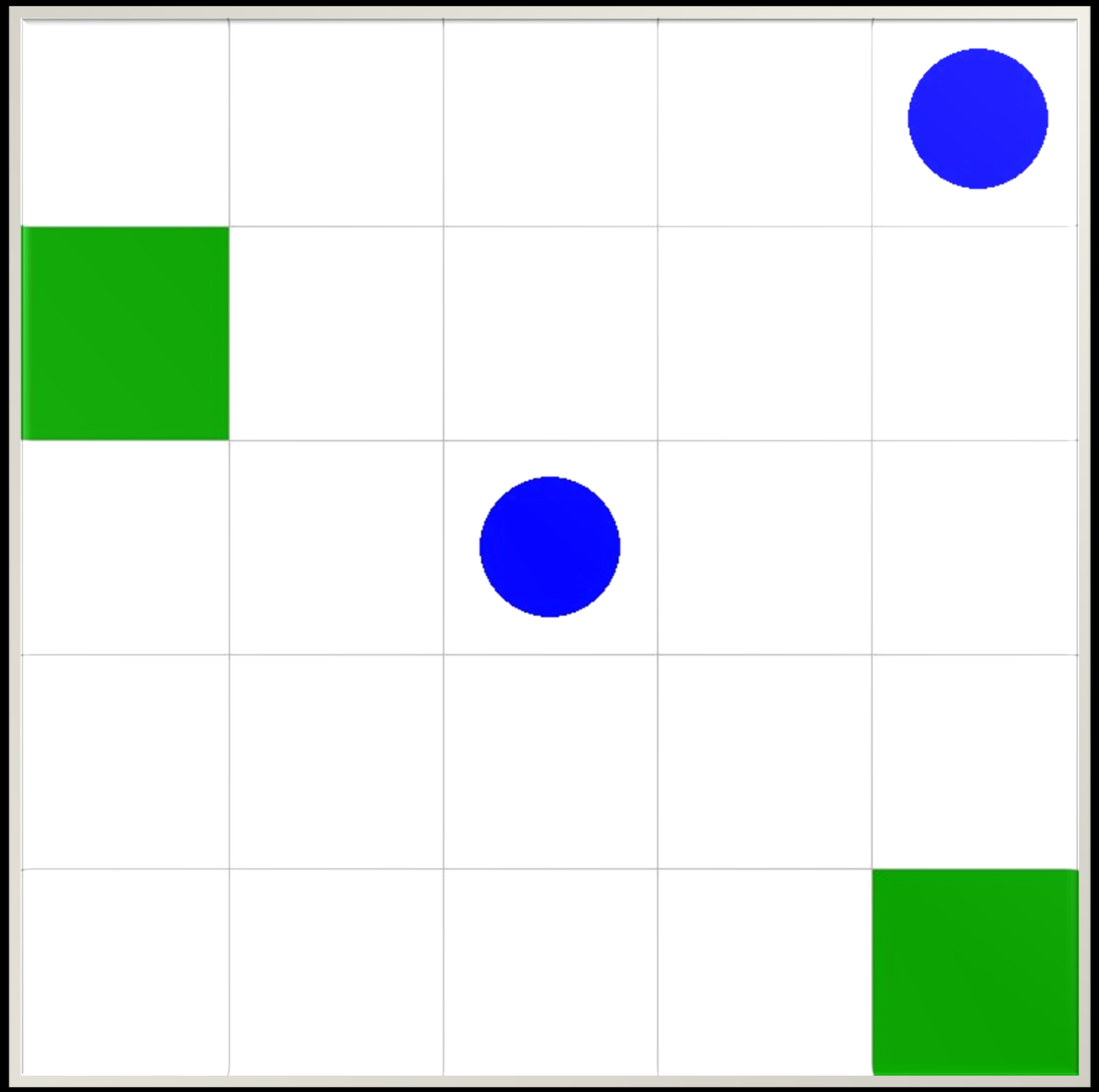
C v1.0 M

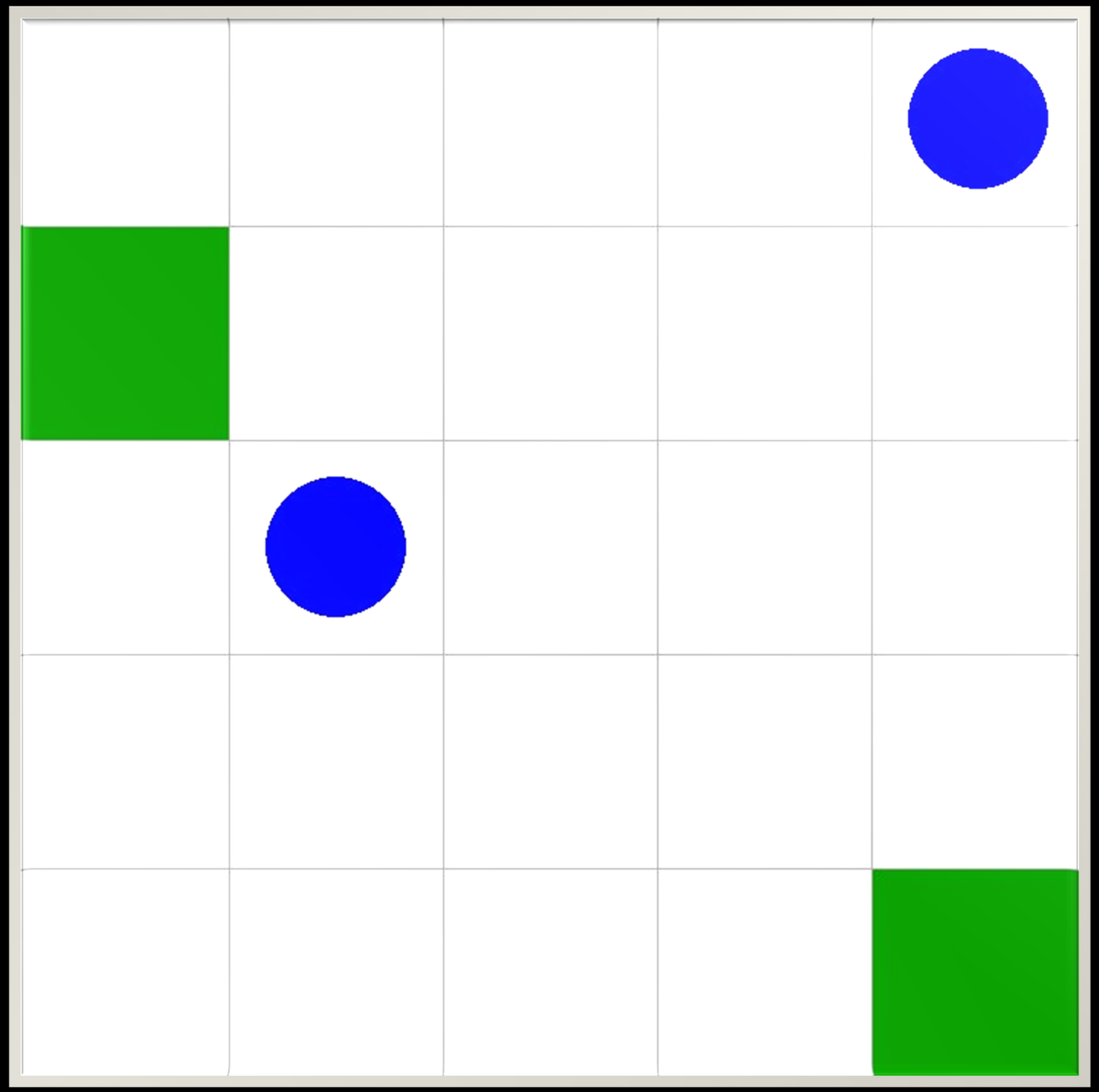
$$[\,1, 3, 3, 1, 1, 1, 3, 4\,]$$

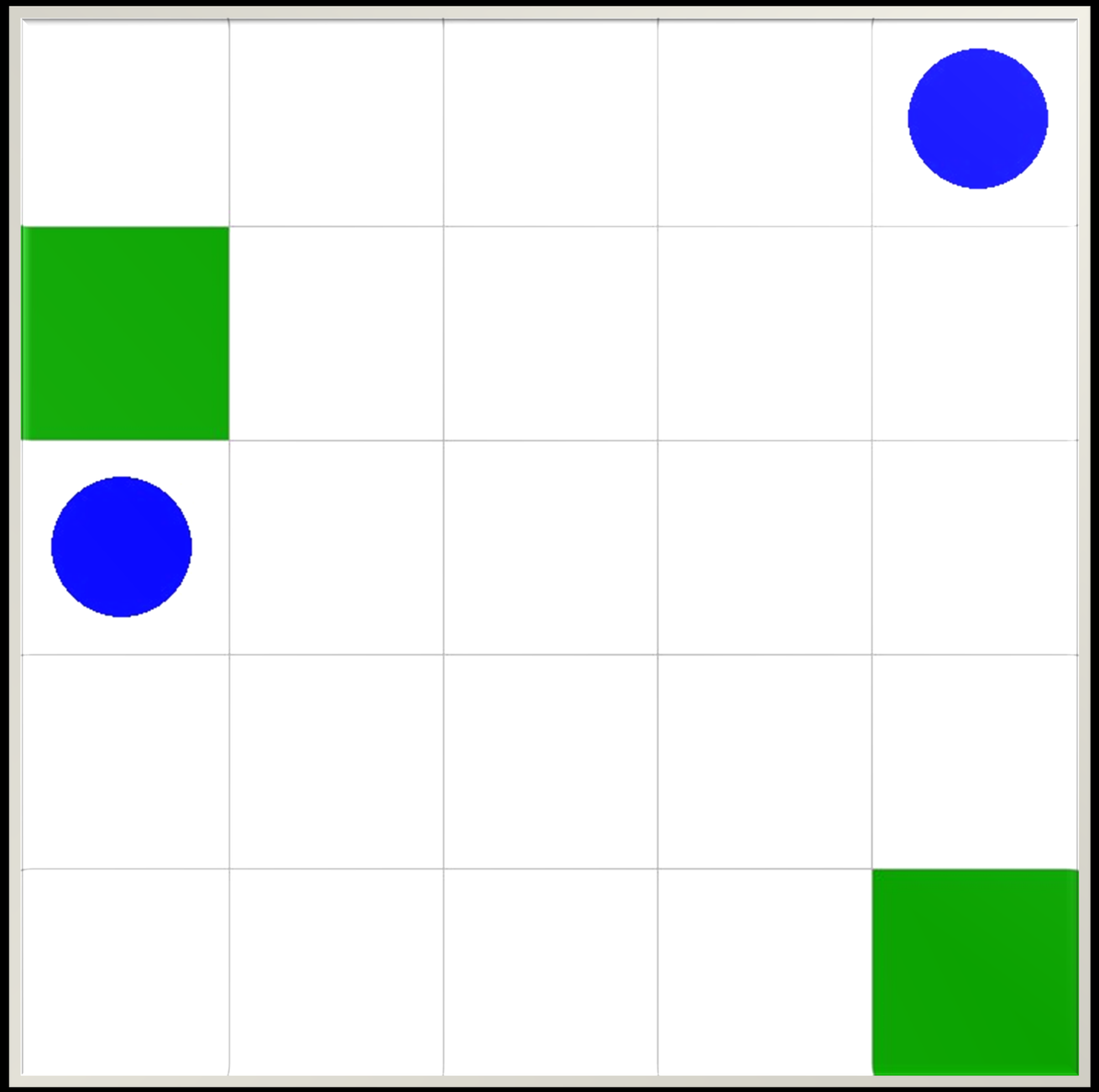
v1.5

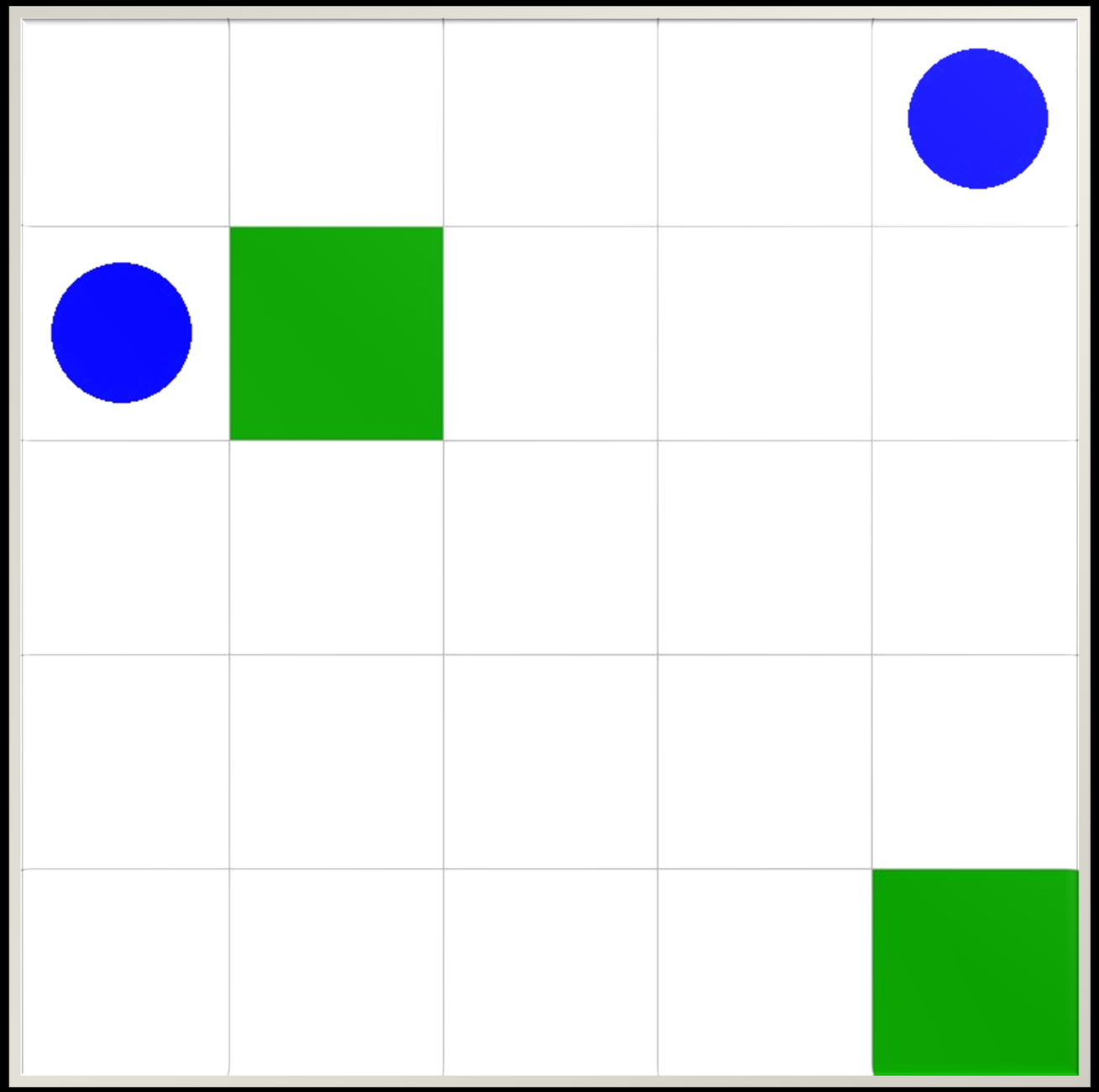
$$\frac{[\,1, 3, 3, 1, 1, 1, 3, 4\,]}{g - 1}$$

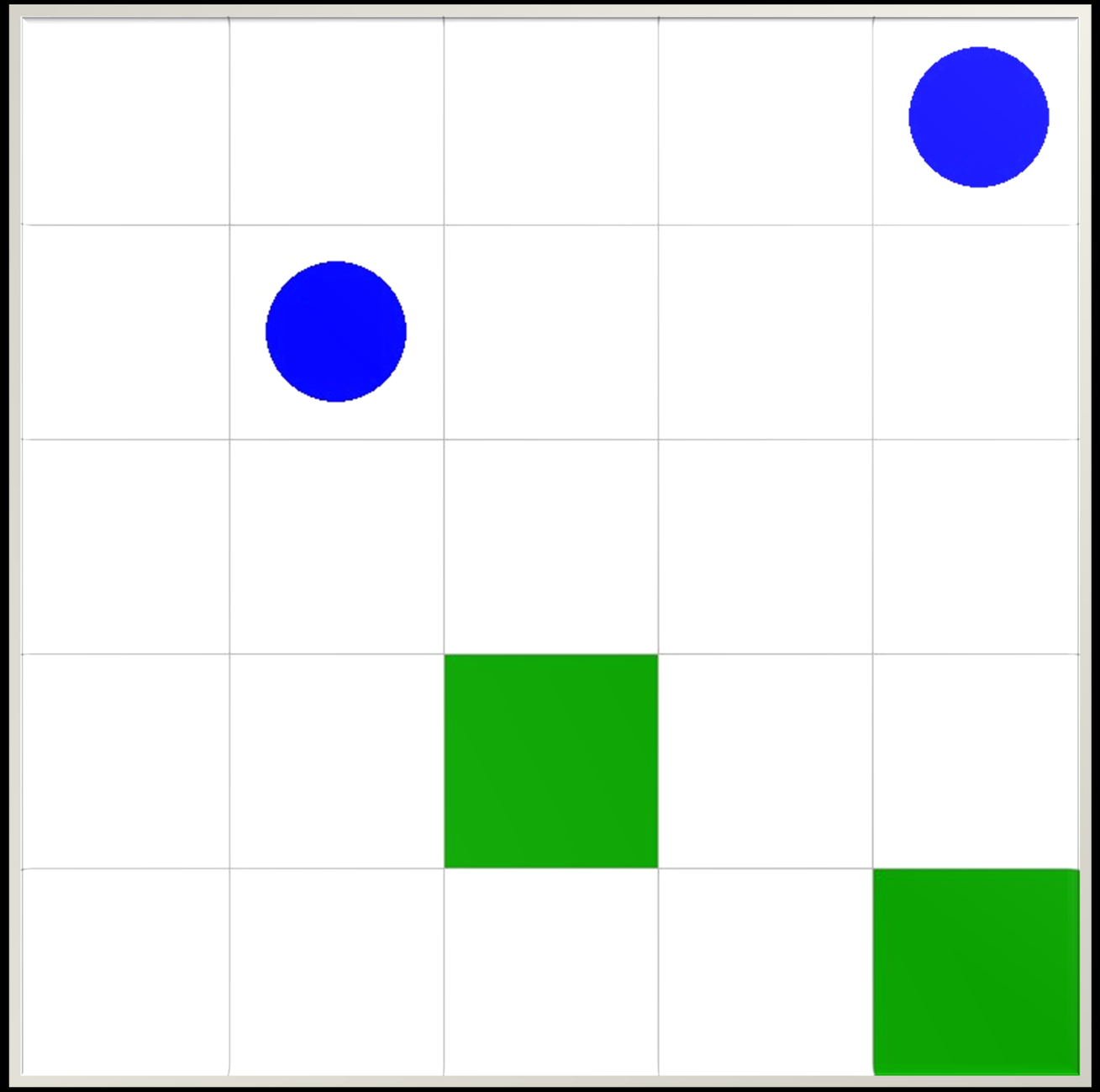
$$[\,0.25, 0.75, 0.75, 0.25, 0.25, 0.25, 0.75, 1\,]$$

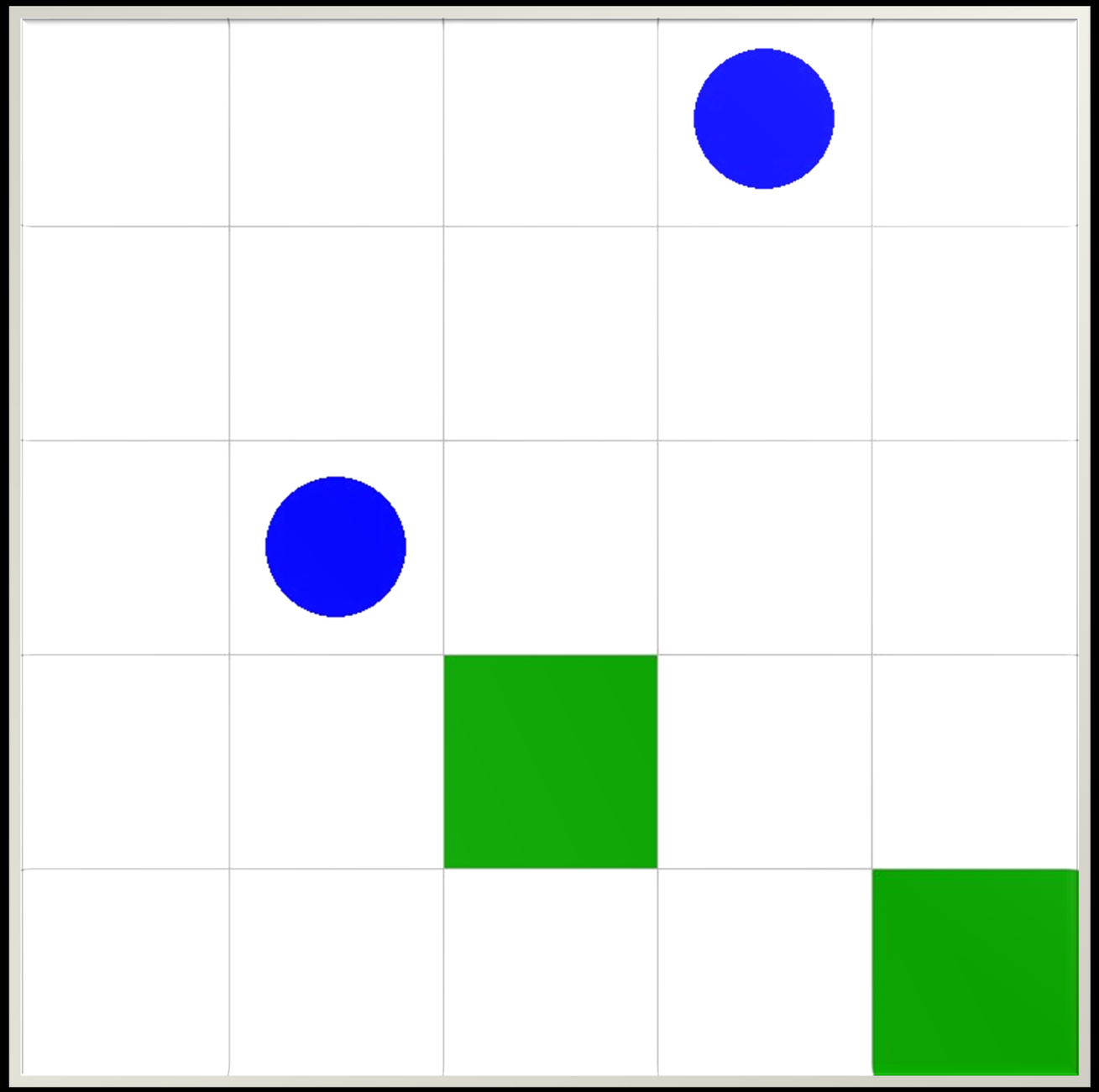


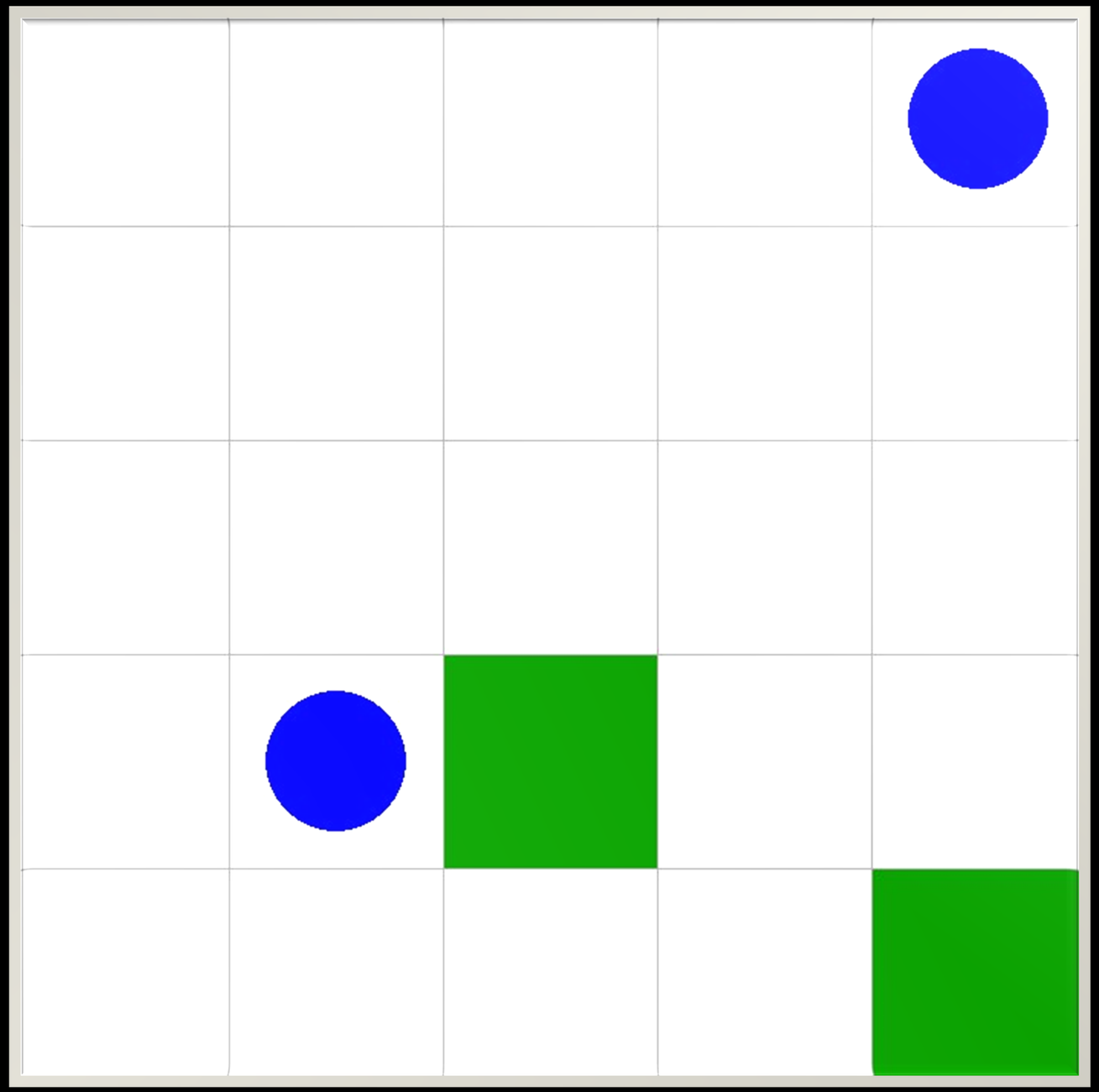


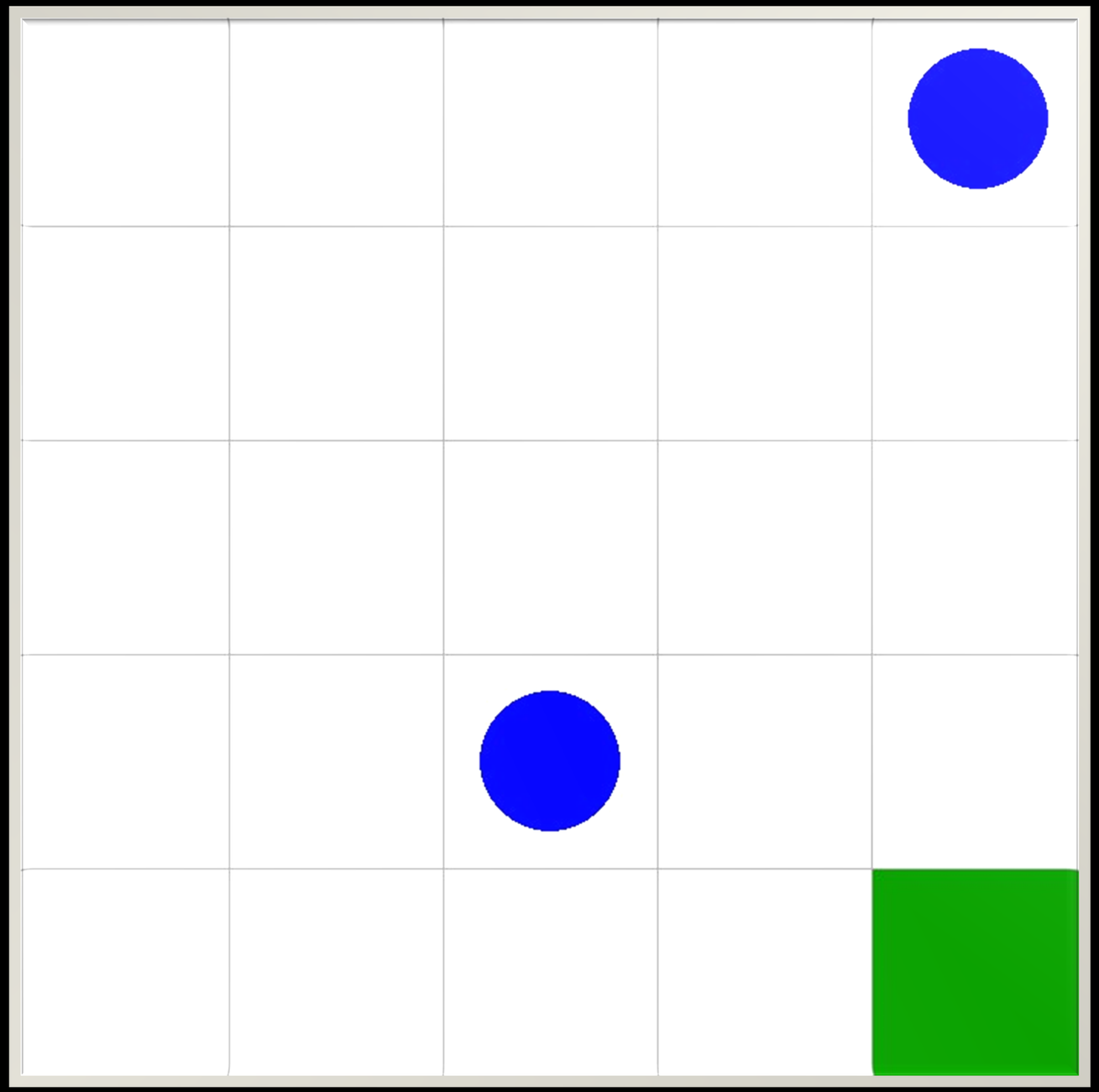


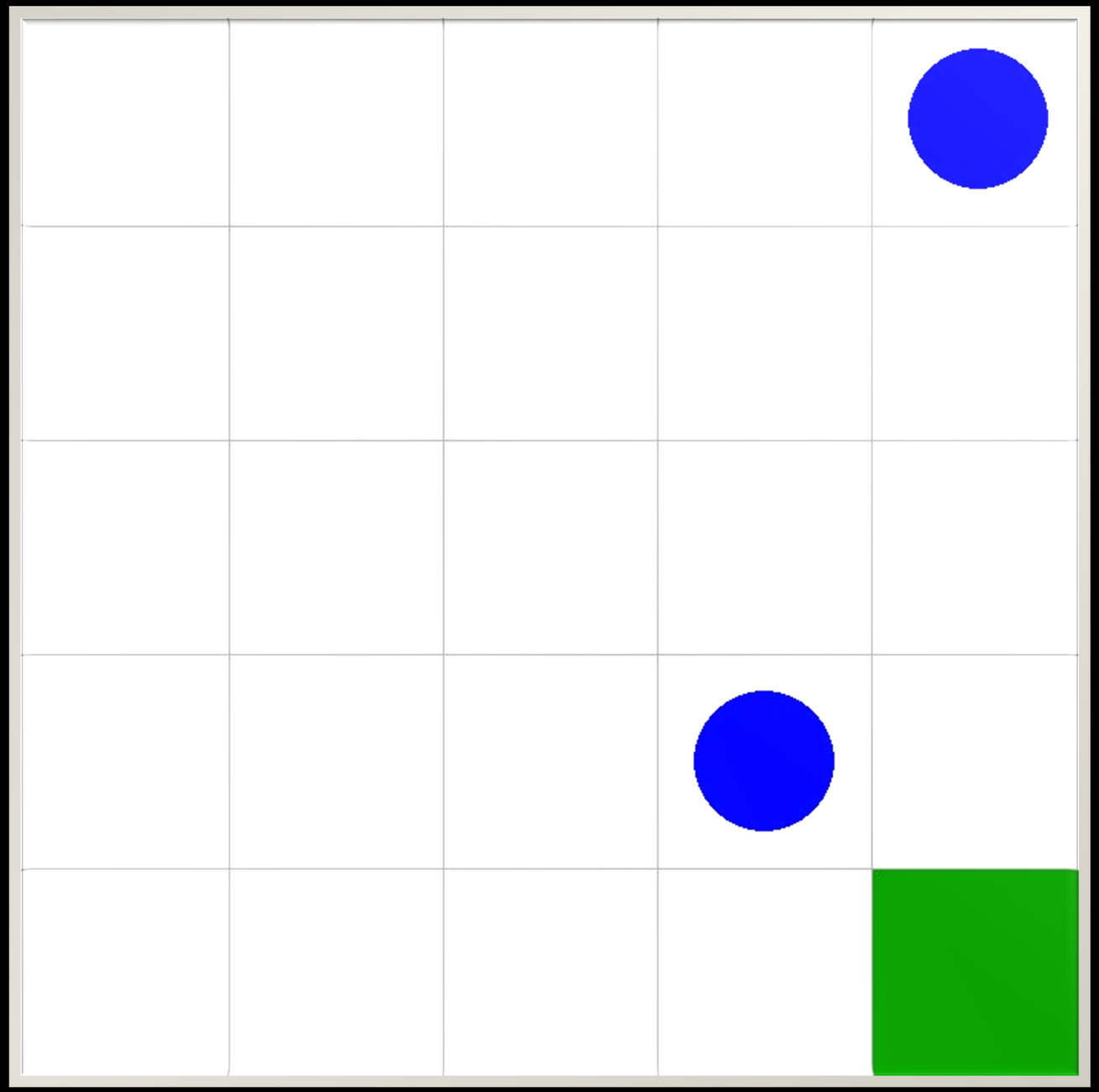


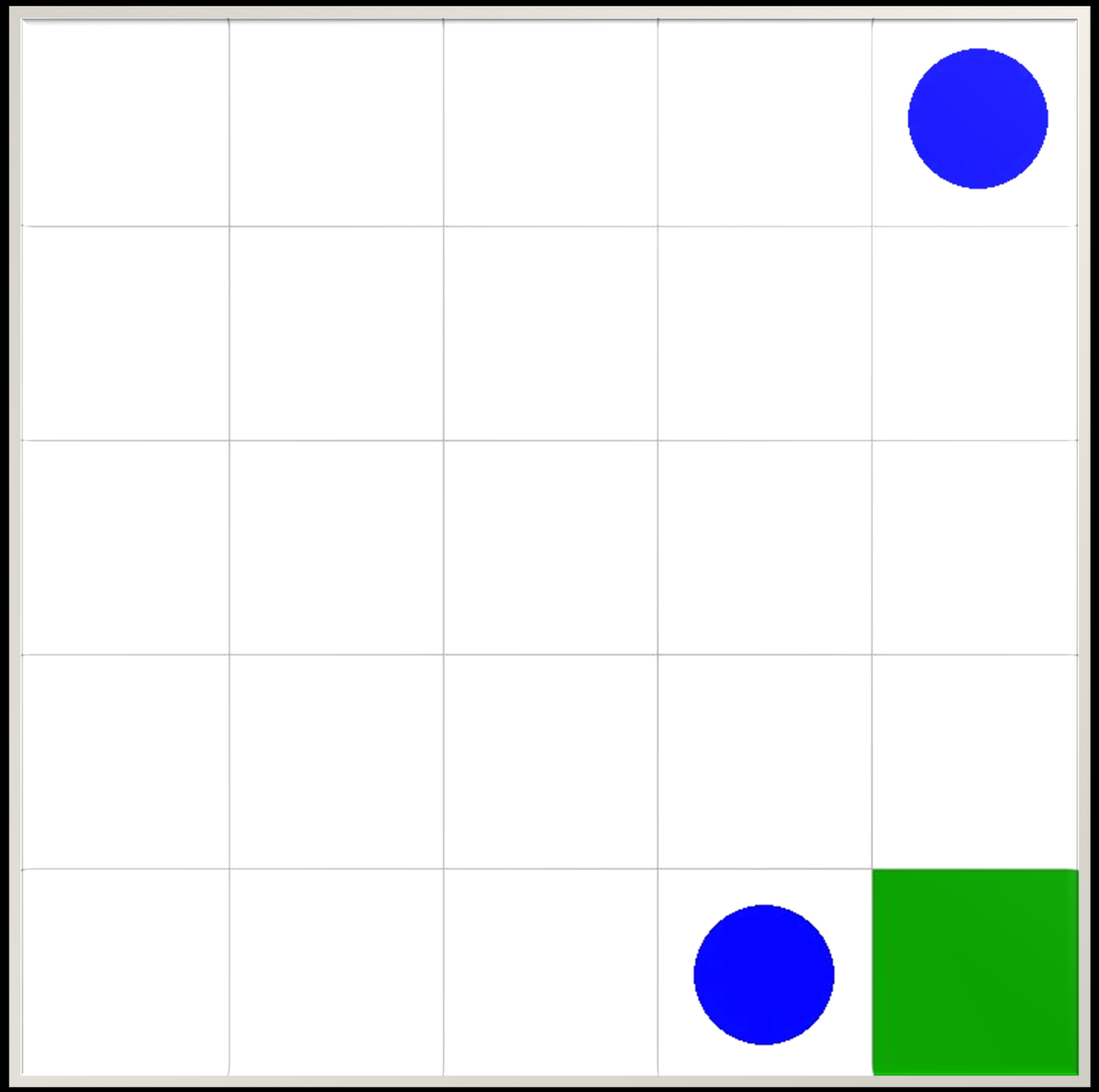






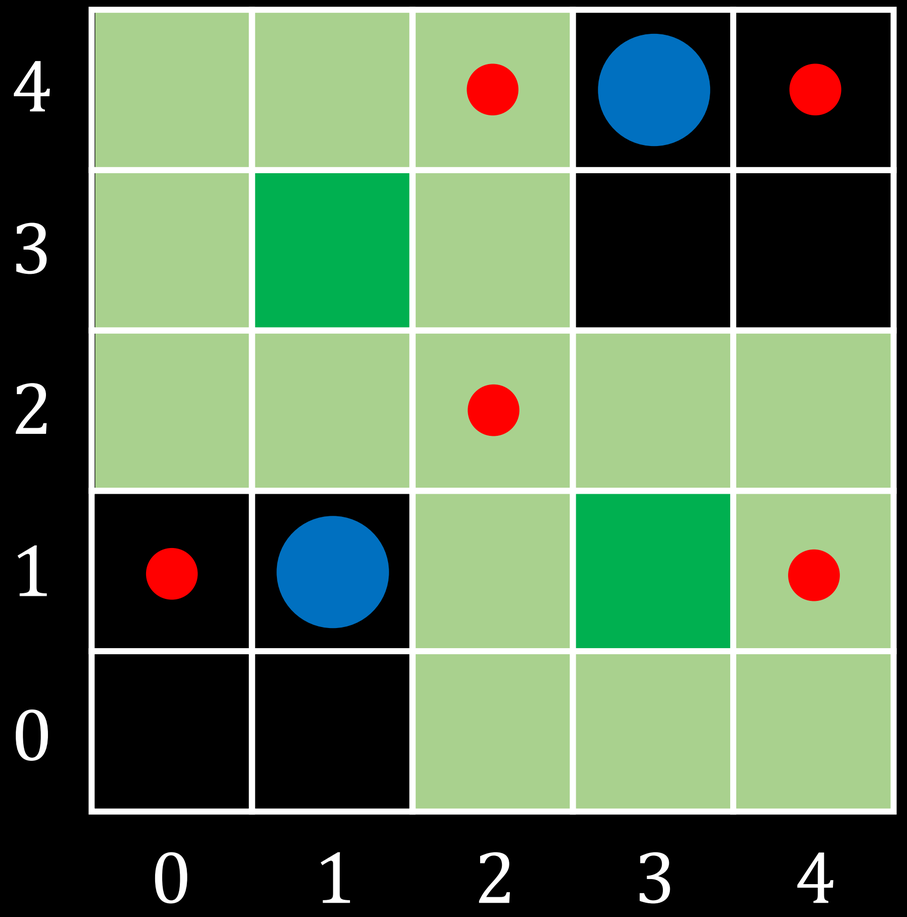






v2.0 & v2.5

$g = 5$
$I = 1$
$p = 1$



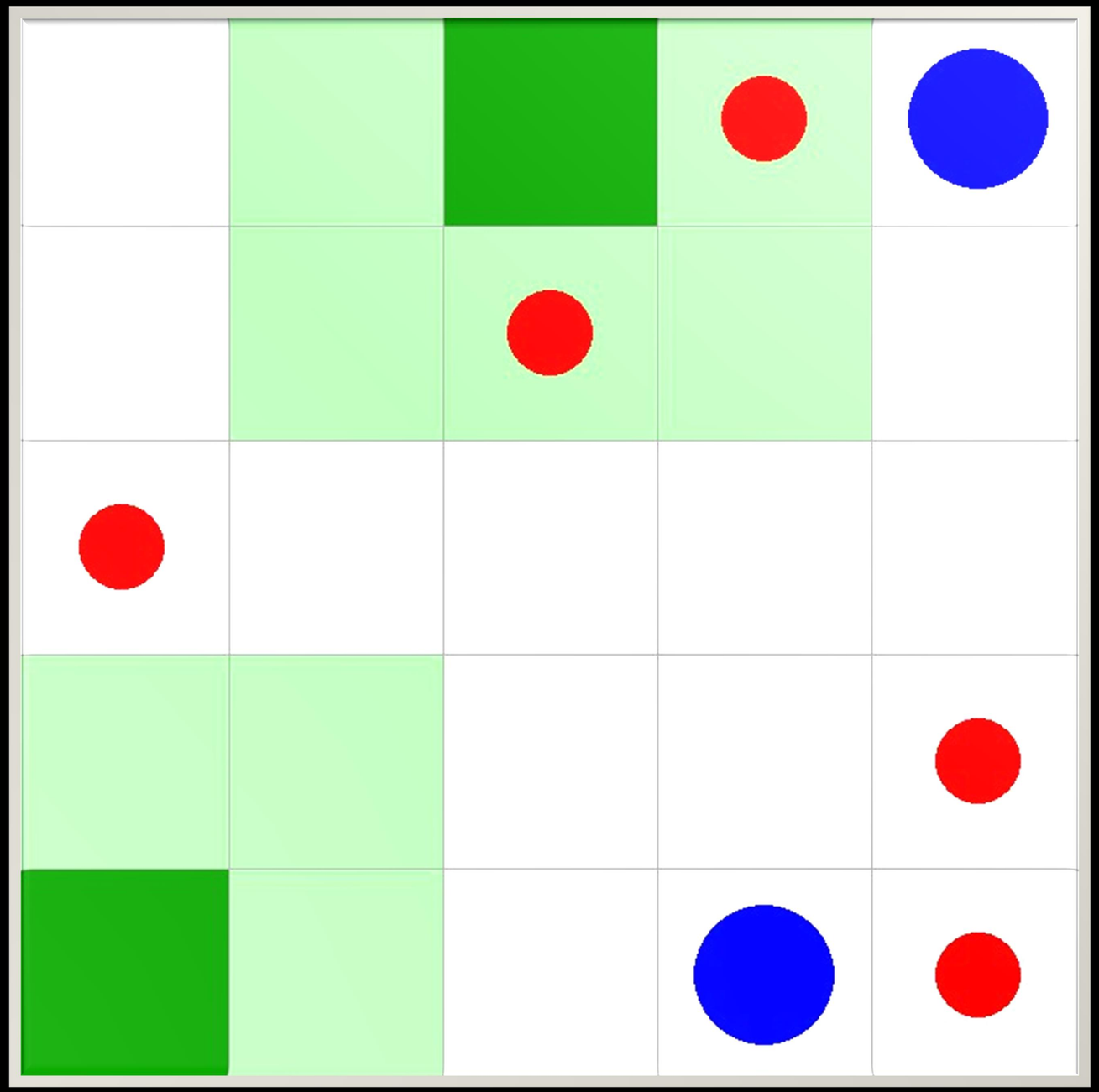
v2.0

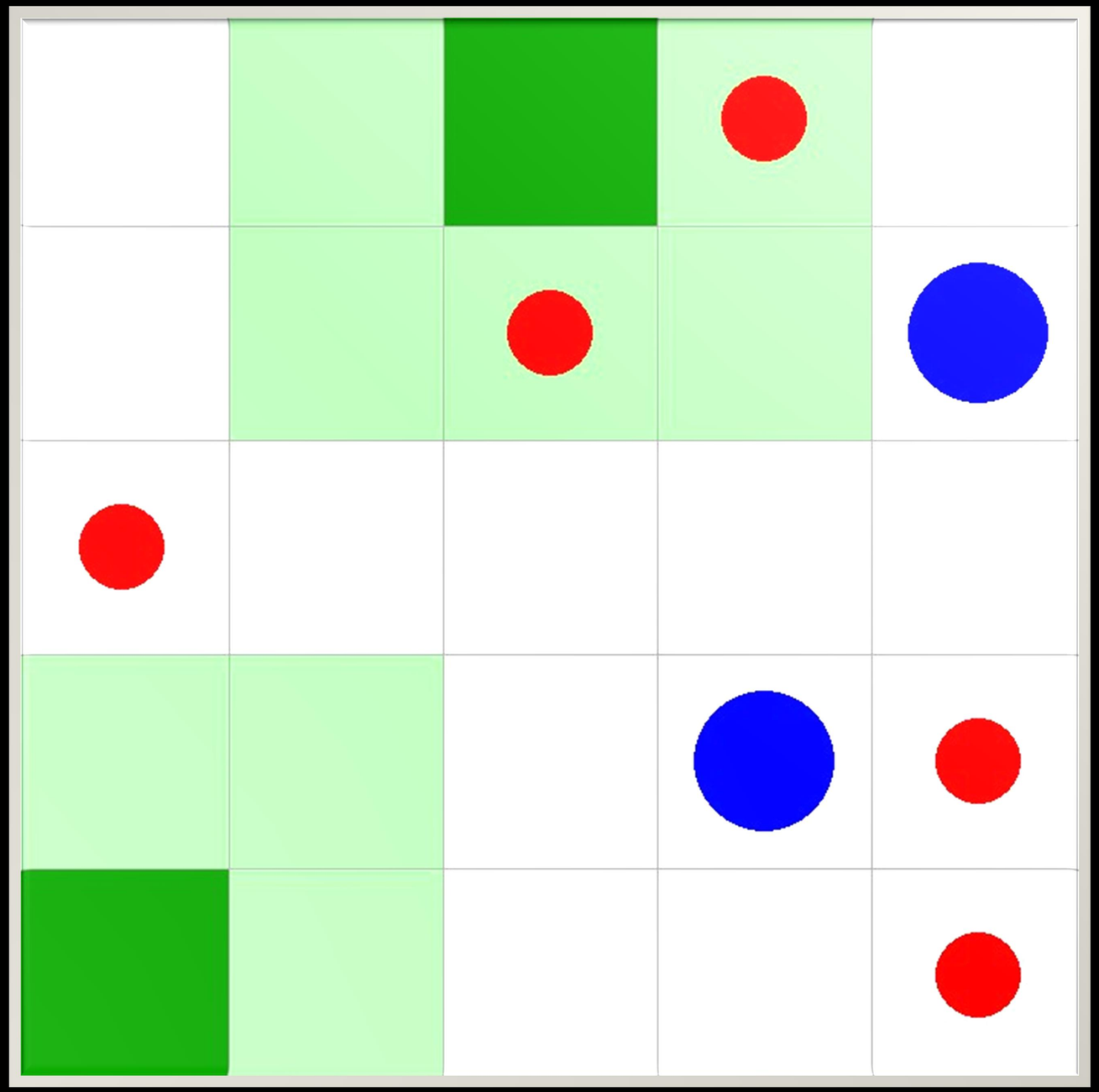
$[\, C, M, N \,]$

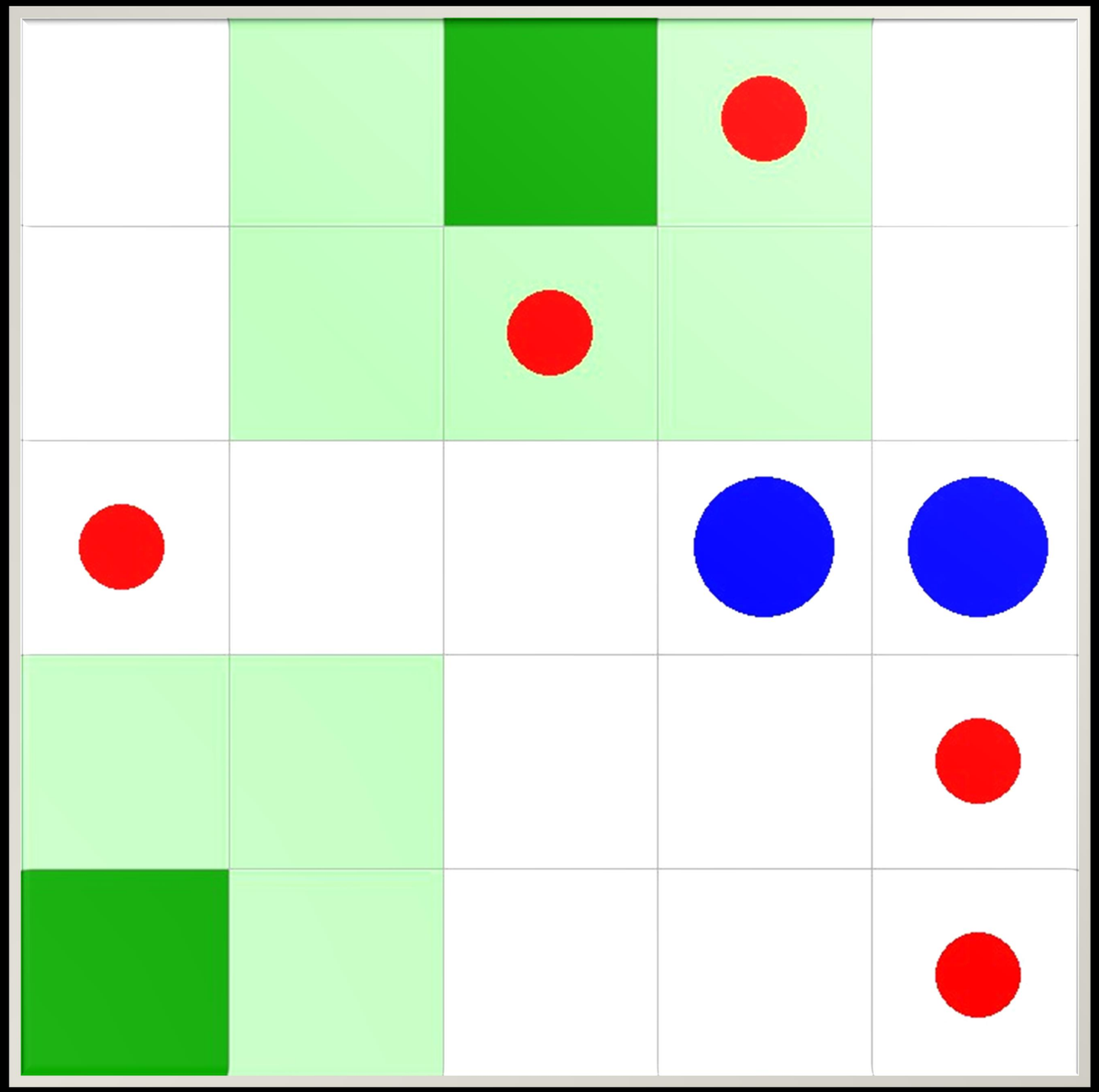
v2.5

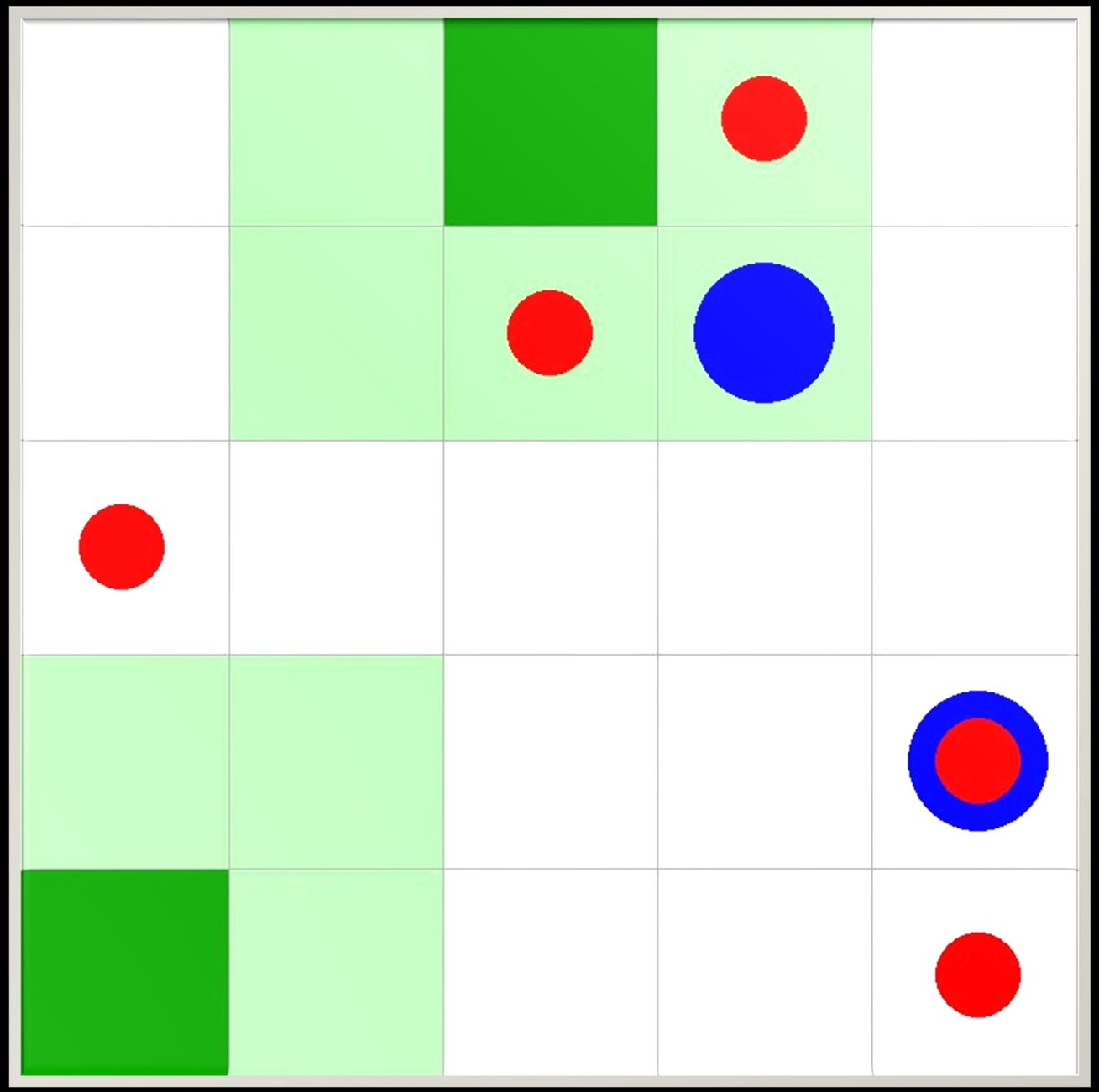
[1, 3, 2, 3, 1, 2, 1, 1, 3, 4]

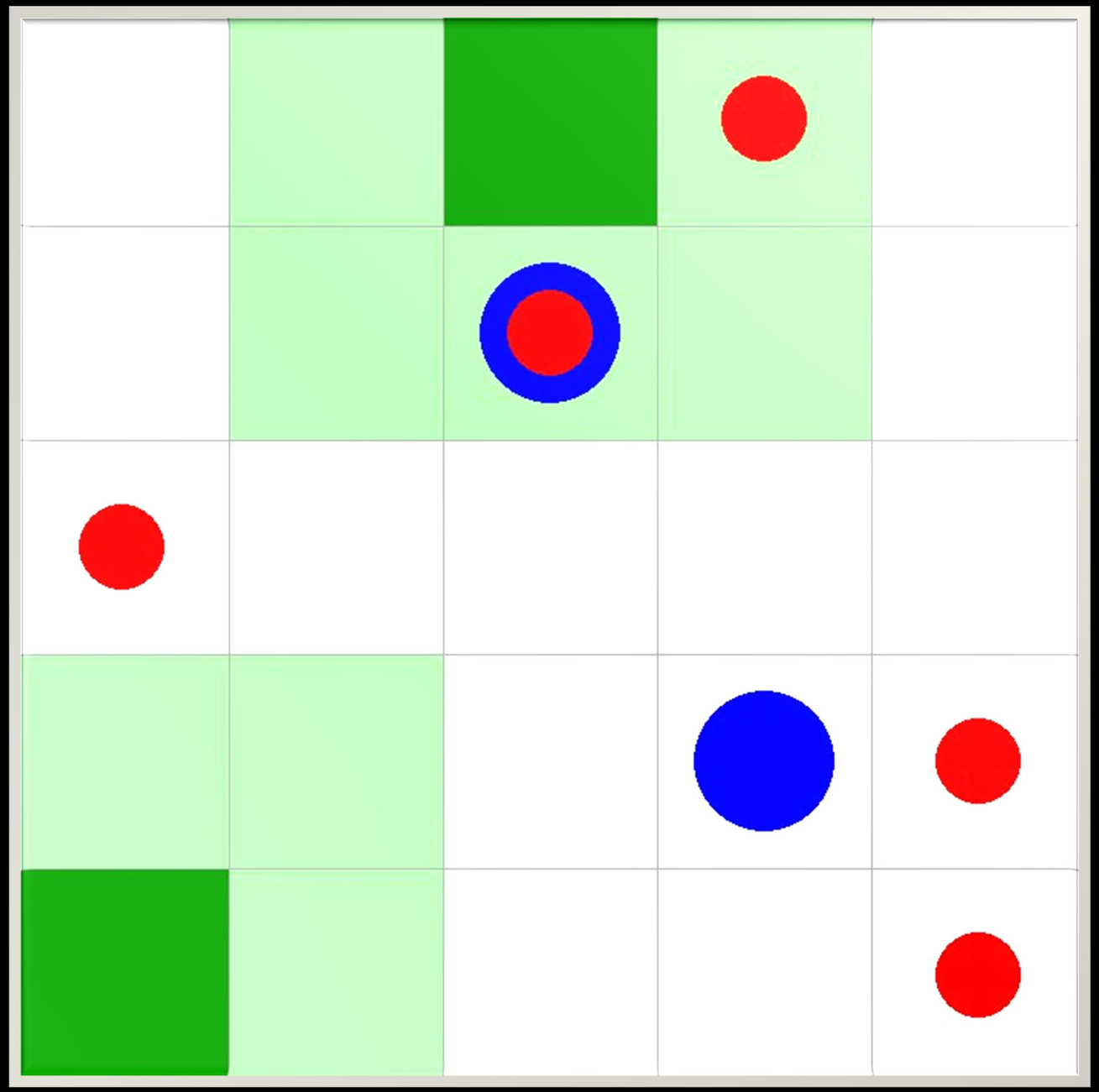
Normalize!

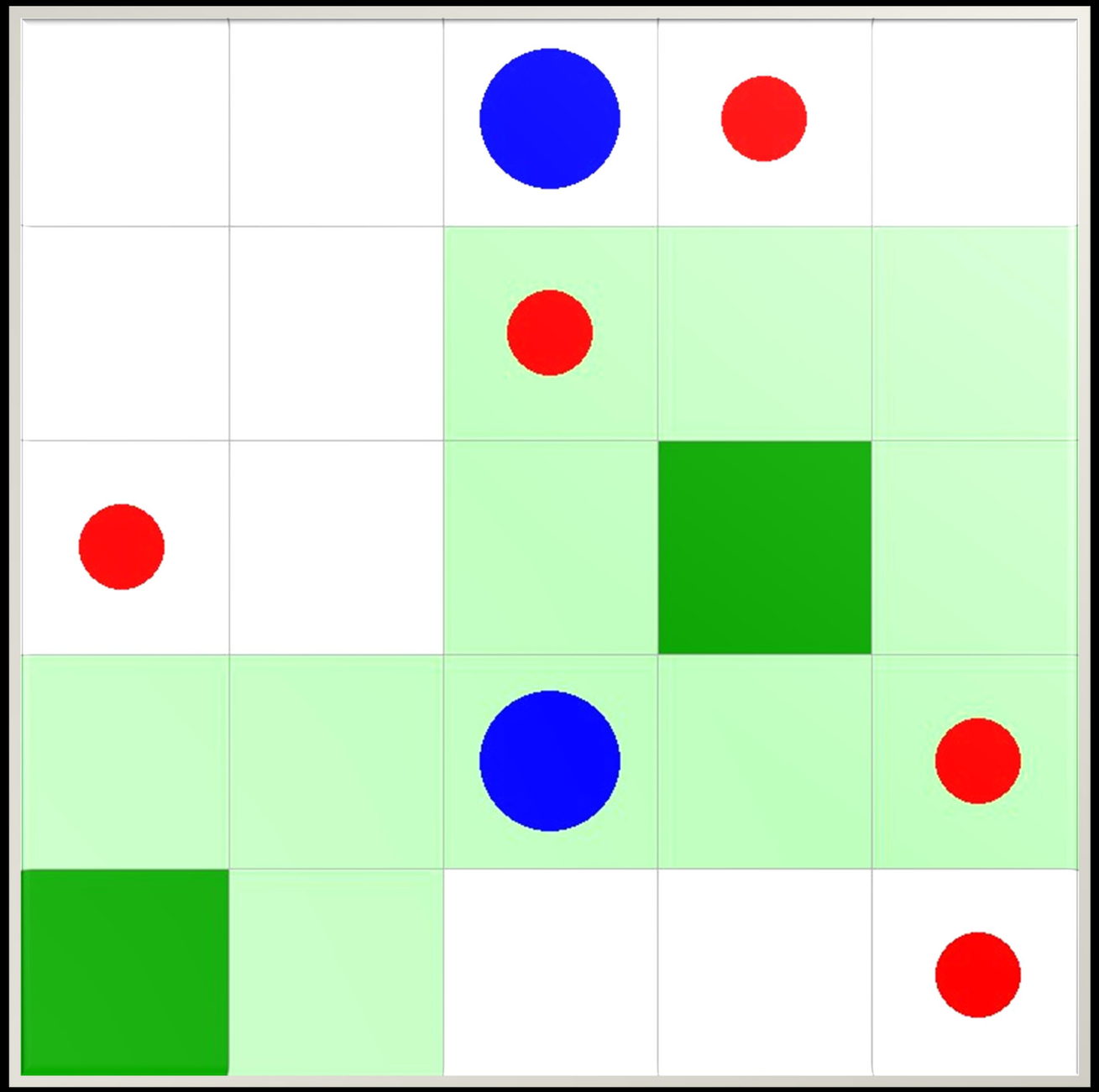


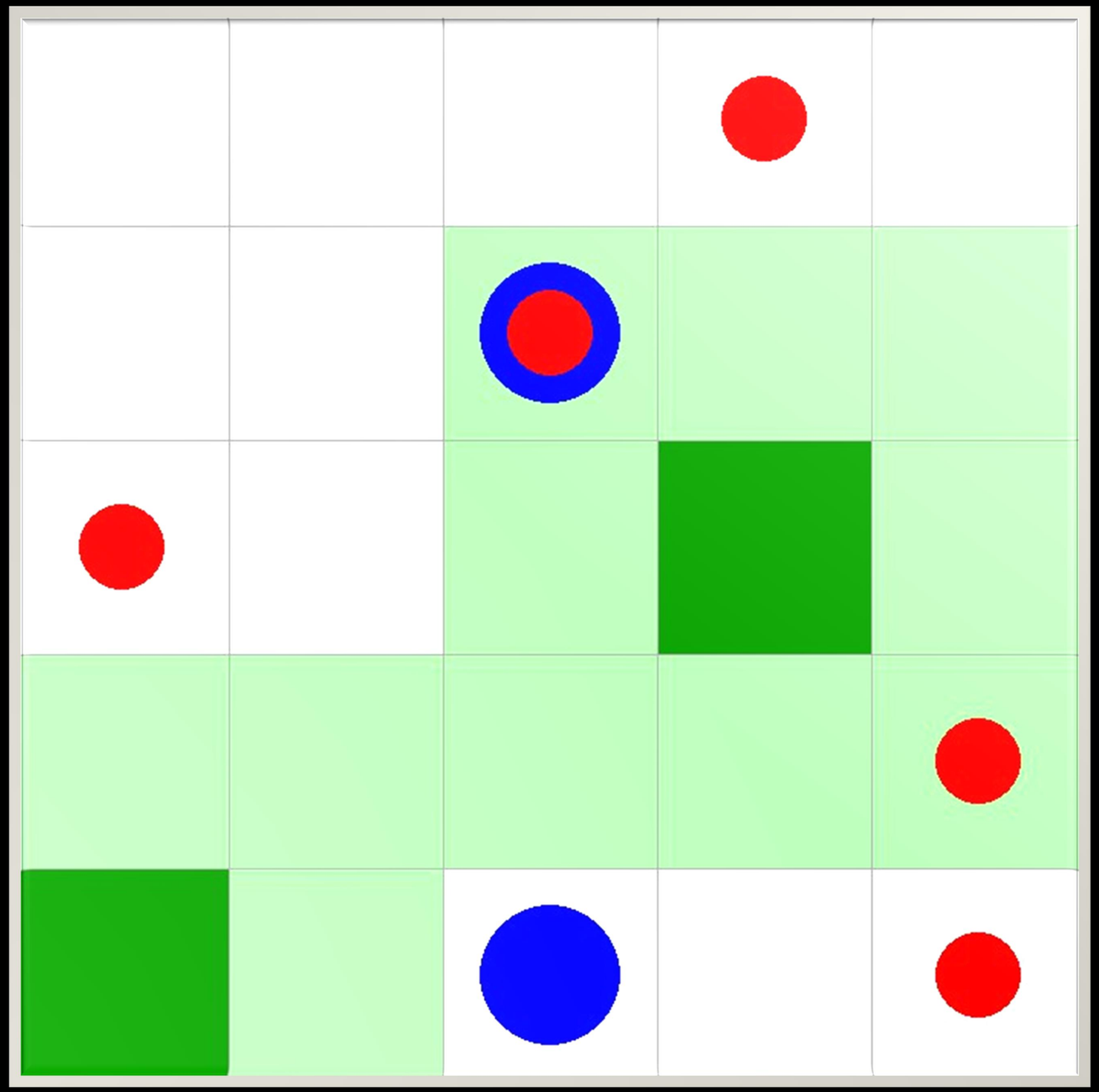


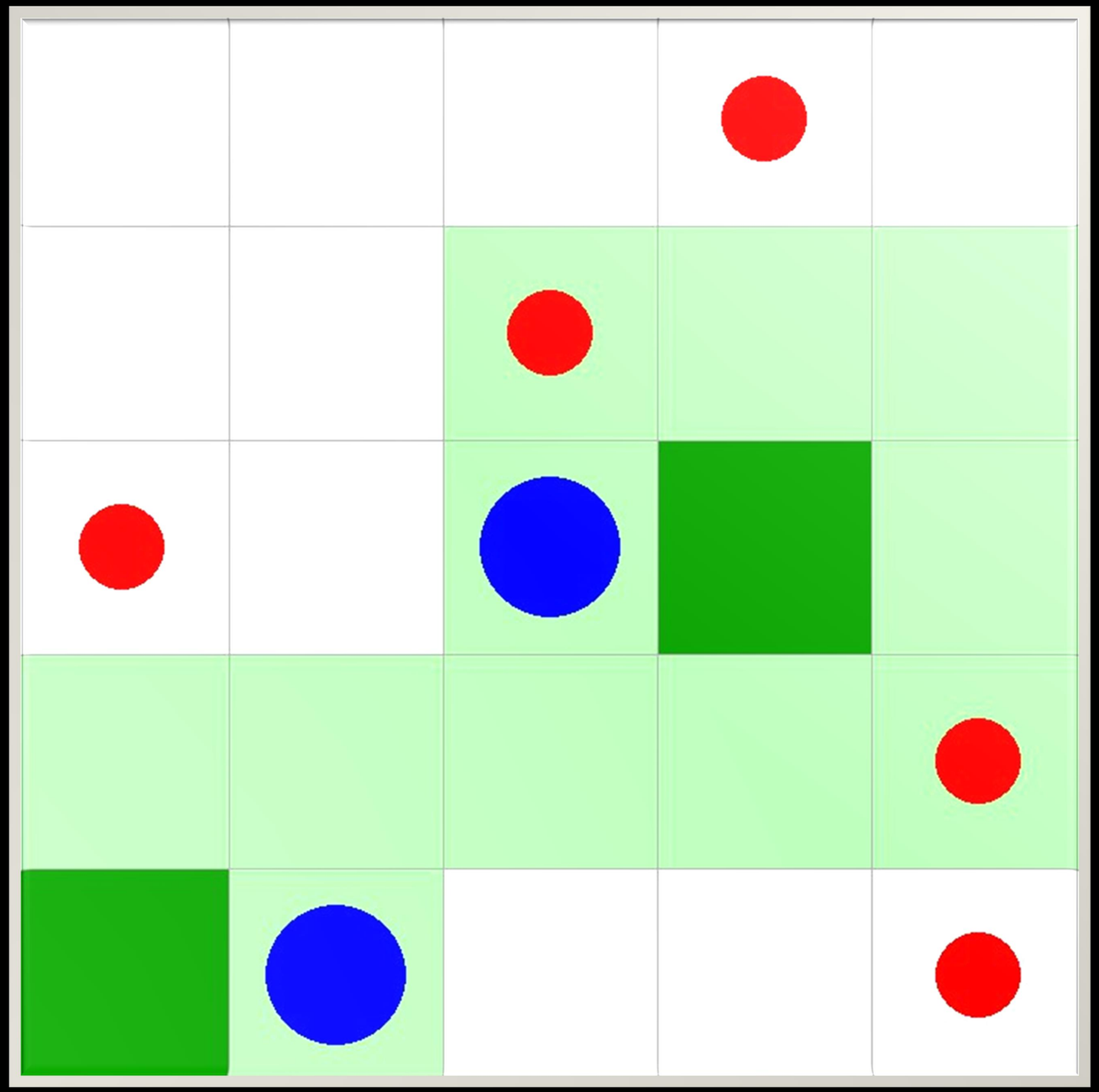


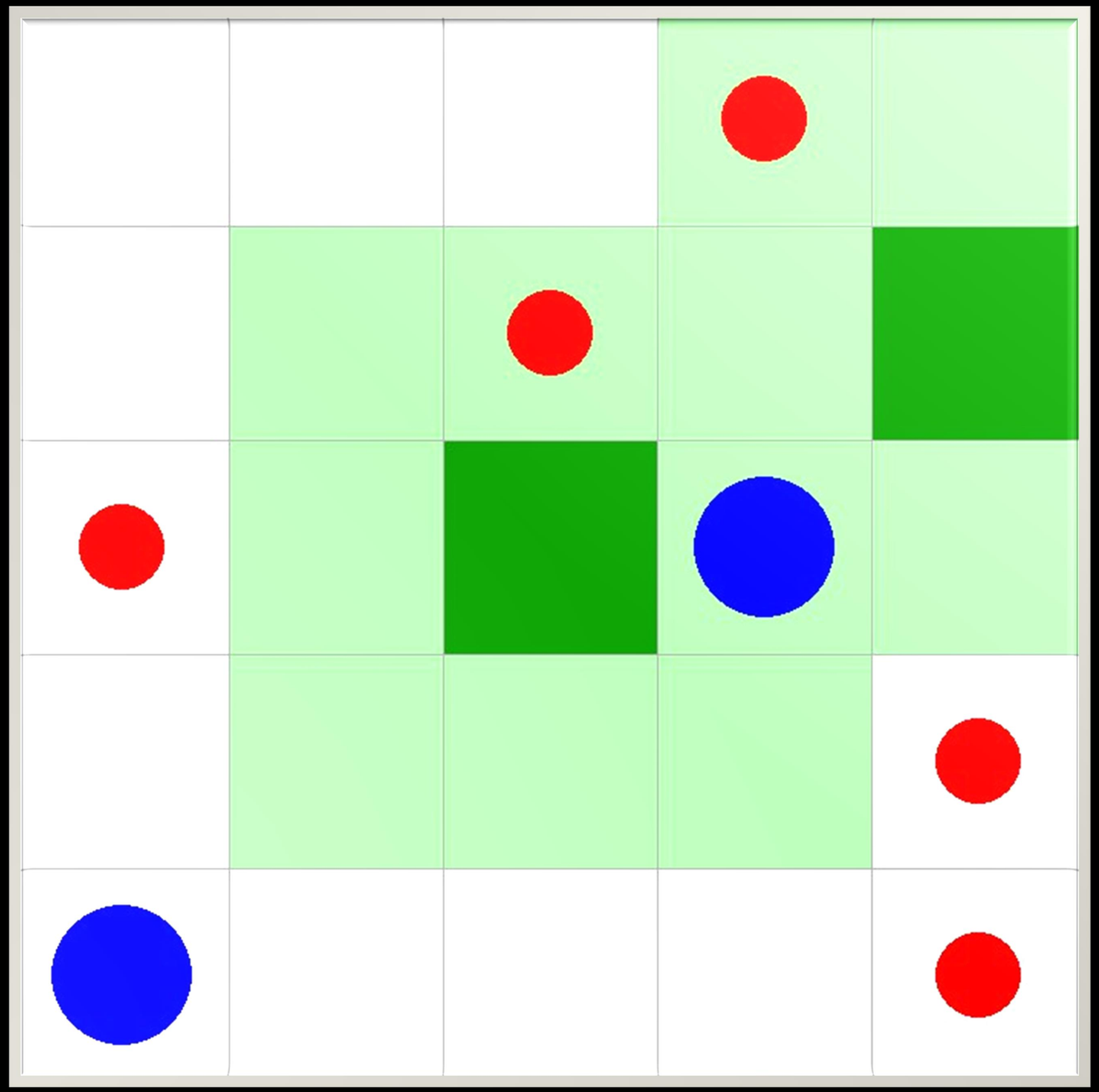


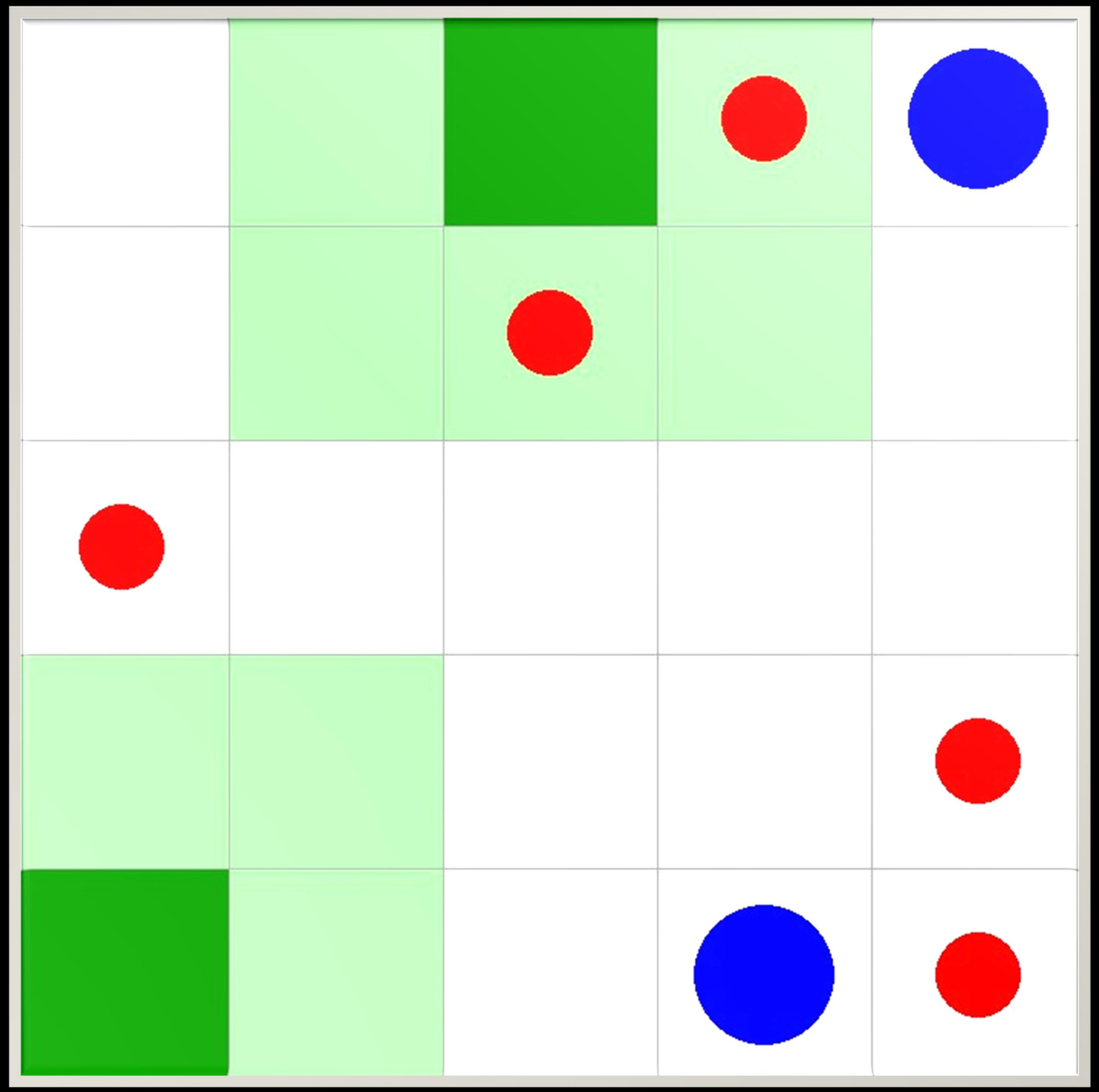


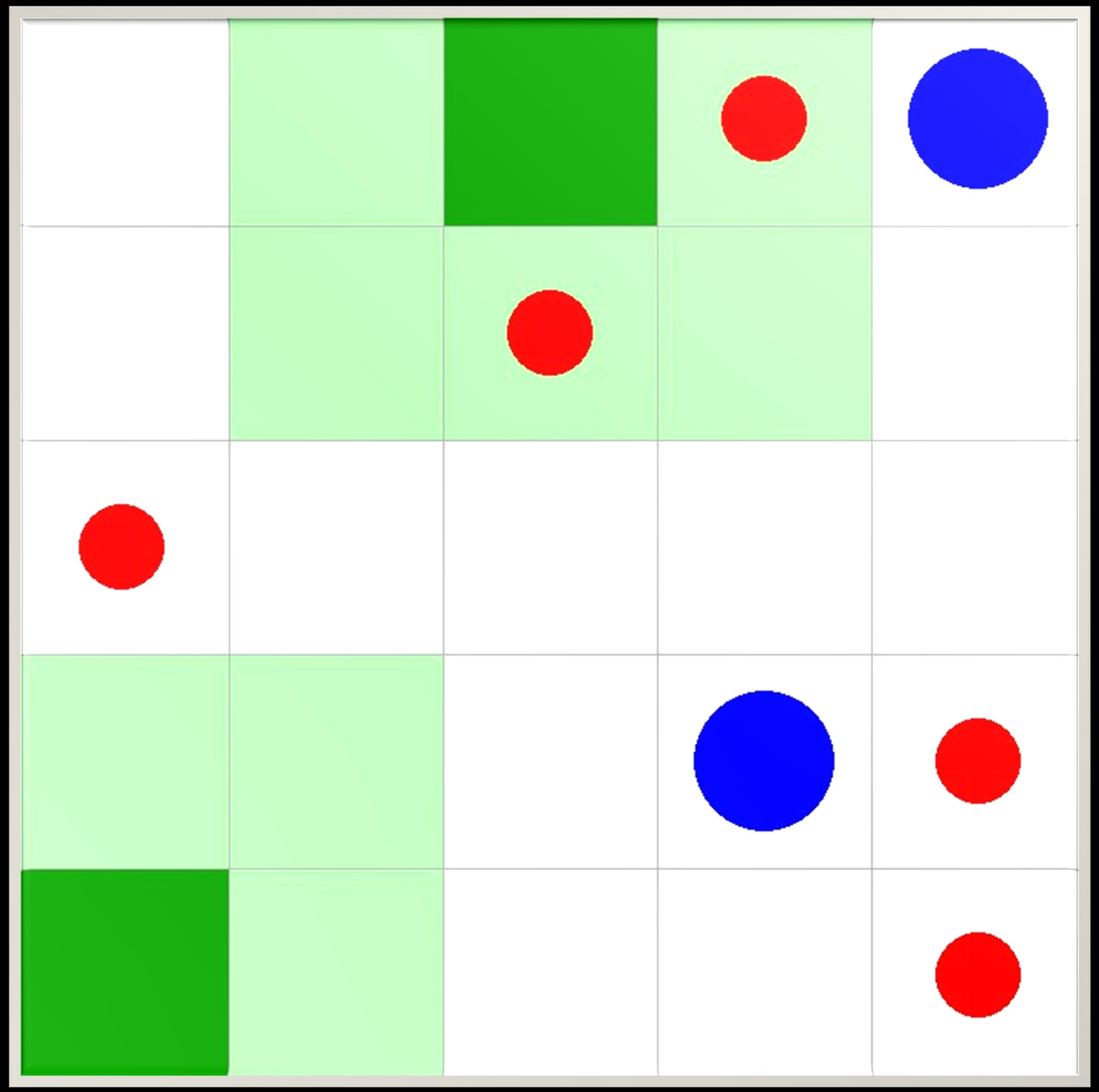


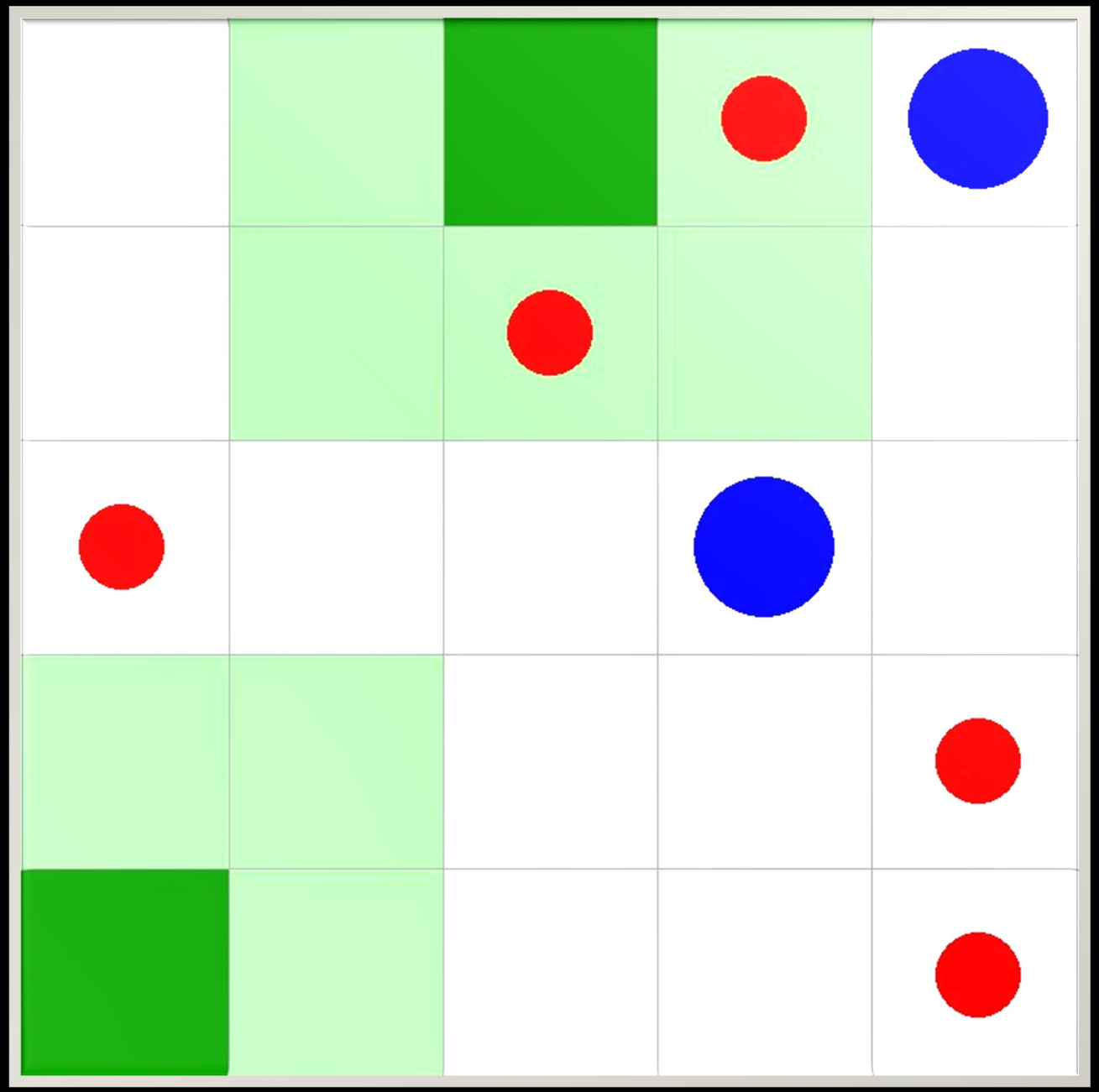


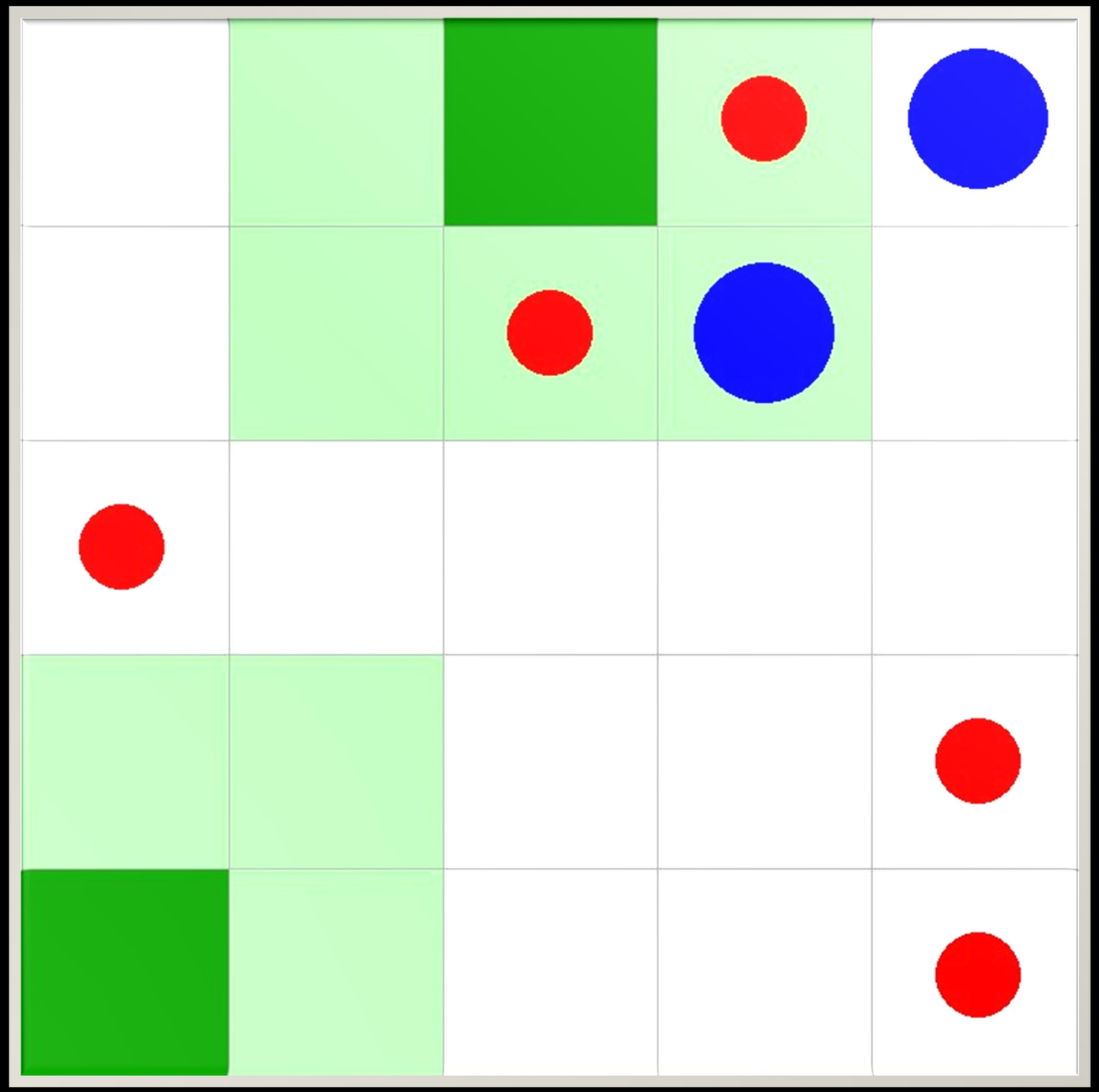


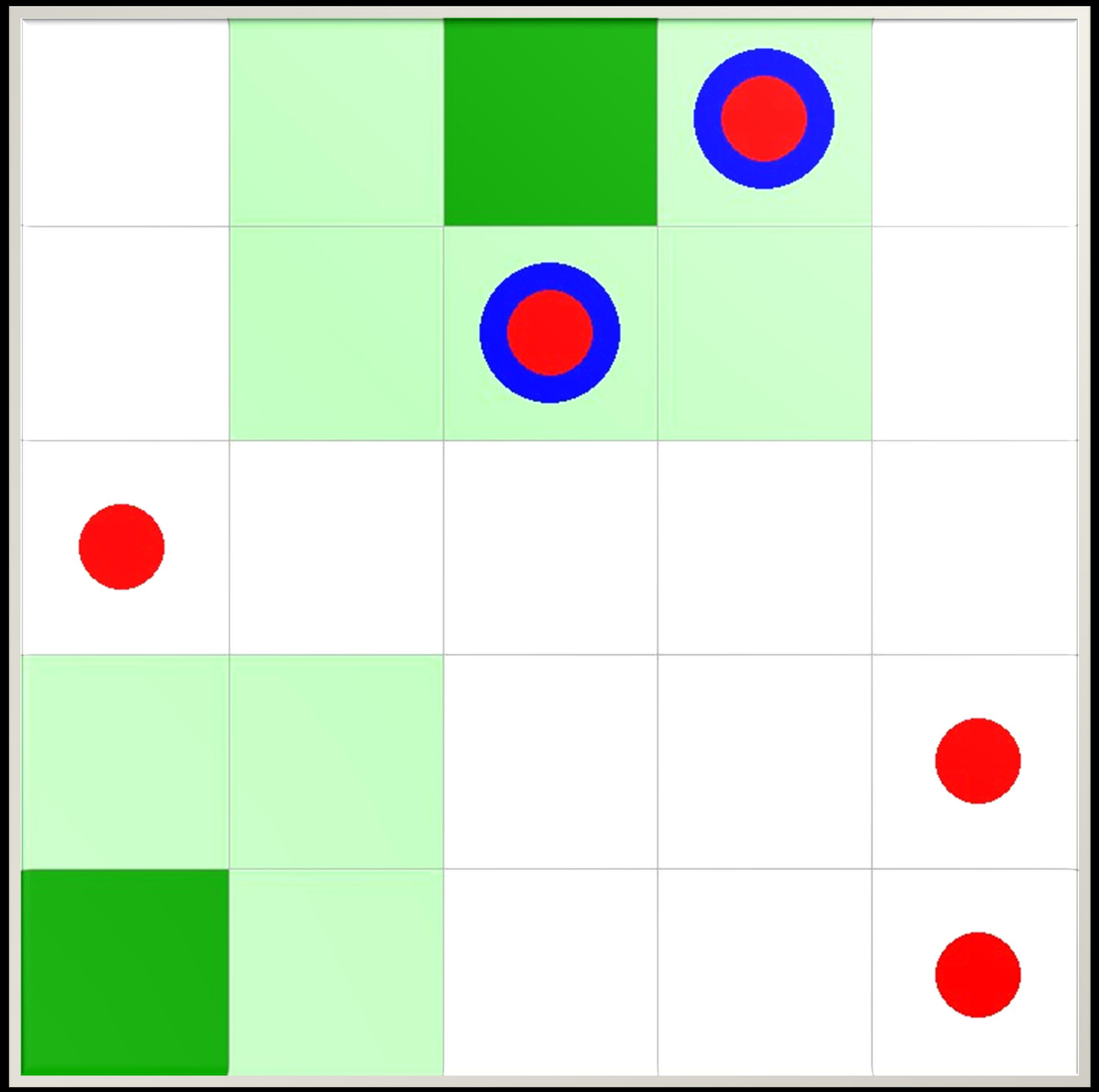


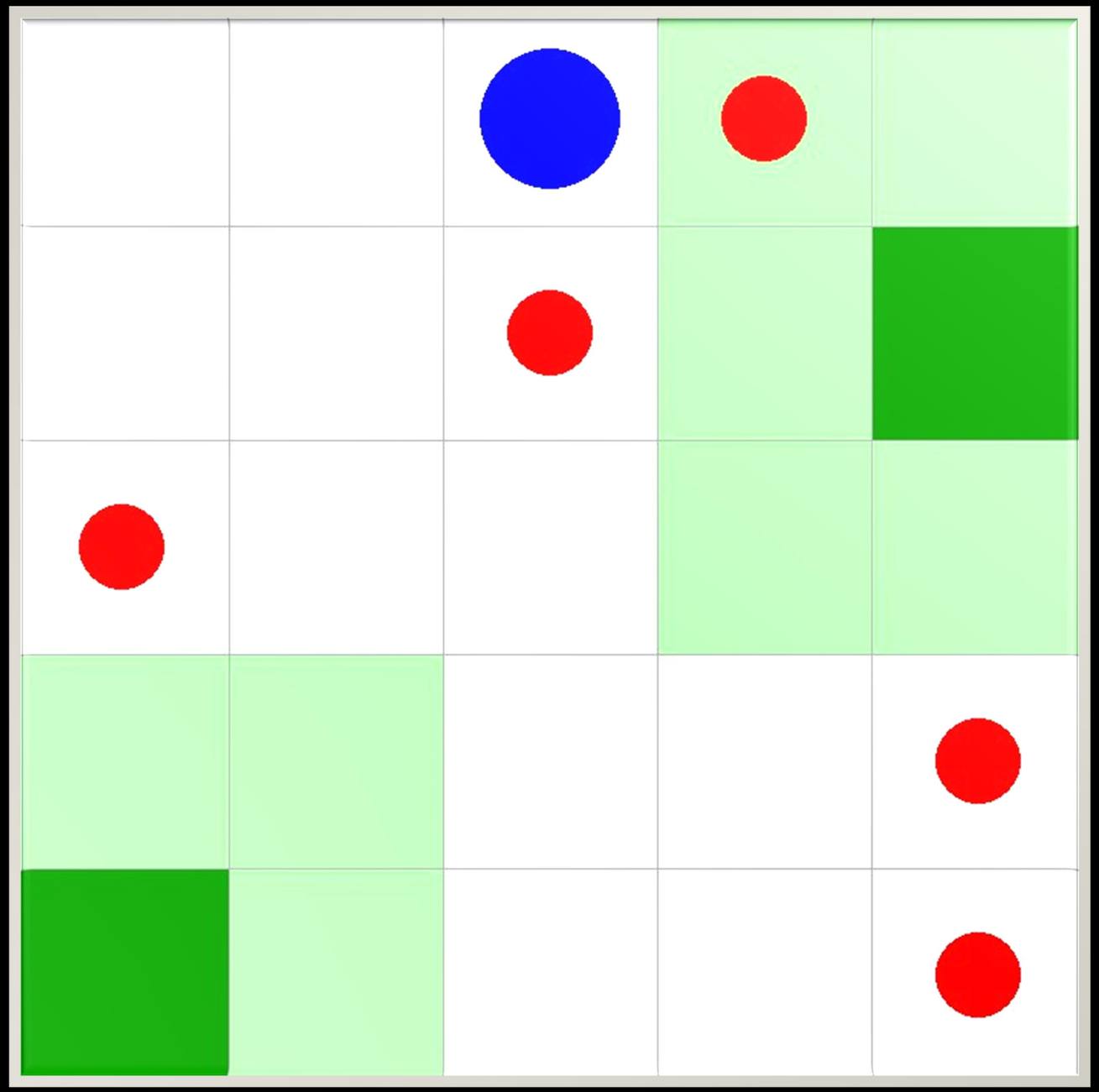


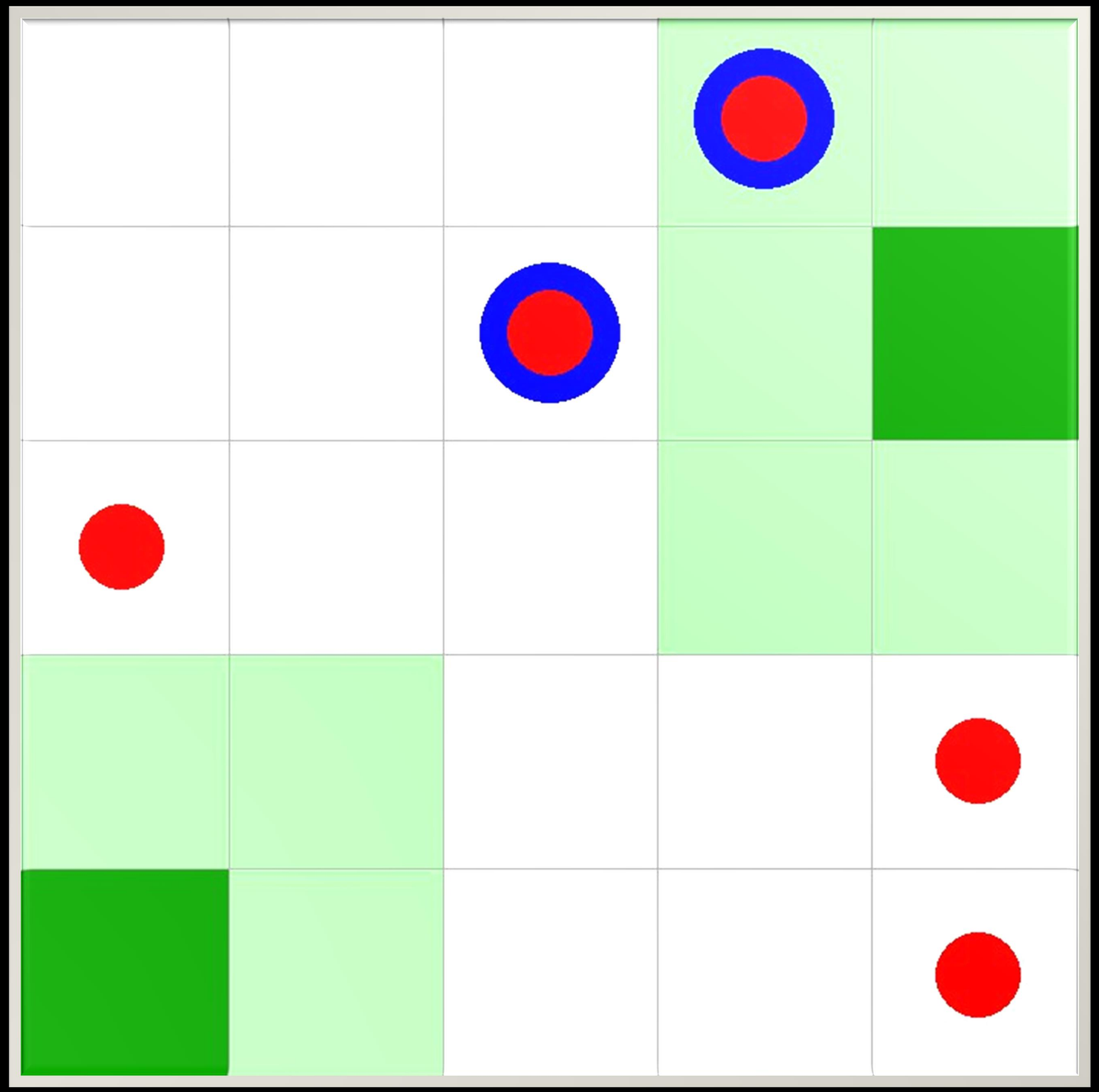


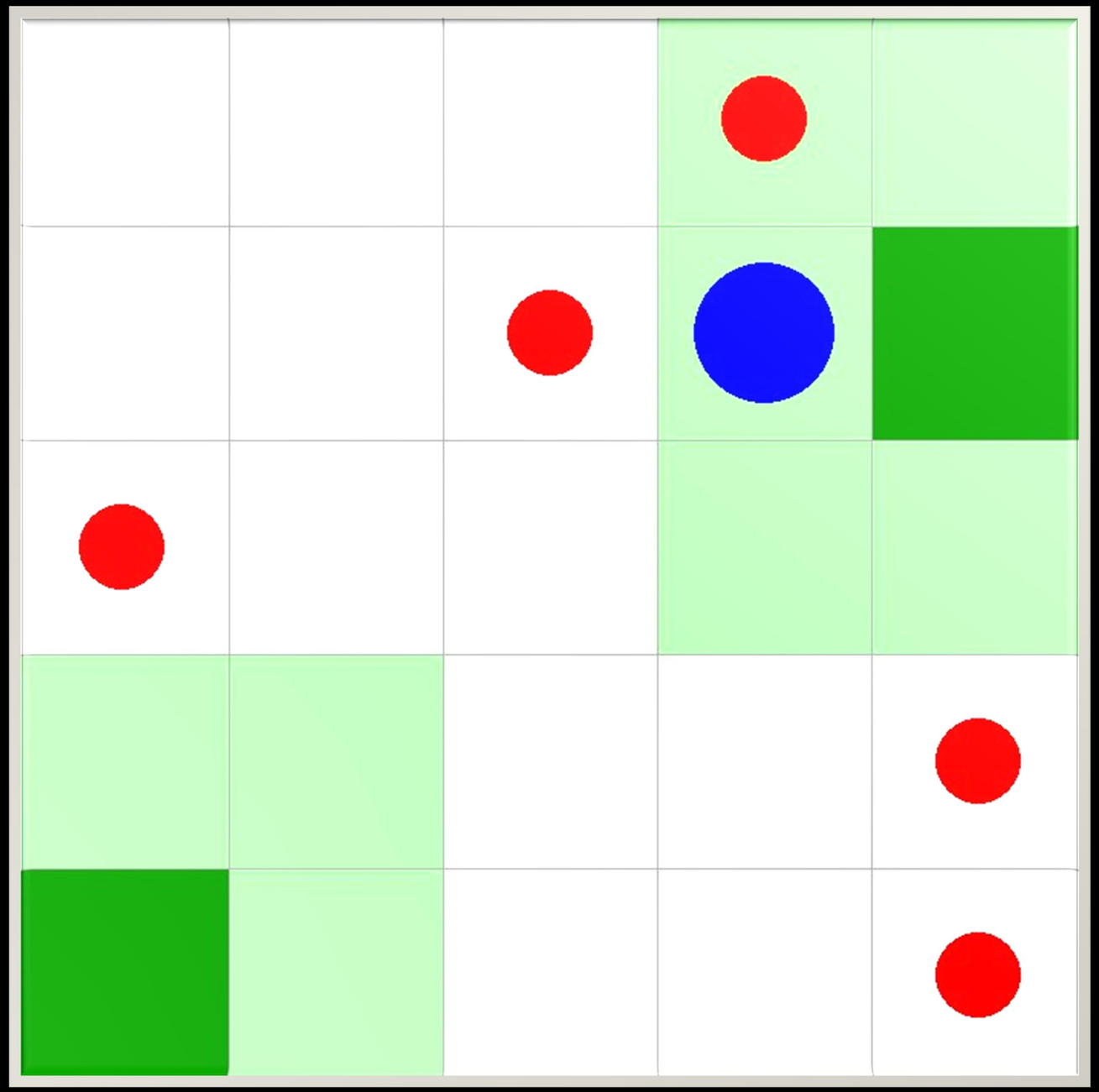


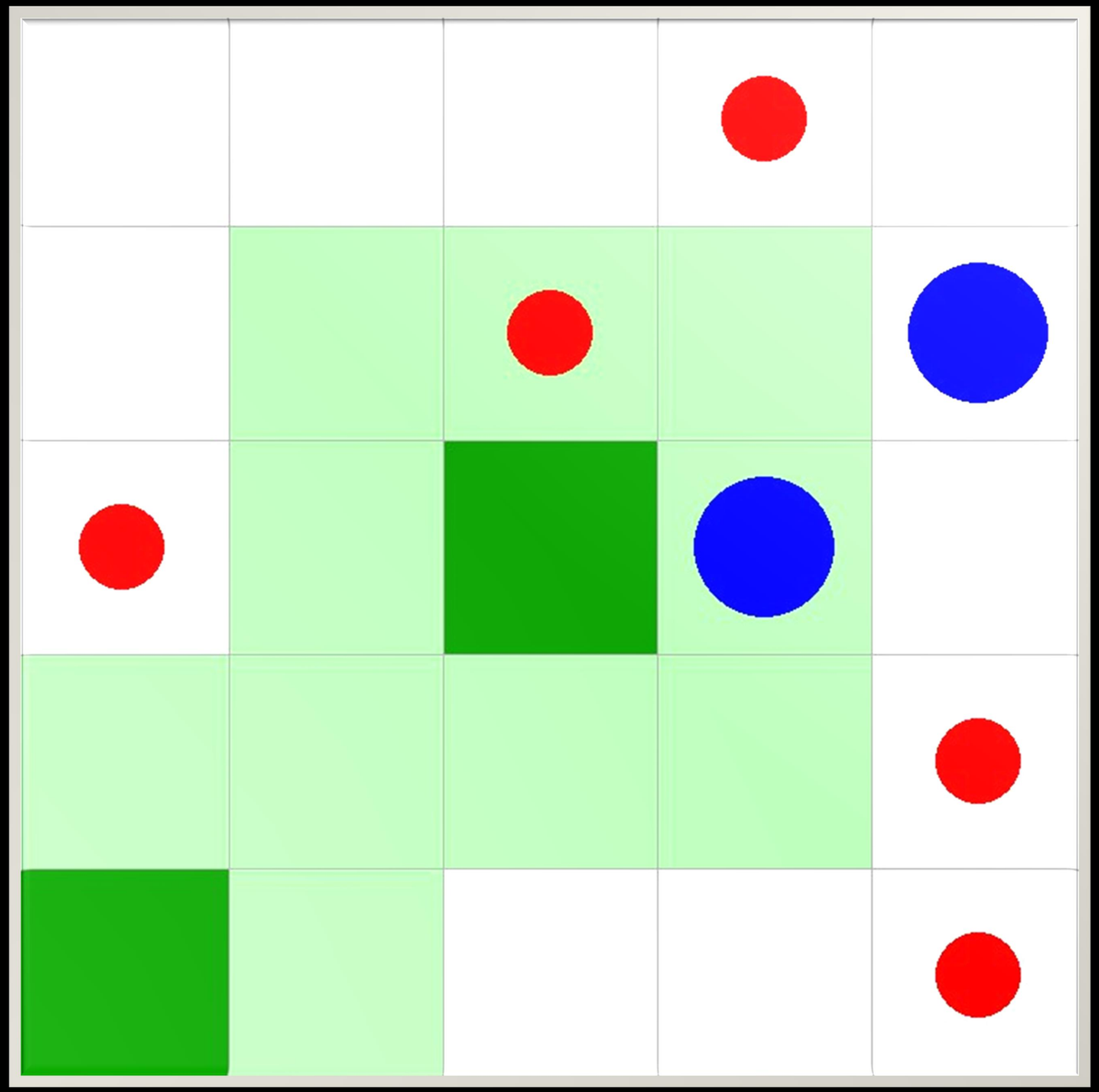






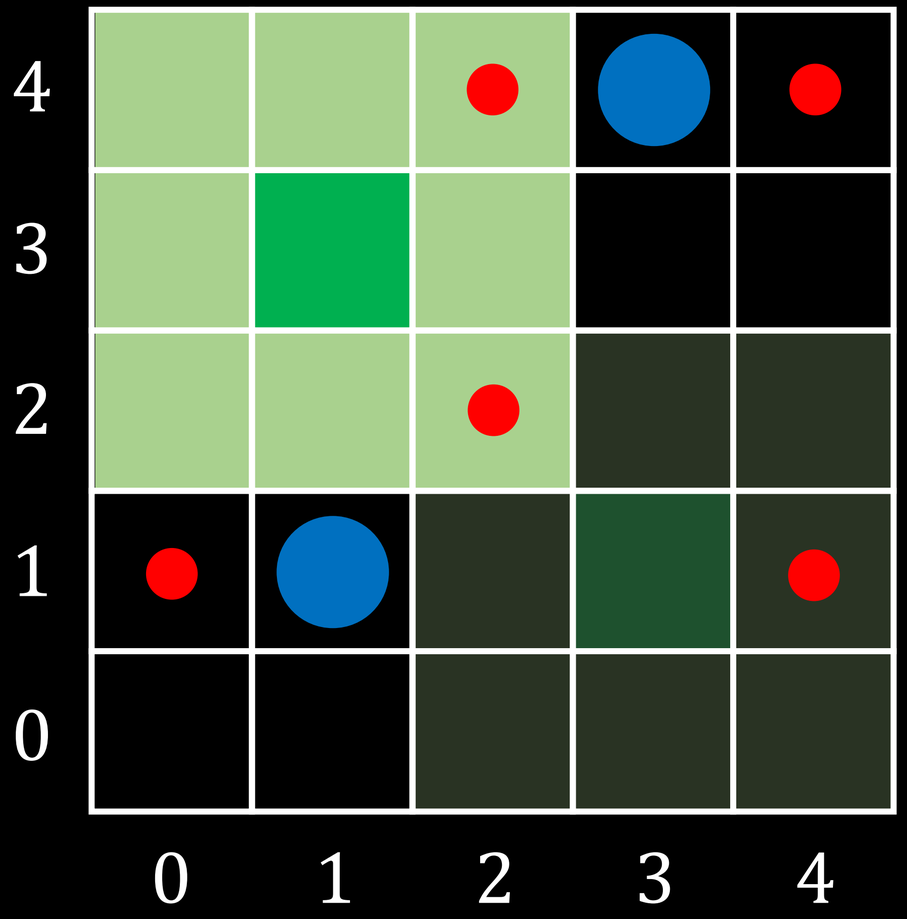






v3.0 & v3.5

$$g = 5$$
$$I = 1$$
$$p = 1$$



v3.0

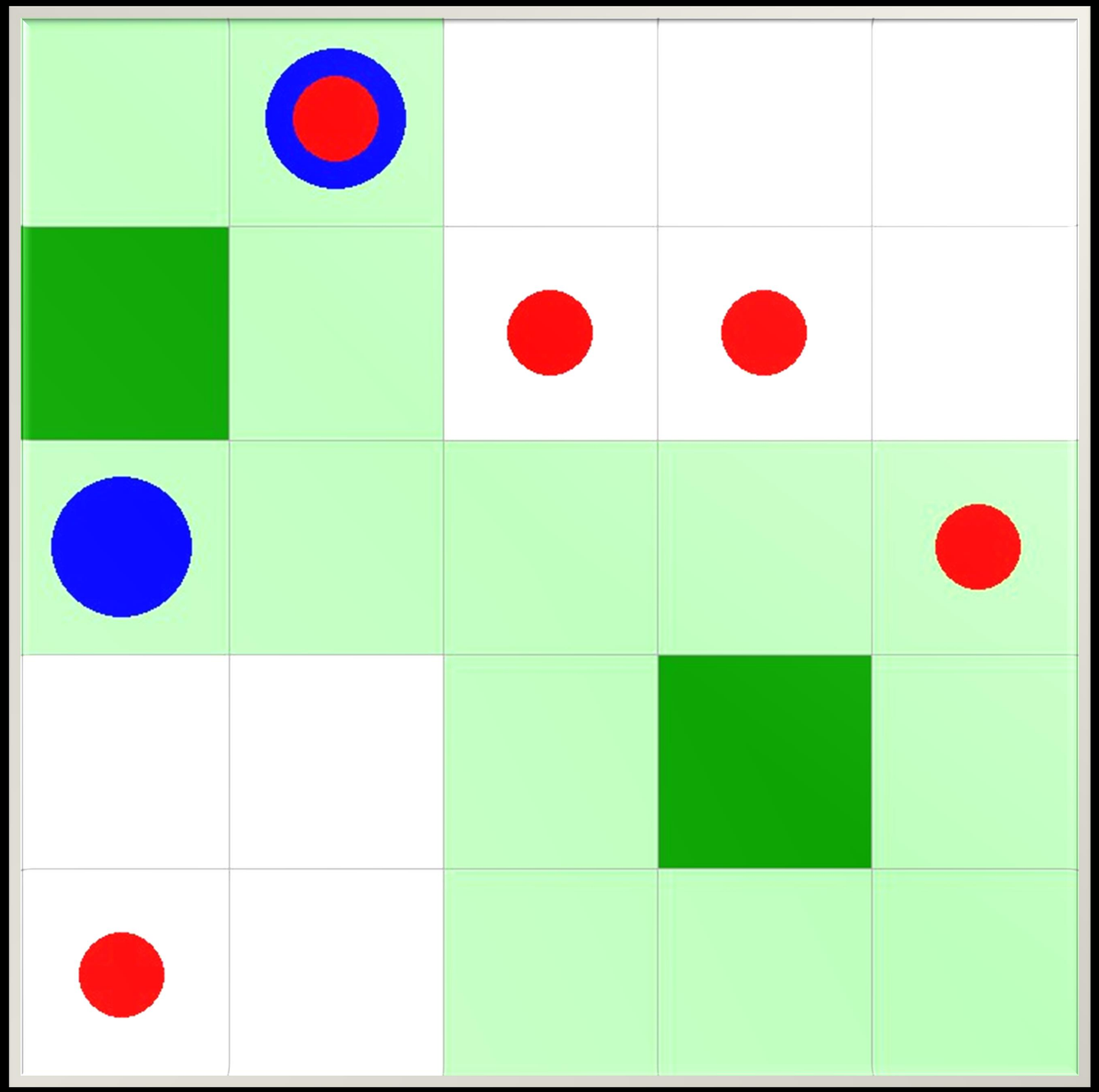
[1, 3, 2, −1, −1, 0, 1, 1, 3, 4]

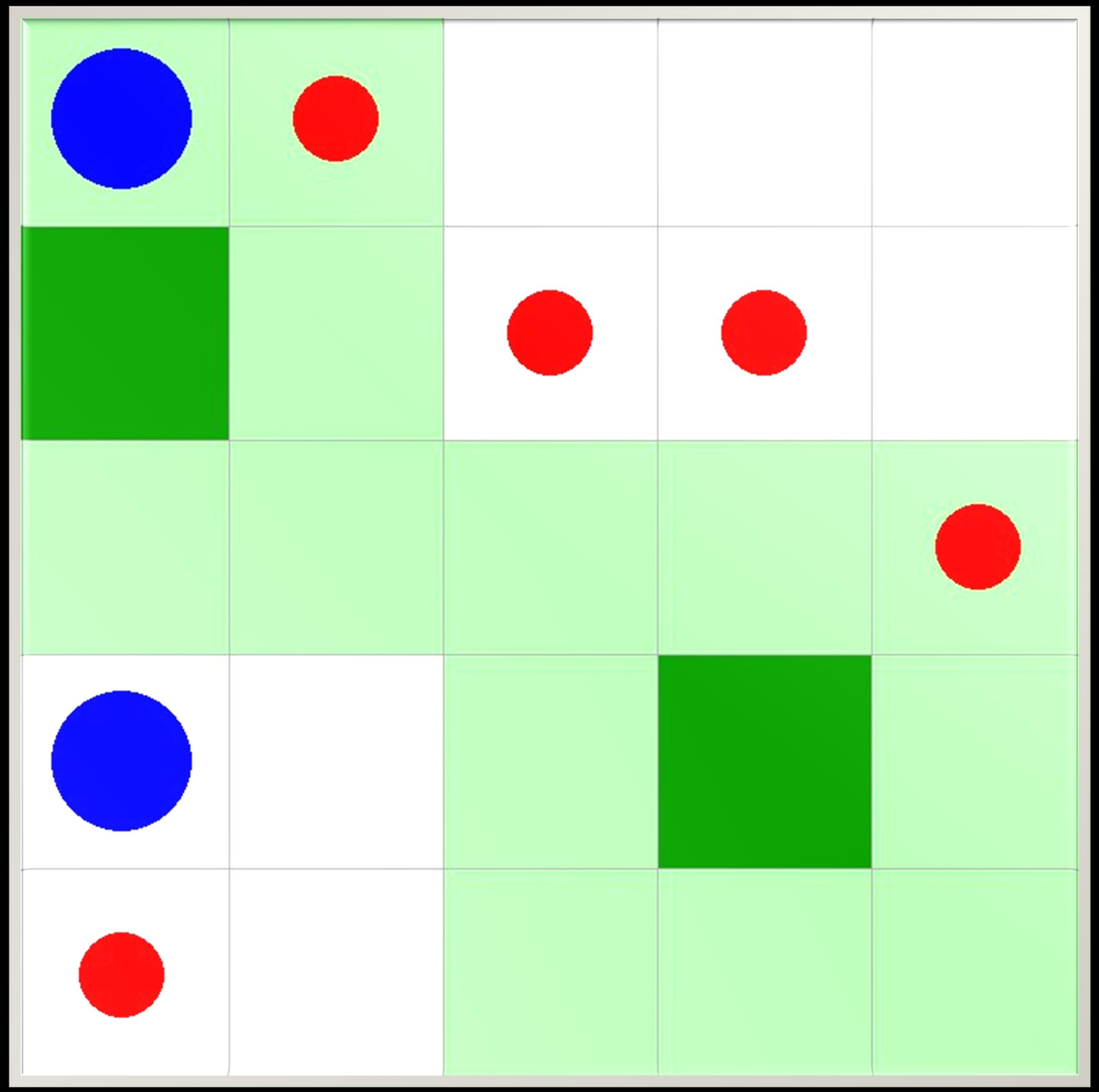
Normalize!

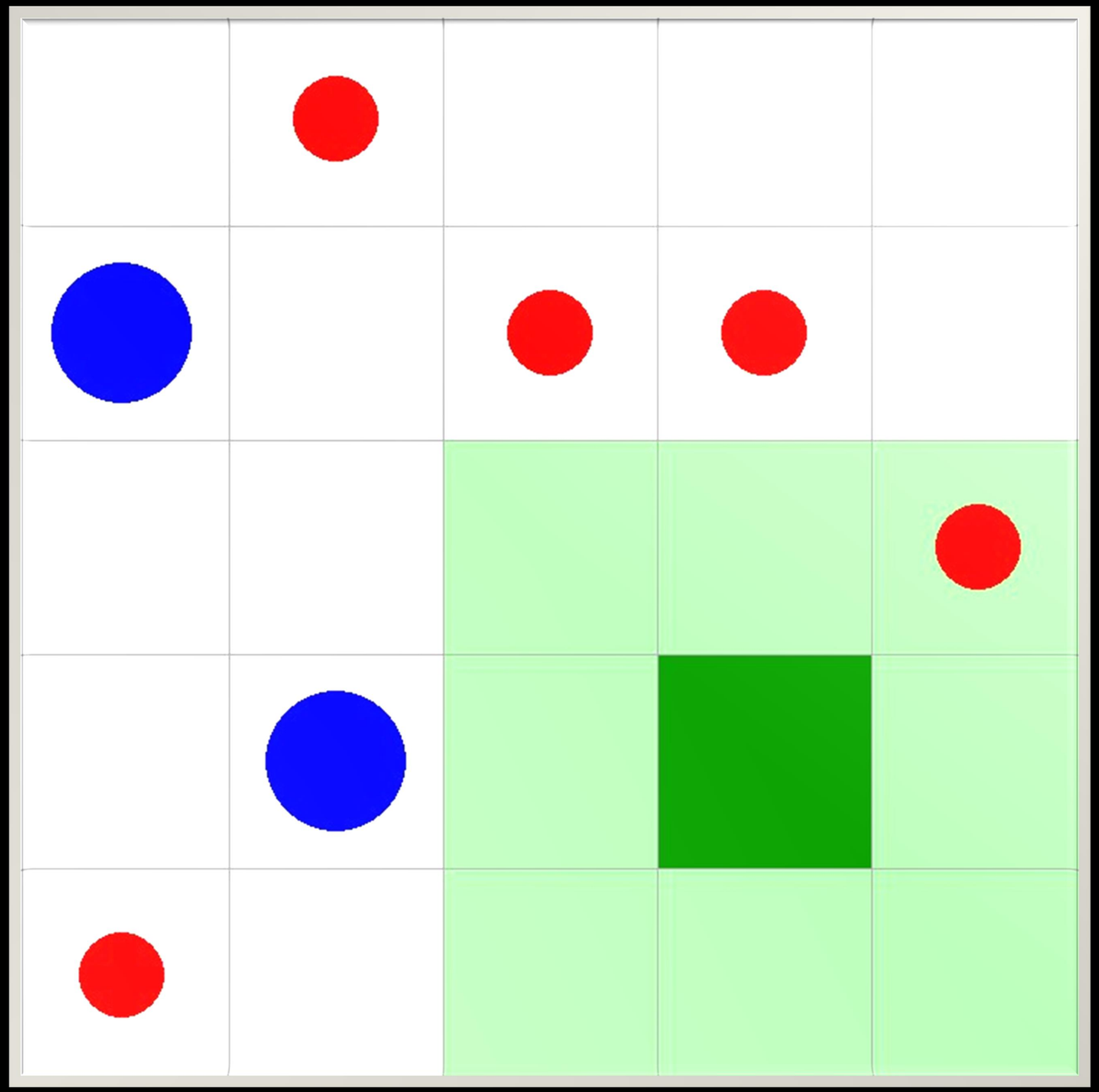
v3.5

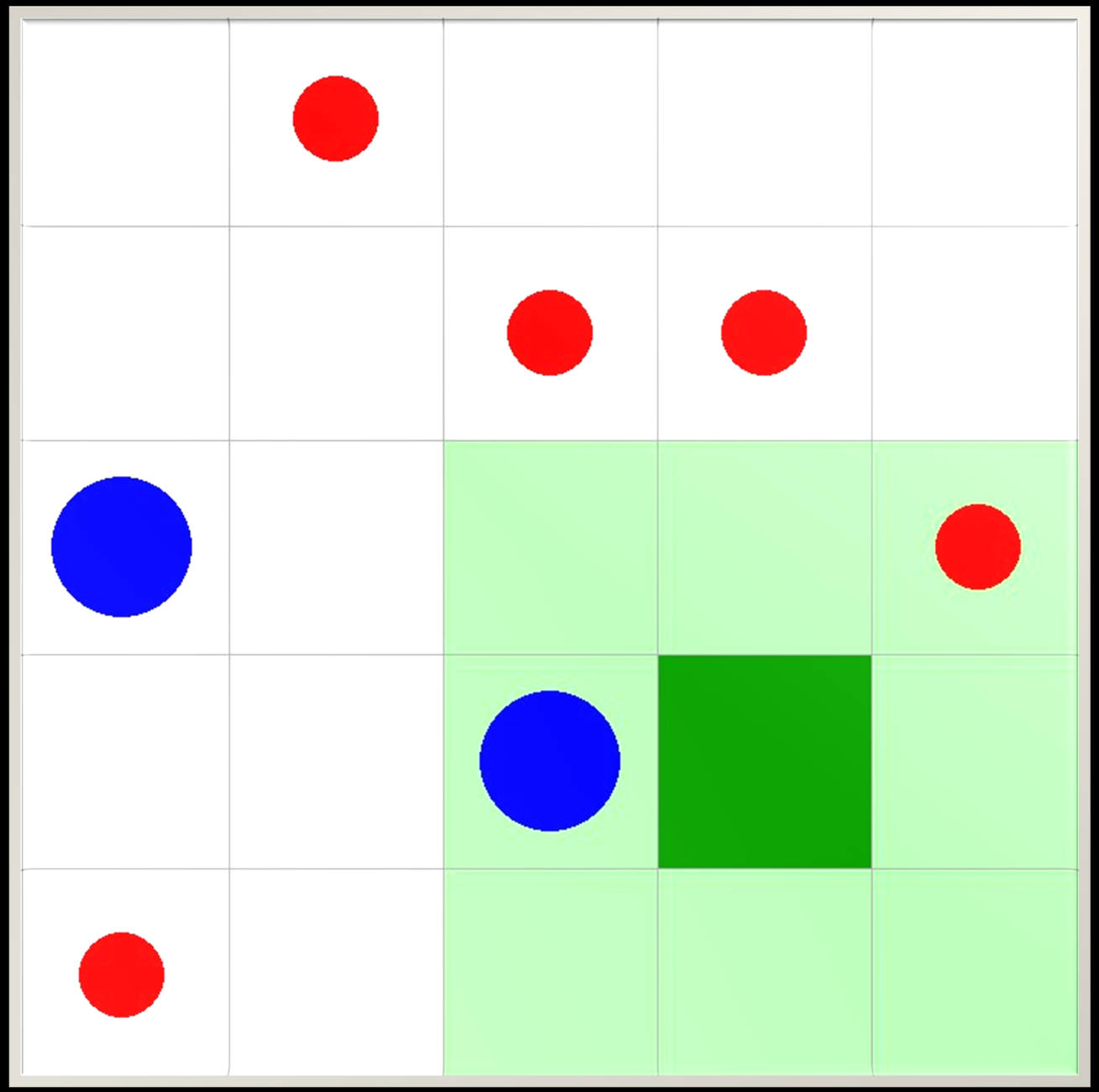
[1, 3, 0, 2, 3, 1, 1, 0, 1, 1, 3, 4]

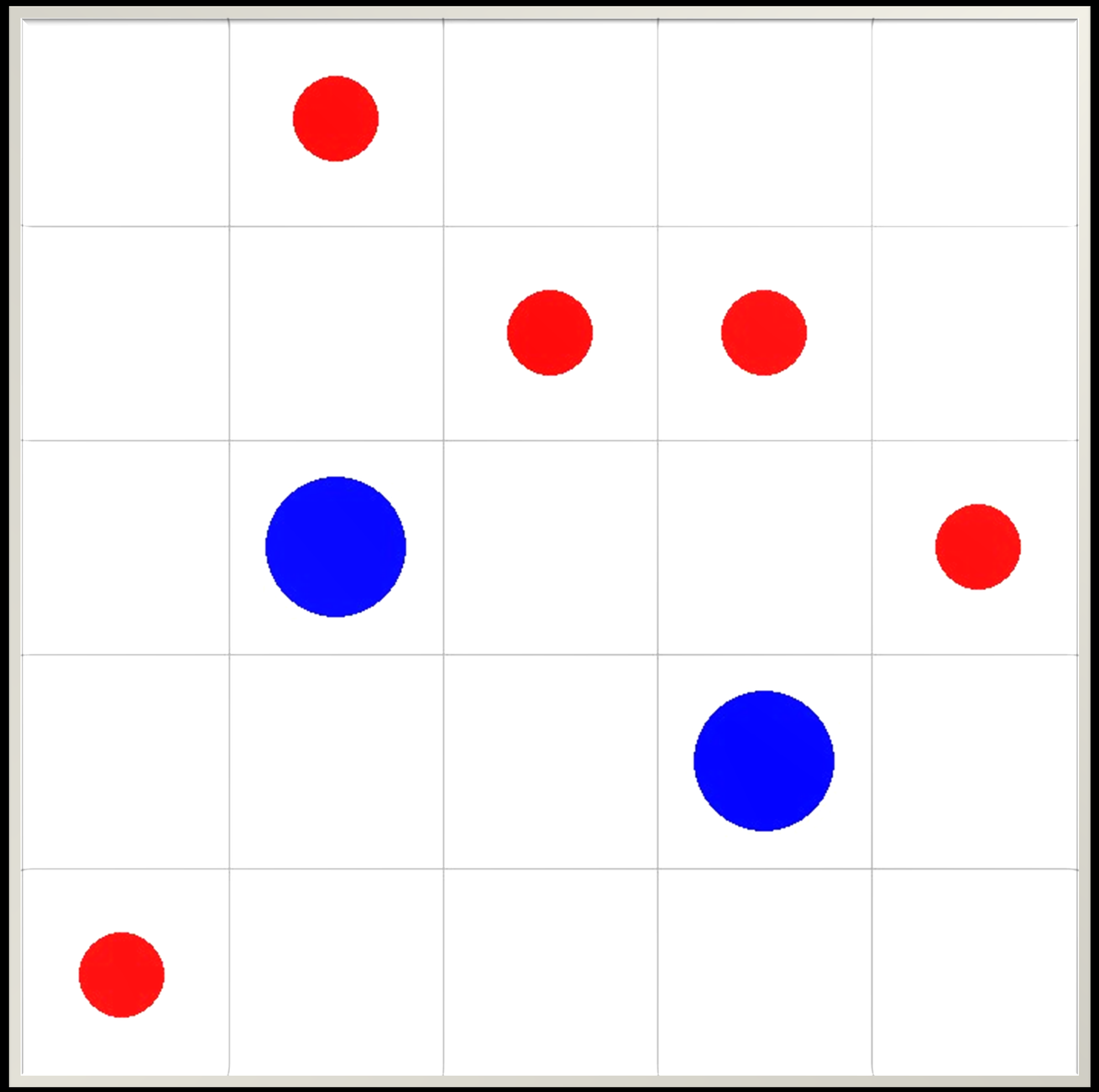
Normalize!

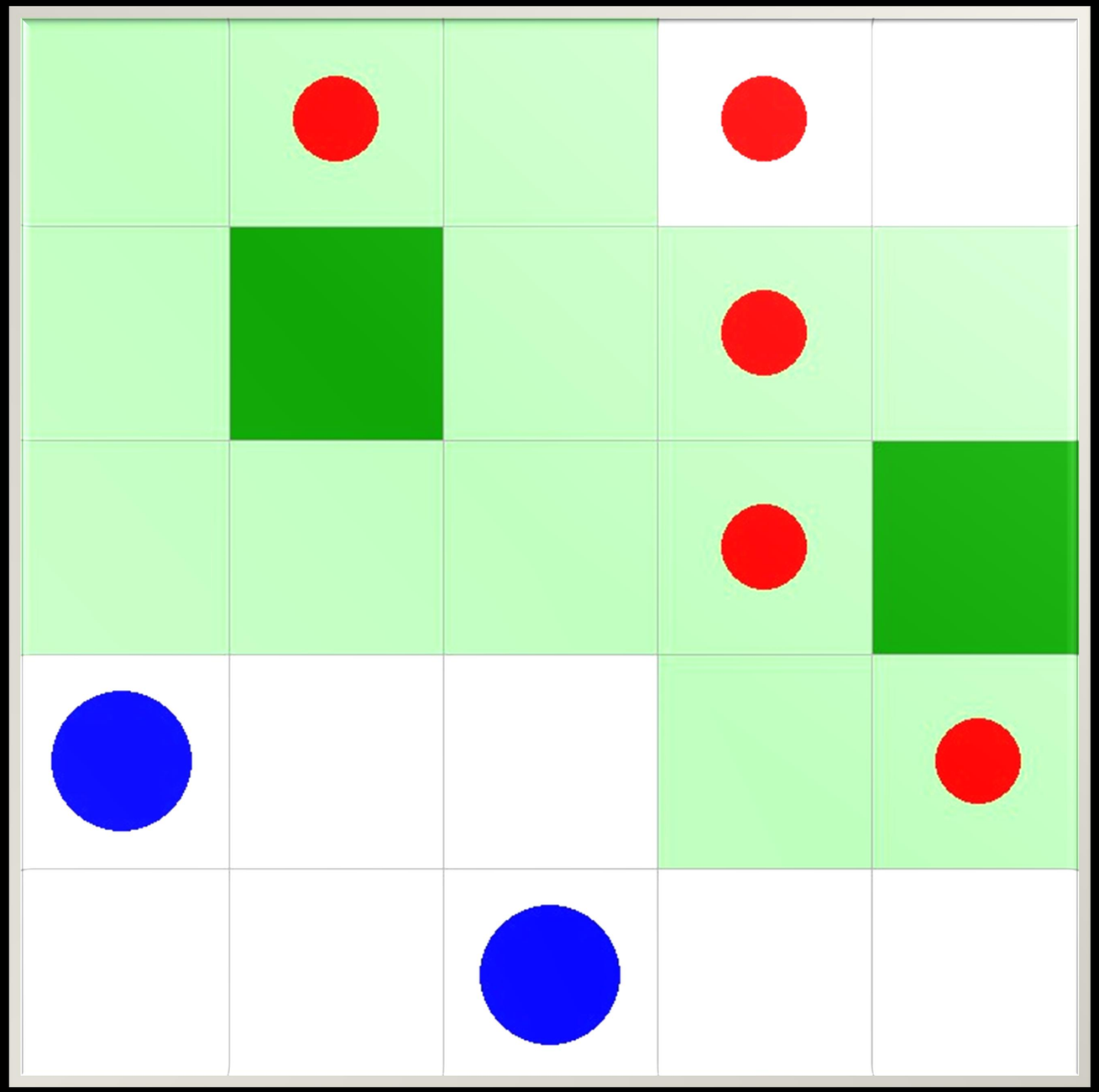


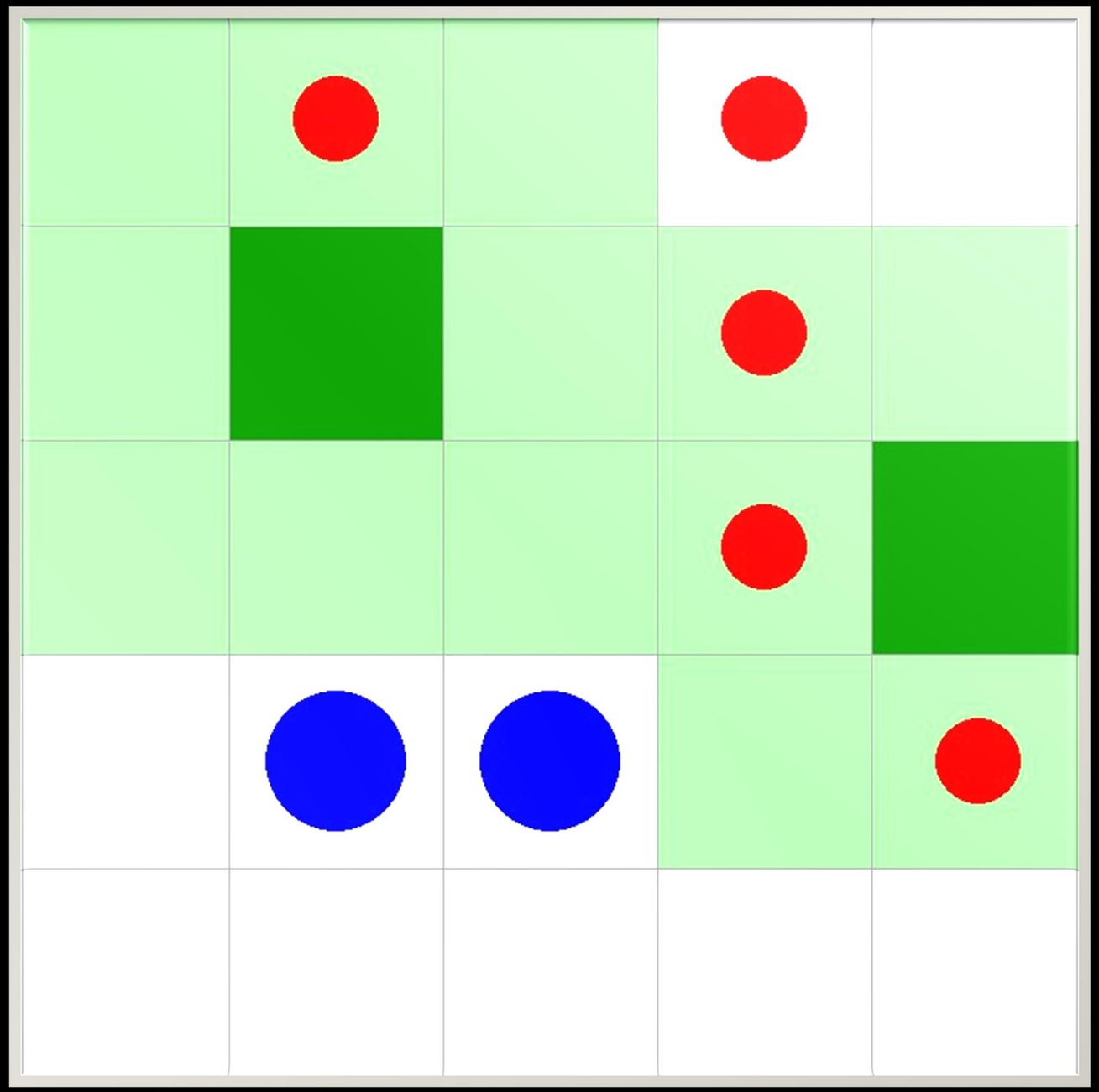


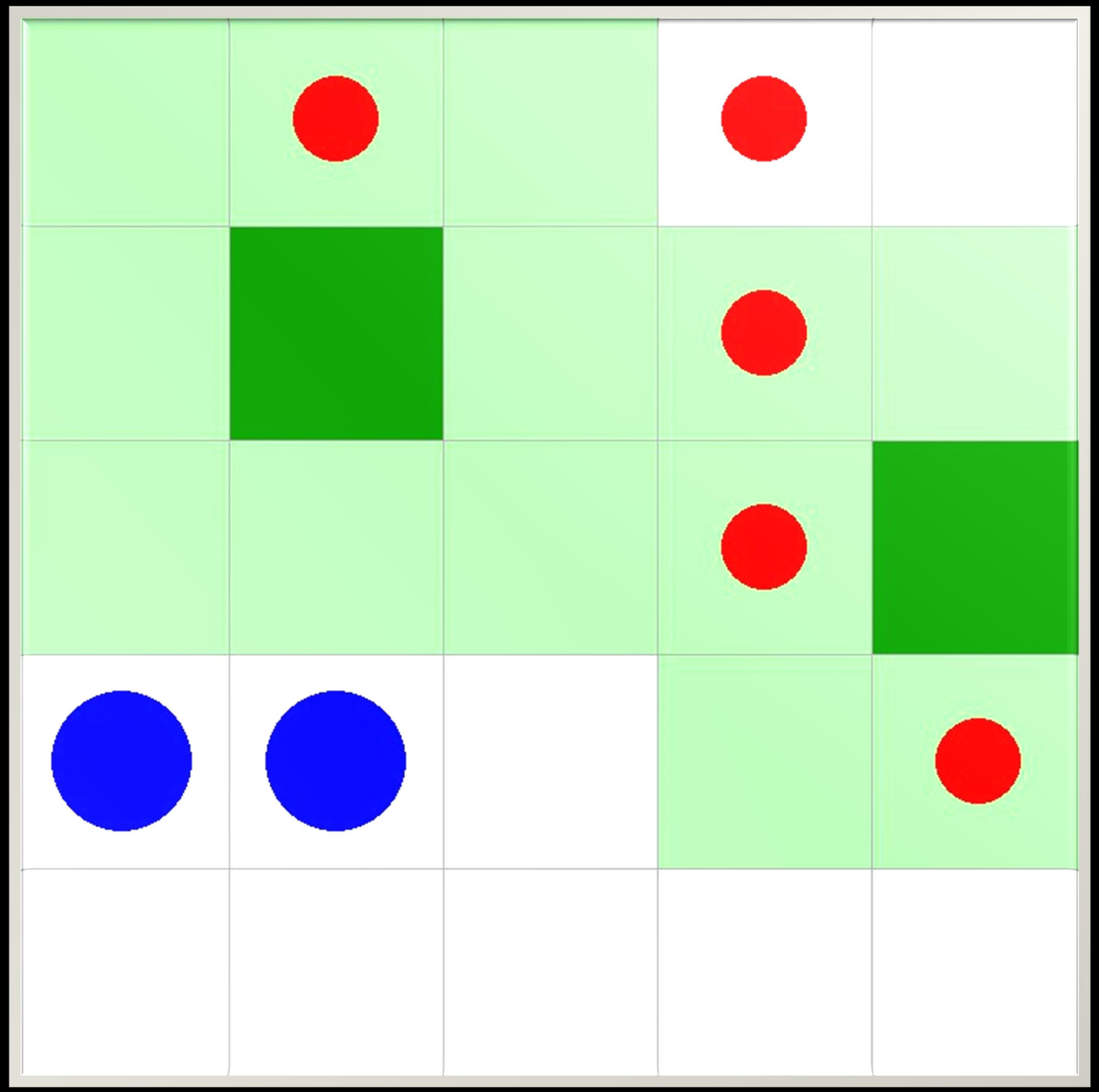


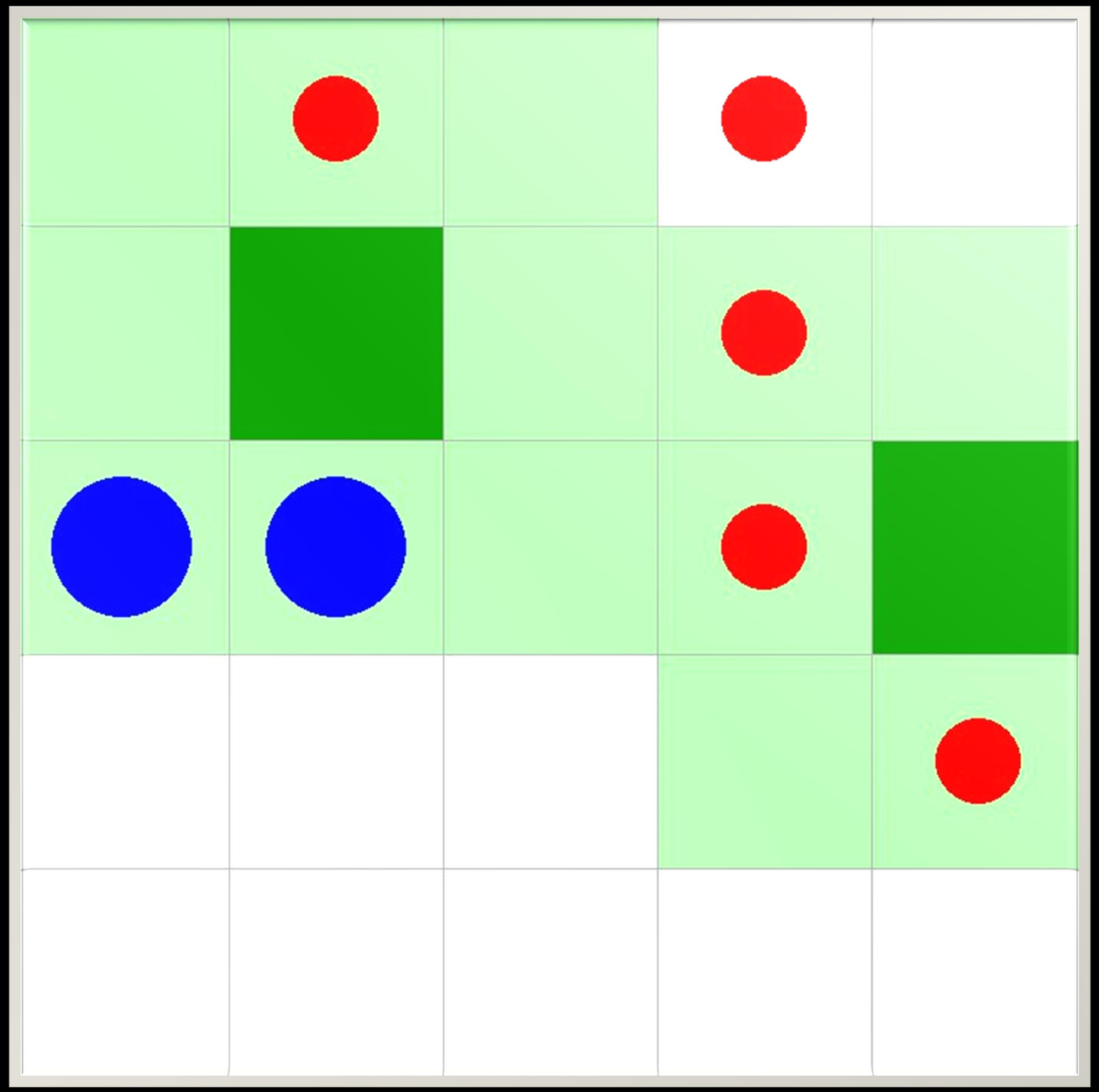


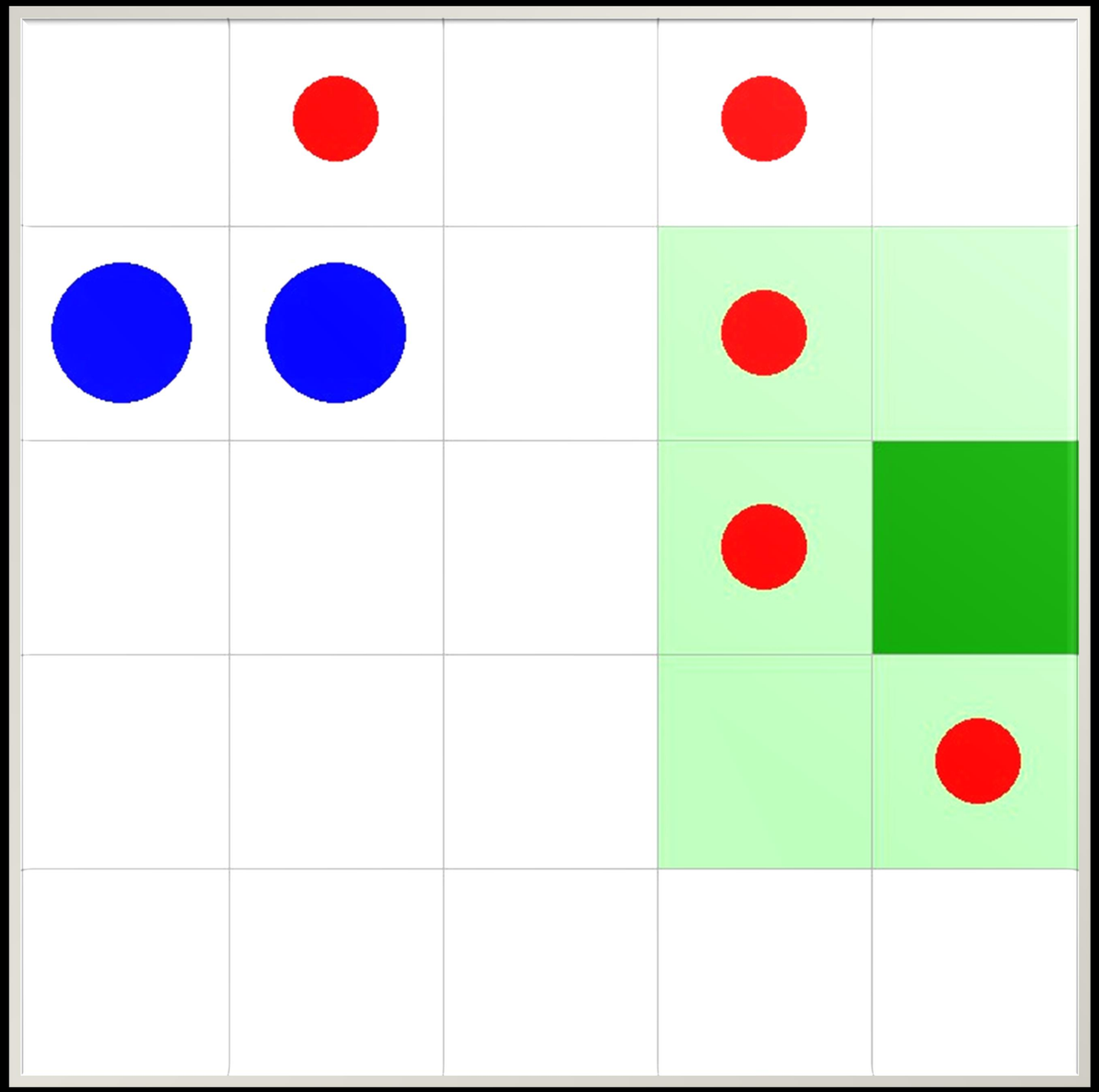


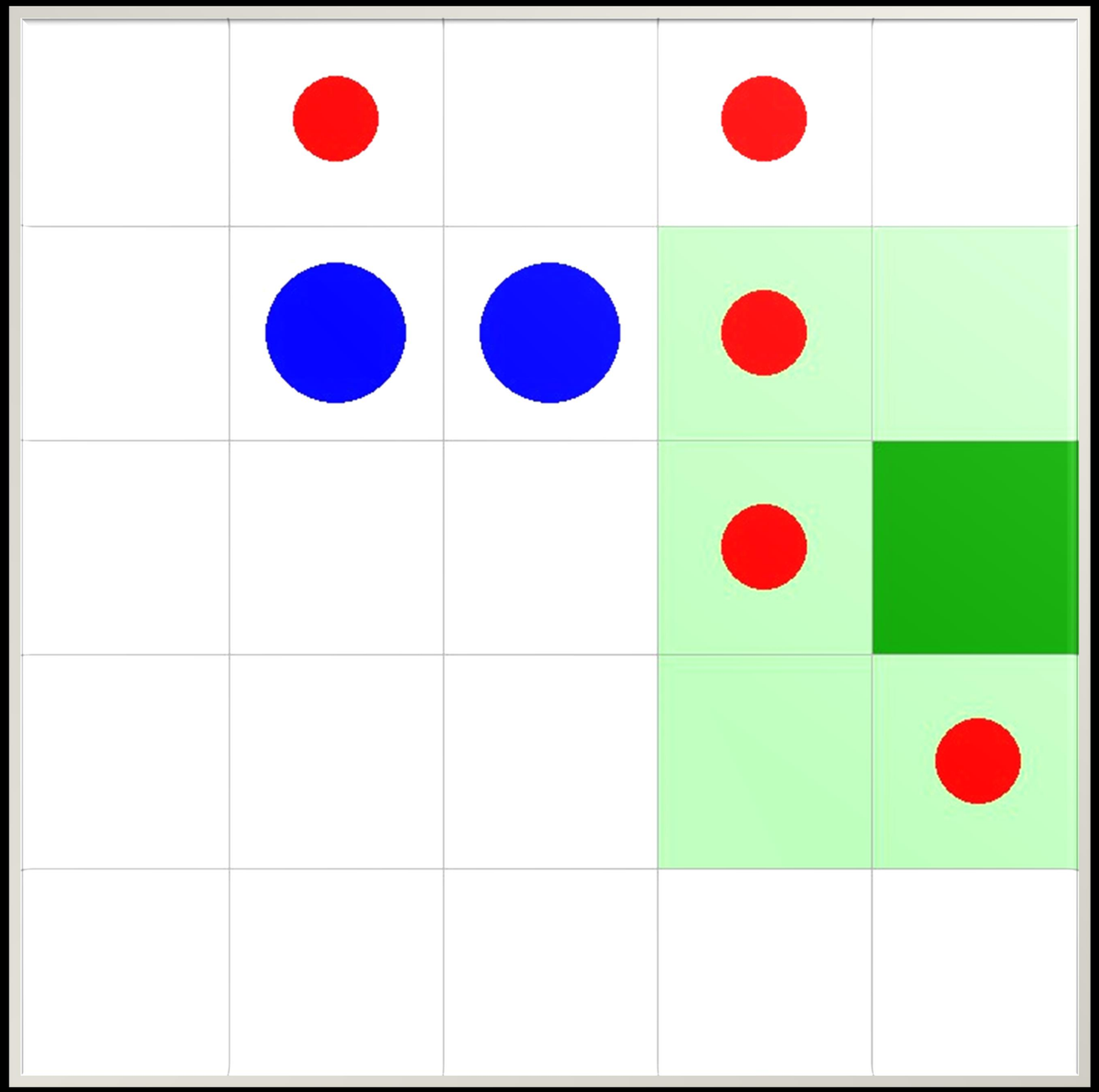


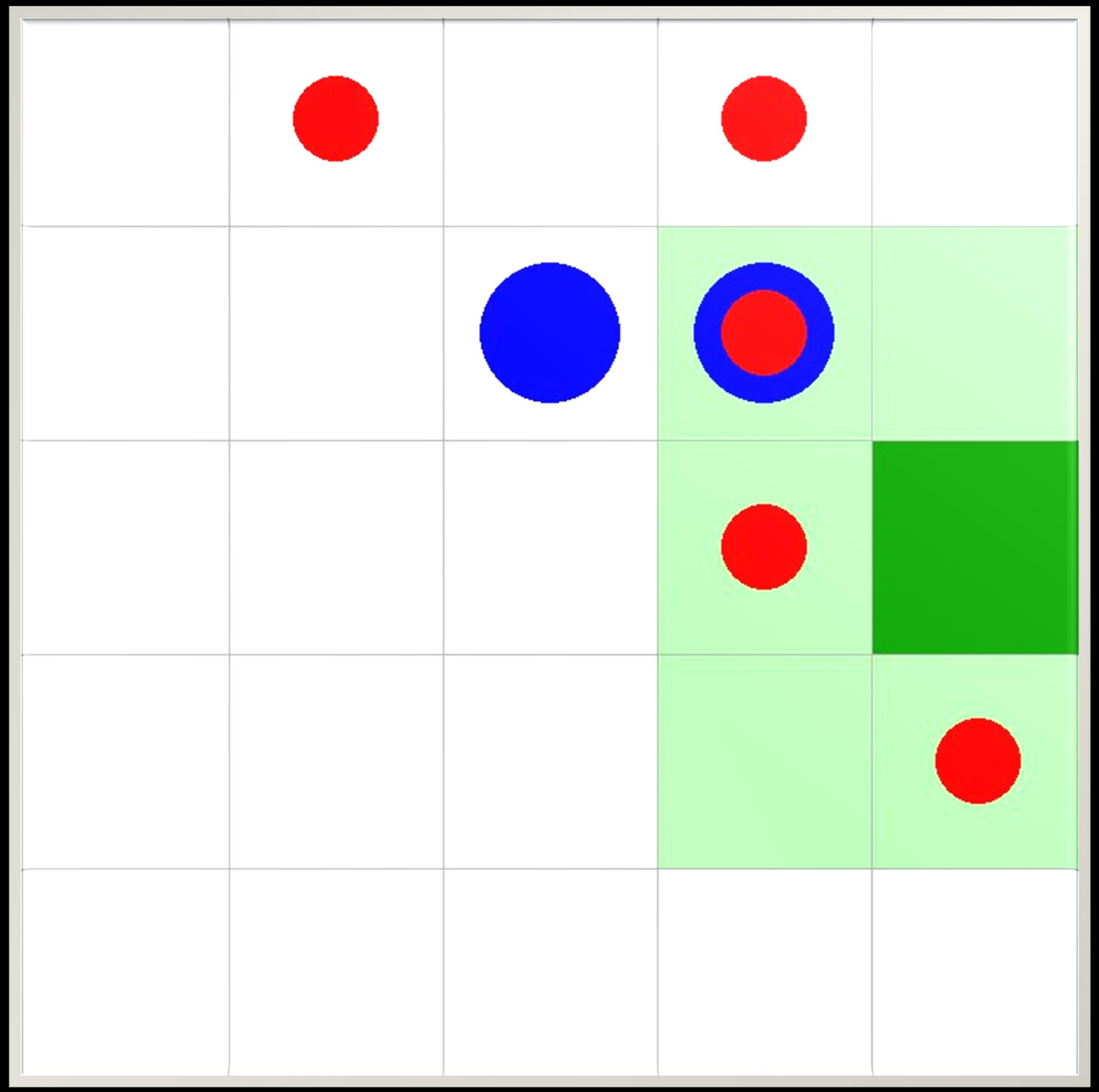


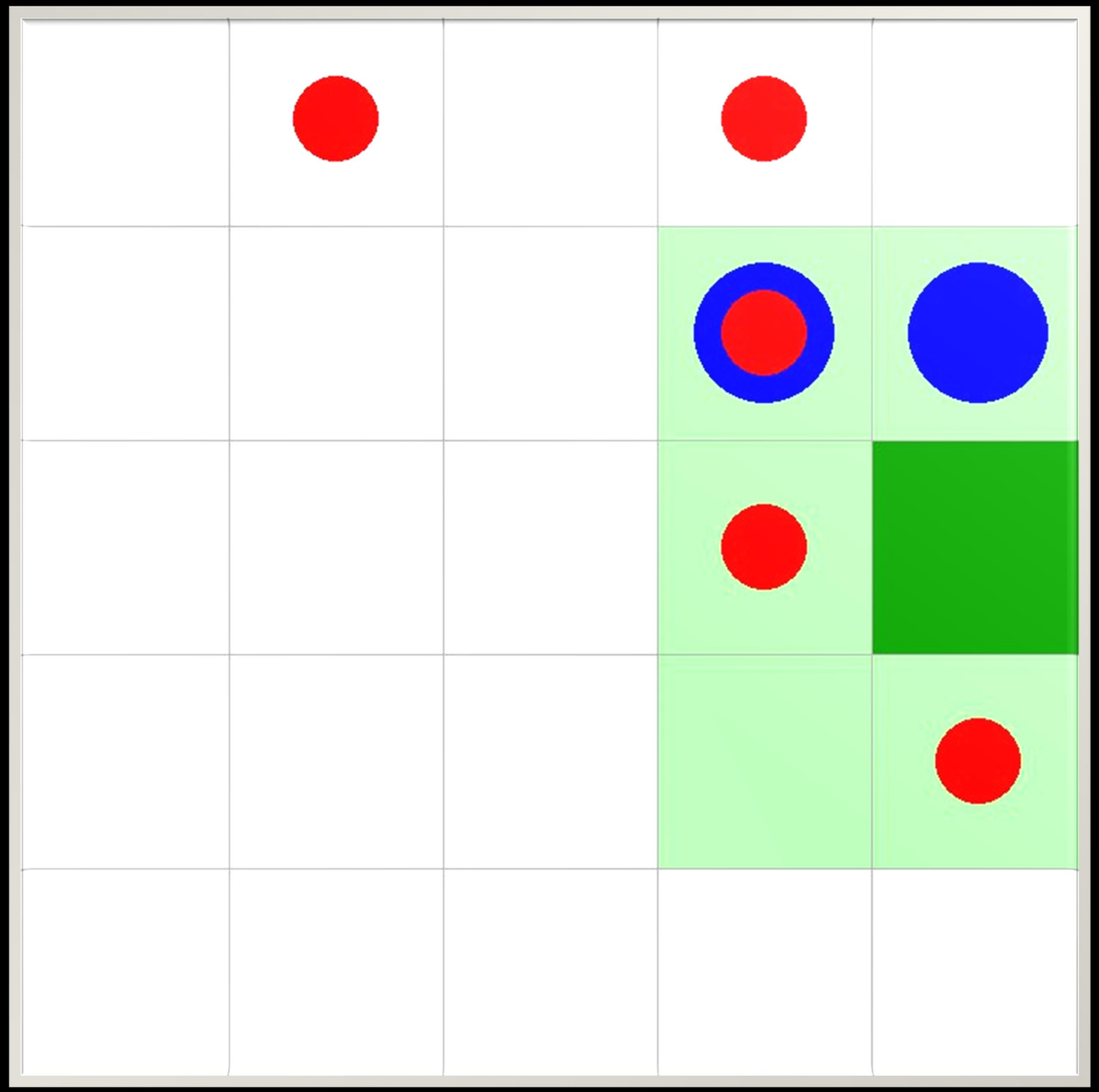


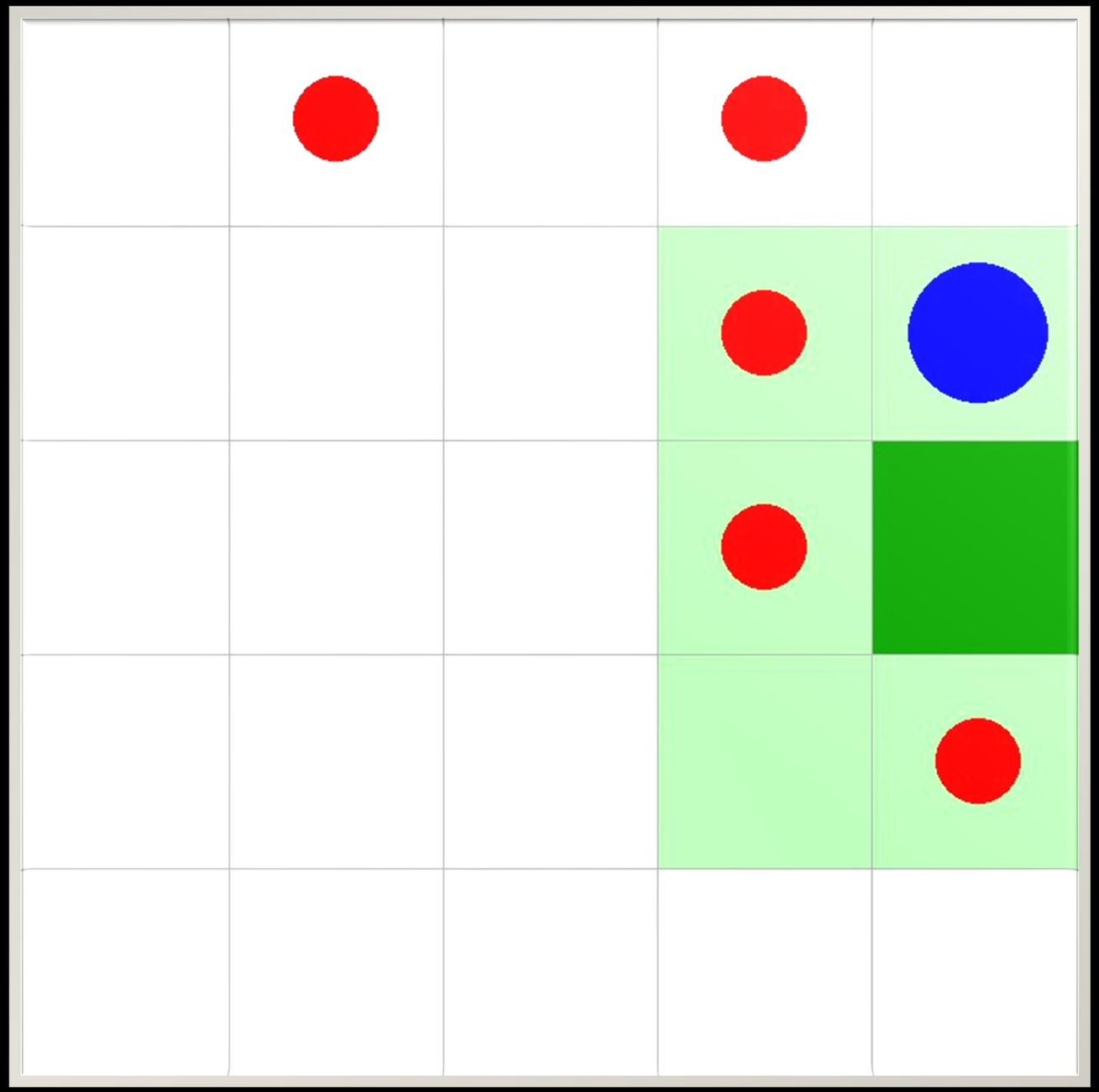


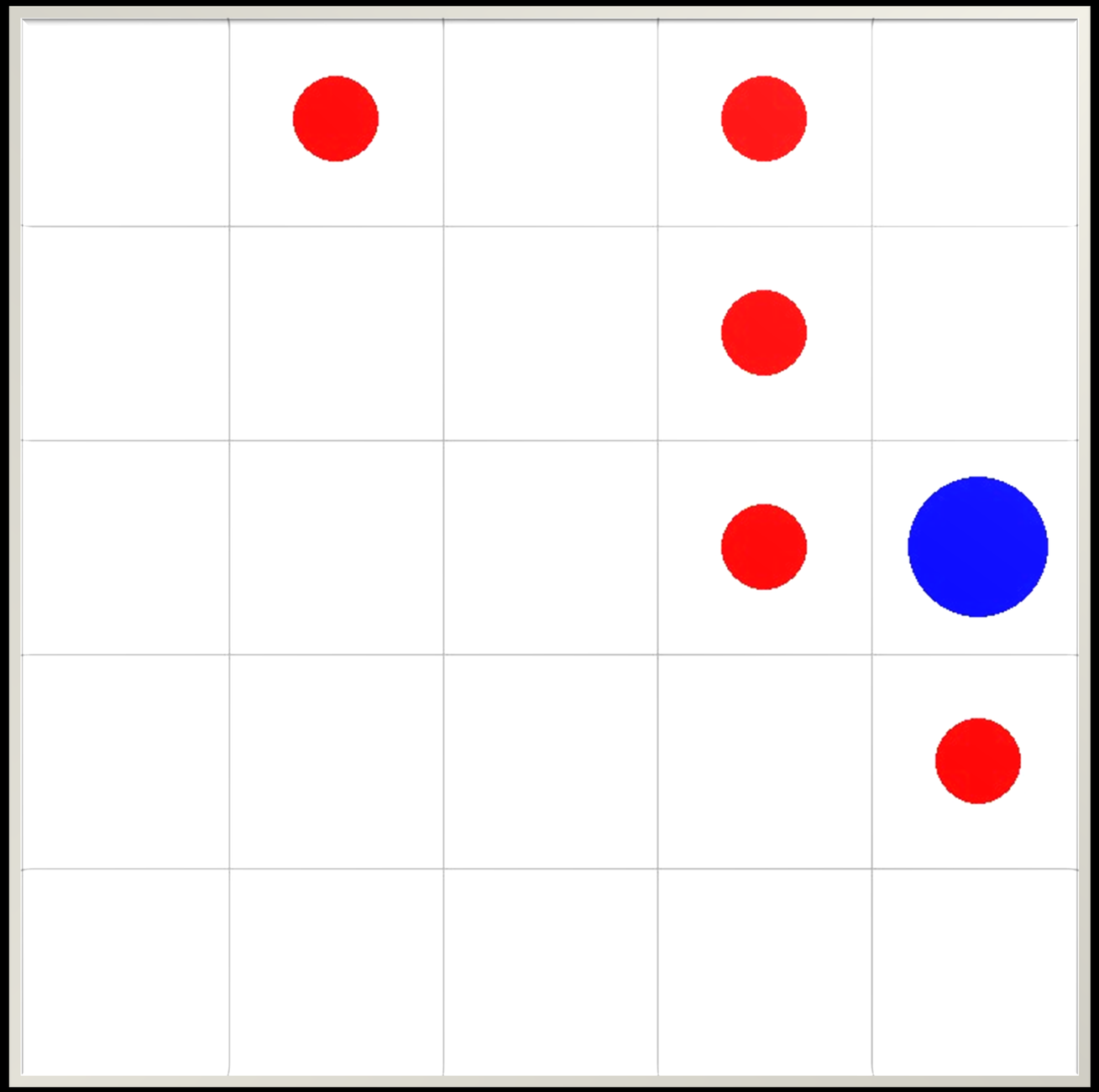












In a nutshell

- Resource extraction game
- MARL to SARL
- DQNs
- Normalization
- Lazy agents
- Exploits
- Beyond 5-by-5 grids?
- Distributed learning & execution

Questions?