



Hamilton Institute



The Curse of Hamilton's Chairs

Ahmed Shalaby

2nd year PhD

Supervisor: Damien Woods





Hamilton Institute



**Maynooth
University**
National University
of Ireland Maynooth

The Curse of Hamilton's Chairs

Thermodynamics of a multistranded one-dimension Scaffolded DNA Computer

Ahmed Shalaby

2nd year PhD

Supervisor: Damien Woods



European
Innovation
Council



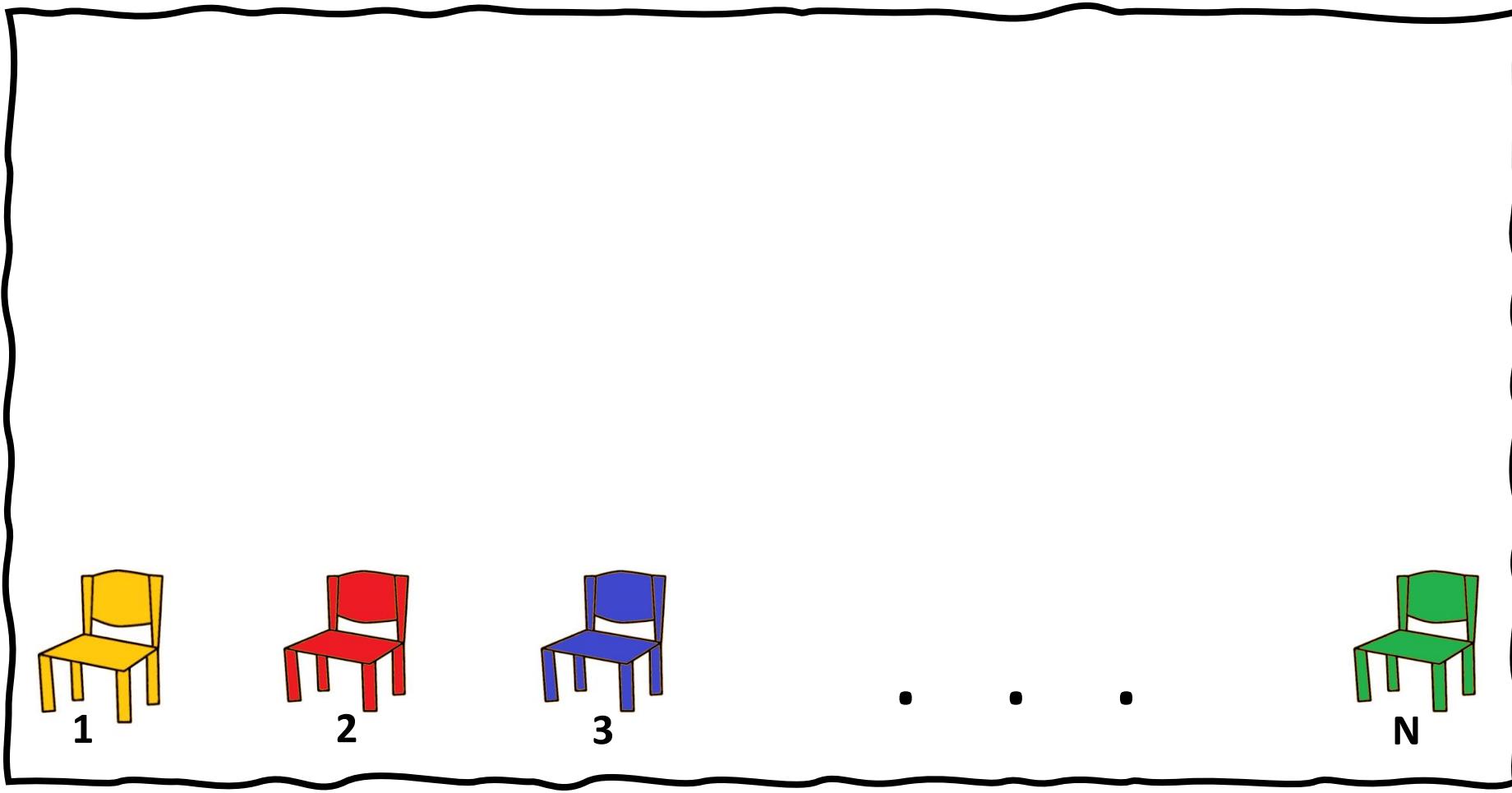
Funded by
the European Union



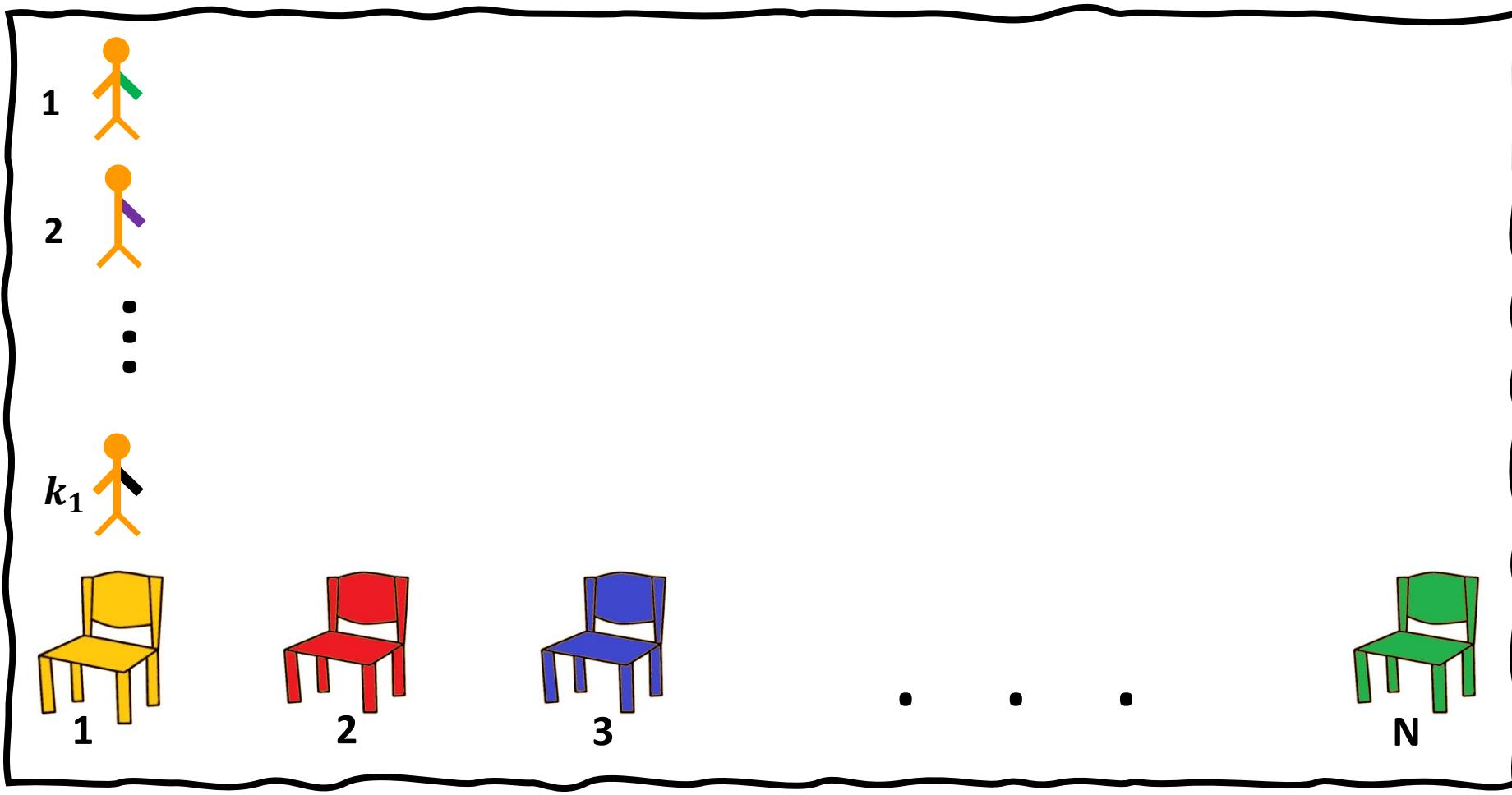


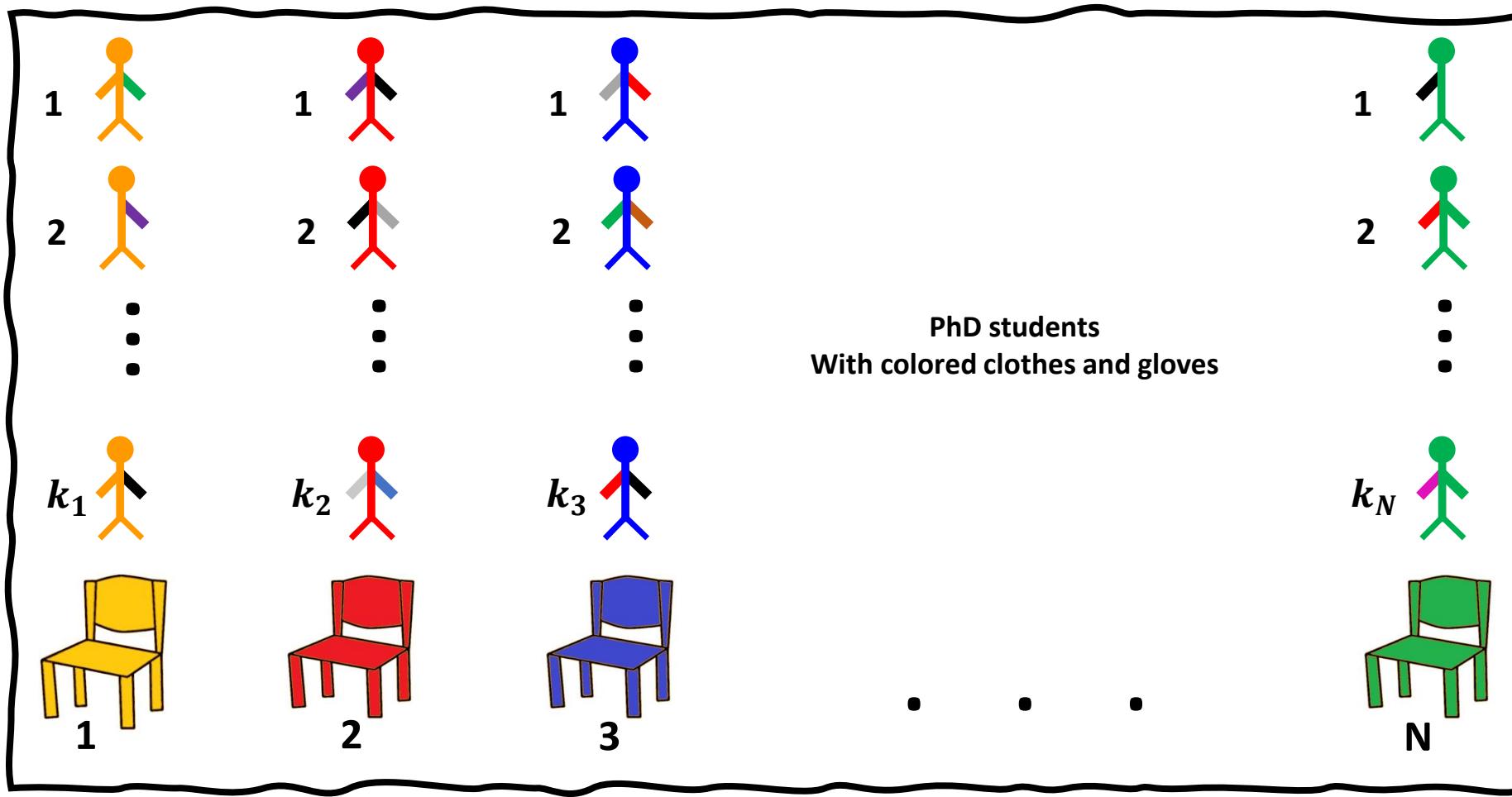


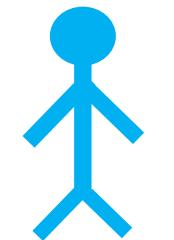
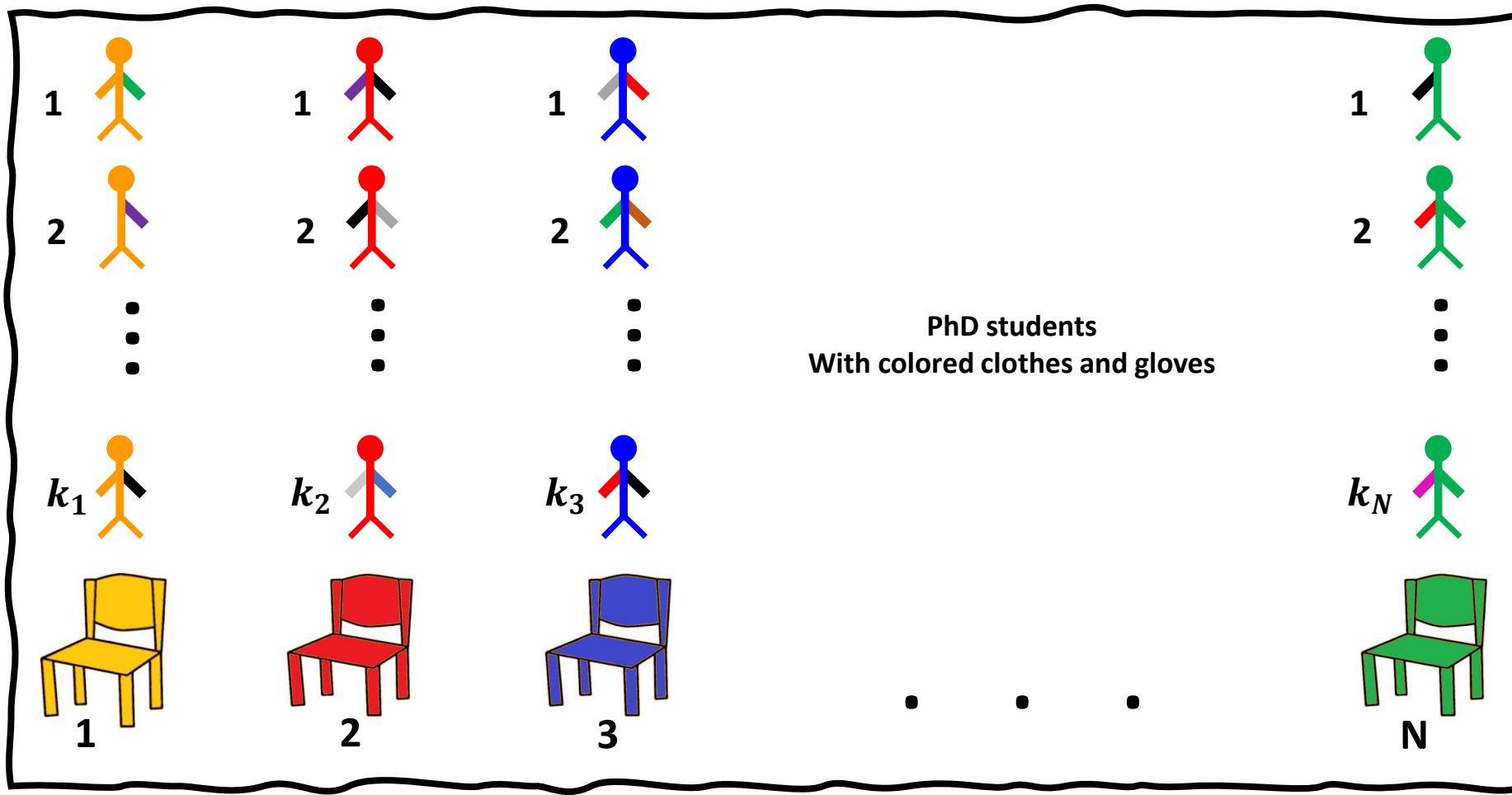
Let's discover the rules of the game



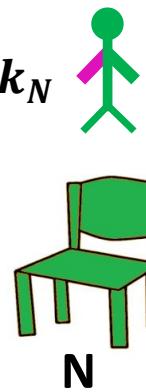
Hamilton



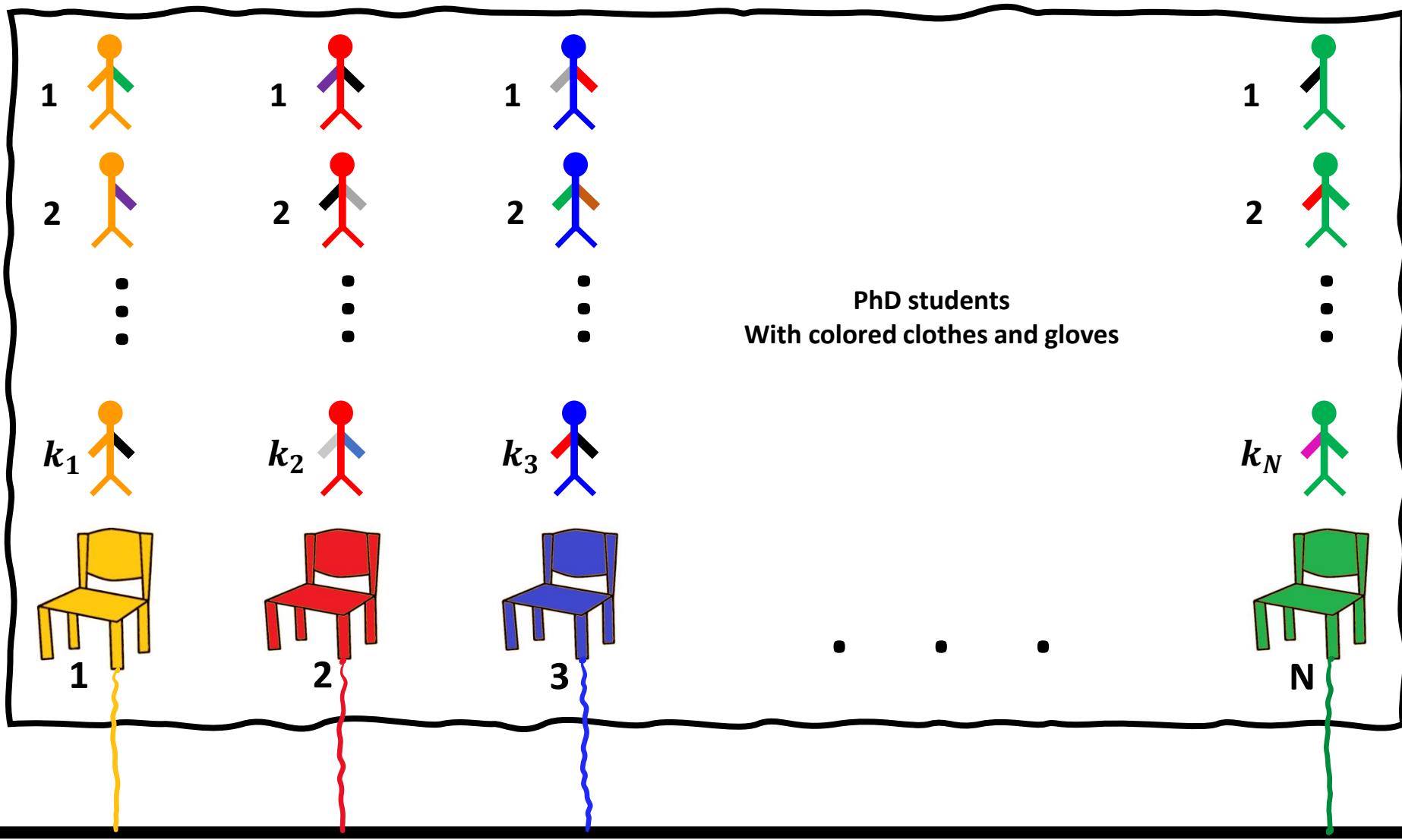




Damien

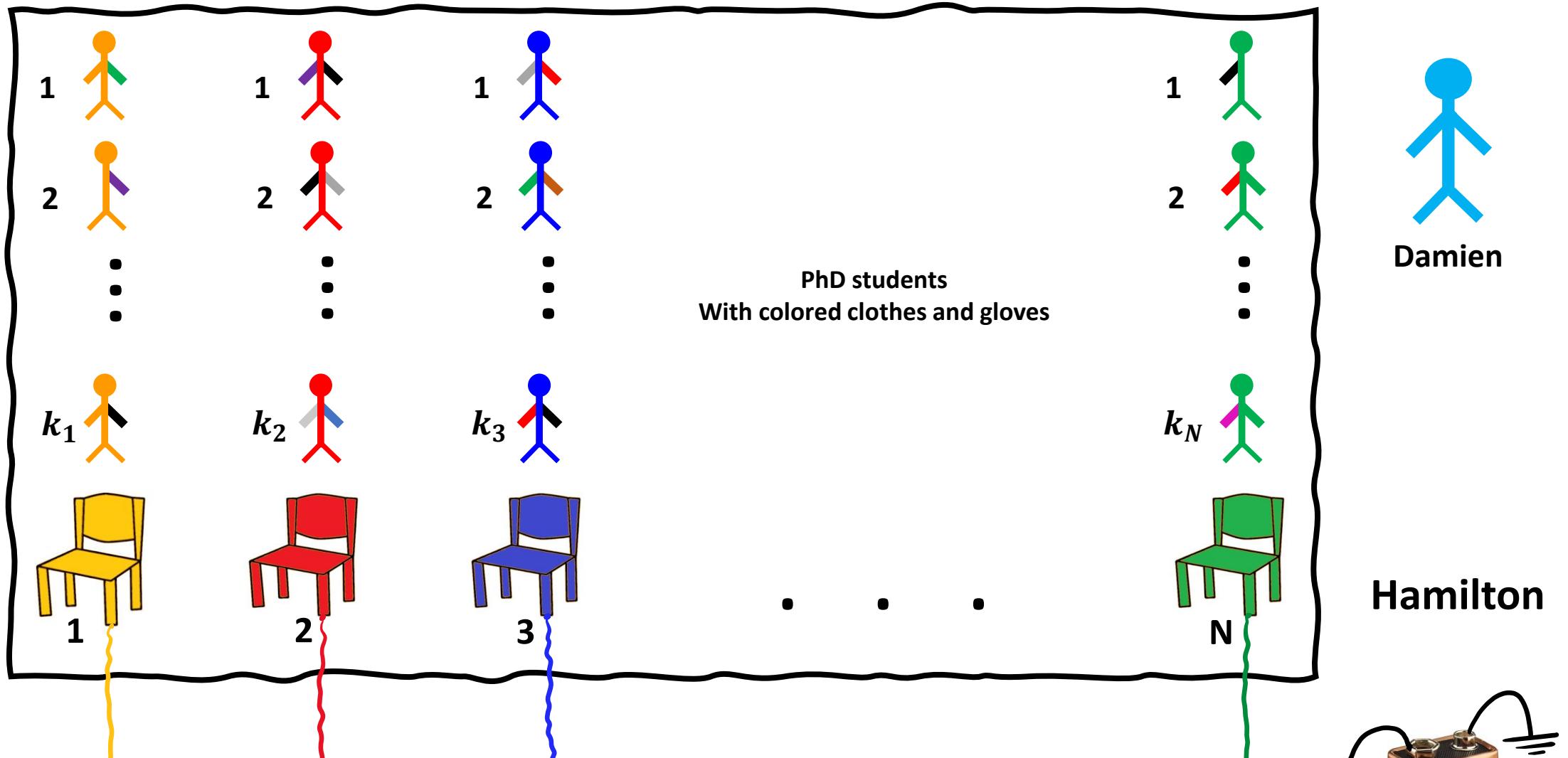


Hamilton

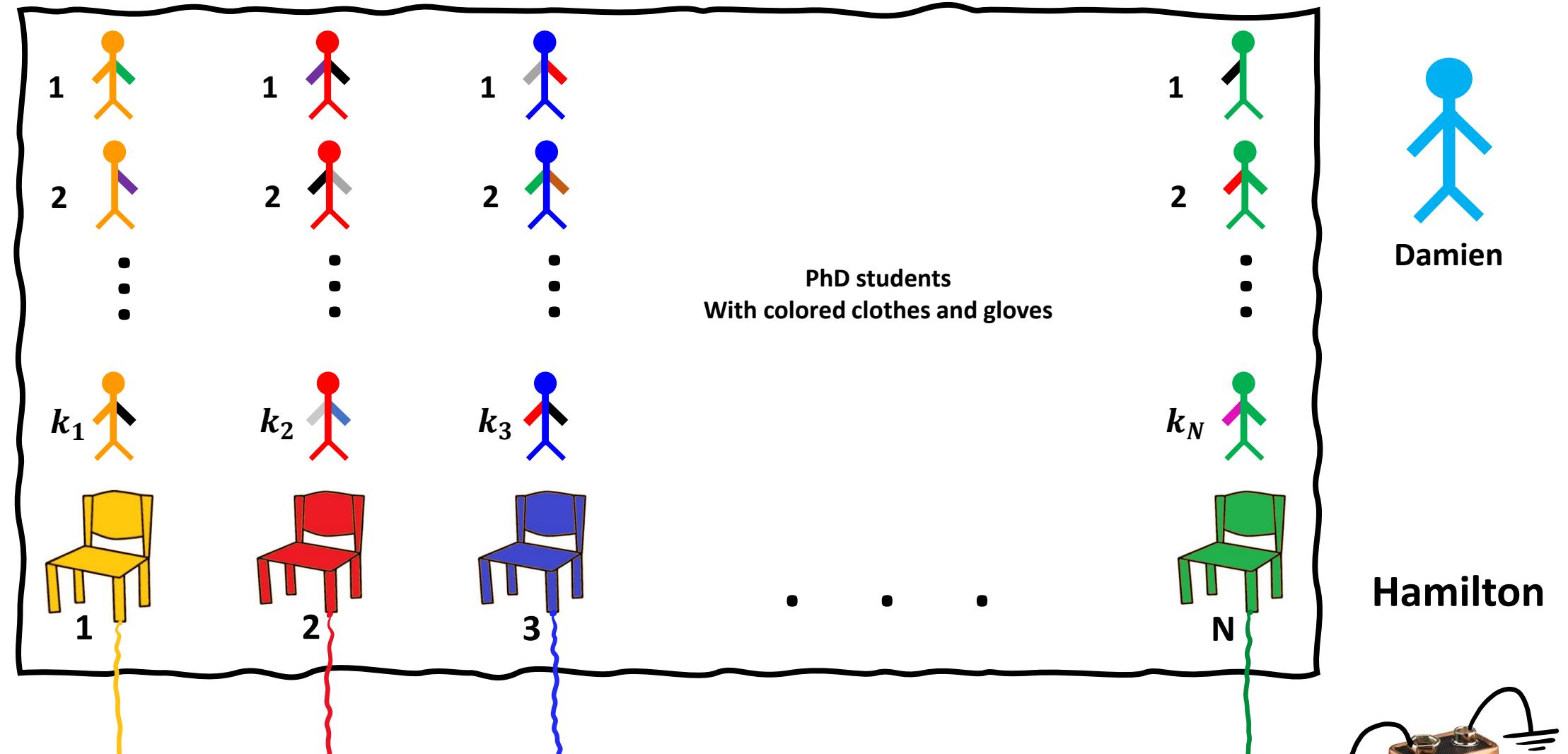


Damien

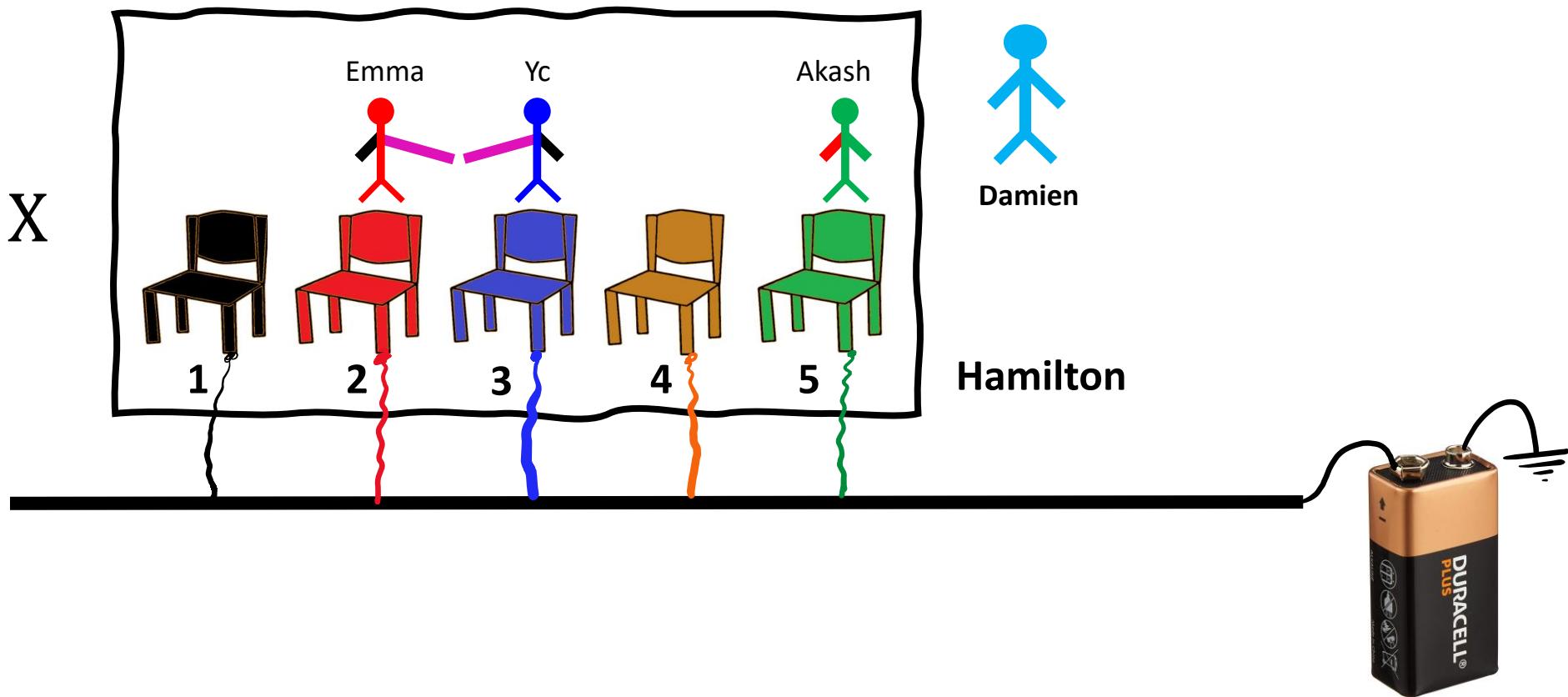
Hamilton

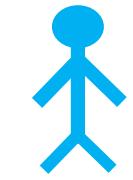
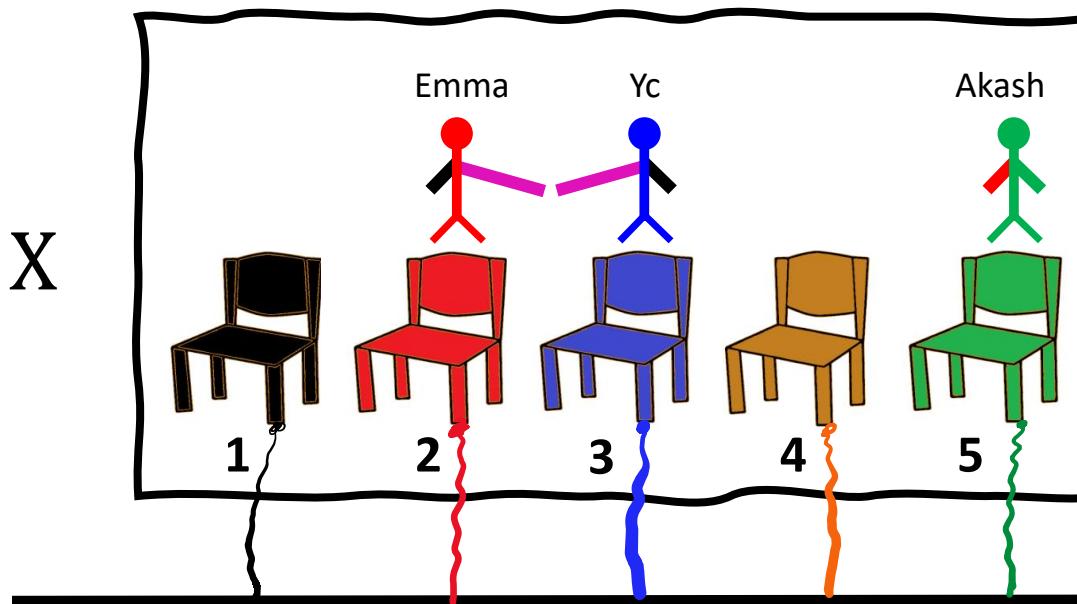


- How many different configurations we will have ?



- How many different configurations we will have ?





Damien

Hamilton



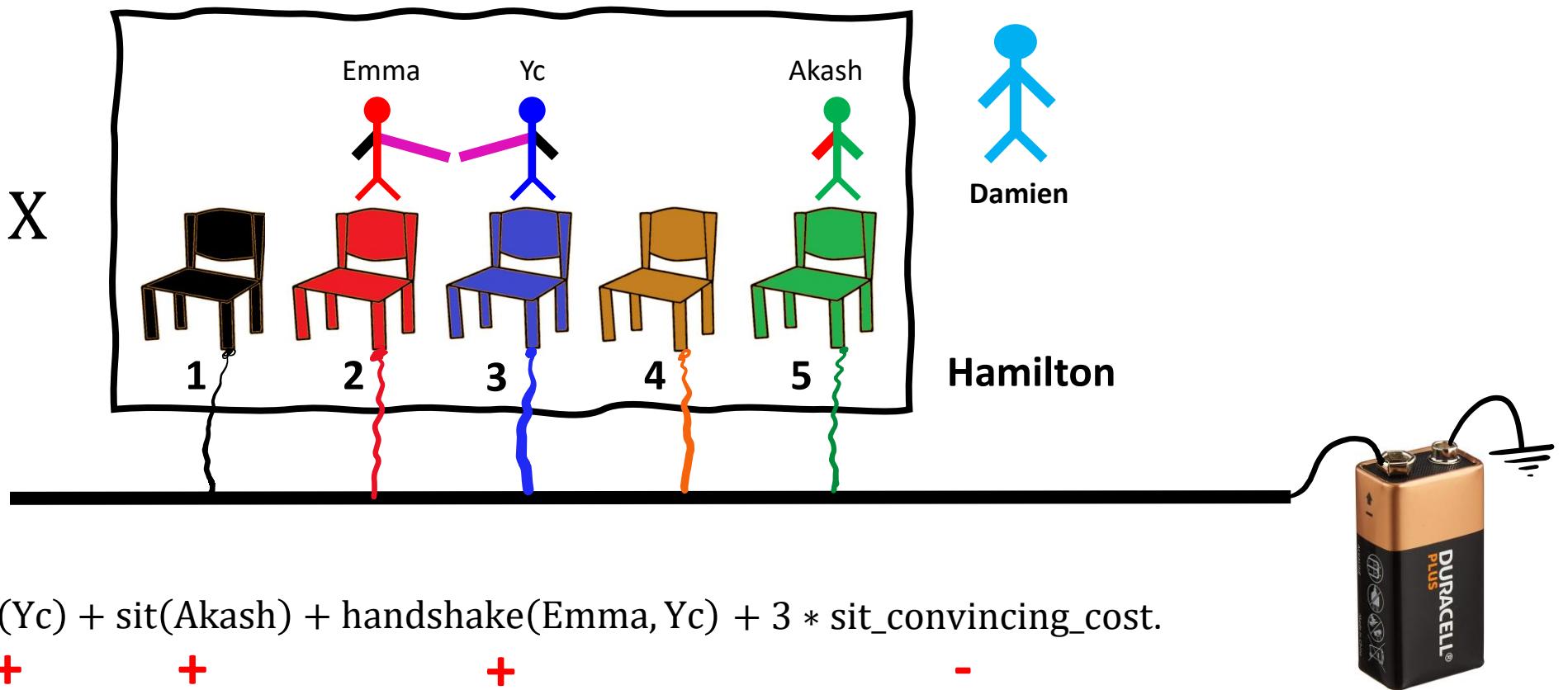
$$E(X) = \text{sit}(\text{Emma}) + \text{sit}(\text{Yc}) + \text{sit}(\text{Akash}) + \text{handshake}(\text{Emma}, \text{Yc})$$

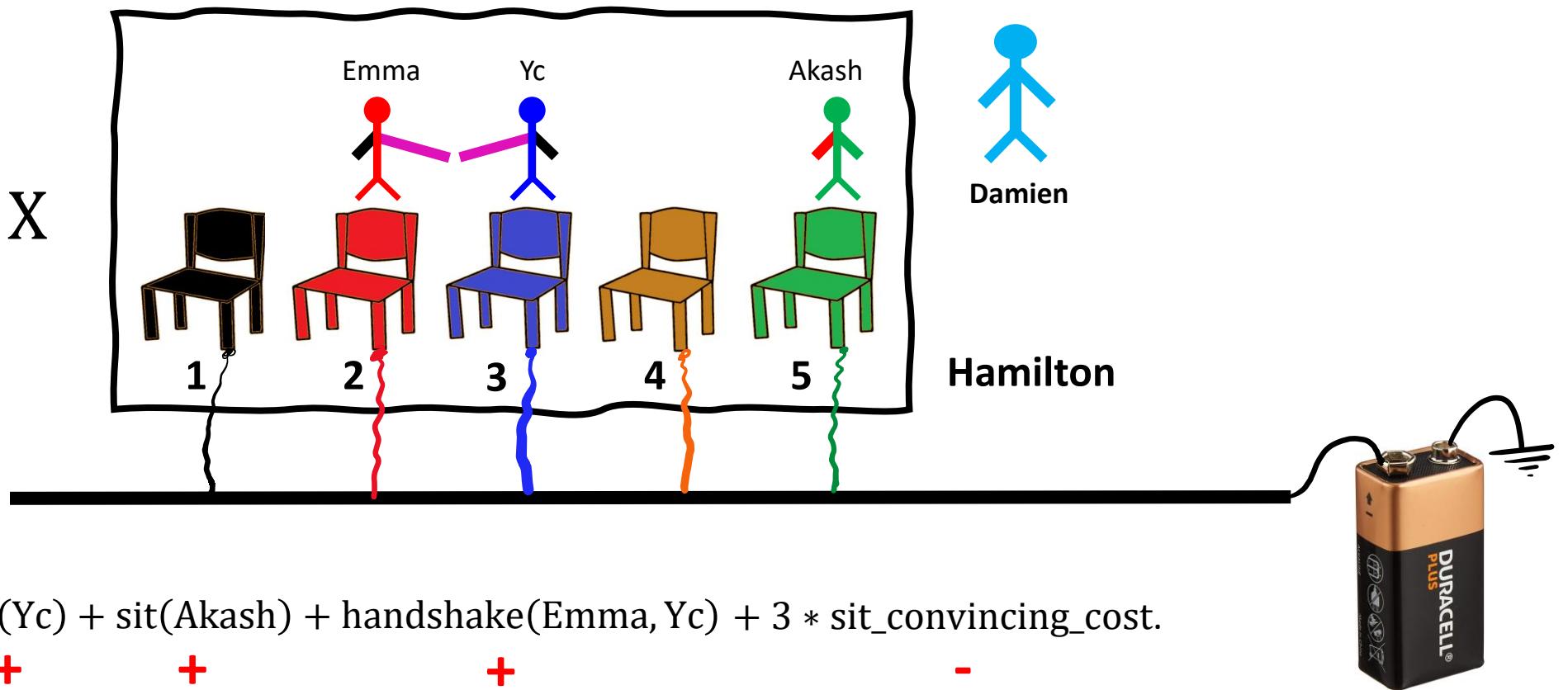
+

+

+

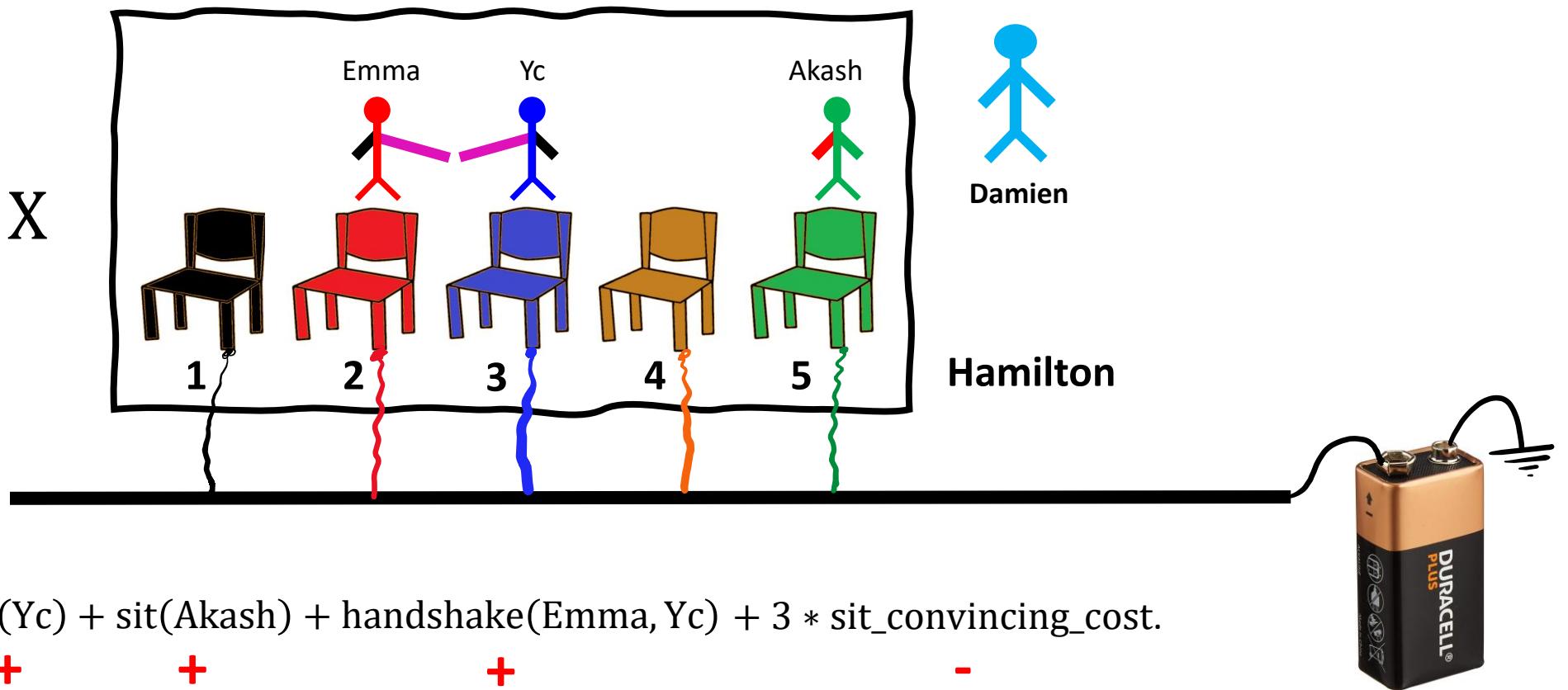
+





We further assume the following:

- $|\text{sit}(p)| > |\text{sit_convincing_cost}|$. (Damien always gains by convincing a PhD student to sit)



$$E(X) = \text{sit}(\text{Emma}) + \text{sit}(\text{Yc}) + \text{sit}(\text{Akash}) + \text{handshake}(\text{Emma}, \text{Yc}) + 3 * \text{sit_convincing_cost}.$$

+ + + + -

$$E(X) = \sum_{p \in X} \text{sit}(p) + \sum_{p_i, p_{i+1} \in X} \text{handshake}(p_i, p_{i+1}) + l * \text{sit_convincing_cost}.$$

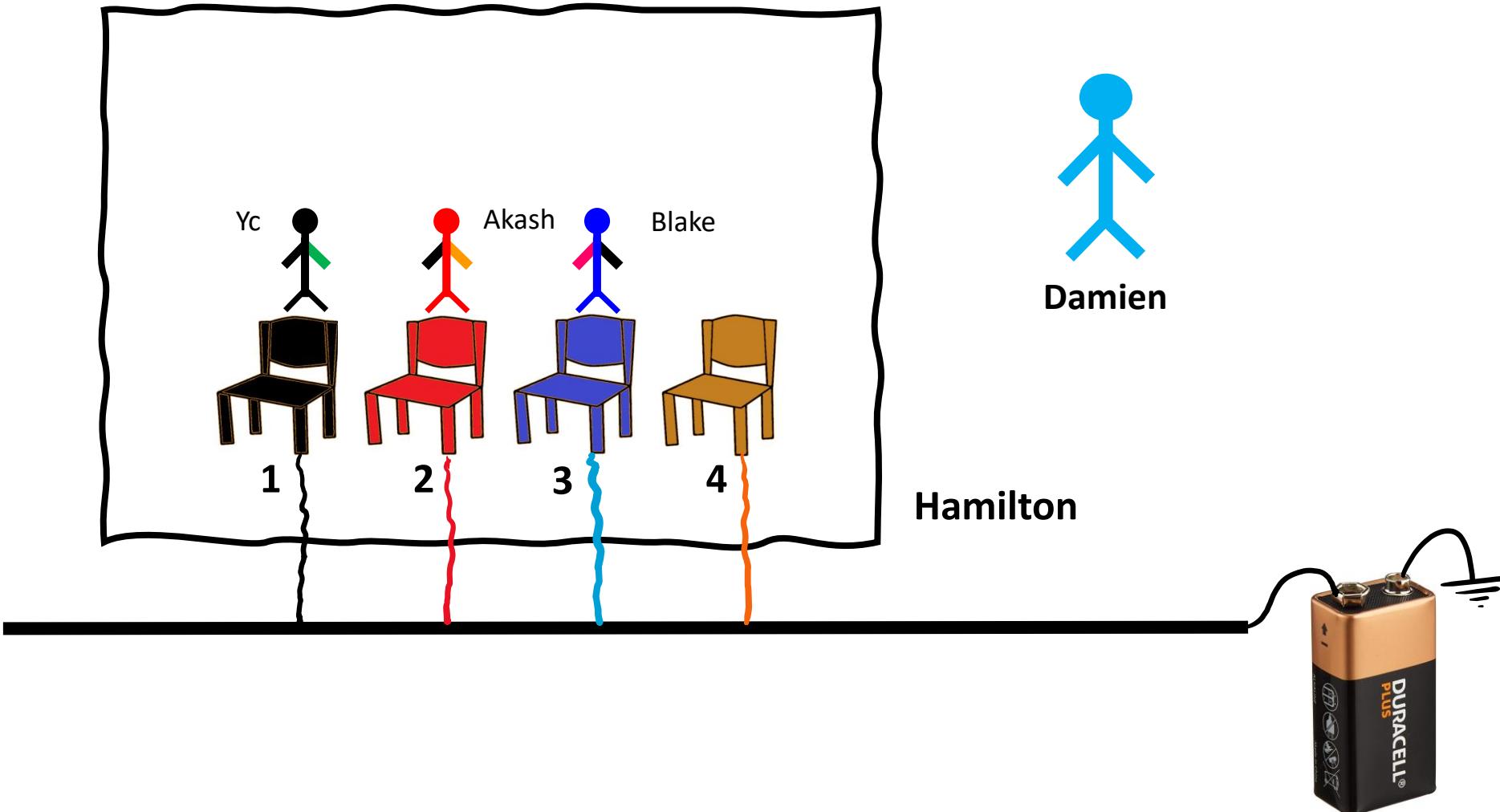
configuration X of size l PhD students

We further assume the following:

- $|\text{sit}(p)| > |\text{sit_convincing_cost}|$. (Damien always gains by convincing a PhD student to sit)

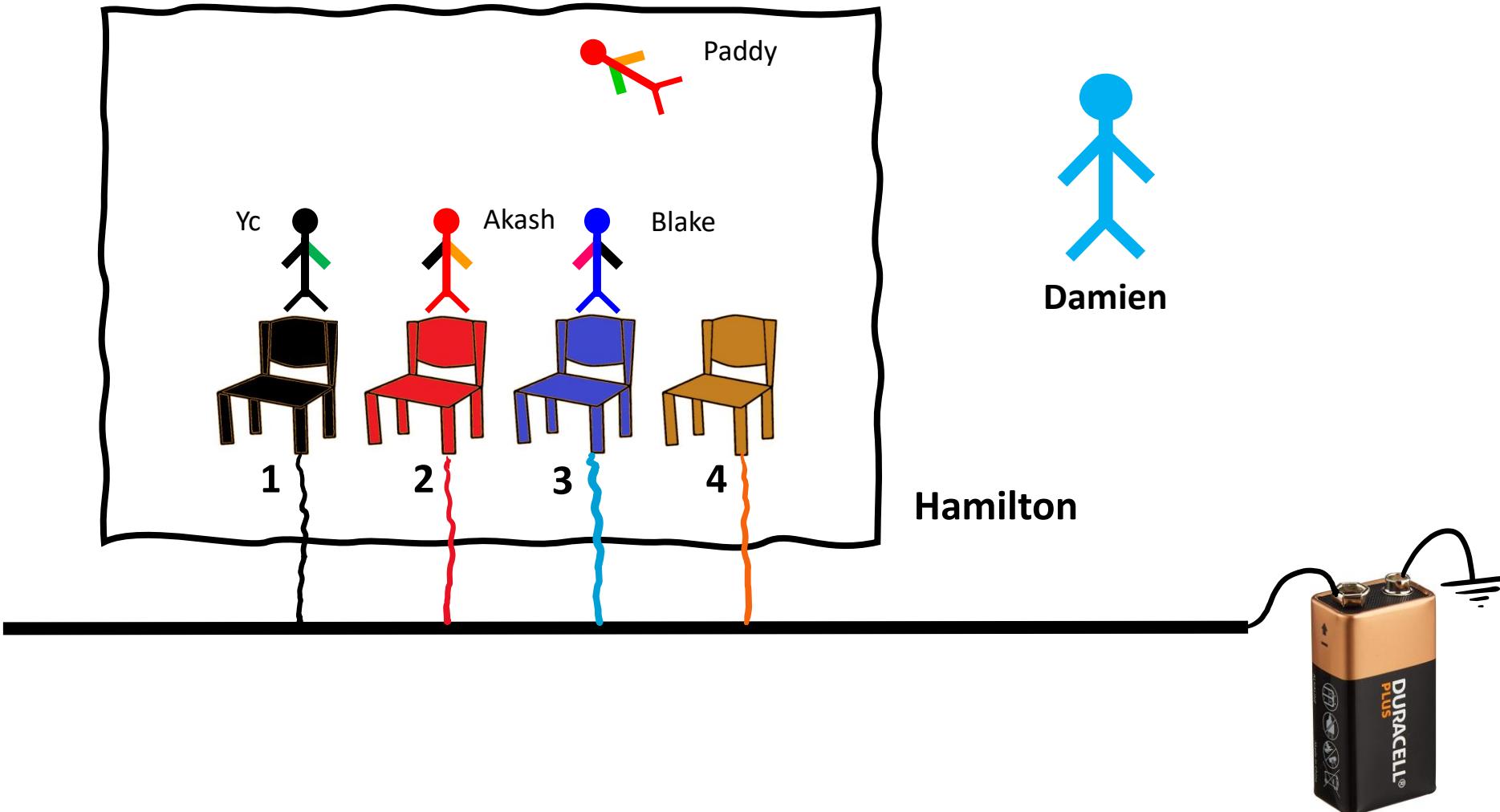
Built-in self improvement mechanism

PhD students' displacement system



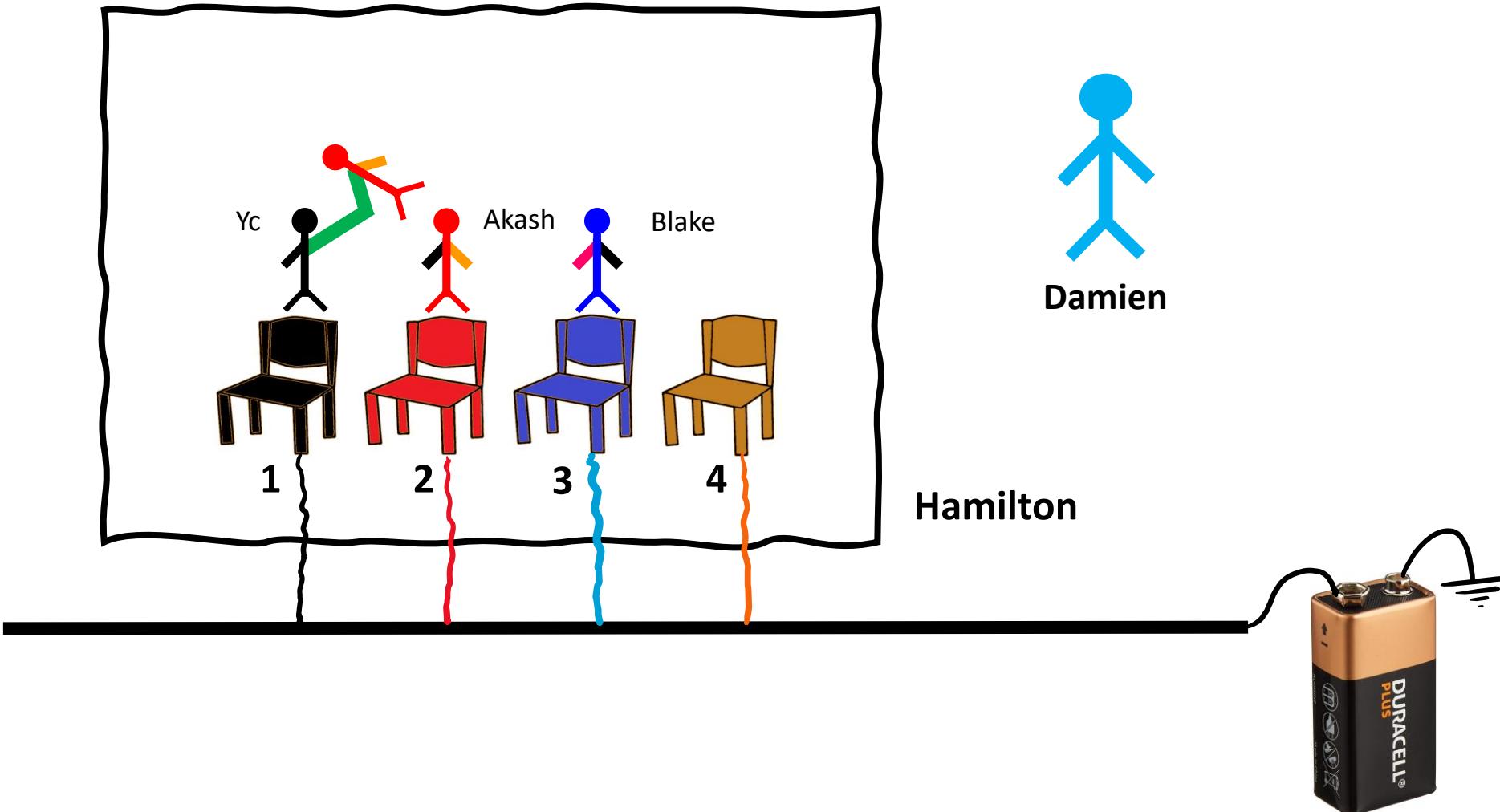
Built-in self improvement mechanism

PhD students' displacement system



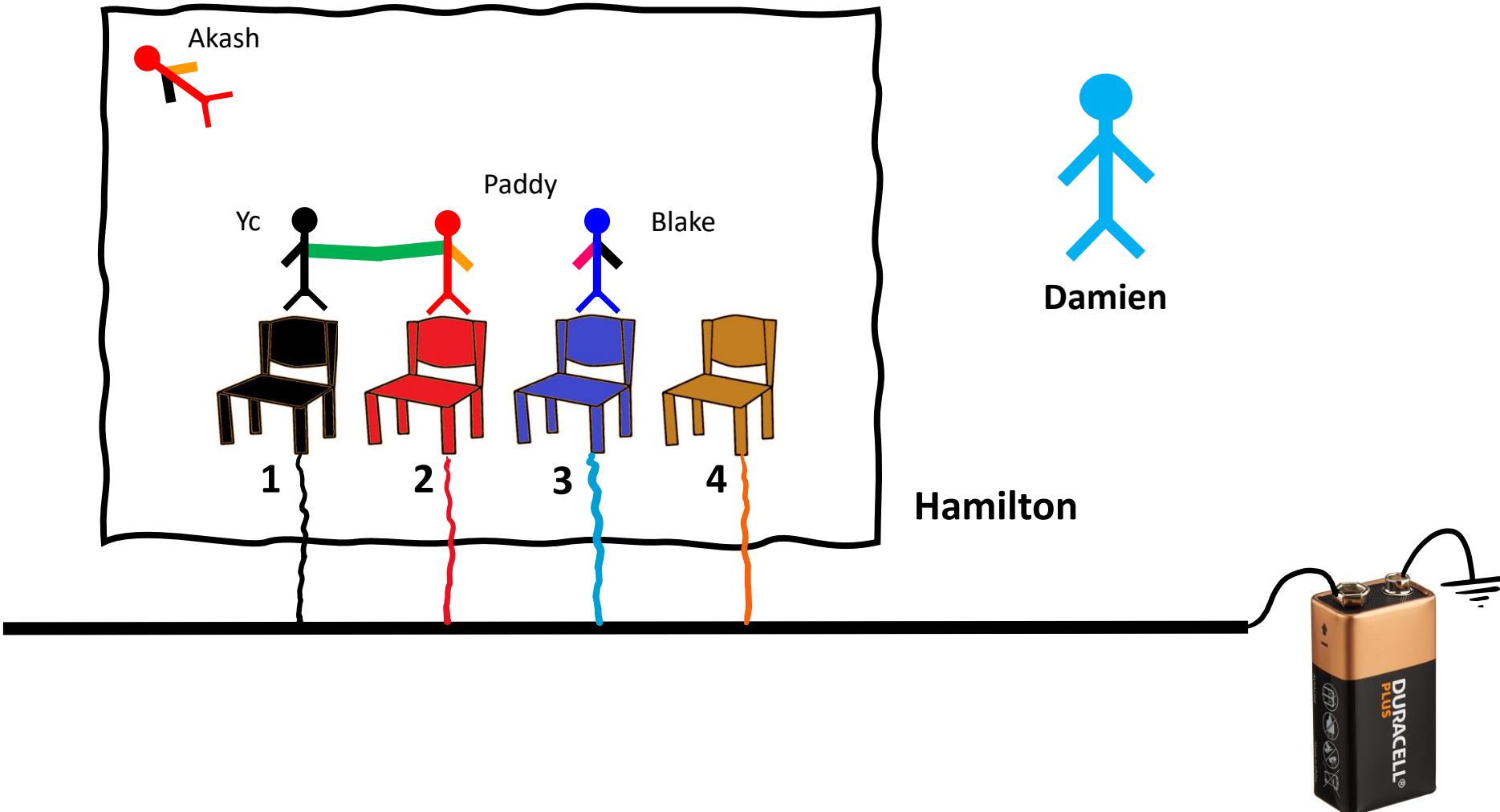
Built-in self improvement mechanism

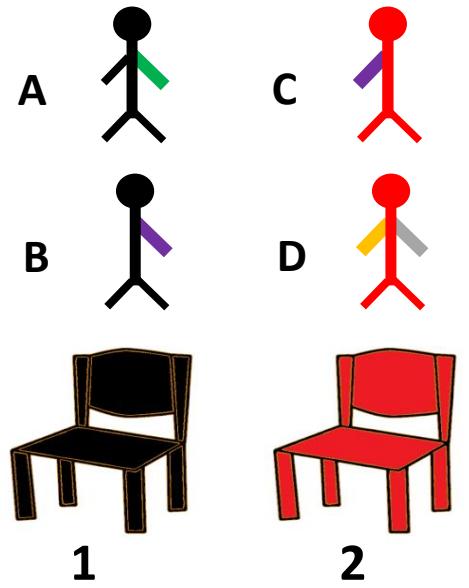
PhD students' displacement system



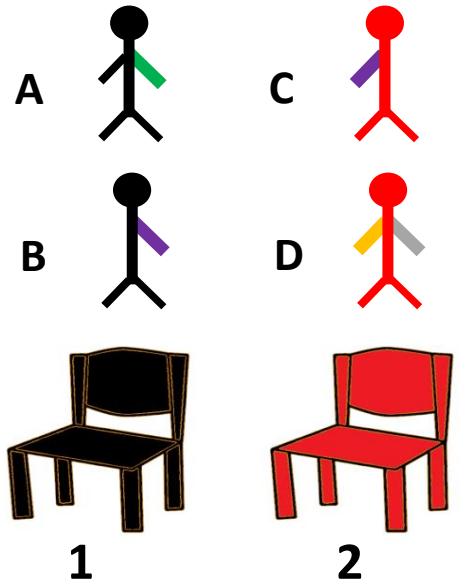
Built-in self improvement mechanism

PhD students' displacement system

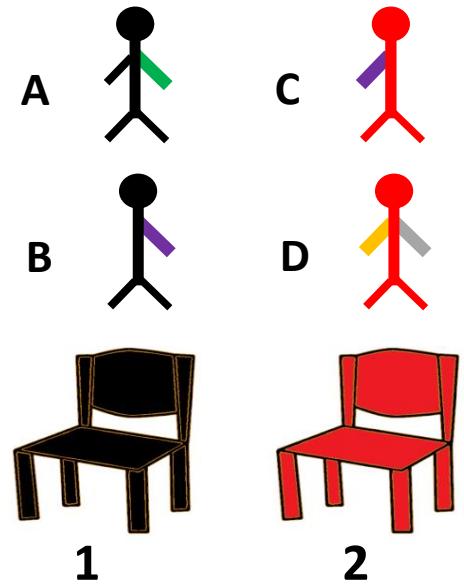




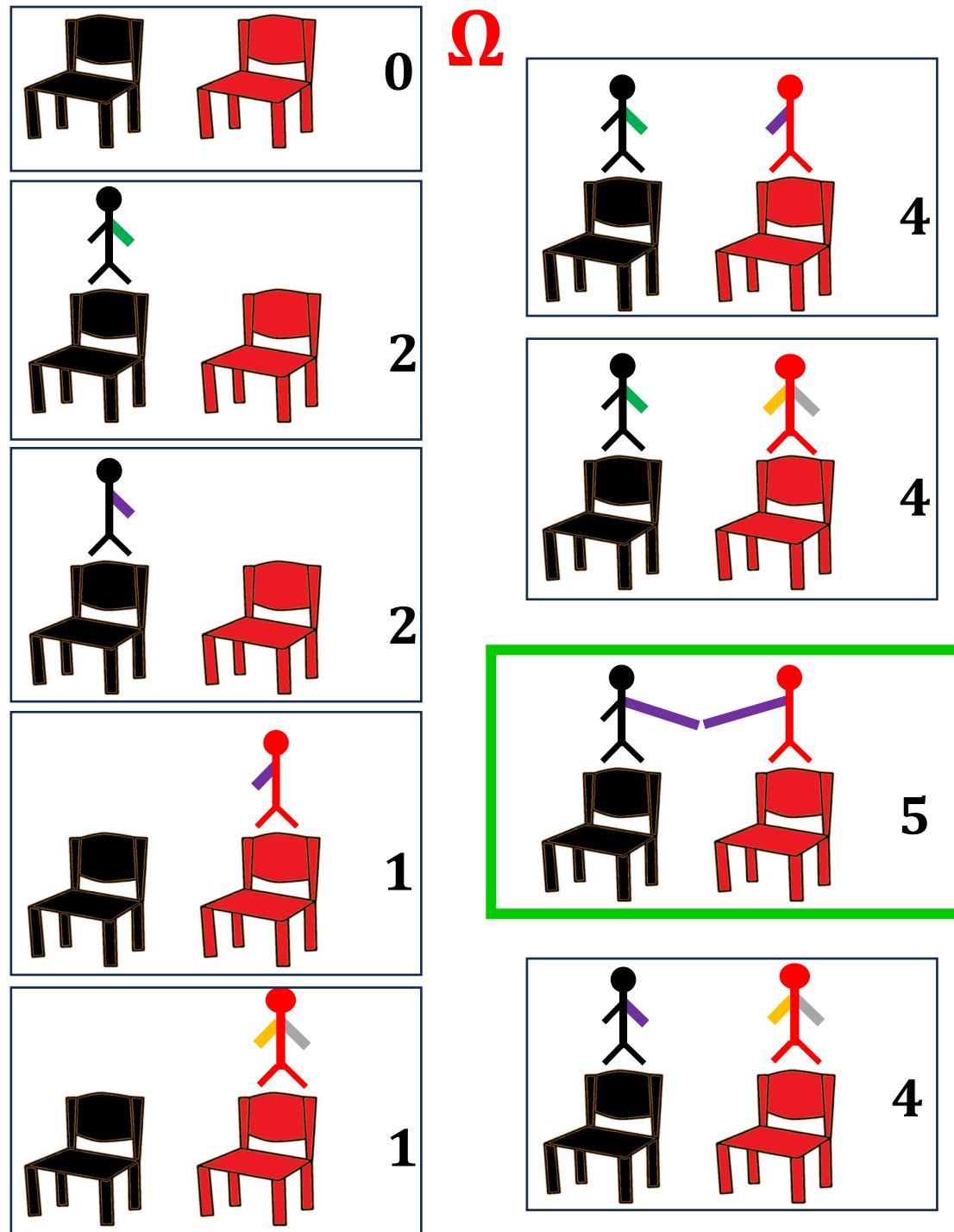
- $\text{sit}(A) = \text{sit}(B) = 3$
- $\text{sit}(C) = \text{sit}(D) = 3$
- $\text{handshake}(B, C) = 1$
- $\text{sit_convincing_cost} = -1$

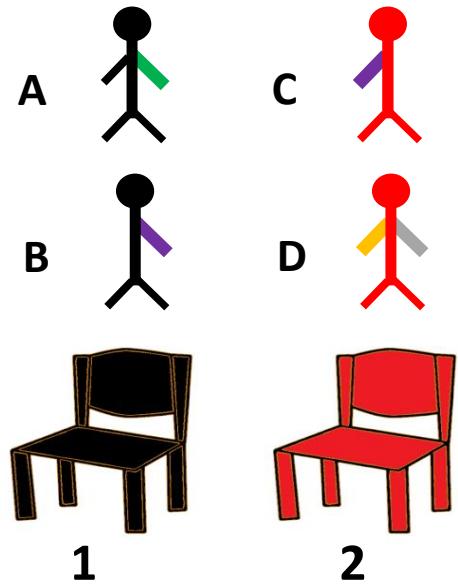
Ω 

- $\text{sit}(A) = \text{sit}(B) = 3$
- $\text{sit}(C) = \text{sit}(D) = 3$
- $\text{handshake}(B, C) = 1$
- $\text{sit_convincing_cost} = -1$

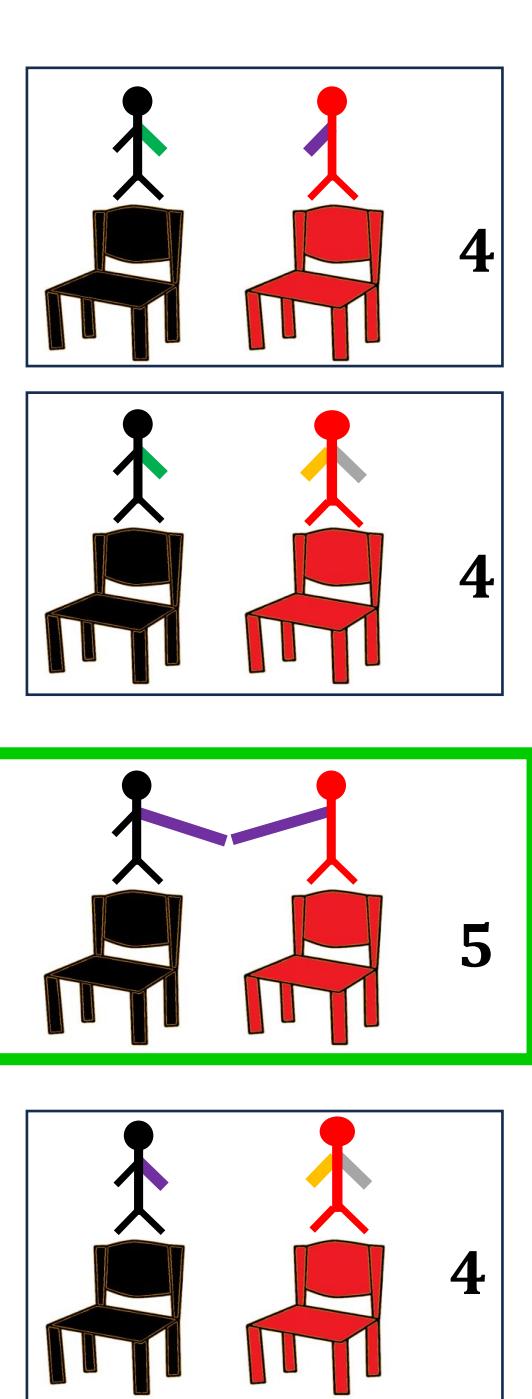
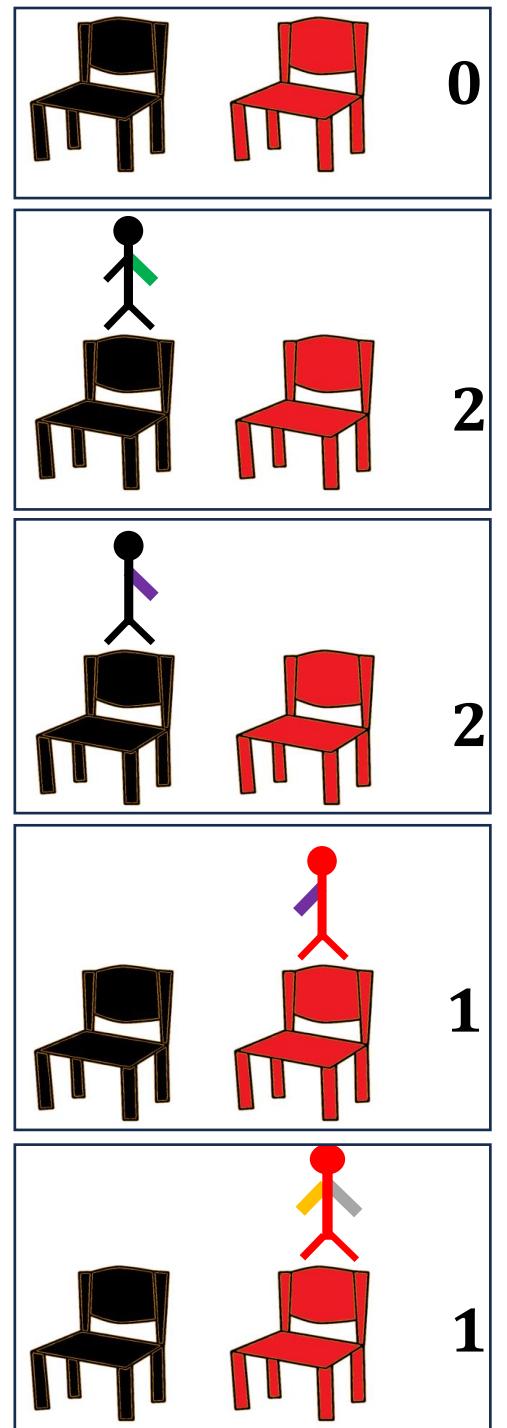


- $\text{sit}(A) = \text{sit}(B) = 3$
- $\text{sit}(C) = \text{sit}(D) = 3$
- $\text{handshake}(B, C) = 1$
- $\text{sit_convincing_cost} = -1$



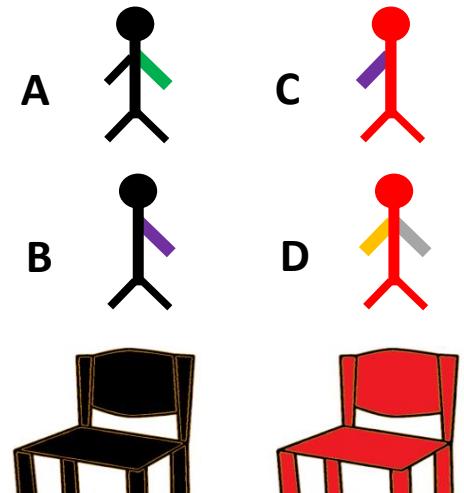


- $\text{sit}(A) = \text{sit}(B) = 3$
- $\text{sit}(C) = \text{sit}(D) = 3$
- $\text{handshake}(B, C) = 1$
- $\text{sit_convincing_cost} = -1$

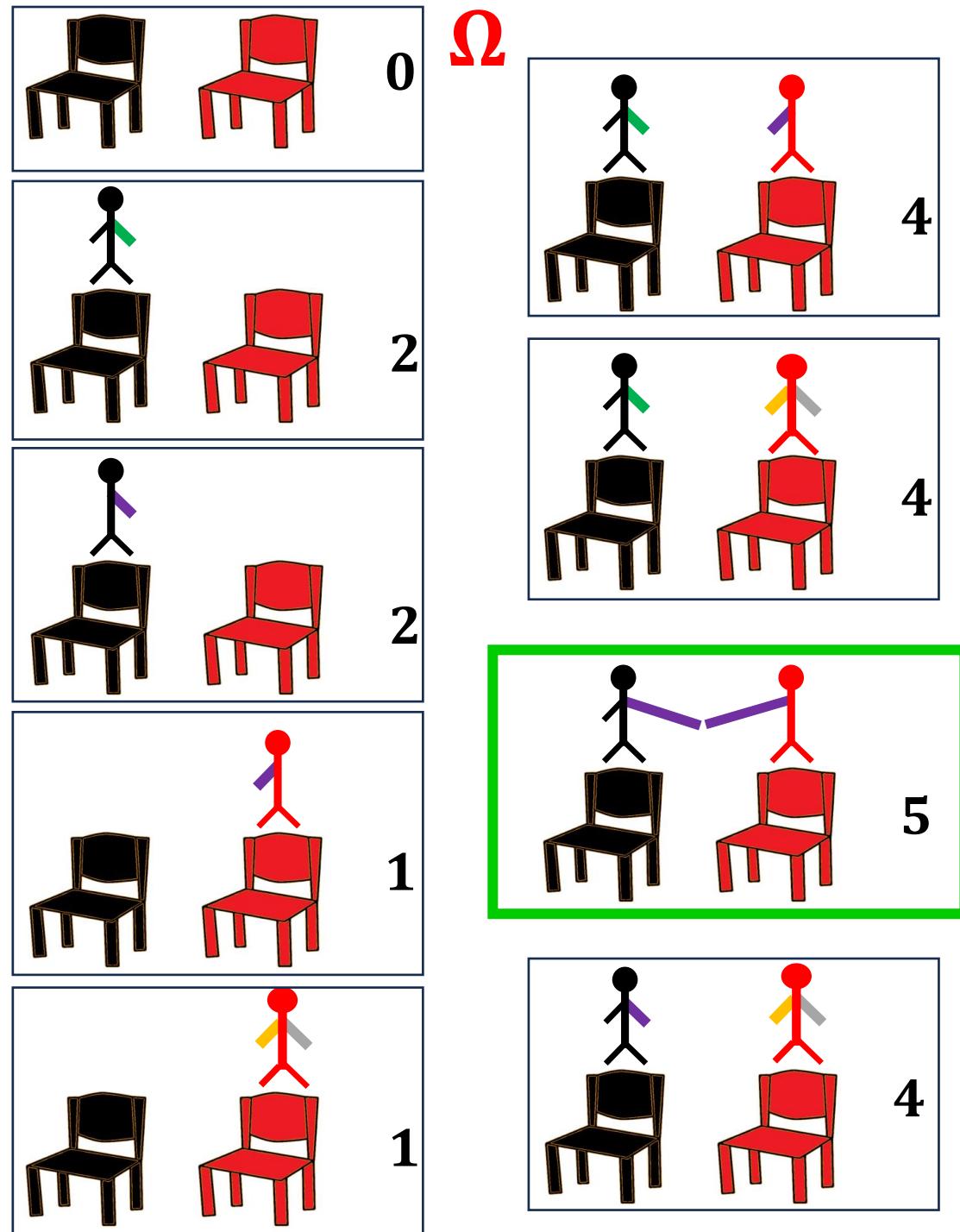


The Maximum Energy M

$$M := \max_{X \in \Omega} \{E(X)\}$$

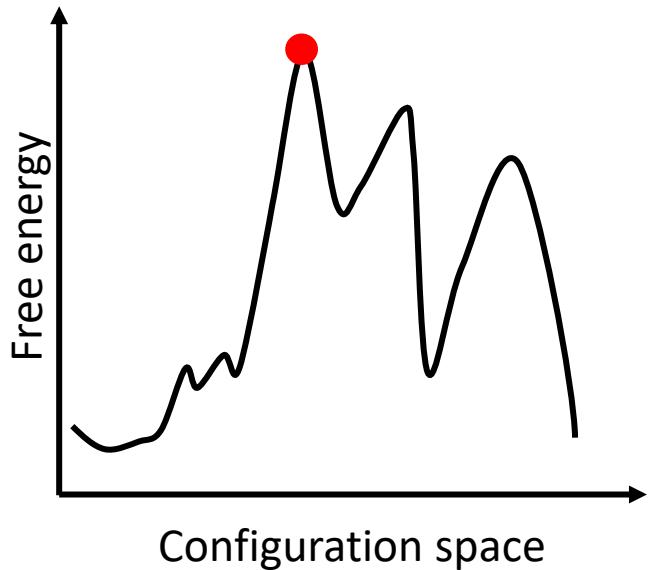


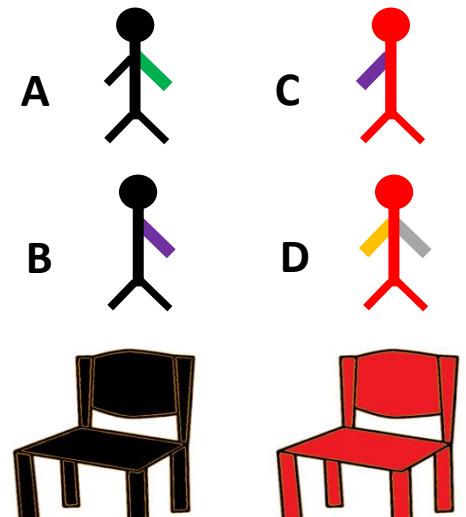
- $\text{sit}(A) = \text{sit}(B) = 3$
- $\text{sit}(C) = \text{sit}(D) = 3$
- $\text{handshake}(B, C) = 1$
- $\text{sit_convincing_cost} = -1$



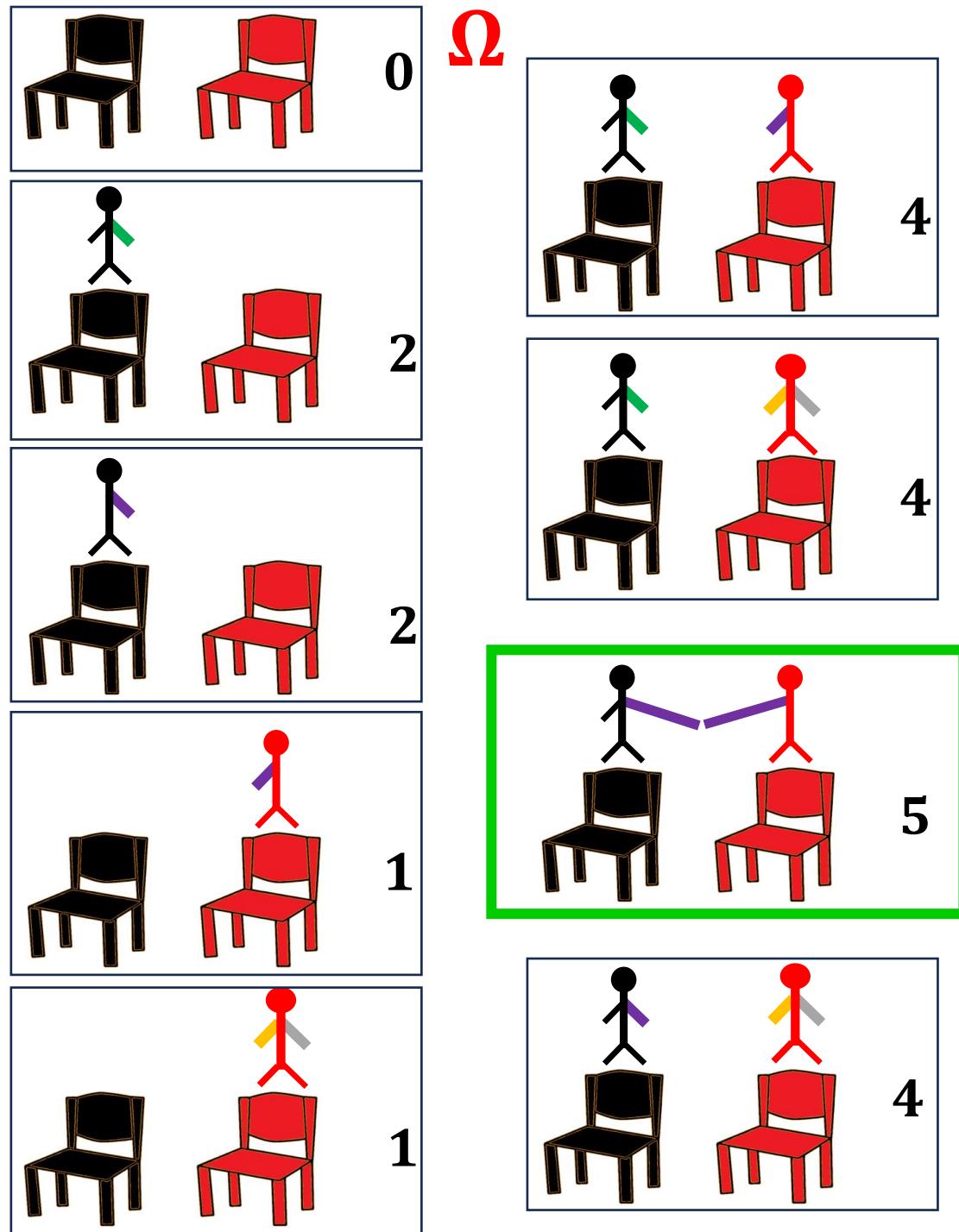
The Maximum Energy M

$$M := \max_{X \in \Omega} \{E(X)\}$$



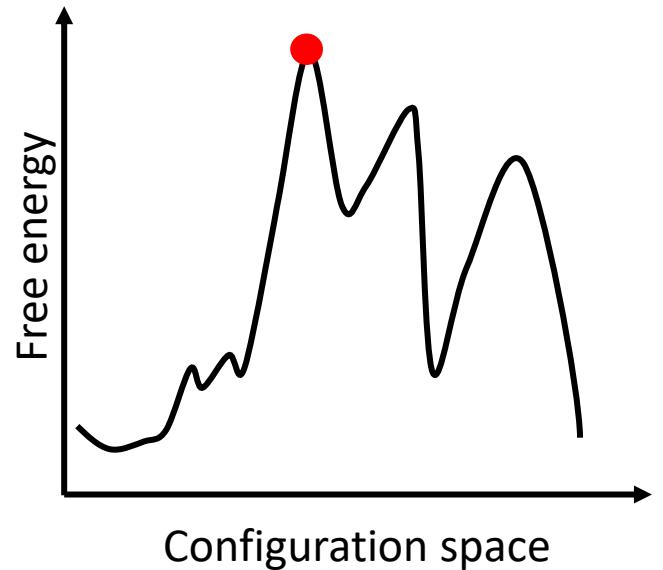


- $\text{sit}(A) = \text{sit}(B) = 3$
- $\text{sit}(C) = \text{sit}(D) = 3$
- $\text{handshake}(B, C) = 1$
- $\text{sit_convincing_cost} = -1$



The Maximum Energy M

$$M := \max_{X \in \Omega} \{E(X)\}$$



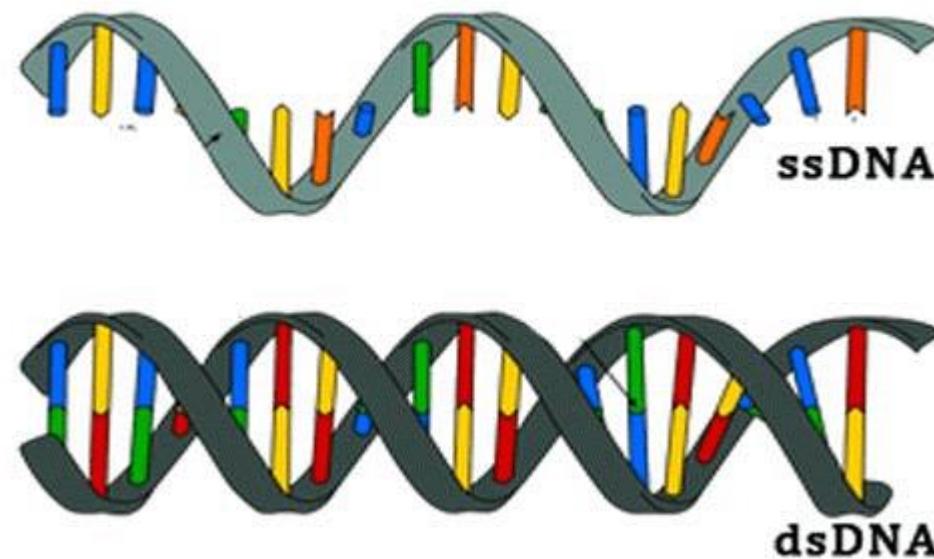
The Partition function Q

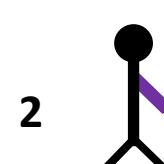
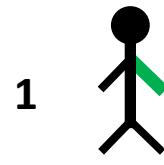
$$Q := \sum_{X \in \Omega} e^{\frac{E(X)}{c}}$$

Where c is some specific constant

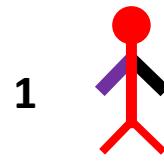
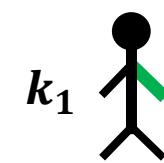
Why are we interested in this ?

Why are we interested in this ?

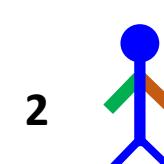
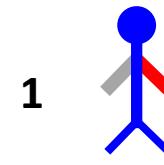
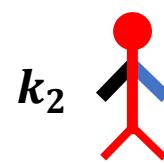




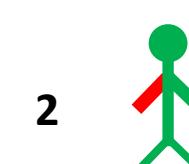
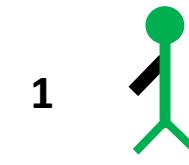
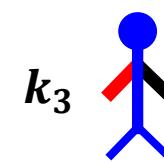
⋮



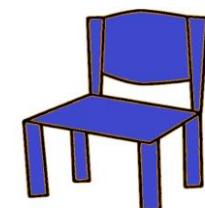
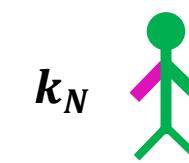
⋮



⋮

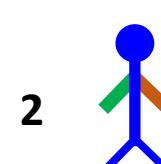
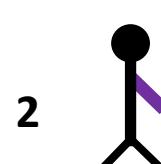
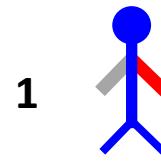
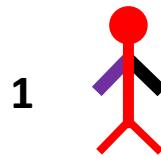
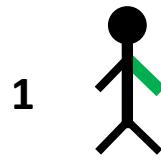


⋮



⋮ ⋮ ⋮

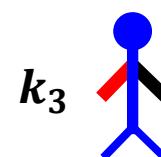
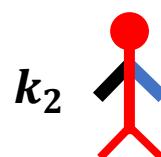
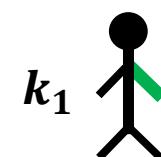


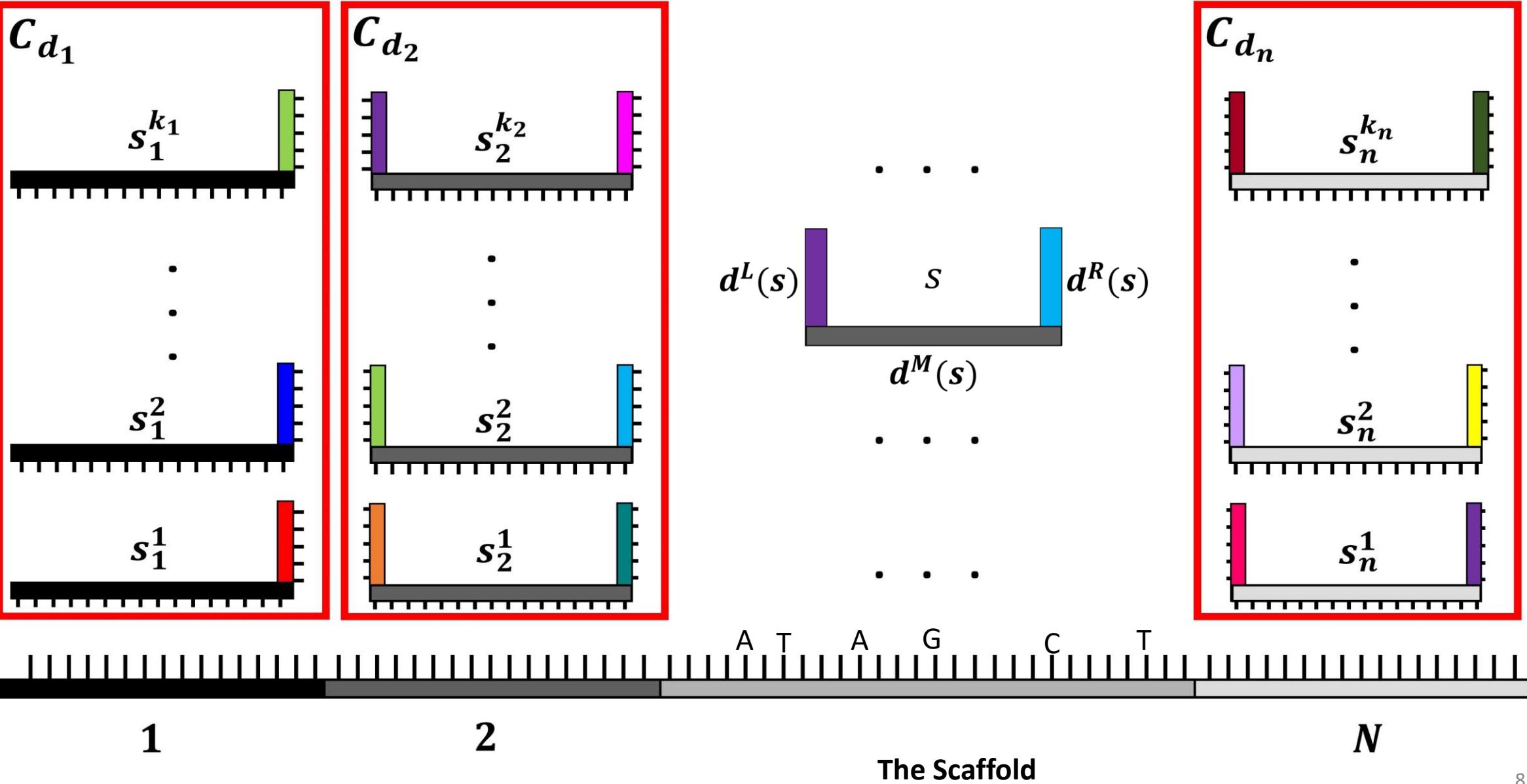


⋮
⋮

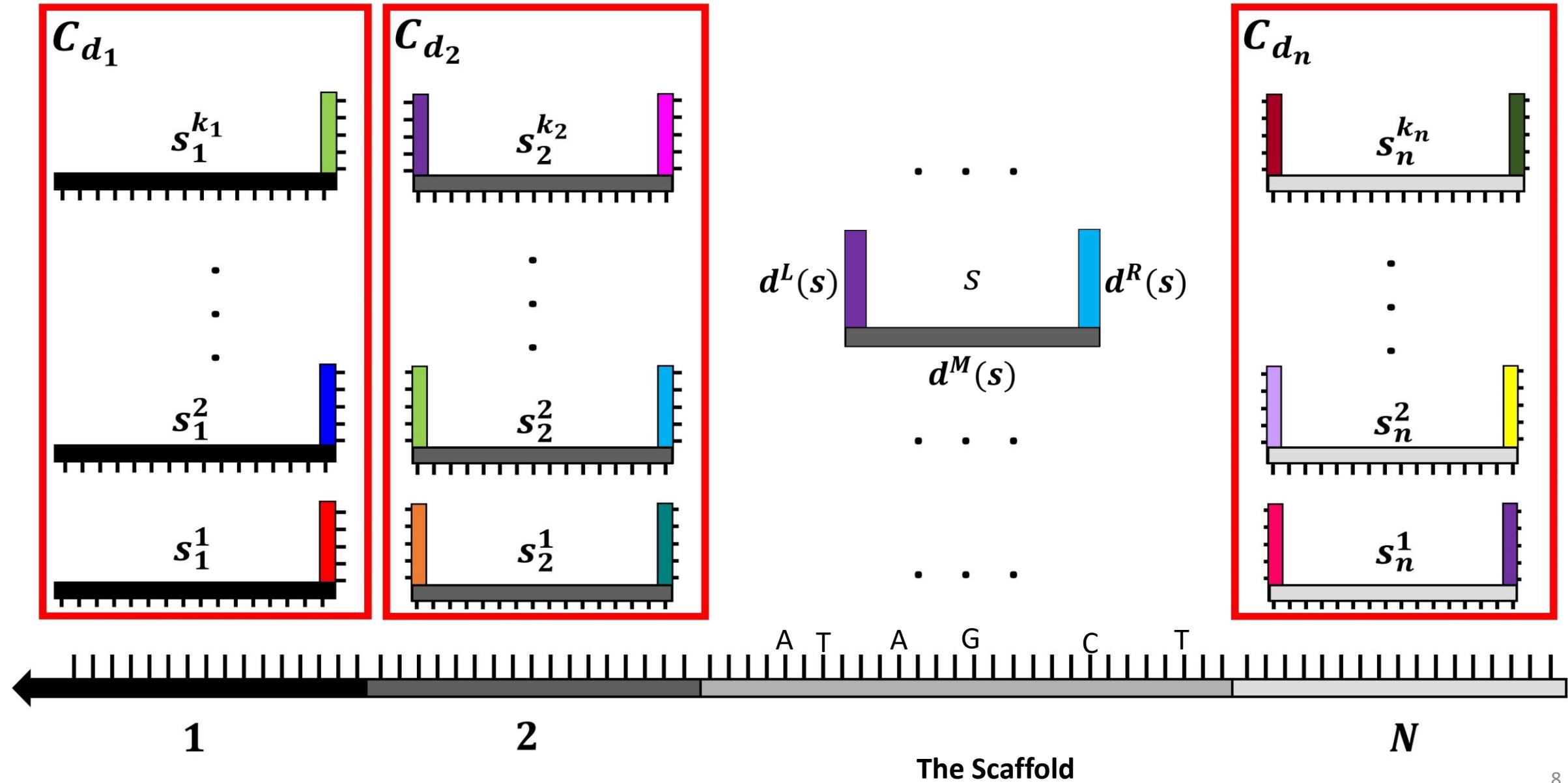
⋮
⋮

⋮
⋮





This is 1D Scaffolded DNA Computer



Built-in algorithmic self correction mechanism

DNA strand displacement

Built-in algorithmic self correction mechanism

DNA strand displacement

Microsoft Research

Materials supplied by Microsoft Corporation may be used for internal review, analysis, or research only. Any editing, reproduction, publication, rebroadcast, public showing, internet, or public display is forbidden and may violate copyright law.

2016 © Microsoft Corporation. All rights reserved.

1D Scaffolded DNA Computer Energy model and thermodynamic features

For any configuration X of size l :

Chair $E(X) = \sum_{p \in X} sit(p) + l * sitcost + \sum_{p_i, p_{i+1} \in X} handshake(p_i, p_{i+1}).$

1D Scaffolded DNA Computer Energy model and thermodynamic features

For any configuration X of size l :

Chair $E(X) = \sum_{p \in X} sit(p) + l * sitcost + \sum_{p_i, p_{i+1} \in X} handshake(p_i, p_{i+1}).$

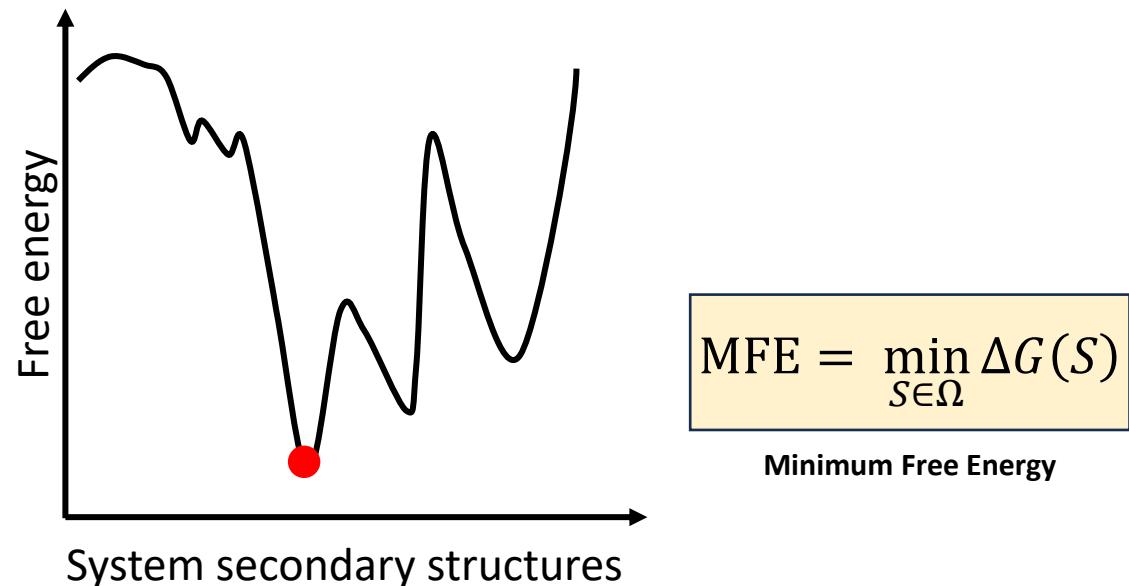
DNA $\Delta G^S(X) = \sum_{s \in X} \Delta G(d^M(s)) + l \cdot \Delta G^{\text{assoc}} + \sum_{s_i, s_{i+1} \in X} \Delta G(d^R(s_i), d^L(s_{i+1})).$

1D Scaffolded DNA Computer Energy model and thermodynamic features

For any configuration X of size l :

Chair $E(X) = \sum_{p \in X} sit(p) + l * sitcost + \sum_{p_i, p_{i+1} \in X} handshake(p_i, p_{i+1}).$

DNA $\Delta G^S(X) = \sum_{s \in X} \Delta G(d^M(s)) + l \cdot \Delta G^{\text{assoc}} + \sum_{s_i, s_{i+1} \in X} \Delta G(d^R(s_i), d^L(s_{i+1})).$

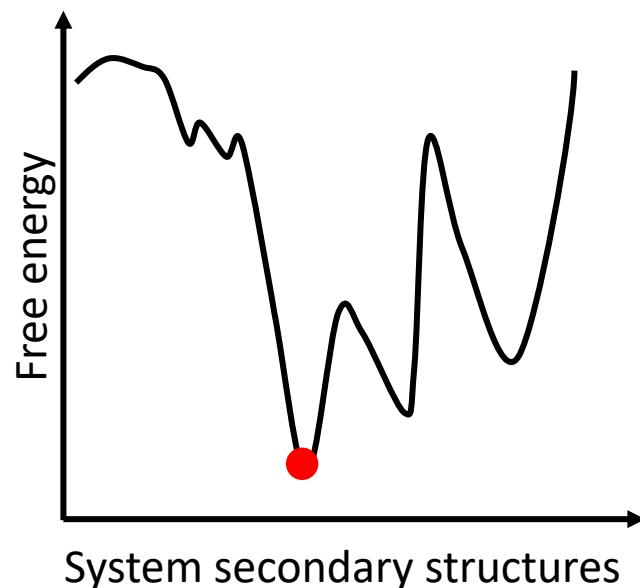


1D Scaffolded DNA Computer Energy model and thermodynamic features

For any configuration X of size l :

Chair $E(X) = \sum_{p \in X} sit(p) + l * sitcost + \sum_{p_i, p_{i+1} \in X} handshake(p_i, p_{i+1}).$

DNA $\Delta G^S(X) = \sum_{s \in X} \Delta G(d^M(s)) + l \cdot \Delta G^{\text{assoc}} + \sum_{s_i, s_{i+1} \in X} \Delta G(d^R(s_i), d^L(s_{i+1})).$



$$\text{MFE} = \min_{S \in \Omega} \Delta G(S)$$

Minimum Free Energy

Boltzmann weighted sum

$$Q = \sum_{S \in \Omega} e^{-\Delta G(S)/k_B T}$$

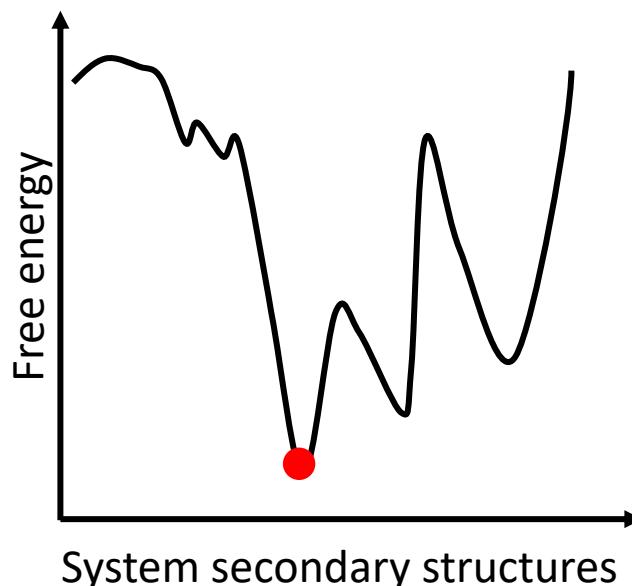
Partition Function

1D Scaffolded DNA Computer Energy model and thermodynamic features

For any configuration X of size l :

Chair $E(X) = \sum_{p \in X} sit(p) + l * sitcost + \sum_{p_i, p_{i+1} \in X} handshake(p_i, p_{i+1}).$

DNA $\Delta G^S(X) = \sum_{s \in X} \Delta G(d^M(s)) + l \cdot \Delta G^{\text{assoc}} + \sum_{s_i, s_{i+1} \in X} \Delta G(d^R(s_i), d^L(s_{i+1})).$



$$\text{MFE} = \min_{S \in \Omega} \Delta G(S)$$

Minimum Free Energy

Boltzmann weighted sum

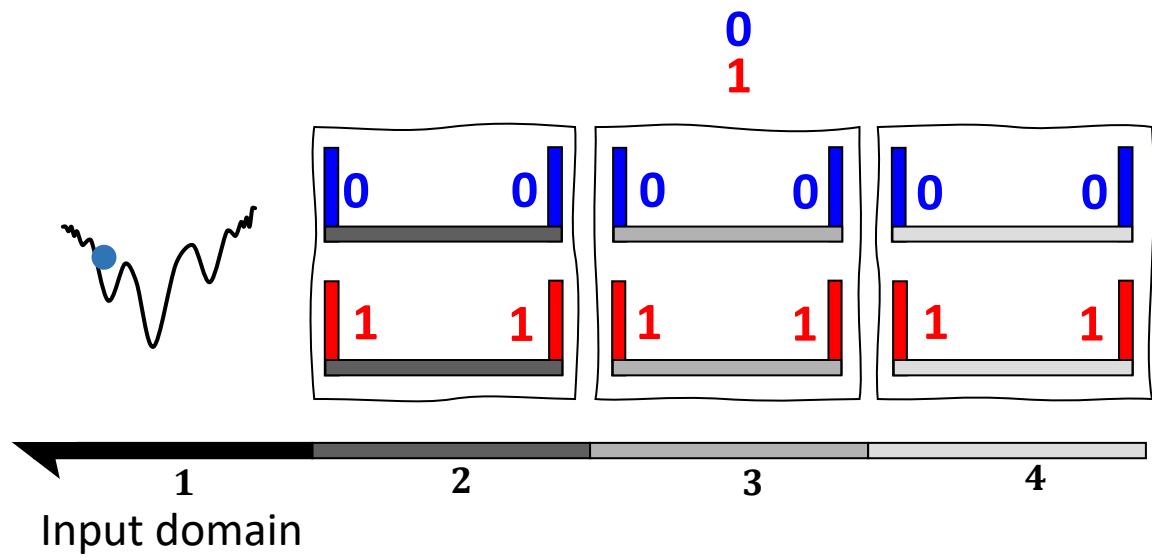
$$Q = \sum_{S \in \Omega} e^{-\Delta G(S)/k_B T}$$

Partition Function

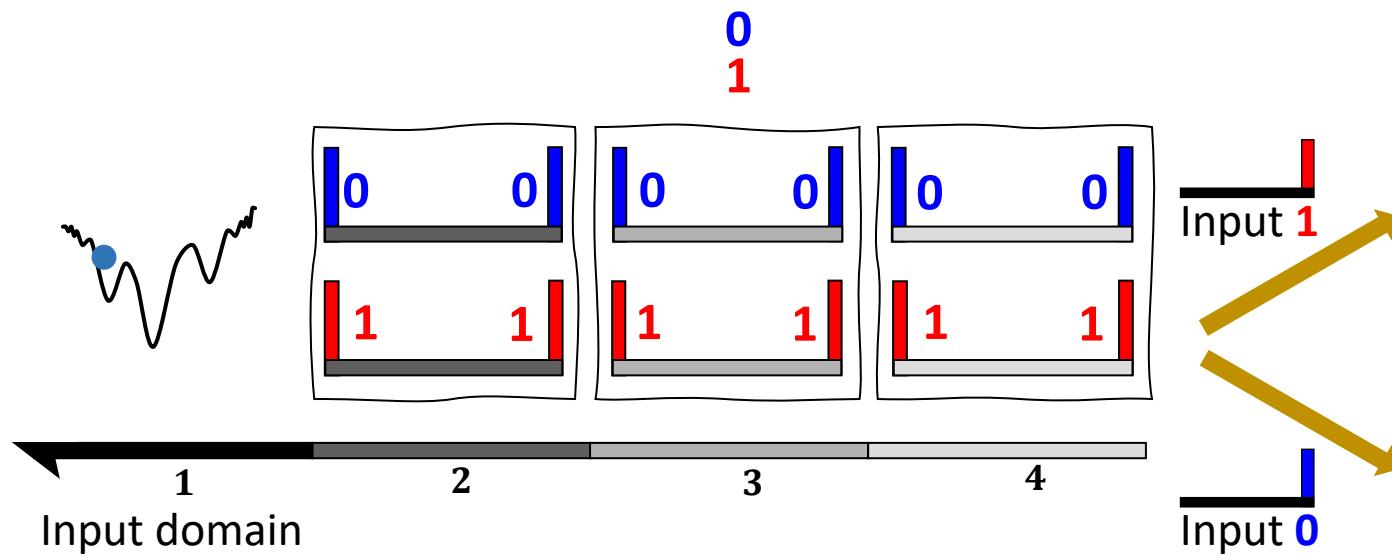


$$\Pr[S] = \frac{e^{-\Delta G(S)/k_B T}}{Q}$$

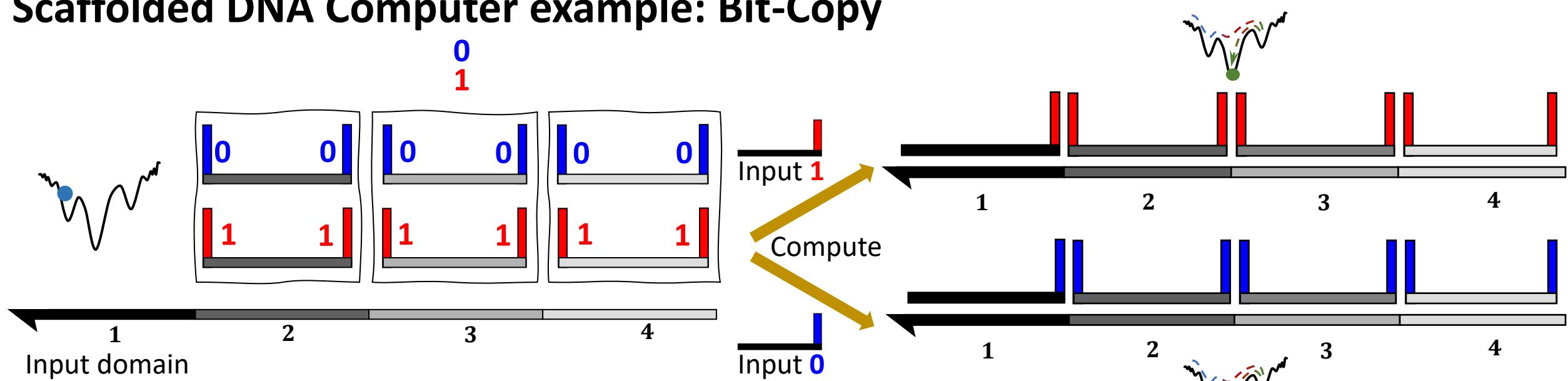
Scaffolded DNA Computer example: Bit-Copy



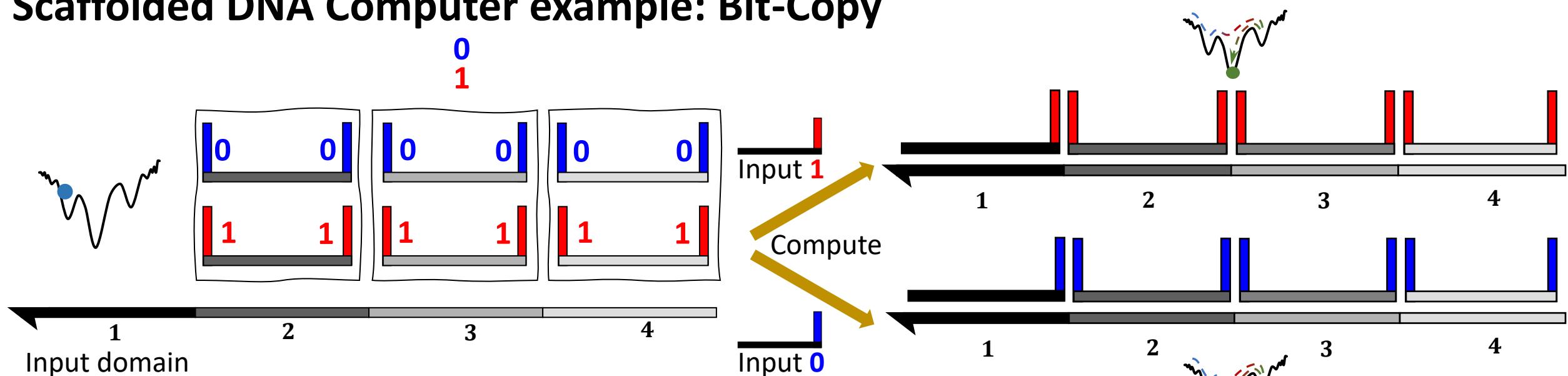
Scaffolded DNA Computer example: Bit-Copy



Scaffolded DNA Computer example: Bit-Copy



Scaffolded DNA Computer example: Bit-Copy



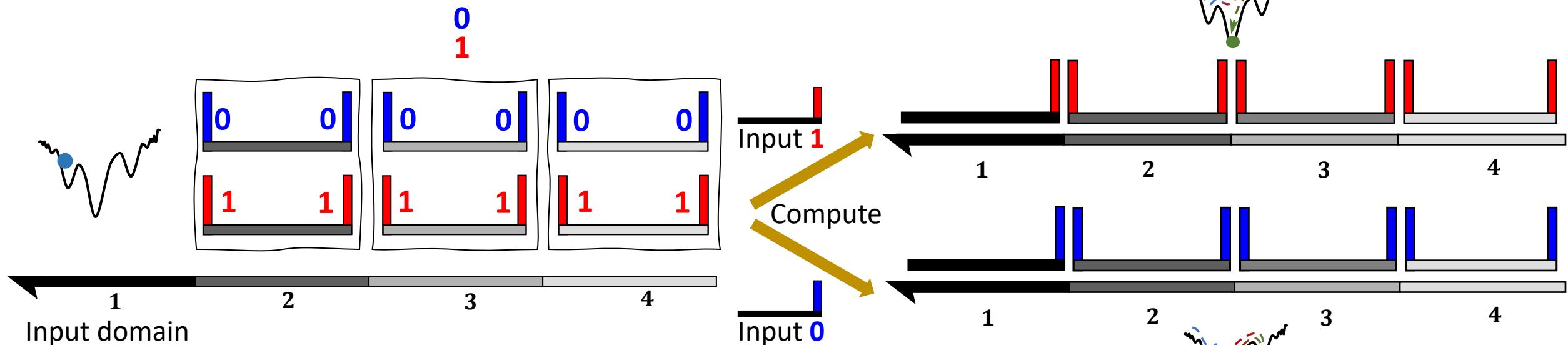
GOAL

$$\Pr[\text{Scaffold configuration}] \gg \sum_c \Pr[c : \text{is another configuration}]$$

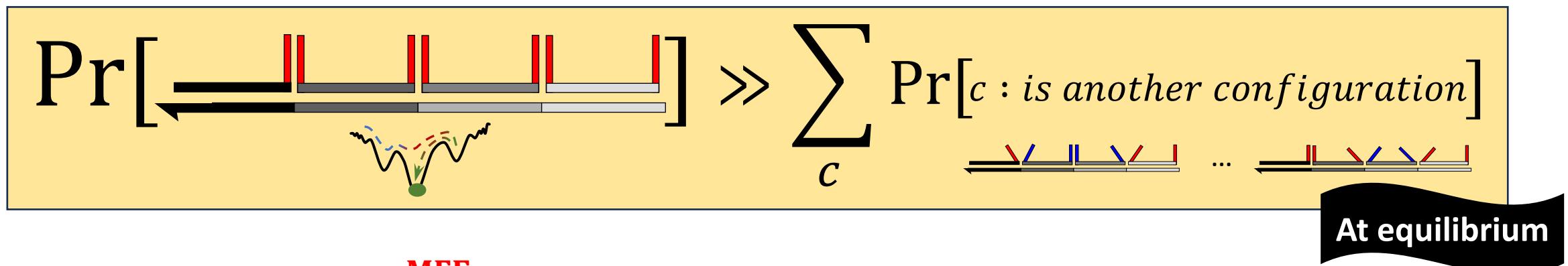
At equilibrium

The equation expresses the goal of the DNA computer: the probability of the target scaffold configuration being present is much greater than the sum of the probabilities of all other configurations. The term "At equilibrium" is shown in a black banner below the equation.

Scaffolded DNA Computer example: Bit-Copy



GOAL



$$\Pr[\text{target}] = \frac{e^{-\Delta G(\text{target}) / k_B T}}{Q}$$

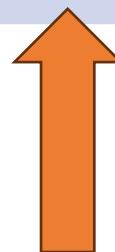
Computational complexity of Minimum Free Energy and the Partition Function

Input Type	MFE	Partition Function
Single Strand	$O(n^3)$	$O(n^3)$
Multiple Strands, Bounded ($\leq s$)	?	$O(n^3)(s - 1)!$
Multiple Strands, Unbounded	$NP - \text{Complete}$?

n bases, s strands

Computational complexity of Minimum Free Energy and the Partition Function

Input Type	MFE	Partition Function
Single Strand	$O(n^3)$	$O(n^3)$
Multiple Strands, Bounded ($\leq s$)	?	$O(n^3)(s - 1)!$
Multiple Strands, Unbounded	$NP - \text{Complete}$?



n bases, s strands

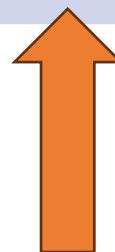
GOAL

At equilibrium

$$\Pr\left[\begin{array}{c} \text{Diagram of a single-stranded RNA molecule with hairpins and bulges} \\ \text{with a wavy line below it representing thermal fluctuations} \end{array}\right] \gg \sum_c \Pr[c : \text{is another configuration}]$$

Computational complexity of Minimum Free Energy and the Partition Function

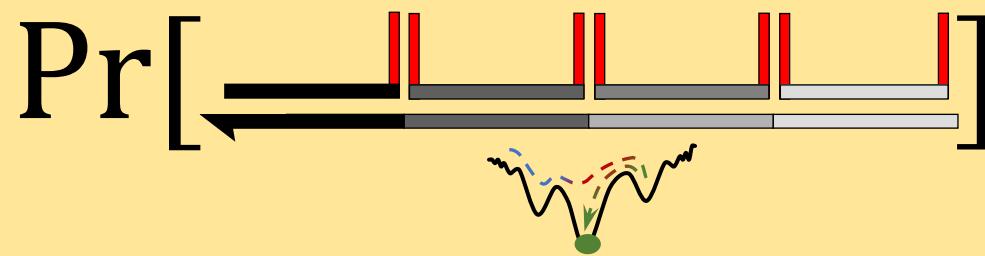
Input Type	MFE	Partition Function
Single Strand	$O(n^3)$	$O(n^3)$
Multiple Strands, Bounded ($\leq s$)	?	$O(n^3)(s - 1)!$
Multiple Strands, Unbounded	$NP - \text{Complete}$?



n bases, s strands

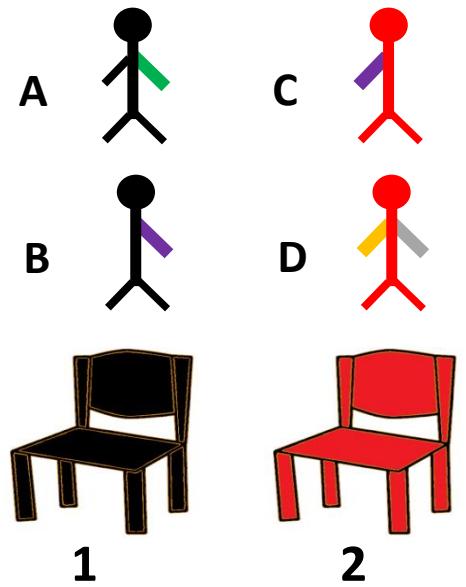
GOAL

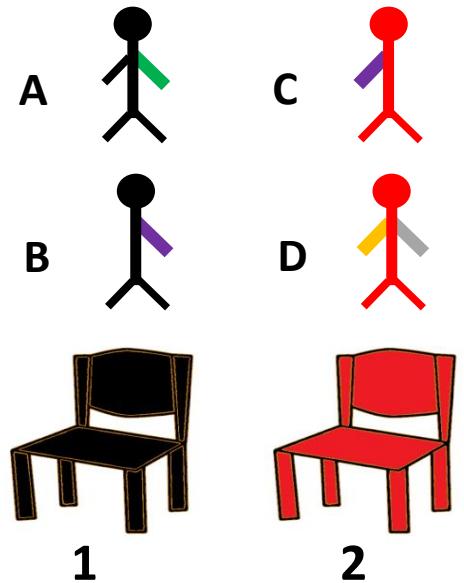
At equilibrium



$$\gg \sum_c \Pr[c : \text{is another configuration}]$$

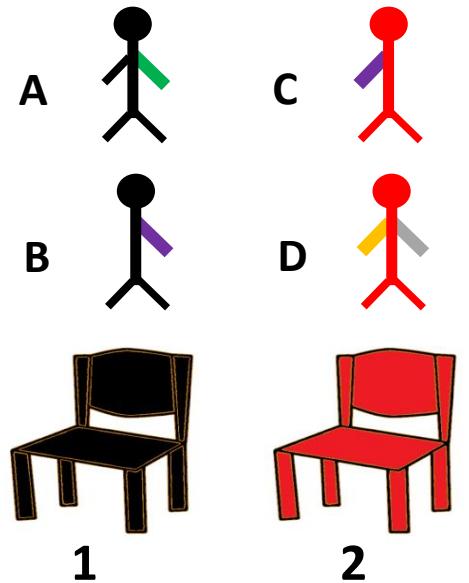
Efficiently





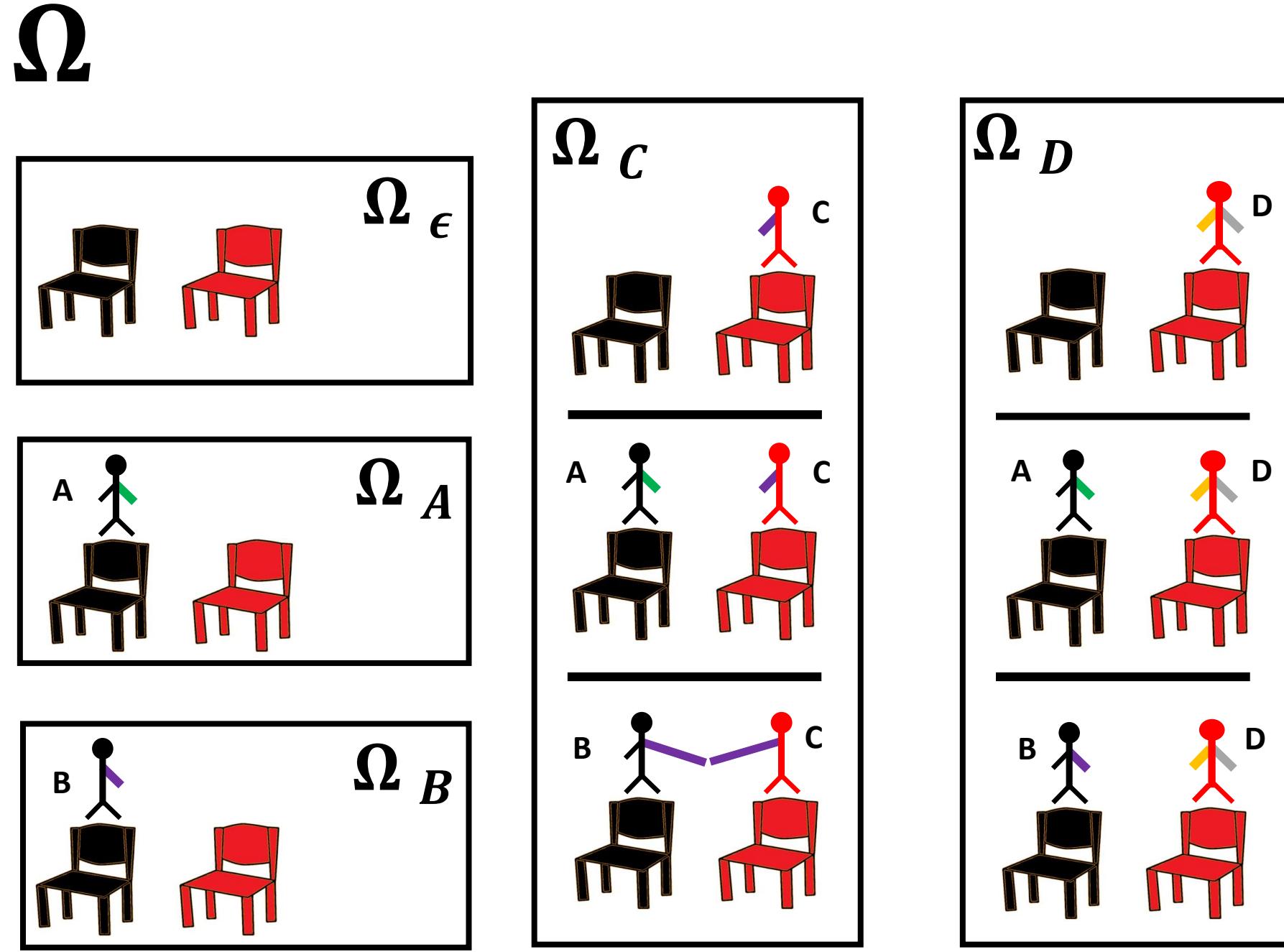
$$|\Omega| = (k + 1)^N = 9$$

$$Q = \sum_{p \in \Omega} e^{-\Delta G(p)/k_B T}$$

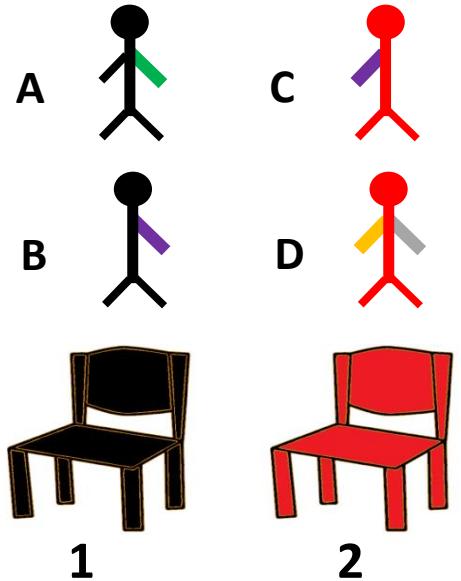


$$|\Omega| = (k + 1)^N = 9$$

$$Q = \sum_{p \in \Omega} e^{-\Delta G(p)/k_B T}$$

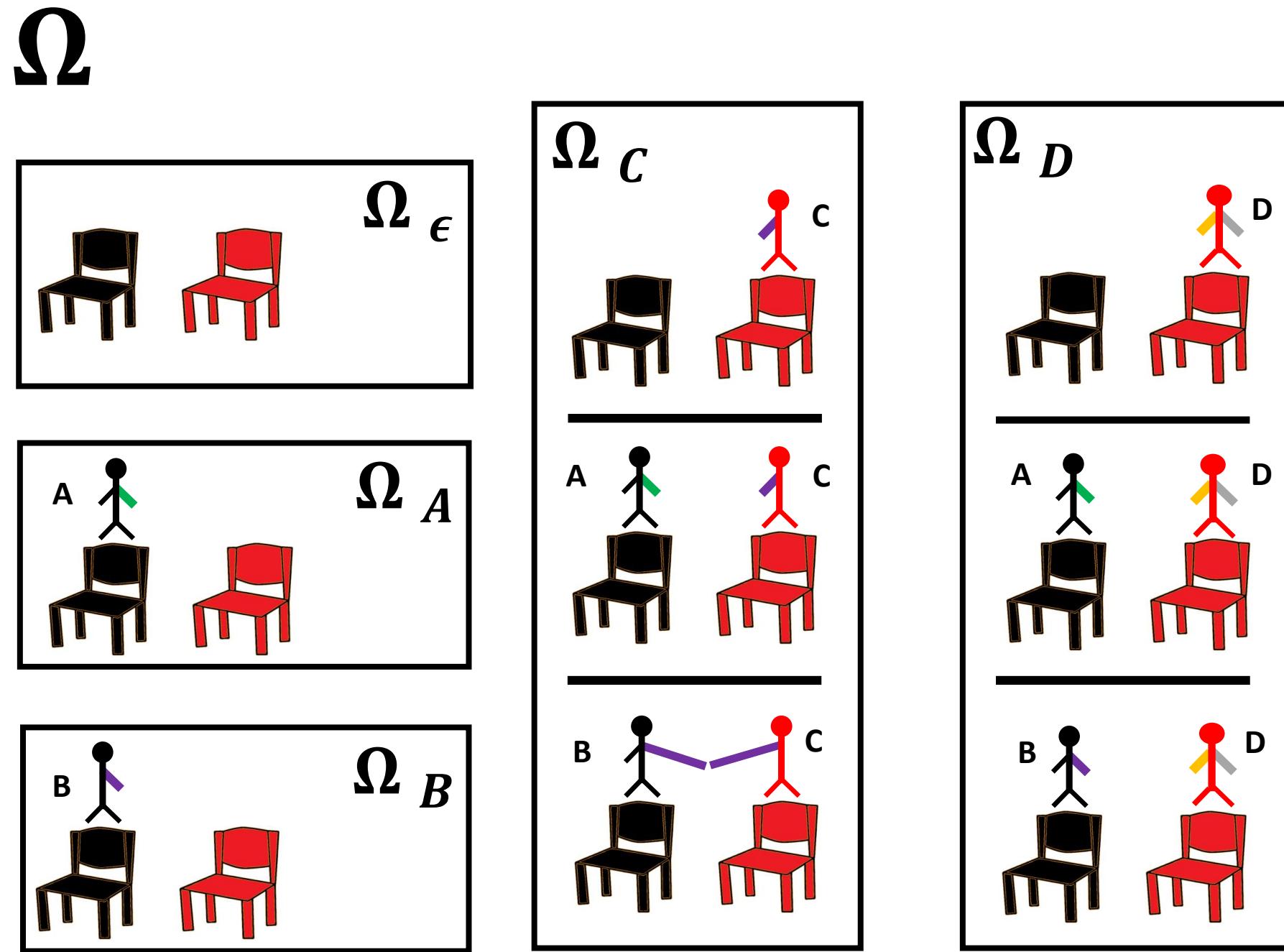


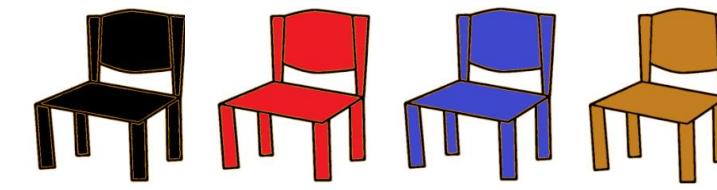
Is there recursive way
to build these classes
from themselves?



$$|\Omega| = (k + 1)^N = 9$$

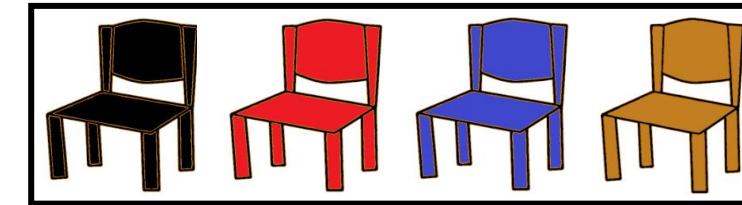
$$Q = \sum_{p \in \Omega} e^{-\Delta G(p)/k_B T}$$



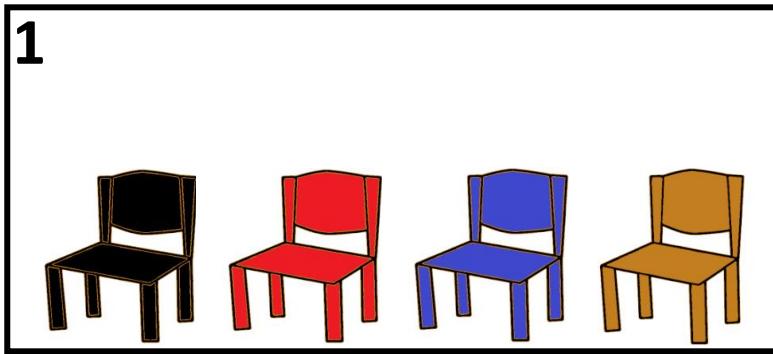
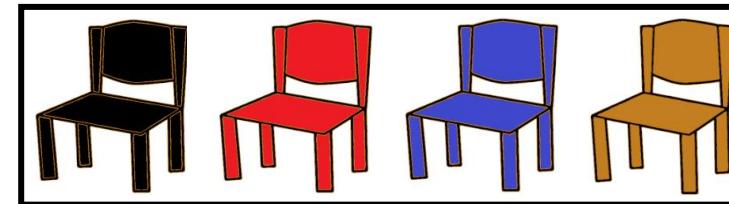


0

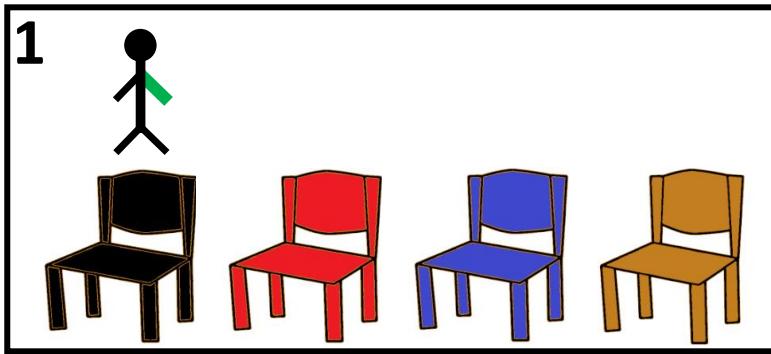
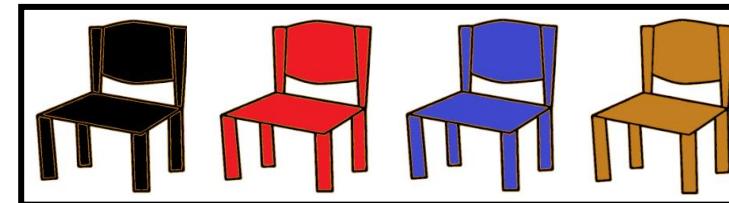
0



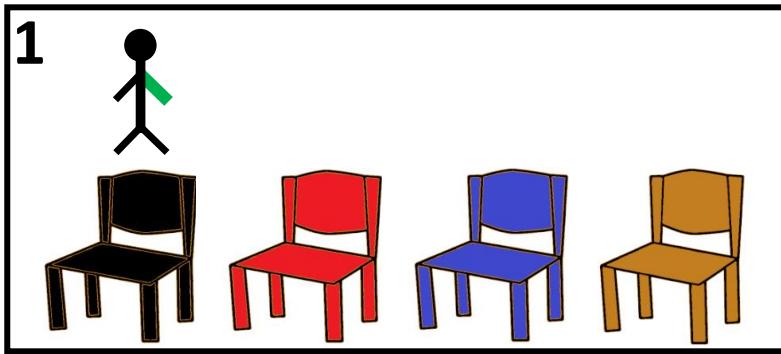
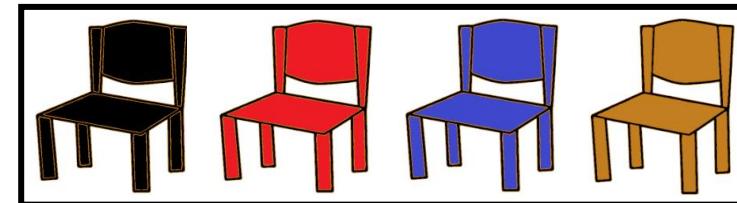
0



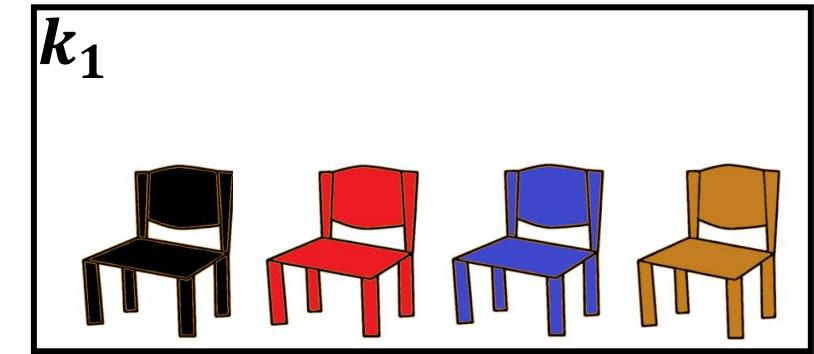
0



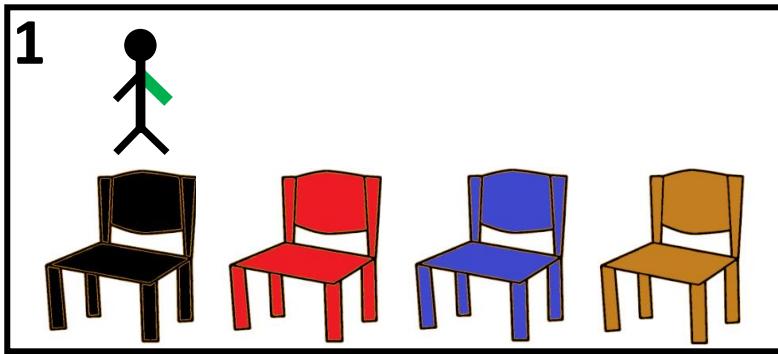
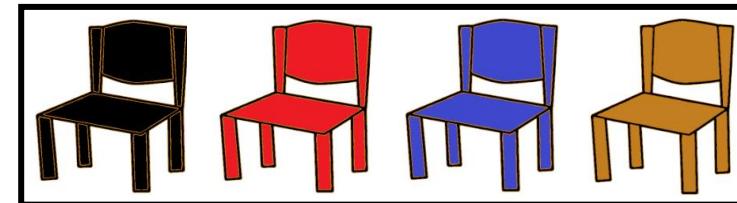
0



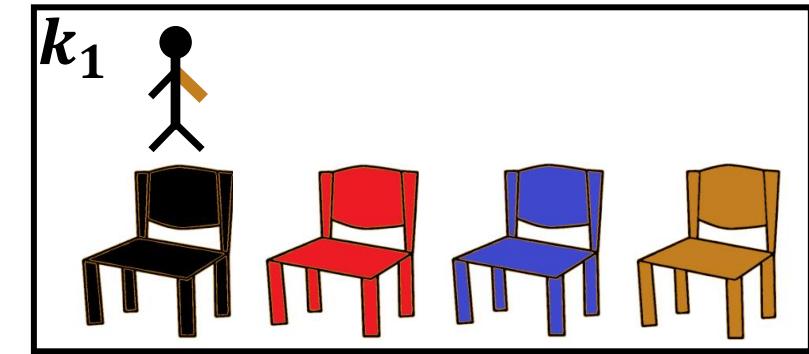
• • •



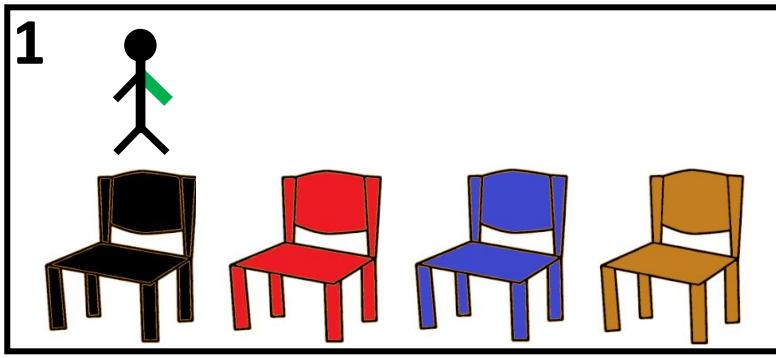
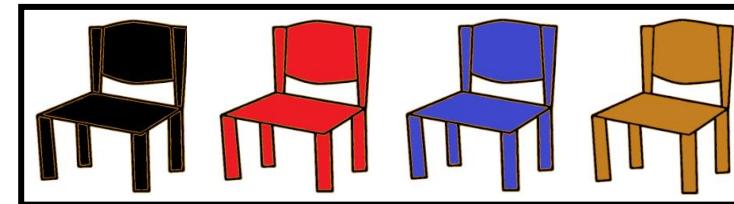
0



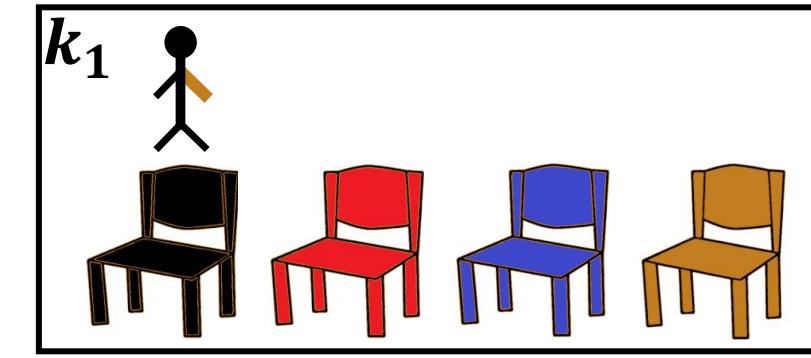
...



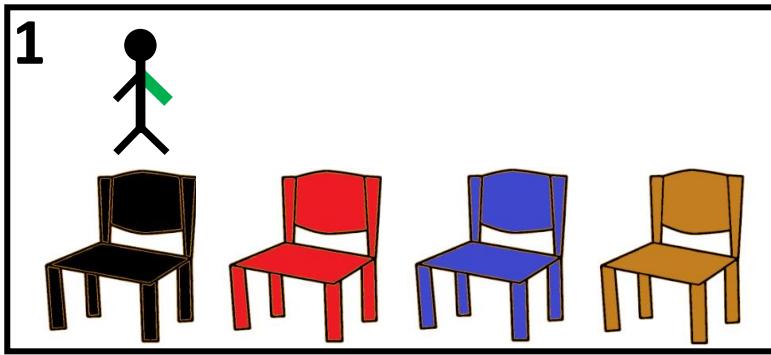
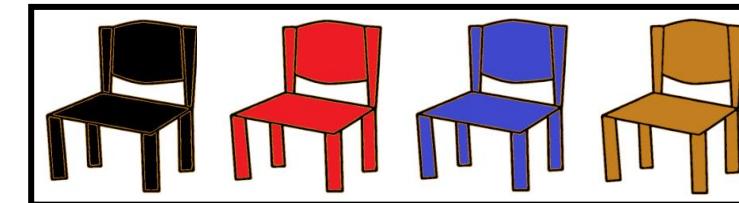
0



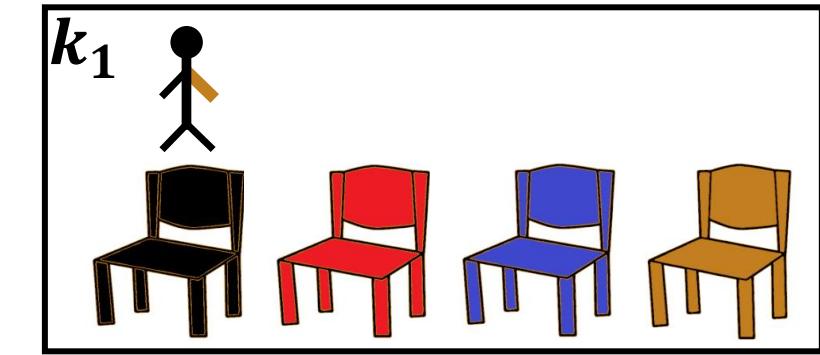
...



0



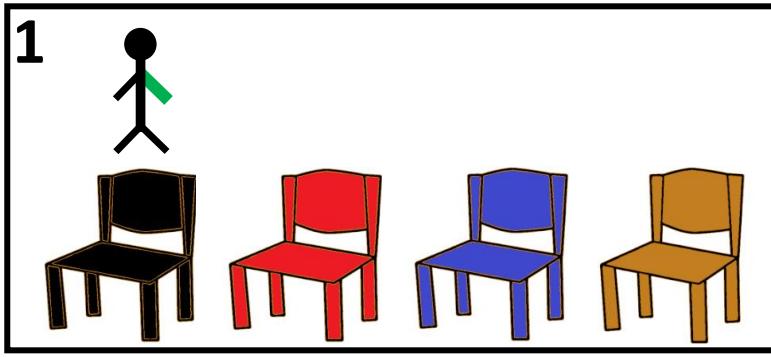
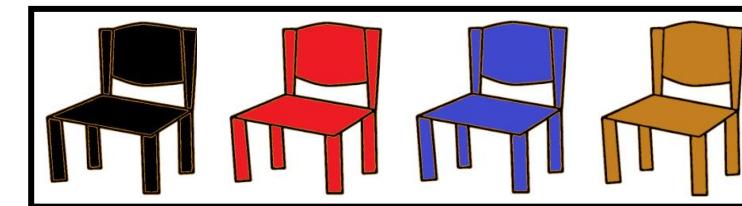
...



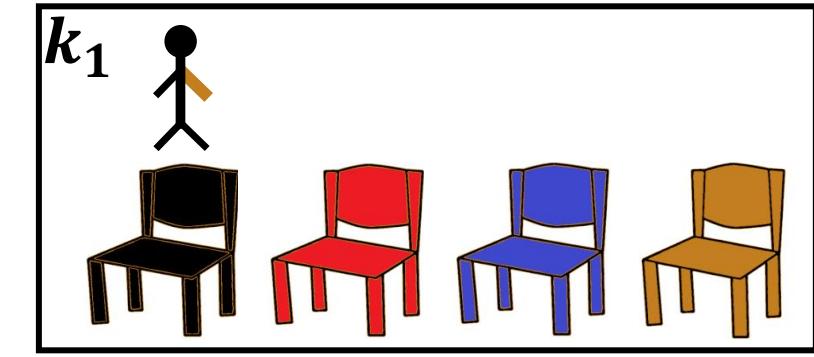
1

2

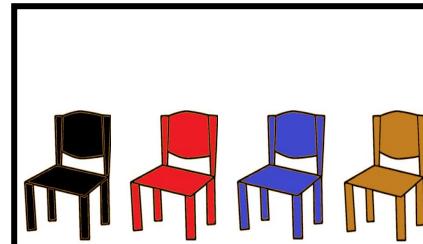
0



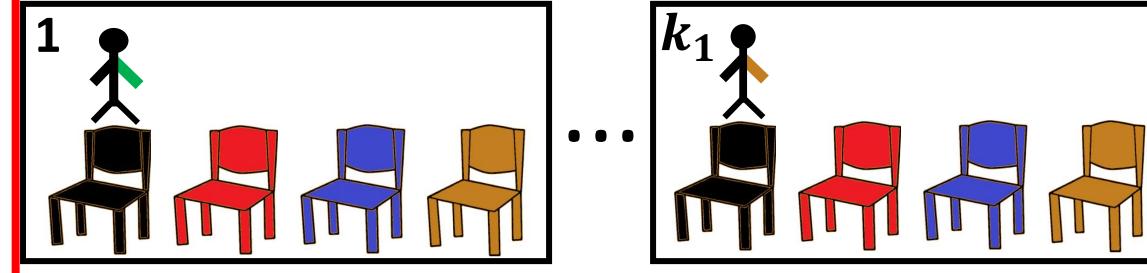
...



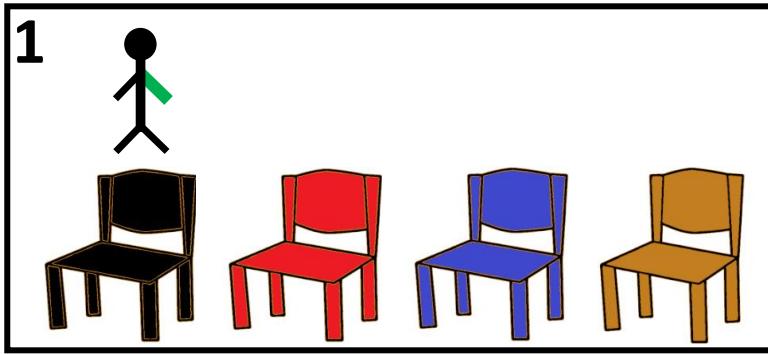
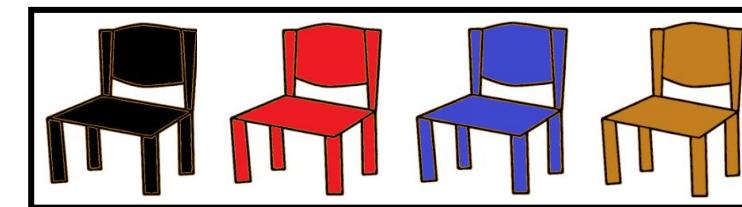
1



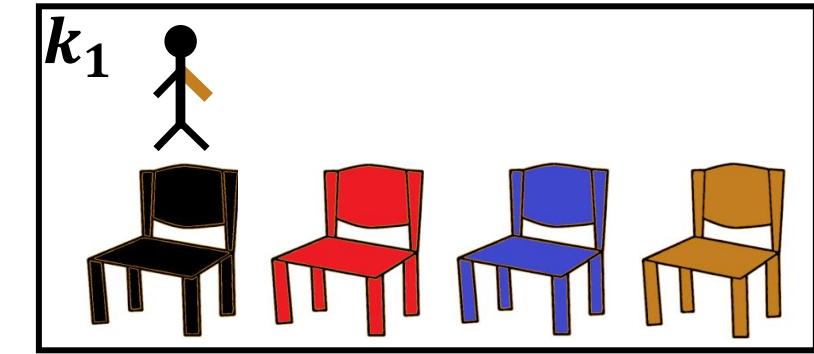
2



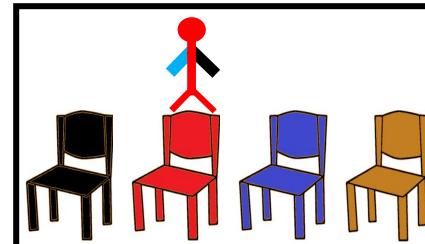
0



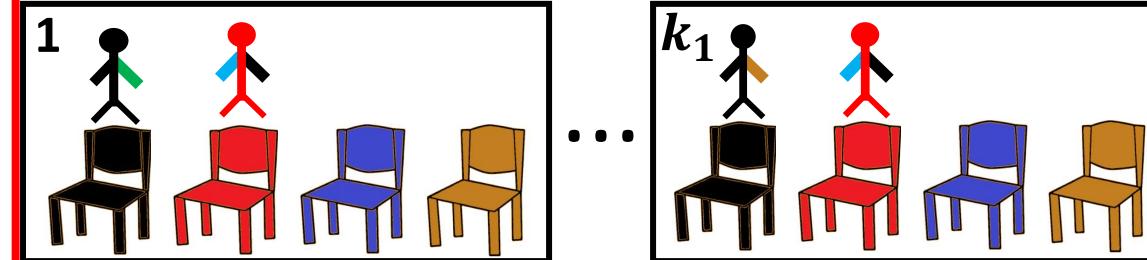
...



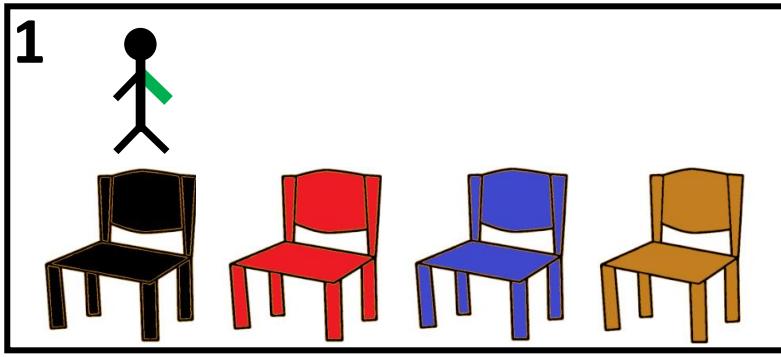
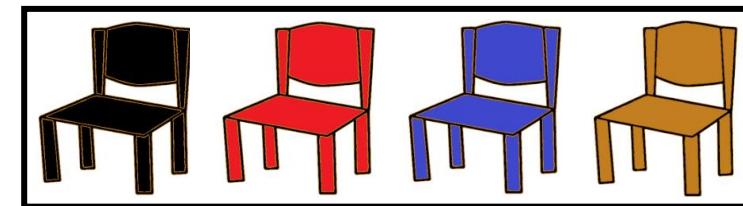
1



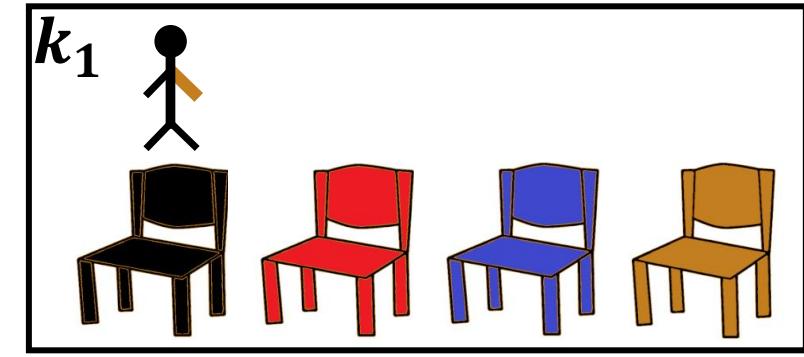
2



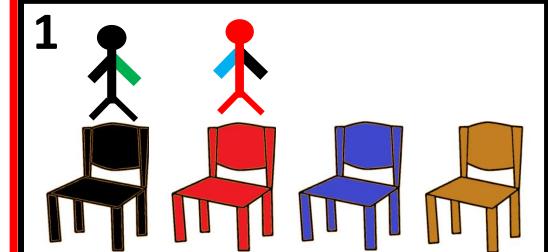
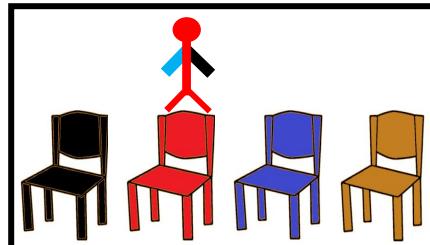
0



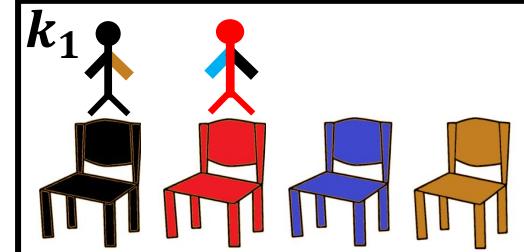
...



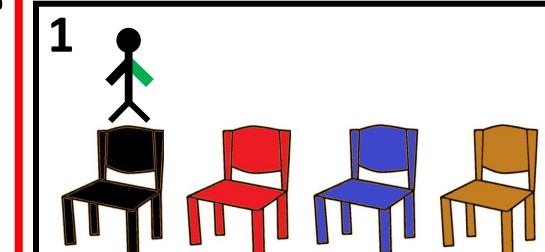
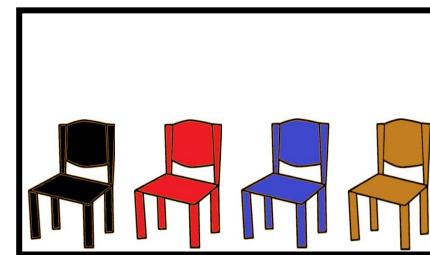
1



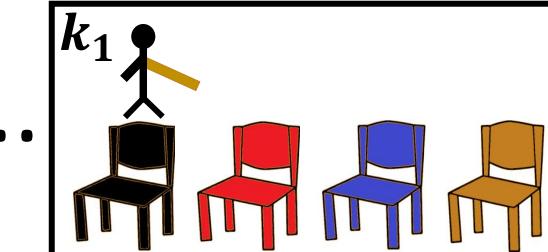
...



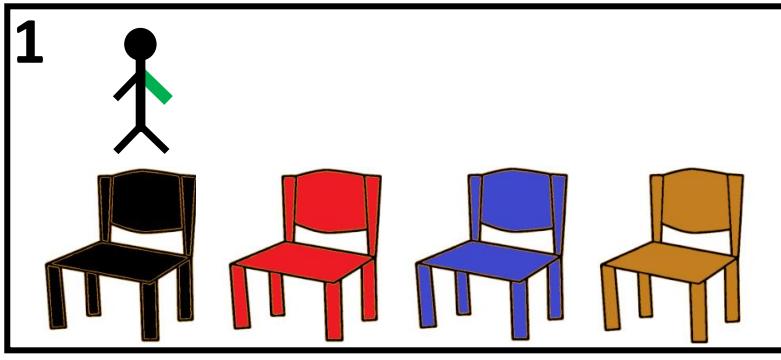
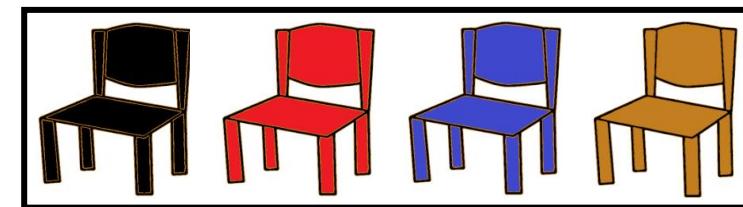
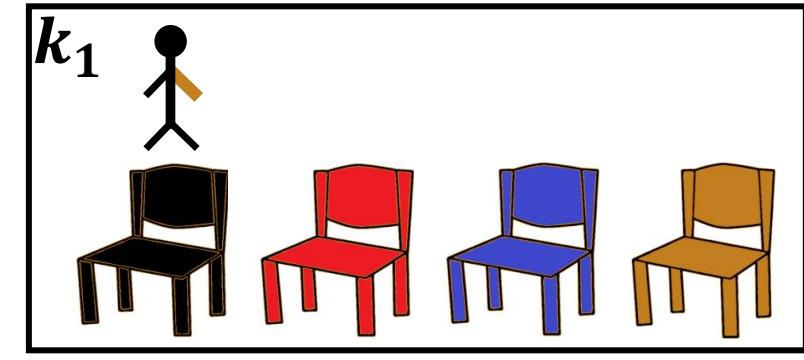
...

 k_2 

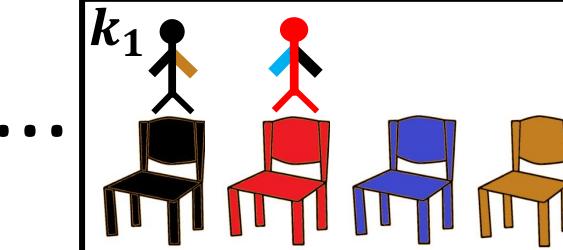
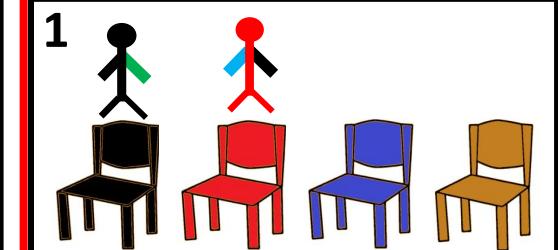
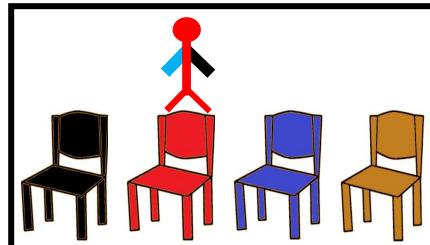
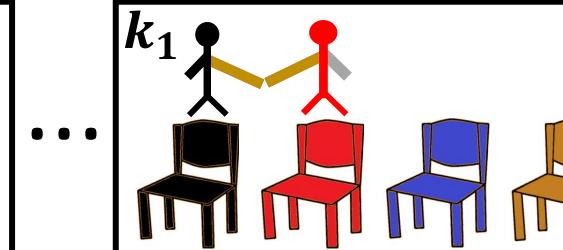
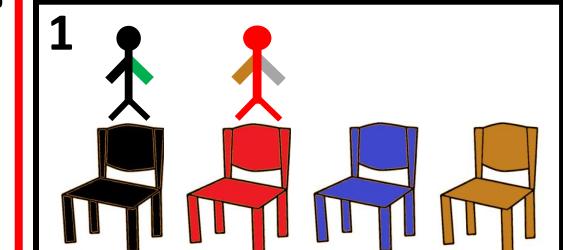
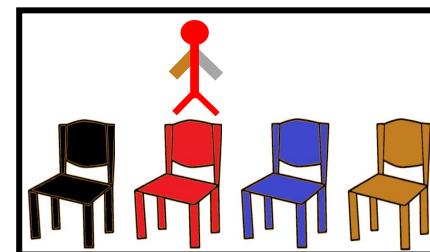
...

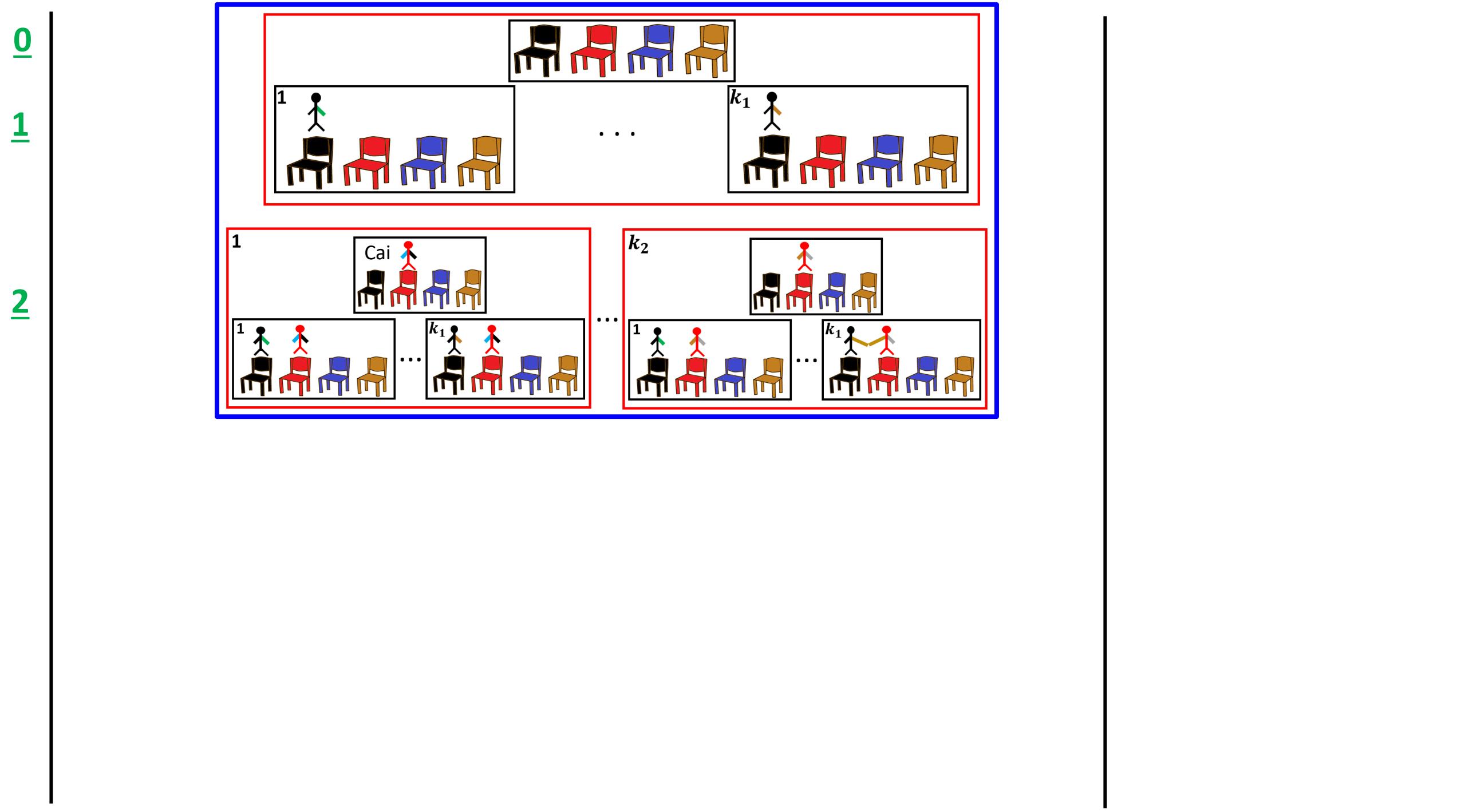


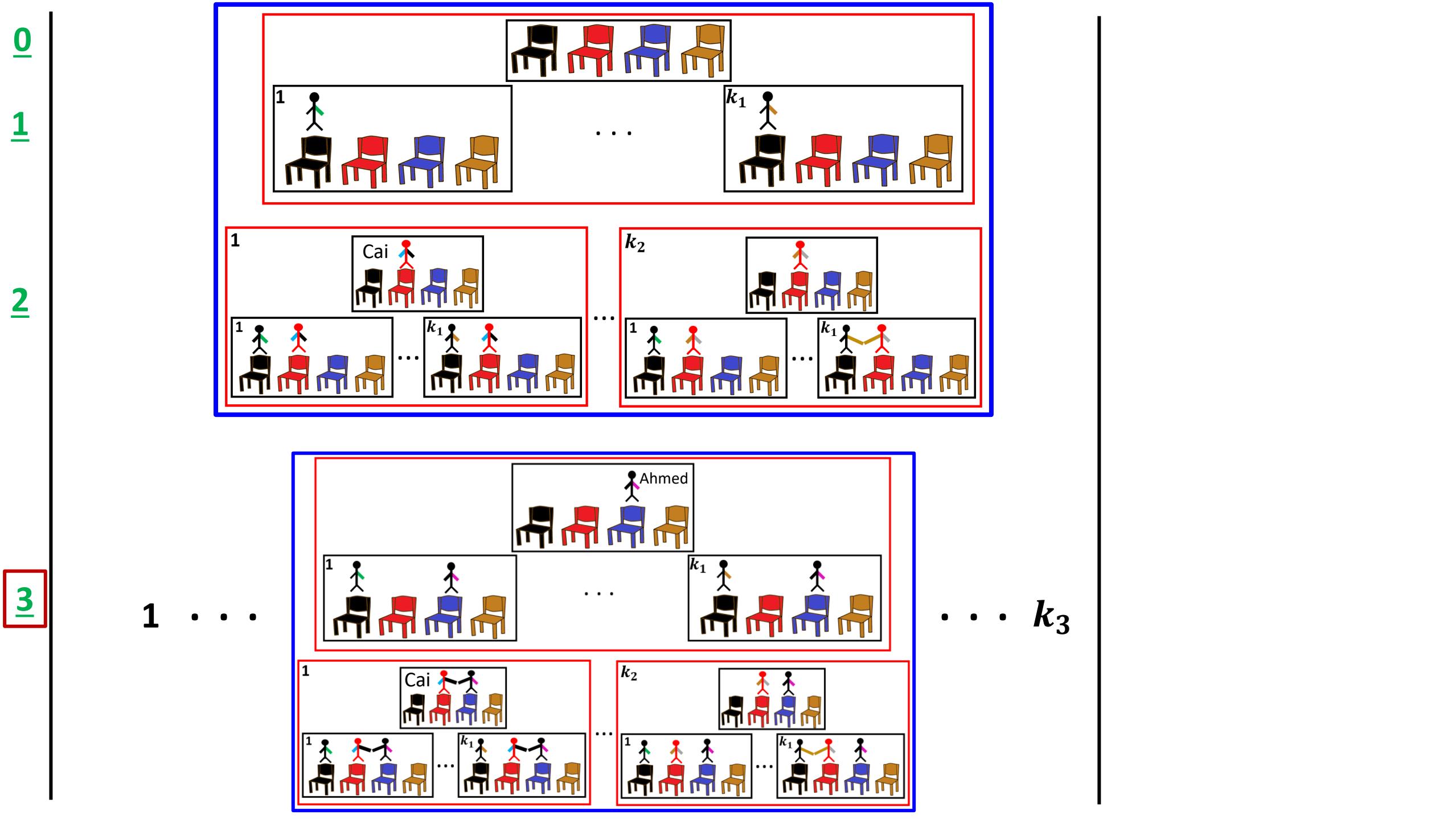
0

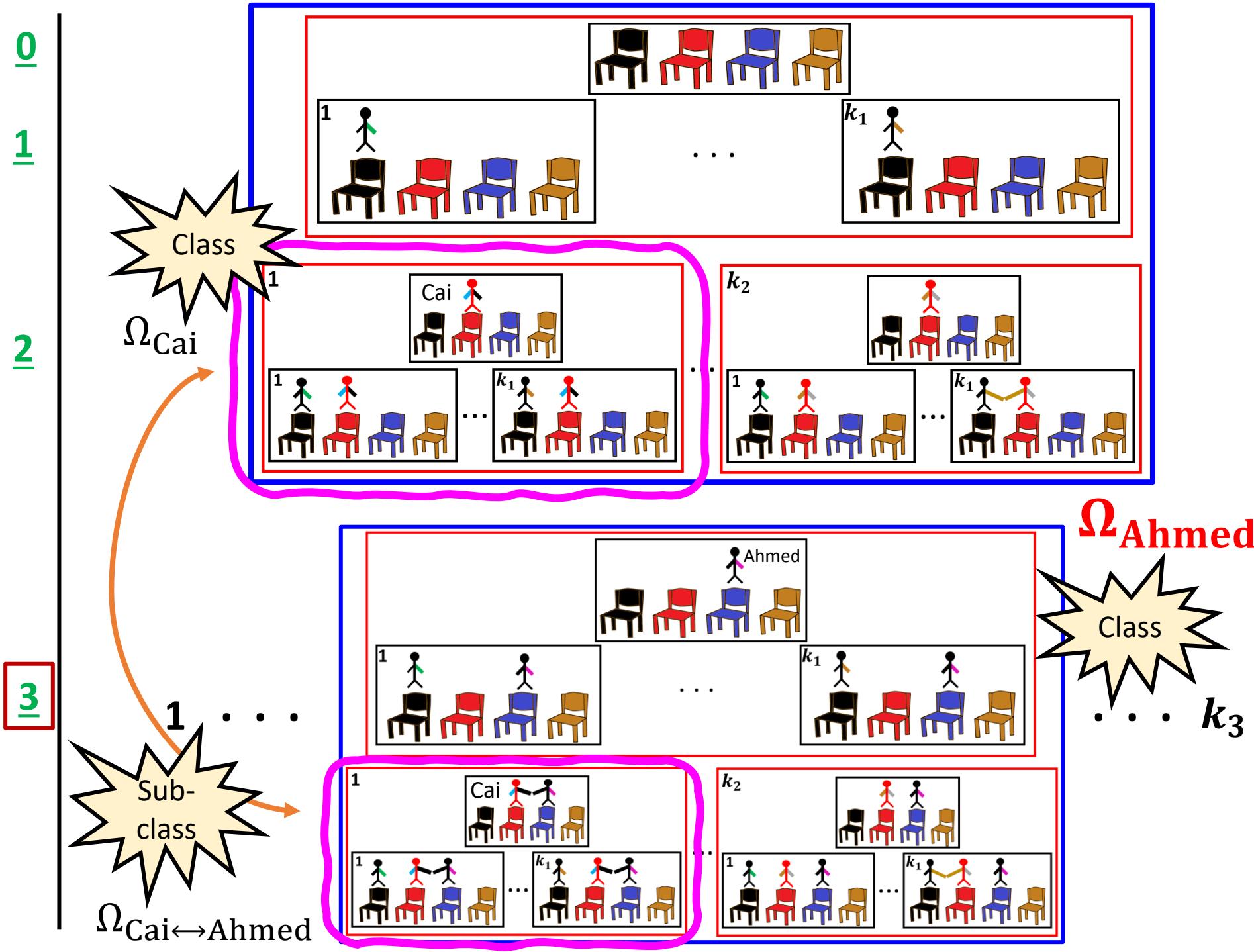
 \dots 

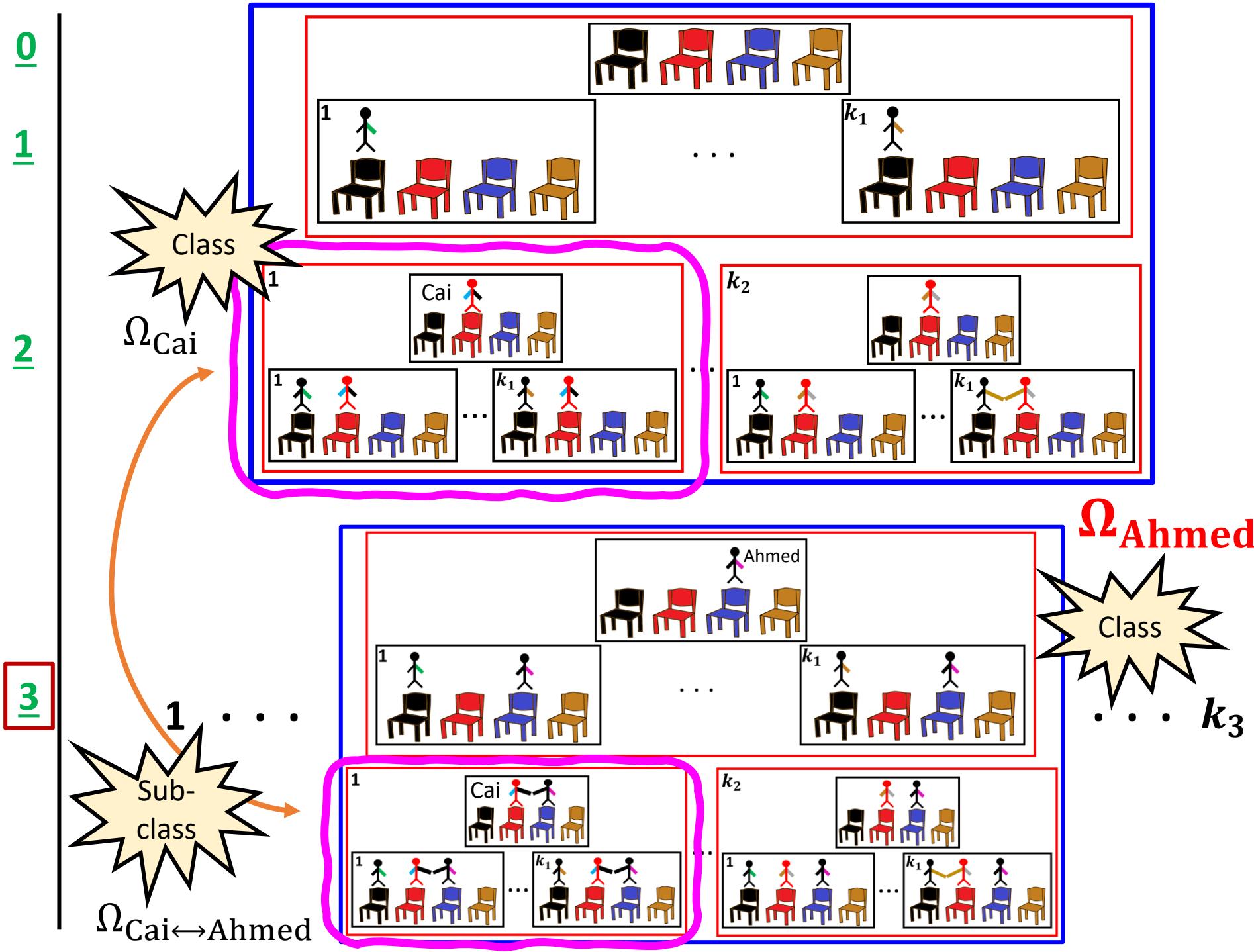
1

 \dots k_2 









Layer = chair

There is a 1-1 correspondence between each **class** of higher layers and a **sub-class** of any class in the current layer.

Ω_{Cai}



$\Omega_{\text{Cai} \leftrightarrow \text{Ahmed}}$

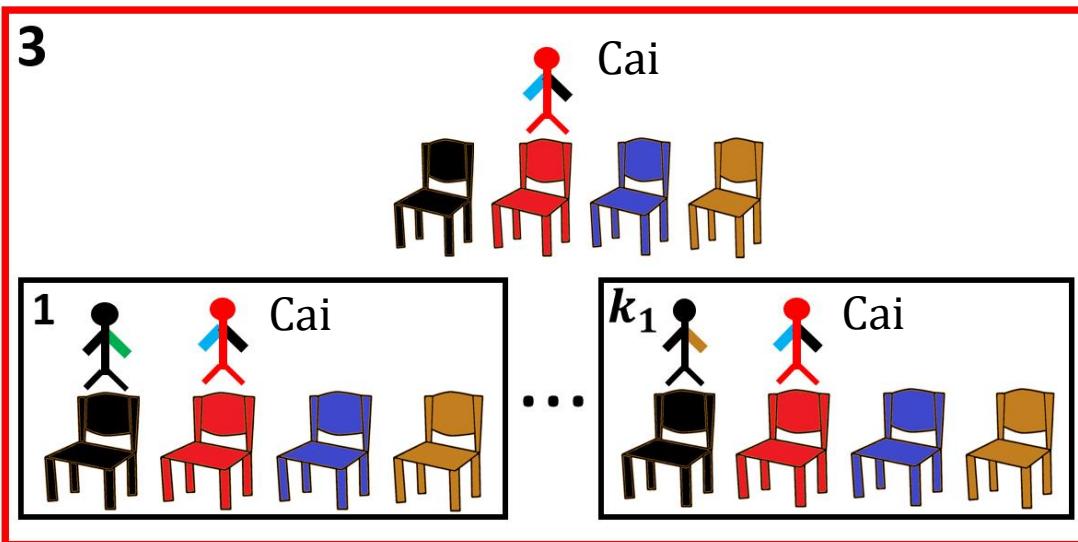
$\subset \Omega_{\text{Ahmed}}$

Can we use this inductive construction process to propagate information through this hierarchy ?

Can we use this inductive construction process to propagate information through this hierarchy ?

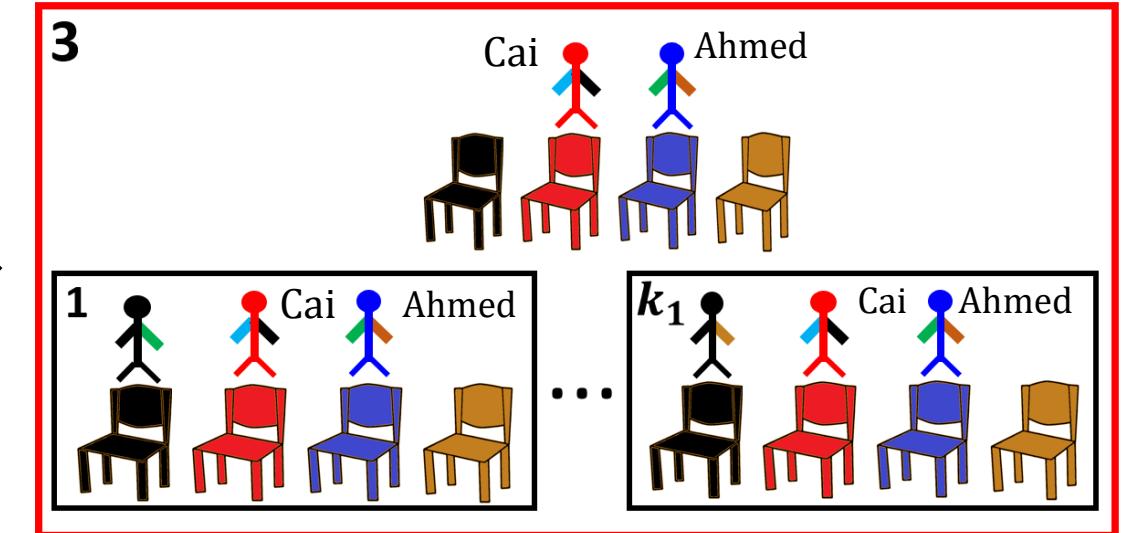
1 - Propagate information from exactly the previous layer (Direct neighbour handshaking possibility).

Ω_{Cai}



Information given: $Q(\Omega_{\text{Cai}})$

$\Omega_{\text{Cai} \rightarrow \text{Ahmed}}$

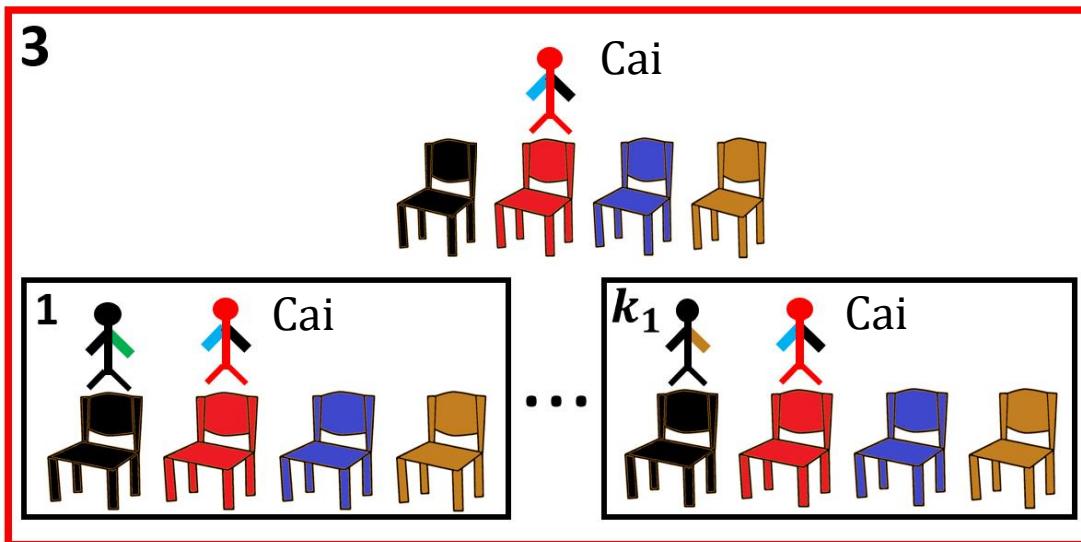


Information needed: $Q(\Omega_{\text{Cai} \leftrightarrow \text{Ahmed}})$

Can we use this inductive construction process to propagate information through this hierarchy ?

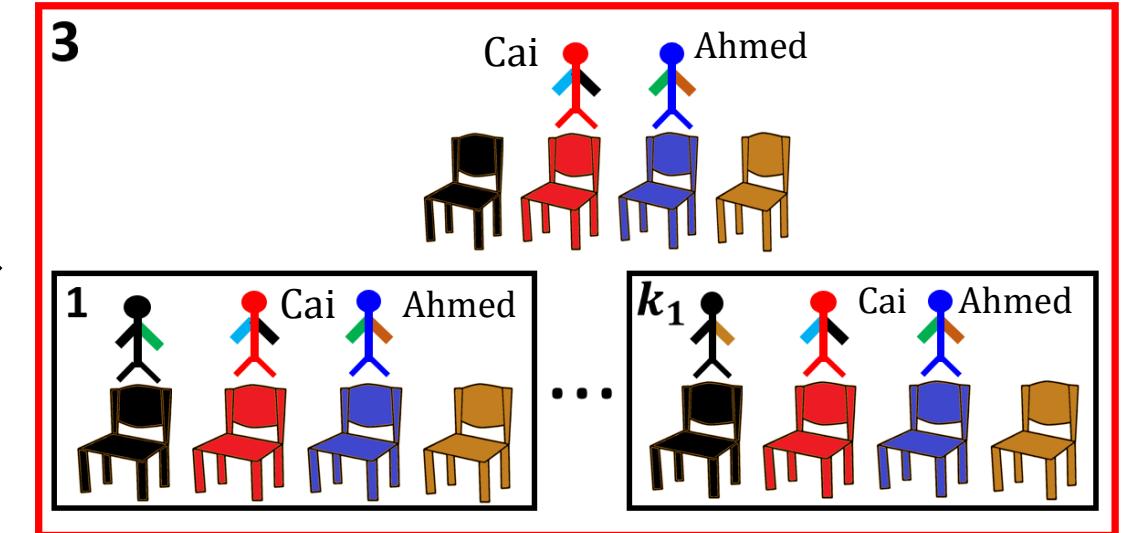
1 - Propagate information from exactly the previous layer (Direct neighbour handshaking possibility).

Ω_{Cai}



Information given: $Q(\Omega_{\text{Cai}})$

$\Omega_{\text{Cai} \rightarrow \text{Ahmed}}$

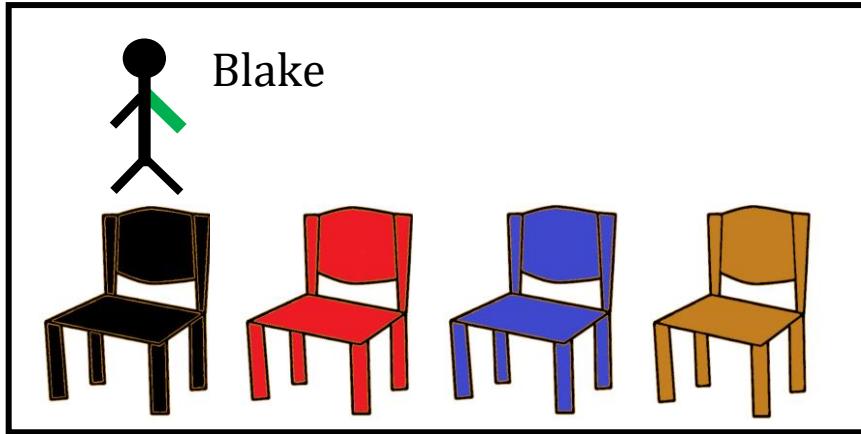


Information needed: $Q(\Omega_{\text{Cai} \leftarrow \text{Ahmed}})$

$$Q(\Omega_{\text{Cai} \leftarrow \text{Ahmed}}) = \dots = e^{\frac{\text{sit(Ahmed)} + \text{Sit_Cost}}{c}} * e^{\frac{\text{handshake(Cai,Ahmed)}}{c}} * Q(\Omega_{\text{Cai}})$$

2 - Propagate information from the rest (**No handshaking possibility**).

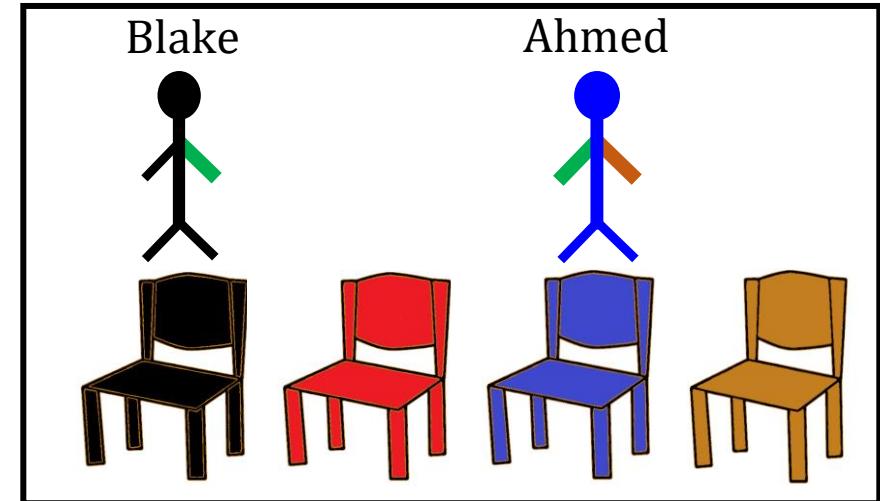
$$\Omega_{\text{Blake}}$$



Information given: $Q(\Omega_{\text{Blake}})$

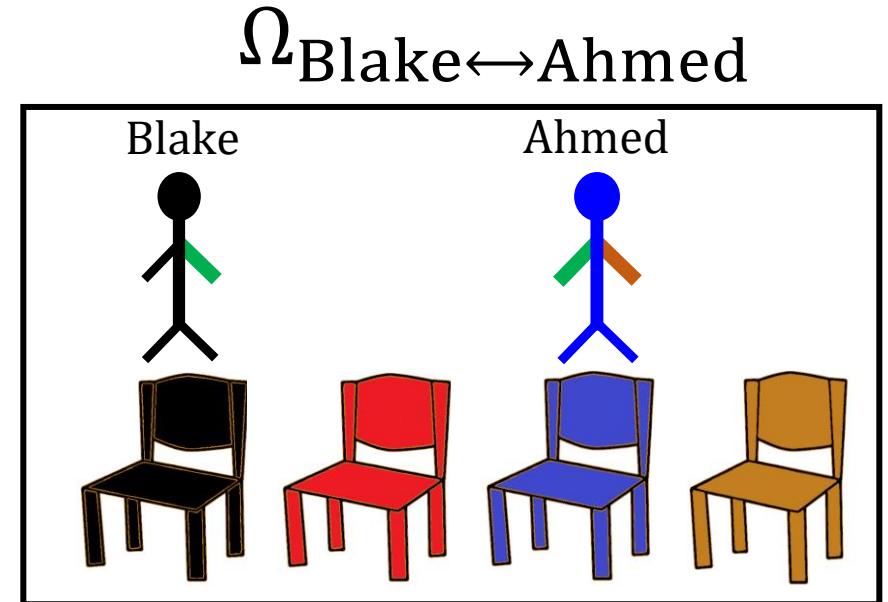
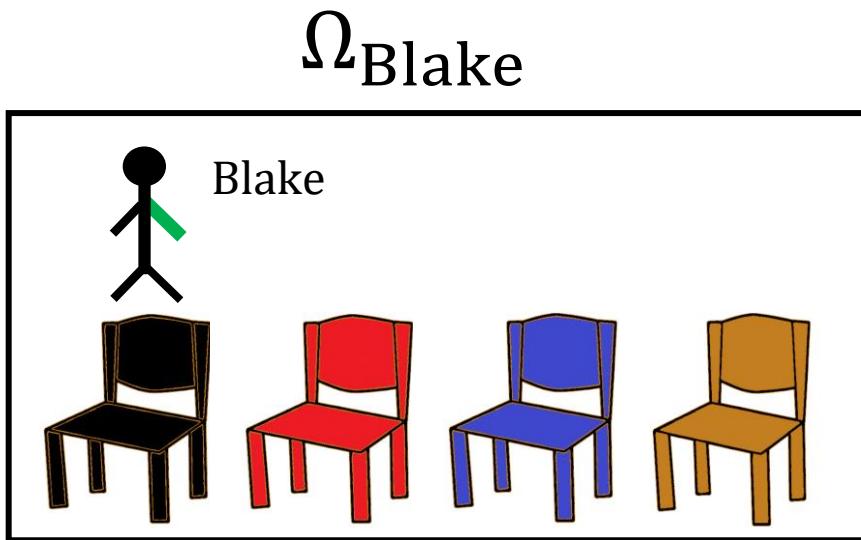


$$\Omega_{\text{Blake} \leftrightarrow \text{Ahmed}}$$



Information needed: $Q(\Omega_{\text{Blake} \rightarrow \text{Ahmed}})$

2 - Propagate information from the rest (**No handshaking possibility**).



$$Q(\Omega_{\text{Blake} \leftrightarrow \text{Ahmed}}) = \dots = e^{\frac{\text{sit(Ahmed)} + \text{sit_cost}}{c}} * Q(\Omega_{\text{Blake}})$$

$$Q(\Omega_p) = e^{\frac{sit(p)+sit_cost}{c}} * \left[\left(\sum_{\substack{p' \text{ is} \\ \text{direct neighbour}}} e^{\frac{handshake(p',p)}{c}} * Q(\Omega_{p'}) \right) + 1 + \sum_{\substack{p' \text{ is not} \\ \text{direct neighbour}}} Q(\Omega_{p'}) \right]$$

$$Q = 1 + \sum_p Q(\Omega_p)$$

$$Q(\Omega_p) = e^{\frac{sit(p)+sit_cost}{c}} * \left[\left(\sum_{\substack{p' \text{ is} \\ \text{direct neighbour}}} e^{\frac{handshake(p',p)}{c}} * Q(\Omega_{p'}) \right) + 1 + \sum_{\substack{p' \text{ is not} \\ \text{direct neighbour}}} Q(\Omega_{p'}) \right]$$

► **Theorem 2.** There is an $O(k^2N)$ time algorithm for the domain-level partition function for a 1D SDC of length N with $\leq k$ computation strands competing at each scaffold domain.

$$Q(\Omega_s) = e^{\frac{-\Delta G(M(s)) + \Delta G^{\text{assoc}}}{k_B T}} * \left[\sum_{s' \in LD_s} \left(e^{\frac{-\Delta G(R(s'), L(s))}{k_B T}} * Q(\Omega_{s'}) \right) + 1 + \sum_{s' \not\in LD_s} Q(\Omega_{p'}) \right]$$

$$Q = 1 + \sum_{s \in T} Q(\Omega_s)$$

■ **Algorithm 2** 1D SDC partition function algorithm. The proof of Theorem 2 argues that this algorithm returns Z^S as defined in Equation (6). Note that arrays are indexed from 1, and recall that k_1, \dots, k_N are the counts of competing strands at scaffold domains d_1, \dots, d_N , and we let s_i^j be the j^{th} strand competing at domain d_i . See Figure 9.

```

1:  $Z_{\text{curr}} = [0, 0, \dots, 0]$            ▷ size  $k = \max(k_1, \dots, k_N)$ , current (partial) partition function
2:  $Z_{\text{prev}} = [0, 0, \dots, 0]$            ▷ size  $k = \max(k_1, \dots, k_N)$ , previous (partial) partition function
3:  $Z^S \leftarrow 1$ ;  $\text{sum}_a \leftarrow 0$ 
4: for  $i \leftarrow 1 \dots N$  do
5:    $\text{sum}_a \leftarrow \text{sum}_a + \sum_{i \in \{1, \dots, k\}} Z_{\text{prev}}[i]$            ▷  $\text{sum}_a$ : rightmost summation Equation (7)
6:    $Z_{\text{prev}} \leftarrow Z_{\text{curr}}$ 
7:    $Z_{\text{curr}} = [0, 0, \dots, 0]$ 
8:   for  $j \leftarrow 1 \dots k_i$  do           ▷ each iteration computes Equation (7) for a strand
9:      $t_1 = e^{-(\Delta G(d^M(s_i^j)) + \Delta G^{\text{assoc}})/k_B T}$ 
10:    if  $i = 1$  then           ▷ first domain where is no neighbors at all
11:       $Z_{\text{curr}}[j] = t_1$ 
12:    else
13:       $t_2 \leftarrow 0$ 
14:      for  $m \leftarrow 1 \dots k_{i-1}$  do
15:         $t_2 \leftarrow t_2 + \left( e^{-(\Delta G(d^R(s_{i-1}^m), d^L(s_i^j))) / k_B T} \right) \cdot Z_{\text{prev}}[m]$ 
16:      end for
17:       $Z_{\text{curr}}[j] \leftarrow t_1 + t_2 + \text{sum}_a$ 
18:    end if
19:     $Z^S \leftarrow Z^S + Z_{\text{curr}}[j]$            ▷ computing Equation (6)
20:  end for
21: end for
22: return  $Z^S$ 

```

► **Theorem 1.** There is an $O(k^2N)$ time algorithm to determine the domain-level MFE for a 1D SDC of length N with $\leq k$ computation strands competing at each scaffold domain.

$$M^S = \min_{s \in C_{d_N}} \{M_s^S\}$$

$$M_s^S = \Delta G(d^M(s)) + \Delta G^{\text{assoc}} + \min_{s' \in L_s} \{M_{s'}^S + \Delta G(d^R(s'), d^L(s))\}$$

■ **Algorithm 1** 1D SDC MFE algorithm. The proof of Theorem 1 proof shows that this algorithm returns the value M^S defined in Equation (4). Note that arrays are indexed from 1. Notation: k_1, \dots, k_N are the counts of competing strands at scaffold domains d_1, \dots, d_N . Let s_i^j be the j^{th} strand competing at domain d_i .

```

1:  $M_{\text{curr}} = [0, 0, \dots, 0]$                                 ▷ size  $k = \max(k_1, \dots, k_N)$  for current MFEs
2:  $M_{\text{prev}} = [0, 0, \dots, 0]$                                 ▷ size  $k = \max(k_1, \dots, k_N)$  for previous MFEs
3: for  $i \leftarrow 1 \dots N$  do                                         ▷ index scaffold domains
4:    $M_{\text{prev}} \leftarrow M_{\text{curr}}$ 
5:   for  $j \leftarrow 1 \dots k_i$  do                                     ▷ index computational strands at scaffold domain  $d_i$ 
6:     if  $i = 1$  then                                              ▷ first scaffold domain, has no left neighbour
7:        $M_{\text{curr}}[j] \leftarrow \Delta G(d^M(s_i^j))$ 
8:     else                                                               ▷  $O(k)$  steps to choose min and bind scaffold + entropic penalty
9:        $M_{\text{curr}}[j] \leftarrow [\min_{m \in \{1, 2, \dots, k_{i-1}\}} (M_{\text{prev}}[m] + \Delta G(d^R(s_{i-1}^m), d^L(s_i^j)))$ 
10:      +  $\Delta G(d^M(s_i^j)) + \Delta G^{\text{assoc}}]$ 
11:     end if
12:   end for
13: end for
14:  $M^S \leftarrow \min_{k' \in \{1, 2, \dots, k_N\}} M_{\text{curr}}[k']$           ▷  $O(k)$  steps implement Equation (4) giving  $M^S$ 
15: return  $M^S$ 

```

GOAL

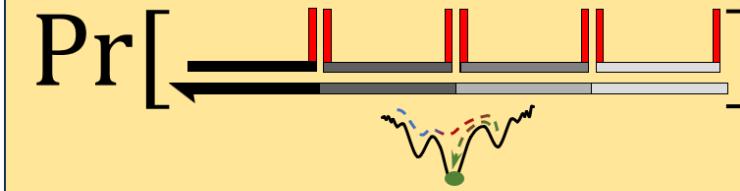
$$\Pr[\text{[Diagram of a system with red vertical bars and a wavy line with a green dot at its minimum]}] \gg \sum_c \Pr[c : \text{is another configuration}]$$

At equilibrium

Efficiently

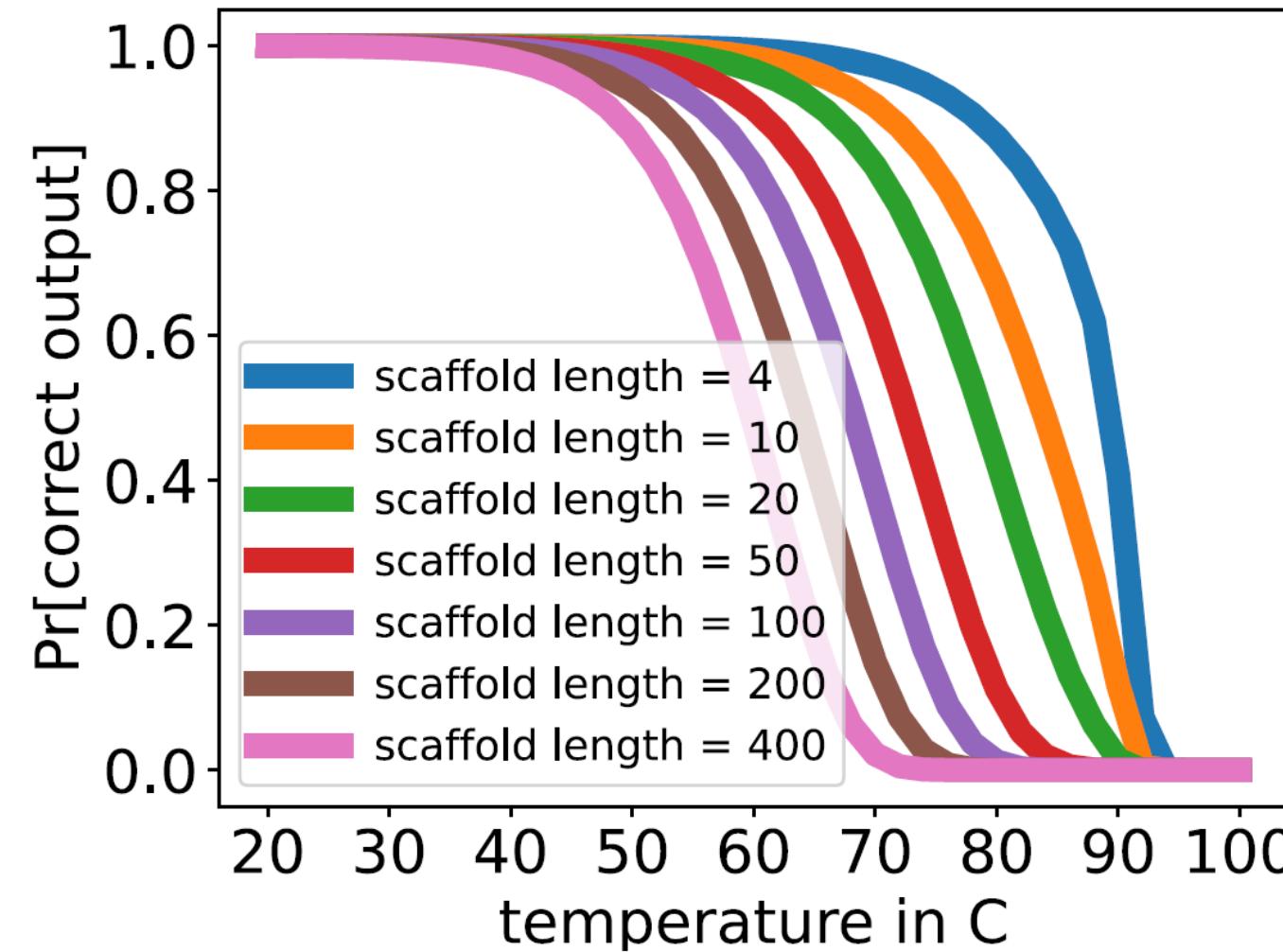


GOAL



At equilibrium

Efficiently



Benefits of Domain Based models !!

Previous: 1 simulation of length 13
Now: 280 simulations in 20 min [800 strands]

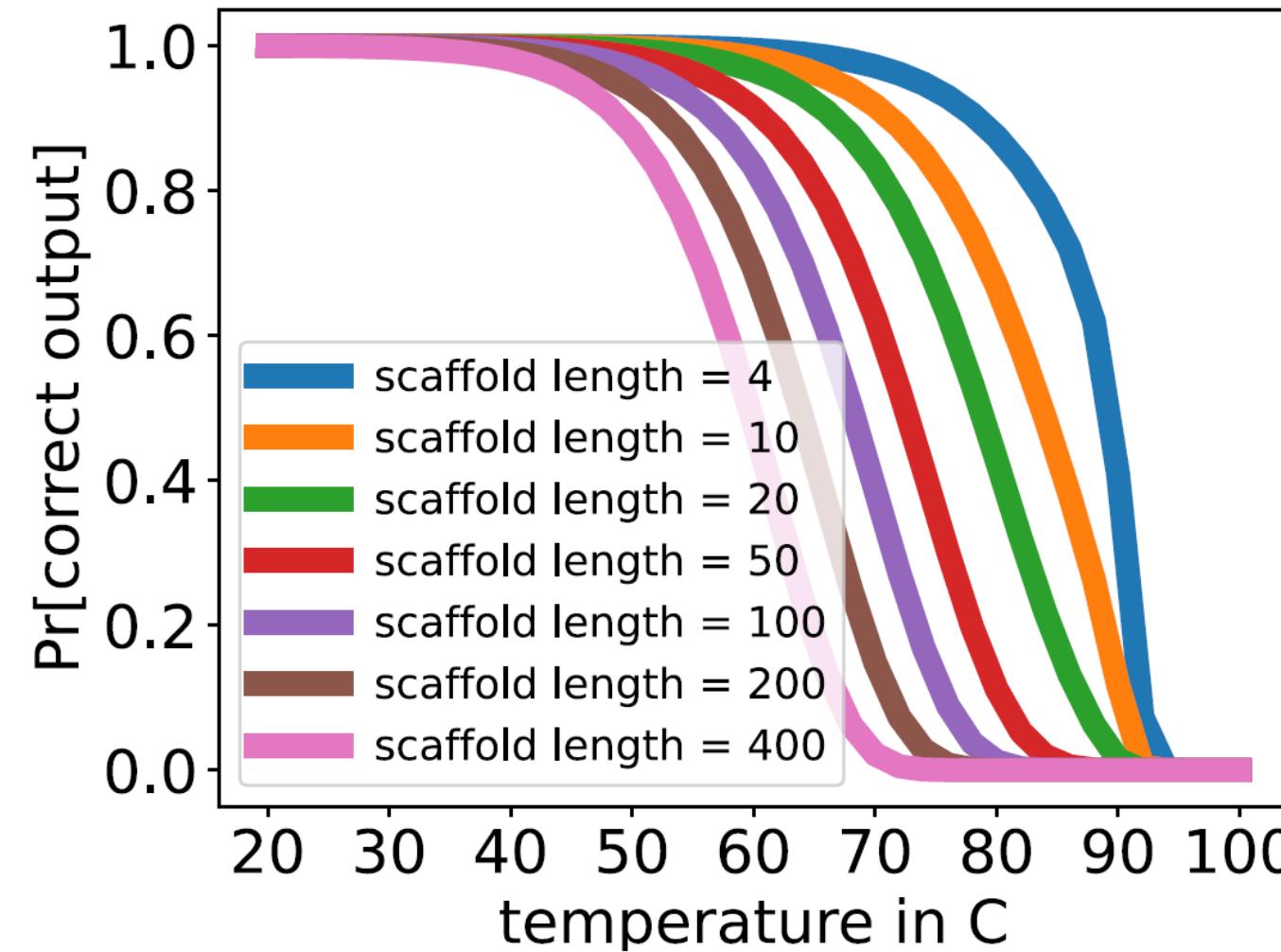
GOAL



$$\Pr[\text{correct configuration}] \gg \sum_c \Pr[c : \text{is another configuration}]$$

At equilibrium

Efficiently



Benefits of Domain Based models !!

Previous: 1 simulation of length 13
Now: 280 simulations in 20 min [800 strands]

Thanks



Hamilton Institute



**Maynooth
University**
National University
of Ireland Maynooth

dna.hamilton.ie

We're hiring!
Postdoc, PhD



European
Innovation
Council



Funded by
the European Union