Effective Attacks and Provable Defenses for Website Fingerprinting

Tao Wang
Ian Goldberg

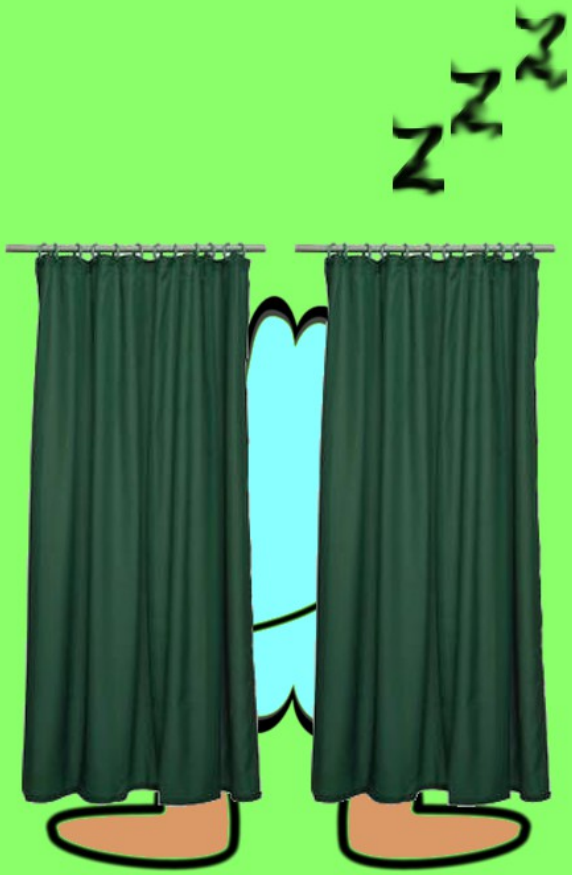
University of Waterloo

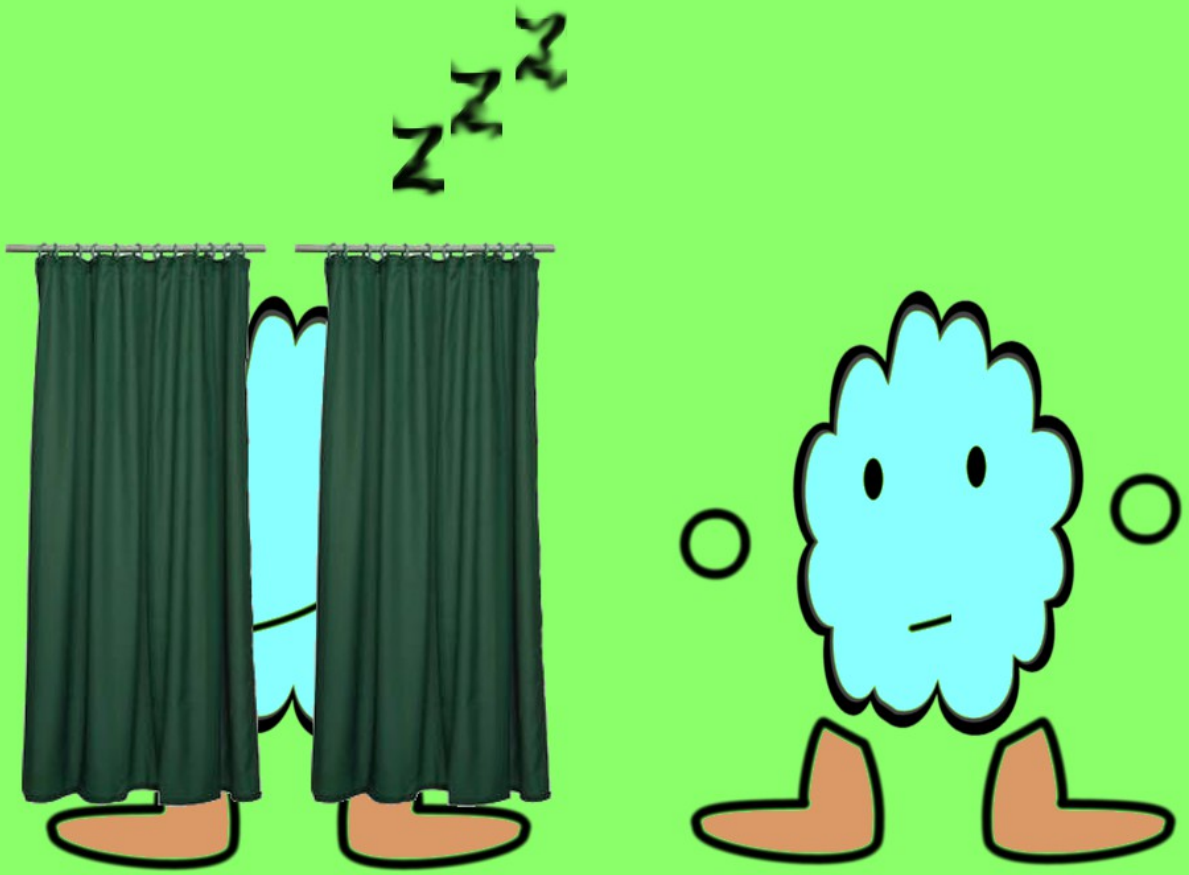


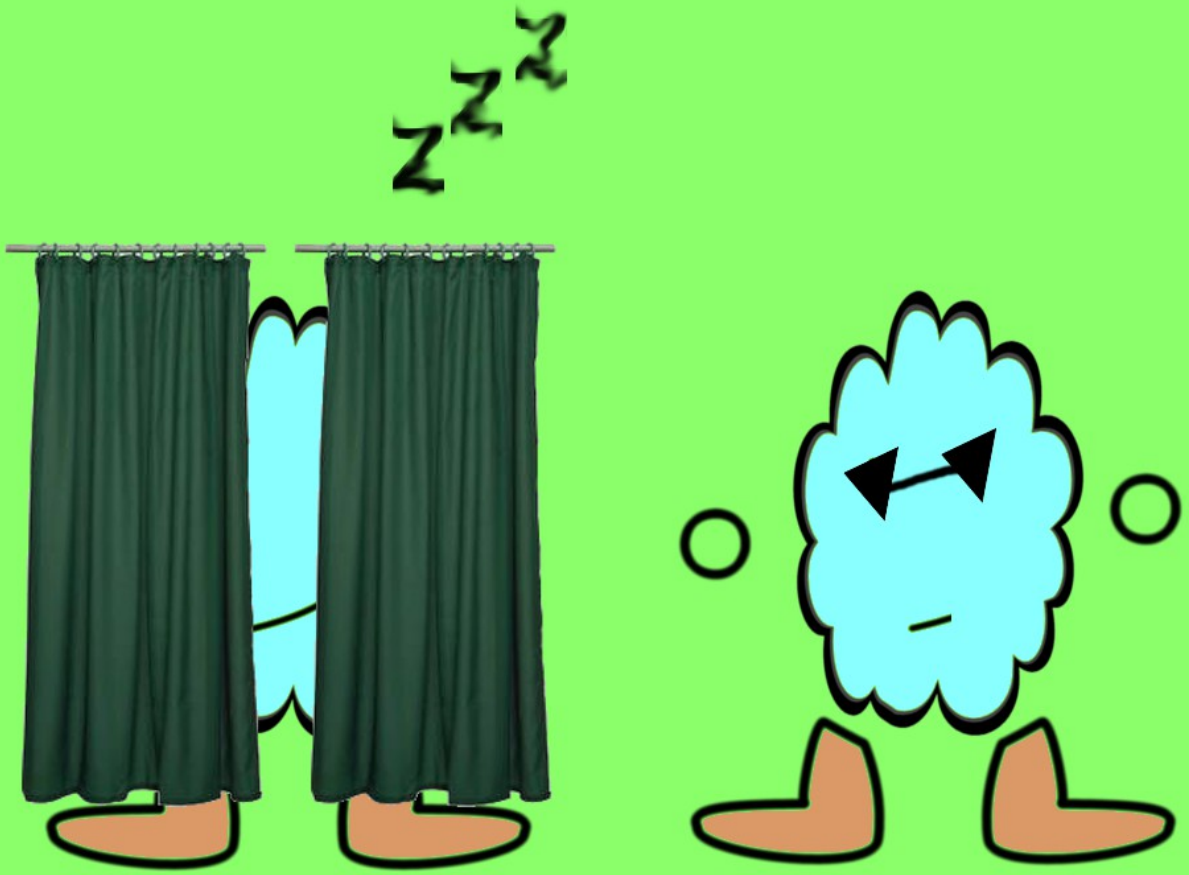
Xiang Cai
Rishab Nithyanand
Rob Johnson

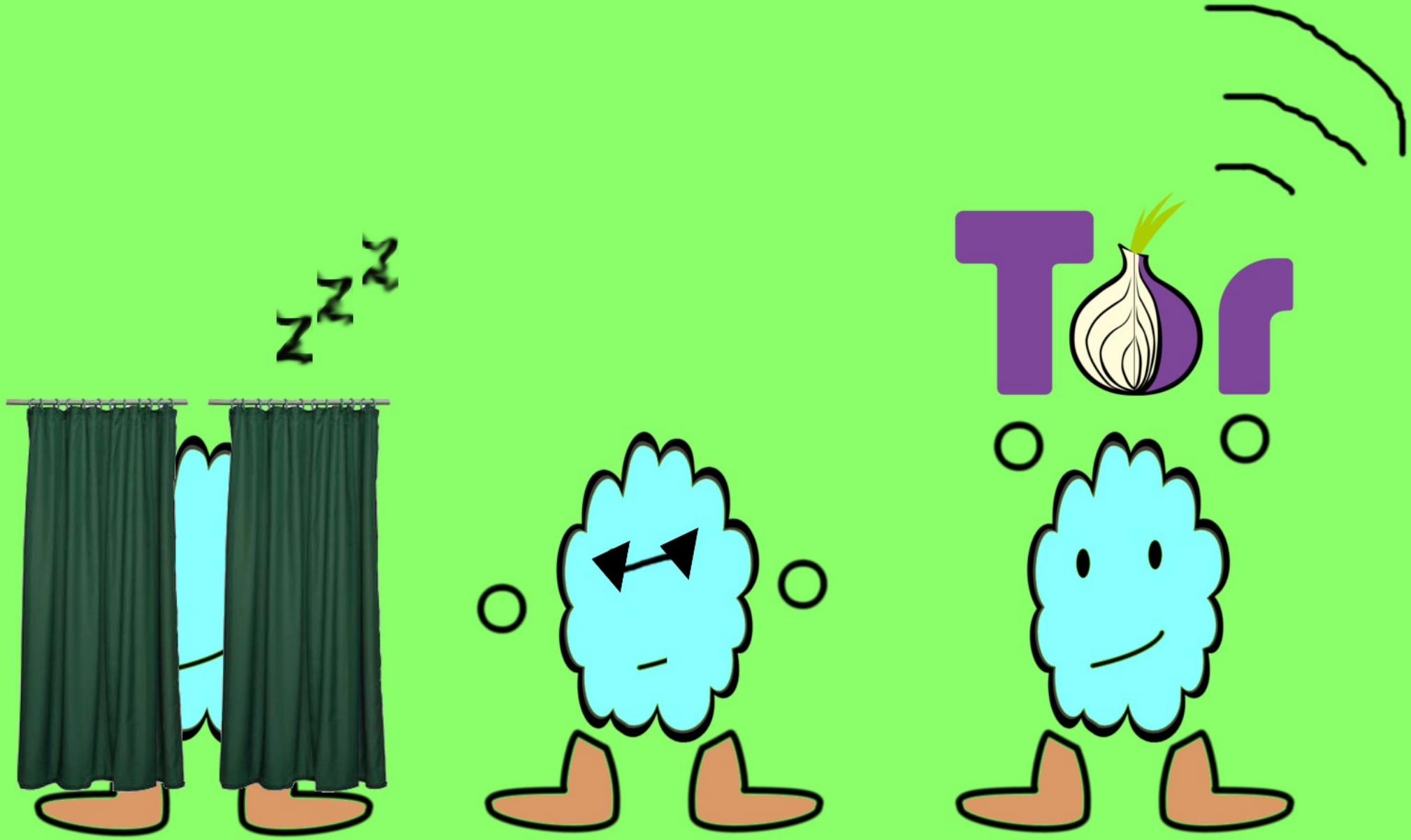
Stony Brook University





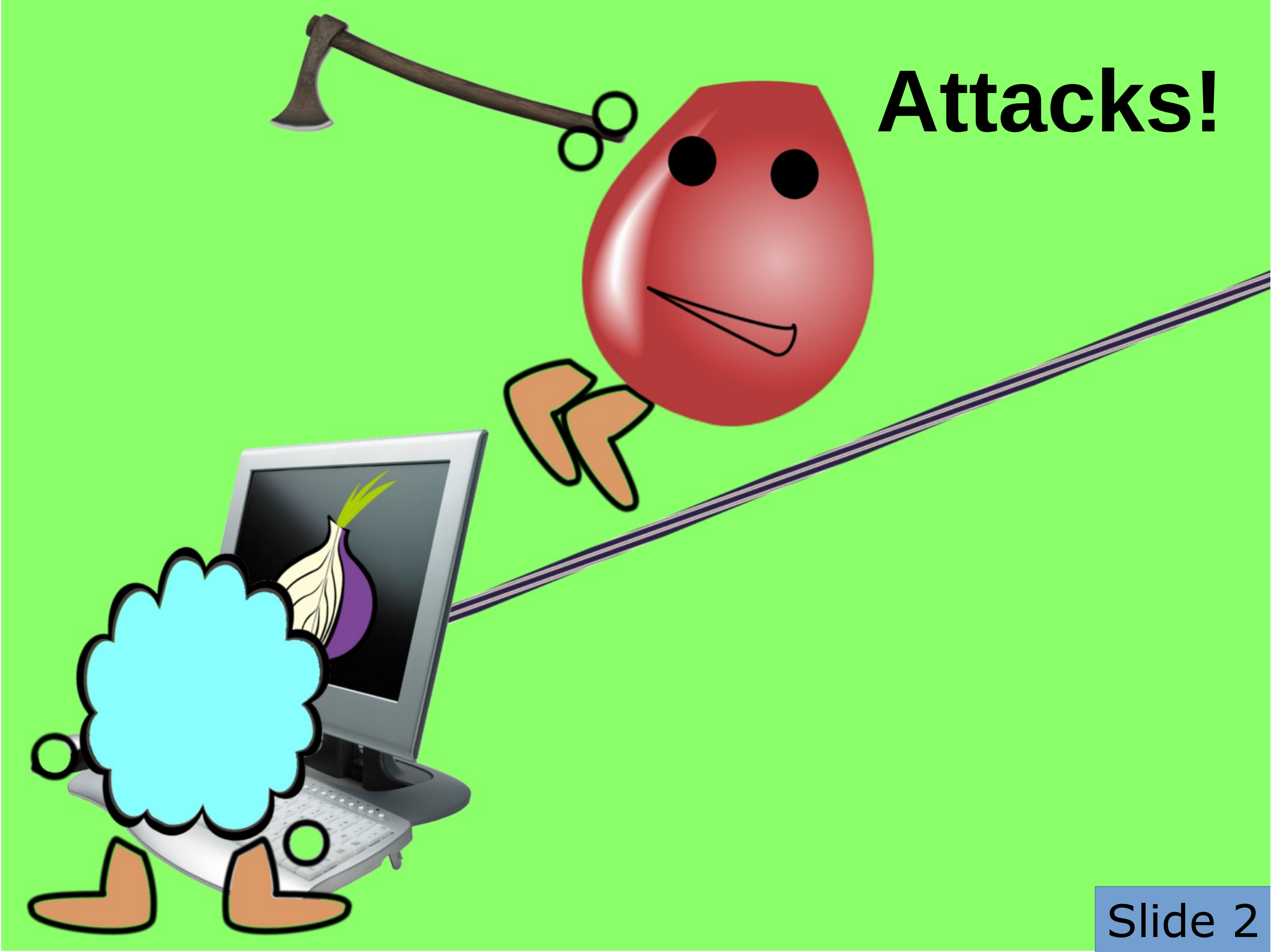





Tor



Attacks!



Effective Attacks and Provable Defenses for Website Fingerprinting

Tao Wang
Ian Goldberg

University of Waterloo



Xiang Cai
Rishab Nithyanand
Rob Johnson

Stony Brook University

Effective Attacks and Provable Defenses for Website Fingerprinting

Tao Wang
Ian Goldberg

University of Waterloo

Xiang Cai
Rishab Nithyanand
Rob Johnson

Stony Brook University





Effective Attacks and Provable Defenses for Website Fingerprinting



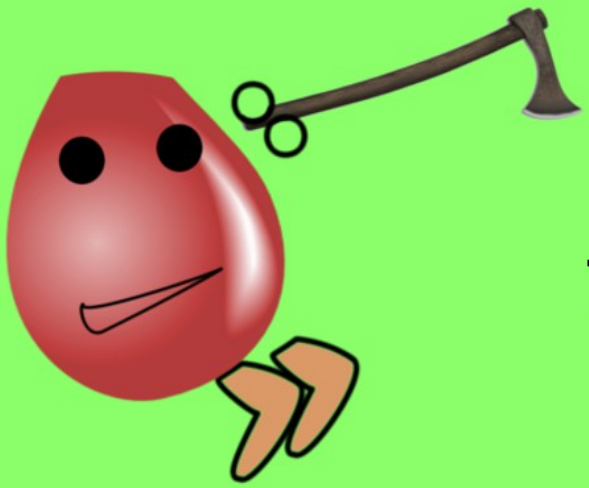
Tao Wang
Ian Goldberg

University of Waterloo

Xiang Cai
Rishab Nithyanand
Rob Johnson

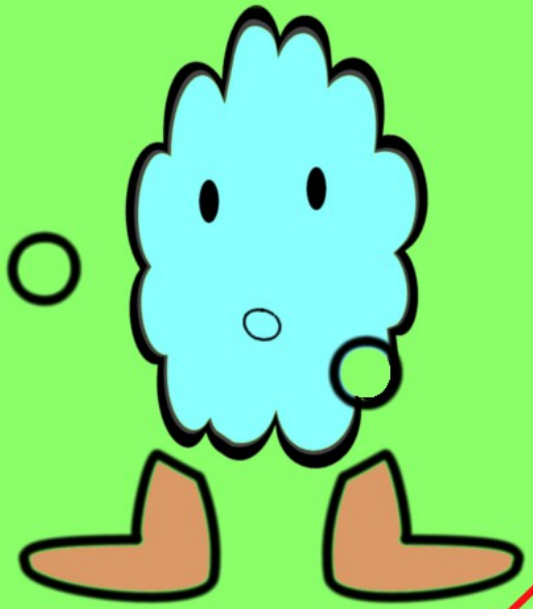
Stony Brook University

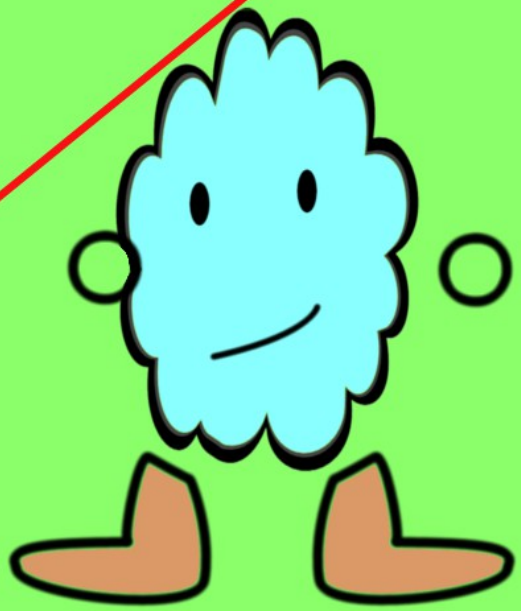


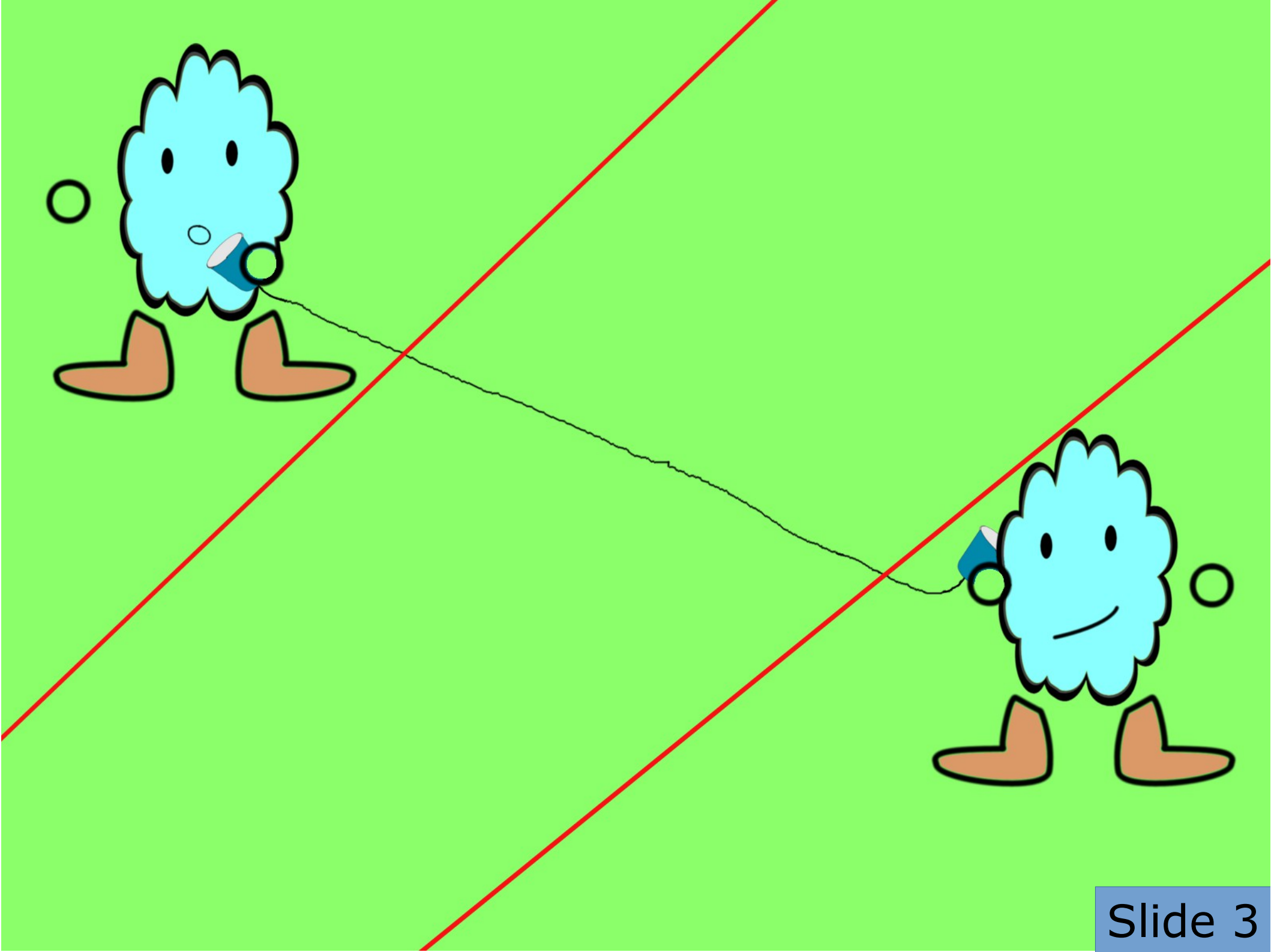




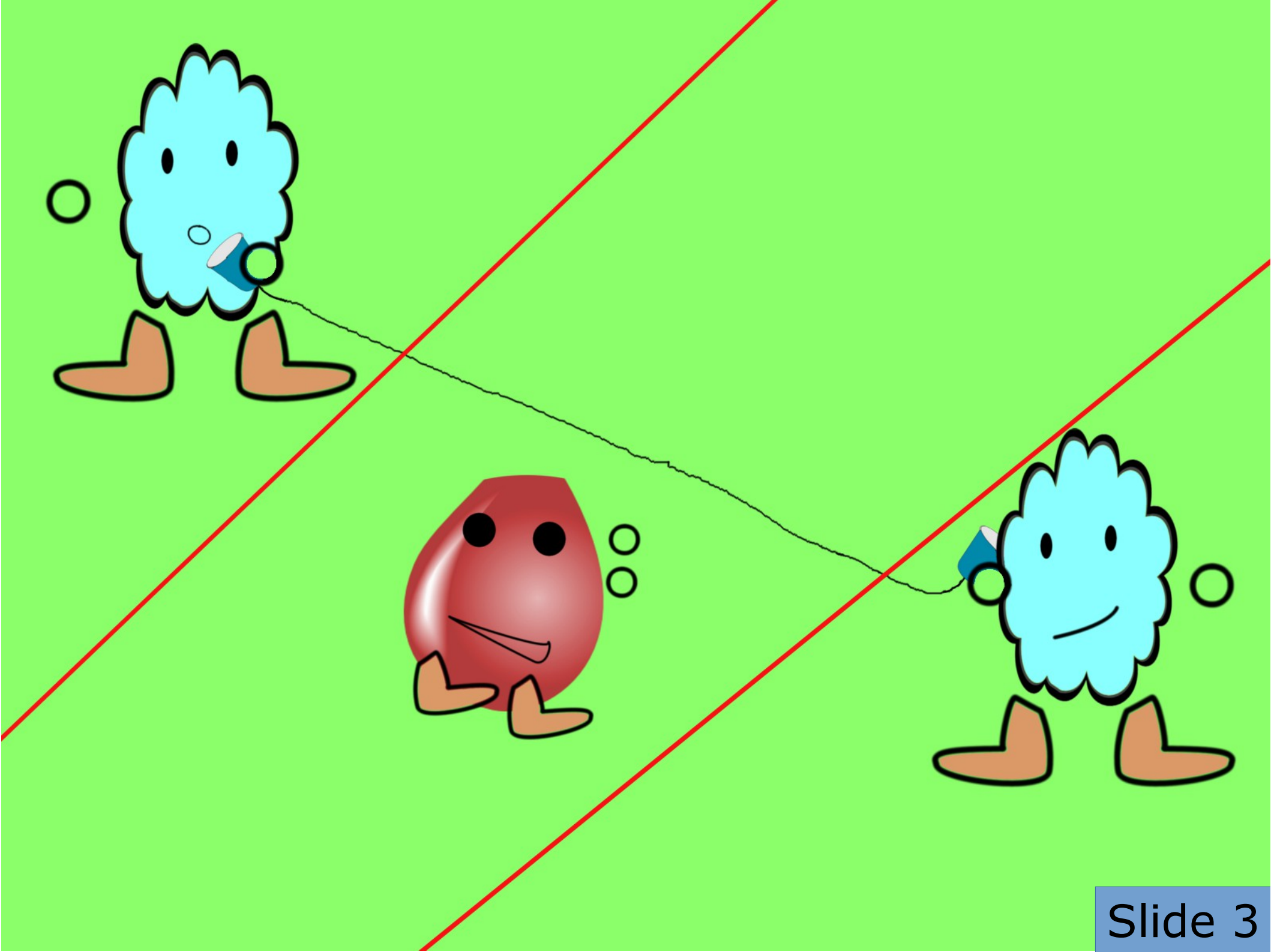
This guy



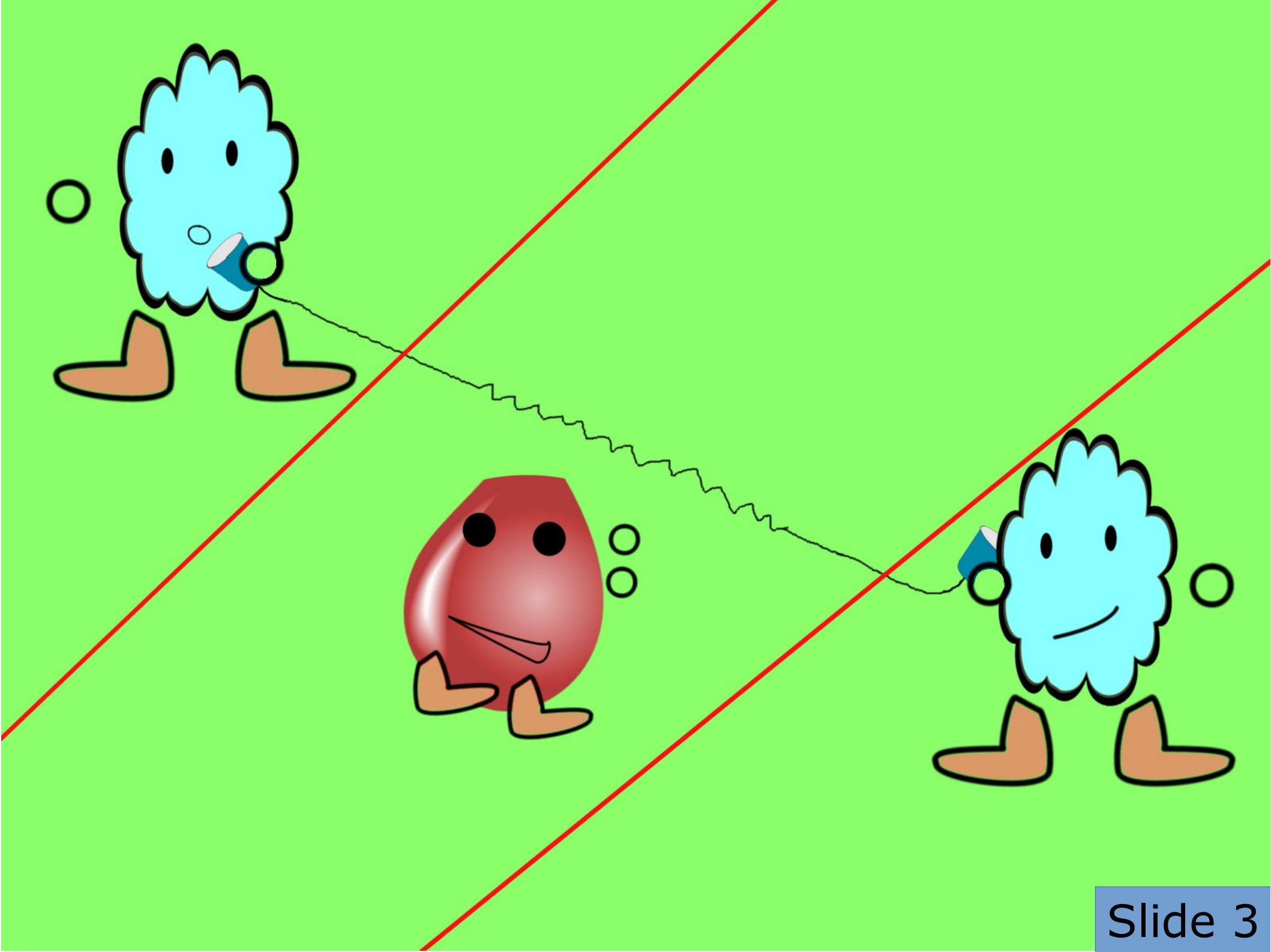




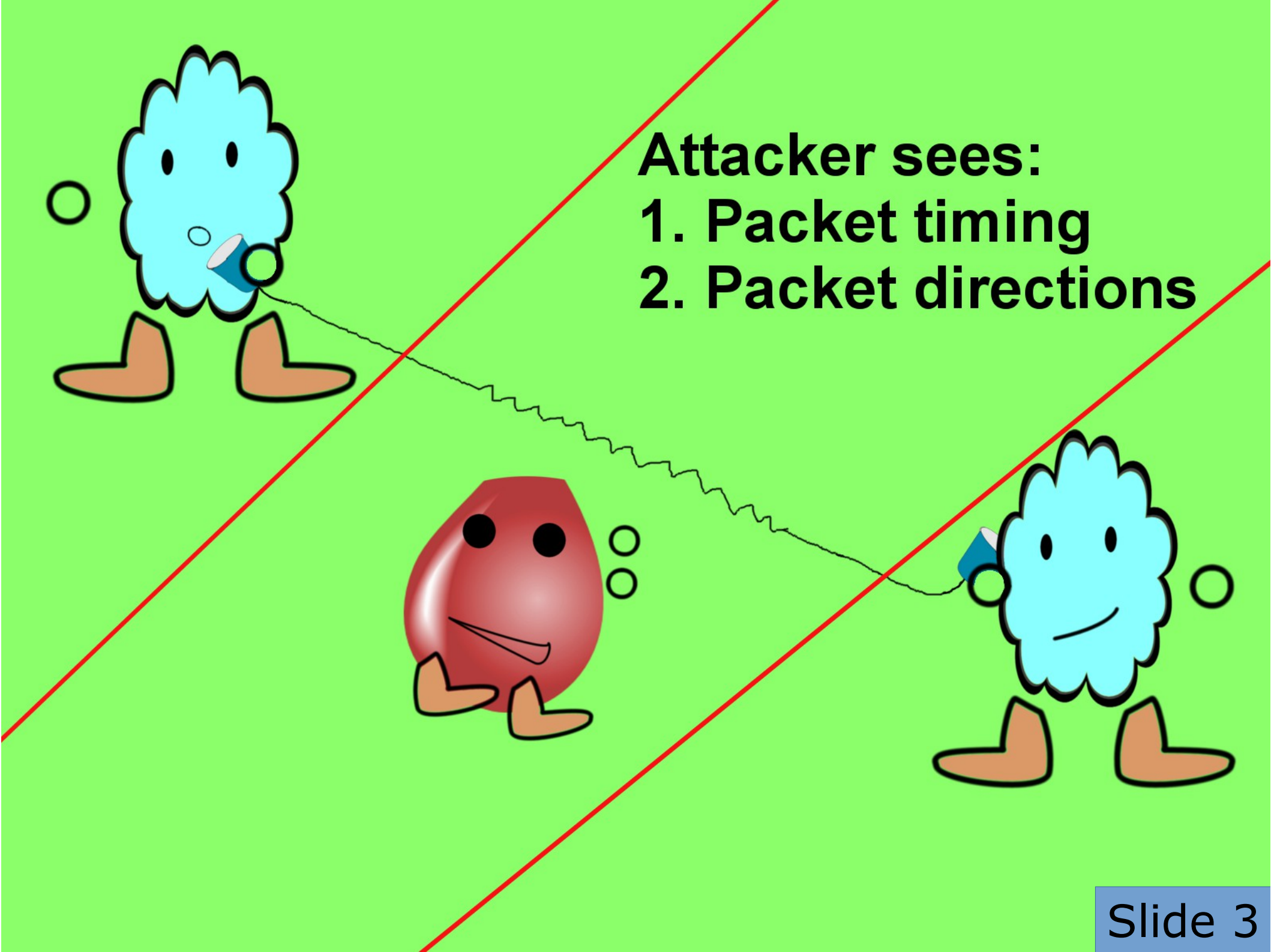
Slide 3



Slide 3



Slide 3


Attacker sees:
1. Packet timing
2. Packet directions

Our attack



Speed



Our attack



Accuracy



Open World





Our attack



Open World

???

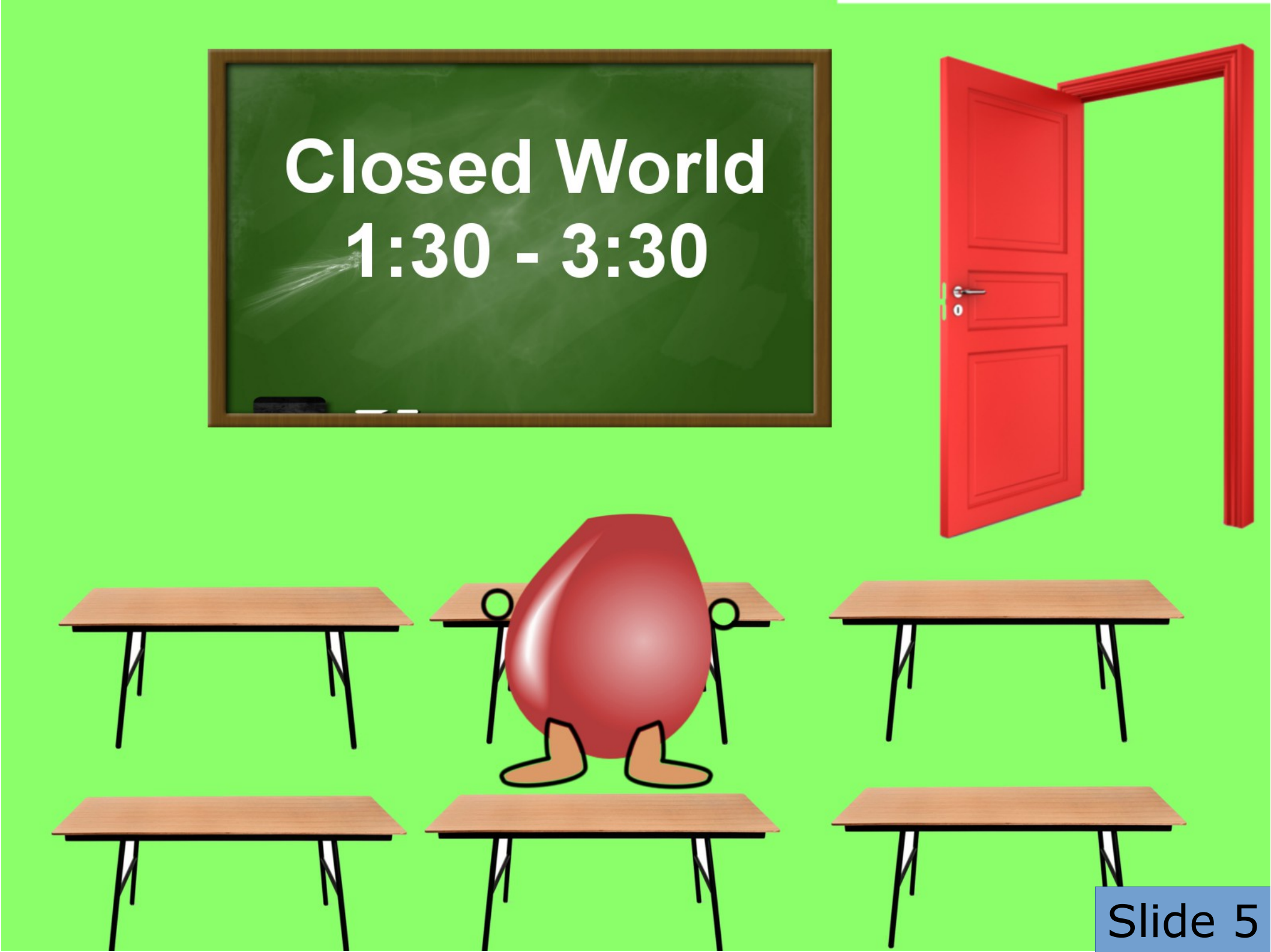


???

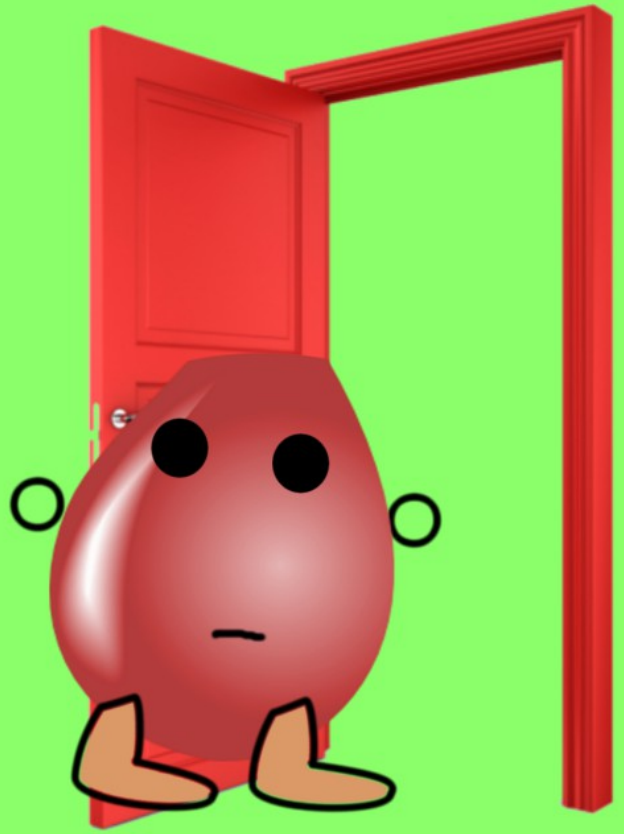


???

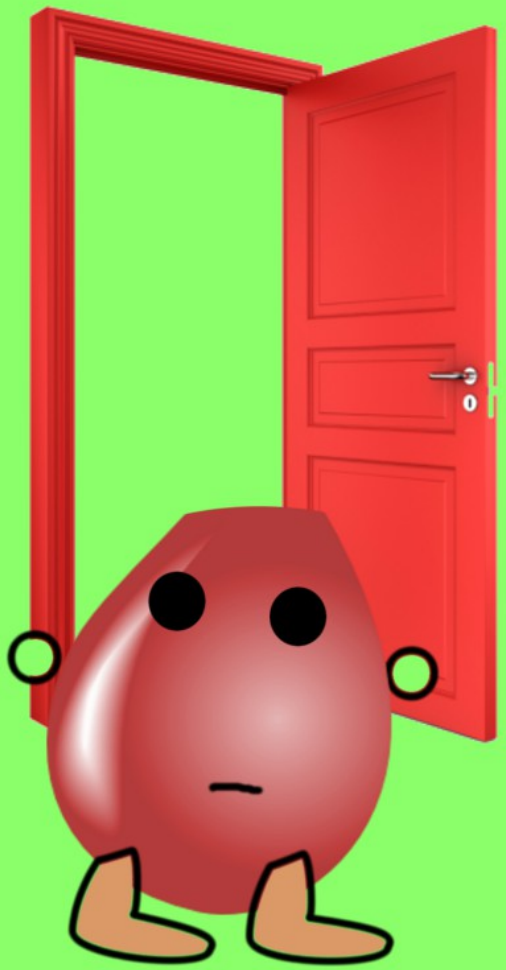



Closed World
1:30 - 3:30

Closed World
1:30 - 3:30



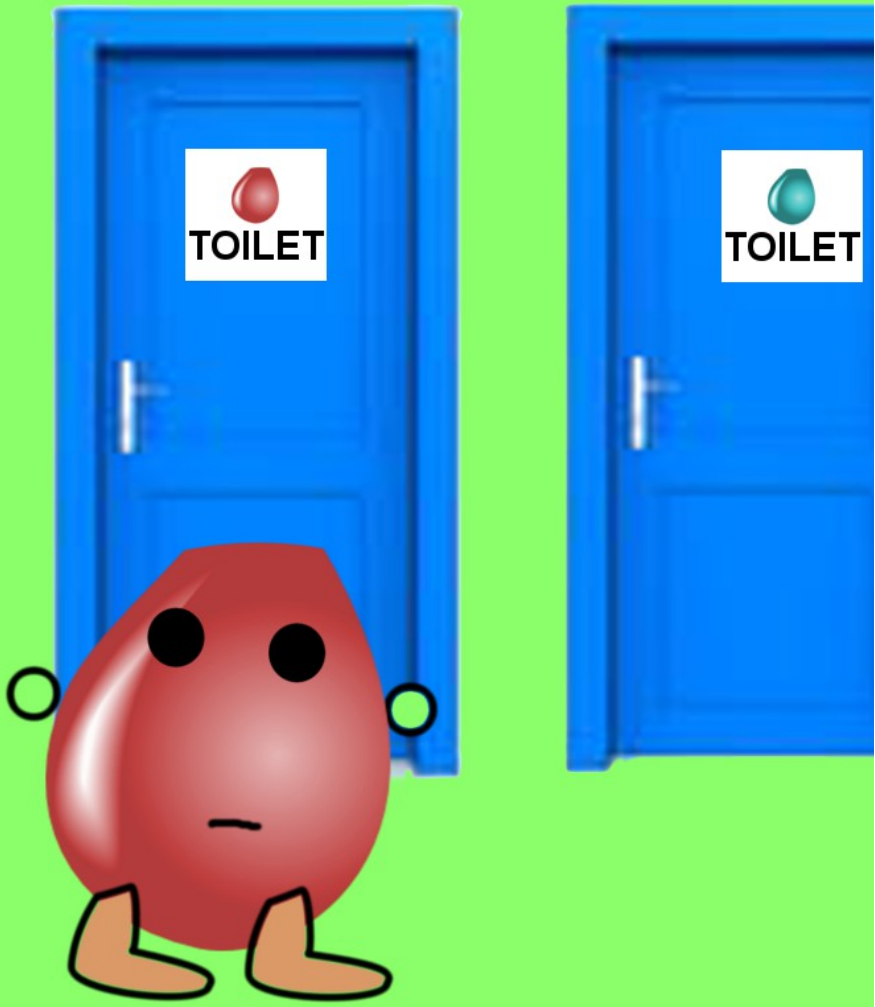





TOILET


TOILET

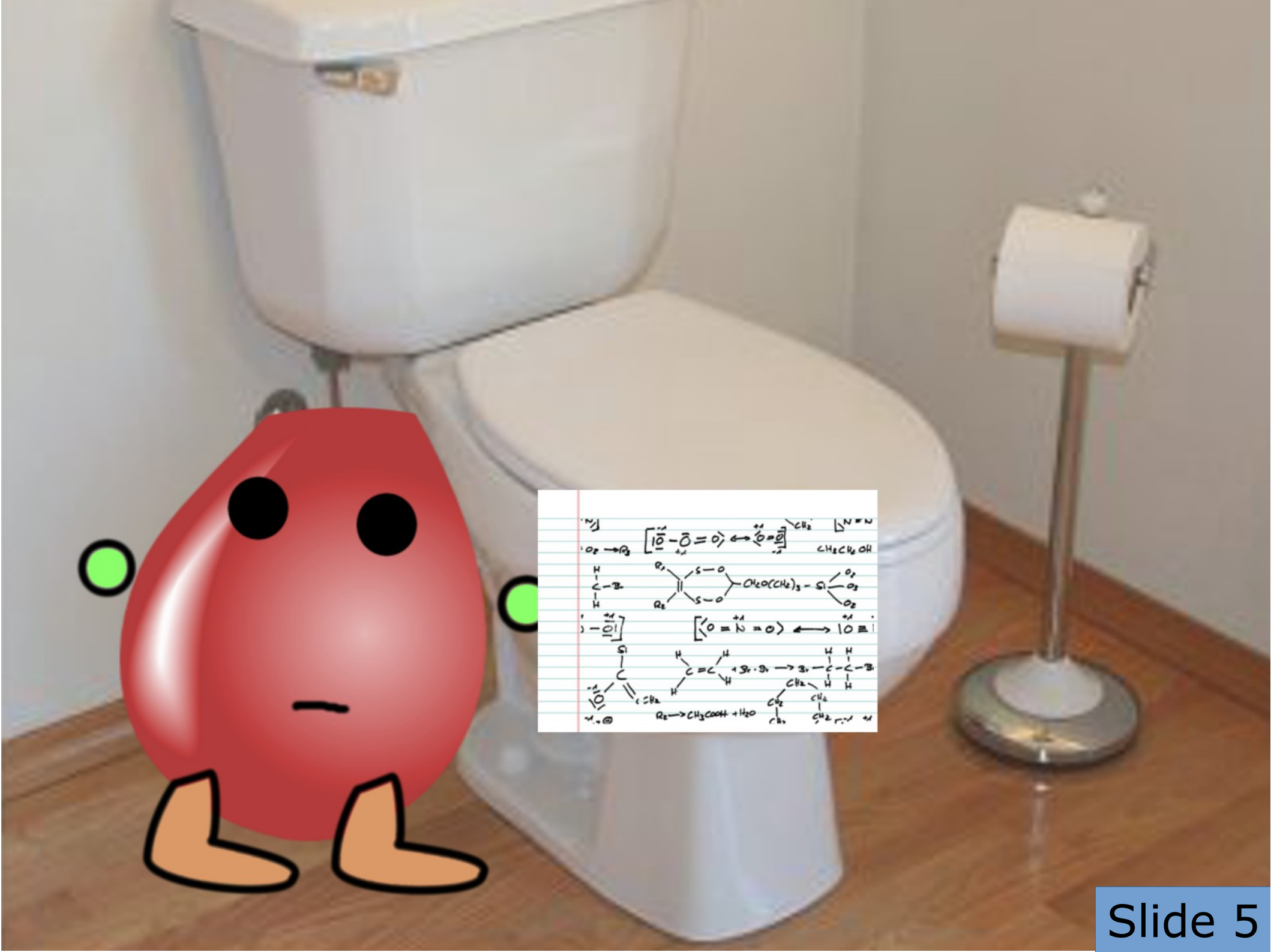



TOILET
TOILET


Slide 5

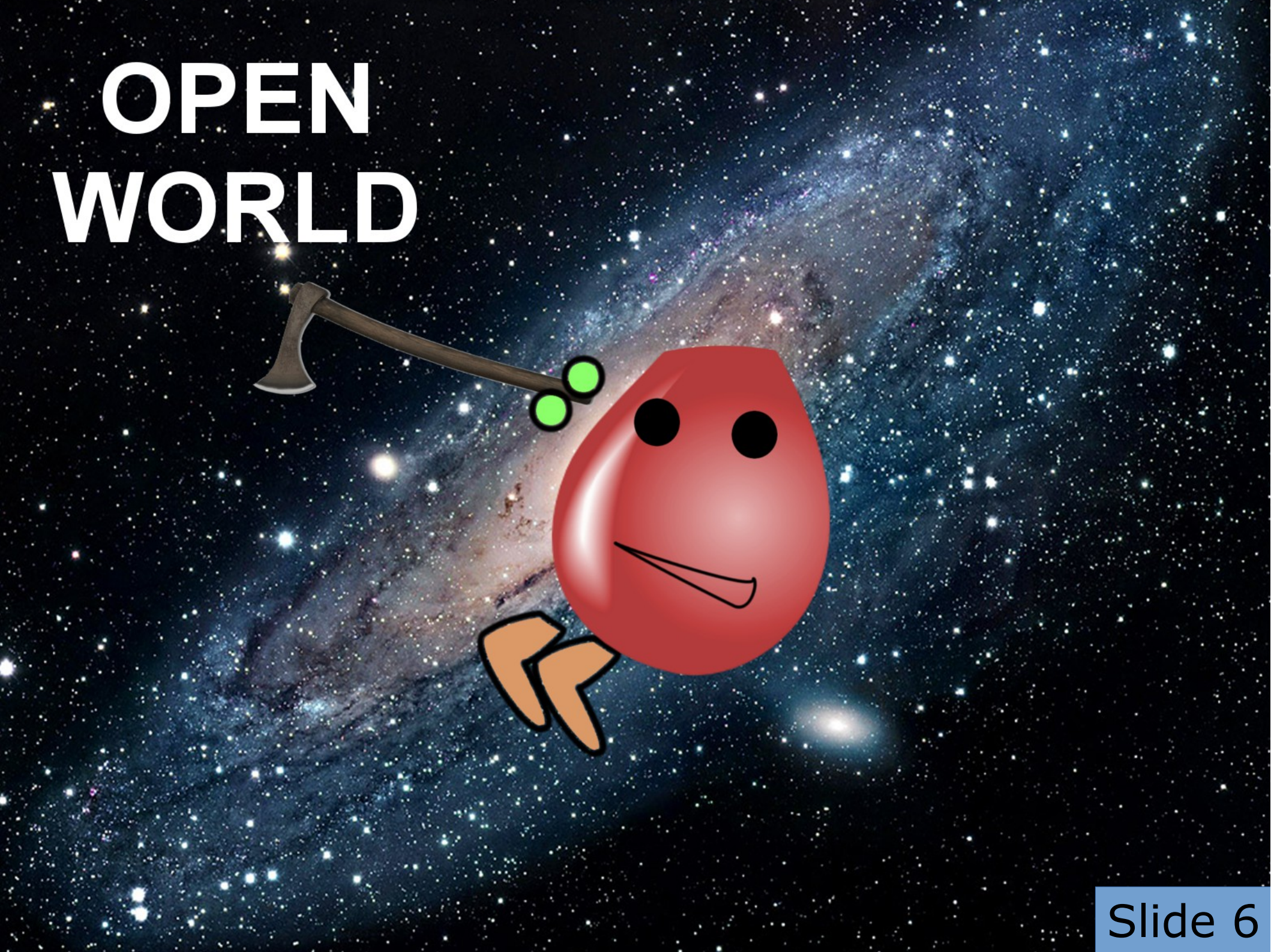

Dig
Dig


Slide 5



Slide 5

OPEN WORLD



CLOSED WORLD

reddit

Monitored Sites









YouTube

Non-Monitored Sites

Gmail by Google

OPEN WORLD

reddit

Monitored Sites









YouTube

Non-Monitored Sites

Gmail by Google

OPEN WORLD

reddit

Monitored Sites







Classifier Answers:
1. Twitter
2. Facebook
3. Others

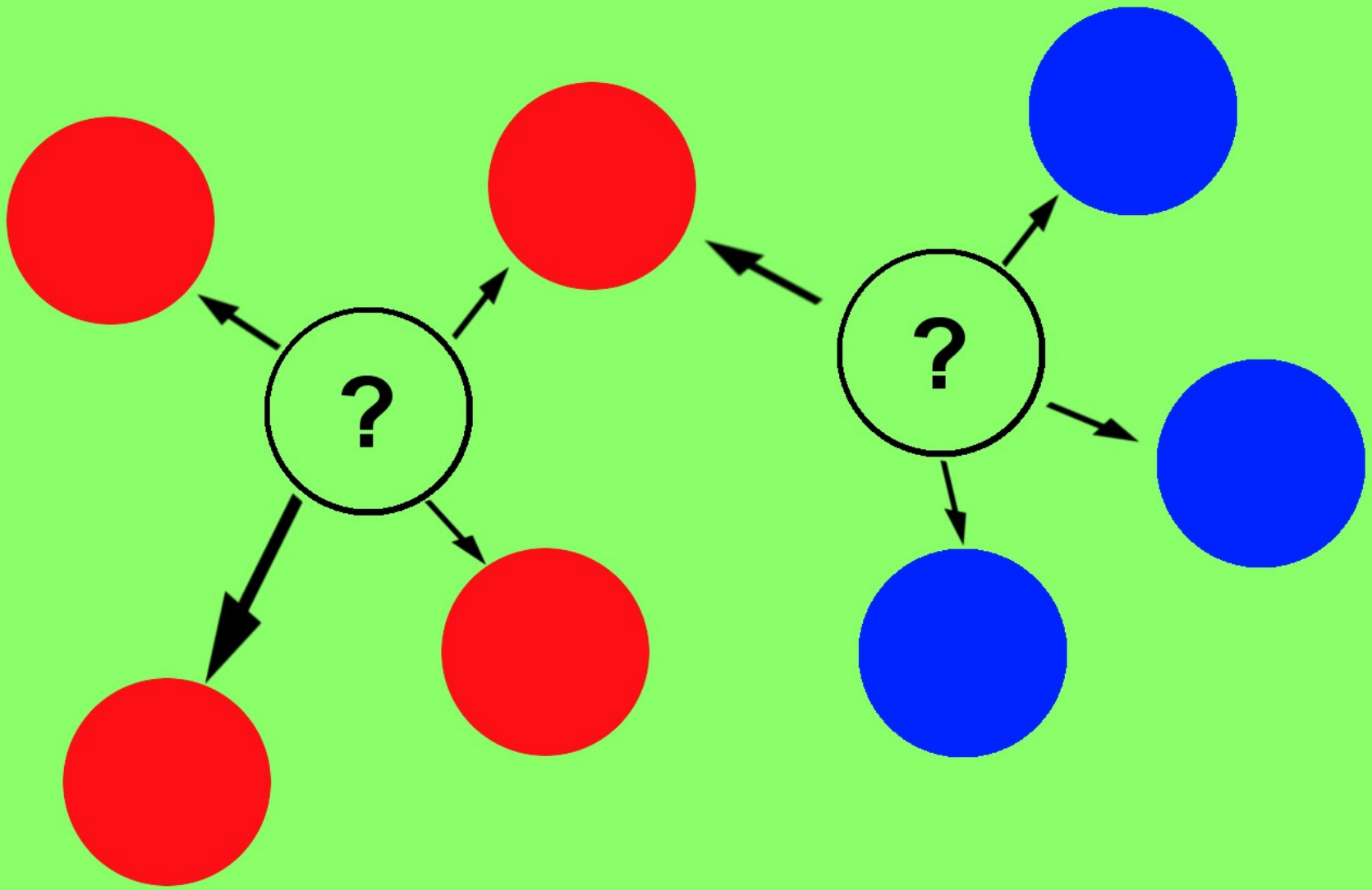


YouTube

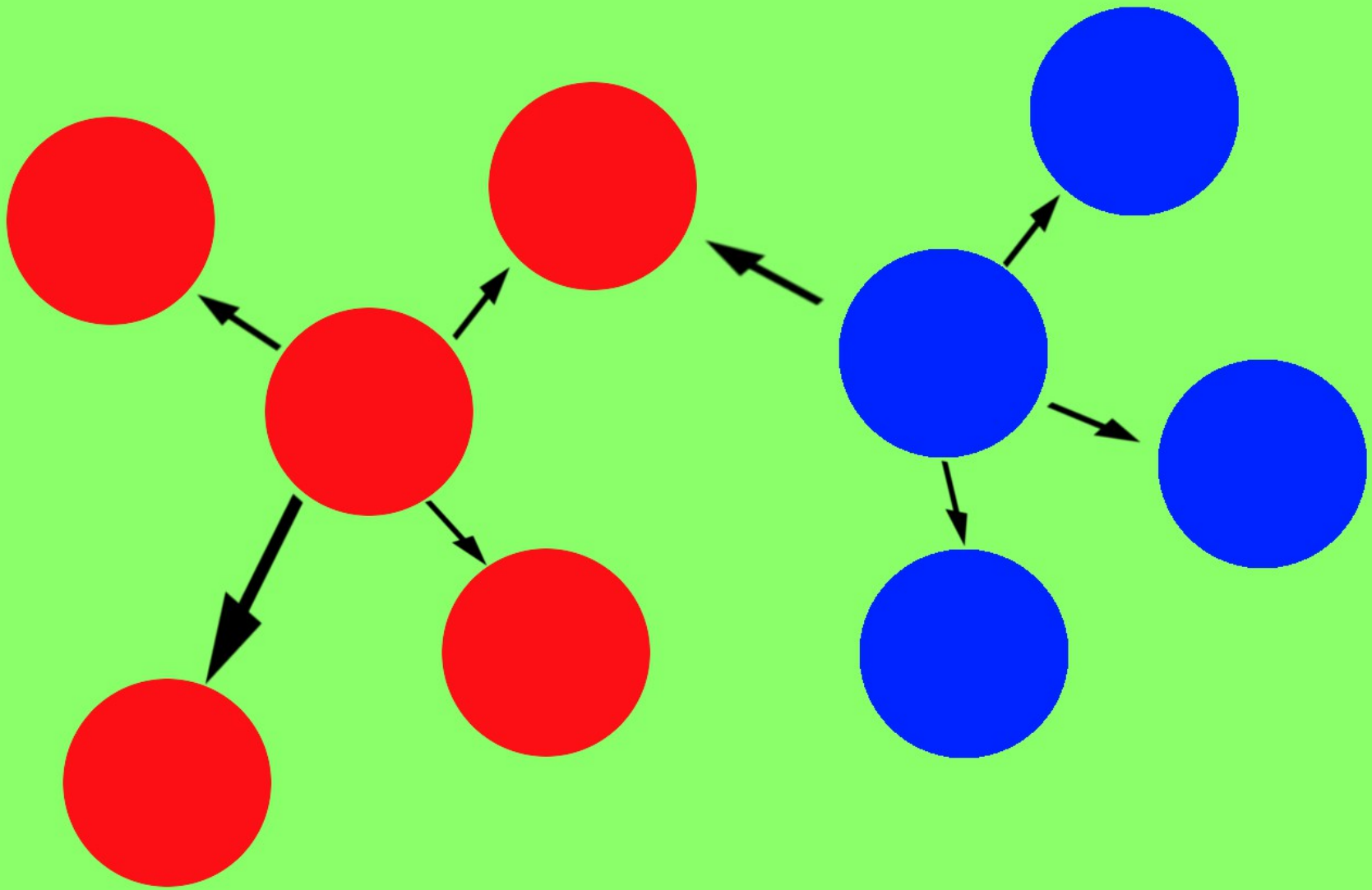
Non-Monitored Sites

Gmail by Google

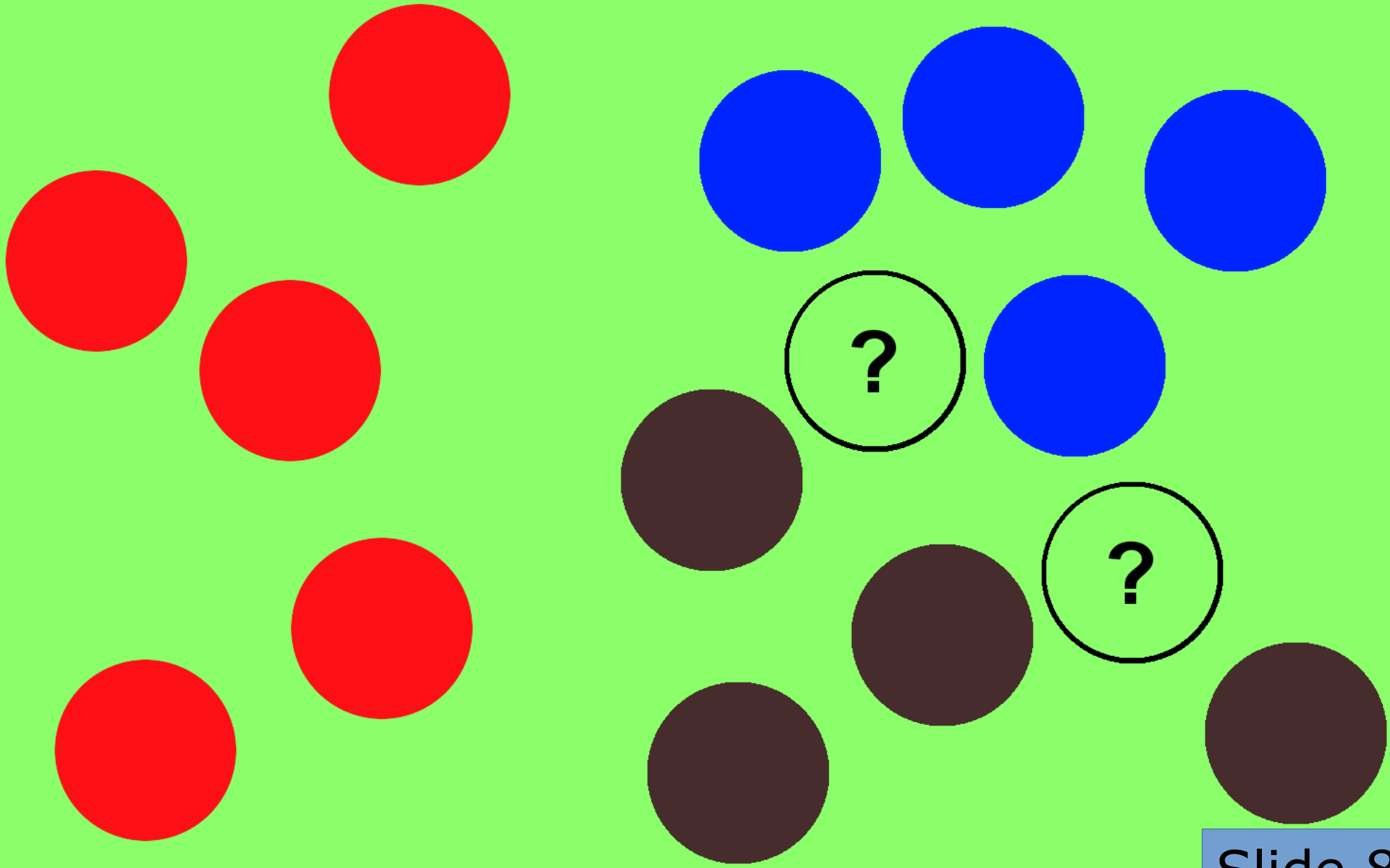
k-Nearest Neighbours



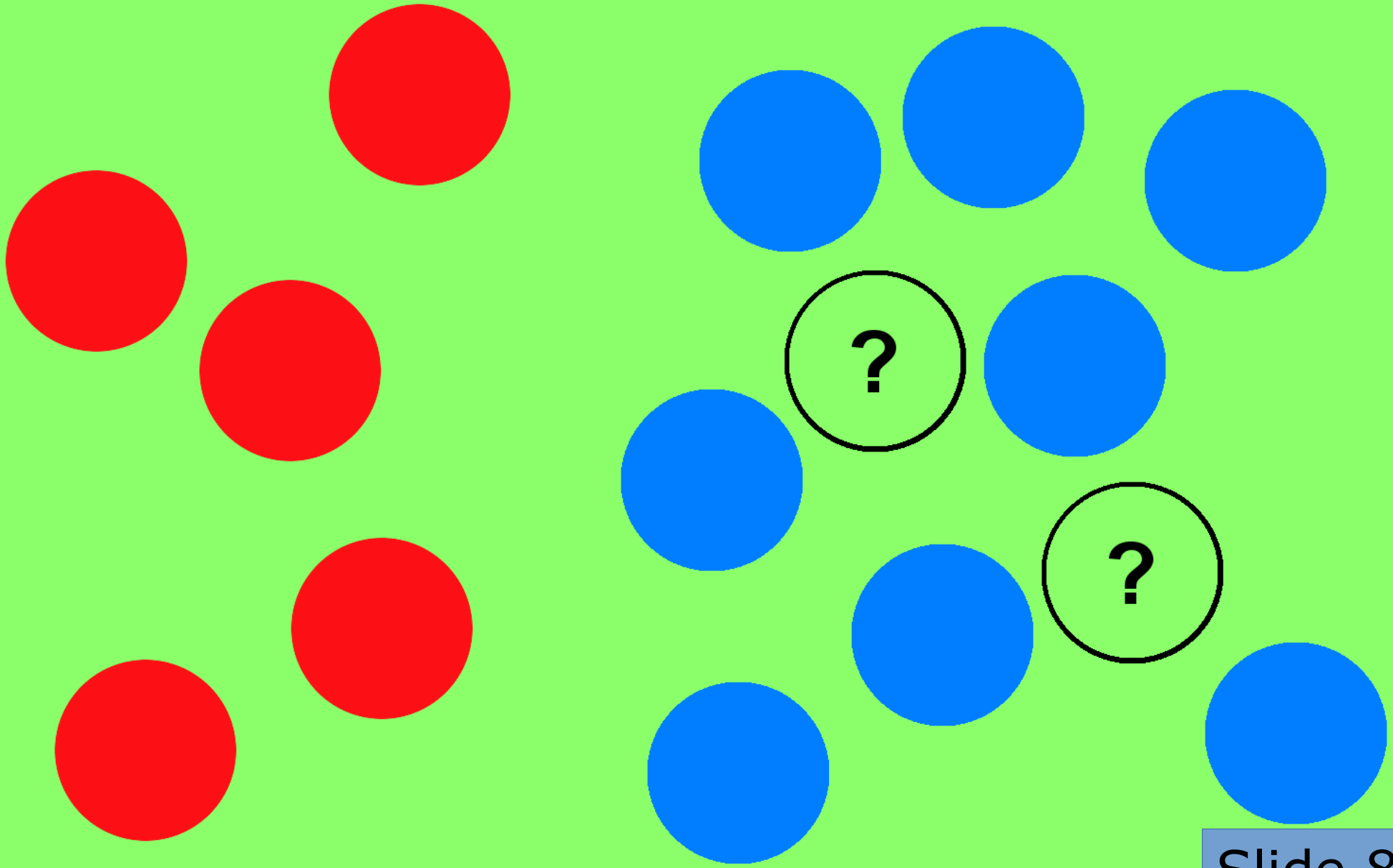
k-Nearest Neighbours



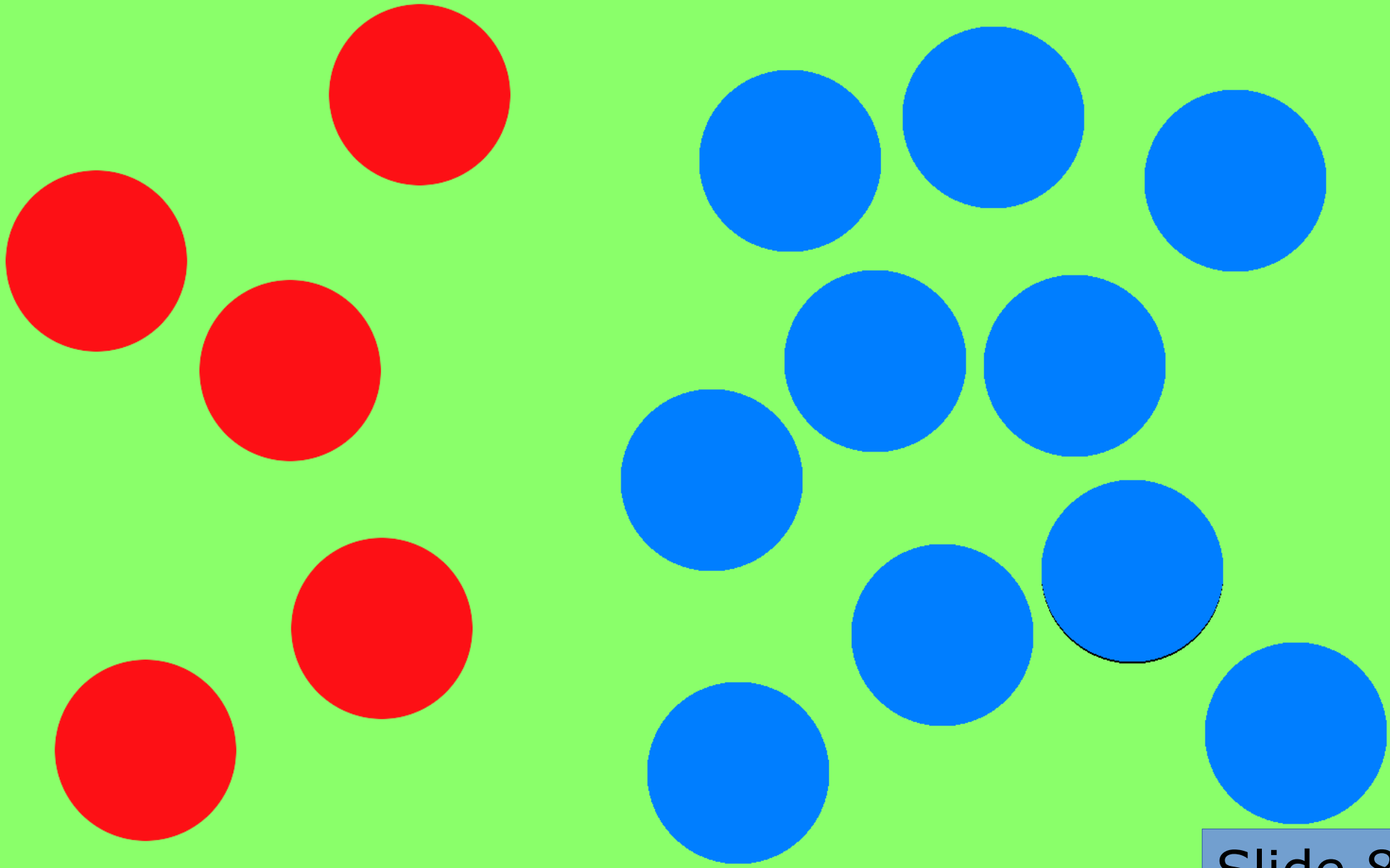
k-Nearest Neighbours



k-Nearest Neighbours



k-Nearest Neighbours



kNN should be multi-modal













Distance

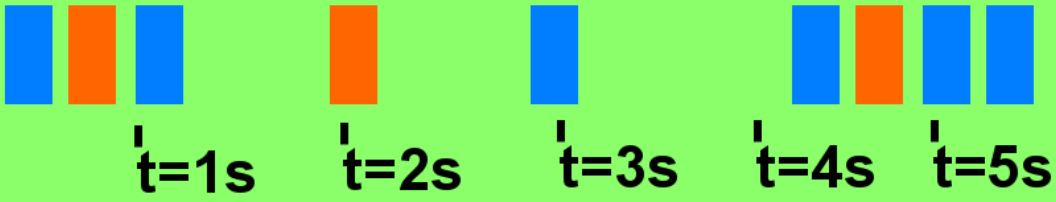






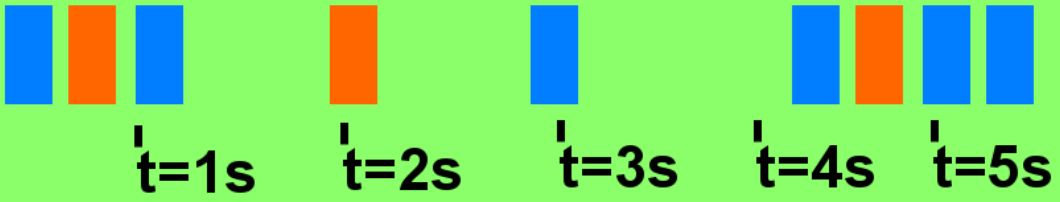

reddit
YouTube

Packet sequence features



Total time

Packet sequence features

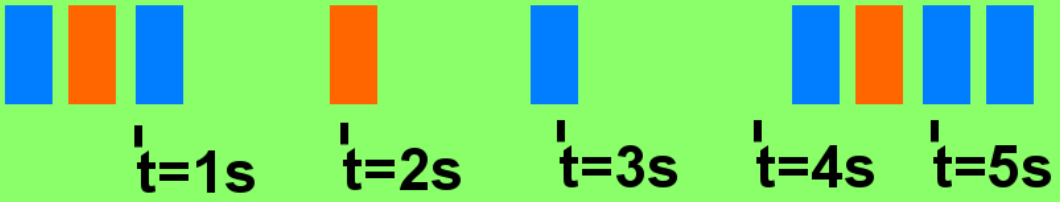


Total time



Packet count

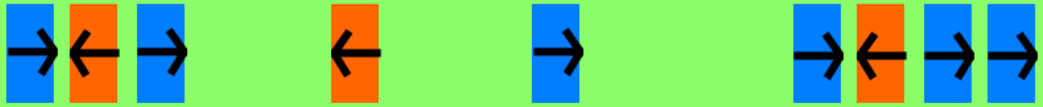
Packet sequence features



Total time

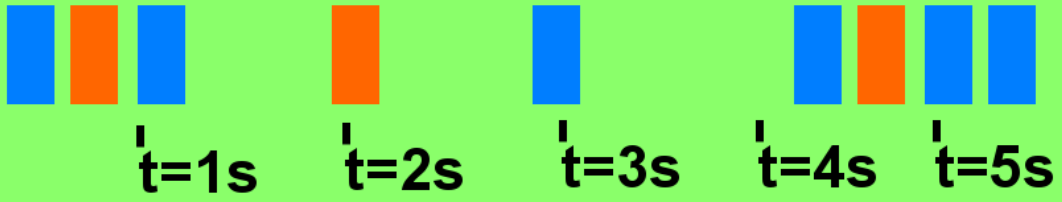


Packet count



Packet dir.

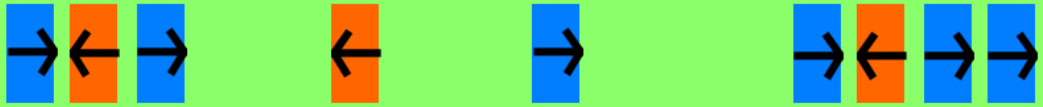
Packet sequence features



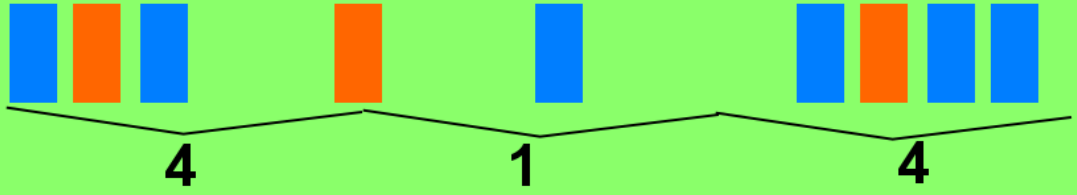
Total time



Packet count

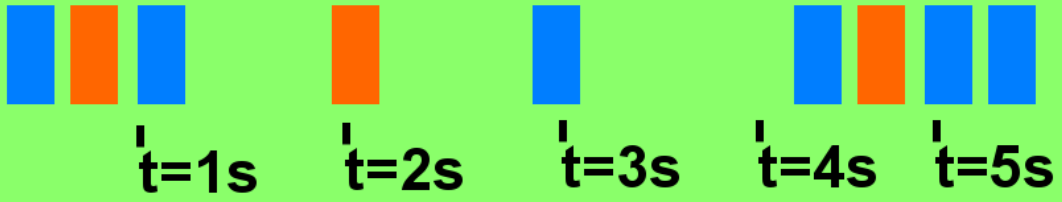


Packet dir.



Packet density

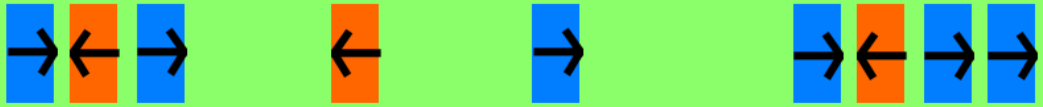
Packet sequence features



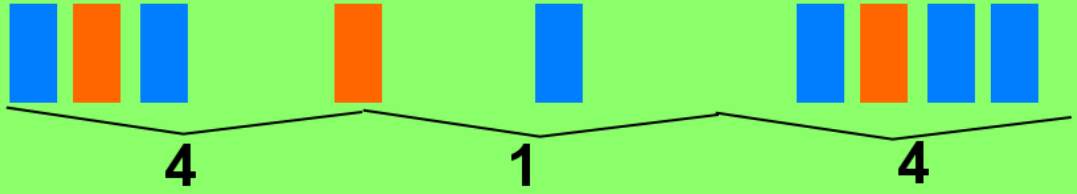
Total time



Packet count



Packet dir.

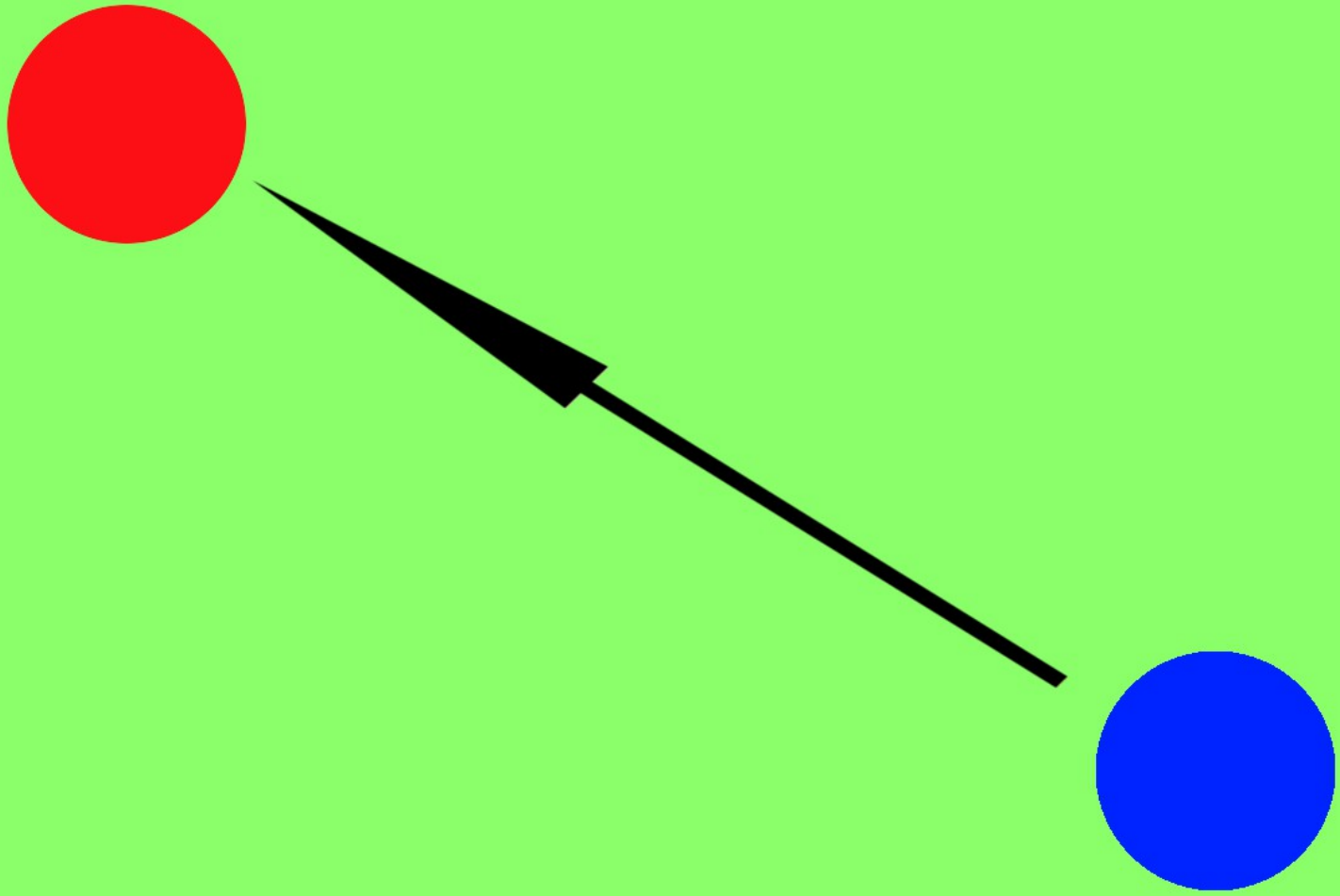


Packet density

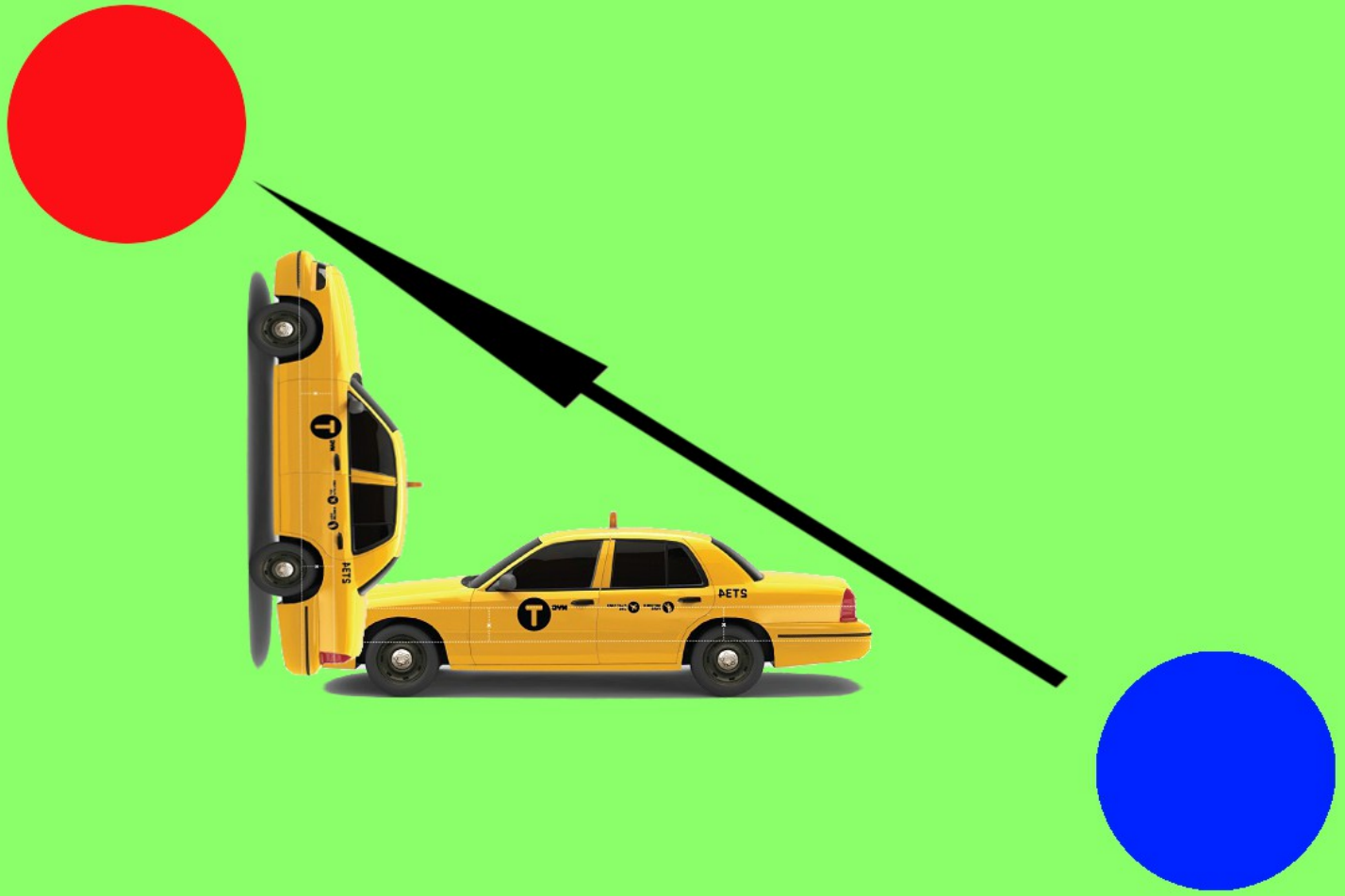


Burst

Distance training



Distance training

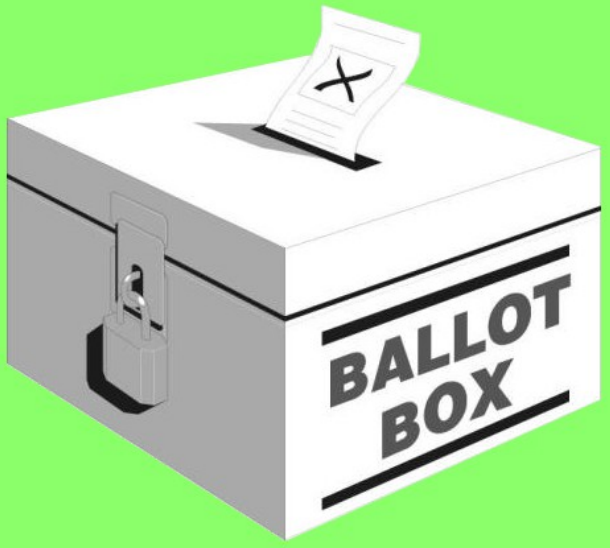


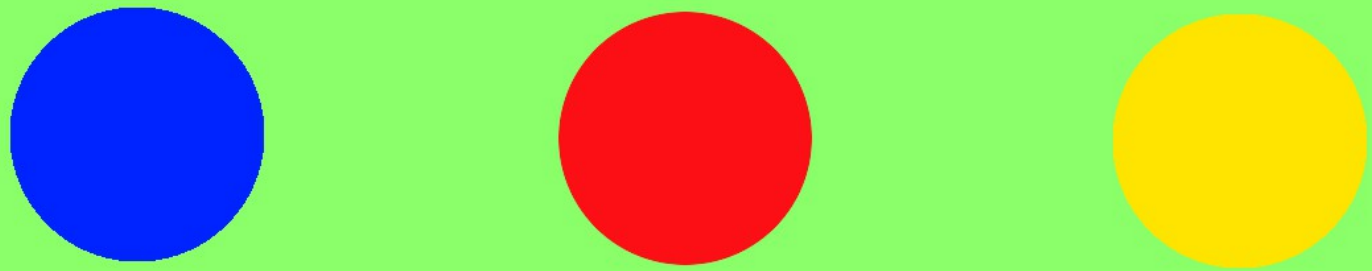
Distance training

$$\sum w_i \left| f_{b_i} - f_{r_i} \right|$$



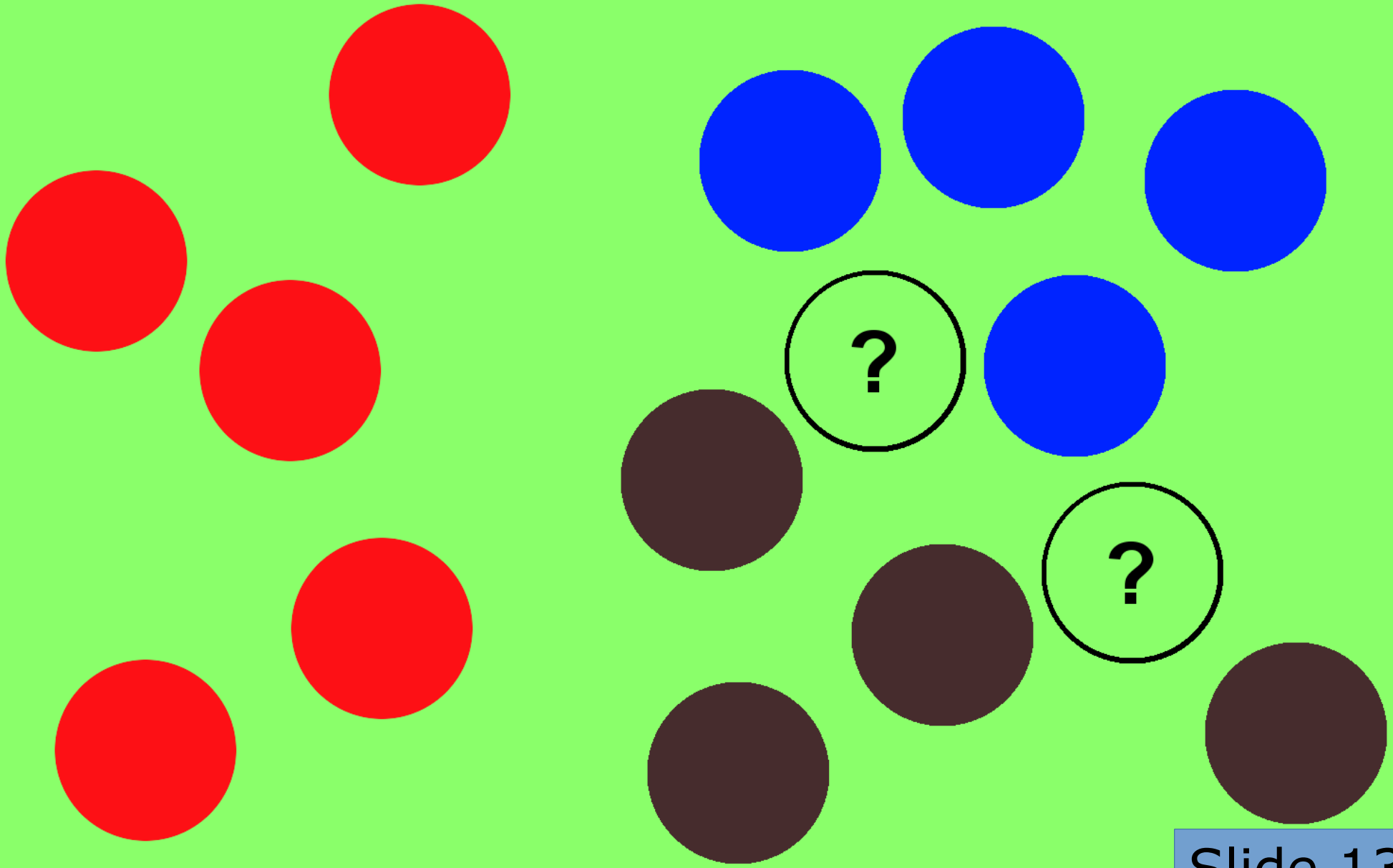
Distance training



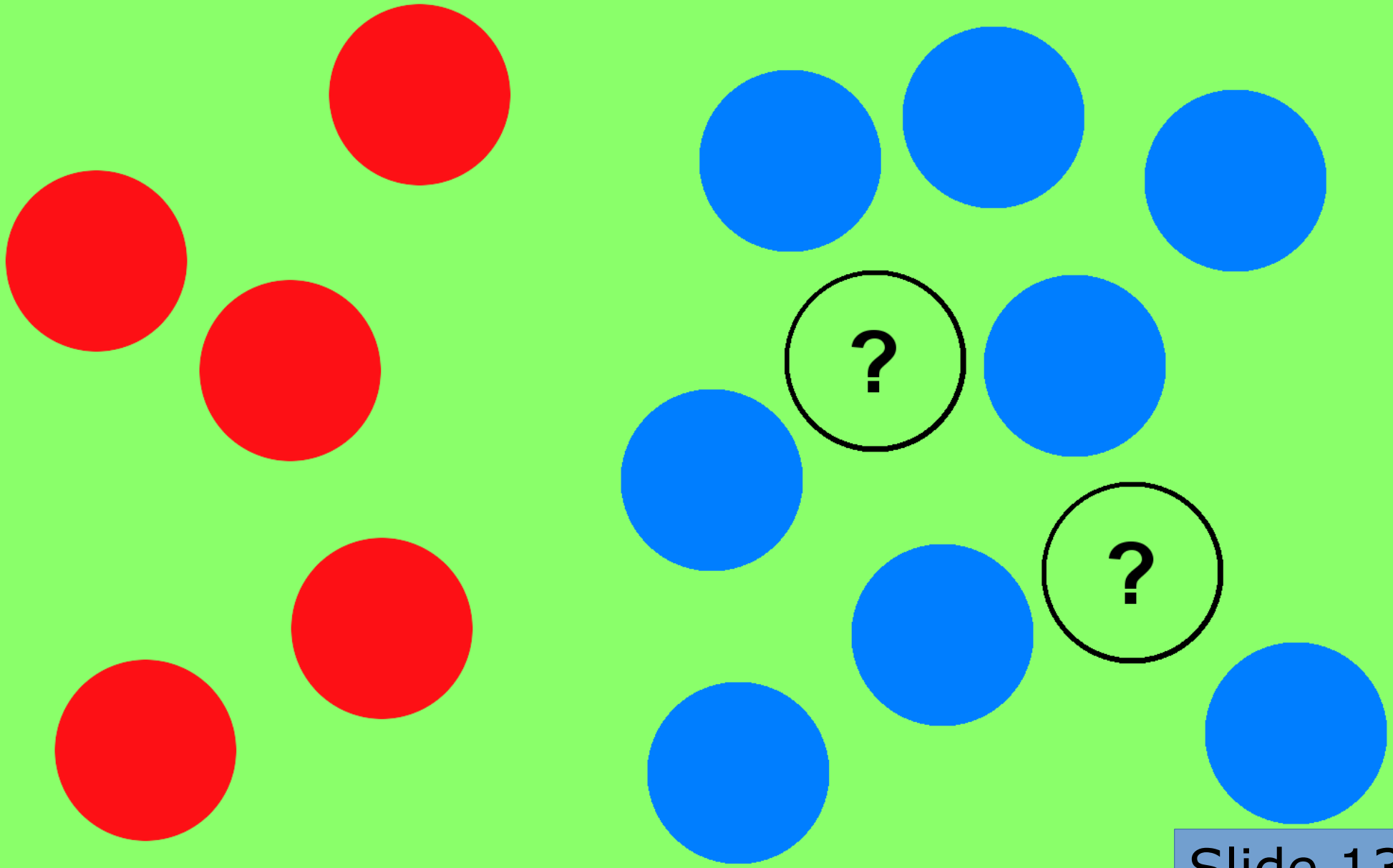


Each point gets a vote on weight changes

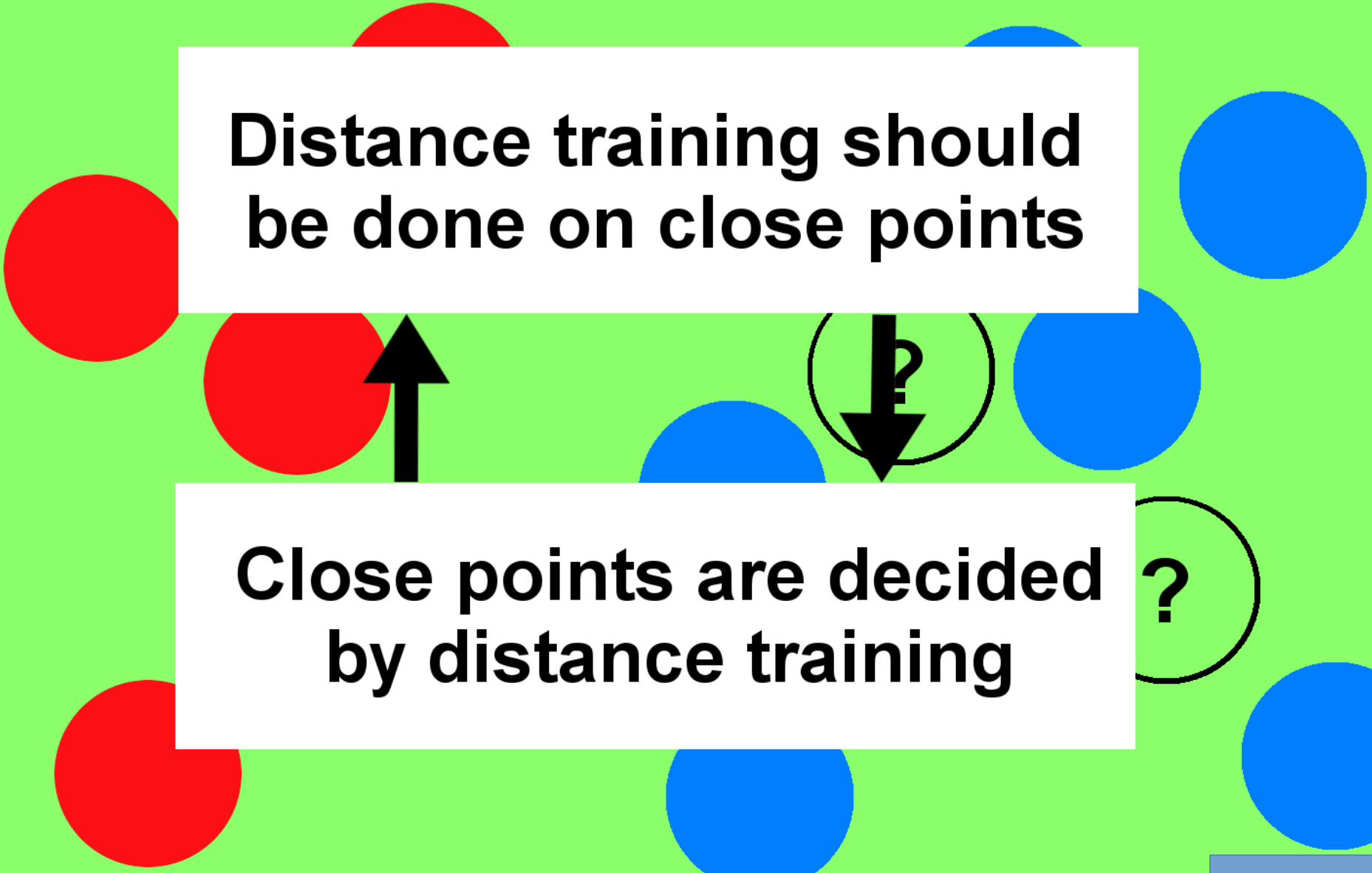
Distance training



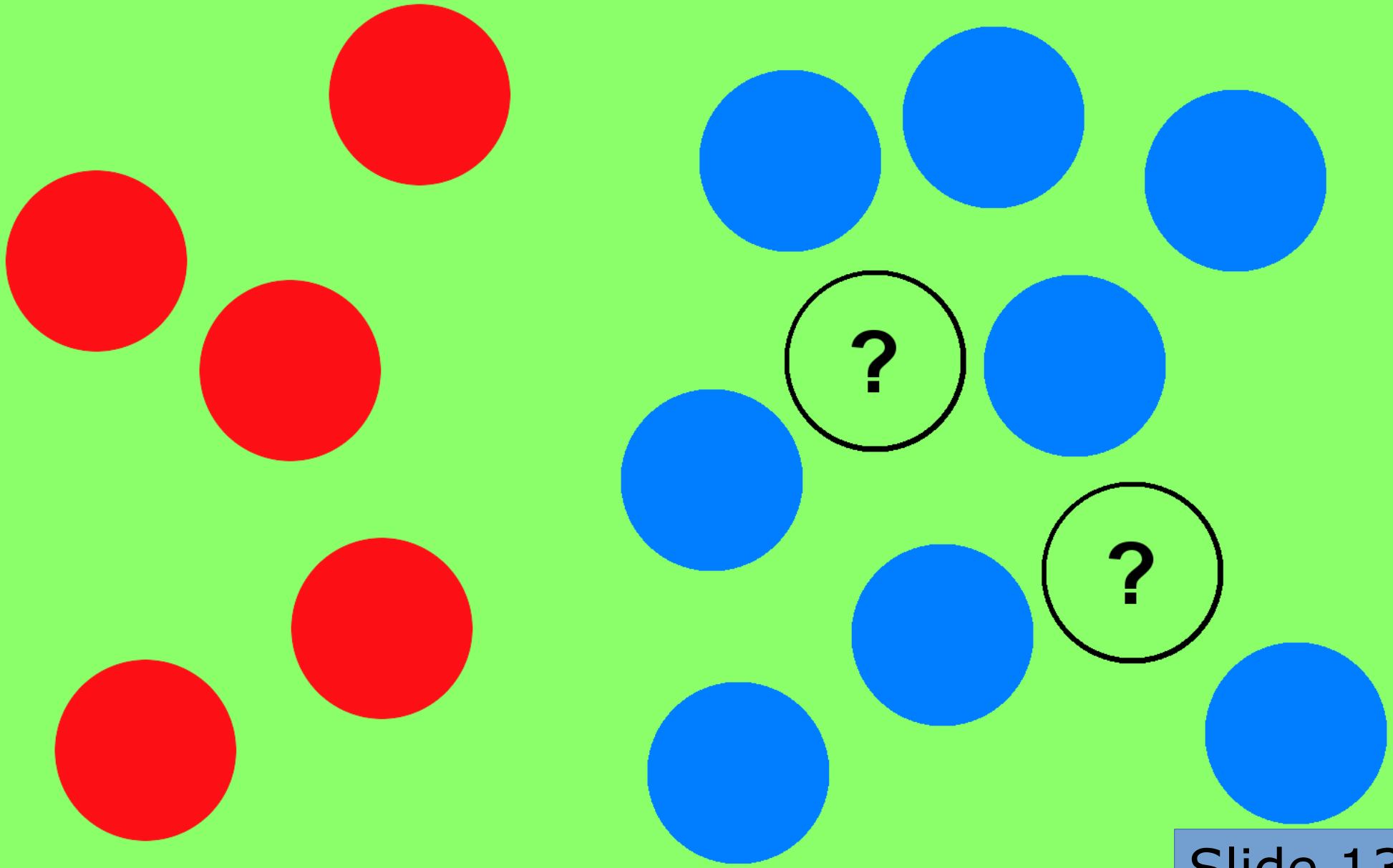
Distance training



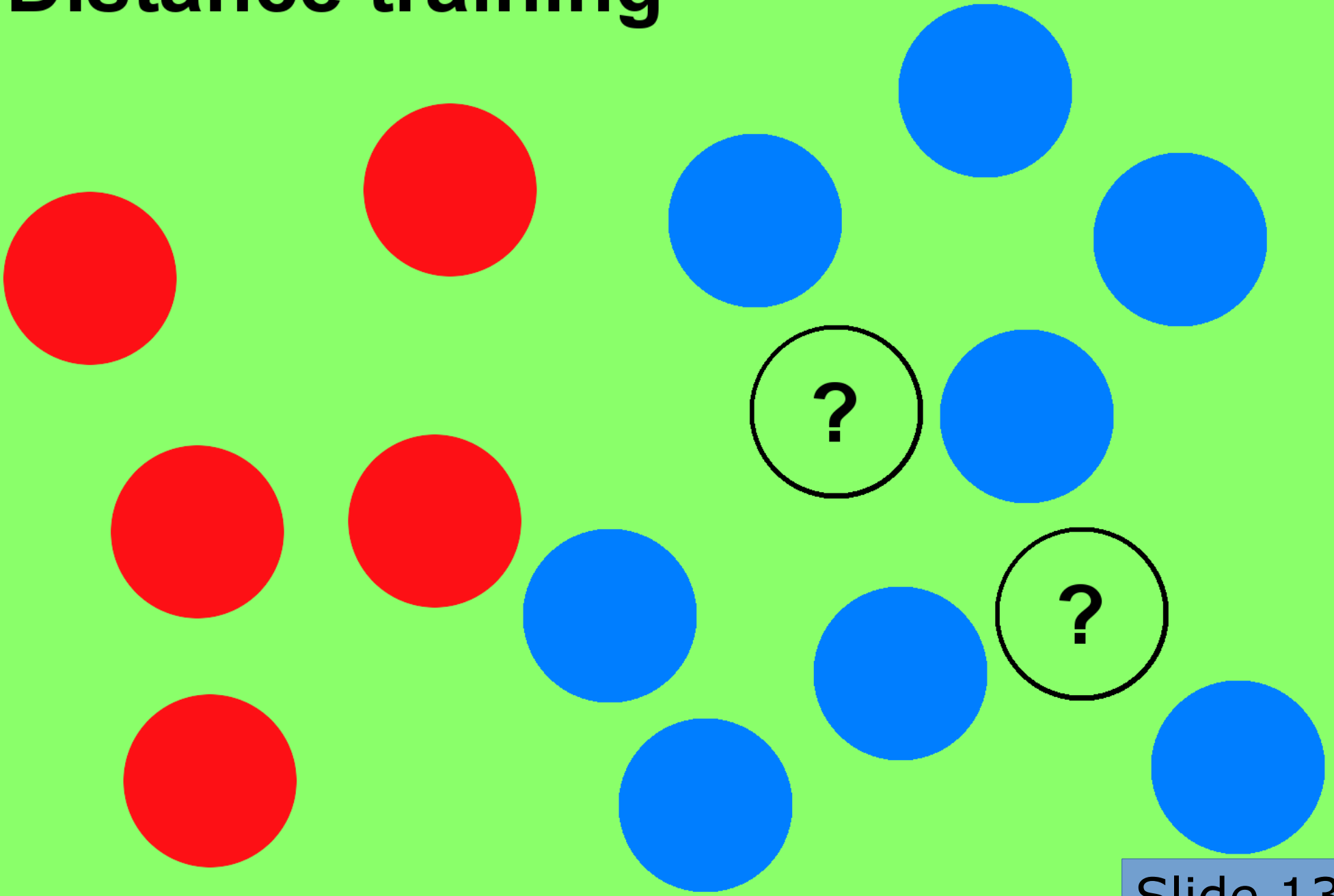
Distance training



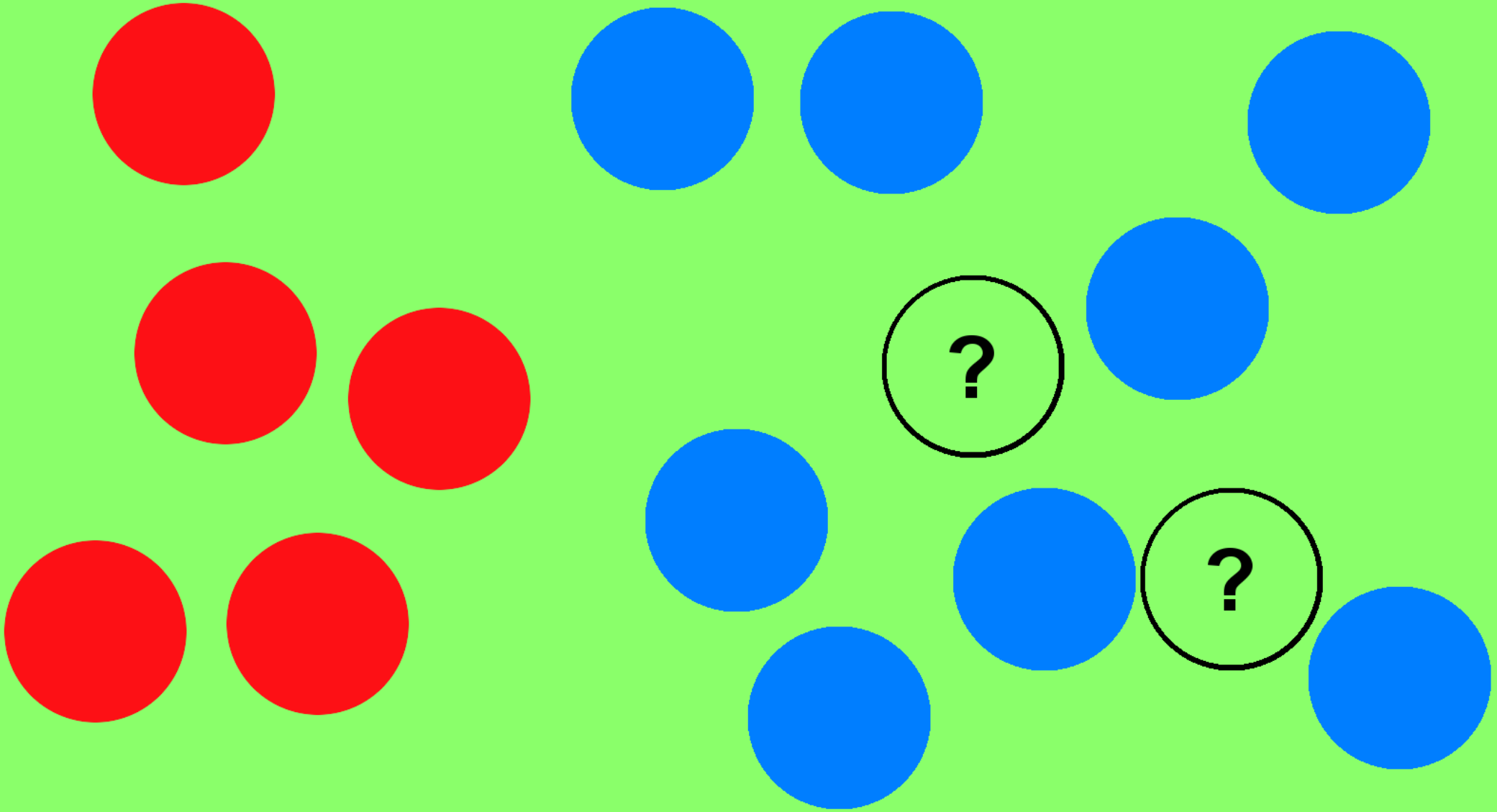
Distance training



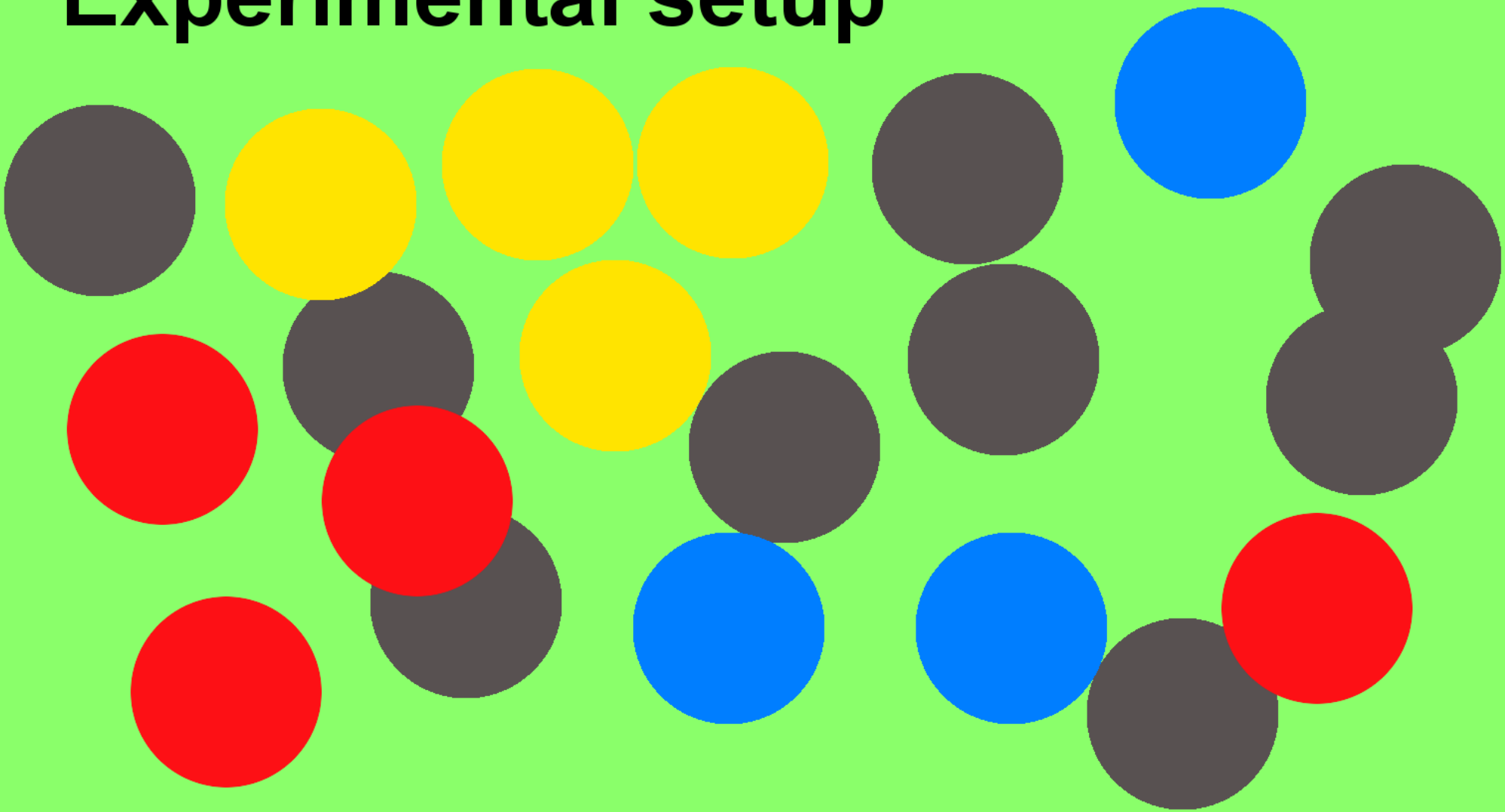
Distance training



Distance training

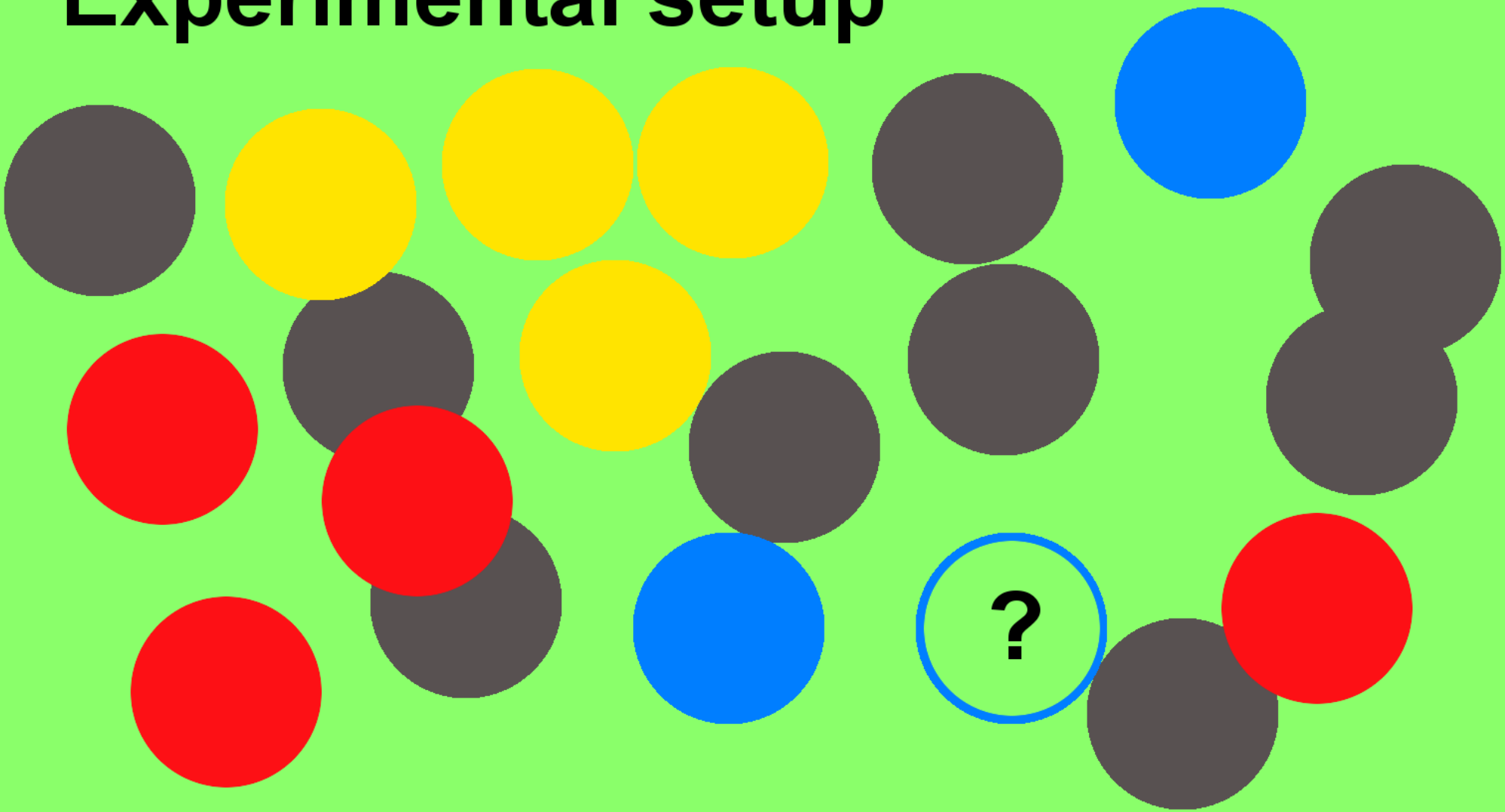


Experimental setup



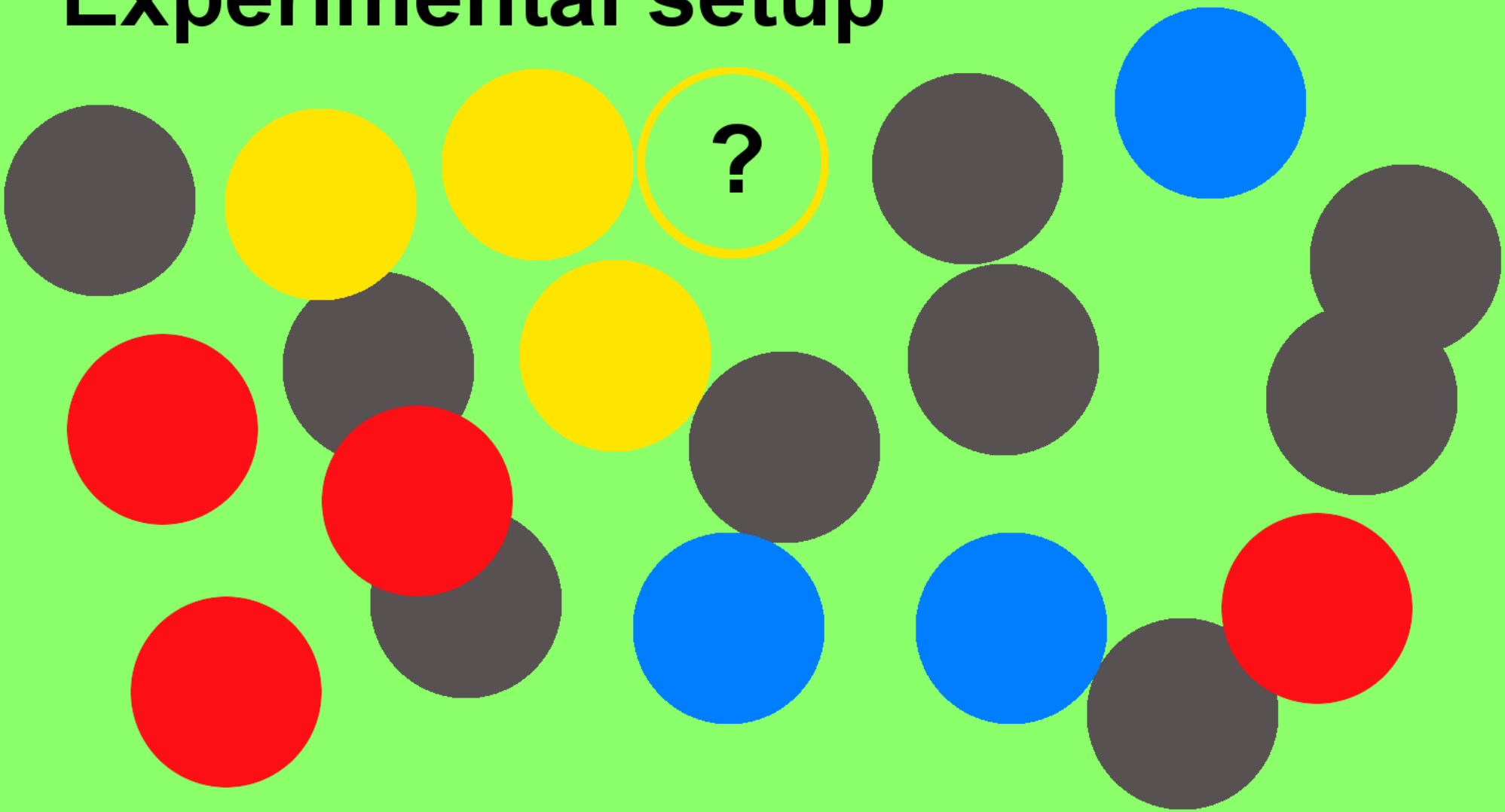
100 colors, 90 circles each (monitored sites)
9000 grey circles (non-monitored sites)

Experimental setup



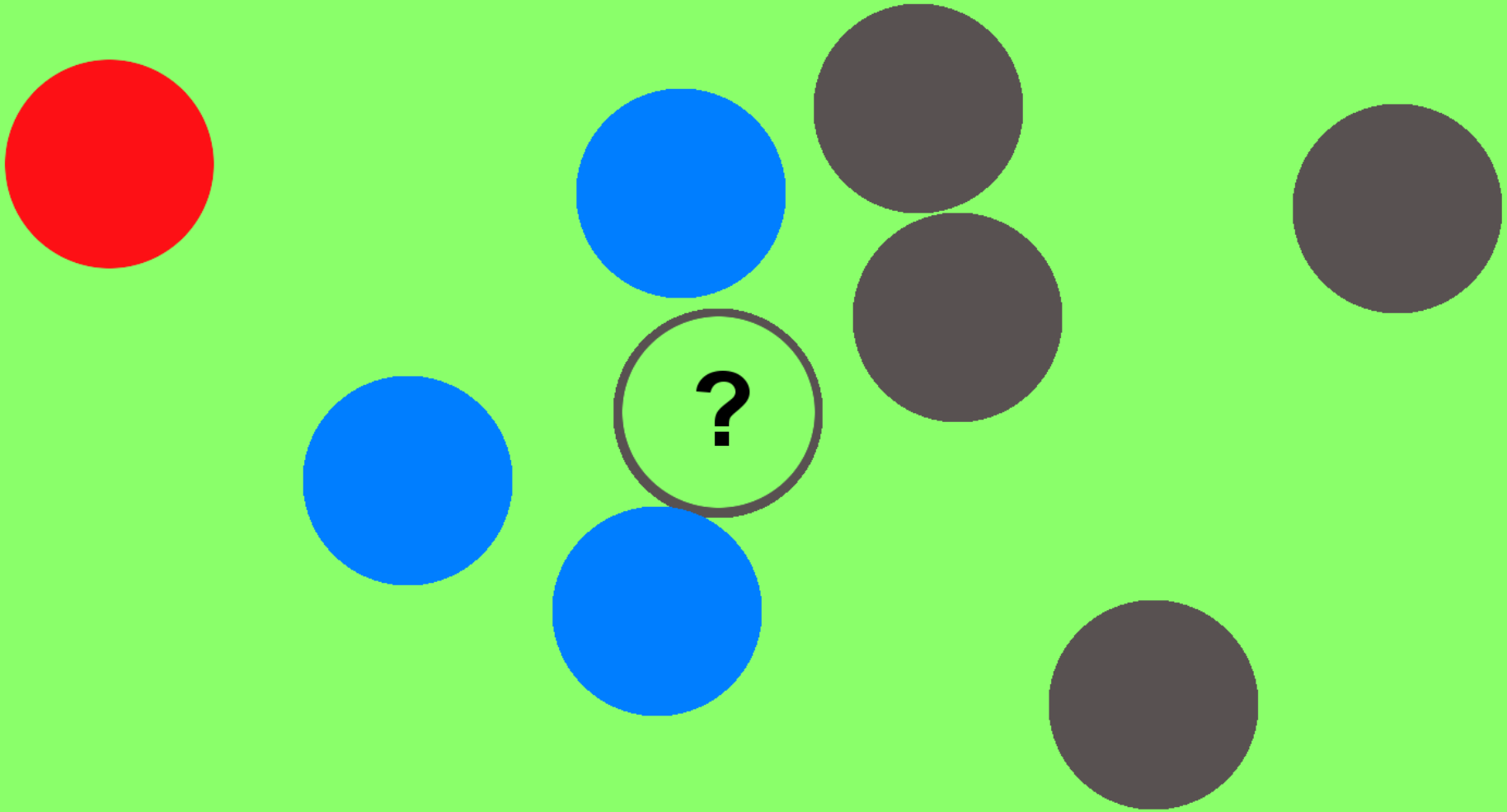
100 colors, 90 circles each (monitored sites)
9000 grey circles (non-monitored sites)

Experimental setup

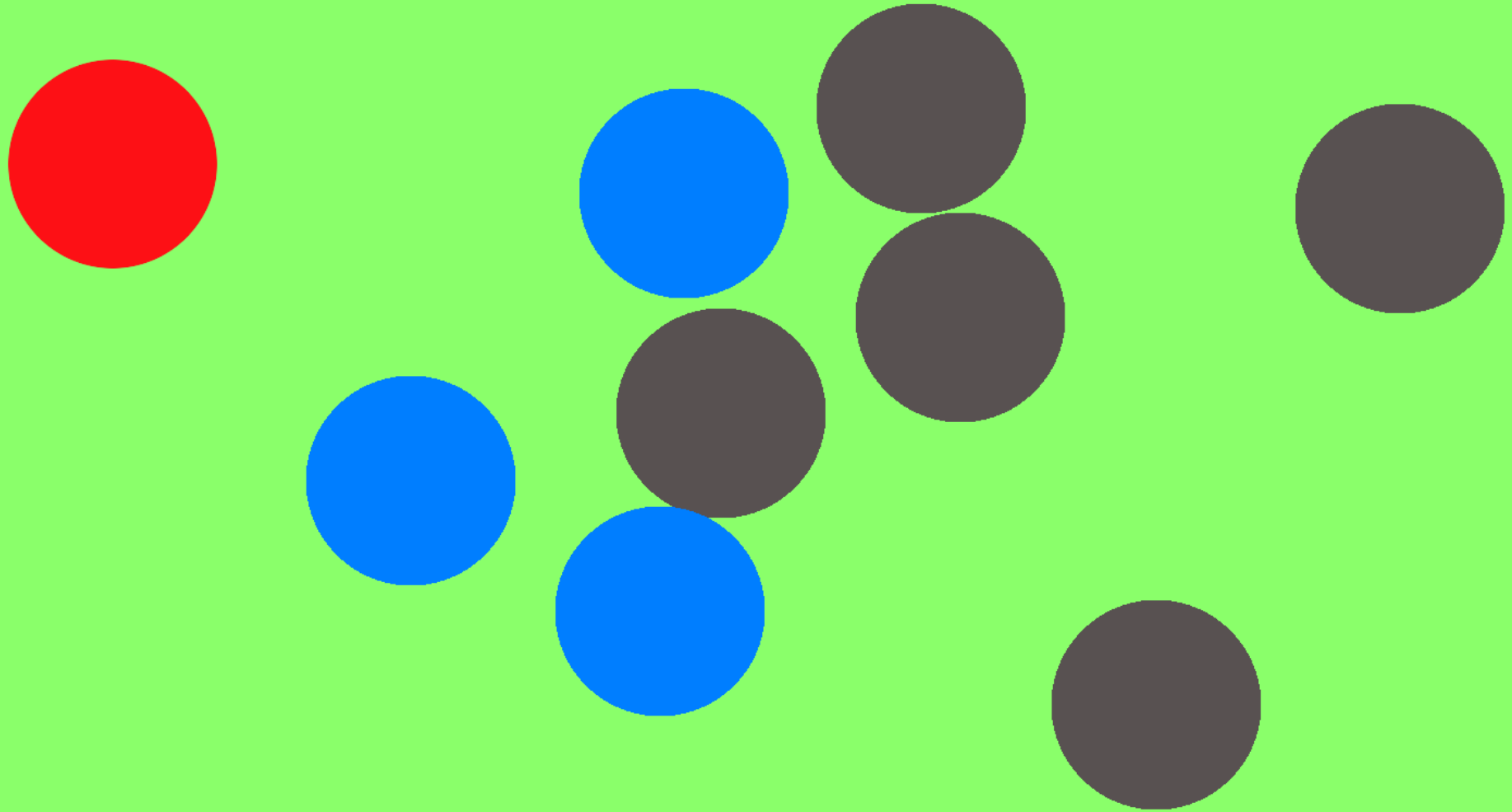


100 colors, 90 circles each (monitored sites)
9000 grey circles (non-monitored sites)

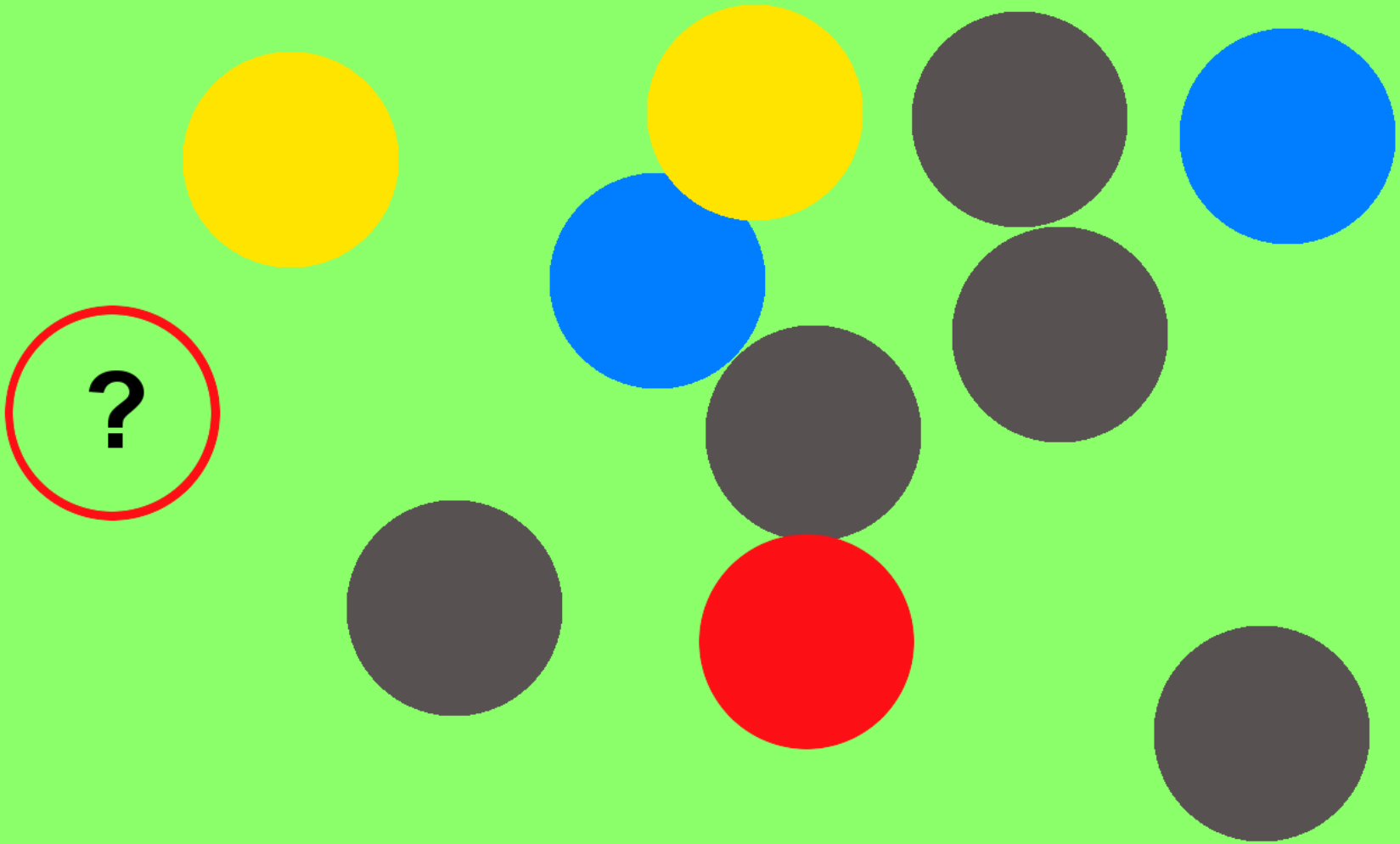
No consensus = false



No consensus = false

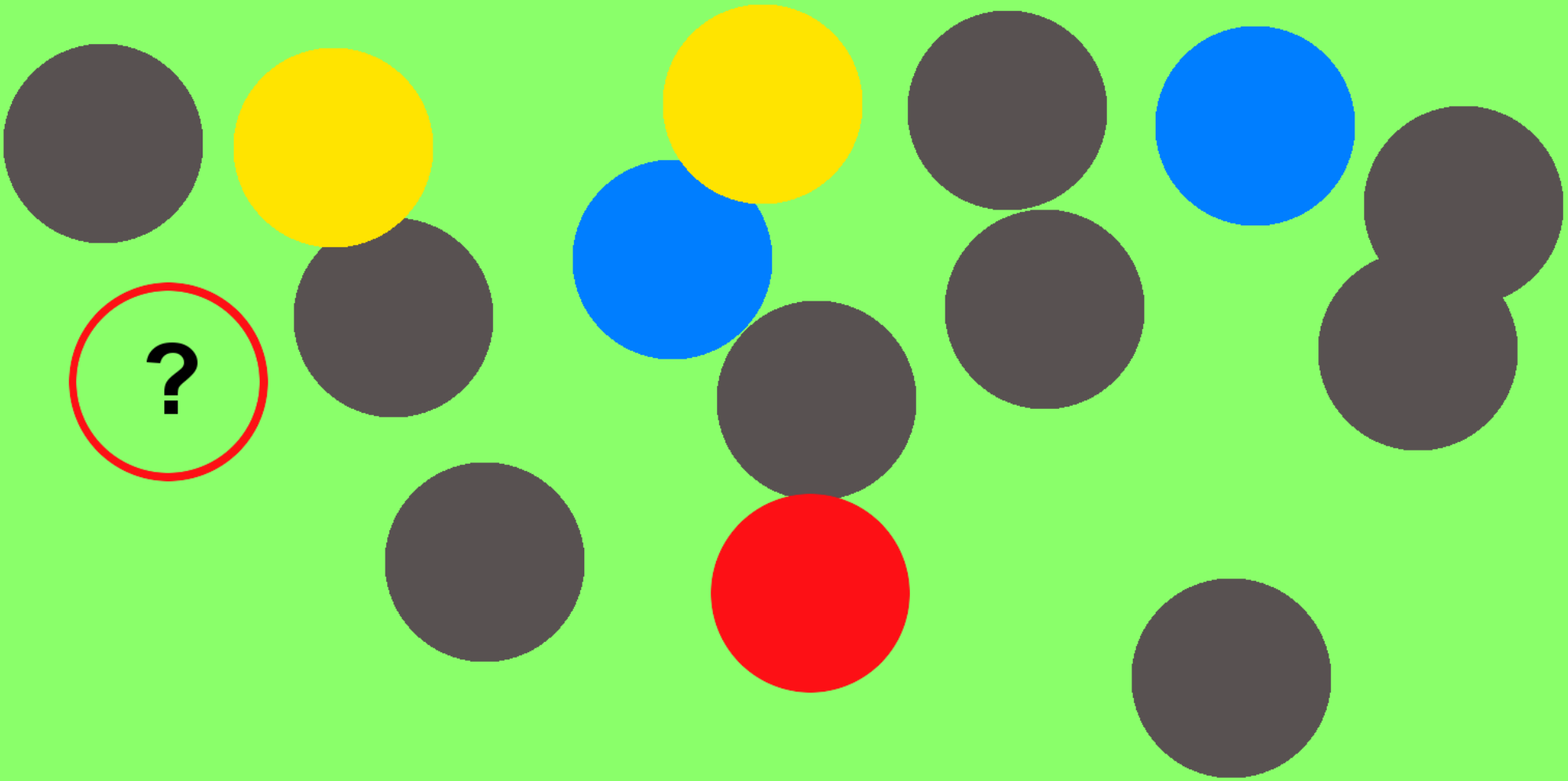


Parameter 1: Grey circle number



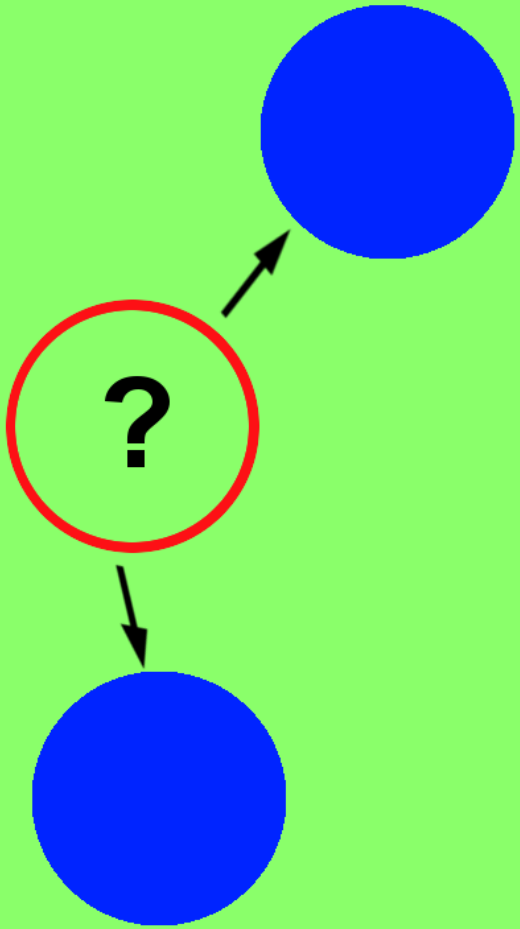
Range: 100 to 5000

Parameter 1: Grey circle number



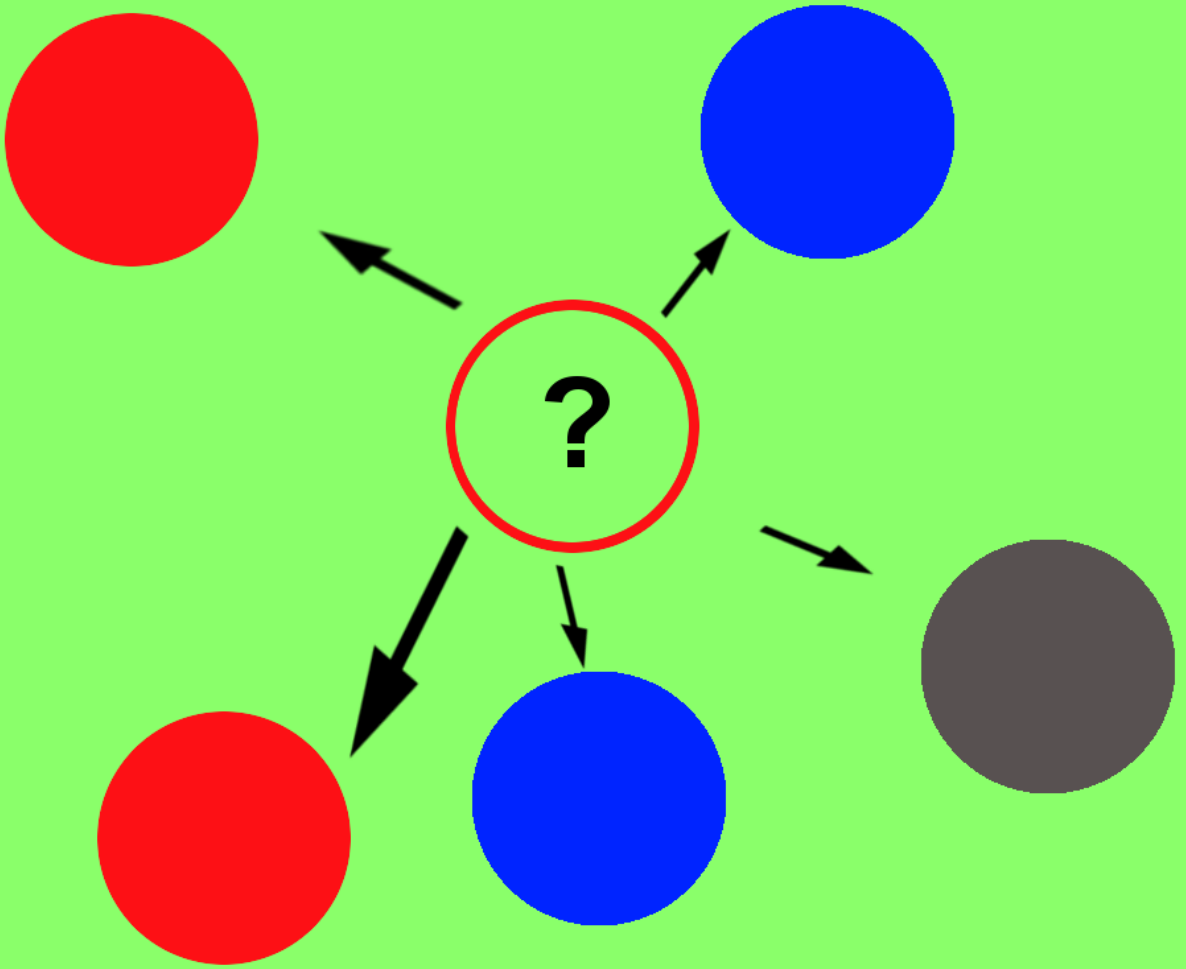
Range: 100 to 5000

Parameter 2: Neighbour number



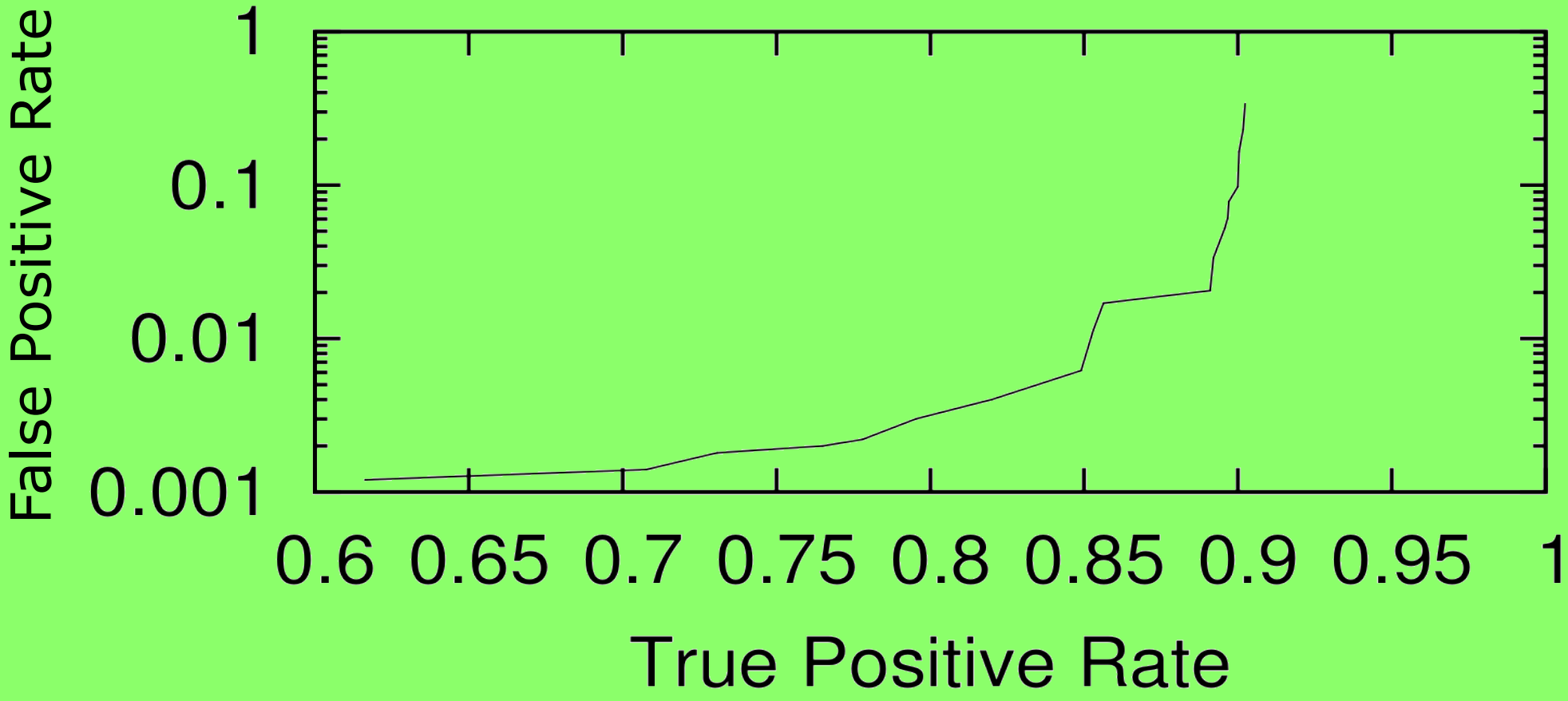
Range: 1 to 15

Parameter 2: Neighbour number



Range: 1 to 15

Accuracy



TPR 85% +/- 4% at FPR 0.6% +/- 0.4%

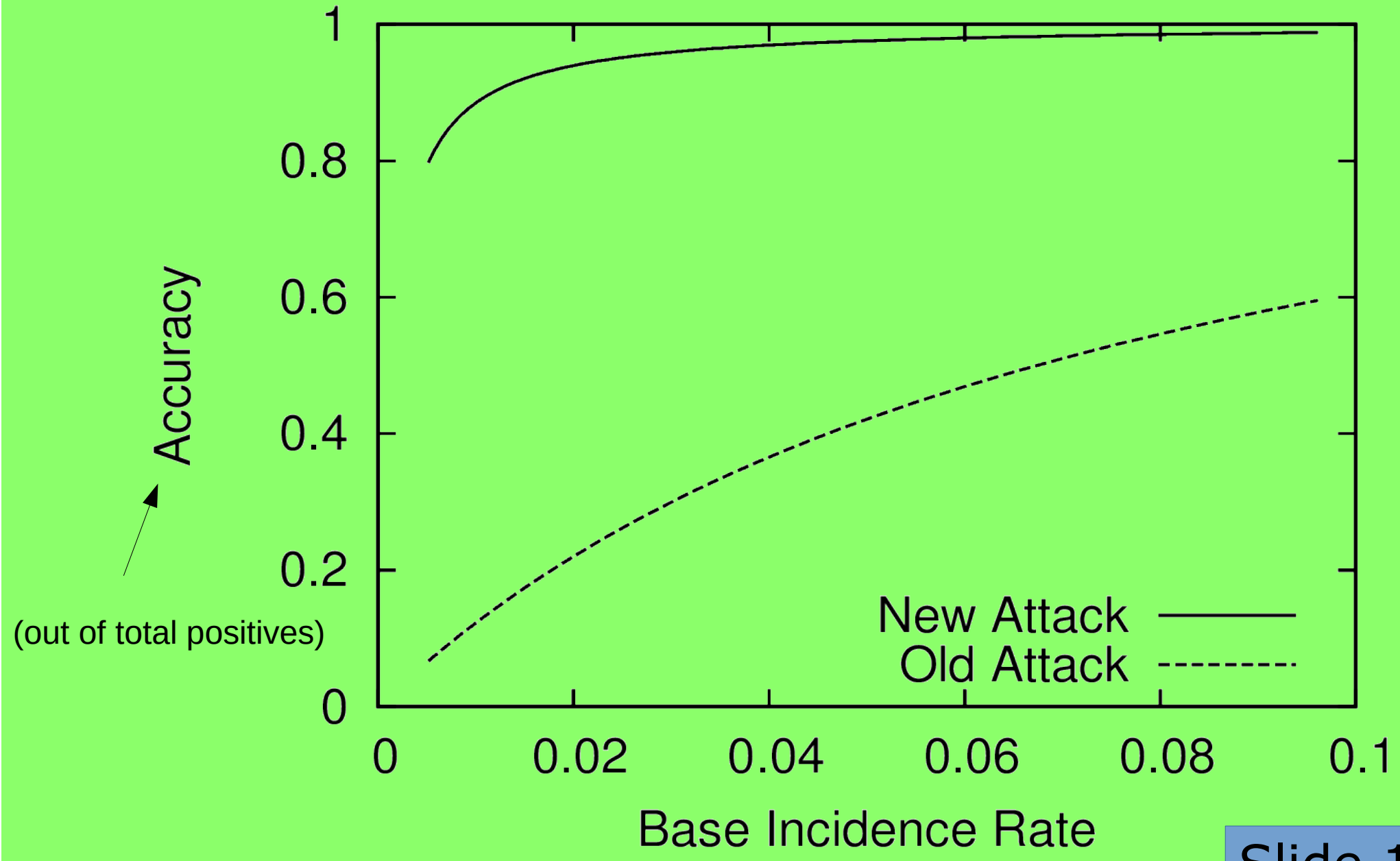
TPR 76% +/- 6% at FPR 0.1% +/- 0.1%

Accuracy

	TPR	FPR	Problem size
Panchenko	57%	0.2%	5
Cai	83%	0.1%	1
Wang	97%	0.2%	1
Wang	83%	6%	100
New	85%	0.6%	100
	76%	0.1%	100

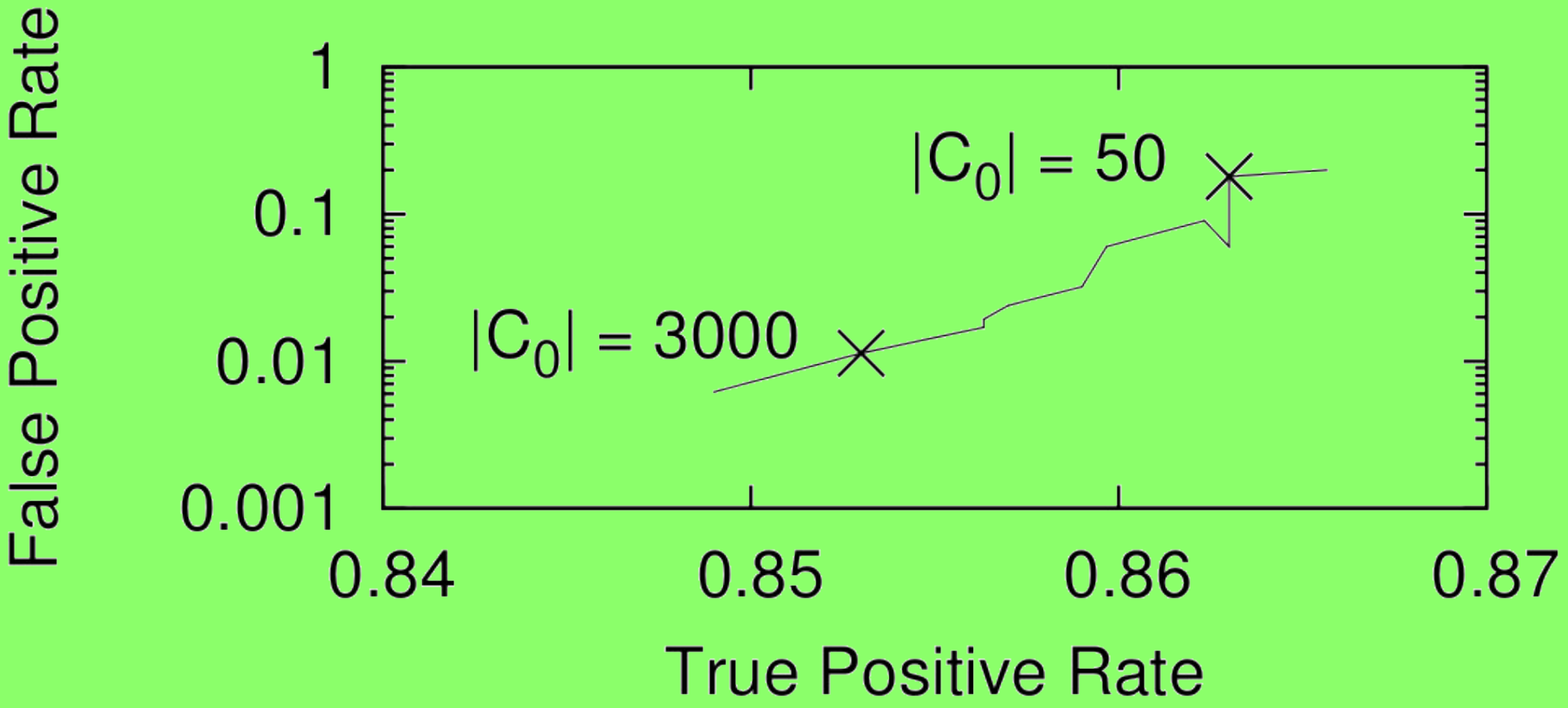
The FPR is around 10-60 times better at roughly the same TPR

Accuracy


1
0.8
0.6
0.4
0.2
0
Accuracy
(out of total positives)
0
0.02
0.04
0.06
0.08
0.1
Base Incidence Rate
New Attack
Old Attack

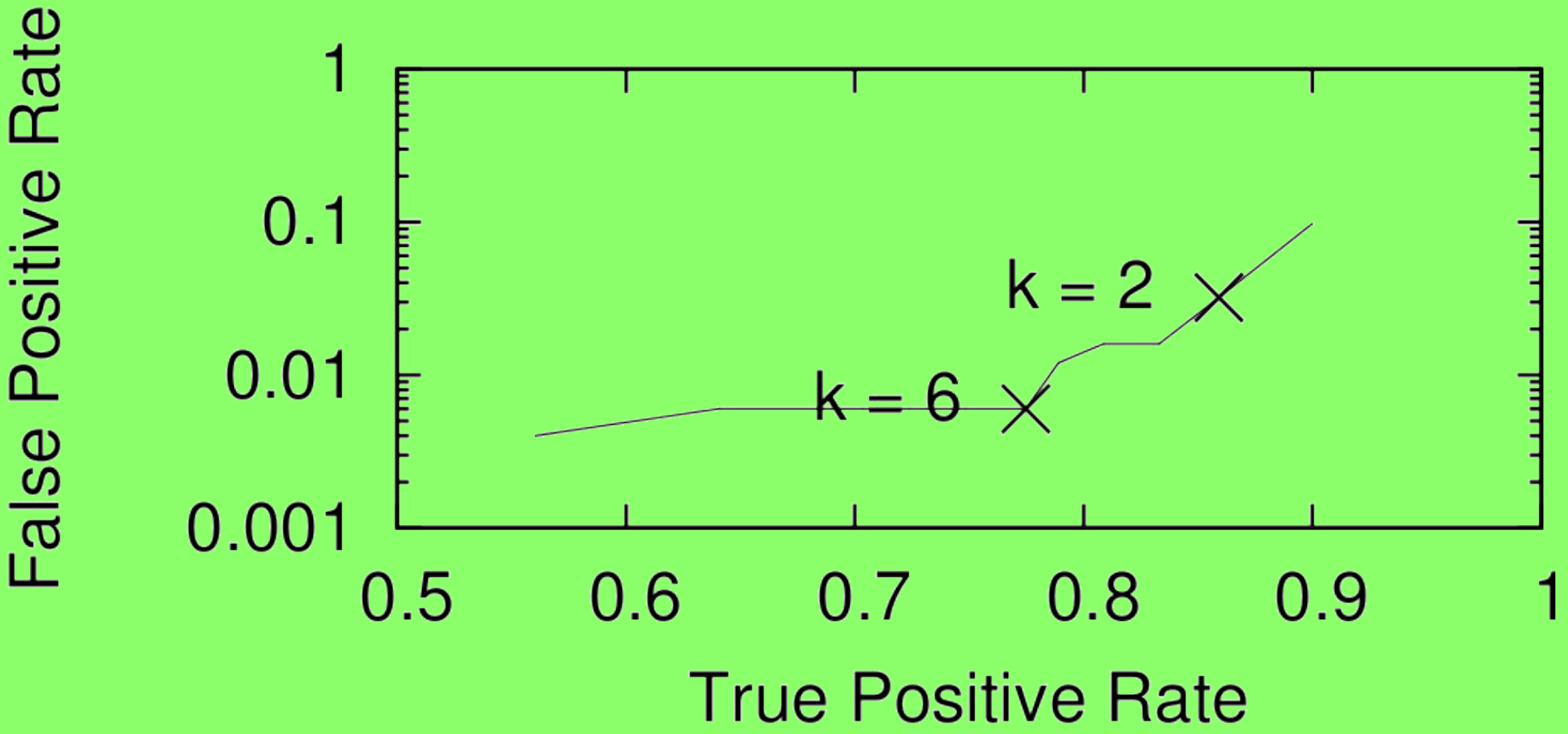
Accuracy

How many grey circles to throw in?

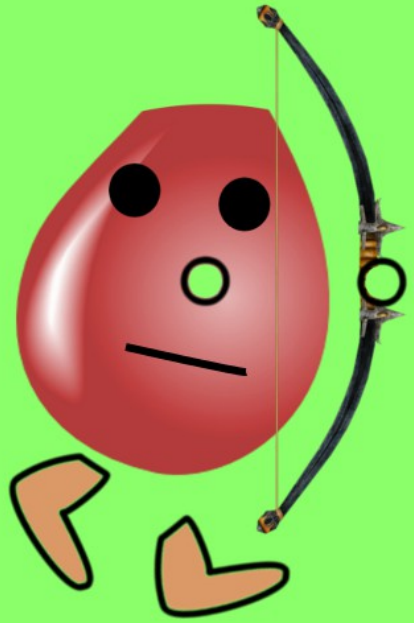


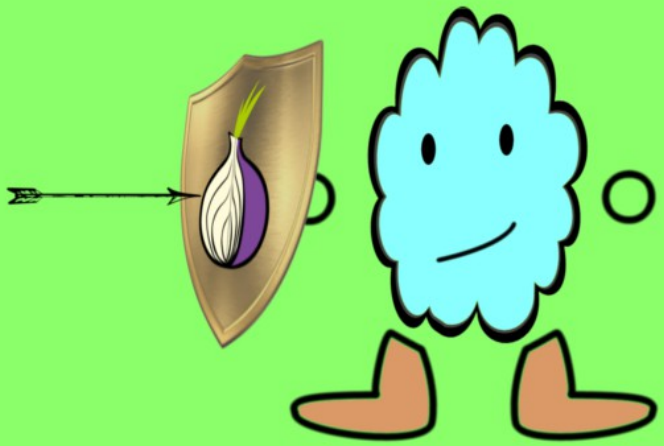
Accuracy

How many neighbours (k)?

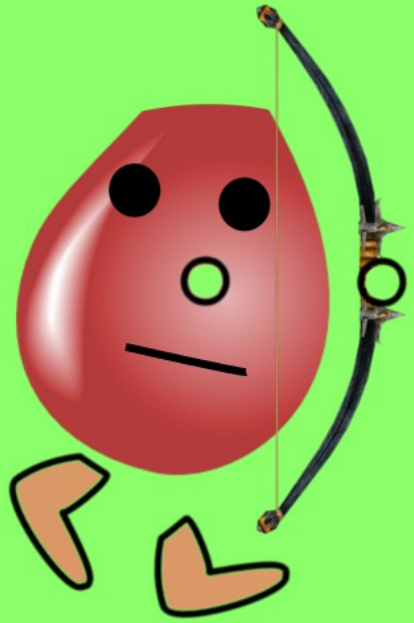


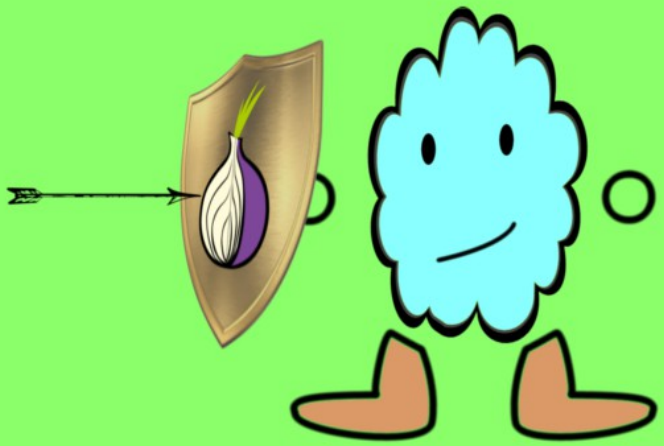
Attack vs Defense

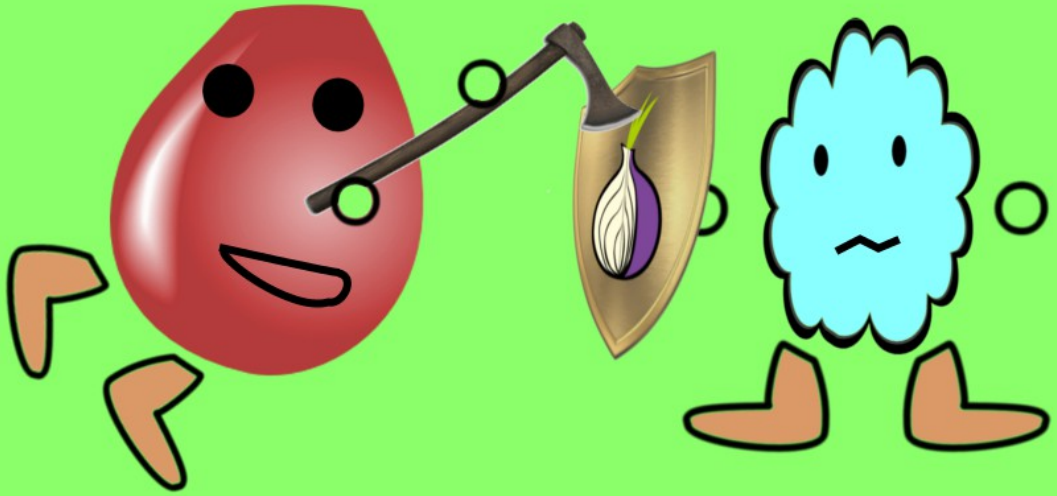




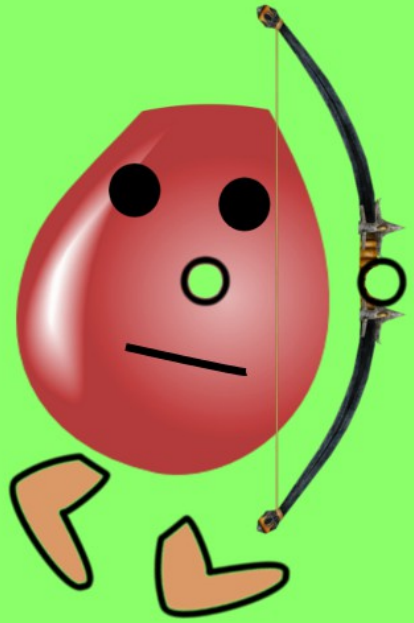
Attack vs Defense

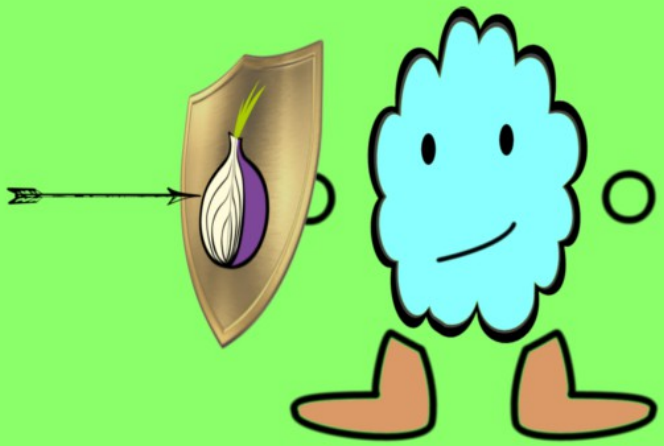


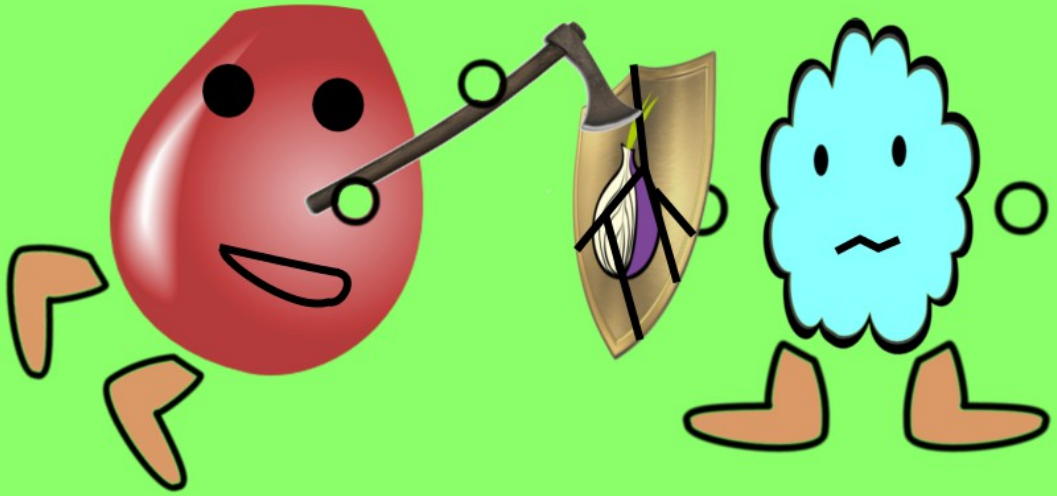




Attack vs Defense

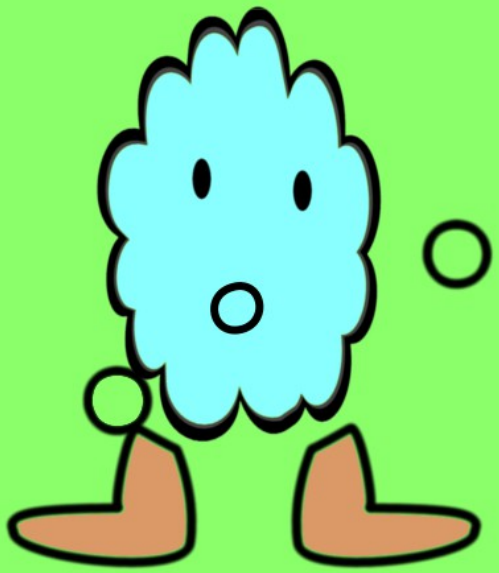


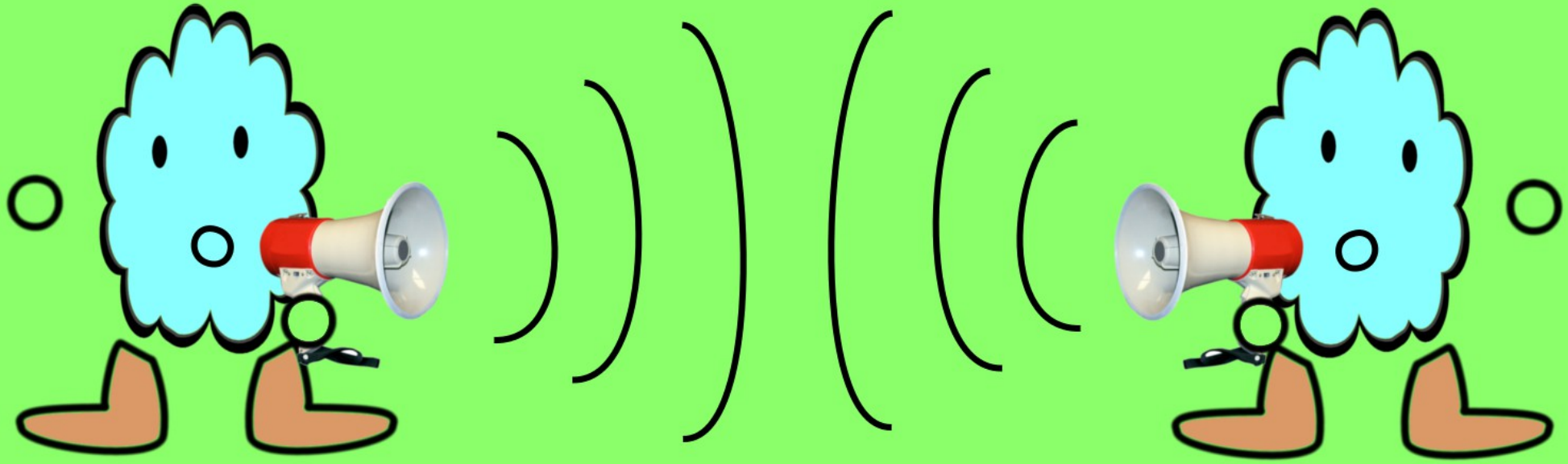


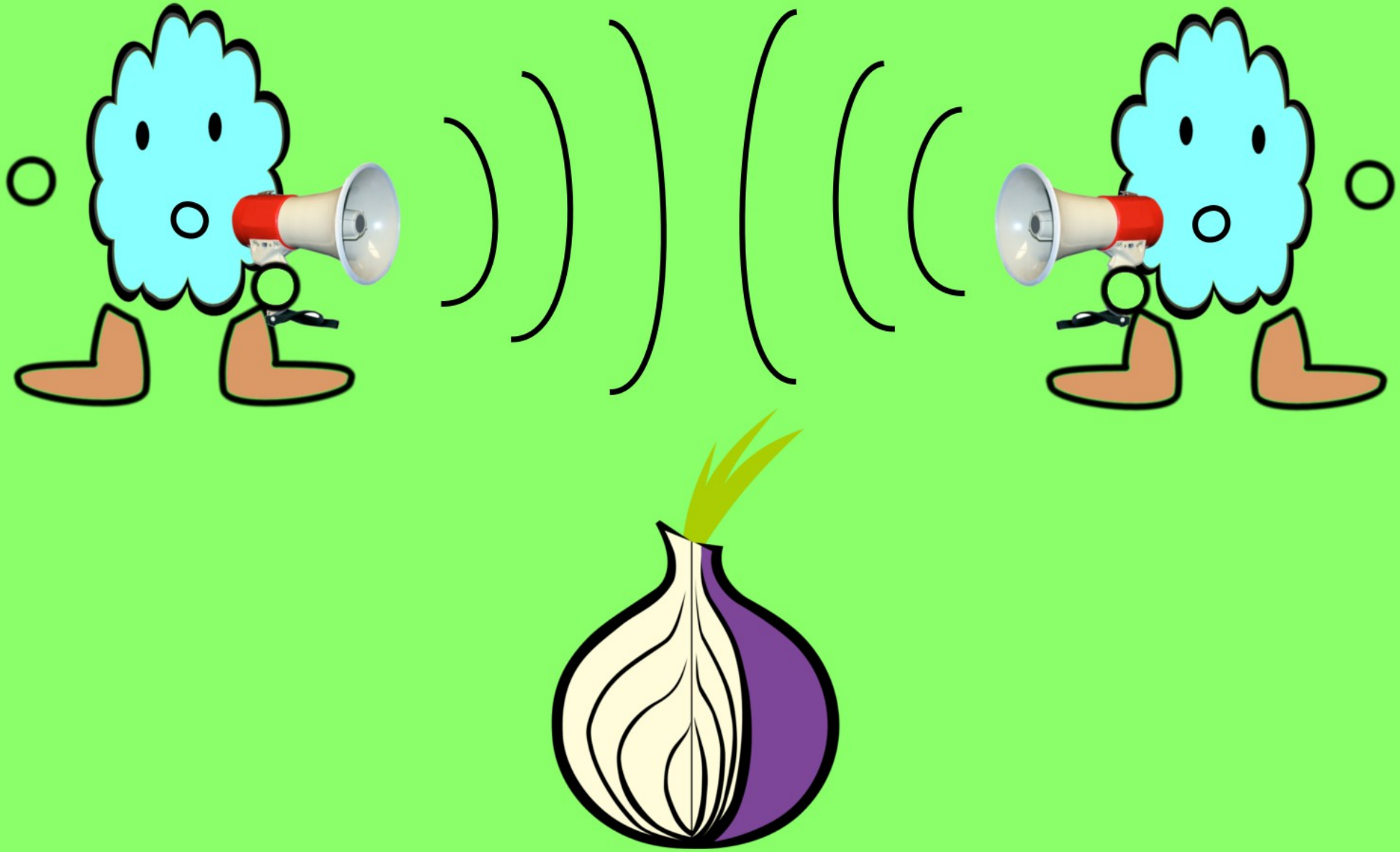


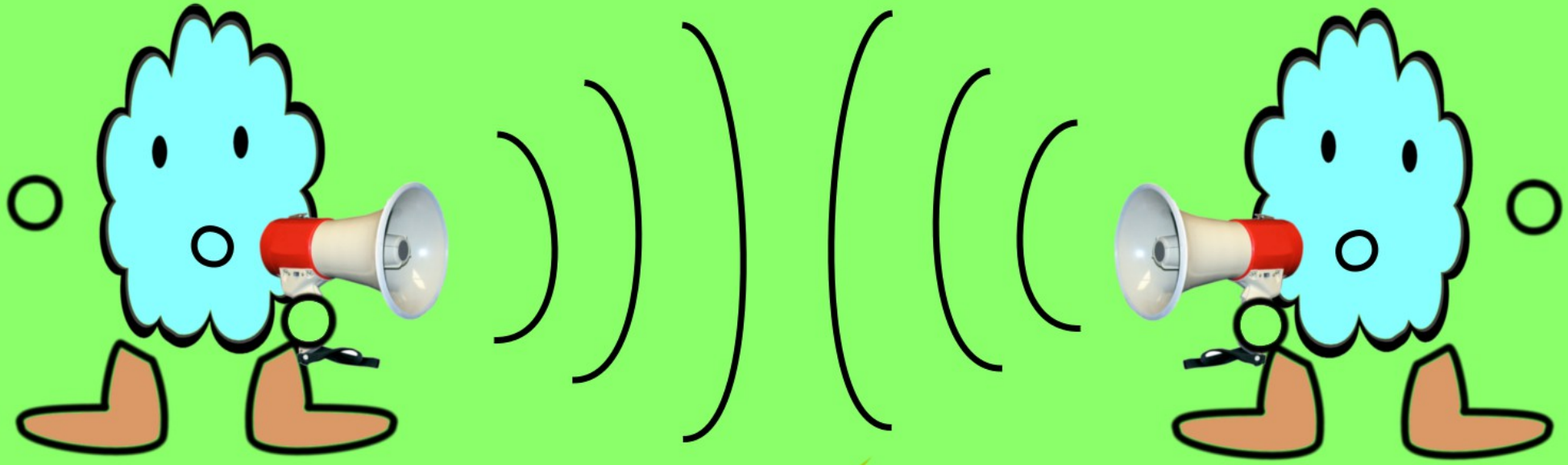

Provable
Defense





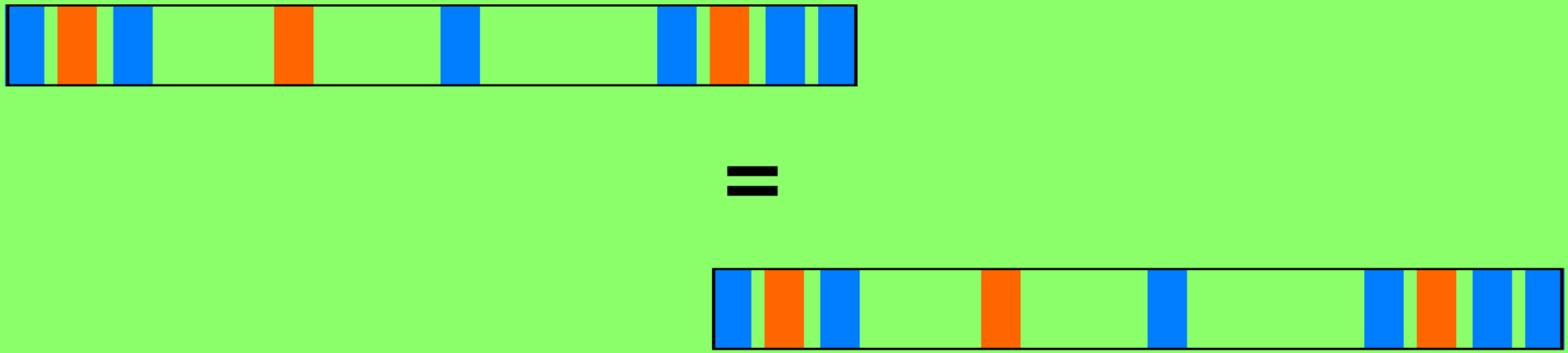








TOR SMASH!



A defended packet sequence is indistinguishable from another one.

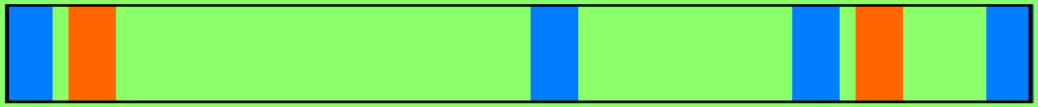


=

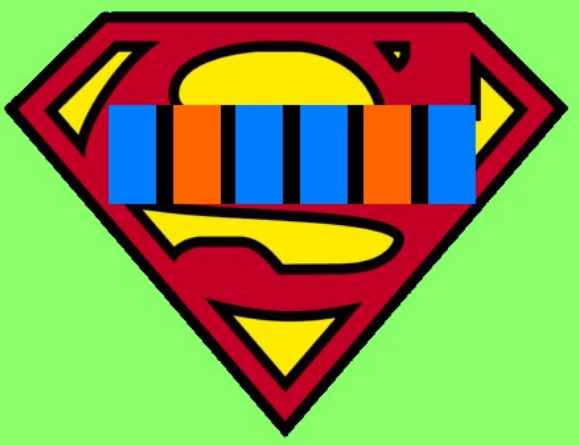
Covering packet sequence



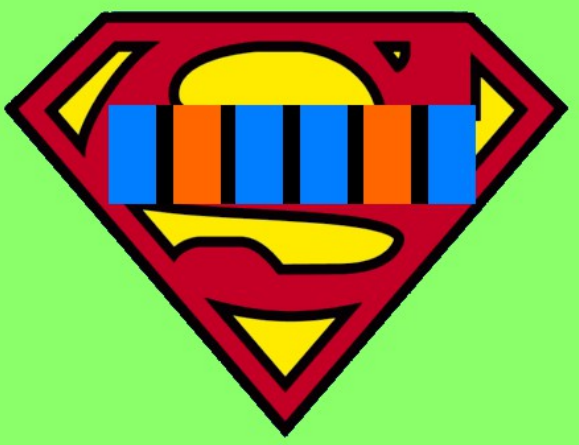
Real packet sequence



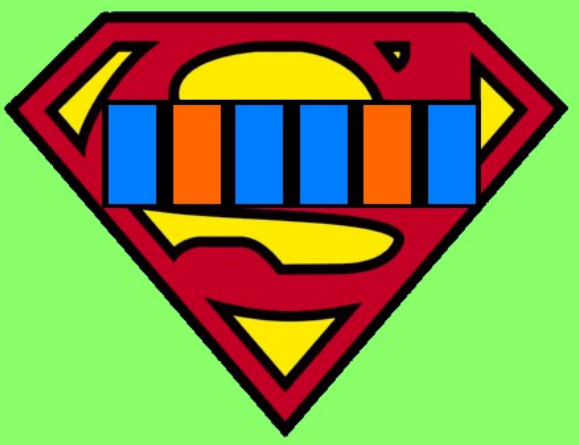
A defended packet sequence is indistinguishable from another one.



Supersequence



Shortest Common Supersequence



Shortest Common Supersequence



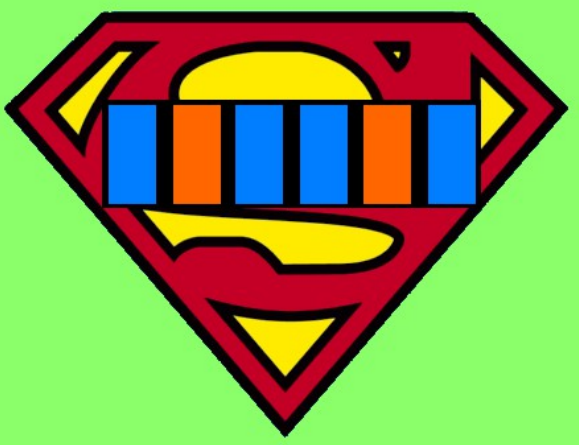






Shortest Common Supersequence





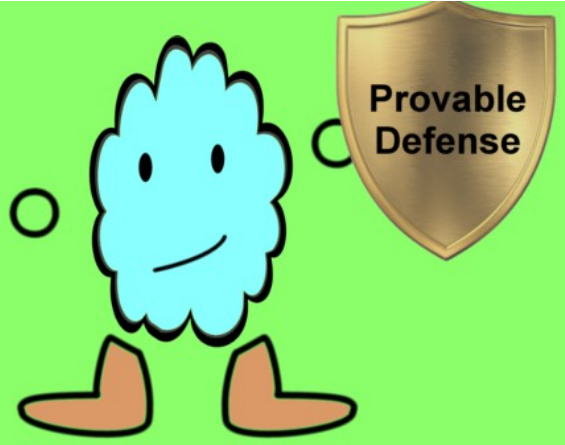
Shortest Common Supersequence







Clairvoyance!



No info

1. No info:
   One cover
   Pre-computed



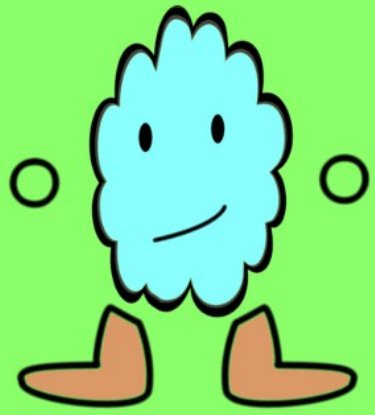
Site info

Lower overhead

More Practical



Full info



No info

2. Site info:
Multiple covers
Pre-computed



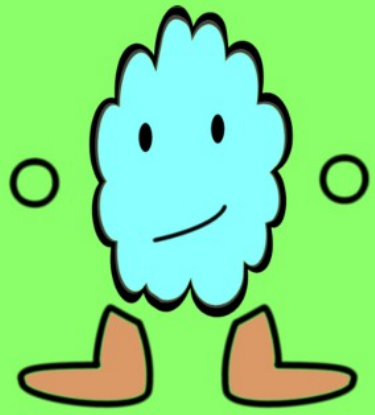
Site info

Lower overhead

More Practical



Full info



No info



Site info



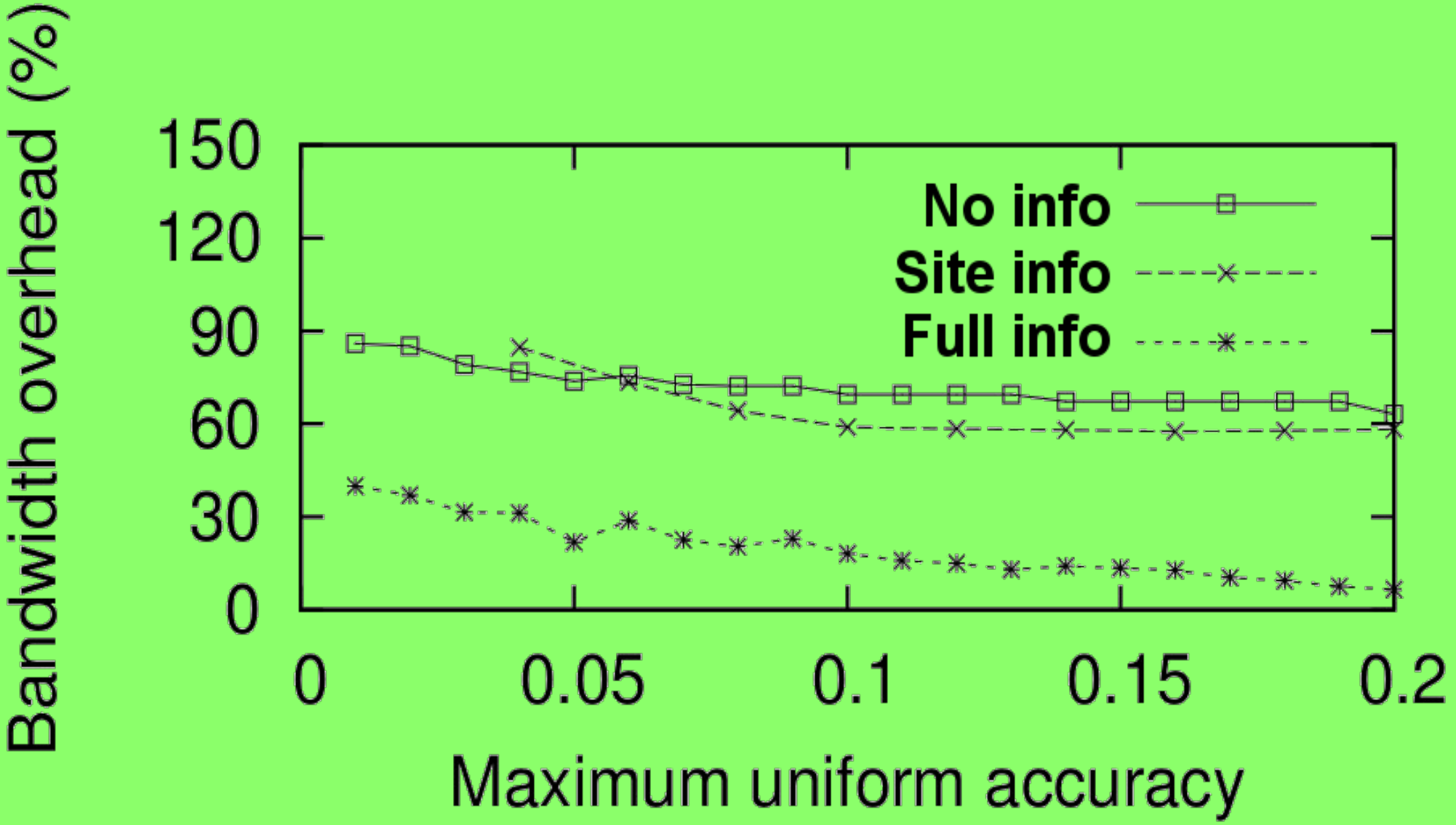
Full info

Lower overhead

More Practical

3. Full info:
Multiple covers
Compute at load

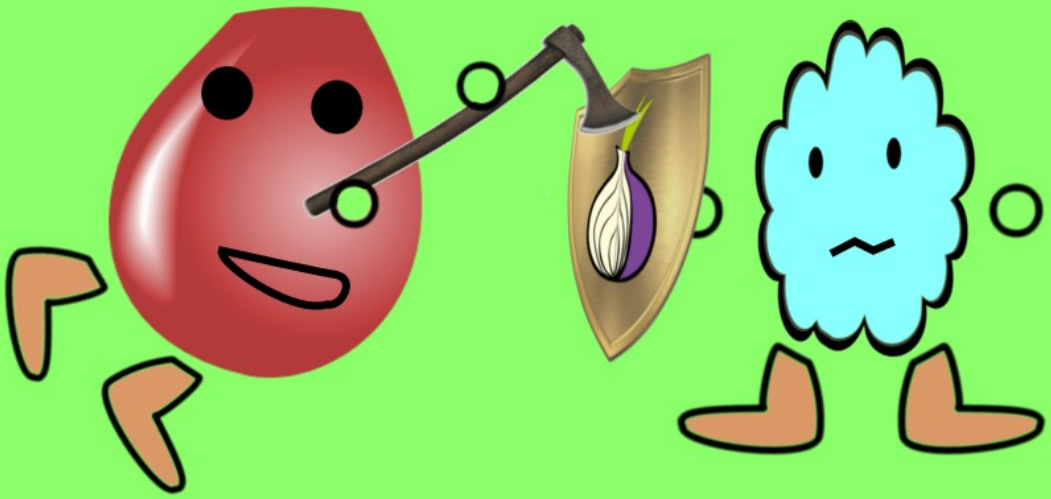
Results


Bandwidth overhead (%)
150
120
90
60
30
0
0
0.05
0.1
0.15
0.2
Maximum uniform accuracy
No info
Site info
Full info

Attack versus Defense

Defense	Overhead	Accuracy	Max accuracy
Traffic morphing	50%	0.82	/
HTTPOS (split)	5%	0.86	/
Decoy page	130%	0.30	/
BuFLO	190%	0.10	/
Provable (No info)	**96%**	**0.09**	**0.1**
Provable (Site info)	**59%**	**0.07**	**0.1**

Takeaway



Website Fingerprinting attacks can and will break unproven defenses.



Website Fingerprinting





Website Fingerprinting

Chew



Google

Defense testimonials

Luo 11

Extensive evaluation of HTTPOS on live web traffic shows that it can successfully prevent the state-of-the-art attacks from inferring private information from encrypted HTTP flows.

<6% accuracy

Cai 12

Our attack is able to defeat [Luo et al.]'s implementation of HTTPOS.

76% accuracy

Defense testimonials

Wright 11

We... thwart statistical traffic analysis algorithms by optimally morphing one class of traffic to look like another class.

5% accuracy

Lu 12

Traces of the ordering information remain even under traffic morphing and they can be extracted for identification.

78% accuracy

Defense testimonials

Dyer 12

[The defense] should, intuitively, defeat any [traffic analysis] classifier by removing all side-channel information.

Dyer 12

Surprisingly, [Panchenko et al.]'s classifier still identifies sites with an average accuracy of 27.3%.

27% accuracy