

Reward Based Decision Making

Alireza Soltani and Xiao-Jing Wang Physics Department and Volen Center for Complex Systems Brandeis University, Waltham, MA

Sloan-Swartz Meeting 2004, CSHL



Outline Adaptive Decision Making Spiking Network Model Reward Gated Plasticity Simulation Results Conclusion



Adaptive decision making and matching task
Spiking network model for decision making



● Outline	Adaptive decision makir
Adaptive Decision Making	•
Spiking Network Model Reward Gated Plasticity	Spiking network model f
Simulation Results	
Conclusion	Reward gated plasticity

on making and matching task

k model for decision making

Alireza Soltani and Xiao-Jing Wang July 24, 2004



● Outline	Ada
Adaptive Decision Making	
Spiking Network Model	
Reward Gated Plasticity	Spi
Simulation Results	
Conclusion	

- Adaptive decision making and matching task
- Spiking network model for decision making
- Reward gated plasticity
- Simulation results



Adaptive Decision Making

Real World

Matching Law

Remarks and Motive

Matching Task in Monkeys

Matching Behavior

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Adaptive Decision Making



Adaptive Decision	Making
-------------------	--------

Real World
Matching Law

- Remarks and Motive
- Matching Task in Monkeys
- Matching Behavior

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

There are many situations where the mapping between an action and its outcome is not fixed.



Adaptive Decis	ion Making
----------------	------------

- Real WorldMatching Law
- Remarks and Motive
- Matching Task in Monkeys
- Matching Behavior

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

There are many situations where the mapping between an action and its outcome is not fixed.

Social and economic interactions



● Outline	The
Adaptive Decision Making	
● Real World	betw
 Matching Law 	
Remarks and Motive	
 Matching Task in Monkeys 	
 Matching Behavior 	- 0-
Spiking Network Model	
Reward Gated Plasticity	
	<u> </u> Га
Simulation Results	■ Fo
Conclusion	

There are many situations where the mapping between an action and its outcome is not fixed.Social and economic interactions

Foraging behavior in natural environment



0	utl	ine	

Adaptive Decision Making
● Real World
 Matching Law
Remarks and Motive
 Matching Task in Monkeys
 Matching Behavior
Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

There are many situations where the mapping between an action and its outcome is not fixed.

Foraging behavior



Adaptive Decision Making

Real World

Matching Law

Remarks and Motive

Matching Task in Monkeys

Matching Behavior

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

The animal allocates its time between competing *choices* in a manner that it matches the relative *return* from those choices.

$$\frac{C_i}{C_1 + C_2 + \dots + C_n} = \frac{R_i}{R_1 + R_2 + \dots + R_n}$$

 $return = (reward magnitude) \times (rate)$

Herrnstein R.J., "The Matching Law, Papers in Psychology and Economics"



• Why do they match?

Outline

Adaptive Decision Making

Real World

Matching Law

Remarks and Motive

Matching Task in Monkeys

Matching Behavior

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion



Adaptive Decision Making

Real World

Matching Law

Remarks and Motive

Matching Task in Monkeys

Matching Behavior

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

• Why do they match?

• What is the local choice rule?



Outline Why do they match? Adaptive Decision Making Real World Matching Law Remarks and Motive Matching Task in Monkeys Matching Behavior Spiking Network Model Reward Gated Plasticity What is the neural basis of the matching behavior?

Simulation Results

Conclusion

Remarks and Motive



Adaptive Decision Making

Real World

Matching Law

Remarks and Motive

Matching Task in Monkeys

Matching Behavior

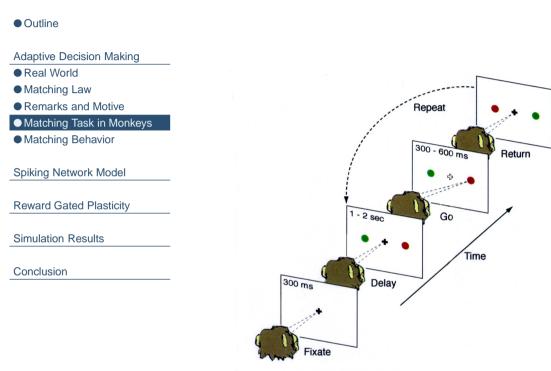
Spiking Network Model

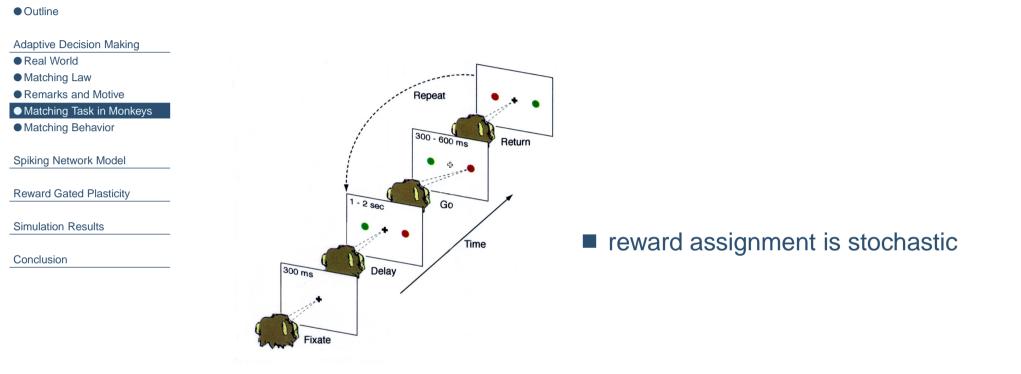
Reward Gated Plasticity

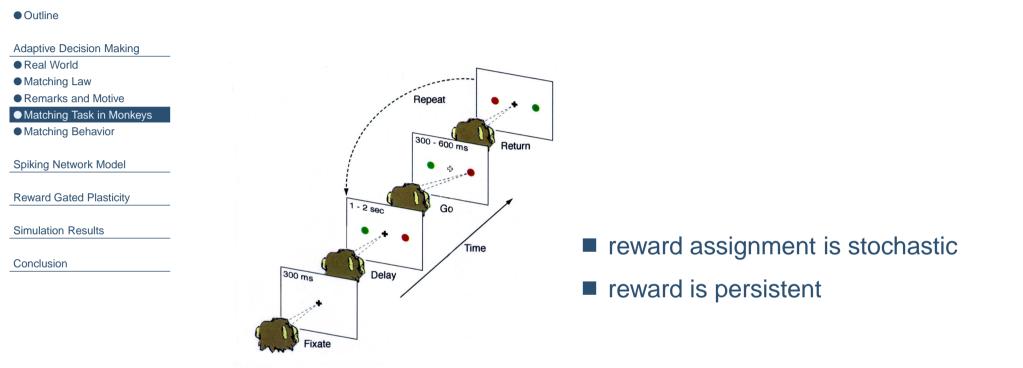
Simulation Results

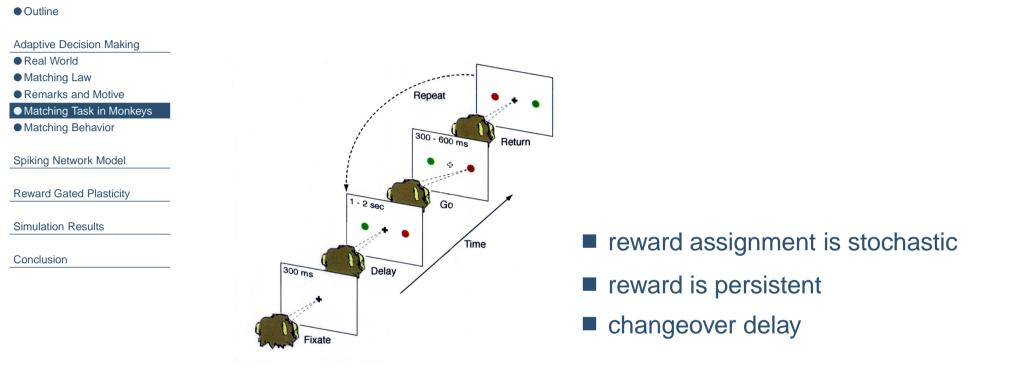
Conclusion

What is the neural basis of the matching behavior?

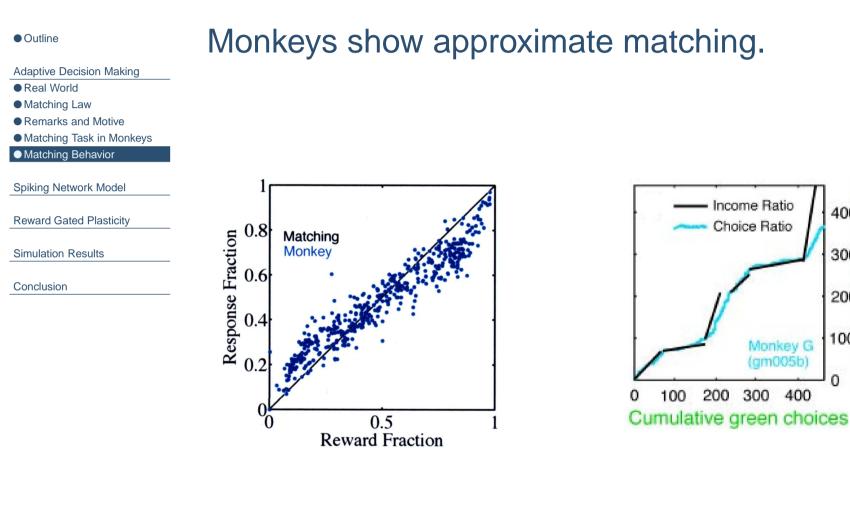












Sugrue L.P., Corrado G.S., Newsome W.T., Science 2004

400

300

200

100

C

Cumulative



Adaptive Decision Making

Spiking Network Model

Network Behavior

• Stochastic Response

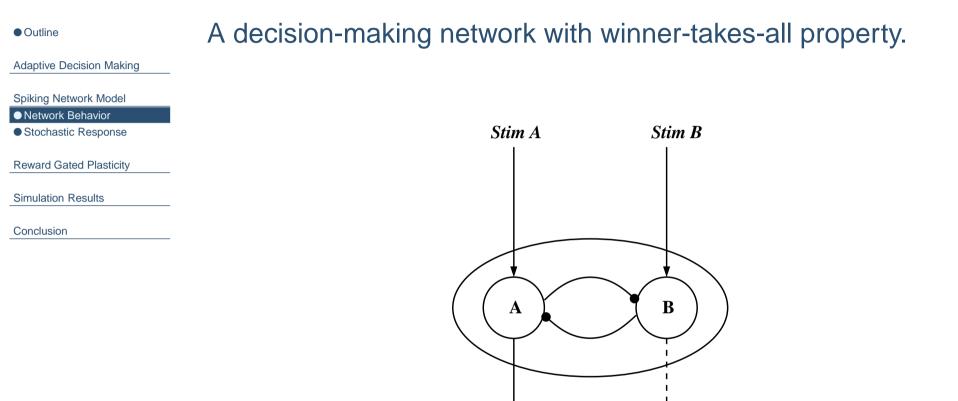
Reward Gated Plasticity

Simulation Results

Conclusion

Spiking Network Model







Adaptive Decision Making

Spiking Network Model

Network Behavior

Stochastic Response

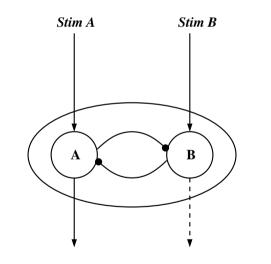
Reward Gated Plasticity

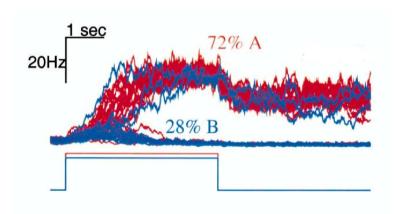
Simulation Results

Conclusion



- 1. Strength of the inputs
- 2. Intrinsic noise in the neurons spiking





Wang X.J., Neuron 2002



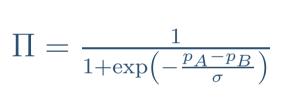
Outline
 Adaptive Decision Making
 Spiking Network Model
 Network Behavior
 Stochastic Response

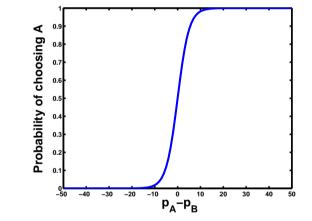
Reward Gated Plasticity

Simulation Results

Conclusion

The stochastic network behavior as a function of synaptic strengths is approximately sigmoidal.





 p_A : Probability of afferent synapses to population A being in potentiated state.

 σ : Sensitivity of the network to the biased inputs



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

• Three-factor Rule

• Learning Rule

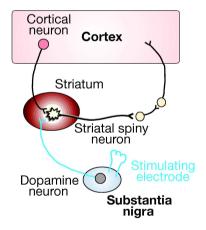
• Stochastic Learning

Simulation Results

Conclusion

Reward Gated Plasticity

● Outline	Three factors for dopamine-dependent plasticity:
Adaptive Decision Making	presynaptic activity, postsynaptic activity, dopamine (reward signal).
Spiking Network Model	
Reward Gated Plasticity	
 Three-factor Rule 	
Learning Rule	
Stochastic Learning	Cortical neuron Cortex
Simulation Results	
Conclusion	Striatum
	Striatal spiny



Reynolds J.N., Hyland B.I., Wickens J.R., Nature 2001 Reynolds J.N., Wickens J.R., Neural Network 2002

Outline Adaptive Decision Making Spiking Network Model	Three factors for dopamine-dependent pl presynaptic activity, postsynaptic activity,	,
Reward Gated Plasticity Three-factor Rule Learning Rule Stochastic Learning Simulation Results Conclusion	• $(pre \uparrow post \uparrow) + Dopamine \rightarrow LTP$	Cortical neuron Cortex Striatum Striatal spiny neuron Stimulating electrode

Reynolds J.N., Hyland B.I., Wickens J.R., Nature 2001 Reynolds J.N., Wickens J.R., Neural Network 2002

Dopamine (neuron

Substantia nigra

● Outline	Three factors for dopamine-dependent plasticity:
Adaptive Decision Making	presynaptic activity, postsynaptic activity, dopamine (reward signal).
Spiking Network Model	
Reward Gated Plasticity ● Three-factor Rule	
Learning RuleStochastic Learning	Cortical
Simulation Results	$ (pre \uparrow post \uparrow) + Dopamine \rightarrow LTP $
Conclusion	$ (pre \uparrow post \uparrow) + No Dopamine \rightarrow LTD $
	Striatal spiny neuron

Reynolds J.N., Hyland B.I., Wickens J.R., Nature 2001 Reynolds J.N., Wickens J.R., Neural Network 2002 electrode

Substantia nigra

Dopamine (neuron \bigcirc

● Outline	Three factors for dopamine-dependent p	plasticity:
Adaptive Decision Making	presynaptic activity, postsynaptic activity, dopamine (reward signal).	
Spiking Network Model		
Reward Gated Plasticity		
Three-factor Rule		
 Learning Rule Stochastic Learning 		Cortical
Simulation Results	$ (pre \uparrow post \uparrow) + Dopamine \rightarrow LTP $	neuron Cortex
Conclusion	$\blacksquare (pre \uparrow post \uparrow) + No \text{ Dopamine} \rightarrow LTD$	Striatum
	$\blacksquare (pre \uparrow post \downarrow) + \text{Dopamine} \rightarrow \text{LTD/LTP}$	Striatal spiny neuron

Reynolds J.N., Hyland B.I., Wickens J.R., Nature 2001 Reynolds J.N., Wickens J.R., Neural Network 2002

Dopamine (neuron \bigcirc

Substantia nigra

● Outline	Three factors for dopamine-dependent plasticity:	
Adaptive Decision Making	presynaptic activity, postsynaptic activity, dopamine (reward signal).	
Spiking Network Model		
Reward Gated Plasticity		
Three-factor RuleLearning Rule		
 Stochastic Learning 		Cortical neuron Cortex
Simulation Results	$ (pre \uparrow post \uparrow) + Dopamine \rightarrow LTP $	
Conclusion	$\blacksquare (pre \uparrow post \uparrow) + No \text{ Dopamine} \rightarrow LTD$	Striatum
	$\blacksquare (pre \uparrow post \downarrow) + Dopamine \rightarrow LTD/LTP$	Striatal aniau
	■ $(pre \uparrow post \downarrow)$ +No Dopamine \rightarrow No	Striatal spiny neuron Stimulating
	change	Dopamine neuron Substantia nigra

Reynolds J.N., Hyland B.I., Wickens J.R., Nature 2001 Reynolds J.N., Wickens J.R., Neural Network 2002

Learning Rule

Outline

Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Three-factor Rule

Learning Rule

Stochastic Learning

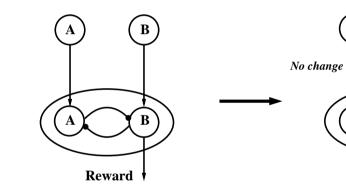
Simulation Results

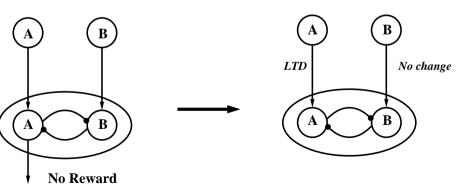
Conclusion

Assumptions:

Presynaptic side is always active.

Postsynaptic side is active only for the winner population.





B

B

Α

LTP



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Three-factor Rule

Learning Rule

Stochastic Learning

Simulation Results

Conclusion

If the condition for learning is met, only a fraction of synapses are changed.

Amit D.J., Fusi S., Neural Comp. 1994

Fusi S., Biol. Cybernetic 2002



Outline
 Adaptive Decision Making
 Spiking Network Model
 Reward Gated Plasticity
 Three-factor Rule
 Learning Rule
 Stochastic Learning
 Simulation Results
 Conclusion

If the condition for learning is met, only a fraction of synapses are changed.

--- Potentiation: $\Delta p_A = (1 - p_A)q_+$

Amit D.J., Fusi S., Neural Comp. 1994

Fusi S., Biol. Cybernetic 2002



Outline
 Adaptive Decision Making
 Spiking Network Model
 Reward Gated Plasticity
 Three-factor Rule
 Learning Rule
 Stochastic Learning

If the condition for learning is met, only a fraction of synapses are changed.

Simulation Results

Conclusion

• • Potentiation: $\Delta p_A = (1 - p_A)q_+$

Depression: $\Delta p_A = -p_A q_-$

Amit D.J., Fusi S., Neural Comp. 1994

Fusi S., Biol. Cybernetic 2002



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

- Global Behavior
- Local Behavior
- Reward Tracking
- Choice-Triggered Averages
- Robustness

Conclusion

Simulation Results



Model mimics the monkey's matching behavior.

Adaptive Decision Making

Spiking Network Model

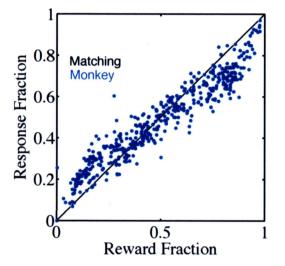
Reward Gated Plasticity

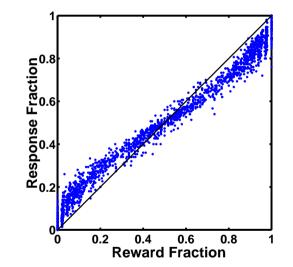
Simulation Results

Global Behavior

- Local Behavior
- Reward Tracking
- Choice-Triggered Averages
- Robustness

Conclusion





Sugrue L.P., Corrado G.S., Newsome W.T., Science 2004



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

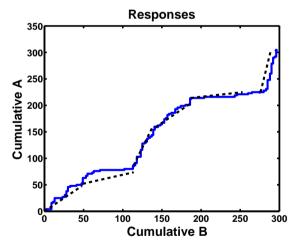
Simulation Results

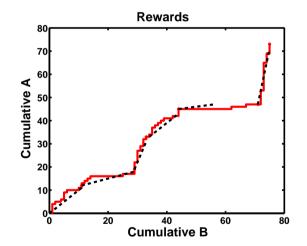
Global Behavior

- Local Behavior
- Reward Tracking
- Choice-Triggered Averages
- Robustness

Conclusion

Model is able to track changes in the reward schedule.







Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Global Behavior

Local Behavior

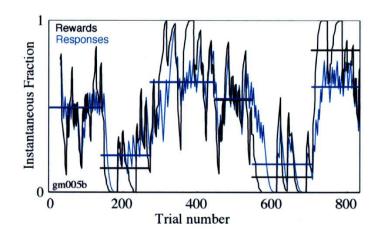
Reward Tracking

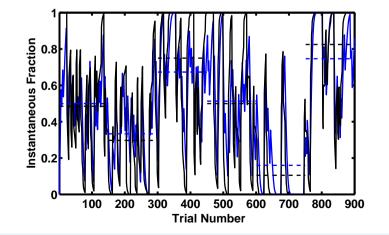
Choice-Triggered Averages

Robustness

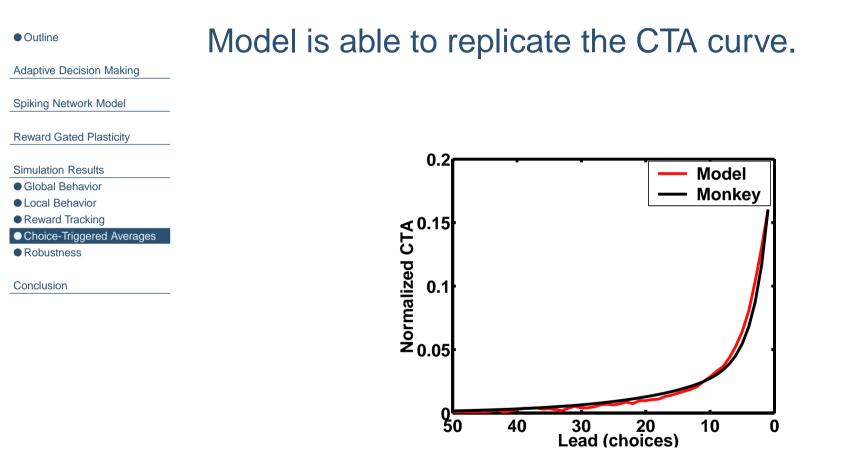
Conclusion

Model tracks random changes in the reward delivery.









Sugrue L.P., Corrado G.S., Newsome W.T., Science 2004



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Global Behavior

Local Behavior

Reward Tracking

Choice-Triggered Averages

Robustness

Conclusion



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Global Behavior

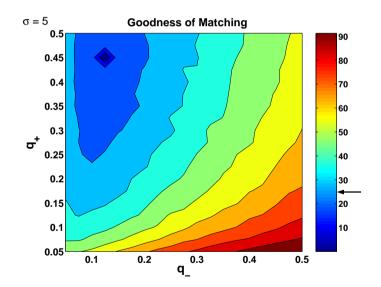
Local Behavior

Reward Tracking

• Choice-Triggered Averages

Robustness

Conclusion





Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Global Behavior

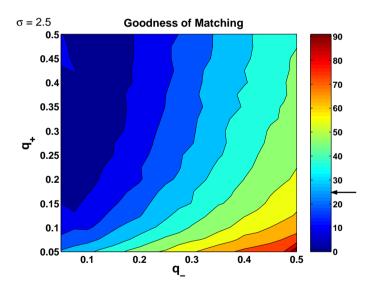
Local Behavior

Reward Tracking

Choice-Triggered Averages

Robustness

Conclusion





Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Global Behavior

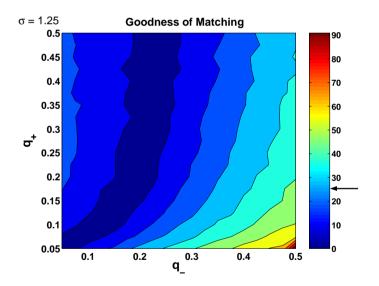
Local Behavior

Reward Tracking

Choice-Triggered Averages

Robustness

Conclusion





Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Global Behavior

Local Behavior

Reward Tracking

Choice-Triggered Averages

Robustness

Conclusion

Model shows matching behavior and high reward harvesting rate for a wide range of parameters.

Model is able to perform the matching task for different overall reward rates.



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Conclusion

Acknowledgment

Conclusion



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Conclusion

Acknowledgment

An adaptive decision-making network can be modeled using the main characteristics of working memory circuits.



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Conclusion

Acknowledgment

An adaptive decision-making network can be modeled using the main characteristics of working memory circuits.

Our model is able to replicate all of the behavioral data.



Outline
 Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

ConclusionAcknowledgment

- An adaptive decision-making network can be modeled using the main characteristics of working memory circuits.
- Our model is able to replicate all of the behavioral data.
- Functional form of choice dependence on reward history can be captured using particular values of parameters.



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Conclusion
 Acknowledgment

An adaptive decision-making network can be modeled using the main characteristics of working memory circuits.

Our model is able to replicate all of the behavioral data.

Functional form of choice dependence on reward history can be captured using particular values of parameters.

The simplest form of learning rule is very robust against changes in the network and environment.



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Conclusion
 Acknowledgment

An adaptive decision-making network can be modeled using the main characteristics of working memory circuits.

Our model is able to replicate all of the behavioral data.

Functional form of choice dependence on reward history can be captured using particular values of parameters.

The simplest form of learning rule is very robust against changes in the network and environment.

Prediction ...



S
L
Wang

Stefano Fusi Leo Sugrue Nanglab members



Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Network Architecture

• Full Network

Introducing Changeover Delay

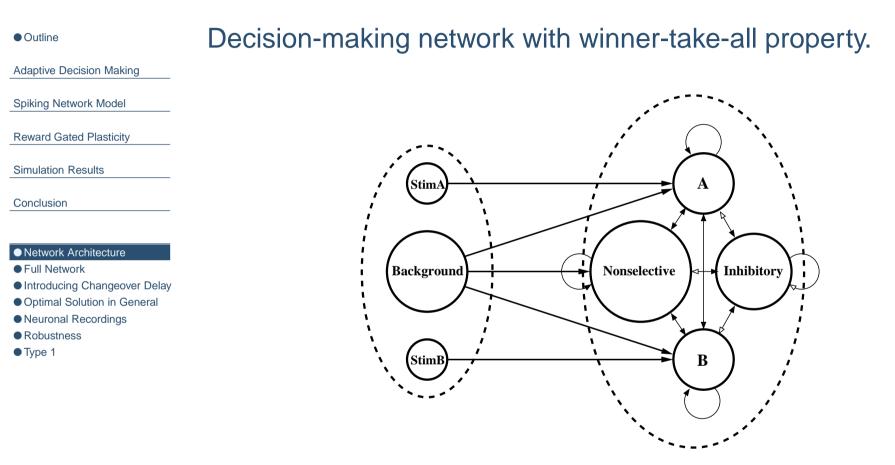
• Optimal Solution in General

Neuronal Recordings

Robustness

• Type 1



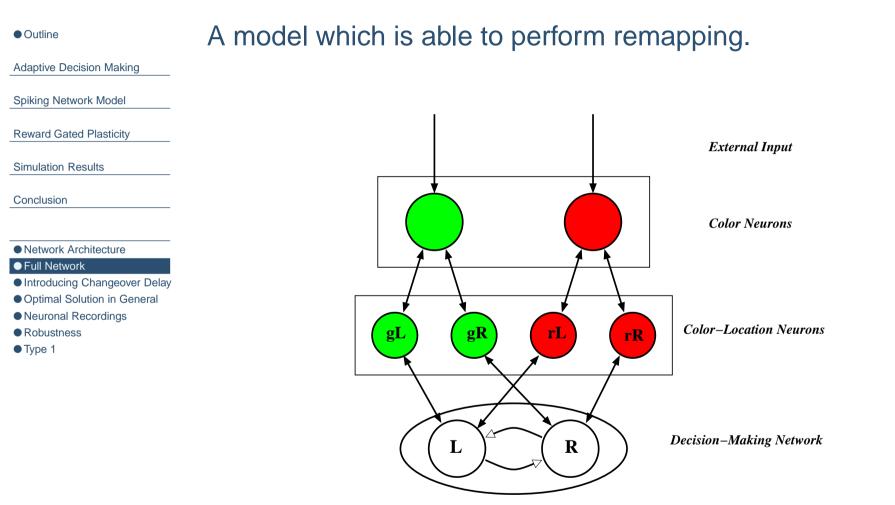






Wang X.J., Neuron 2002





Introducing Changeover Delay

Outline

Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Network Architecture

Full Network

Introducing Changeover Delay

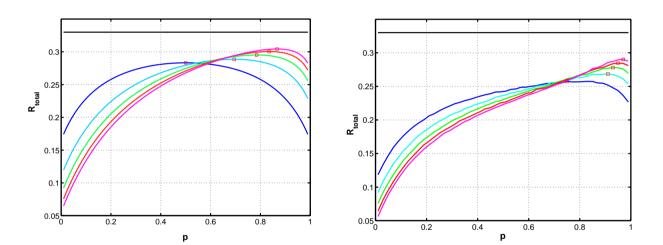
• Optimal Solution in General

Neuronal Recordings

Robustness

• Type 1

Introducing changeover delay, results in steeper total reward harvesting rate.



Optimal Solution in General

Outline
Adaptive Decision Making
Adaptive Decision Making
After Spiking Network Model
Option
Reward Gated Plasticity
Simulation Results
Conclusion
Onclusion
Network Architecture
Full Network
Introducing Changeover Delay
Optimal Solution in General
Neuronal Recordings
Robustness

• Type 1

In any matching task where the reward assignment on targets are independent of each other, a form of matching is an optimal solution.

Optimal Solution in General

Outline

Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Network Architecture

Full Network

Introducing Changeover Delay

Optimal Solution in General

Neuronal Recordings

Robustness

• Type 1

In any matching task where the reward assignment on targets are independent of each other, a form of matching is an optimal solution.

$$\frac{\partial R_1}{\partial p_1} = \frac{\partial R_2}{\partial p_2}$$

Optimal Solution in General

Outline

Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Network Architecture

Full Network

Introducing Changeover Delay

Optimal Solution in General

Neuronal Recordings

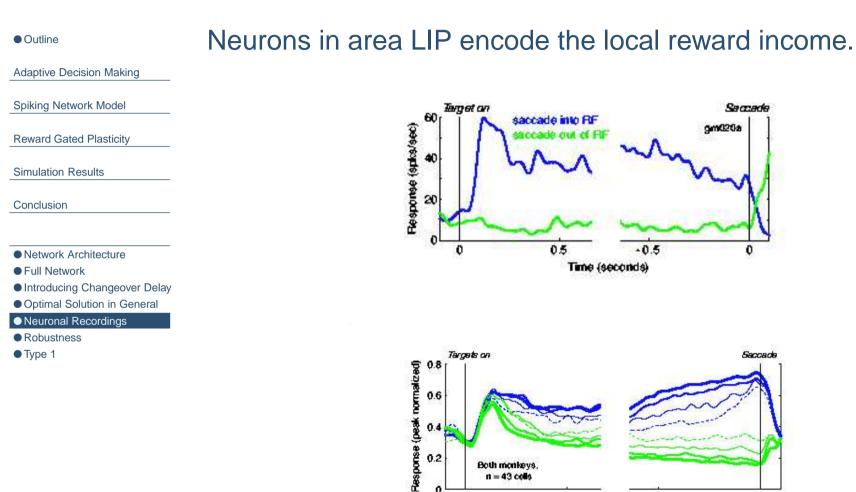
Robustness

• Type 1

In any matching task where the reward assignment on targets are independent of each other, a form of matching is an optimal solution.

 $\frac{\partial R_1}{\partial p_1} = \frac{\partial R_2}{\partial p_2}$ $\frac{p_{opt}}{1 - p_{opt}} = \sqrt{\frac{T_1'}{T_2'}} \times \frac{R_1}{R_2}$





n = 43 cells

0.5

Time (seconds)

- 0.5

0

0

Alireza Soltani and Xiao-Jing Wang July 24, 2004

Reward Based Decision Making - p. 30/32

Saccade

0



● Outline	Moc
Adaptive Decision Making	para
Spiking Network Model	•
Reward Gated Plasticity	
Simulation Results	
Conclusion	
Network Architecture	
Full Network	
Introducing Changeover Delay	
 Optimal Solution in General 	
Neuronal Recordings	
● Robustness	

• Type 1

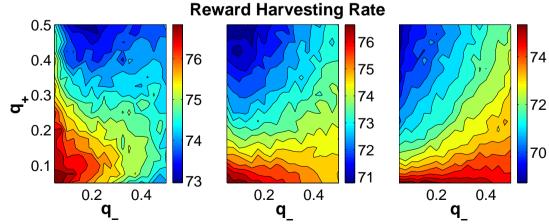
del is able to perform the task for different network ameters and different overall reward rates.



Outline Model is able parameters a
 Adaptive Decision Making parameters a
 Spiking Network Model
 Reward Gated Plasticity
 Simulation Results
 Onformation Conclusion
 Network Architecture
 Full Network
 Optimal Solution in General
 Neuronal Recordings
 Type 1

Model is able to perform the task for different network parameters and different overall reward rates.

Goodness of Matching σ $\sigma = 1.25$ $\sigma = 2.5$ $\sigma = 5$ 0.5 60 80 0.4 50 60 40 60 40 30 0.2 40 20 20 0.1 10 0.2 0.2 0.2 0.4 0.4 0.4





Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Network Architecture

Full Network

Introducing Changeover Delay

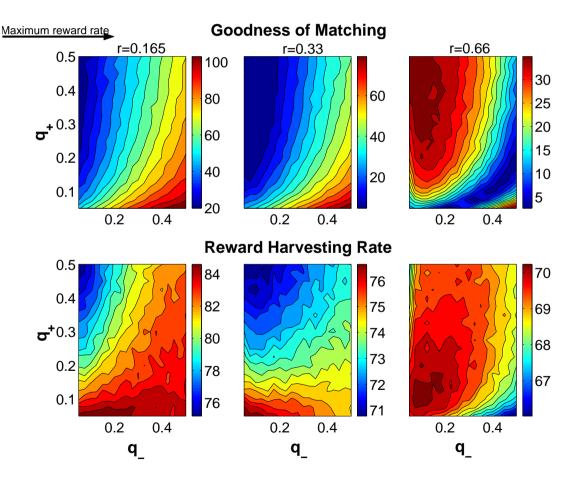
Optimal Solution in General

Neuronal Recordings

Robustness

• Type 1

Model is able to perform the task for different network parameters and different overall reward rates.





Not so robust.

Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Network Architecture

Full Network

Introducing Changeover Delay

• Optimal Solution in General

Neuronal Recordings

Robustness

● Type 1



Not so robust.

Adaptive Decision Making

Spiking Network Model

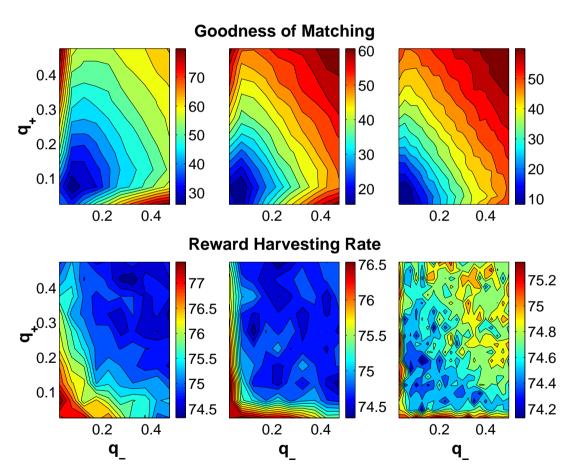
Reward Gated Plasticity

Simulation Results

Conclusion

- Network Architecture
- Full Network
- Introducing Changeover Delay
- Optimal Solution in General
- Neuronal Recordings
- Robustness

● Type 1





Not so robust.

Adaptive Decision Making

Spiking Network Model

Reward Gated Plasticity

Simulation Results

Conclusion

Network Architecture

• Full Network

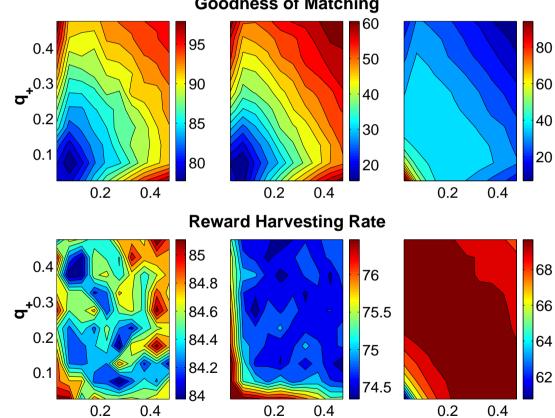
Introducing Changeover Delay

• Optimal Solution in General

Neuronal Recordings

Robustness

● Type 1



q_

q_

Goodness of Matching

q_