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 and matching task

Outline

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

Outline

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

Outline

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

Adaptive Decision Making

Real World

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

Real World

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

- Social and economic interactions

Real World

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

Real World

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

Foraging behavior

Matching Law

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

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

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

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

Remarks and Motive

- Why do they match?

Remarks and Motive

- Why do they match?

- What is the local choice rule?

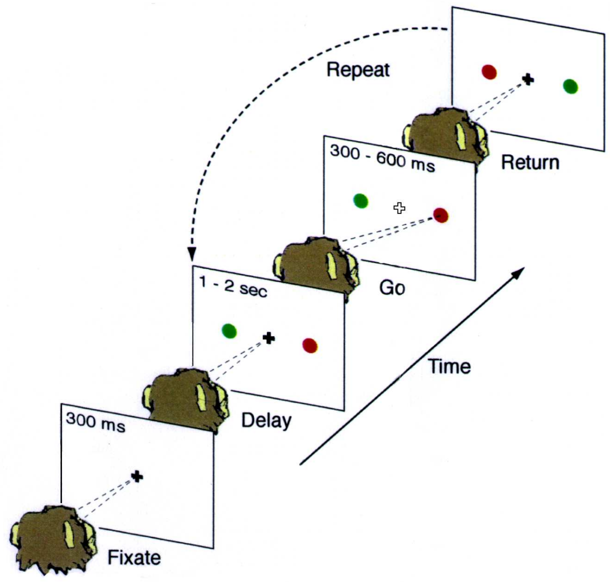
Remarks and Motive

- Why do they match?
- What is the local choice rule?
- What is the neural basis of the matching behavior?

Remarks and Motive

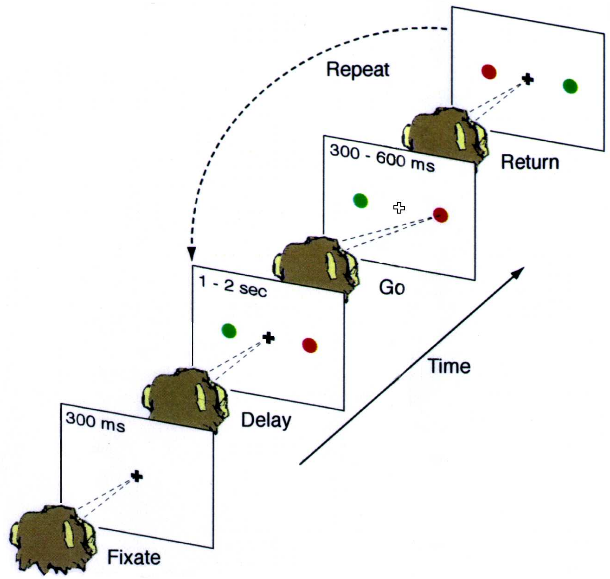
What is the neural basis of the matching behavior?

Matching Task in Monkeys



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

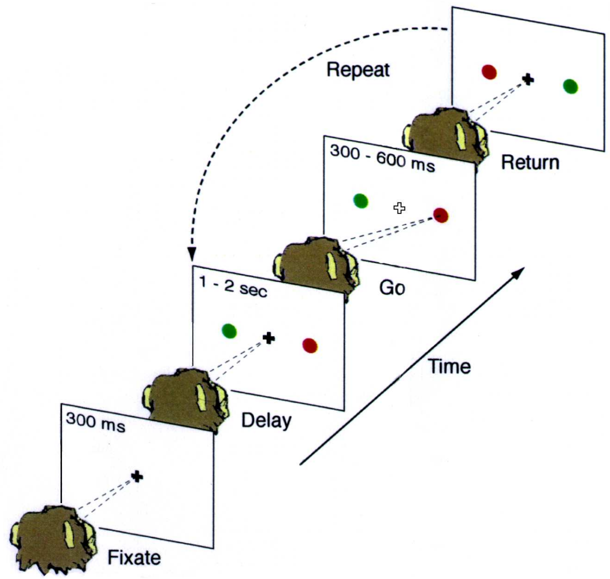
Matching Task in Monkeys



- reward assignment is stochastic

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

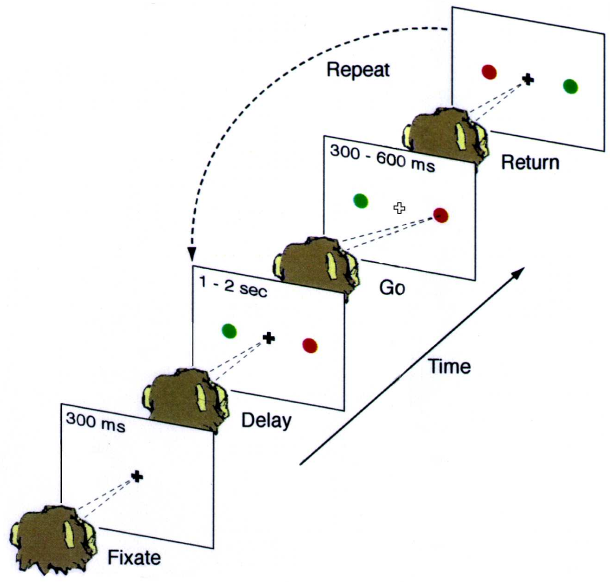
Matching Task in Monkeys



- reward assignment is stochastic
- reward is persistent

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

Matching Task in Monkeys

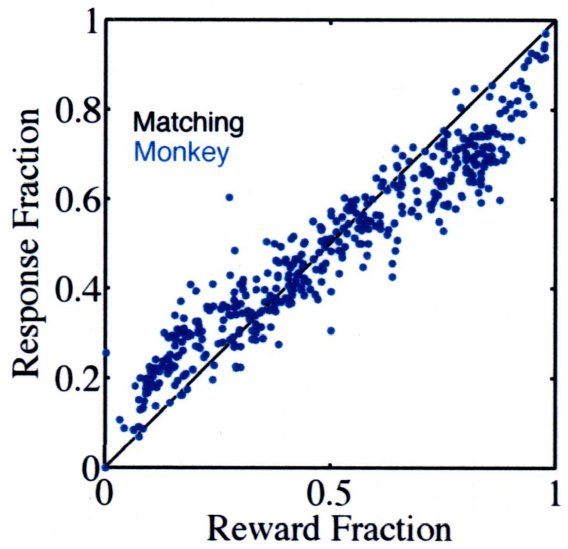


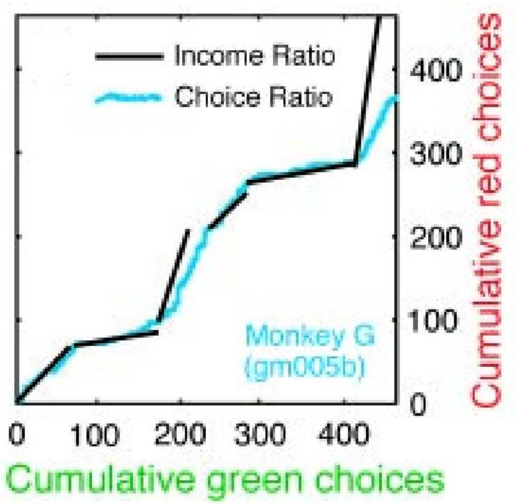
- reward assignment is stochastic
- reward is persistent
- changeover delay

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

Matching Behavior

Monkeys show approximate matching.



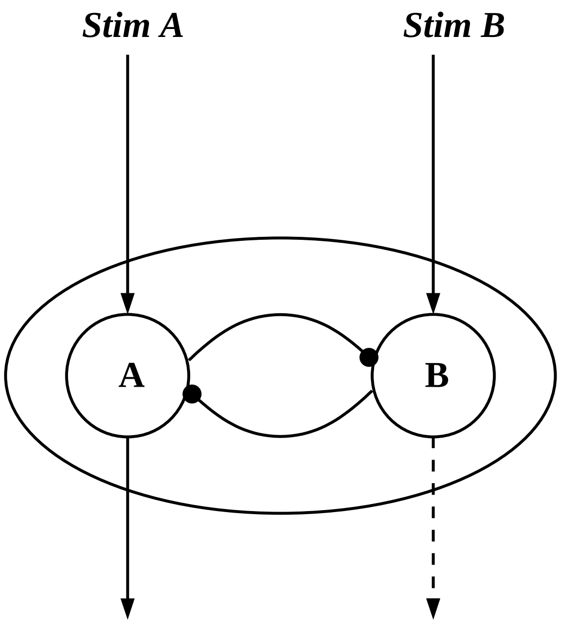


Sugrue L.P., Corrado G.S., Newsome W.T., Science 2004

Spiking Network Model

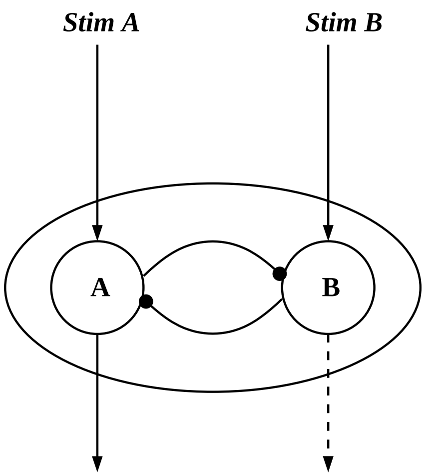
Network Behavior

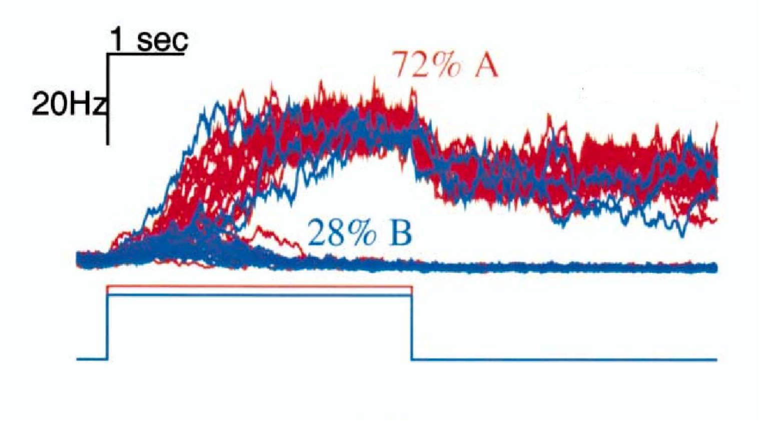
A decision-making network with winner-takes-all property.



Network Behavior

The network choice is determined by:
1. Strength of the inputs
2. Intrinsic noise in the neurons spiking



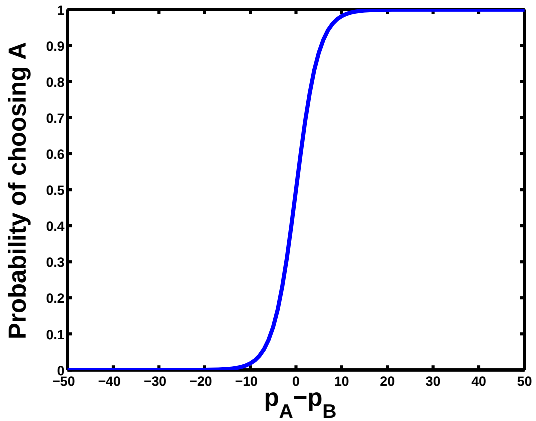


Wang X.J., Neuron 2002

Stochastic Response

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

$$\Pi = \frac{1}{1+\exp\left(-\frac{p_A - p_B}{\sigma}\right)}$$



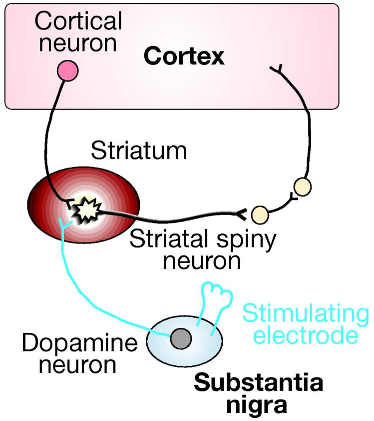
p_A : Probability of afferent synapses to population A being in potentiated state.

σ : Sensitivity of the network to the biased inputs

Reward Gated Plasticity

Three-factor Rule

Three factors for dopamine-dependent plasticity:
presynaptic activity, postsynaptic activity, dopamine (reward signal).



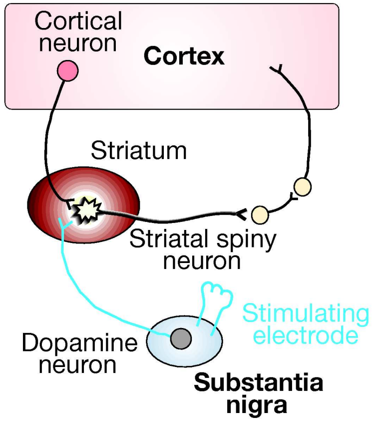
Reynolds J.N., Hyland B.I., Wickens J.R., Nature 2001

Reynolds J.N., Wickens J.R., Neural Network 2002

Three-factor Rule

Three factors for dopamine-dependent plasticity:
presynaptic activity, postsynaptic activity, dopamine (reward signal).

- $(pre \uparrow post \uparrow)+$ Dopamine $\rightarrow$ LTP



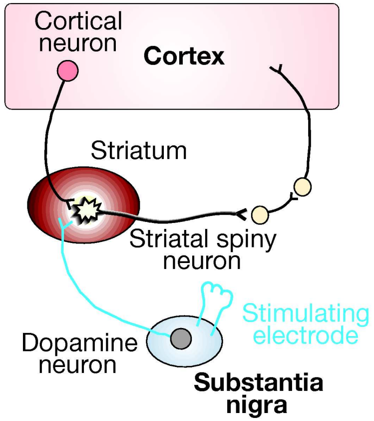
Reynolds J.N., Hyland B.I., Wickens J.R., Nature 2001

Reynolds J.N., Wickens J.R., Neural Network 2002

Three-factor Rule

Three factors for dopamine-dependent plasticity:
presynaptic activity, postsynaptic activity, dopamine (reward signal).

- $(pre \uparrow post \uparrow)+$ Dopamine $\rightarrow$ LTP
- $(pre \uparrow post \uparrow)+$No Dopamine $\rightarrow$ LTD



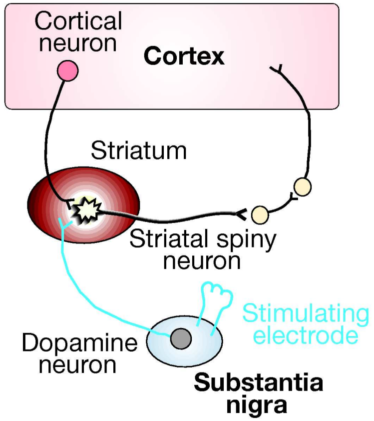
Reynolds J.N., Hyland B.I., Wickens J.R., Nature 2001

Reynolds J.N., Wickens J.R., Neural Network 2002

Three-factor Rule

Three factors for dopamine-dependent plasticity:
presynaptic activity, postsynaptic activity, dopamine (reward signal).

- $(pre \uparrow post \uparrow)+$ Dopamine $\rightarrow$ LTP
- $(pre \uparrow post \uparrow)+$No Dopamine $\rightarrow$ LTD
- $(pre \uparrow post \downarrow)+$ Dopamine $\rightarrow$ LTD/LTP



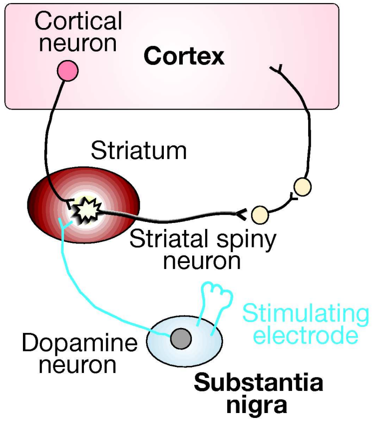
Reynolds J.N., Hyland B.I., Wickens J.R., Nature 2001

Reynolds J.N., Wickens J.R., Neural Network 2002

Three-factor Rule

Three factors for dopamine-dependent plasticity:
presynaptic activity, postsynaptic activity, dopamine (reward signal).

- $(pre \uparrow post \uparrow)+$ Dopamine $\rightarrow$ LTP
- $(pre \uparrow post \uparrow)+$No Dopamine $\rightarrow$ LTD
- $(pre \uparrow post \downarrow)+$ Dopamine $\rightarrow$ LTD/LTP
- $(pre \uparrow post \downarrow)+$No Dopamine $\rightarrow$No change



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

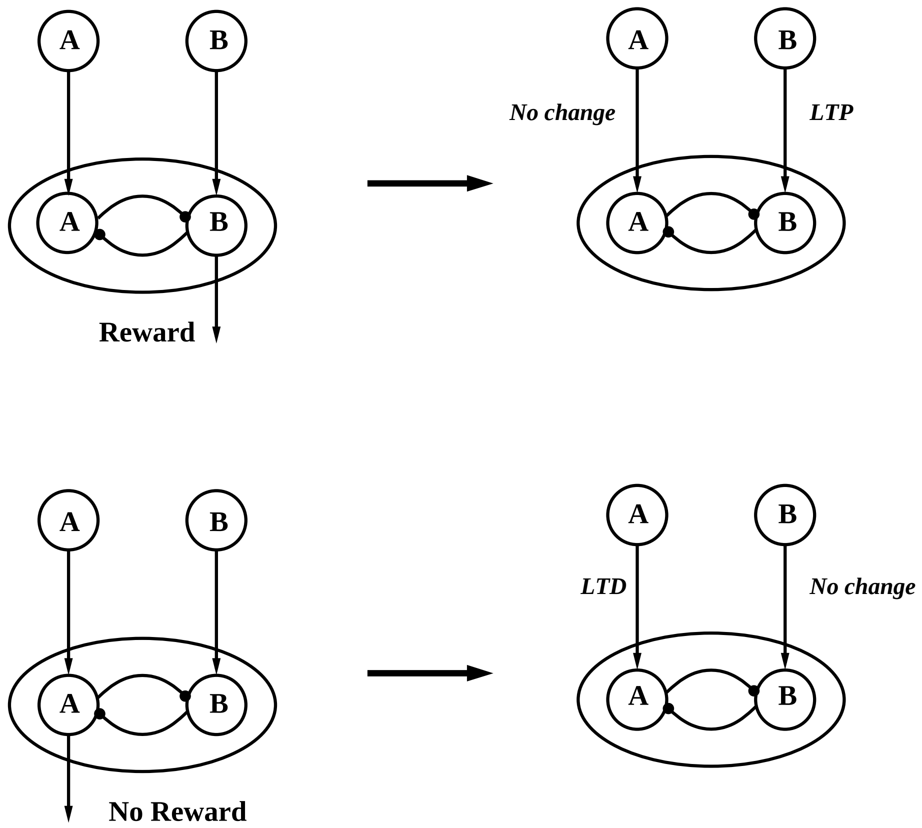
Reynolds J.N., Wickens J.R., Neural Network 2002

Learning Rule

Assumptions:

Presynaptic side is always active.

Postsynaptic side is active only for the winner population.



Stochastic Learning

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

Stochastic Learning

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

Stochastic Learning

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

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

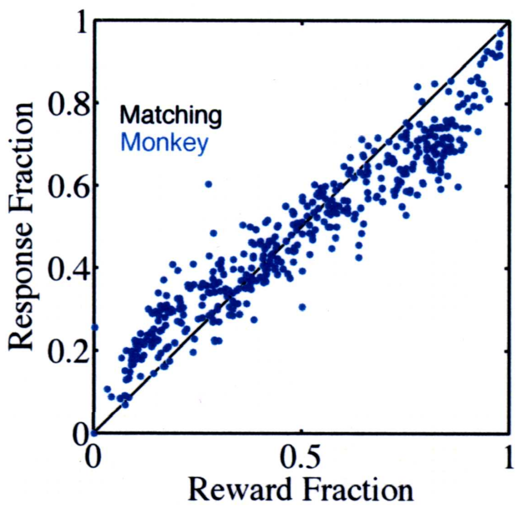
Amit D.J., Fusi S., Neural Comp. 1994

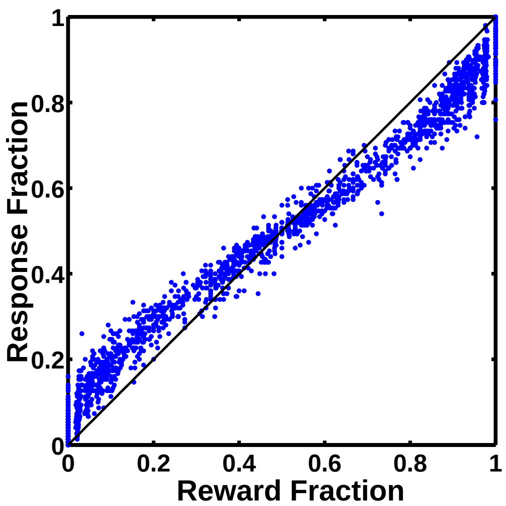
Fusi S.,Biol. Cybernetic 2002

Simulation Results

Global Behavior

Model mimics the monkey's matching behavior.

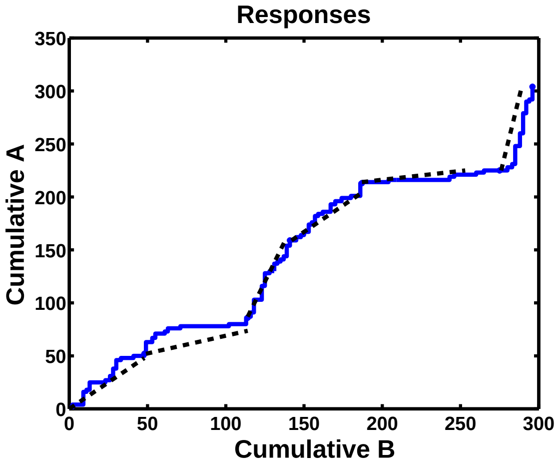


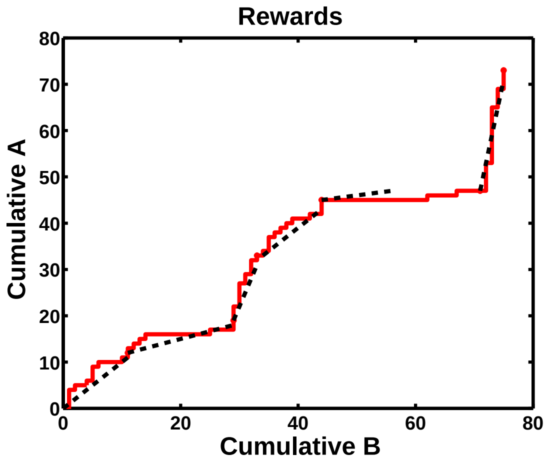


Sugrue L.P., Corrado G.S., Newsome W.T., Science 2004

Local Behavior

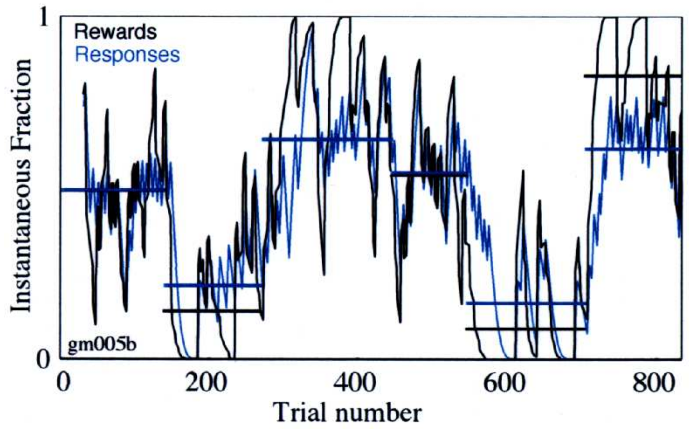
Model is able to track changes in the reward schedule.

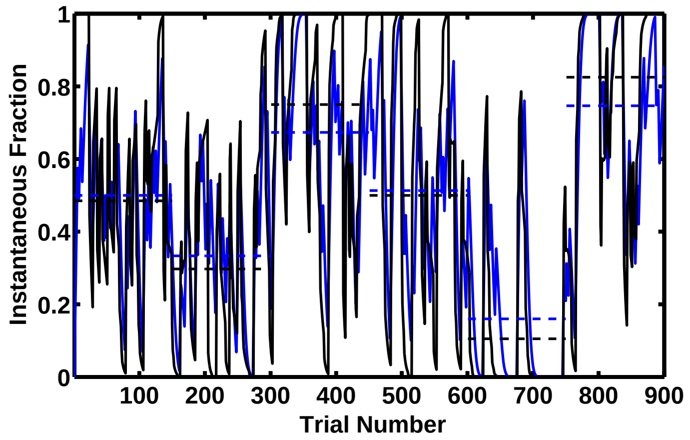




Reward Tracking

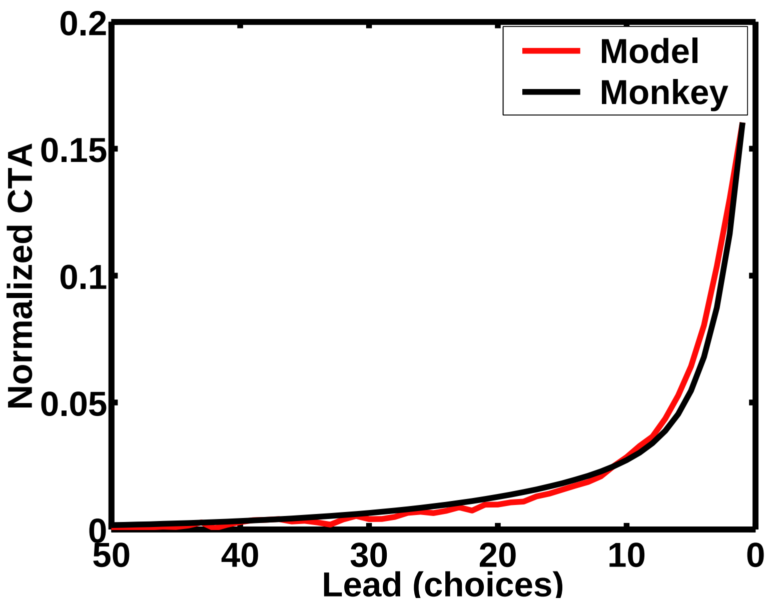
Model tracks random changes in the reward delivery.





Choice-Triggered Averages

Model is able to replicate the CTA curve.



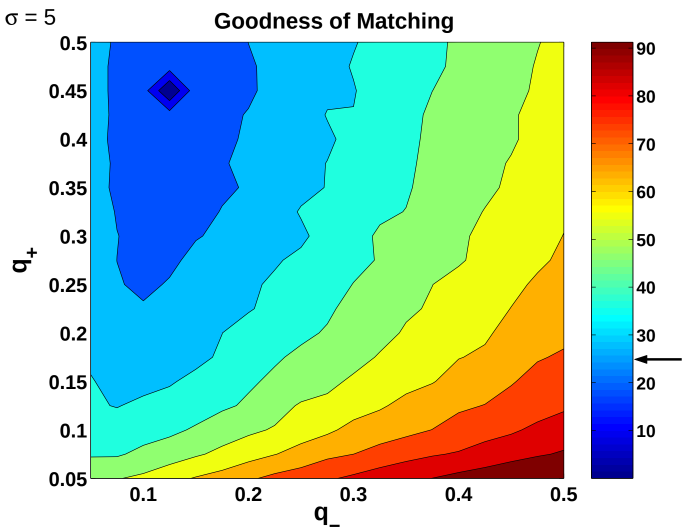
Sugrue L.P., Corrado G.S., Newsome W.T., Science 2004

Robustness

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

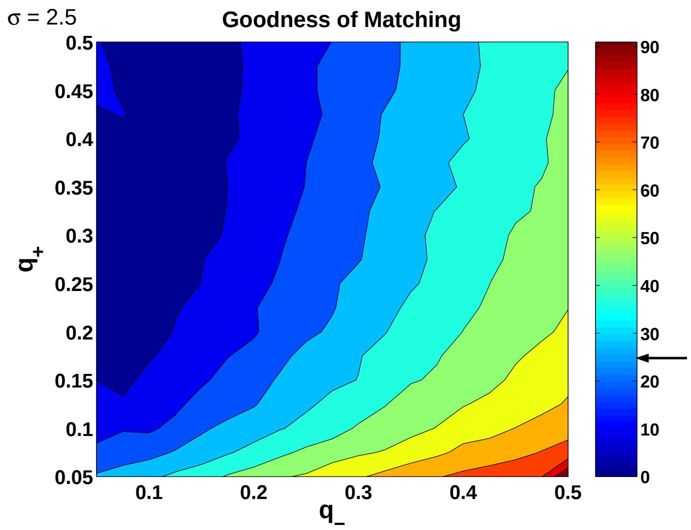
Robustness

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



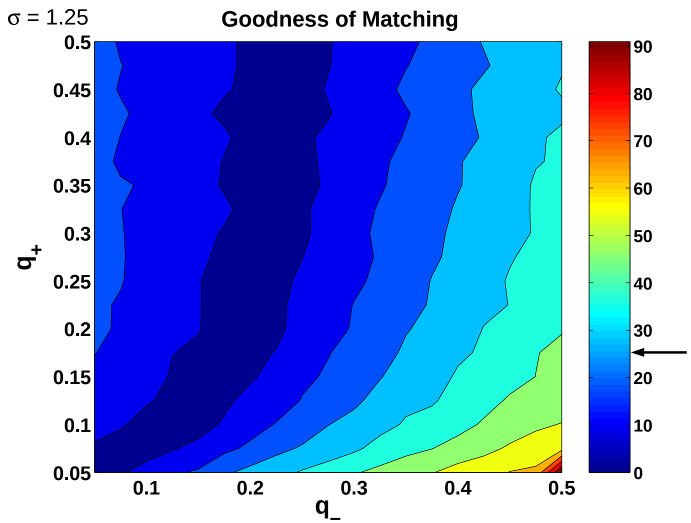
Robustness

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



Robustness

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



Robustness

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.

Conclusion

Conclusion

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

Conclusion

- 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.

Conclusion

- 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.

Conclusion

- 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.

Conclusion

- 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 ...

Acknowledgment

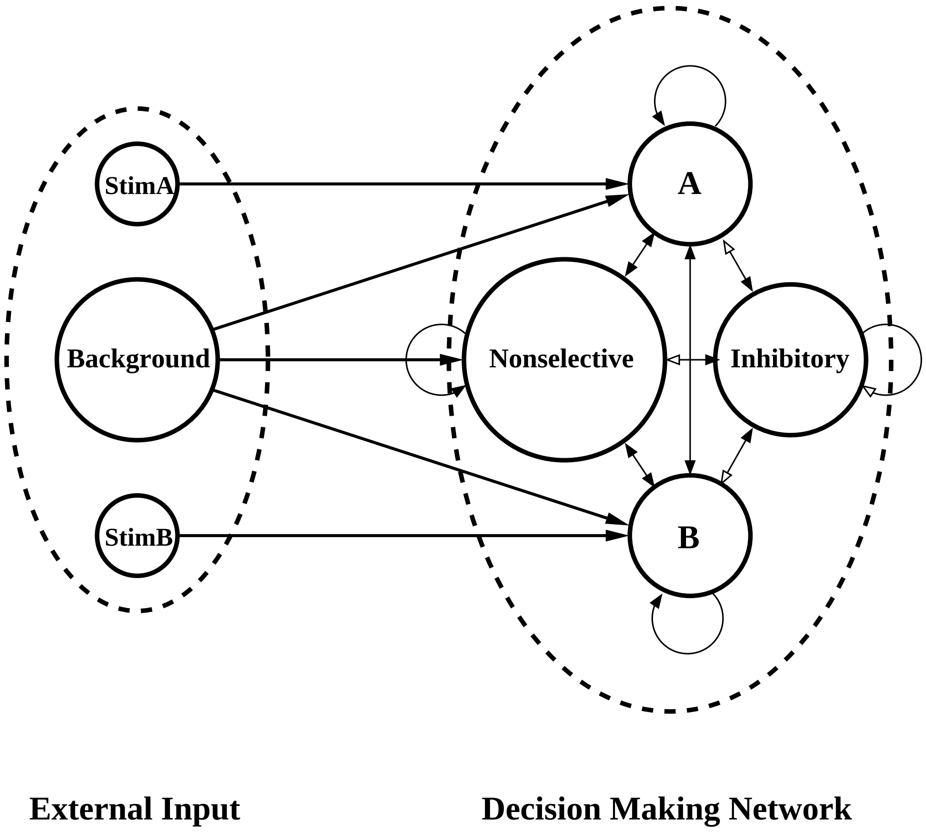
Stefano Fusi

Leo Sugrue

Wanglab members

Network Architecture

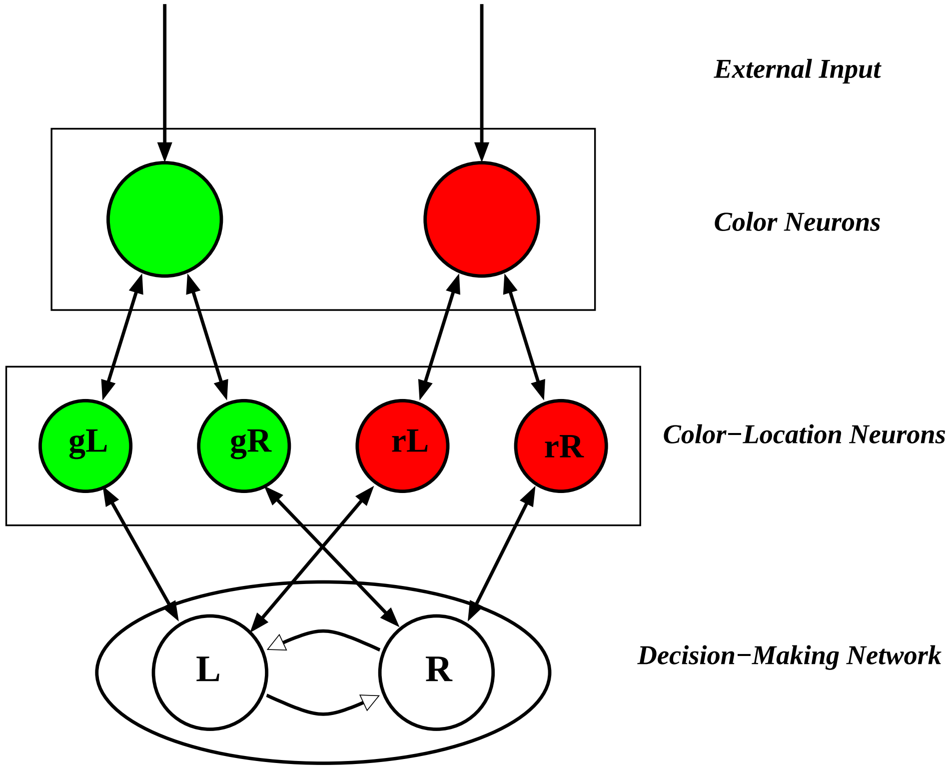
Decision-making network with winner-take-all property.



Wang X.J., Neuron 2002

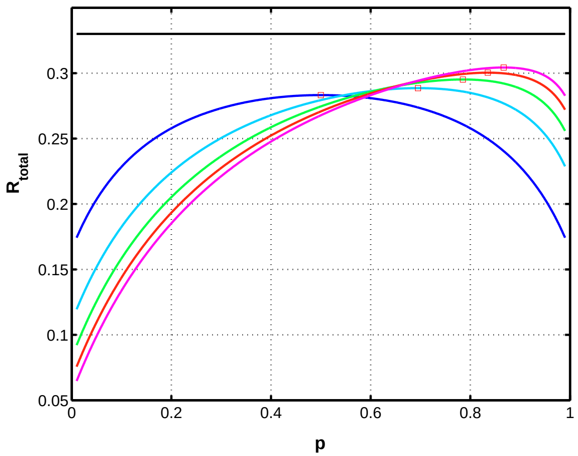
Full Network

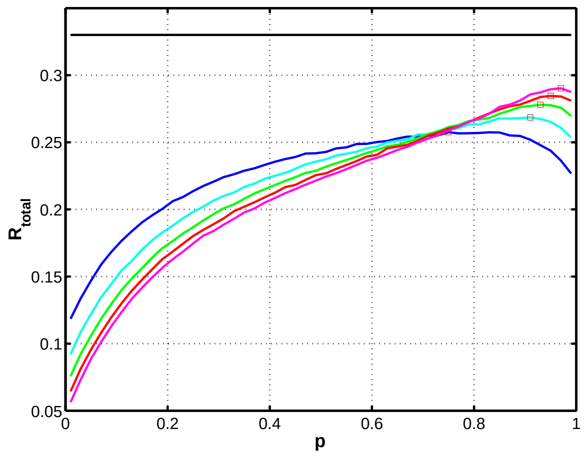
A model which is able to perform remapping.



Introducing Changeover Delay

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





Optimal Solution in General

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

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

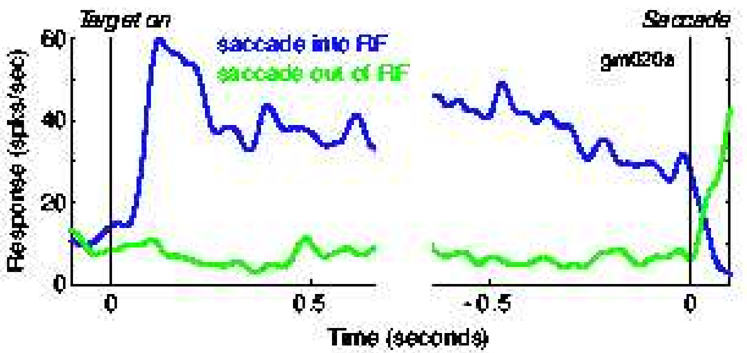
In any matching task where the reward assignment on targets are independent of each other, a form of matching is an optimal solution.

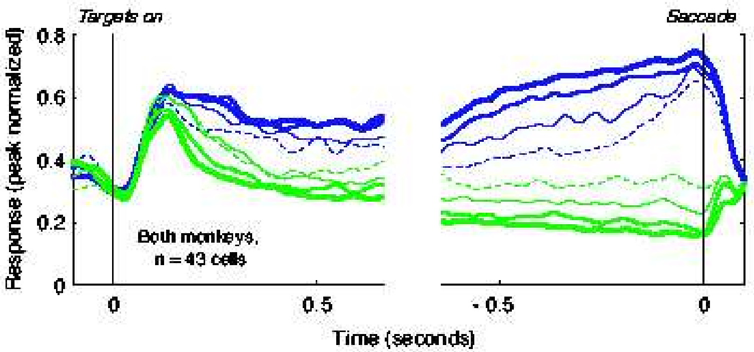
$$\frac{\partial R_1}{\partial p_1} = \frac{\partial R_2}{\partial p_2}$$

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

Neuronal Recordings

Neurons in area LIP encode the local reward income.



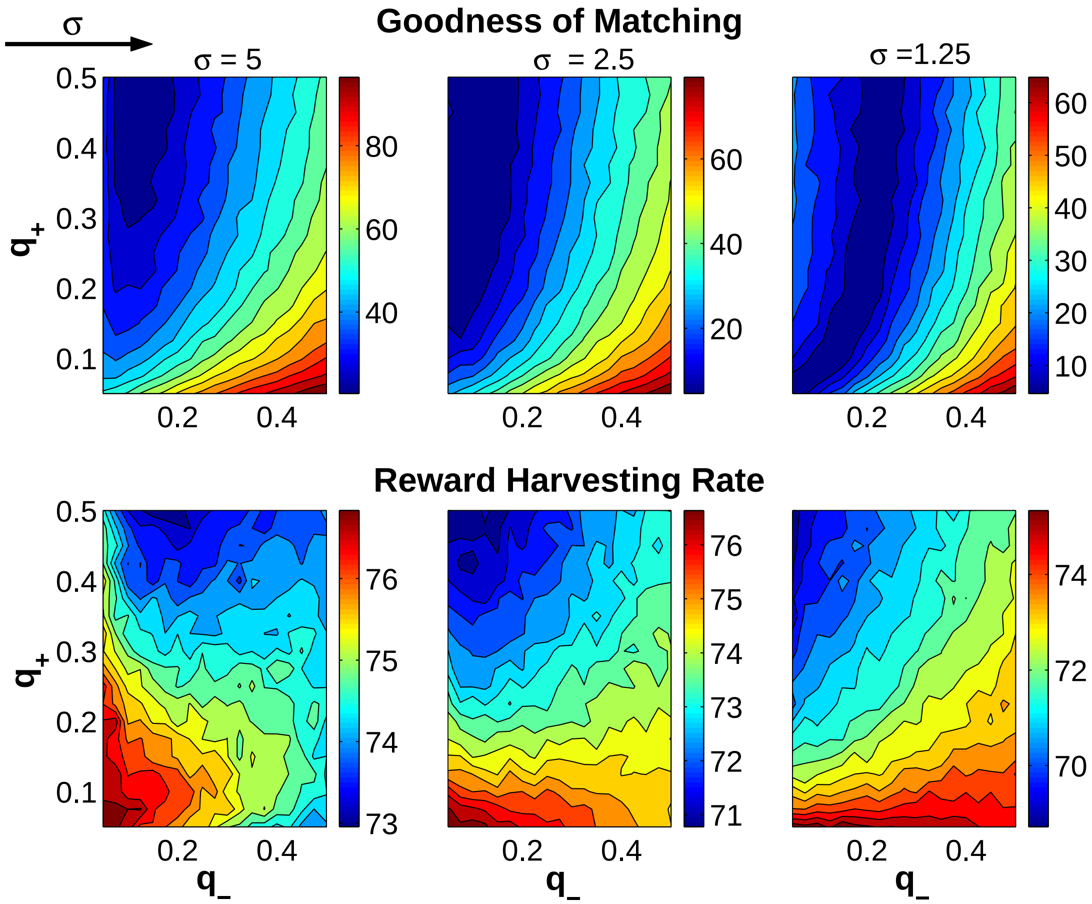


Robustness

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

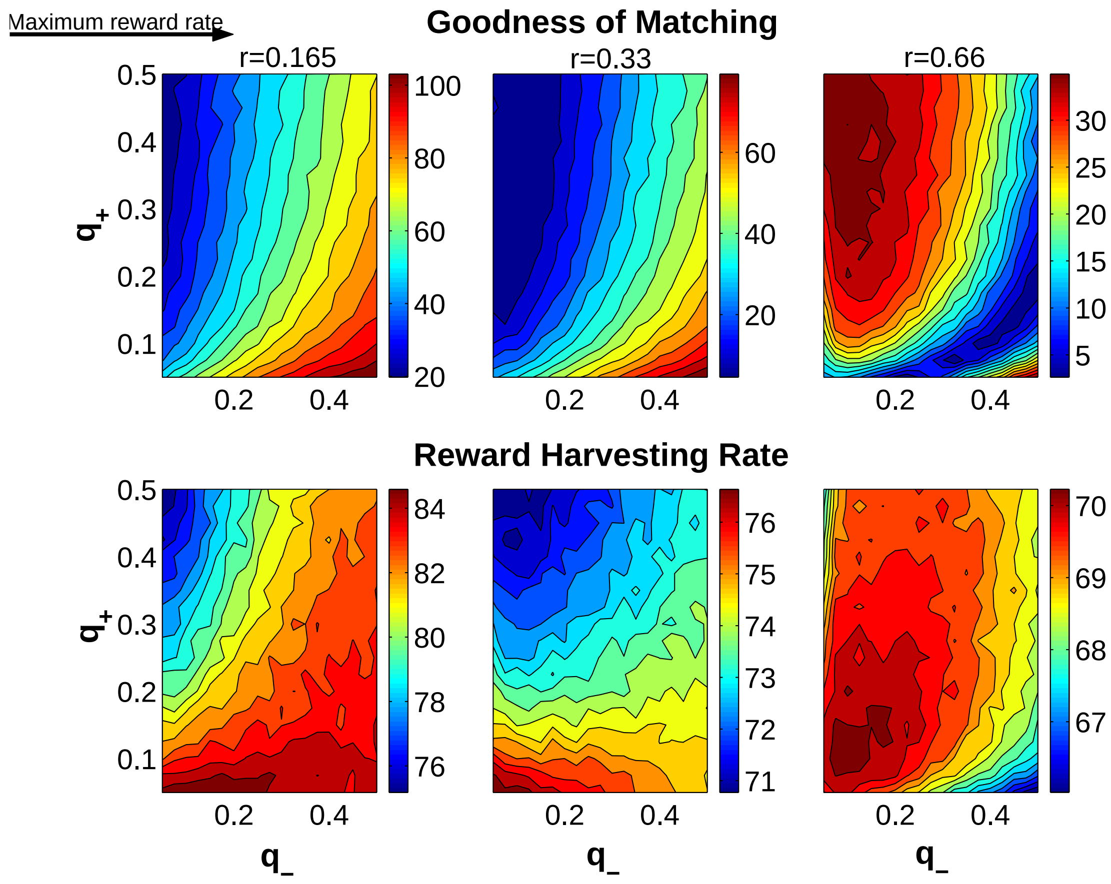
Robustness

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



Robustness

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

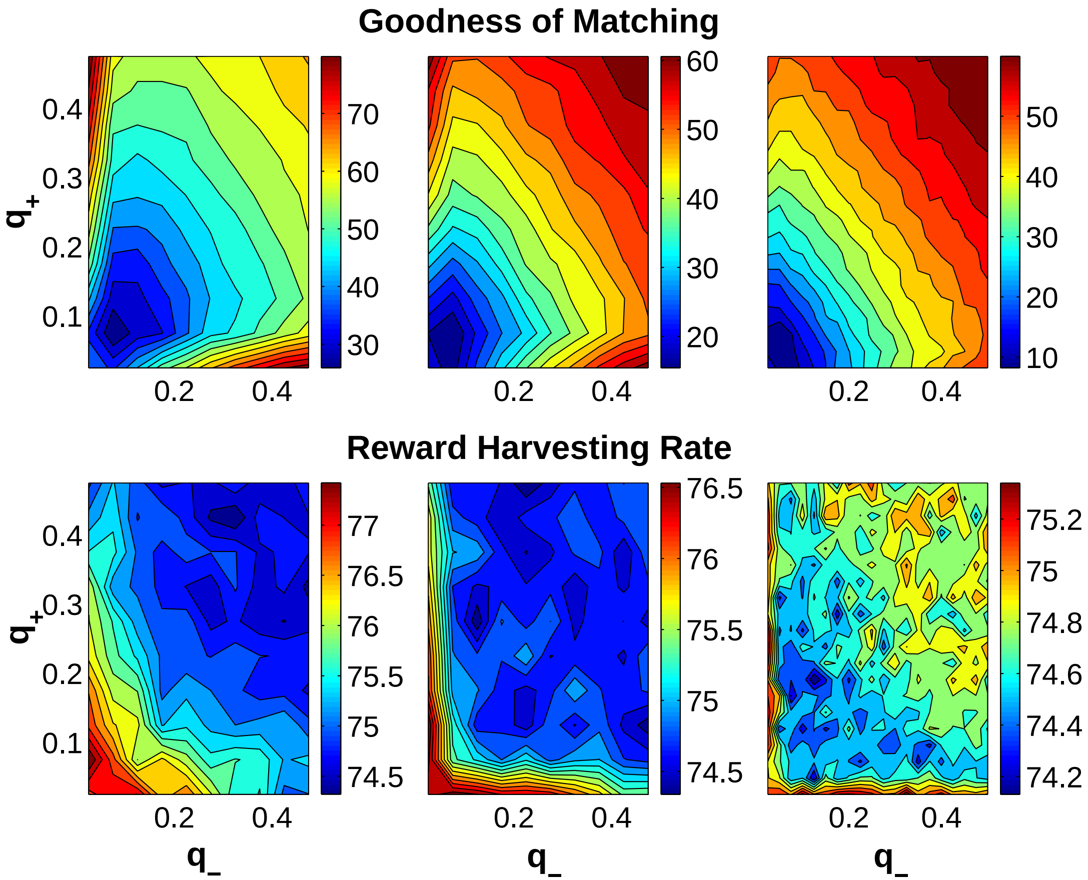


Type 1

Not so robust.

Type 1

Not so robust.



Type 1

Not so robust.