Thinking fast or slow? A reinforcement learning approach

Wouter Kool[a], Samuel J. Gershman[a,b], & Fiery A. Cushman[a]

a) Department of Psychology, Harvard University

b) Center for Brain Science, Harvard University

Dual process theory and reinforcement learning

Theories of judgement and decision making posit existence of *two* systems:
Kahneman (2003)

System 2	vs.	System 1
Habitual		Goal-directed
Automatic		Deliberative
Computationally cheap		Computationally expensive

Recent advances in computer science and reinforcement learning:
Daw et al. (2011)

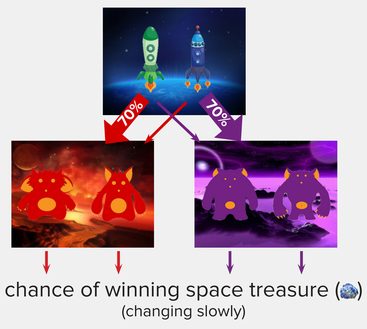
Model-free vs. **Model-based**

Often assumed that systems engage in a cost-benefit trade-off, but direct evidence for this has been sparse

Accuracy-efficiency tradeoff between System 1 and System 2?

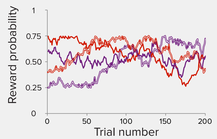
Daw et al. (2011) 2-step task

Task



chance of winning space treasure
(changing slowly)

Reward distribution



Win: Loss:

Bounds = [0.25 0.75]
Drift rate (σ) = 0.025

RL model

Q_{MB}( ) = 0.7 · max ( ) + 0.3 · max ( )

Q_{MF}( ) = Q_{MF}( ) + α·RPE

Q_{net} = $\boldsymbol{w}$ · Q_{MB} + (1 - $\boldsymbol{w}$) · Q_{MF}

w indicates model-basedness (degree of system 2)

Behavioral predictions

Experiment 1. Stakes manipulation $n = 98$

Stakes manipulation

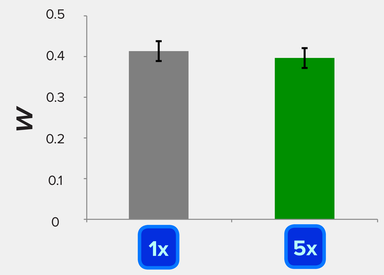
1x 5x

50% 50%

Prediction:
If model-based planning is costly, participants should plan more when stakes are high

$w_{5x} > w_{1x}$

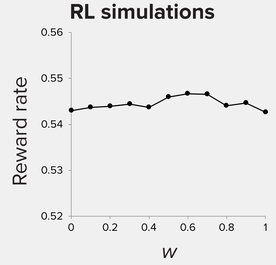
Results



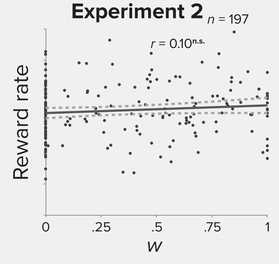
No accuracy-demand tradeoff in Daw 2-step task

Does w predict reward?

RL simulations

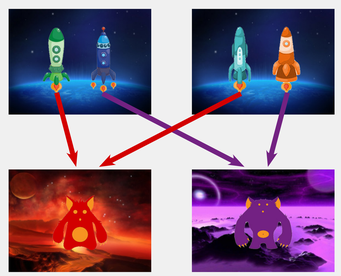


Experiment 2 $n = 197$



Novel paradigm

Task



chance of winning pieces of space treasure:

Range = [,]

Drift rate = 2

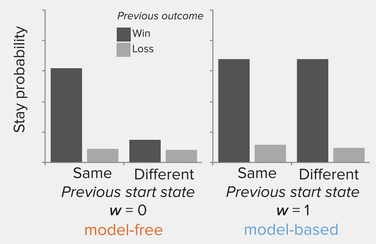
RL model

Q_{MB}( ) = Q( ) = Q_{MB}( )

Q_{MF}( ) = Q_{MF}( ) + α·RPE

Q_{net} = $\boldsymbol{w}$ · Q_{MB} + (1 - $\boldsymbol{w}$) · Q_{MF}

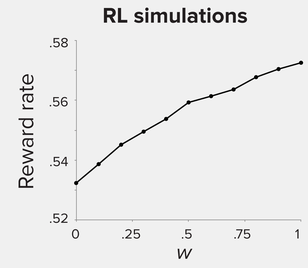
Behavioral predictions



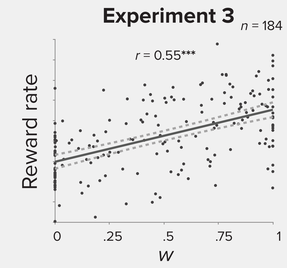
Accuracy-demand tradeoff in novel 2-step task

Does w predict reward?

RL simulations



Experiment 3 $n = 184$



Experiment 4. New paradigm with stakes $n = 94$

Stakes manipulation

1x 5x

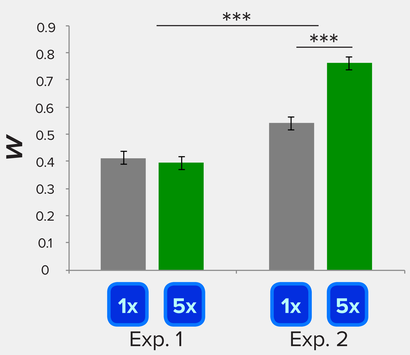
50% 50%

Prediction:
$w_{5x} > w_{1x}$

Correlations

Experiment	Parameter	r	p
Exp 1.	w_{1x}	.54	< .001
	w_{5x}	.32	< .001
Exp. 2	w_{1x}	-.01	.92
	w_{5x}	-.02	.81

Results



Conclusion

Cost-benefit arbitration between multiple RL systems

Flexible adapation based on reward-advantage

In progress: $\boldsymbol{w}$'s relationship with outgroup bias and psychiatric symptoms

Contact and details

wkool@fas.harvard.edu wouterkool.com

*** $p < 0.001$