Peeks and Keeps
Optional Risk Reinforcement Learning

Nathaniel Phillips, Jörg Rieskamp
Economic Psychology
University of Basel

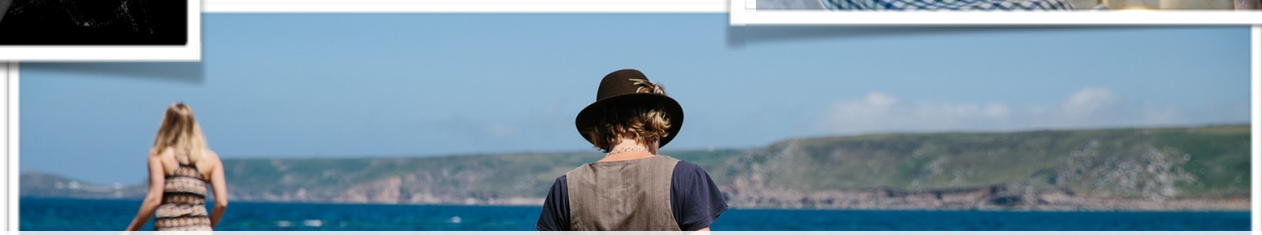
Hansjörg Neth, Wolfgang Gaissmaier
Social Psychology and Decision Sciences
University of Konstanz

Ben Newell, Daniel Navarro
University of New South Wales

DGPS 2016
Leipzig, Germany

REINFORCEMENT LEARNING

...approach to lea
...the total amo
...with a complex,
Sutton & Barto

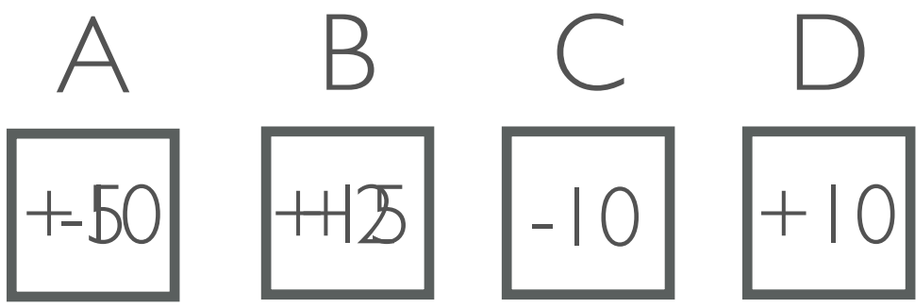


A

B

C

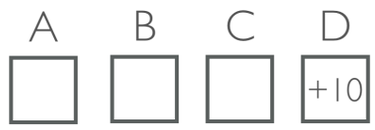
N-ARMED BANDIT



Trial	1	2	3	4	5	…	100
Selection	A	B	C	B	D	…	A
Outcome	**-5**	**+2**	**-10**	**+15**	**+10**	…	**+10**
Rewards	-5	-3	-13	+2	25	…	+200

BANDIT -> CLINICAL POPULATIONS

N-ARMED BANDIT

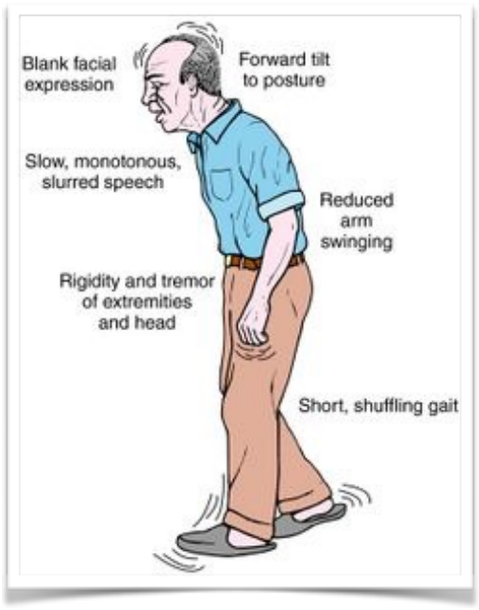


Trial	1	2	3	4	5	…	100
Selection	A	B	C	B	D		
Outcome	-5	+2	-10	+15	+10		
Rewards	-5	-3	-13	+2	25		

Cocaine Users

Cannabis users

Parkinson's



e.g.; Yechiam et al. (2005)

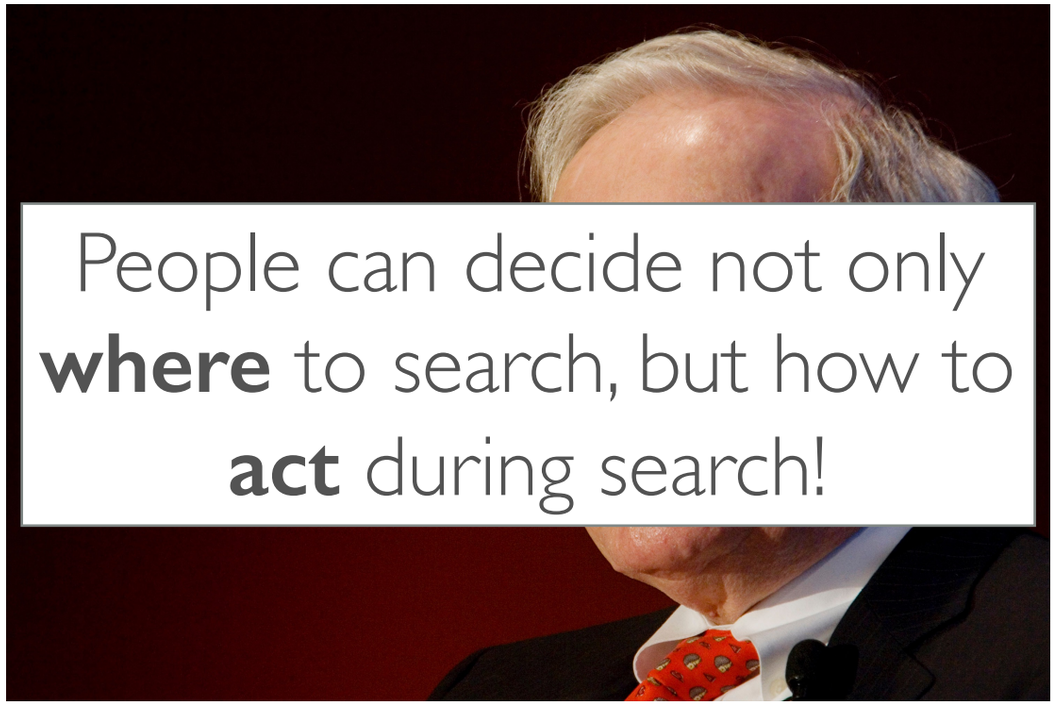
SOMETHING IS MISSING IN THE N-ARMED BANDIT

A good investor is a flexible investor.

Know when to just **watch** (Peek), and know when to **act** (Keep).

Don't **act** (Keep) when you do not have enough information. Don't just **watch** (Peek) when you do have enough information.

Float like a butterfly, then sting like a bee. Then be a butterfly again.



Warren Buffet

OPTIONAL RISK
REINFORCEMENT LEARNING

Optional Risk Decisions from Experience (ORDFE)			
Action		**Information**	**Risk Punishments and Rewards**
Peek		Yes	**No**
Keep		Yes	**Yes**

OPTIONAL RISK RL

Options

Actions	Info?	Risk
Peek 	Yes	No
Keep 	Yes	Yes

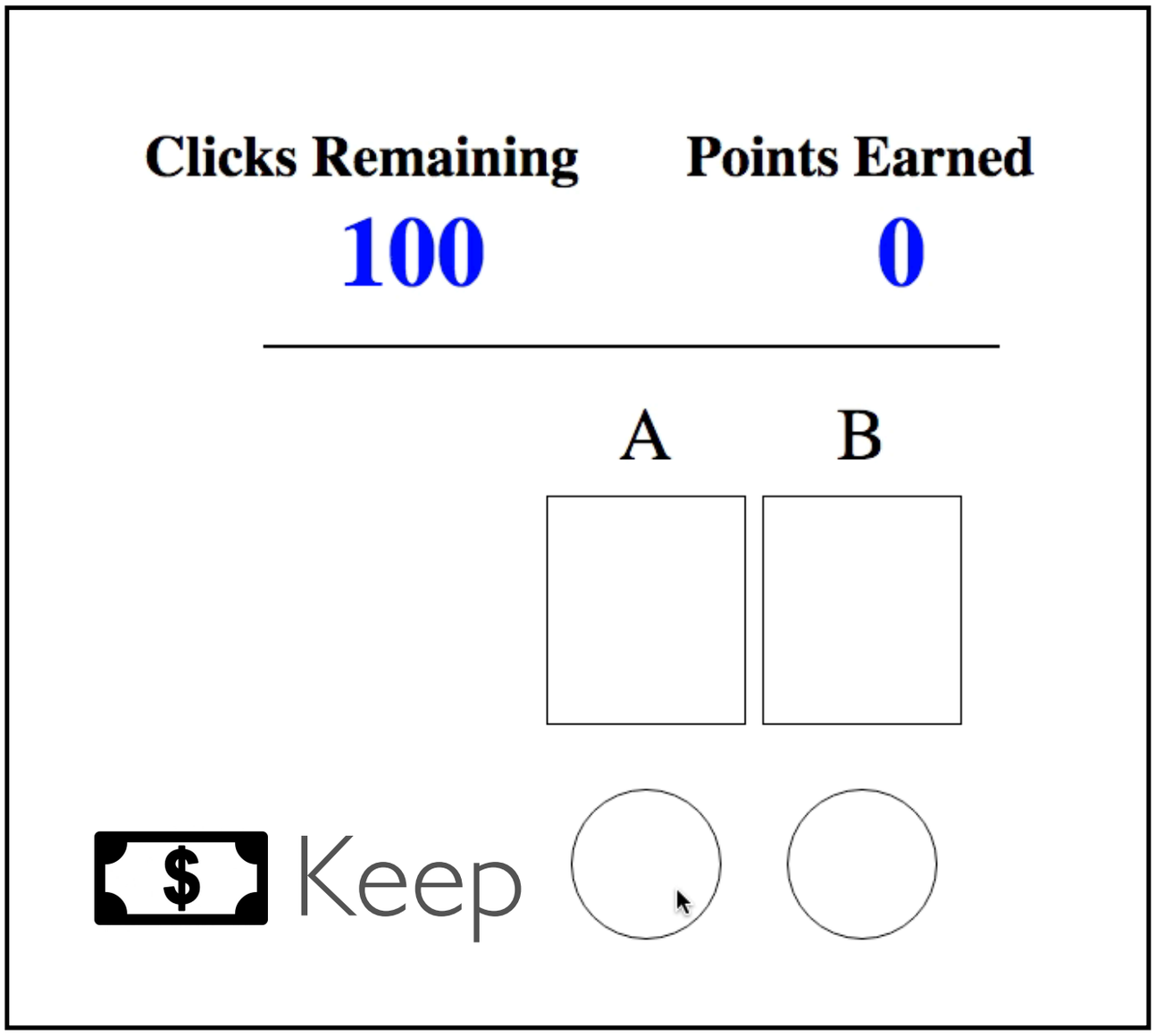
Trial	**1**	**2**	**3**	**4**	**5**	**6**	**…**	**100**
Selection							…	
Action								
Outcome	+10	-5	+8	**+15**	-10	**-5**	…	**+10**
Rewards	0	0	0	+15	+15	+10	…	+200

How, and how well, do people use no-risk 🔍 ***Peeking*** *and risky* 💵 ***Keeping*** *in Optional Risk Reinforcement Learning?*

3 RISK TYPES

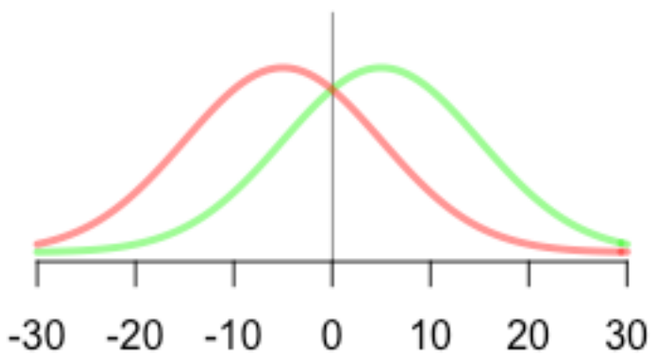
Risk	Domain
 Non-optional Risk	

NON-OPTIONAL RISK

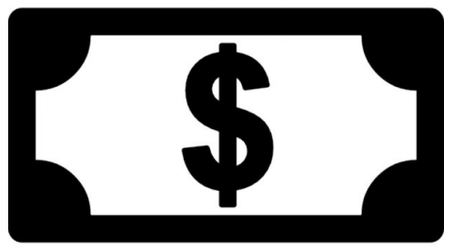
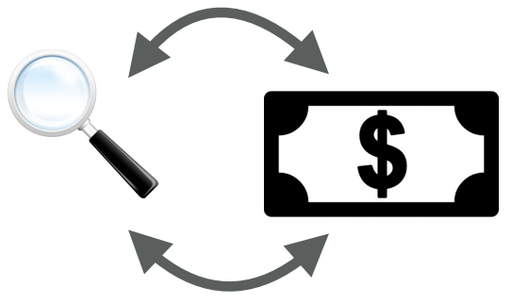


$A \sim N(\mu = +5, \sigma = 10)$

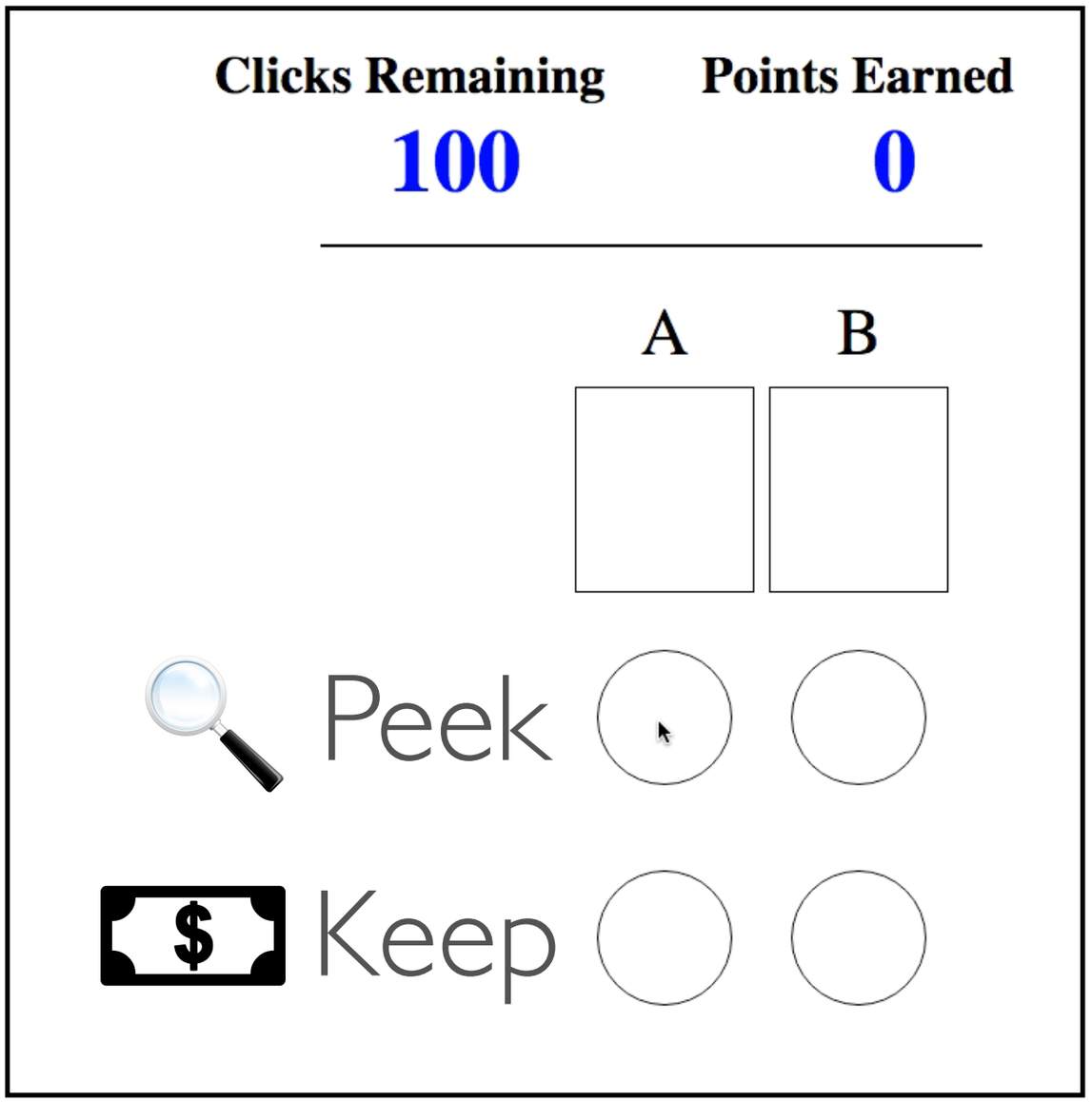
$B \sim N(\mu = -5, \sigma = 10)$



3 RISK TYPES

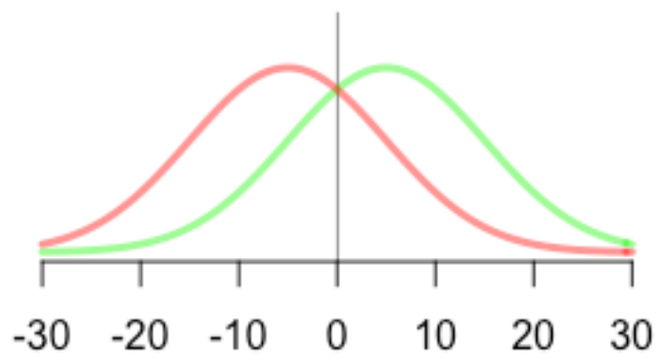
Risk	Domain
 Non-optional Risk	
 Alternating Risk	

ALTERNATING RISK

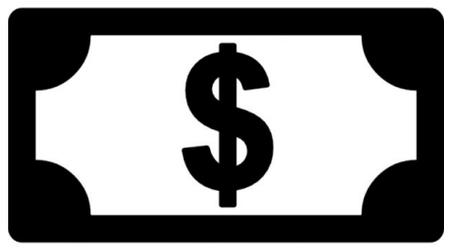
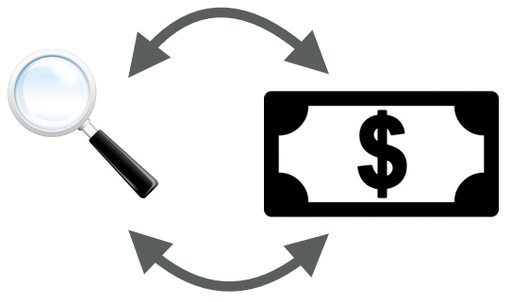
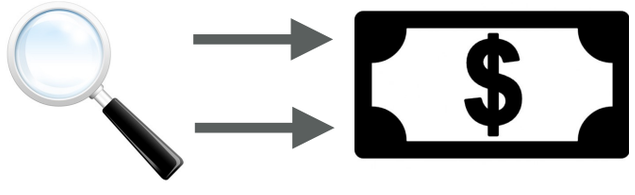


$A \sim N(\mu = +5, \sigma = 10)$

$B \sim N(\mu = -5, \sigma = 10)$



3 RISK TYPES

Task	Domain
 Non-optional Risk	
 Alternating Risk	
 Sequential Risk	

SEQUENTIAL RISK 🔍⇉💵

Clicks Remaining	Points Earned
100	0

A	B

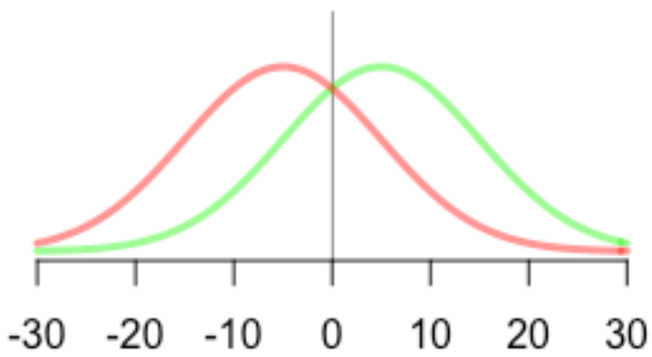
🔍 Peek ○ ○

💵 Keep ○ ○

Note: Once you Keep, you can no longer Peek!

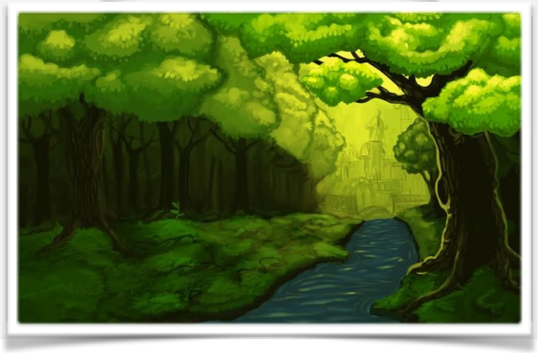
A ~ N(**μ** = **+5**, **σ** = 10)

B ~ N(**μ** = **-5**, **σ** = 10)

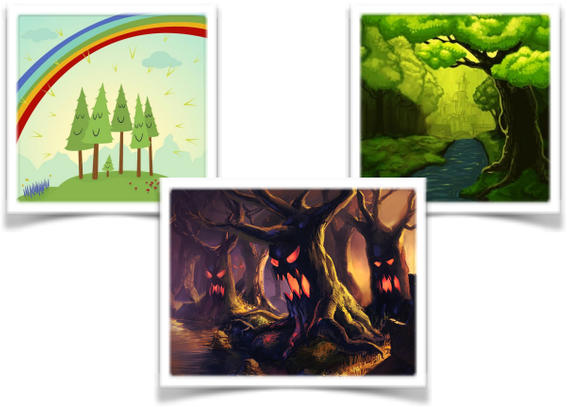


3 ENVIRONMENTS

+ ~ N(**μ** = +5) - ~ N(**μ** = -5)

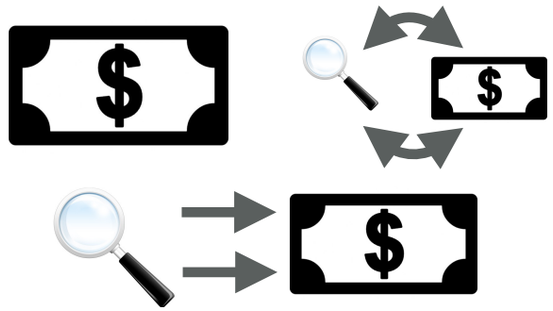
Environment		Stimuli
Safe		— +
Medium		— + — —
Dangerous		— + — — — —

Environments



2 Questions

Risk Types



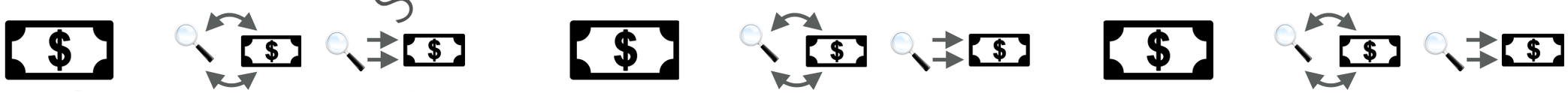
	Question	Answer
1	In which environments do people use optional risk to their benefit and when is it to their detriment?	

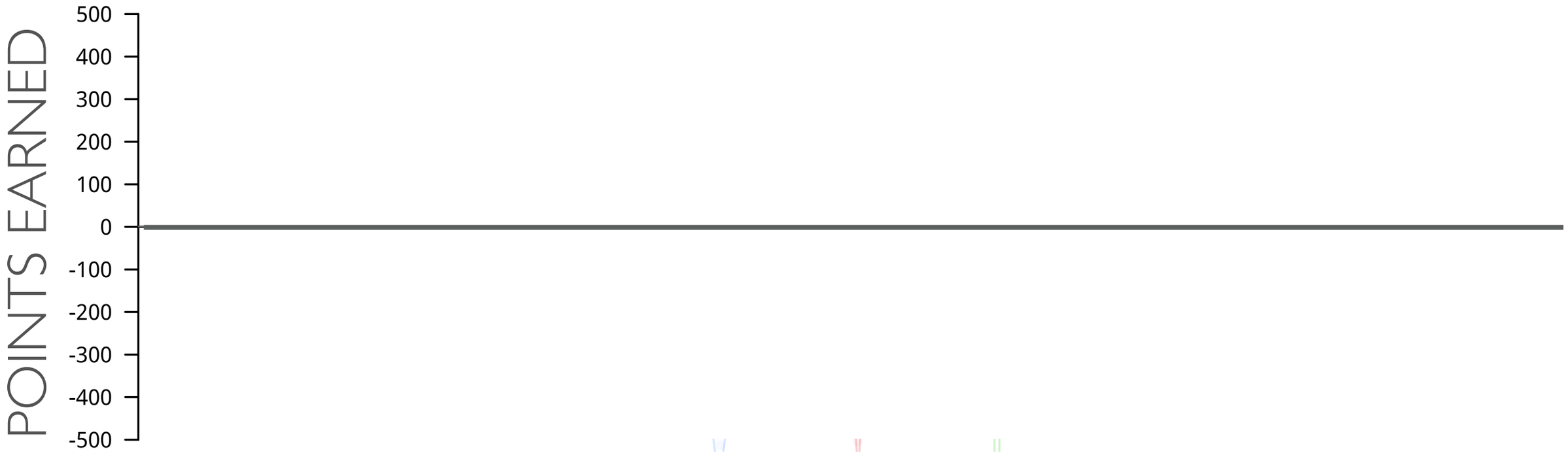
BF (Interaction) = 41.02

Non-Optional

Alternating

Sequential





Safe

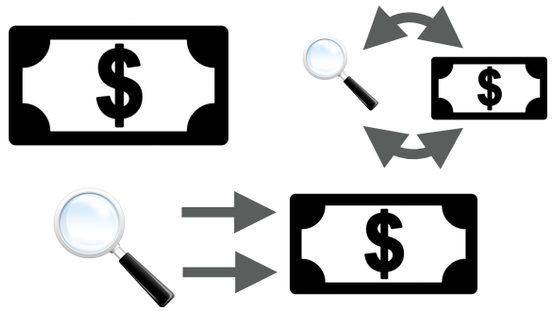
Medium

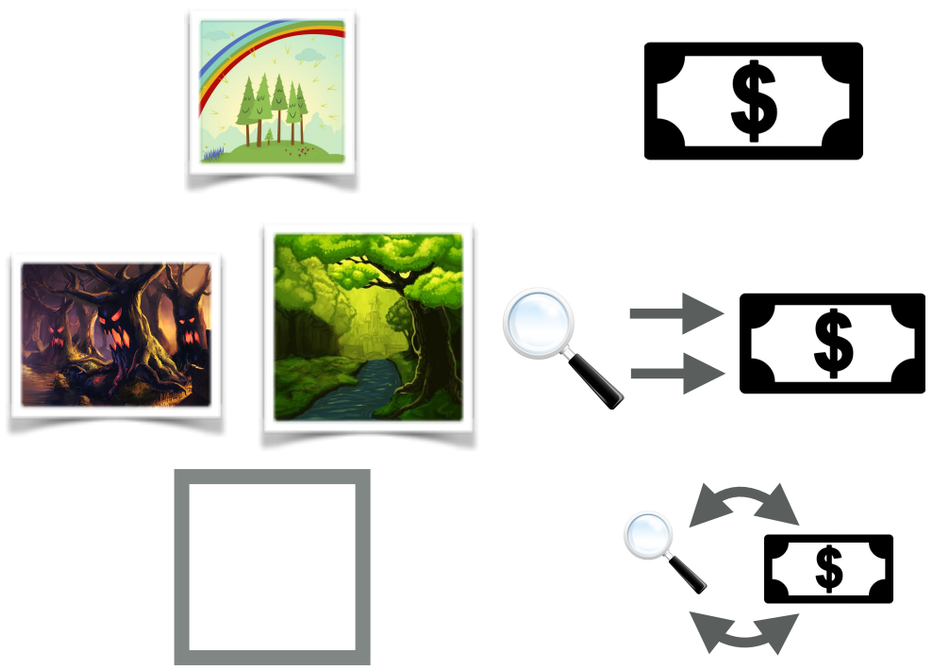
Dangerous

Environments

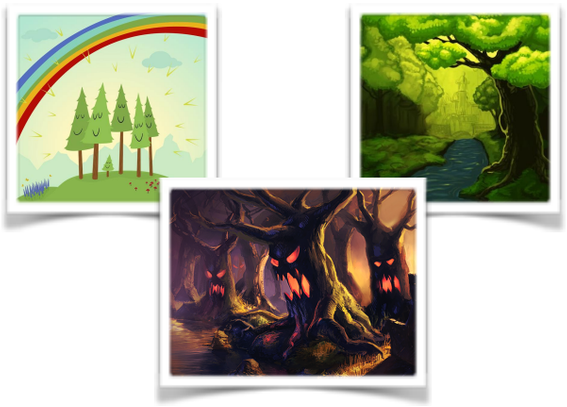
2 Questions

Risk Types



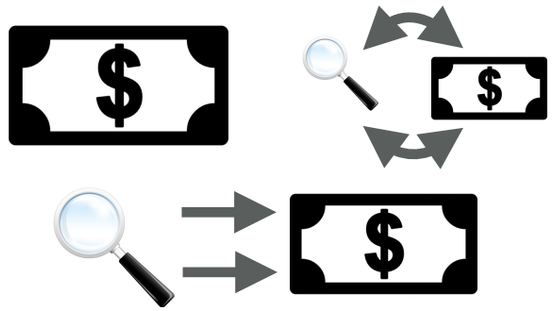
	Question	Answer
1	In which environments do people use optional risk to their benefit and when is it to their detriment?	

Environments



2 Questions

Risk Types



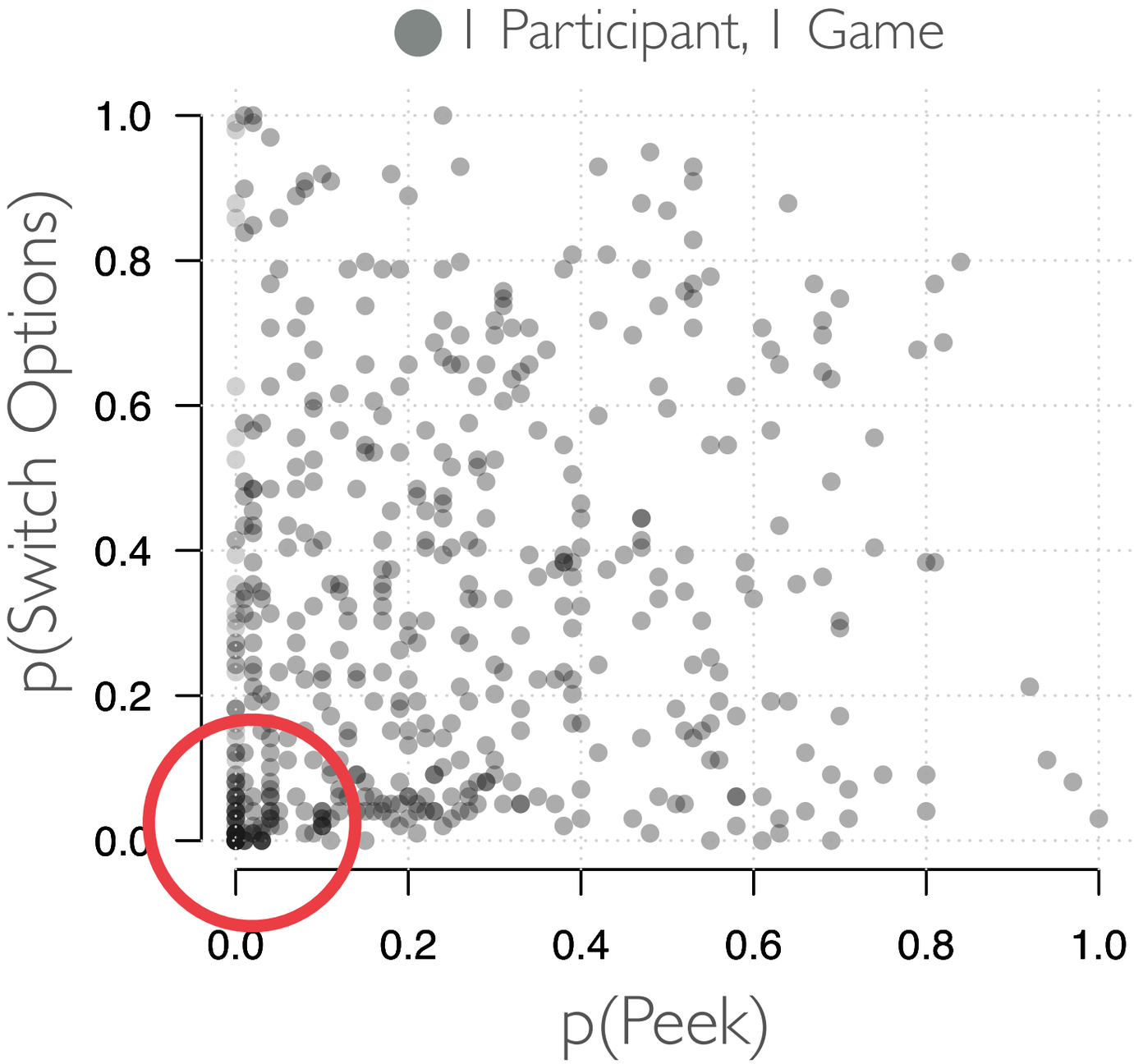
	Question	Answer
2	Could Peeking a new measure of exploration?	

Q2: Could Peeking be a new measure of exploration?

A B C

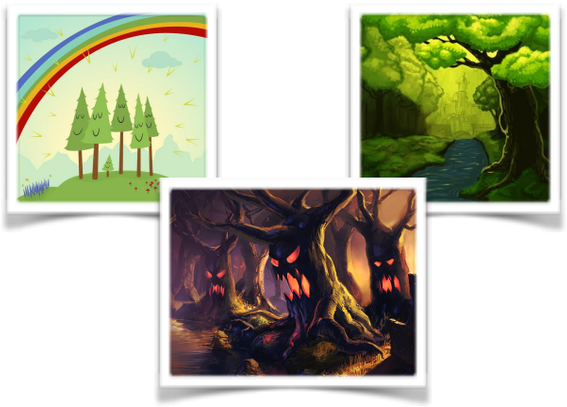
	Exploration	Exploitation
Goal	Select relatively unknown options to obtain information to improve long-term decisions	Select known options to obtain short-term rewards
Behaviour	**High-Switching** A, B, C, A, B, C, A, …	**Low Switching** A, A, A, B, A, A, A, A

Q2: Could Peeking be a new measure of exploration?



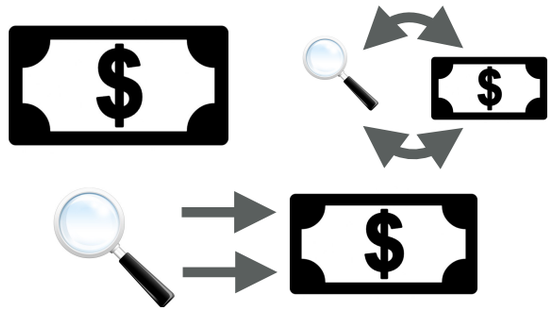
Data	Cor	BF
All	0.32	5328
p(Peek) > 0	0.12	0.53

Environments



2 Questions

Risk Types

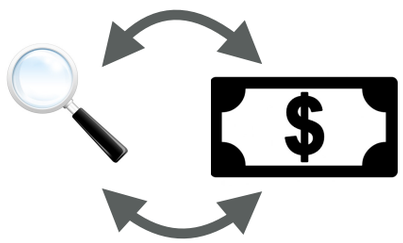


	Question	Answer
2	Could Peeking be a new measure of exploration?	Ja!  ≠ A, B, C, A, B

CONCLUSION

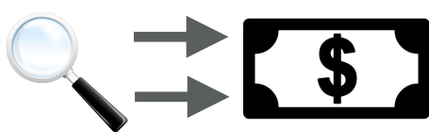
Non-optional Risk

≠



Alternating Risk

≠



Sequential Risk

Peeks and Keeps
Optional Risk Reinforcement Learning

Nathaniel Phillips, Jörg Rieskamp
University of Basel

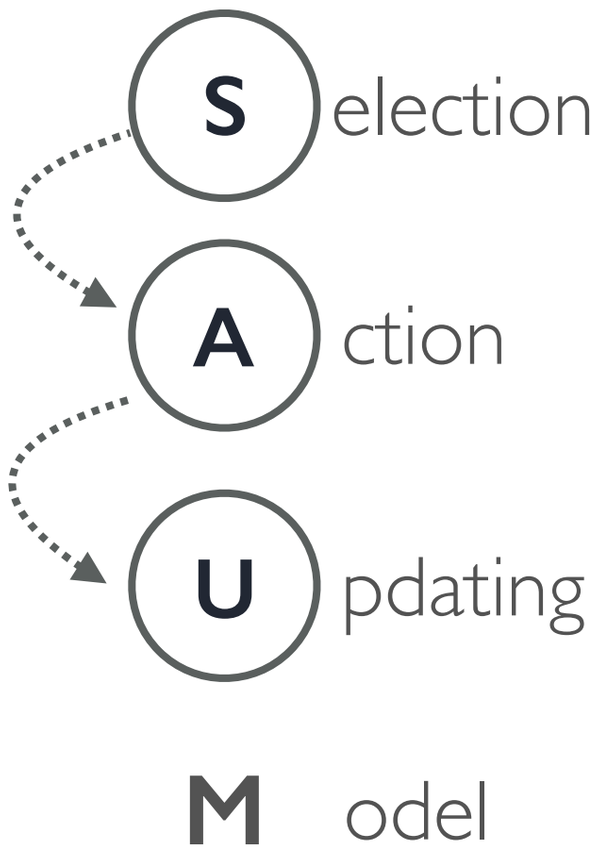
Hansjörg Neth, Wolfgang Gaissmaier
Social Psychology and Decision Sciences
University of Konstanz

Ben Newell, Daniel Navarro
University of New South Wales

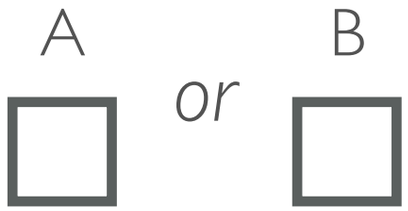
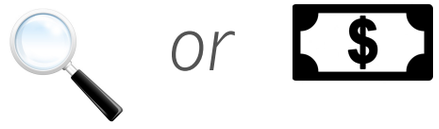
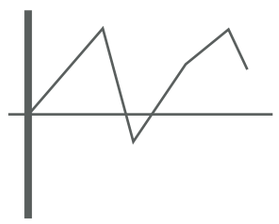
DGPS 2016
Leipzig, Germany

MODELING

Do the cognitive processes underlying search and learning differ between forced risk, alternating risk, and sequential risk?



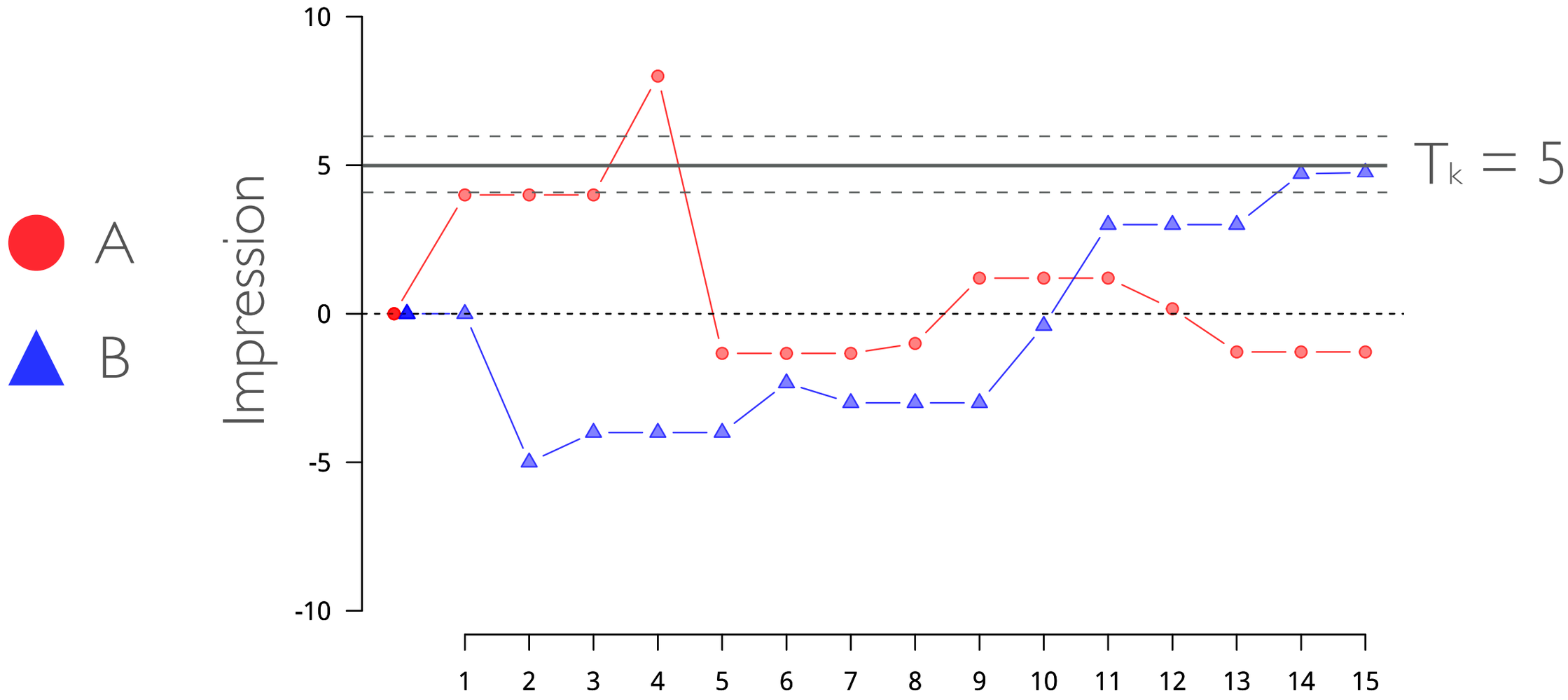
SELECTION, ACTION, UPDATING
SAUM

	Stage	Formulation	Parameters
S	**Selection** *A or B?* 	Softmax $$p(S=i\|T=t)=\frac{e^{I_t(i)\times\theta_t}}{\sum_{j=1}^{n}e^{I_t(i)\times\theta_t}}$$ $$\theta_t = log(t) * \gamma$$	γ 0 ⟼ Inf Random — Greedy
A	**Action** *Peek or Keep?* 	Threshold *Next slide*	Tκ -Inf ⟷ Inf Never Keep — Always Keep
U	**Updating** *New impression?* 	Prediction - Error $$I_t(i) = (1-\frac{1}{s_t(i)}^{\alpha})\times I_t(i) + \frac{1}{s_t(i)}^{\alpha}\times f_t(i)$$	α 0 ⟼ Inf Recency — Primacy

ACTION

Threshold

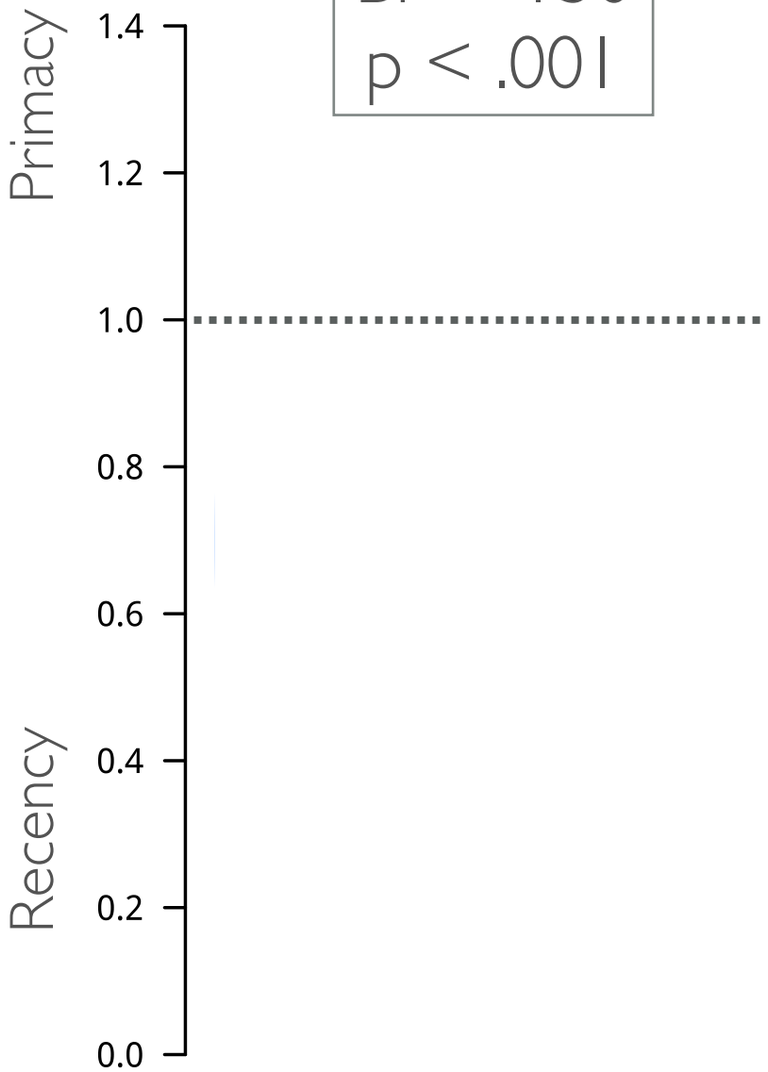
$$p(A = Keep|I_t(i)) = \int_{-\infty}^{I_t(i)} N(\mu = T_\kappa, \sigma = T_\sigma)$$

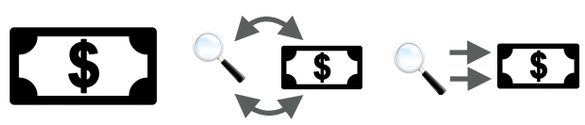


MLE PARAMETER ESTIMATES

α (updating)

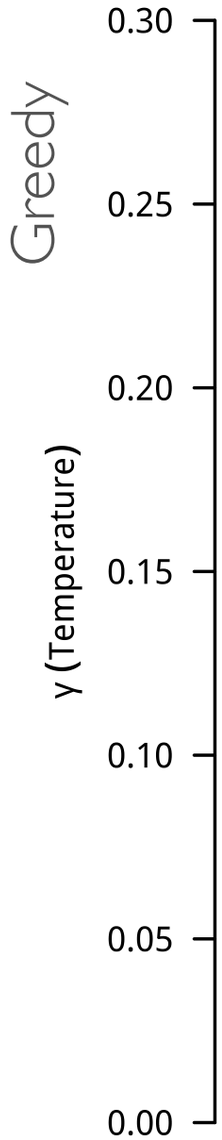
BF > 150
p < .001

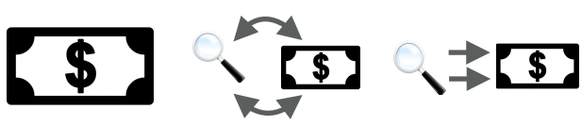




γ (Selection Temperature)

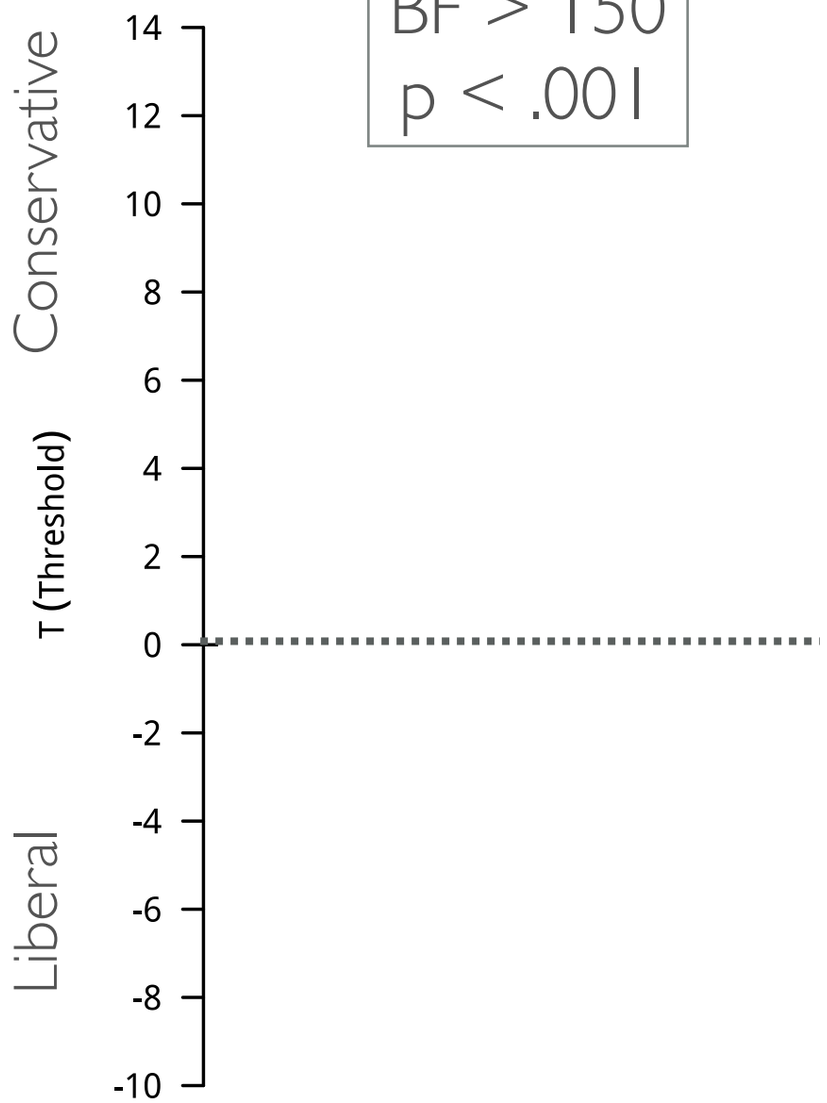
BF = 6.7
p = .002



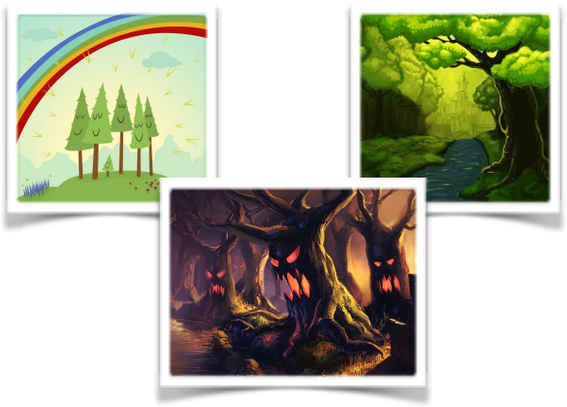


T_τ (Threshold Int)

BF > 150
p < .001

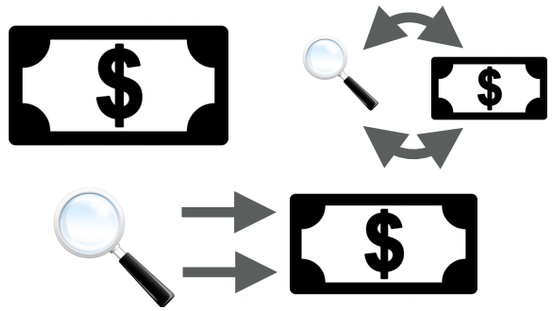


Environments



3 Questions

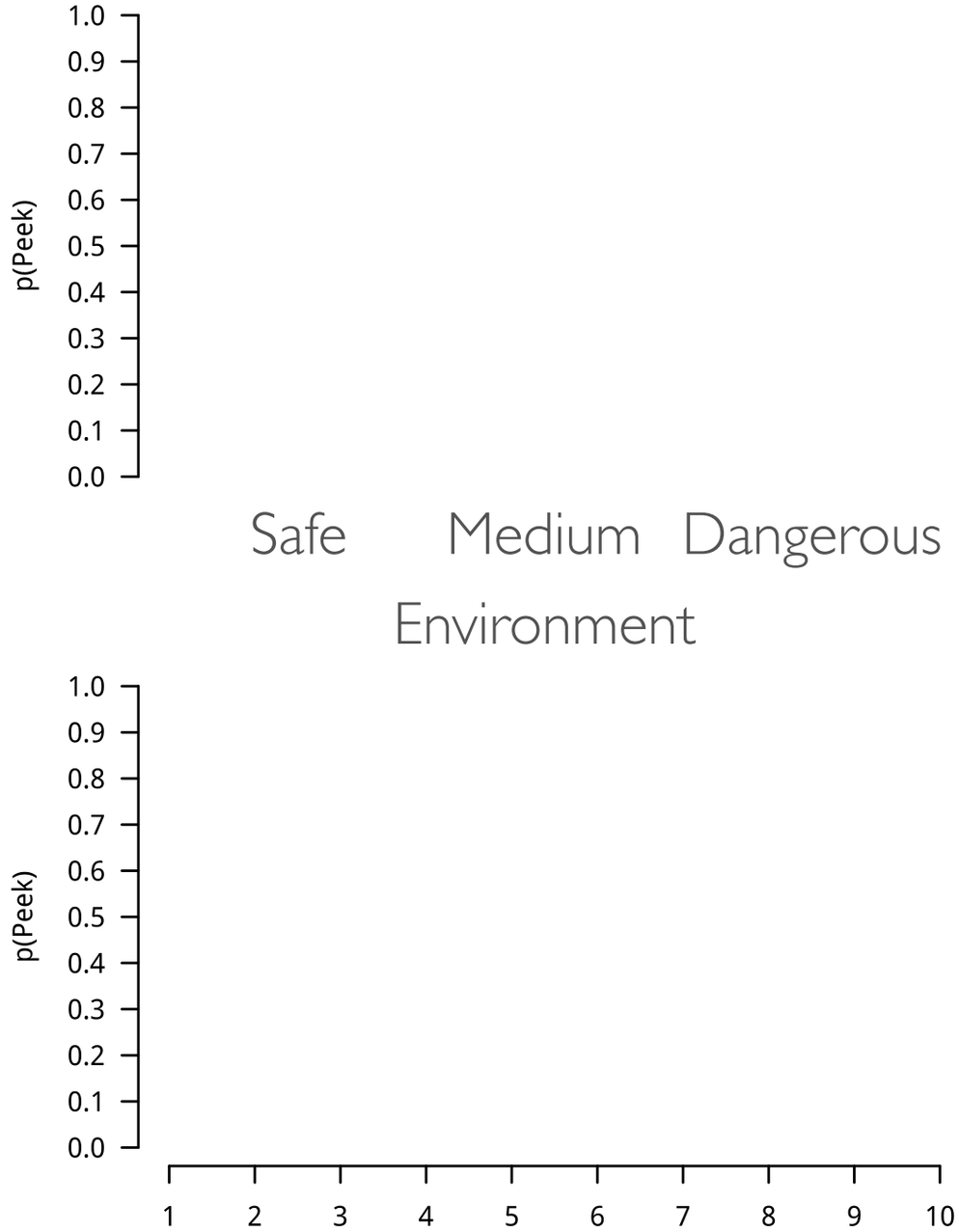
Risk Types



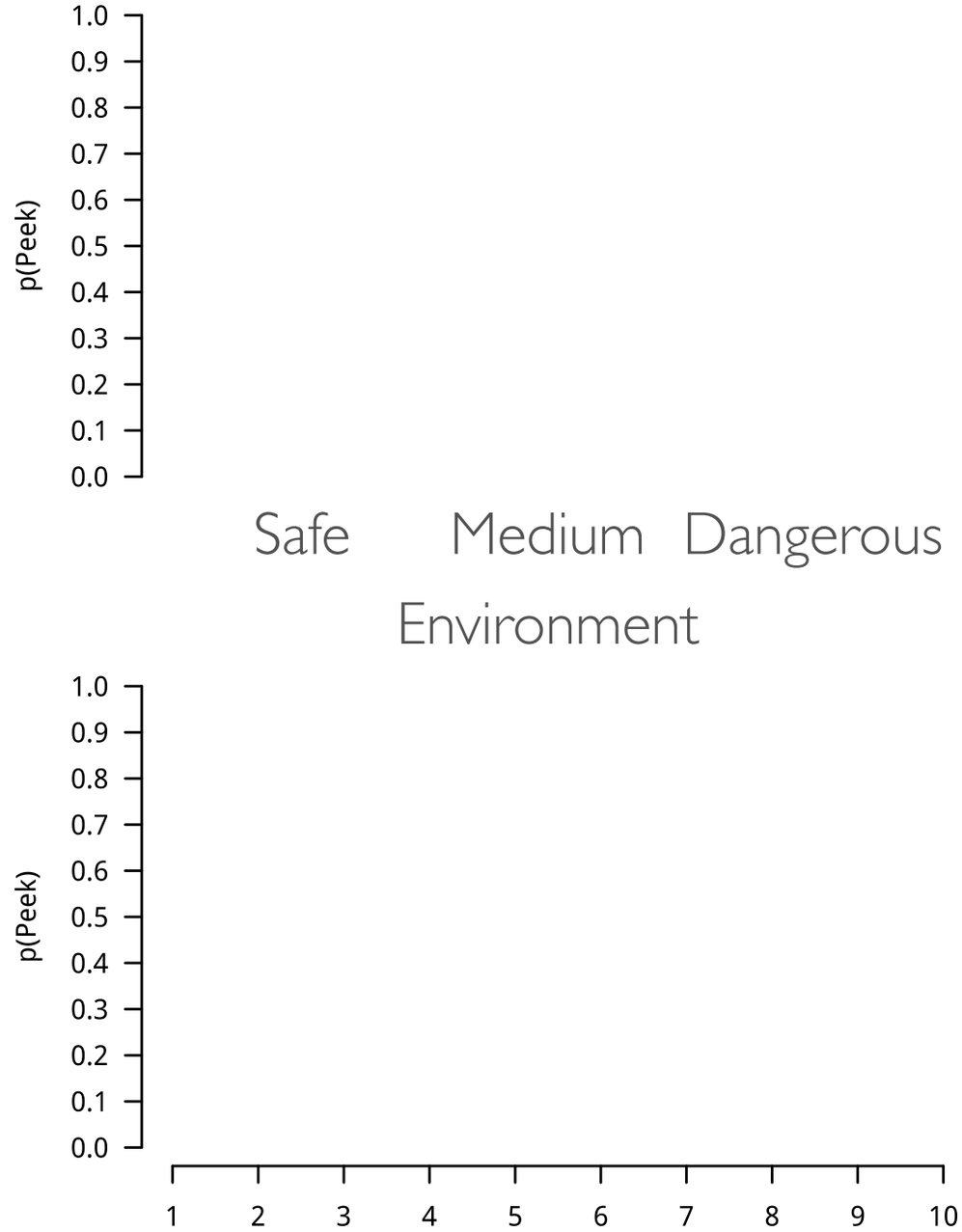
	Question	Answer
2	What affects peeking rates?	

No-Risk Search Rates

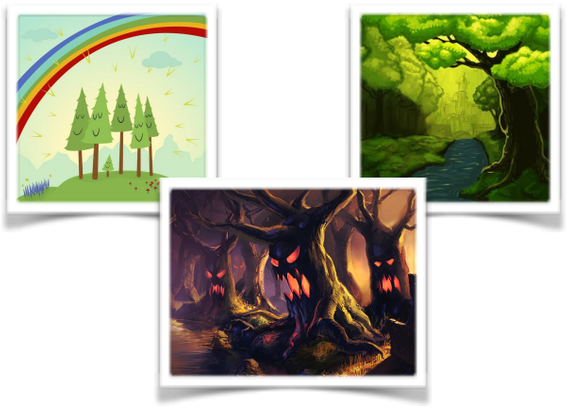
Alternating Risk



Sequential Risk

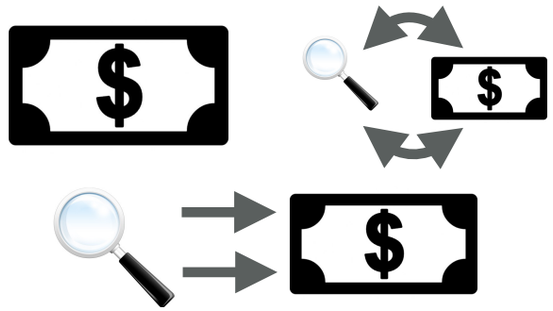


Environments



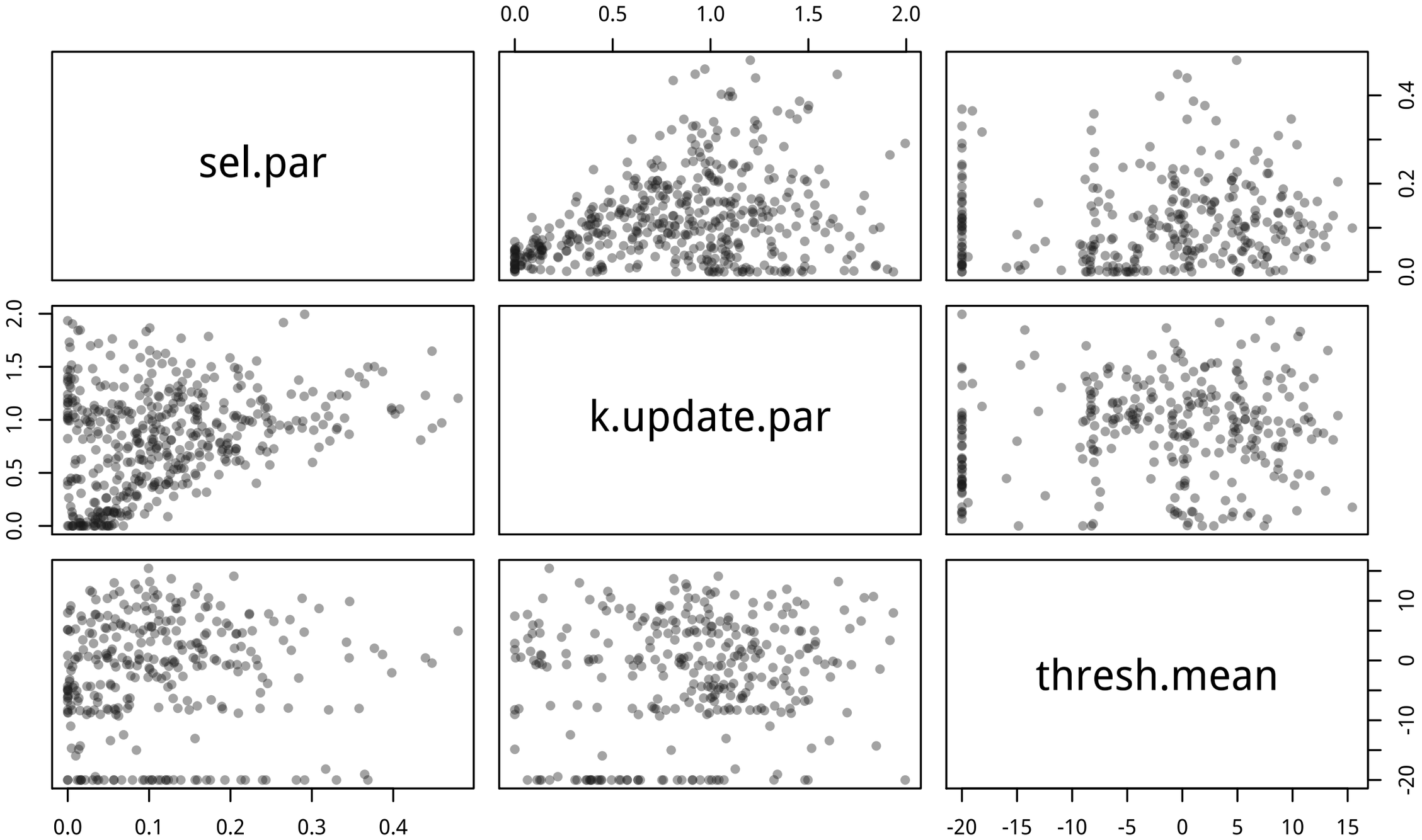
3 Questions

Risk Types

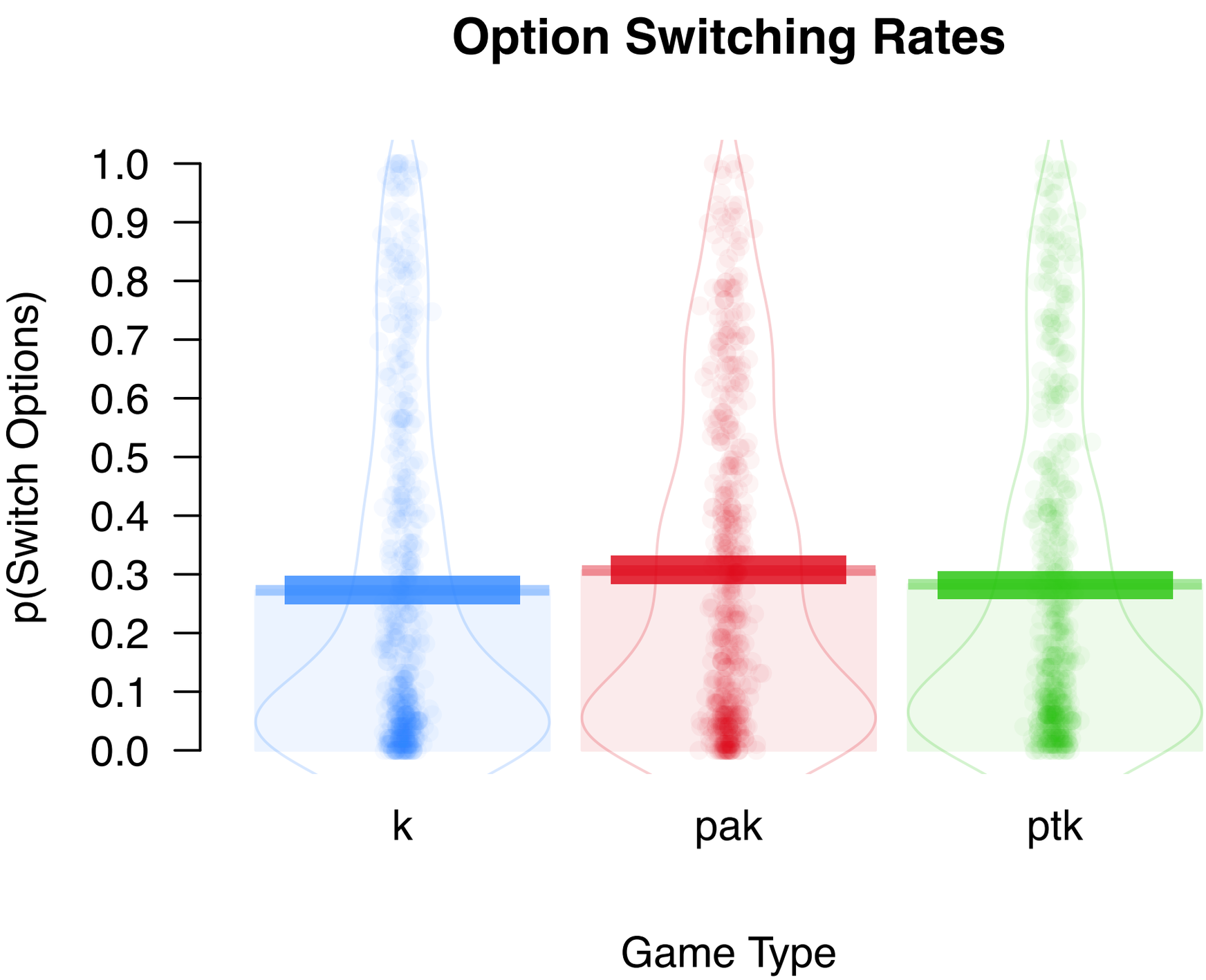


	Question	Answer
2	What affects peeking rates?	

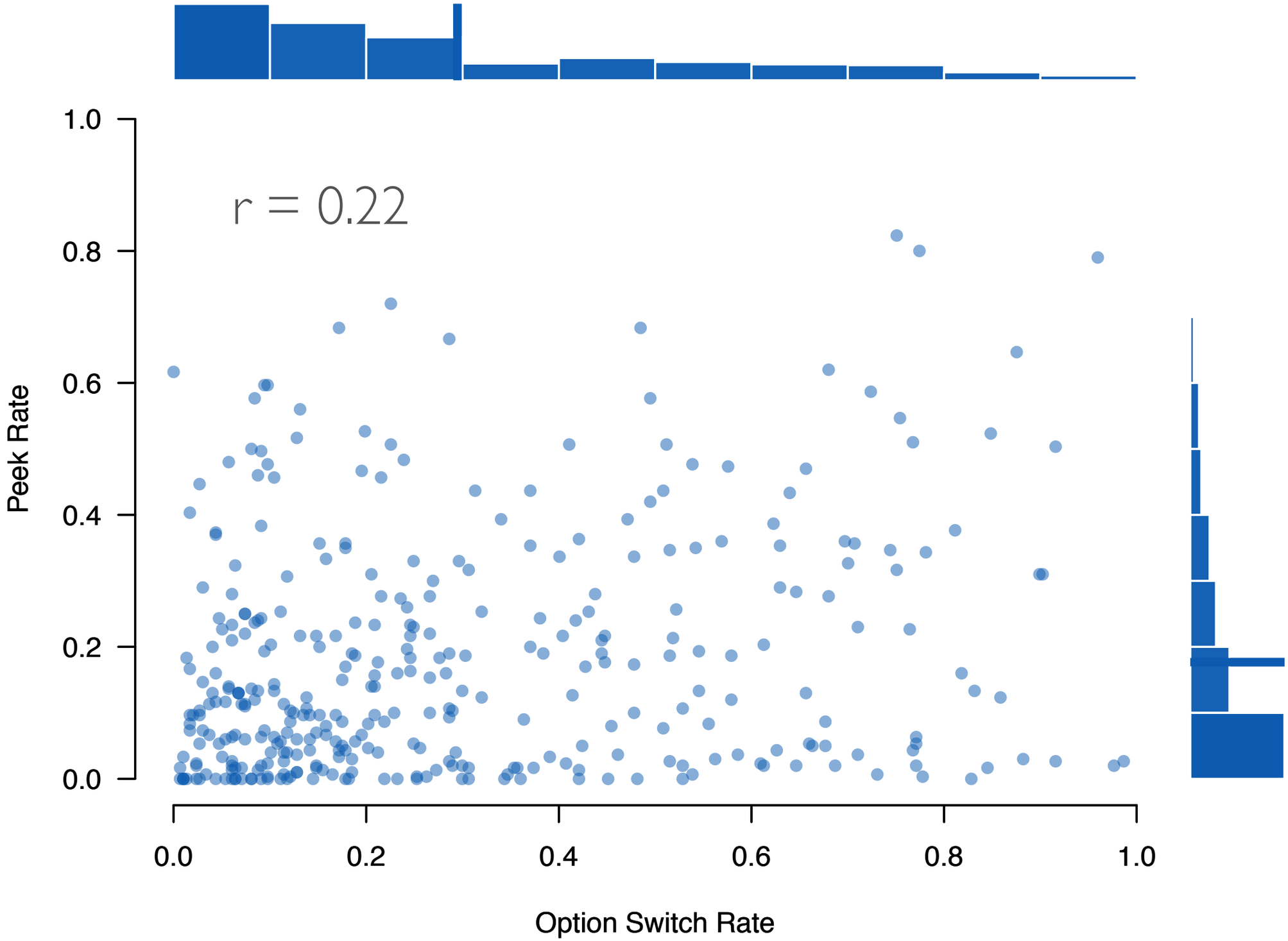
PARAMETER CORRELATIONS



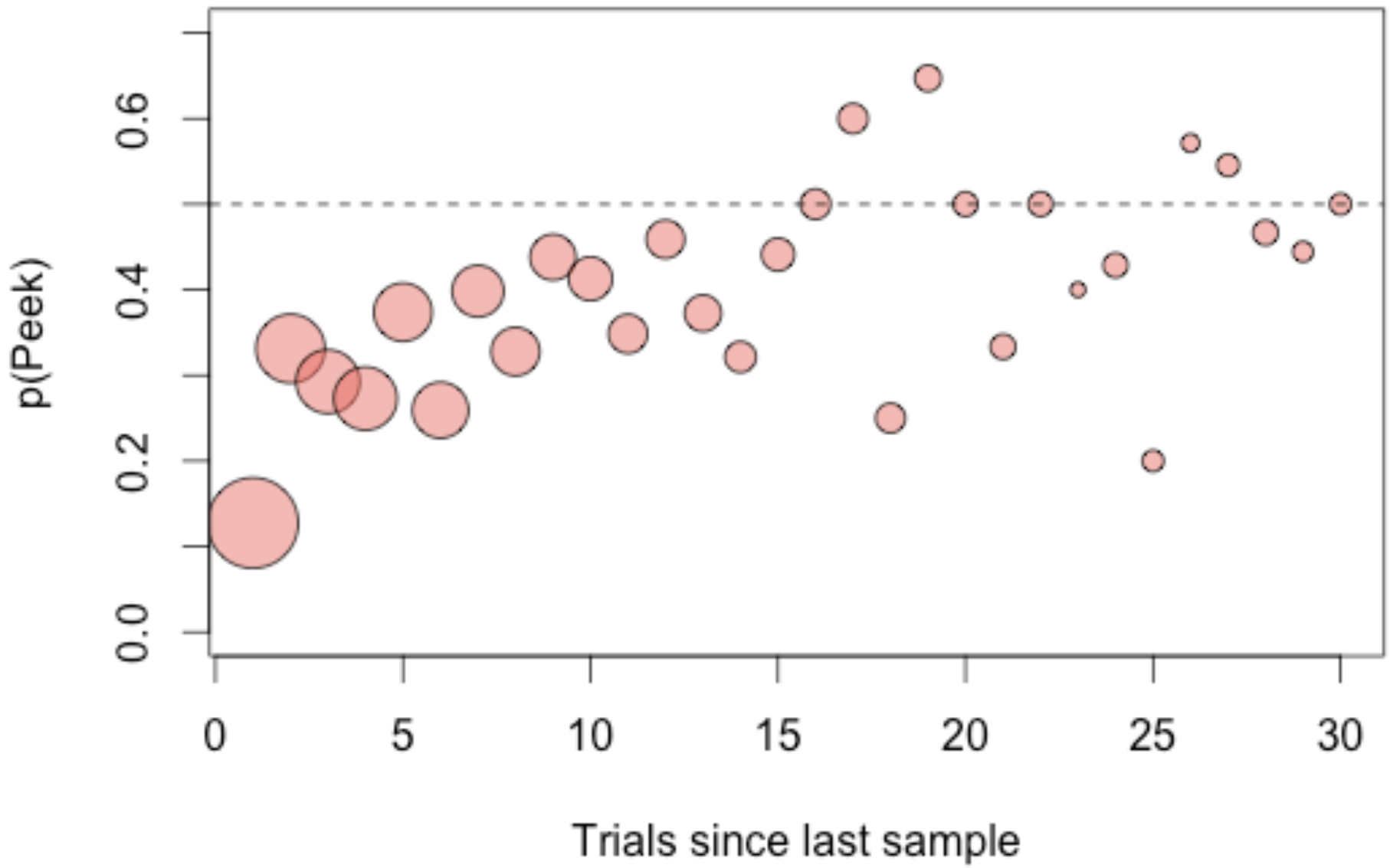
OPTION SWITCHING



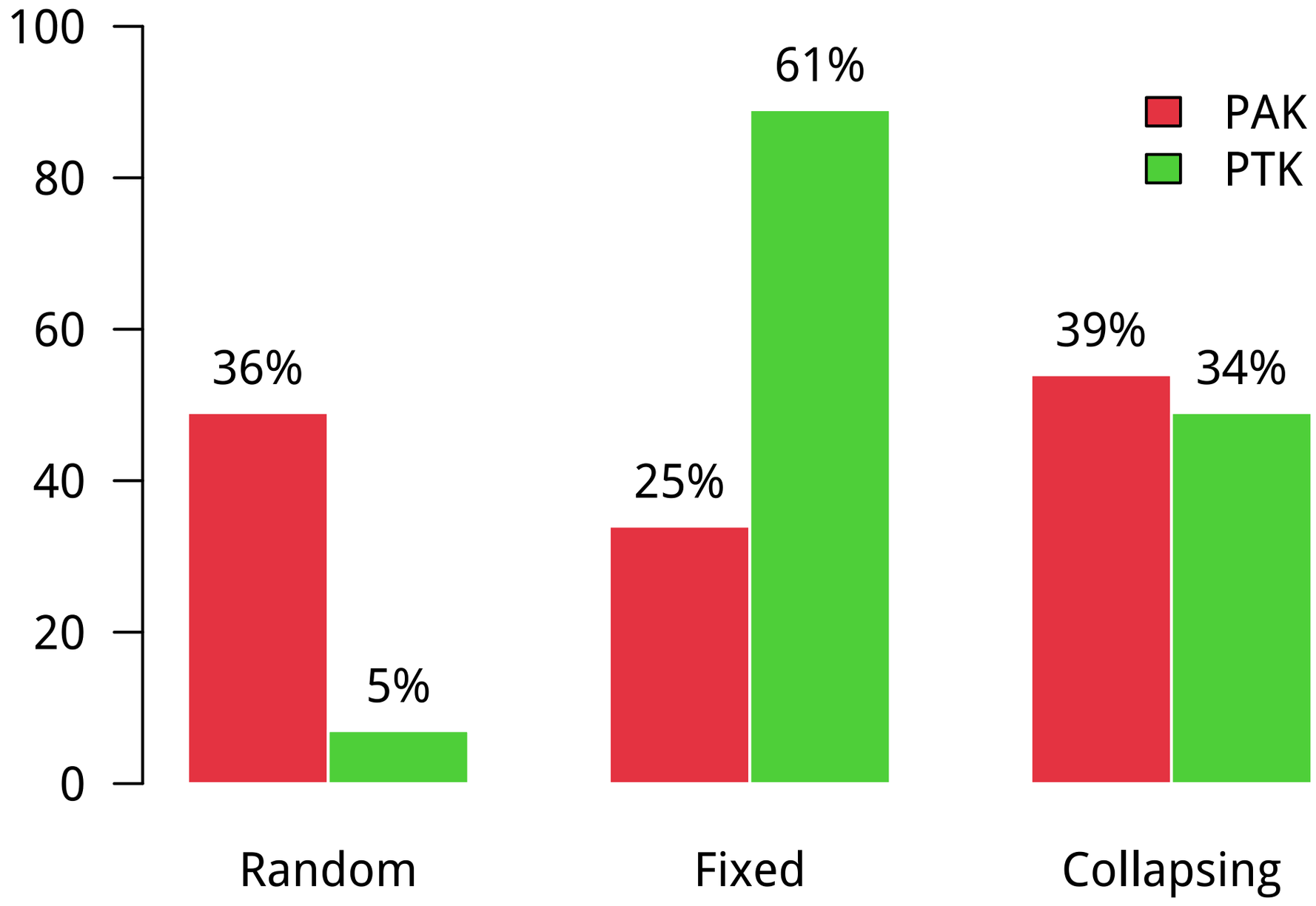
OPTION SWITCHING AND PEEKING



Q3: Is Peeking a novel measure of exploration?



MODEL CLASSIFICATION RESULTS



Calculate ML estimates of each parameter for each participant.
Classify participants based on model with lowest BIC

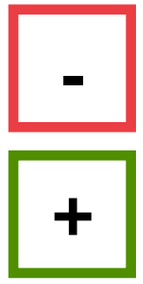
STUDY

- 526 Mturk Participants
- Incentivised performance
- 9 between participant conditions
- 3 identical games per subject

3 Environments

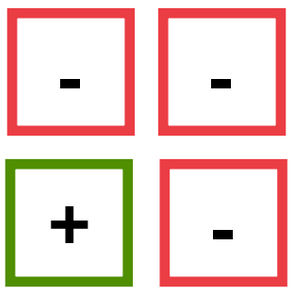
Safe

N = 2



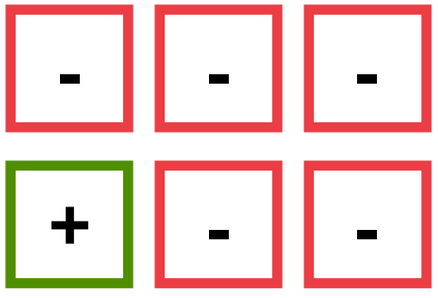
Medium

N = 4



Dangerous

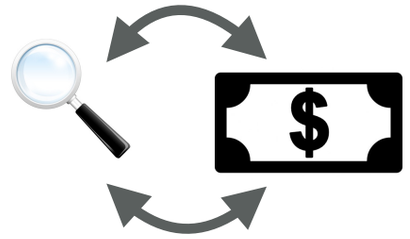
N = 6



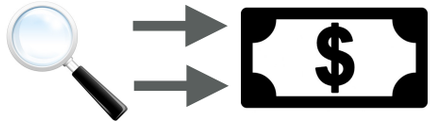
3 Risk Types

Non-optional

Alternating



Sequential

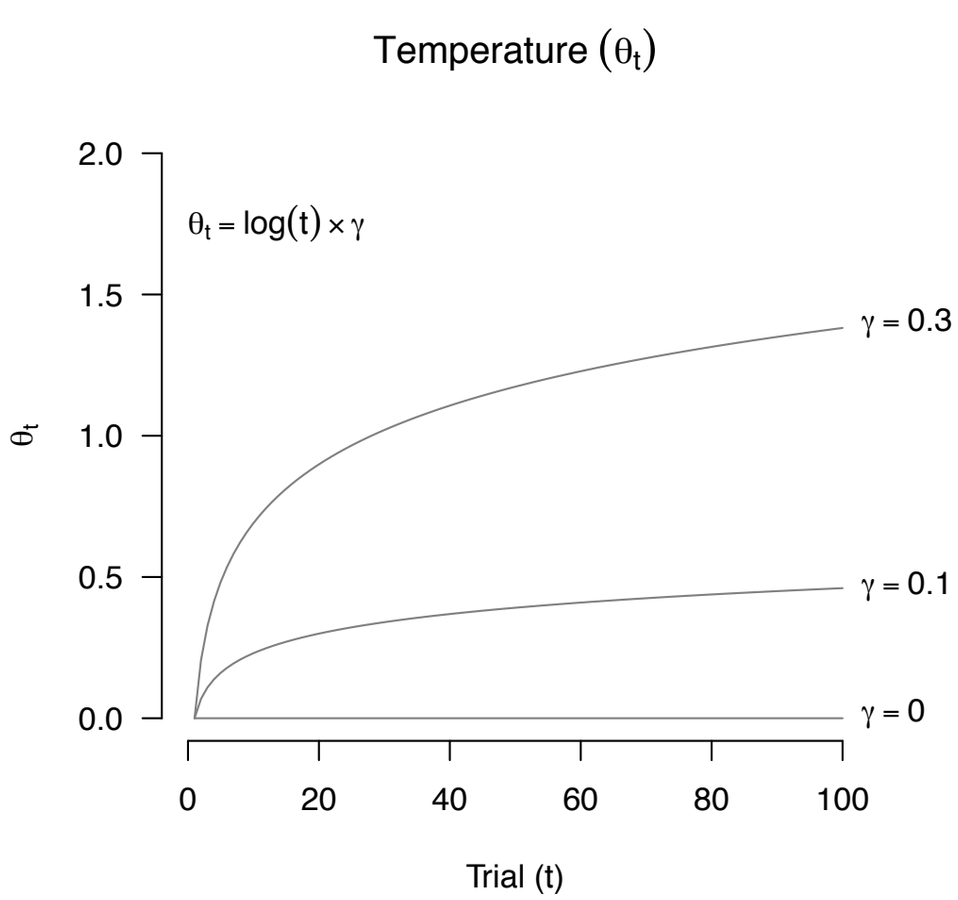


SELECTION

Softmax $$p(S = i|T = t) = \frac{e^{I_t(i) \times \theta_t}}{\sum_{j=1}^{n} e^{I_t(i) \times \theta_t}}$$

Temperature

$$\theta_t = log(t) * \gamma$$



UPDATING

Prediction-error
Learning

$$I_t(i) = (1 - \frac{1}{s_t(i)}^{\alpha}) \times I_t(i) + \frac{1}{s_t(i)}^{\alpha} \times f_t(i)$$

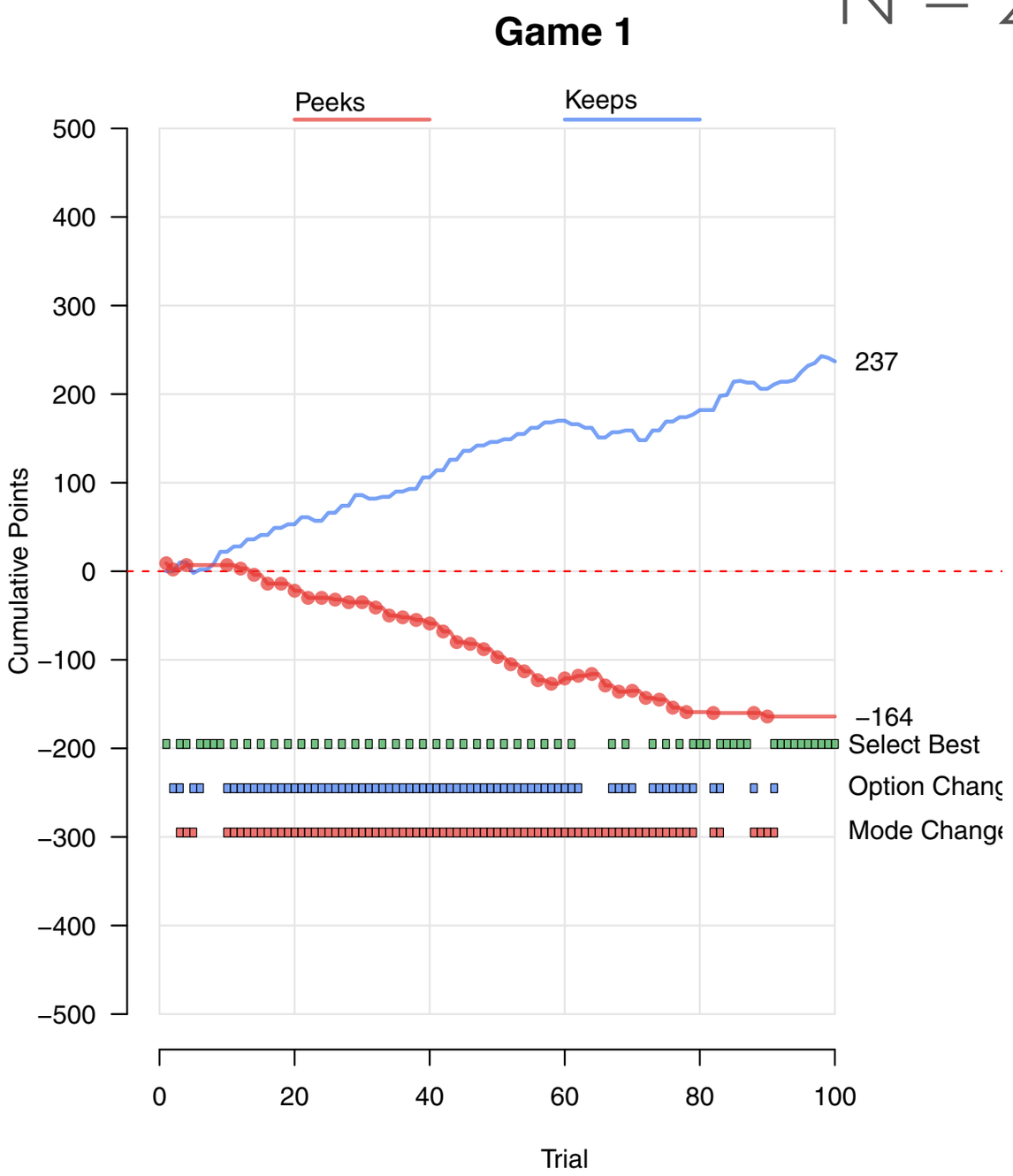
Recency Primacy

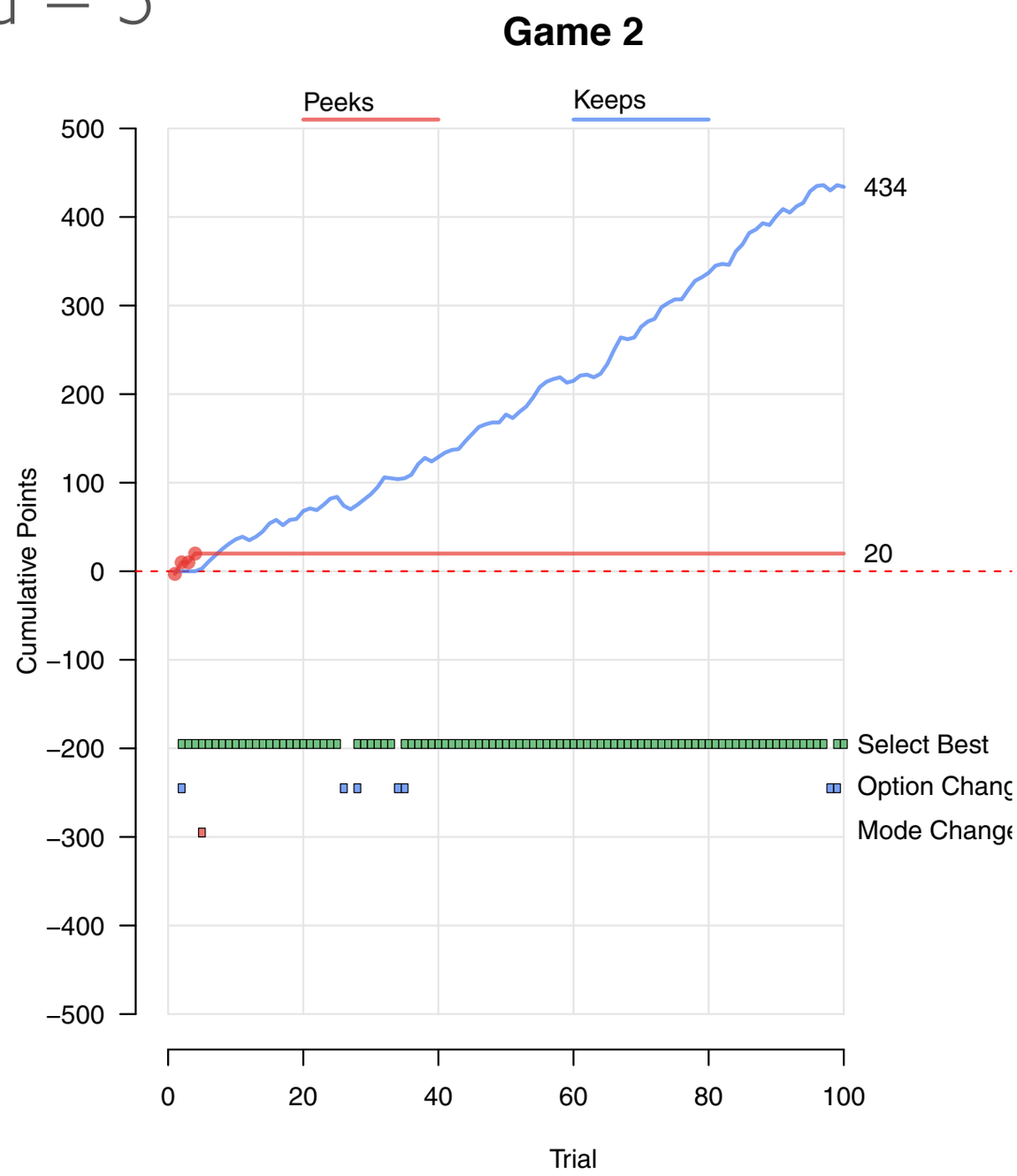
0 1 Inf

α

REPEATED GAMES

N = 2, sd = 5



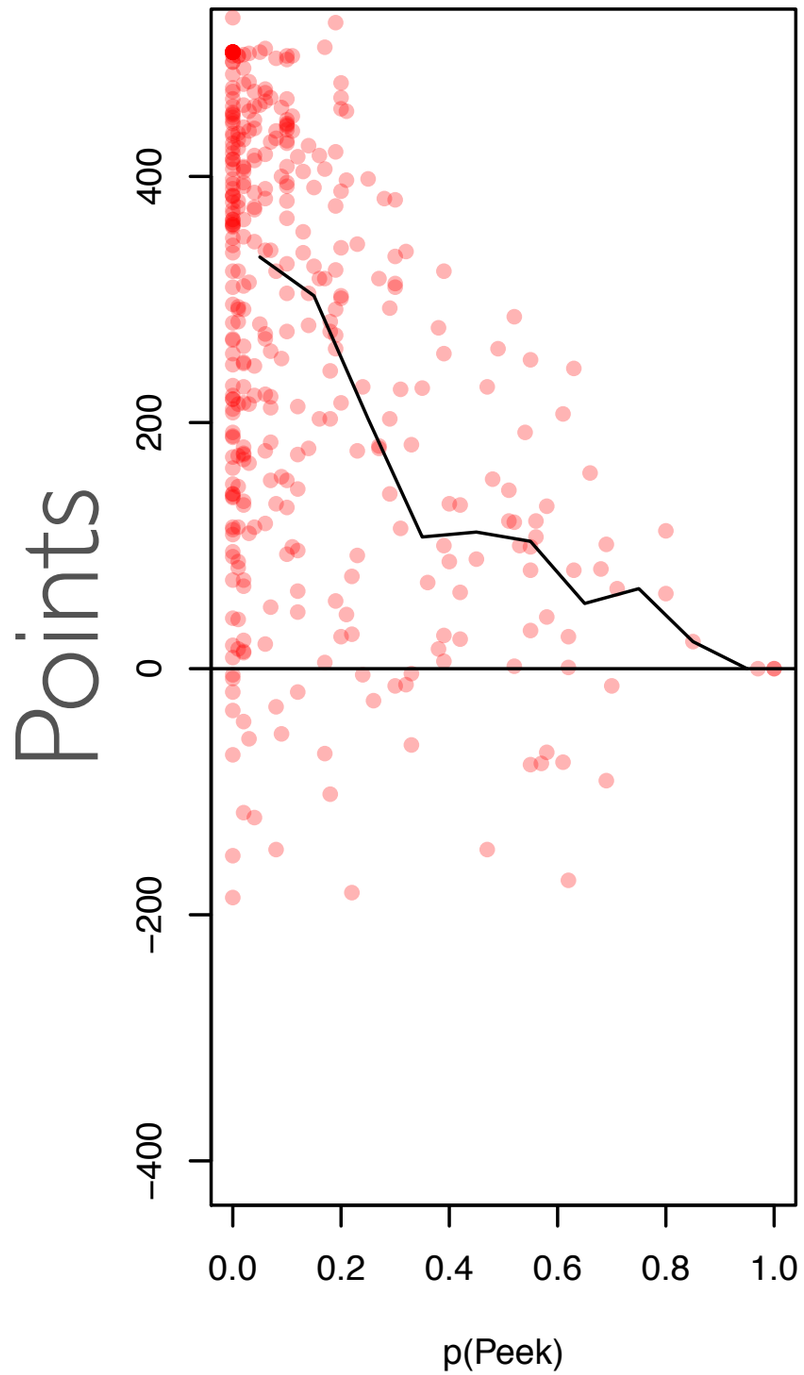


WAS THERE AN OPTIMAL PEEKING RATE IN THE DATA?

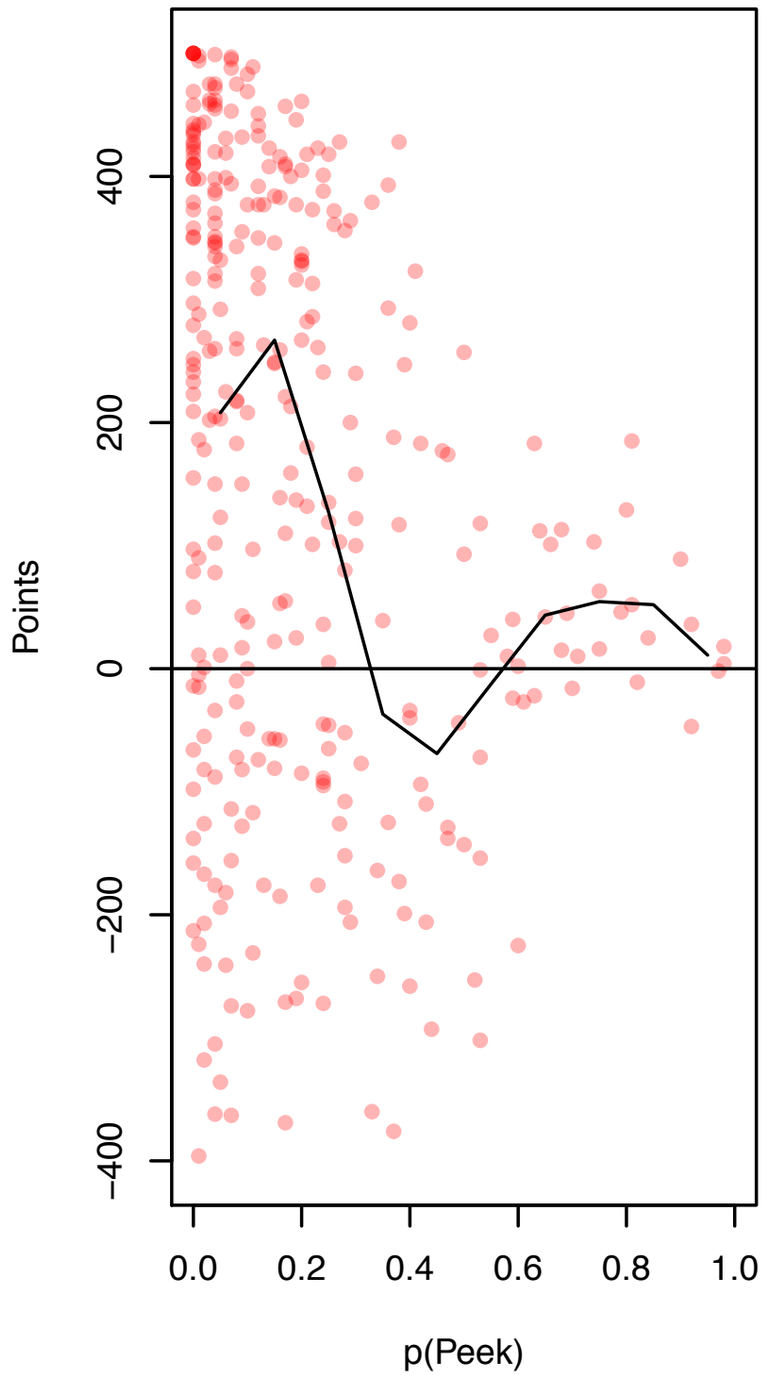
NO

Study 3

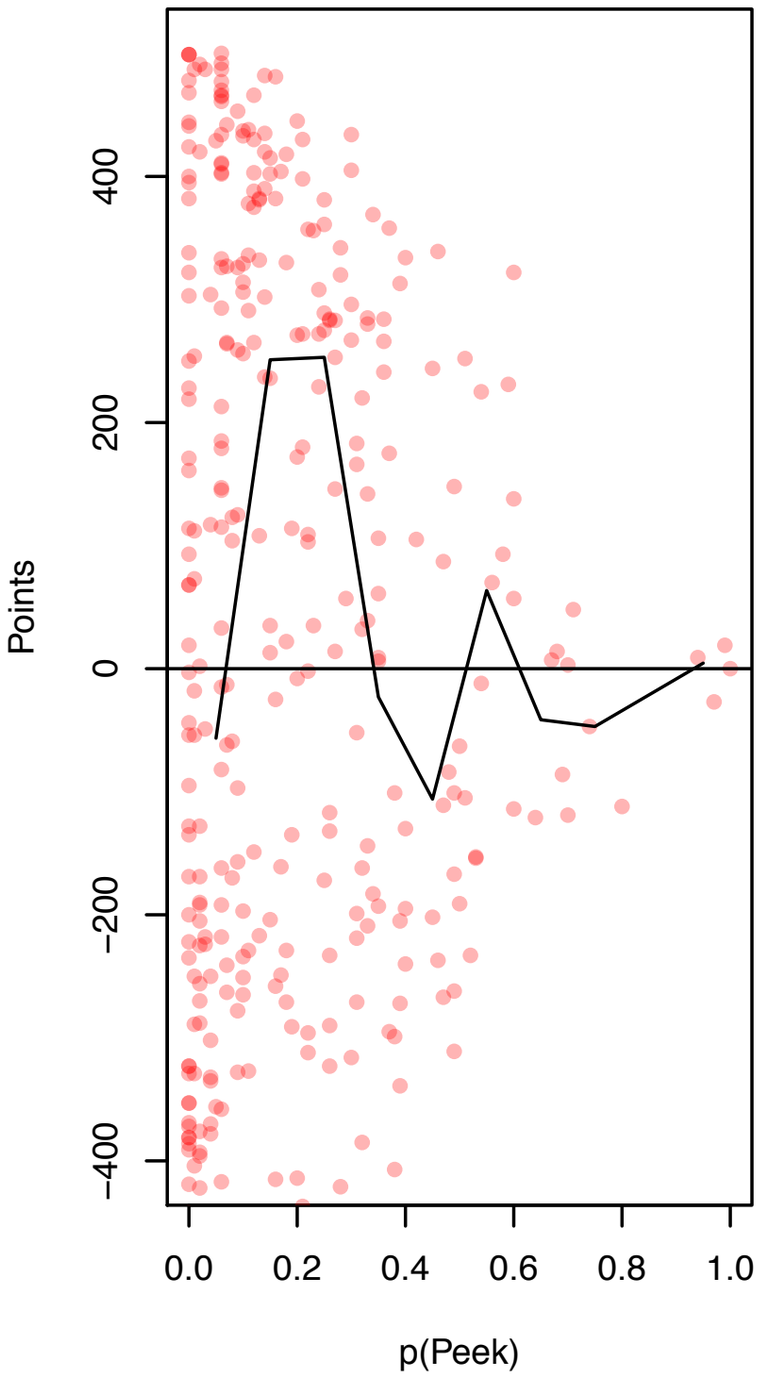
N = 2



N = 4



N = 6

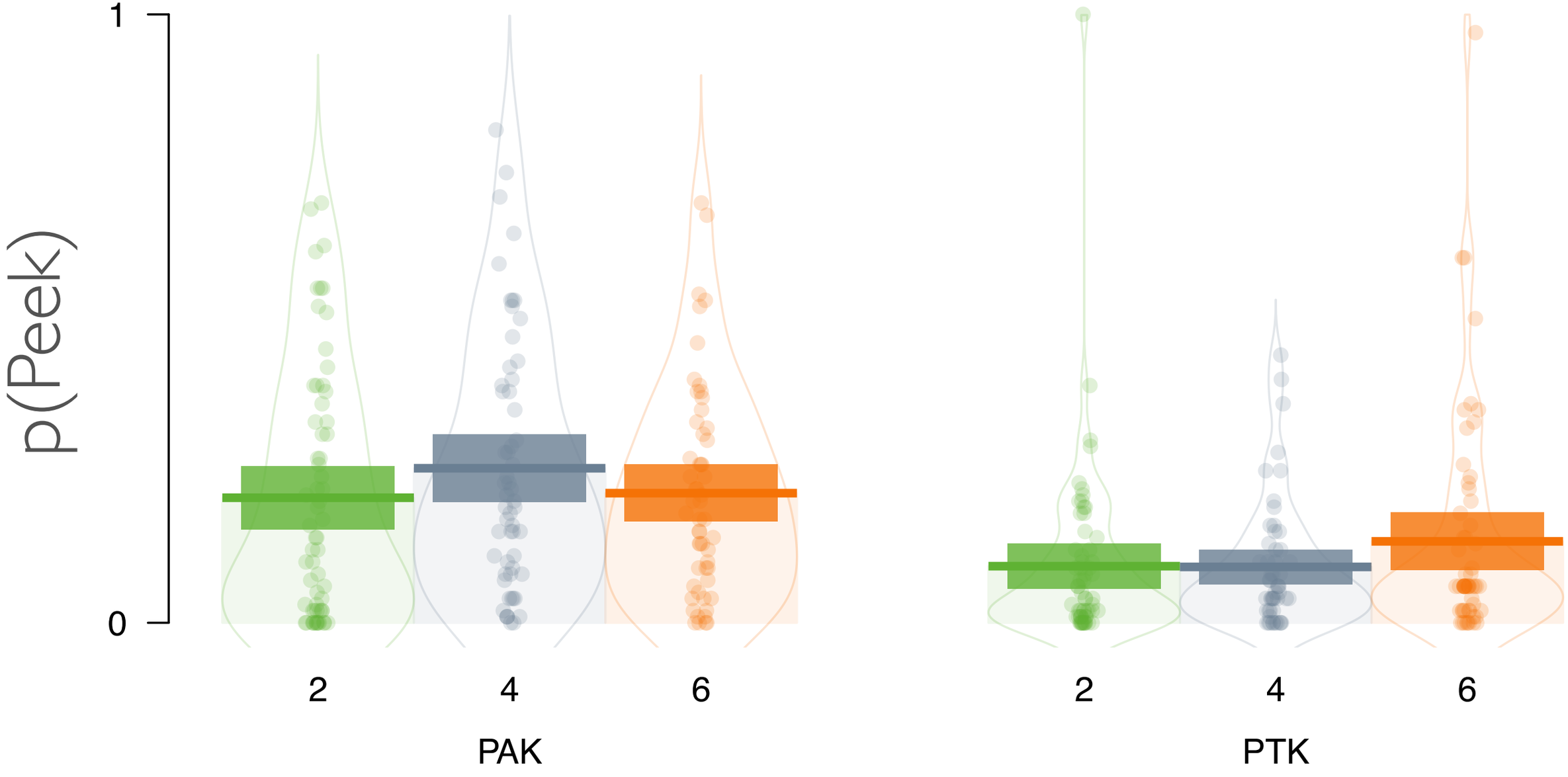


Peeking Rate

ARE PEEKING RATES AFFECTED BY THE ENVIRONMENT?

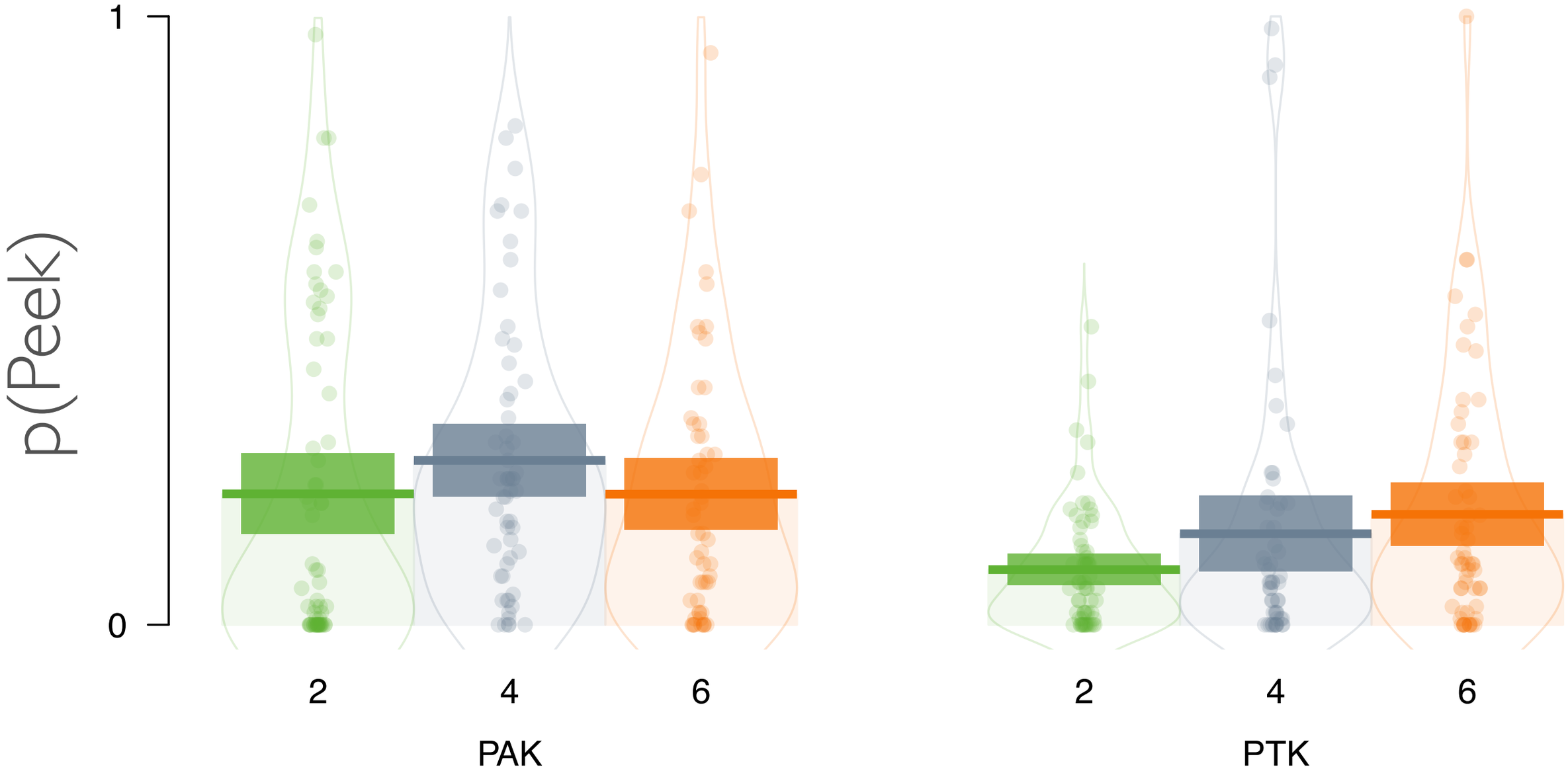
MAYBE

Study 3: Peeking Rates



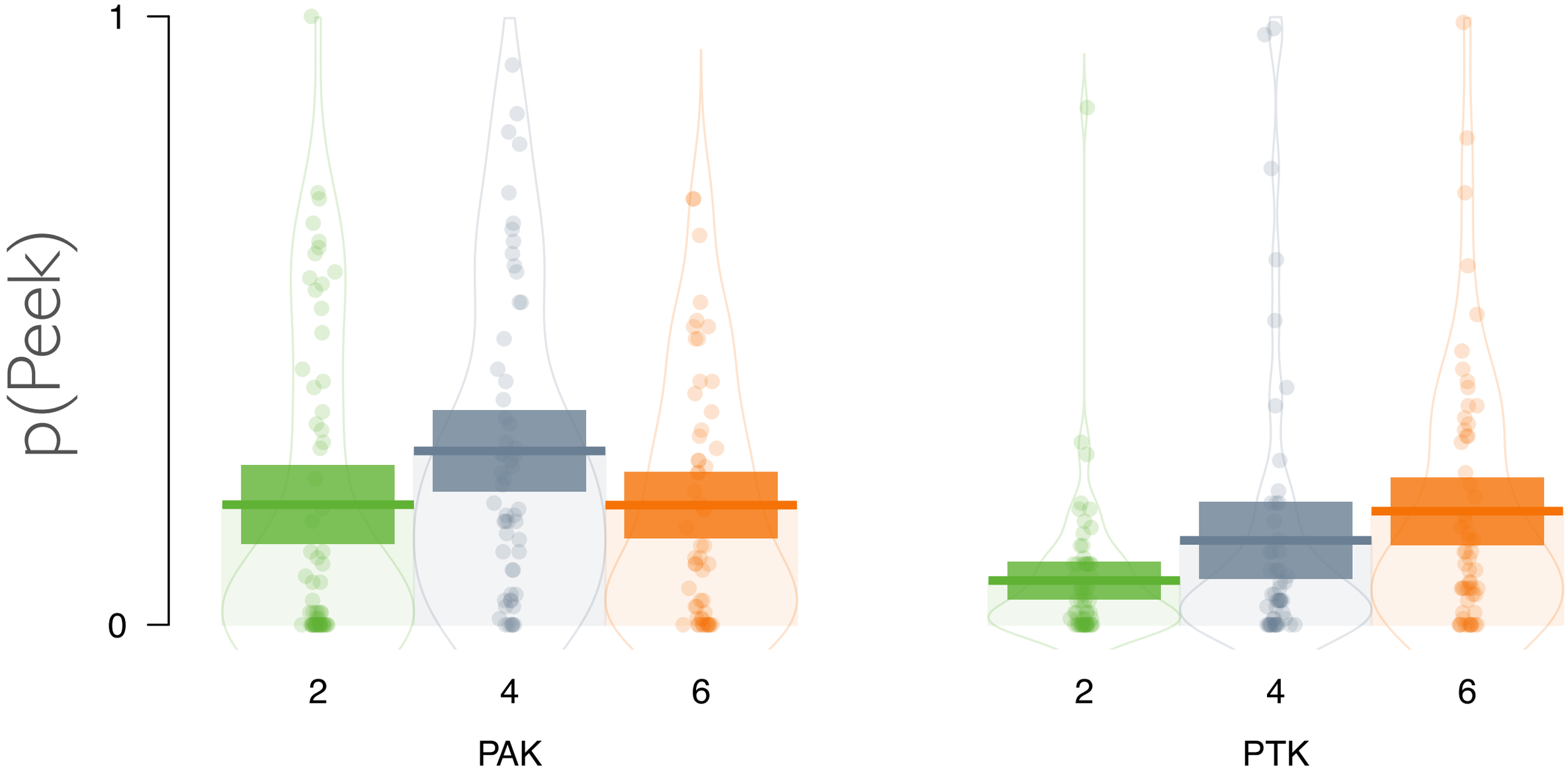
Game 1

Study 3: Peeking Rates



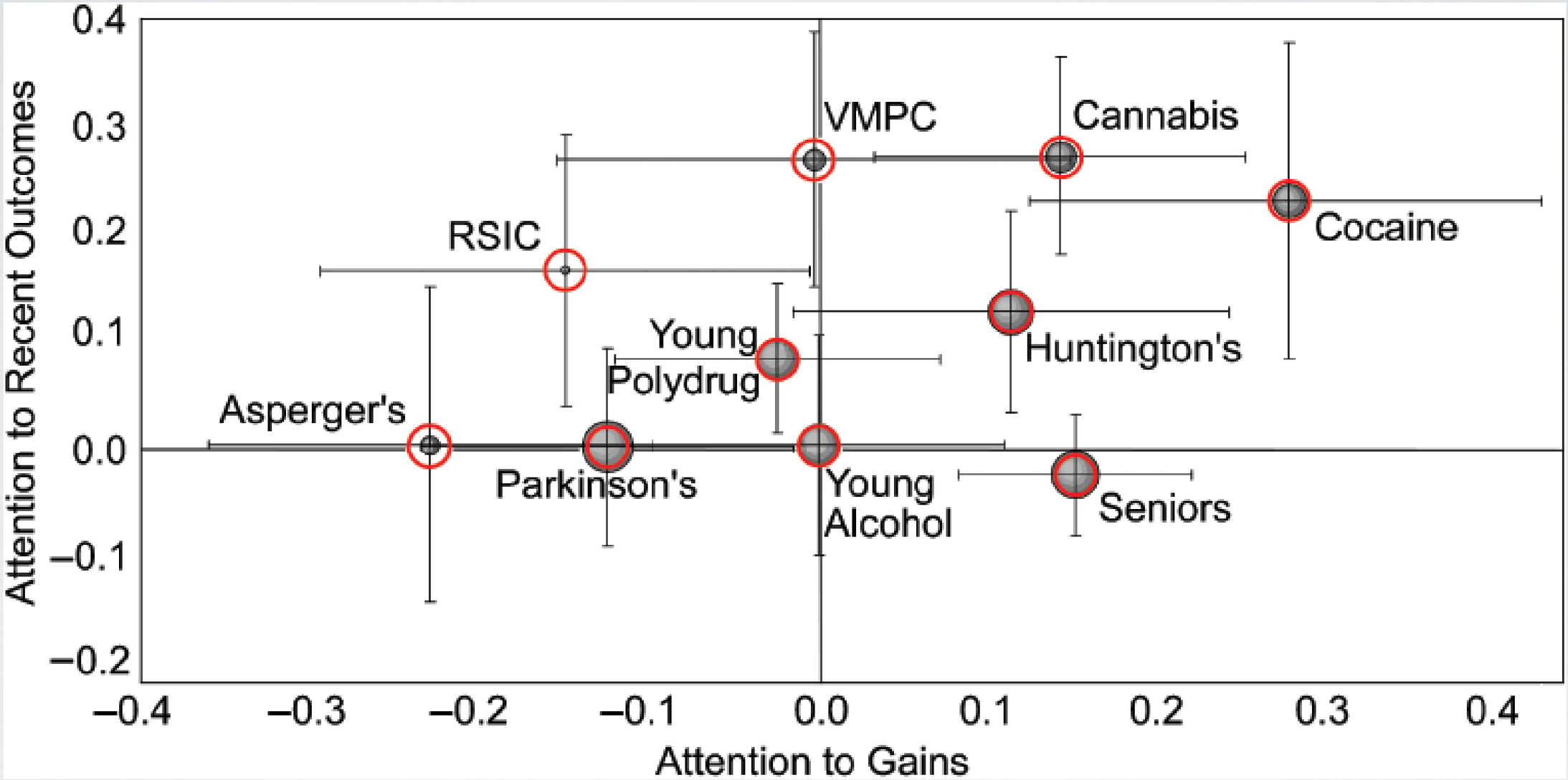
Game 2

Study 3: Peeking Rates



Game 3

INDIVIDUAL BEHAVIOR



Yechiam et al., 2005