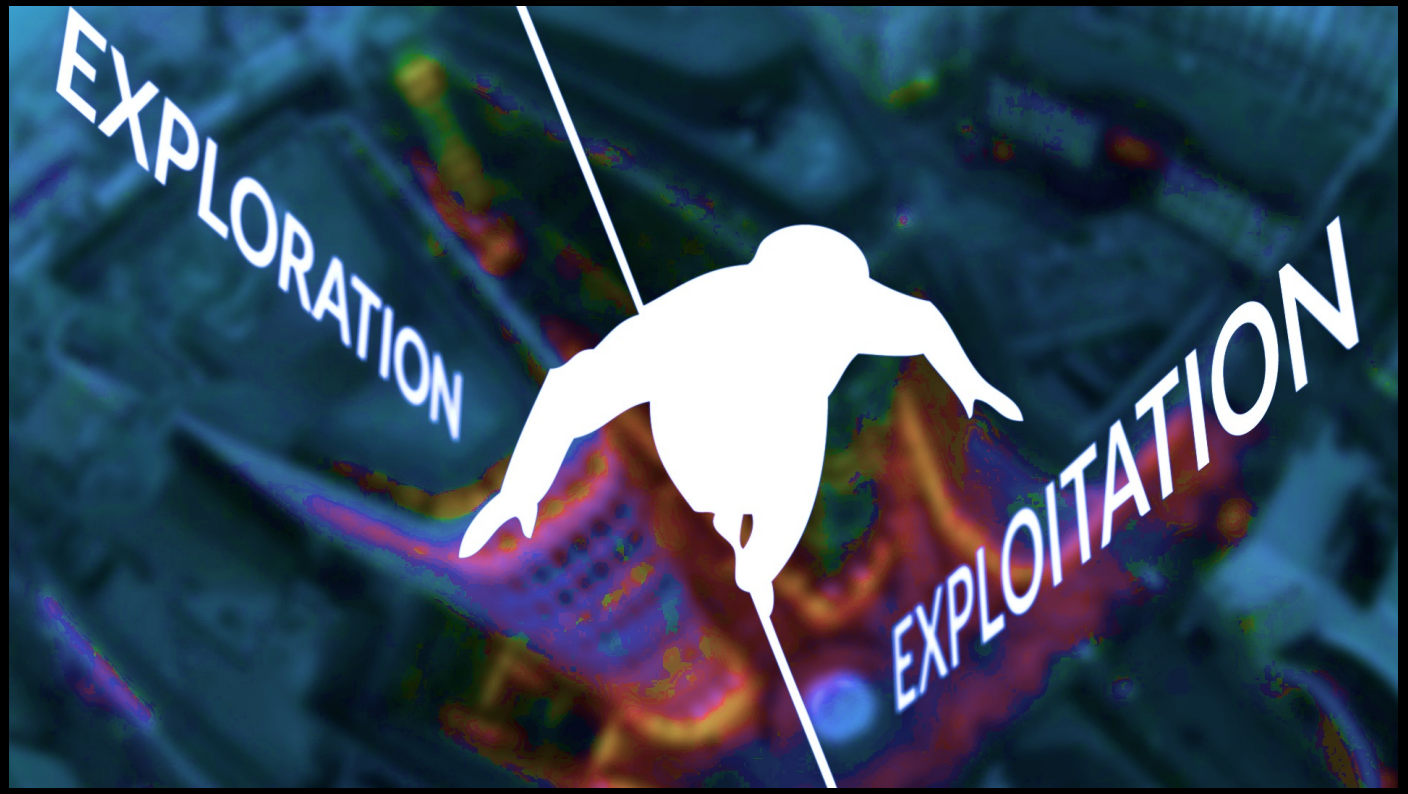
CENTER FOR RESEARCH IN
COGNITION & NEUROSCIENCES

ULB



Fontal Functions Drive Exploratory Strategies

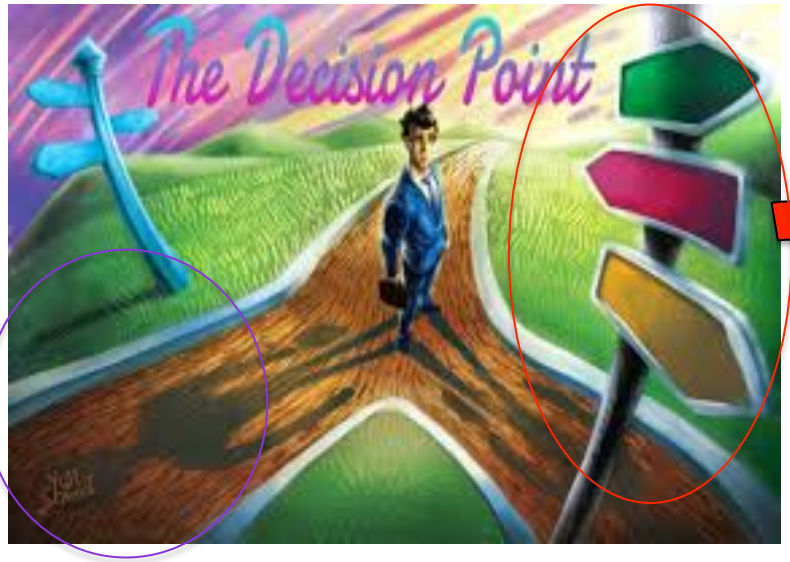
Irene Cogliati Dezza

CO3-Consciousness, Cognition and Computation Group, CRCN-Center for Research in Cognition and Neuroscience, UNI-ULB Neuroscience Institute , Université Libre de Bruxelles

Exploration and Exploitation Dilemma

Exploration and Exploitation trade-off is a dynamic CONFLICT between opposing demands of **gathering** *new information* **and exploiting a particular** information

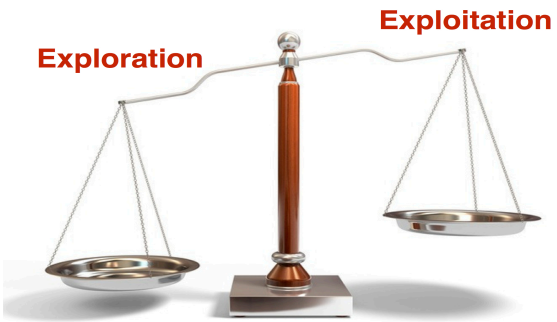
[Daw et al. 2006. Nature]



Current strategy
Exploitation

Sample **other options**
Exploration

Exploration and Exploitation Dilemma



*"For example, when you visit your favorite restaurant, do you choose to order **the same meal** that has brought you pleasure in the past, or do you **explore the menu** for a new dish that could perhaps become your new favorite?"*

[Donahue et al. 2013. Neuron]

Definition

→ Exploitation: "Make the choice of the current option as value as possible"

- Maximize a particular form of reward (for example, a particular meal)
- Doing as well as possible in the current trial ignoring the future, favoring immediate reward

[Dayan et al. 2008. Cognitive Affective,& Behavioral Neurosciences]

→Exploration: "Getting better outcome"

- Taking choices that might not be expected to payoff as much immediately, but might improve the prospect for learning reward on subsequent trials

[Dayan et al. 2008. Cognitive Affective,& Behavioral Neurosciences]

- Exploring is of primary importance for adapting to new situation in a changeable environment: a) **static world**, after learning contingencies, decision-maker has complete knowledge about rules and states and he will only exploit; b) **volatile world**, reward contingencies change and he will explore to gather new information.

Measuring the trade-off in a lab

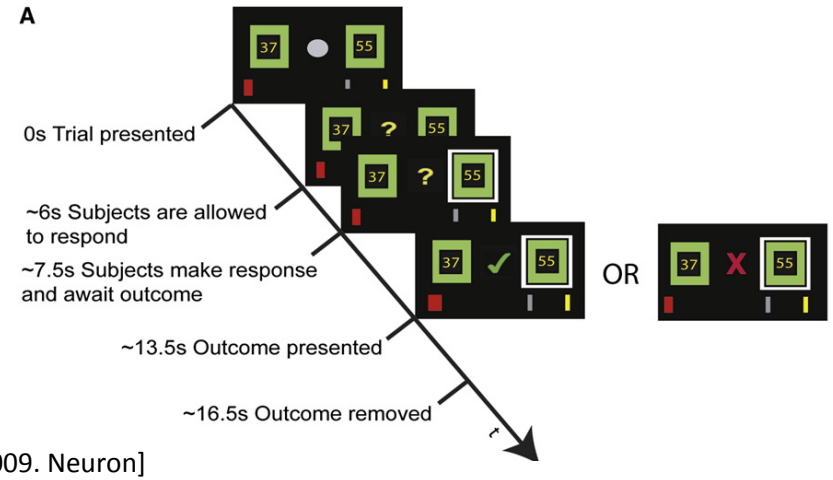
Multi-Armed Bandit Task

" [..] *A gambler has to decide which machine to play, how many times and in which order. When played, each machine provides a random reward from a distribution specific to that machine. The objective of the gambler is to* ***maximize the sum of rewards earned*** *through a sequence of level pulls..* " [Robbins, 1952]

Measuring the trade-off in a lab

2-Armed Bandit Task

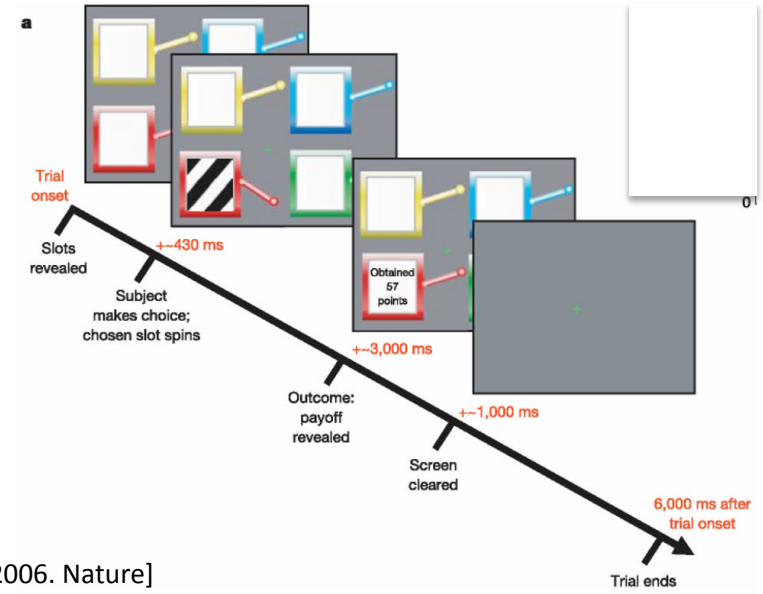
Choosing between right and left options on the basis of past outcomes and the reward magnitudes



[Boorman et al. 2009. Neuron]

4-Armed Bandit Task

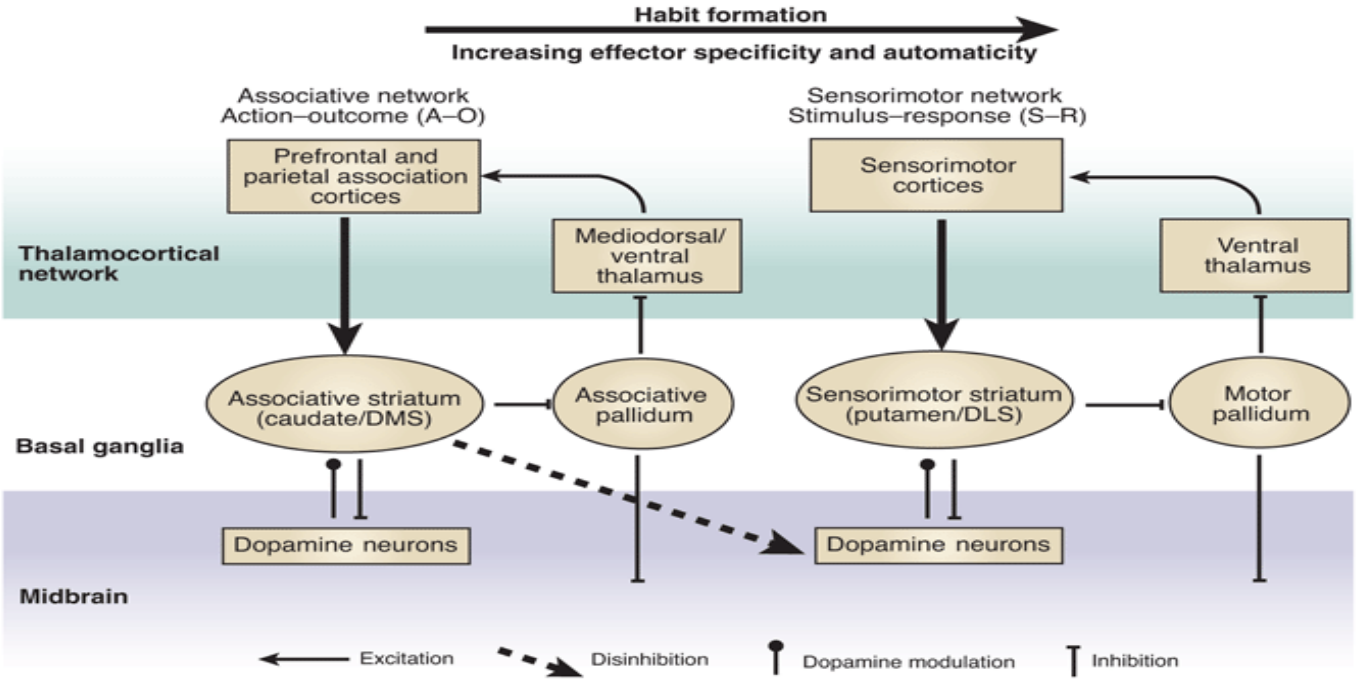
Sequentially choosing between four slot machines



[Daw et al. 2006. Nature]

Driving Exploitation

- Formation of an habitual action (Cortical/ Basal-Ganglia loops): Stimulus-Response behavior despite reward devaluation [Graybiel et al. 2008. Annu. Rev. Neurosci]



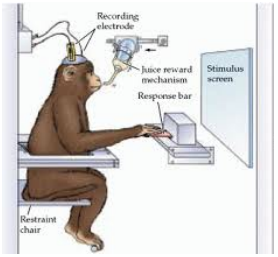
- ACC modulating the shift from exploration to exploitation [Quilodran et al. 2008. Neuron]

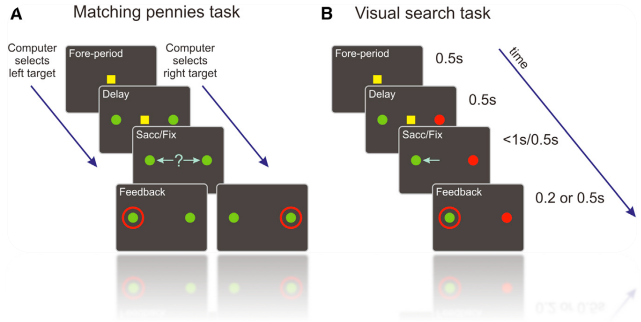
- Computationally: Optimal Bayesian Model [Steyvers et al. 2009. Journal of Mathematical Psychology], reinforcement learning and psychological models (ε-greedy, π-First, latent State model [Lee et al. 2011. Cognitive System research]

Driving Exploration

SEF's neurons were correlated the animal's tendency to explore

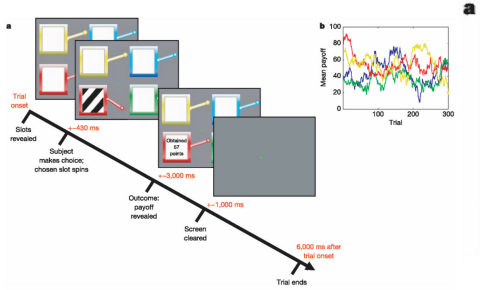
[Donahue et al. 2013. Neuron]

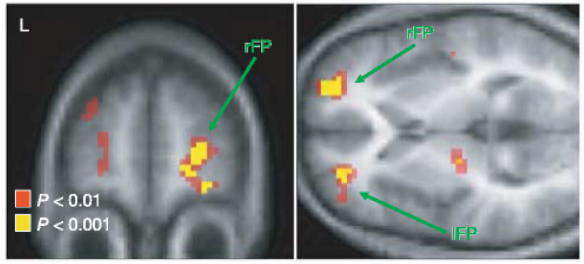




Exploration-related activity in frontopolar cortex (Regions of left and right frontopolar cortex (lFP, rFP) showing significantly increased activation on exploratory compared with exploitative trials

[Daw et al. 2006. Nature]





Exploitation

Most rewarded action
(after learning reward contingencies)

?

Cognitive Processes
Neuronal Substrates
Neuromodulatory factors
Computational mechanisms

Exploration

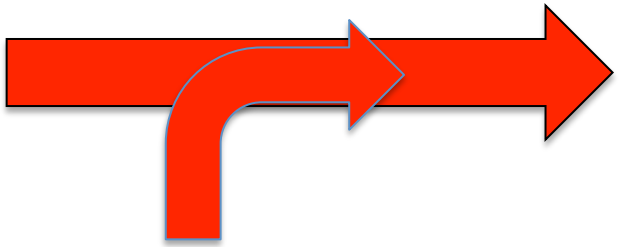
New action never, or less recently, associated to reward

Hypothesis

Frontal processes might play a pivotal role in driving exploratory decisions

Exploitation

Most rewarded action



Exploration

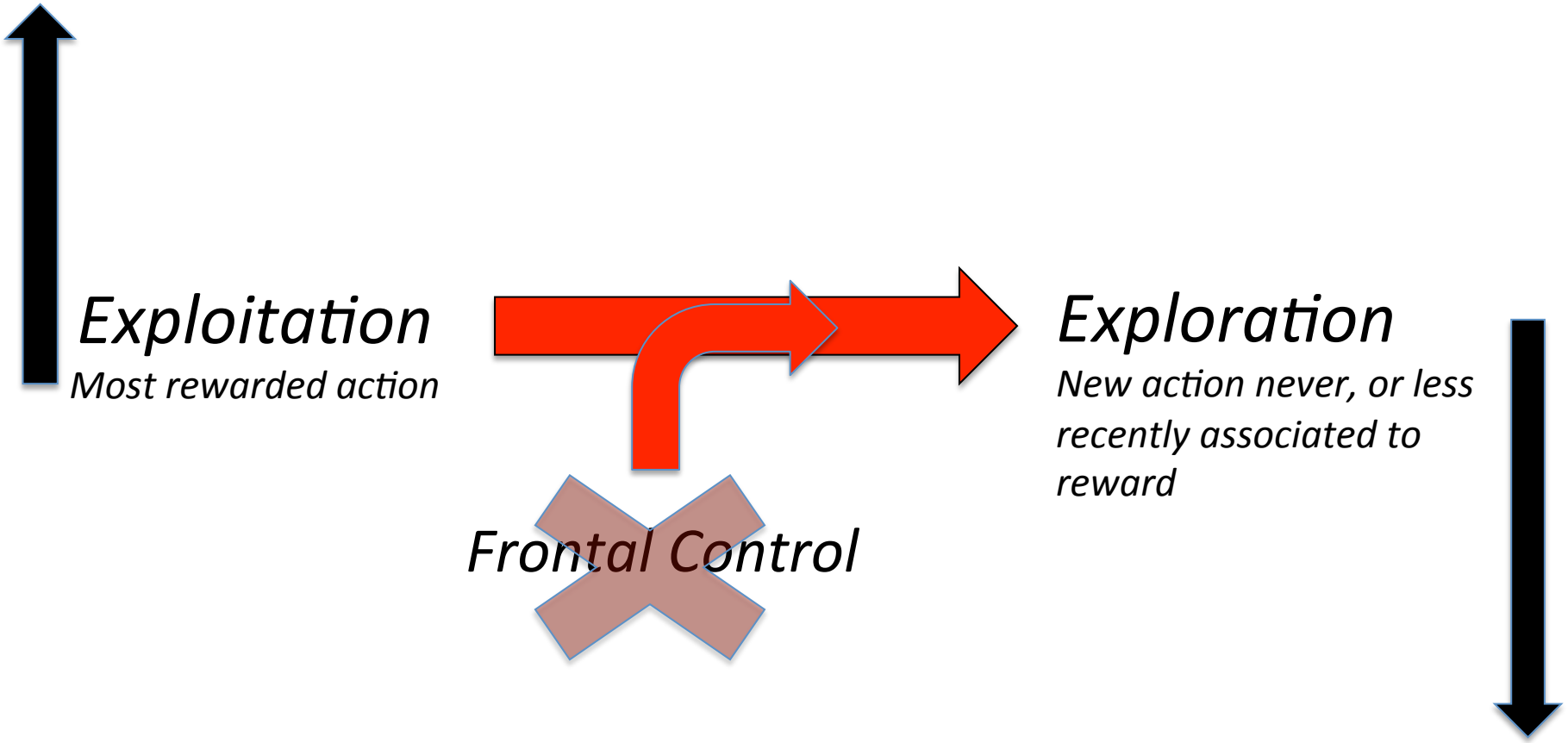
New action never, or less recently associated to reward

Frontal Control

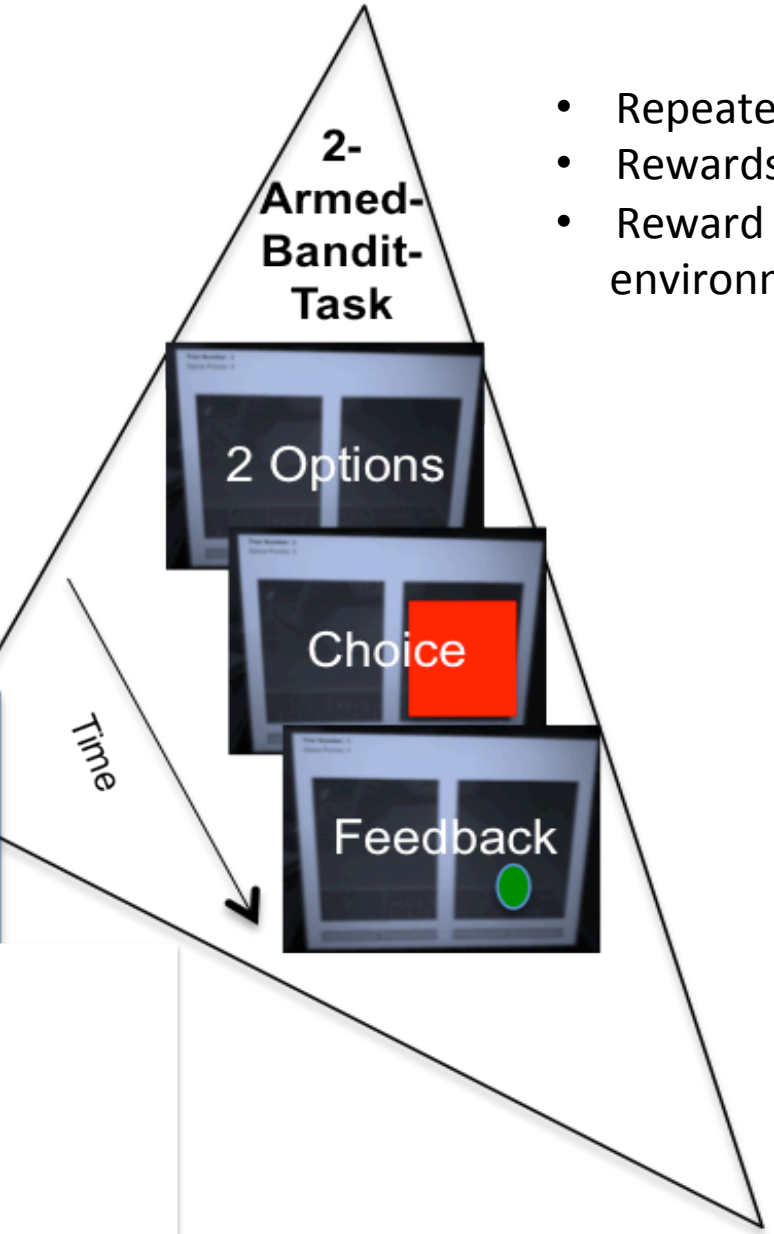
Overloading Approach

- We asked participant to play with the 2-Armed Bandit Task
- Overloaded working memory using a secondary task

[Lavie et al. 2004. Journal of Experimental Psychology; Konstantinuo et al. 2013. Journal of Experimental Psychology]

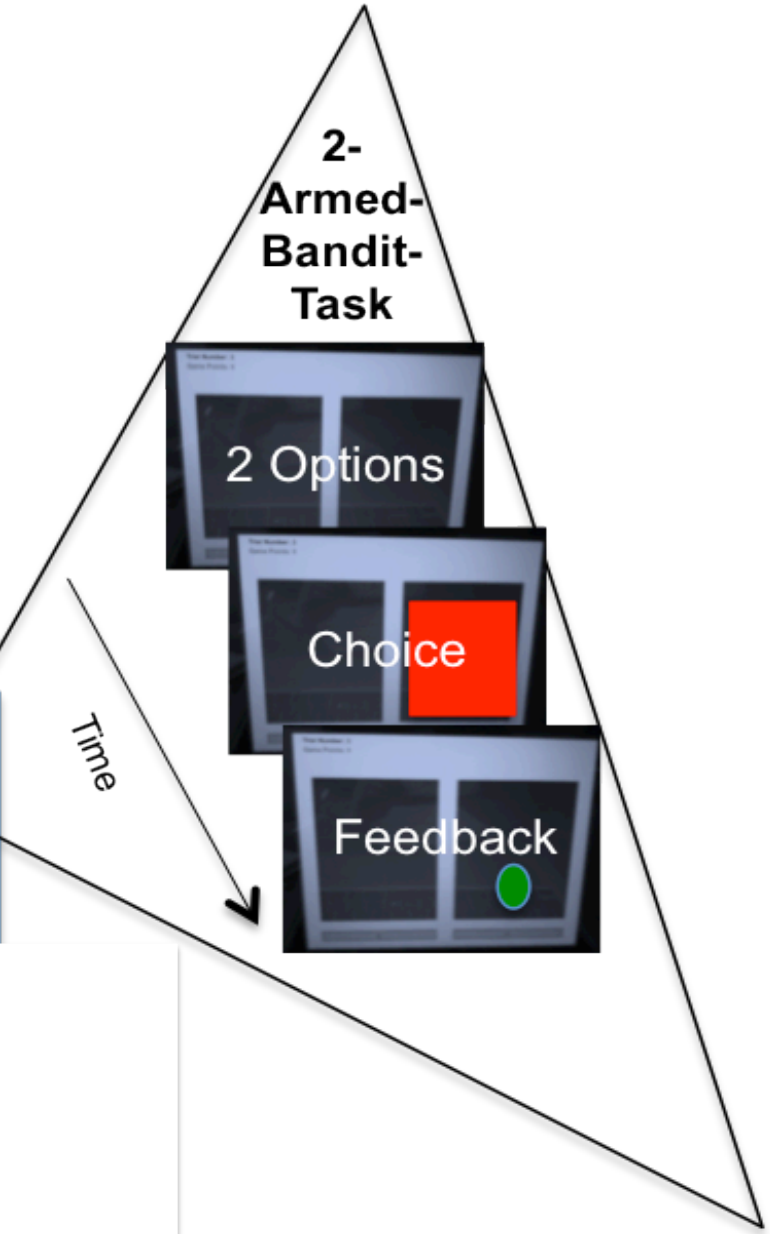


2-Armed Bandit Task

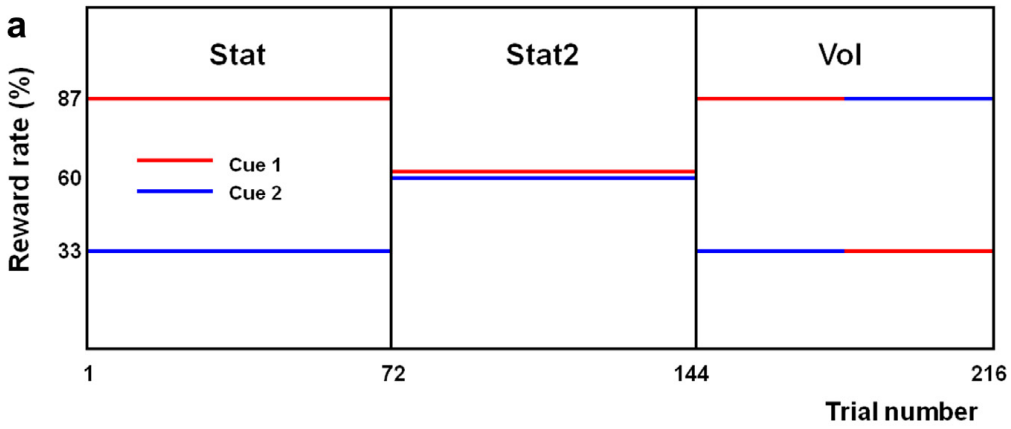


- Repeatedly choose between left and right option (200 trials)
- Rewards are binary (0 point, 1 point)(=/ Boorman's bandit)
- Reward probability changes every 40 trials (simulating a volatile environment) (=/ Daw's bandit)

2-Armed Bandit Task

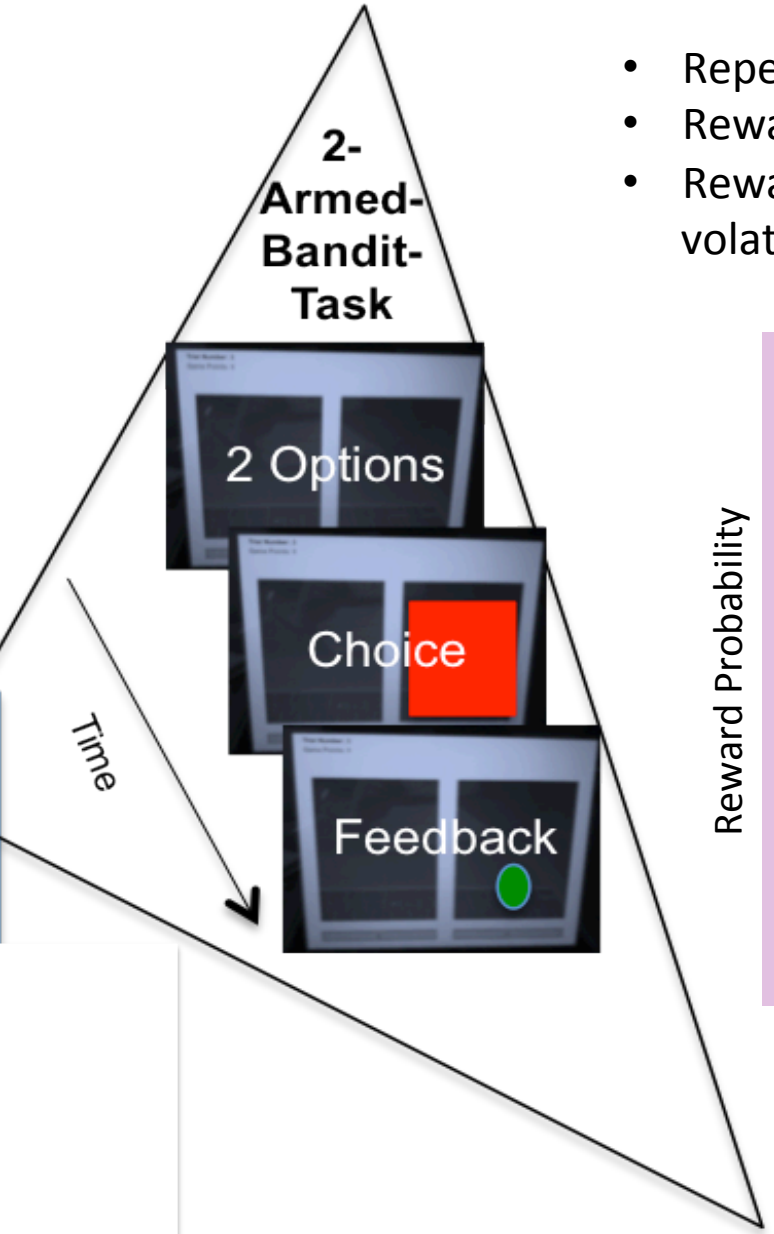


[Silvetti et al. 2012. Cortex]

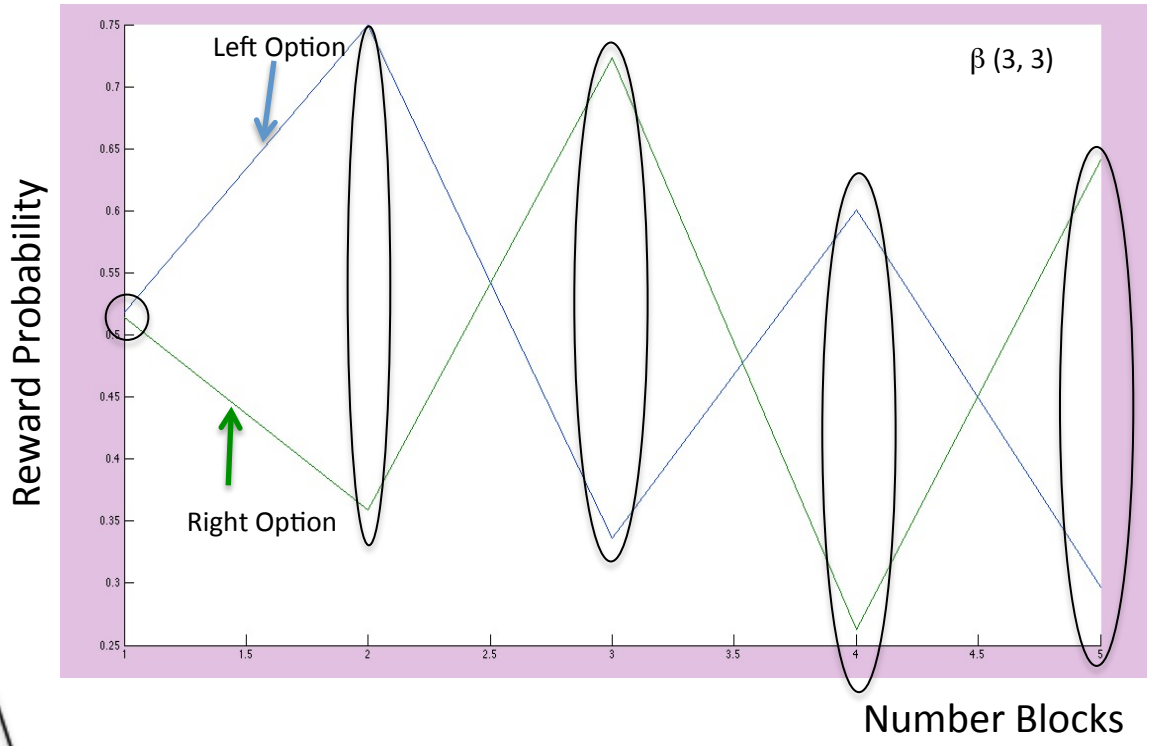


Volatile environment : predictive links between cues and outcomes change over time.

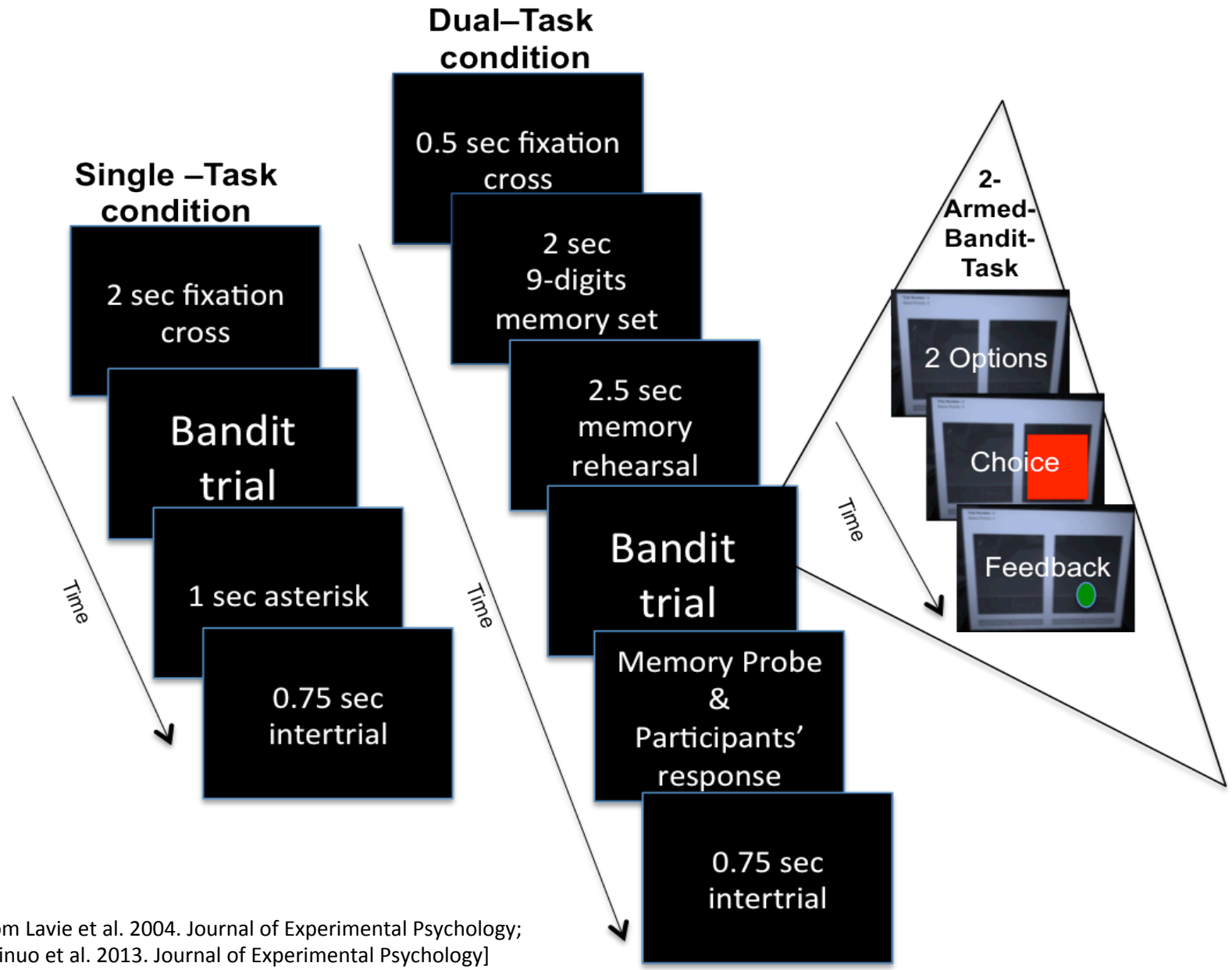
2-Armed Bandit Task



- Repeatedly choose between left and right option (200 trials)
- Rewards are binary (0 point, 1 point)
- Reward probability changes every 40 trials (simulating a volatile environment)

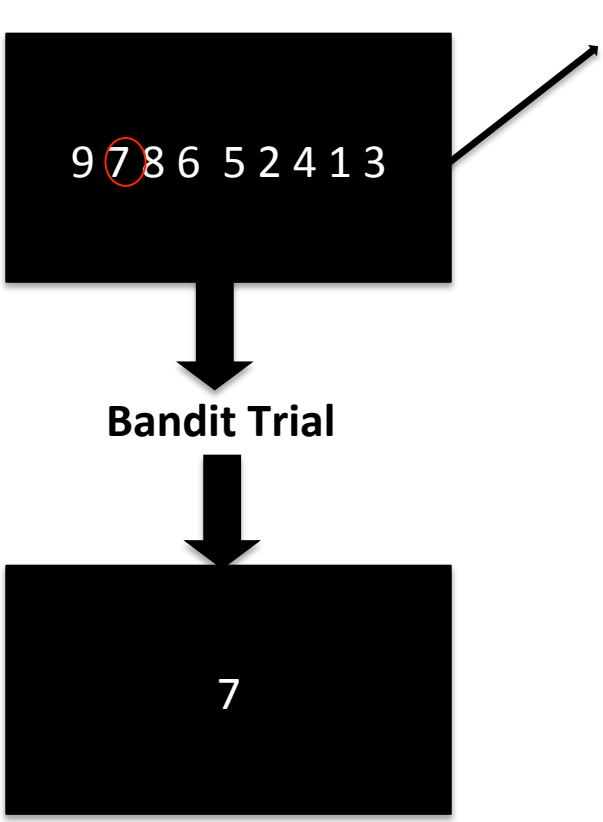


Paradigm



[Adapted from Lavie et al. 2004. Journal of Experimental Psychology; Konstantinuo et al. 2013. Journal of Experimental Psychology]

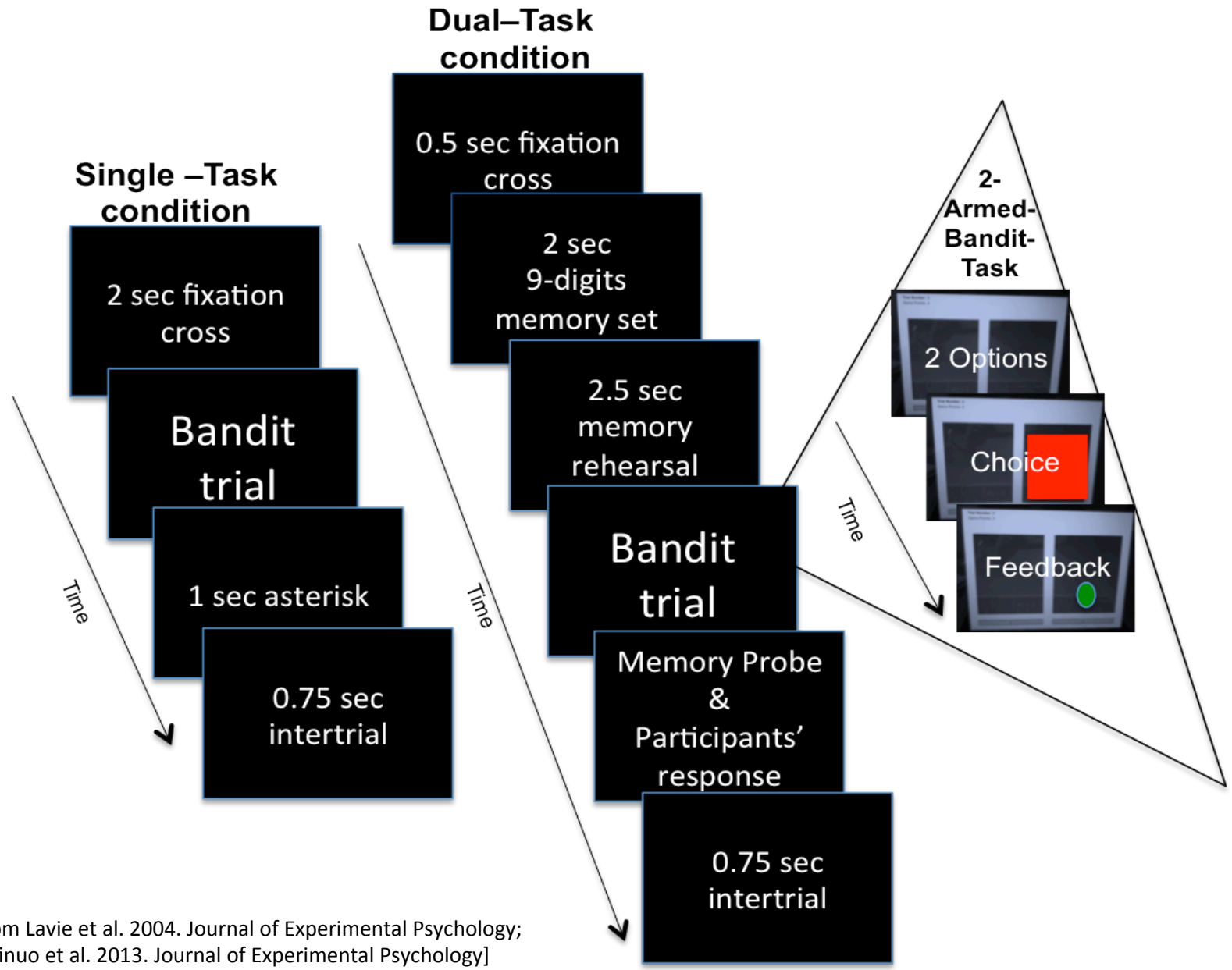
Paradigm



Correct Answer: 8

[Adapted from Lavie et al. 2004. Journal of Experimental Psychology; Konstantinuo et al. 2013. Journal of Experimental Psychology]

Paradigm



[Adapted from Lavie et al. 2004. Journal of Experimental Psychology; Konstantinuo et al. 2013. Journal of Experimental Psychology]

Participants

- ✓ 57 Participants (age from 18-30 y-o)
- ✓ Inclusion criteria: 1) without an history of psychiatry disorders, 2) or sleep problems, 3) without use of psychoactive drugs in the last year, 4) and without the use of alcohol the day before the experiment
- ✓ Written instruction
- ✓ Motivation: pay attention to both tasks (gambling and memory task) as the only strategy to increase their final payoff (ranging from 9 to 12 euros)
- ✓ Condition Order: randomly assigned to single-dual or dual-single order
- ✓ Levenson's LOC SCALE: measure external and internal Locus of Control [Rossier et al. 2002. Ann Med Psych]
- ✓ EYE-BLINKING: 4 minutes

Preliminary Results

- Preliminary result over 44 participants (exploitation power {1-β}=0.8402;exploration power {1-β}=0.7481)
- Age (mean= 22.18, SD=2.423)
- Inclusion criteria memory score (range mean+/-2SD [0.4445+/- 0.1431775])

- Exploratory and Exploitative trials computed using the **Softmax Rule** [Daw et al. 2006. Nature]

$$P_{i,t} = \frac{\exp(\mu_{i,t})}{\Sigma\exp(\mu_{j,t})}$$

Mean reward for option *i* until time *t*

- Choices with higher $P_{i,t}$ was defined as exploitative whereas choices with lower $P_{i,t}$ explorative.
- For Example:

  $P_{1,30}>P_{2,30}$ decision=1: EXPLOITATIVE

  $P_{1,31}>P_{2,31}$ decision=2: EXPLORATIVE

Preliminary Results

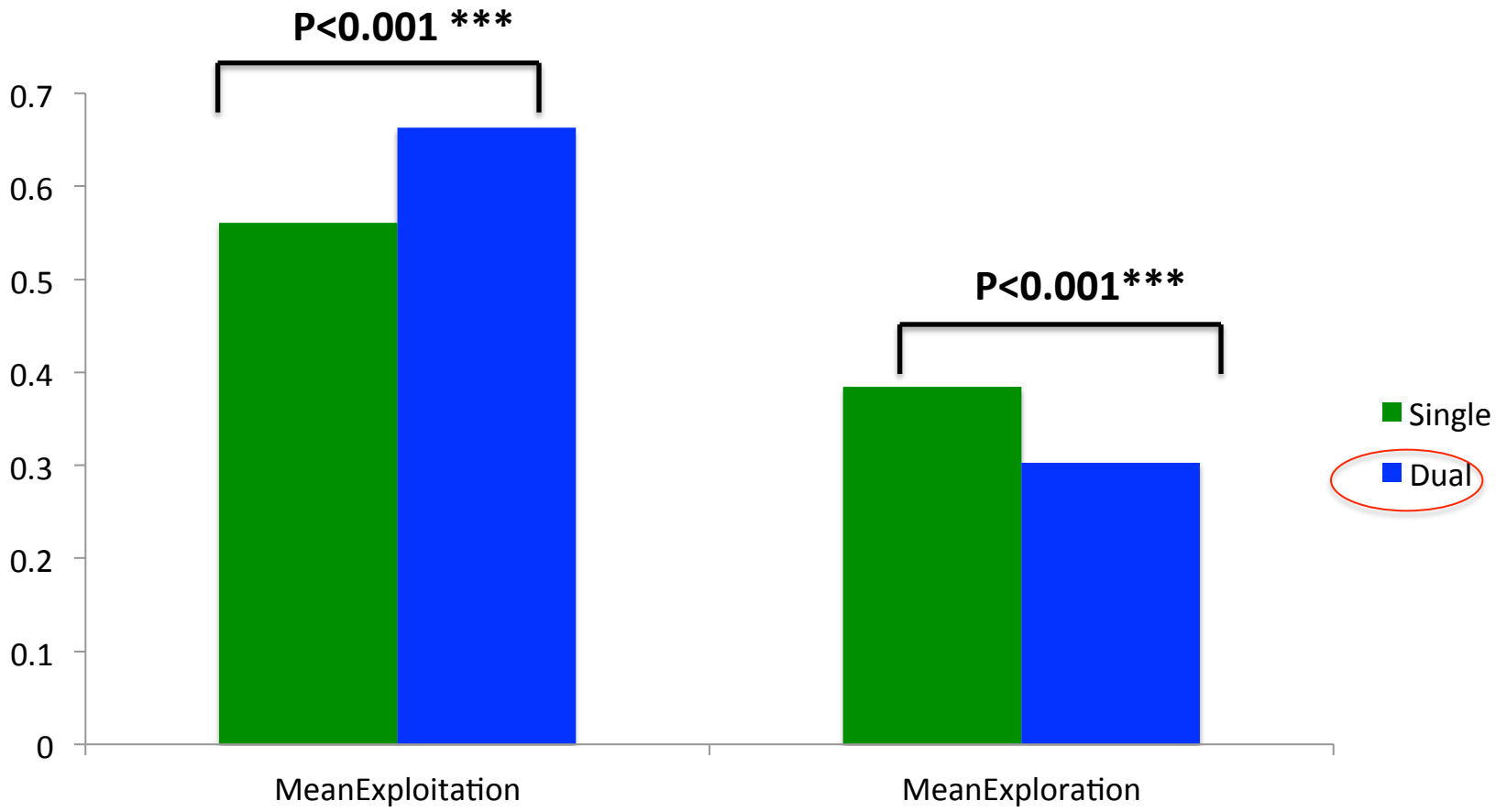
Within Subject Analysis

Exploitation:

SINGLE (mean= 0.560795, SD=0.0693235) vs. **DUAL** Condition (mean=0.662614,SD=0.1417806)

Exploration:

SINGLE (mean= 0.384432, SD=0.0624892) vs. **DUAL** Condition (mean=0.302955,SD=0.1275499)



Preliminary Results

Within Subject Analysis

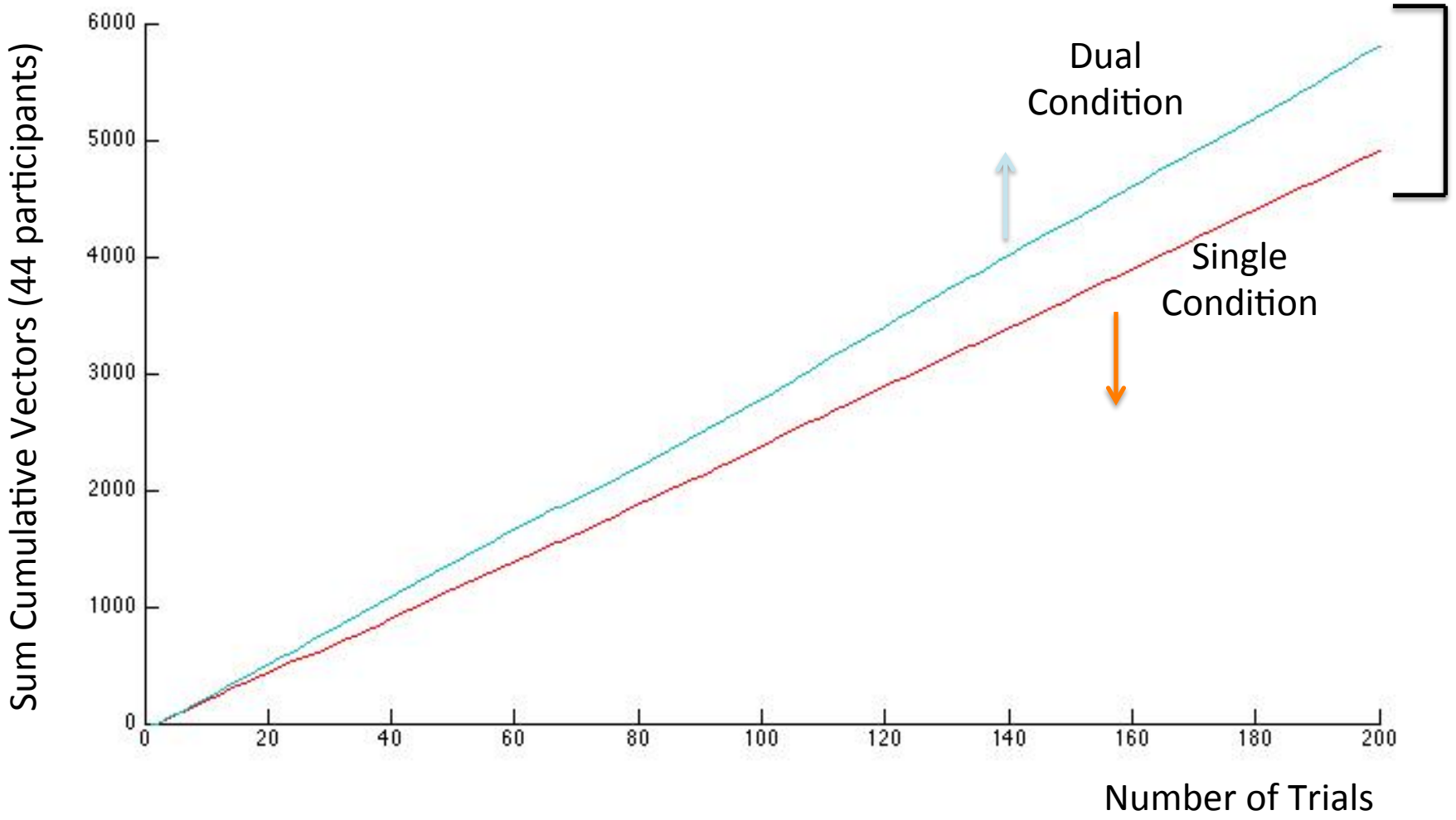
Exploitation

SINGLE (mean= 54.730568, SD=6.3601854)

vs.

DUAL Condition (mean=64.643182 ,SD=14.0877907)

P<0.001 ***



Preliminary Results

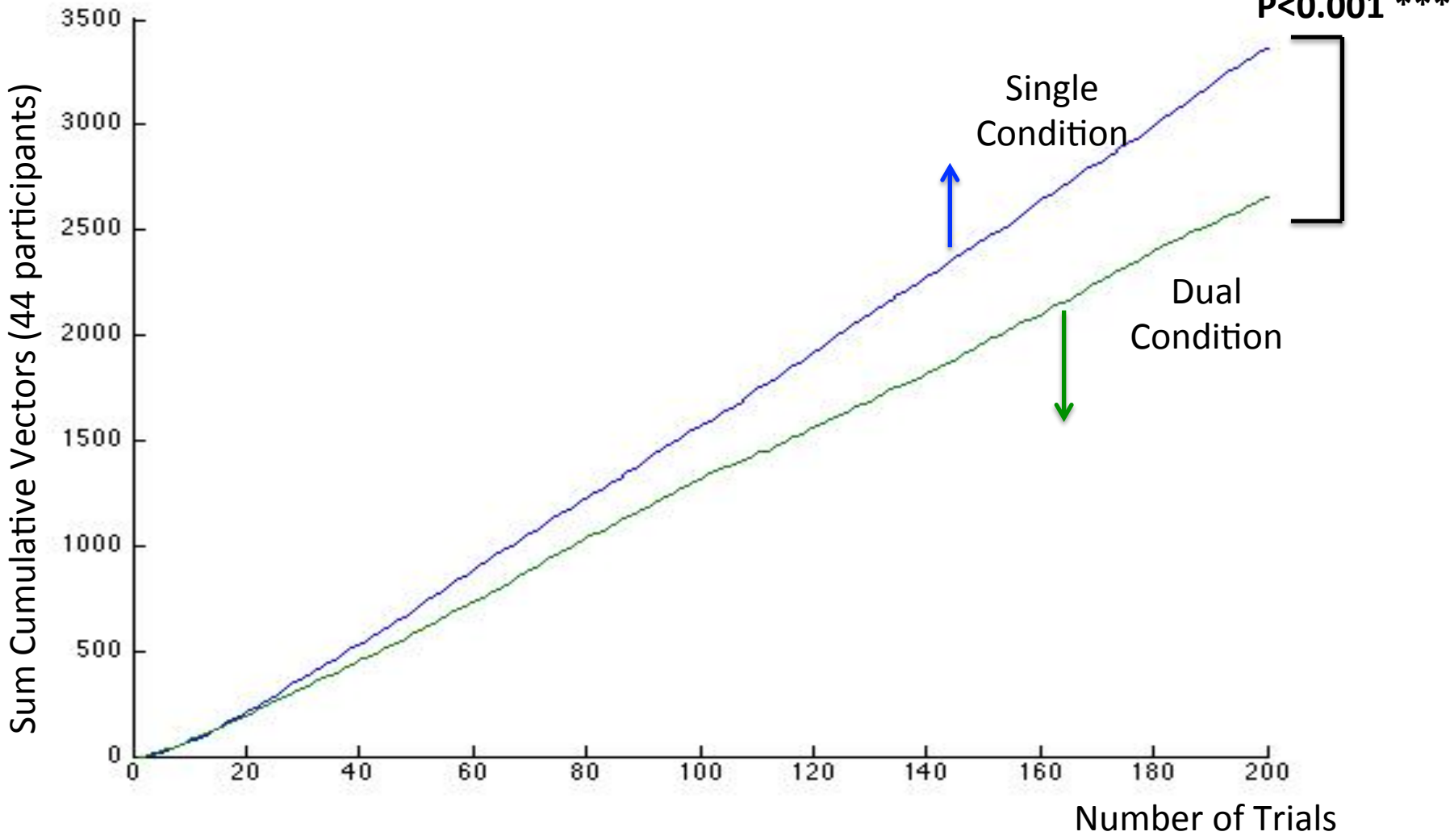
Within Subject Analysis

Exploration

SINGLE (mean= 36.308864, SD=4.6747644)

vs.

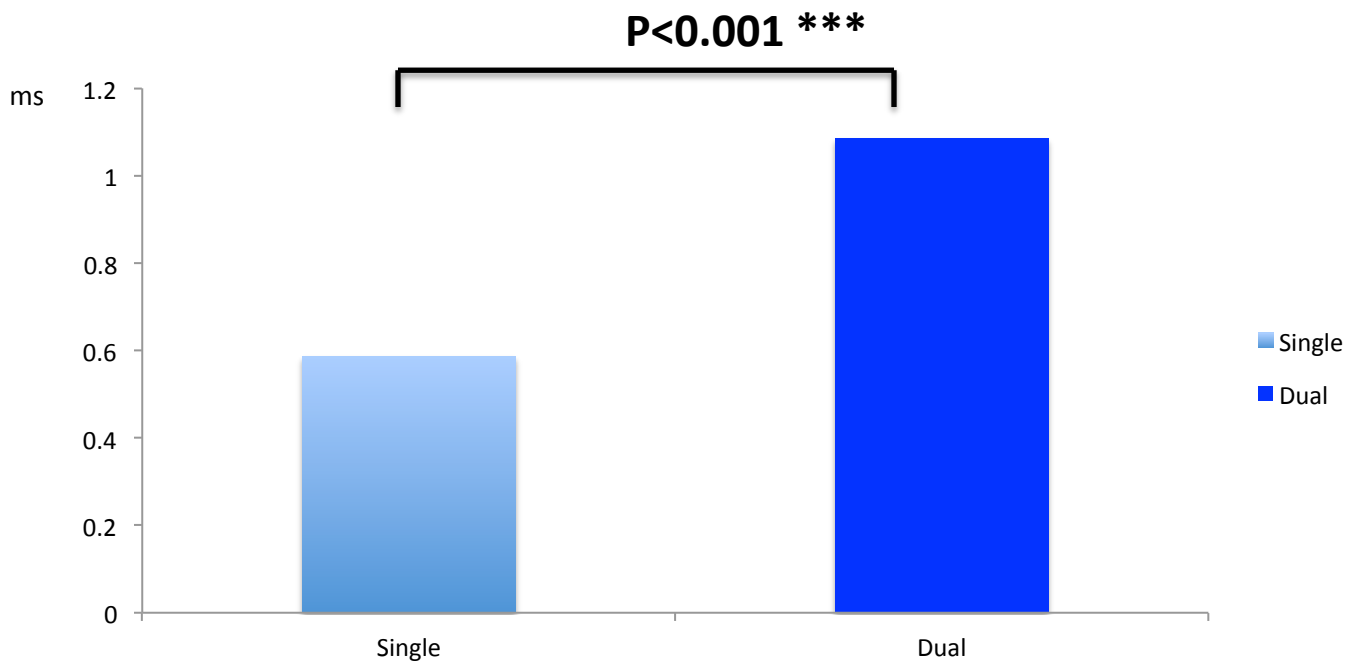
DUAL Condition (mean=29.477500,SD=12.0347735)



Preliminary Results

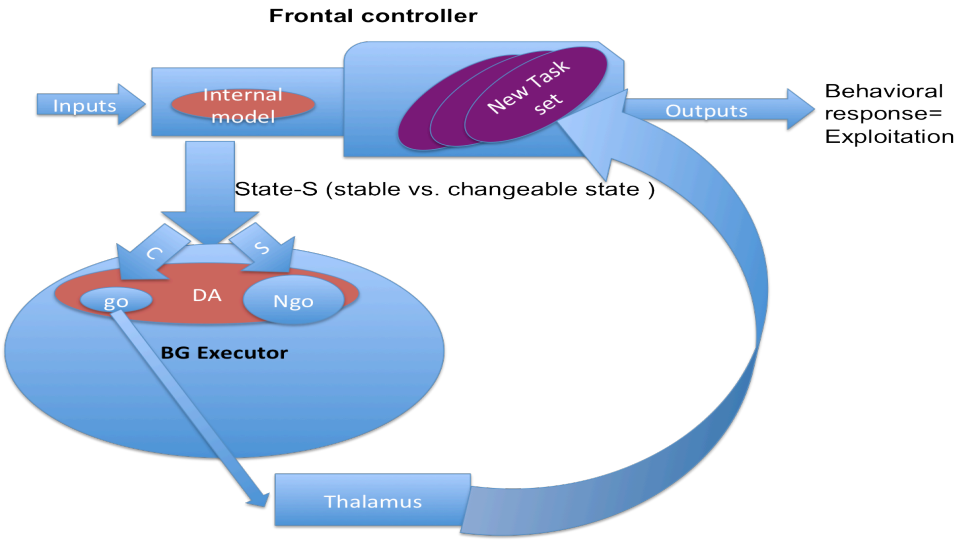
Within Subject Analysis

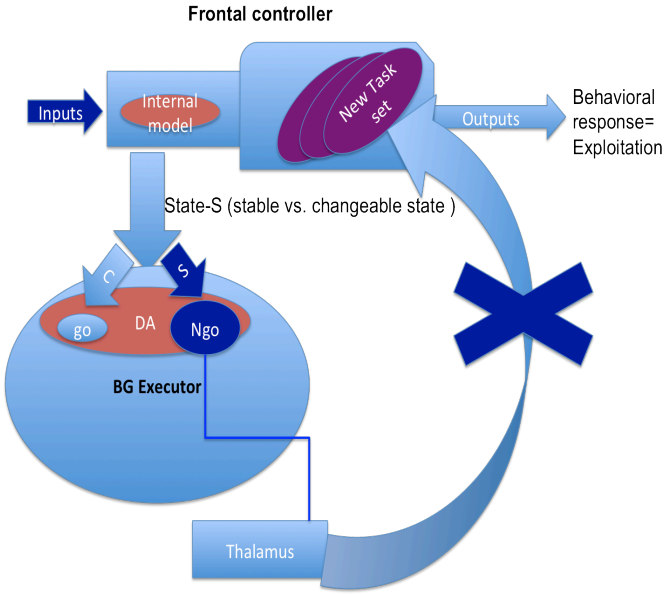
- Increase tendency to exploit in both condition (exploration single vs. exploitation single $p<0.001$; exploration dual vs. exploitation dual $p<0.001$)
- No differences in **condition order** (Exploitation Single $p=0.773$; Exploitation Dual $p=0.337$; Exploration Single $p=0.289$; Exploration Dual $p=0.304$)
- RT Dual (mean=1.085309,SD=0.1018005)roughly double RT Single (mean=0.586261,SD=0.0274280)

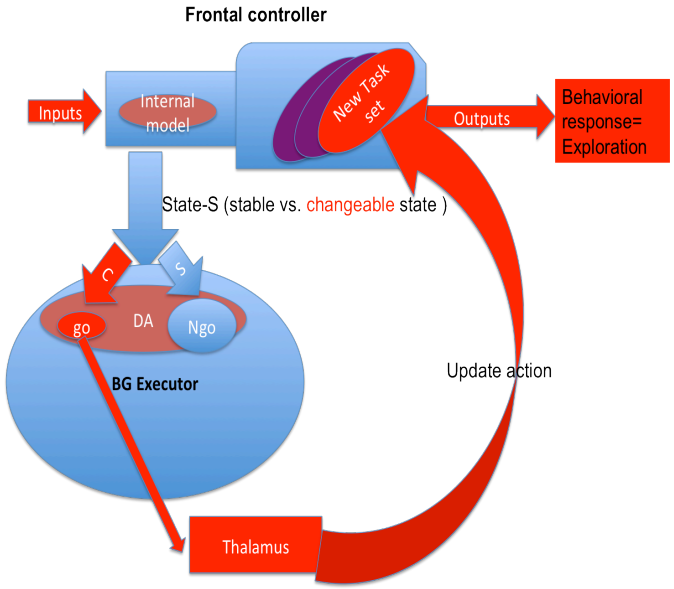


Future Perspectives

- How does frontal function compute exploratory strategy? Model-based model?







- Is the Frontal output dopaminergic in nature?

Besides Neuroscientifc interest

- Addiction as exploitative behavior: repeated choices (compulsive behavior) favoring immediate reward over delayed reward (impulsive system), and decreasing control over behavior (reflective system)

- Understanding the cognitive and neuronal mechanisms behind this switch could be helpful to develop therapies (cognitive and pharmacological) where repeated and compulsive behaviors are the common hallmark (i.e. drug and behavioral addiction)

To be continued…

Thanks to

Axel Cleeremans

NFSR Research Director
Consciousness, Cognition & Computation Group (**Co3**)
Center for Research in Cognition & Neurosciences (**CRCN**)
ULB Institute of Neurosciences (**UNI**)

Lua Koenig

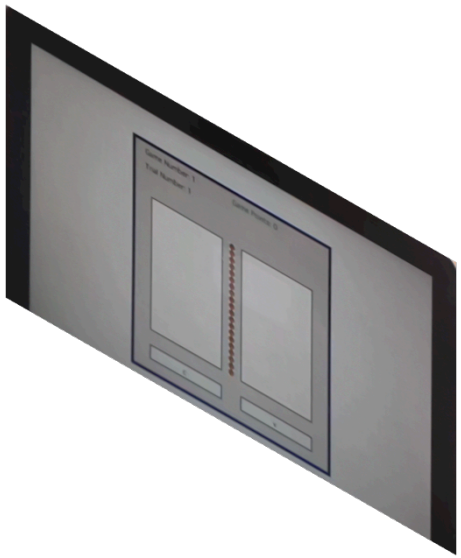
Intern Student CO3

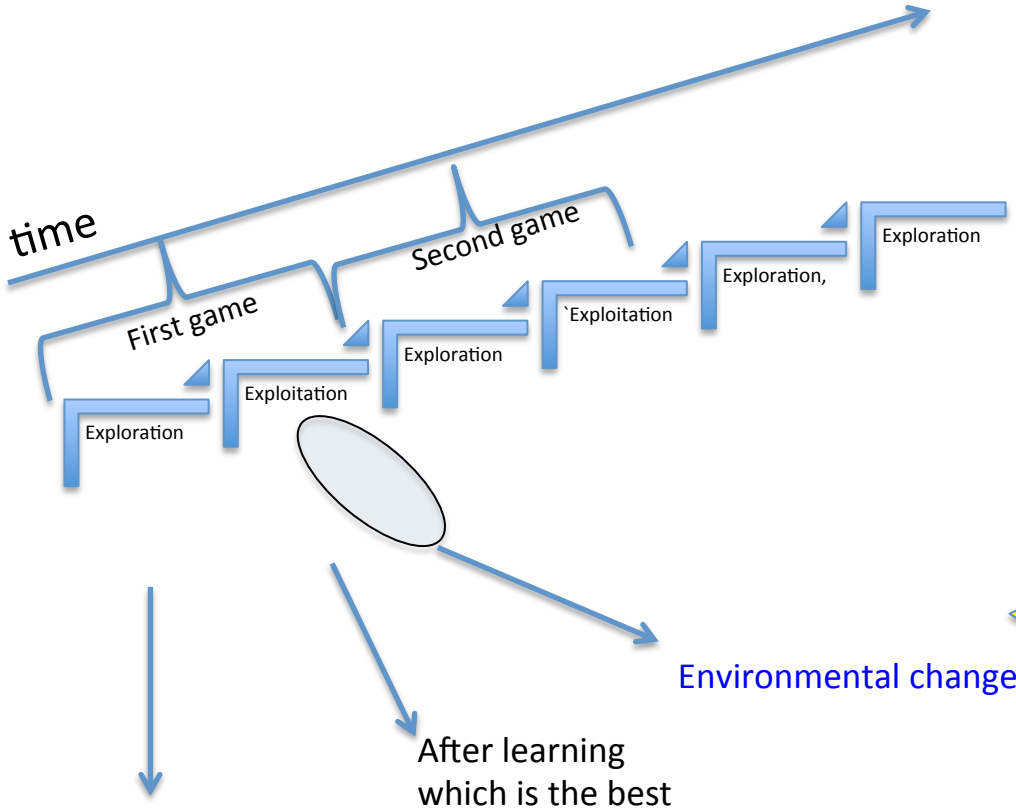
THANK YOU!

Exploratory strategies and frontal functions

- Frontal functions drive Exploration: Explorative decisions are drastically reduced using the secondary task
- We'll expect to find increase RT during explorative decisions compared to exploitative ones, *automatic and habitual actions are fast whereas cognitively controlled actions are slower* [Schneider et al. 2003] (pilot study: *Paired-t-Test* analysis revealed significant differences between exploratory and exploitative RT ($p=0.017$).
- We'll expect to find positive correlation between exploitation and eye-blinking rate, but not with exploration because of its frontal characteristics
- LOC SCALE: internal locus of control and exploration (disagreement with literature)

Bandit Task



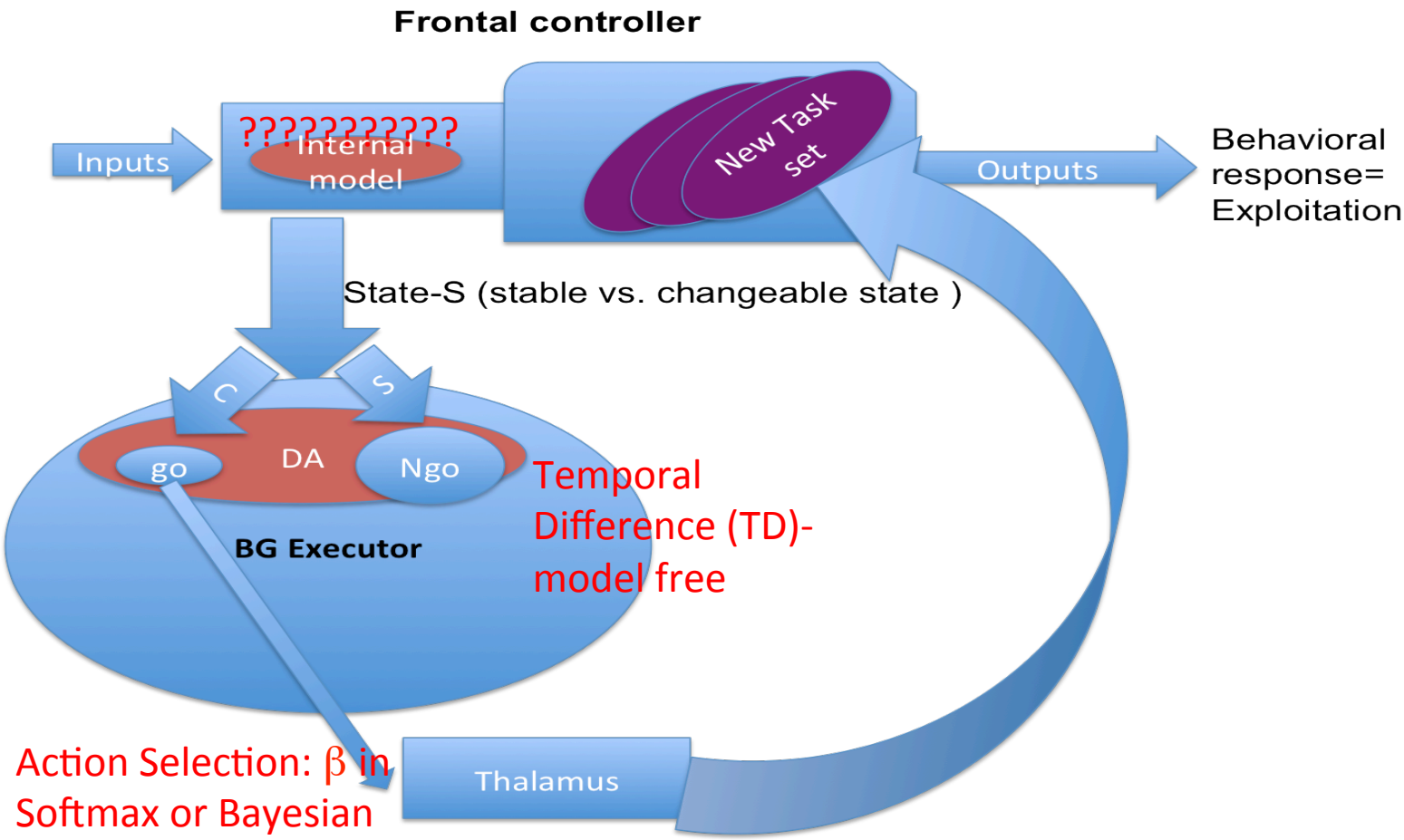


Environmental change

Getting information from the environment, looking for the better alternative.`

After learning which is the best lottery

Conflict: choose the alternative that worked best in the past or accept the risk and explore other alternatives.


Frontal controller
Inputs
???????????
Internal model
New Task set
Outputs
Behavioral response= Exploitation
State-S (stable vs. changeable state)
C
S
go
DA
Ngo
BG Executor
Temporal Difference (TD)-model free
Action Selection: β in Softmax or Bayesian
Thalamus