Spontaneous Detection of Visual Working Memory Failures and Subsequent Performance Recovery

Olga Kozlova[1],& Keisuke Fukuda[1,2]

[1]University of Toronto Mississauga, & [2]University of Toronto

Background & Tasks

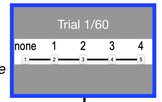
Visual Working Memory (VWM) performance fluctuates from moment to moment with periodic **failures** (Adam et al., 2015; 2017). To behave efficiently with VWM fluctuations, we need to be able to accurately predict VWM performance.

Can we predict and reflect on fluctuations in VWM performance?

After measuring individuals' VWM capacity with a set-size 6 whole report task (Adam et al., 2015):

Earn 600 points!

Bet
Exp.1 & 2: *Until Response (U.R.)*



Bet		Points
0	0+	0
1	1+	1*
2	2+	2*
3	3+	3*
4	4	4*

*0 points if overestimation

Remember
150 ms

Retention
850 ms

Colour Report
U.R.

Confidence Report
U.R.

High Conf
Low Conf
Guess

Report colour & confidence x4

 # Correct ≥ bet | ***# Correct < bet***

3
240 points

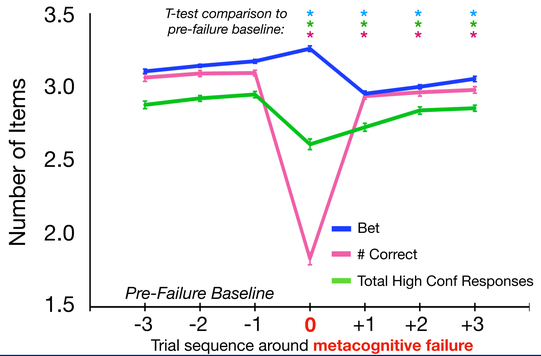
Feedback
500 ms
(Only in Exp. 1)

0
240 points

Metacognitive Failure

Exp. 1: Are we aware of fluctuations in VWM performance?

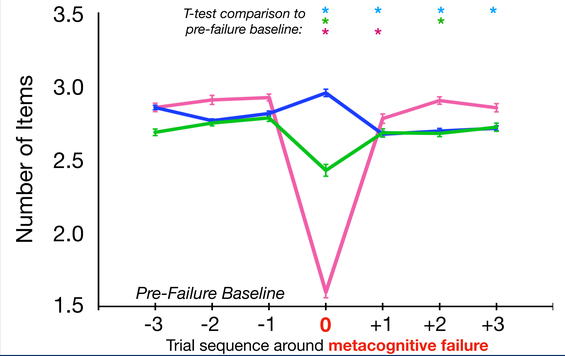
Feedback Condition (N = 56)



NO! Individuals do not anticipate or fully recognize performance drops. Post-Failure: Awareness of sustained performance reduction.

Exp. 2: Is feedback necessary for post-failure recovery?

No Feedback Condition (N = 53)



NO! Individuals spontaneously recognize post-failure reduction in performance.

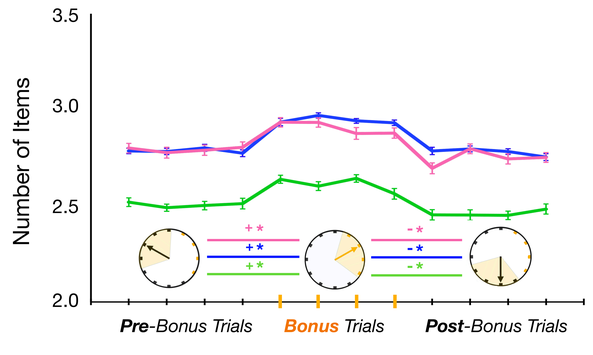
Exp. 3: Can we adapt VWM performance to high stakes?

Exp. 3 Methods

Identical to Exp. 1 except for…

- **Bet** (3 s)
- **BONUS** trials (3x the points)
- "***Try harder*** on ***bonus*** trials!"
- Clock remained throughout the trial

Incentivized Condition (N = 61)



YES! VWM performance is modulated by stakes with temporal symmetry. Predictions and reflections follow the modulation trajectory.

Takeaways

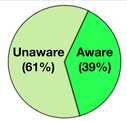
With or without feedback, individuals are **unaware** of upcoming **VWM failures**.

During VWM failures, individuals are more likely to be **unaware** of its **full extent**.

Individuals **spontaneously recognize** the **post-failure reduction** in performance.

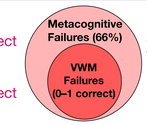
Some individuals **modulate** VWM performance according to **stake structure**. However, this modulation is **independent** of VWM capacity.

Why do individuals only partially recognize a metacognitive failure? Are there two types of failure trials?

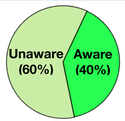


Aware:
Total HC Responses ≤ # Correct

Unaware:
Total HC Responses > # Correct

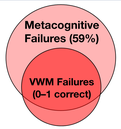


None of the severe VWM failures were accurately **predicted** in Exp. 1.



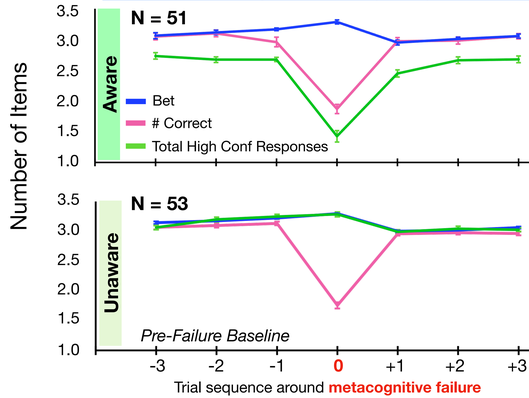
Aware:
Total HC Responses ≤ # Correct

Unaware:
Total HC Responses > # Correct

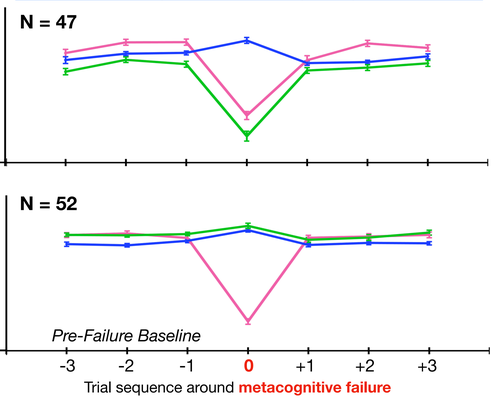


86% of severe VWM failures were not accurately **predicted** in Exp. 2.

Exp. 1 Feedback (N = 56)



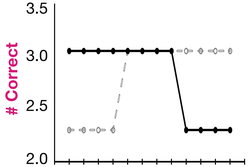
Exp. 2 No Feedback (N = 53)



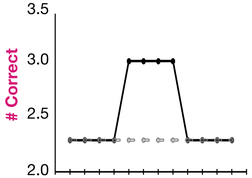
Partial awareness of failure reflects the mixture of less frequent aware and more frequent unaware trials. However, post-failure recovery did not depend on the awareness.

How do individuals vary in their modulation of VWM performance?

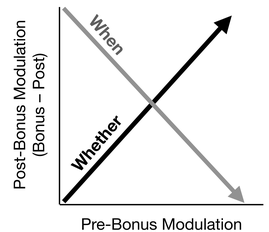
If individuals differ in ***when*** they modulate VWM performance…



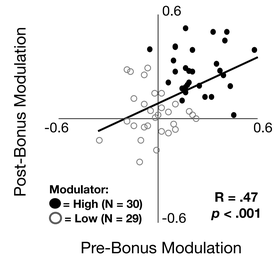
If individuals differ in ***whether*** they modulate VWM performance…



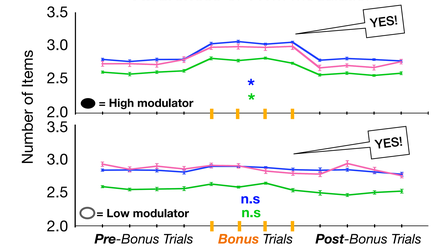
Prediction:



Result:



Awareness of VWM modulation?



Correlation with ***VWM capacity***?