

從知覺到行為：犯罪與心理的距離

「心智」典範的轉移？

沈 勝 昂

中央警察大學 犯罪防治系

Emory University, Ph.D. Clinical Psychology

Licensed Clinical Psychologist

關於行為：什麼是行為的真相？

- 表(現)象的行為
- 古典心理(結構、特質)意涵的行為
- 知覺認知情緒意涵的行為
- 知覺神經認知情緒意涵的行為

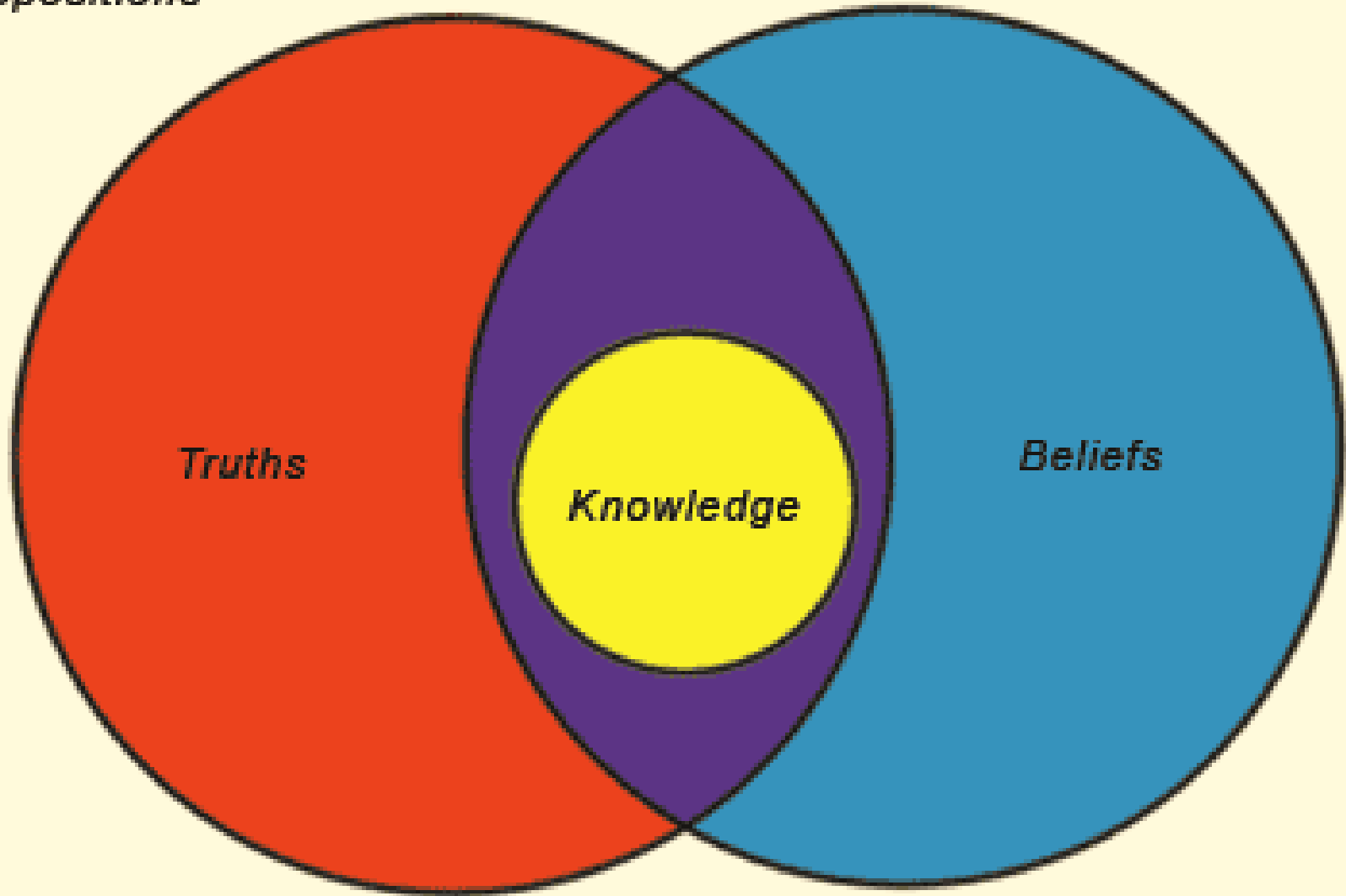


解釋模式

人類行為發生

What is fundamental true?

Propositions



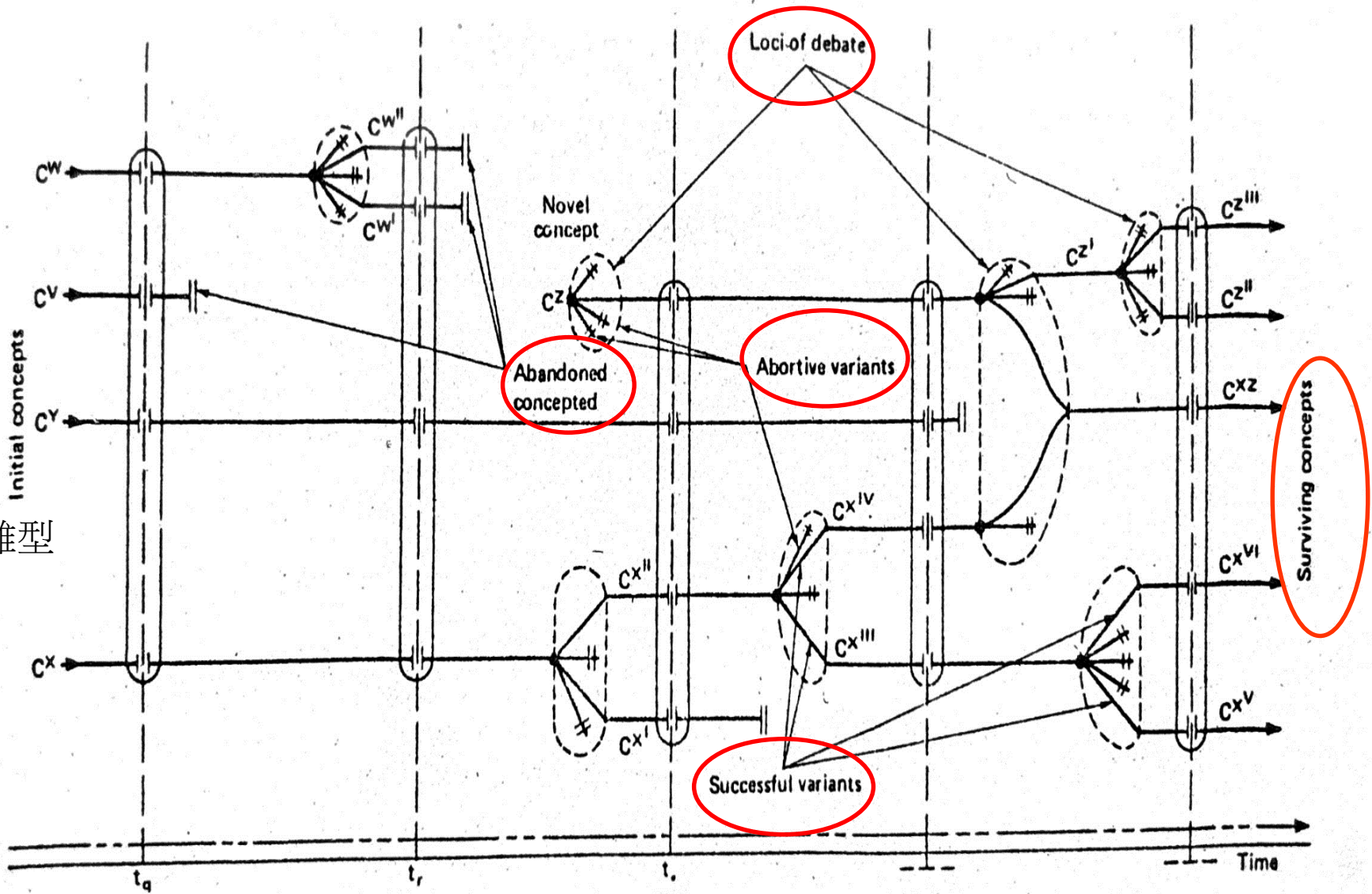
科學概念的演化

一、概念的演化

科學是一群隨著時間而演變的概念，但概念的演變如何發生？

- * 初始概念
- * 變異
- * 淘汰
- * 保留

概念雛型



科學概念的演化(Human Understanding, S. Toulman, 1972)

二、典範的革命

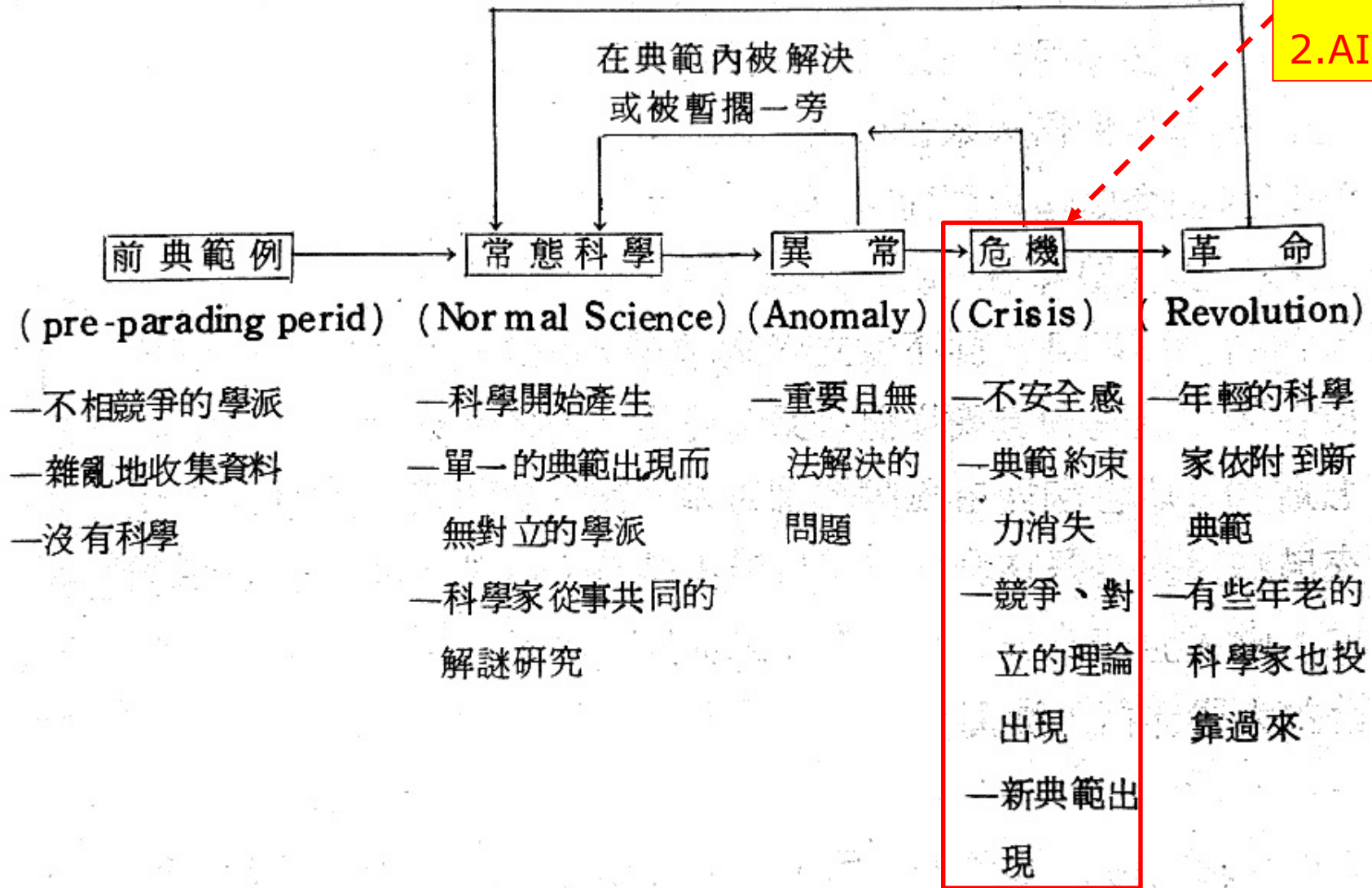
在科學上，使科學家在同一共識下攜手合作的架構稱為典範，典範包括兩個部分：訓練母體(disciplinary matrix)和共同範例(shared examples)

訓練母體：包含一組基本的假設，類似一組心照不宣的默契，他們通常不必接受考驗與證明，譬如原子論，他認為所有的物質都由原(分)子構成;心理學上則指所有的複雜行為都由簡單行為組合而成。

共同範例：科學社群內的科學家共同認可的研究方法，它提供了良好的研究模式，可被應用來解決新的問題

新典範取代舊的，而出現新的常態科學

1.神經科學
2.AI(LLM)



典範的轉移&科學循環的特性(科學革命的結構, Kuhn, 1970)

科學有這樣一個典範嗎？

- * 自然科學有這樣一個典範嗎？
- * 犯罪學有這樣一個典範嗎
- * 心理學有這樣一個典範嗎？

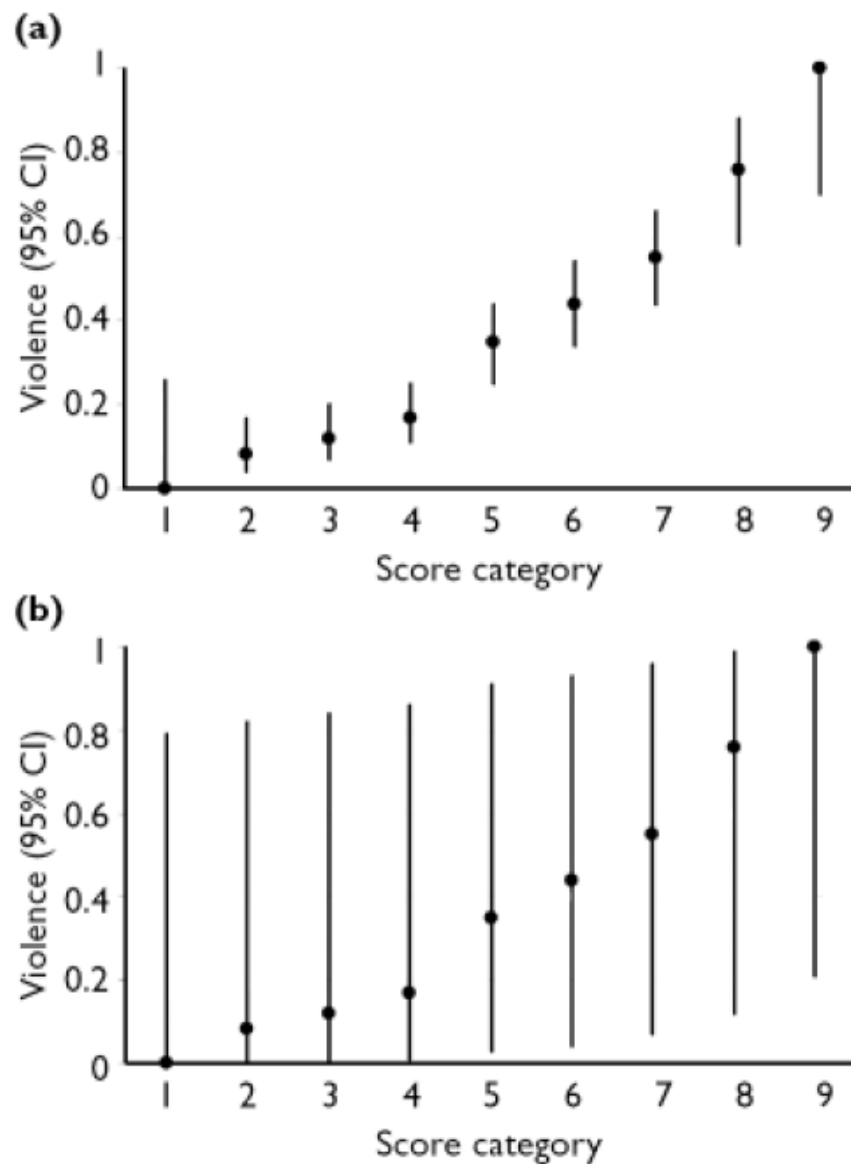


Fig. 1 Risk estimates (95% CI) for group (a) and individual (b) violence with the Violence Risk Appraisal Guide.

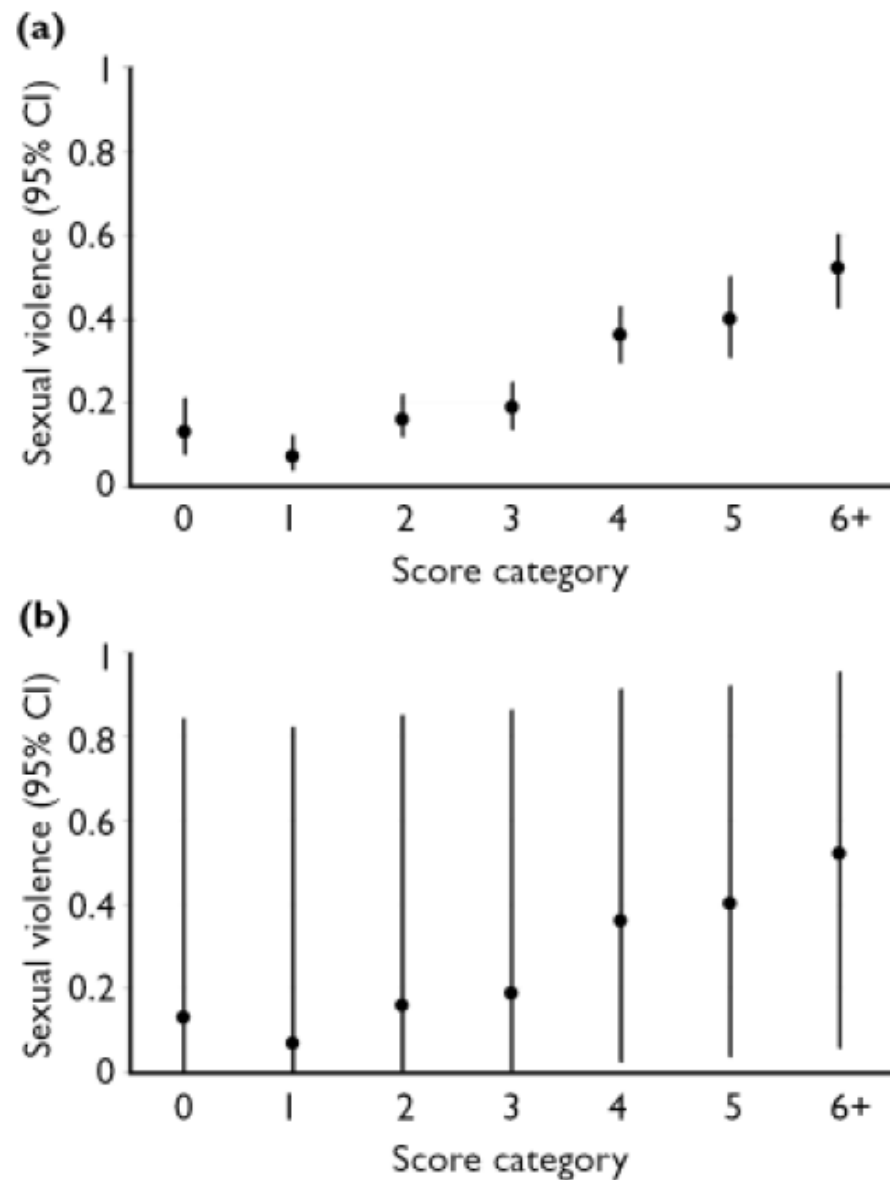


Fig. 2 Risk estimates for (95% CI) for group (a) and individual (b) violence with the Static-99).

(S. Hart, C. Michie & D.J. Cooke, 2007)

動態危險因素與再犯預測

d=.2 (小) AUC=.556
d=.5 (中) AUC=.639
d=.8 (高) AUC=.714
(Rice & Harris, 2005)(Tully & Chou, Browne, 2013)

Table 2. Psychologically Meaningful Risk Factors According to Their Strength of Evidence for Predicting Sexual Recidivism

Variable	Mean <i>d</i>	95% CI	<i>Q</i>	<i>N</i> (<i>k</i>)	Source
Supported					
Sexual preoccupation	0.39	0.23, 0.56	8.31	1,119 (6)	A
Any deviant sexual interest	0.31	0.21, 0.42	21.91	2,769 (16)	A
Sexual preference for children (PPG)	0.32	0.16, 0.47	11.52	1,278 (10)	B
Sexualized violence	0.18	0.04, 0.32	5.14	1,523 (8)	B, C
Multiple paraphilias	0.21	0.01, 0.41	6.71	477 (4)	B
Offense-supportive attitudes	0.22	0.05, 0.38	14.53*	1,617 (9)	B
Emotional congruence with children	0.42	0.16, 0.69	4.32	419 (3)	B
Lack of emotionally intimate relationships with adults					
Never married	0.32	0.21, 0.45	9.62	2,850 (8)	D
Conflicts in intimate relationships	0.36	0.05, 0.66	2.08	298 (4)	B
Lifestyle impulsivity					
General self-regulation problems	0.37	0.26, 0.48	22.85	2,411 (15)	A
Impulsivity, recklessness	0.25	0.06, 0.43	5.35	775 (6)	B
Employment instability	0.22	0.13, 0.30	20.88	5,357 (15)	A
Poor cognitive problem solving	0.22	0.05, 0.40	4.90	1,274 (4)	B, E
Resistance to rules and supervision					
Childhood behavior problems	0.30	0.16, 0.43	7.11	1,996 (8)	B
Noncompliance with supervision	0.62	0.45, 0.79	5.86	2,159 (3)	B
Violation of conditional release	0.50	0.34, 0.65	16.55**	2,151 (4)	B
Grievance/hostility	0.20	0.09, 0.31	13.58	3,139 (11)	B, C, E
Negative social influences	0.26	0.08, 0.44	2.66	1,736 (7)	B, E
Promising					
Hostility toward women	0.29	0.00, 0.58		799 (1)	E
Machiavellianism	1.40	0.48, 2.33		99 (1)	F
Callousness/lack of concern for others	0.29	0.11, 0.47	0.001	1,173 (2)	C, E
Dysfunctional coping					
Sexualized coping	0.43	0.14, 0.74		798 (1)	E
Externalizing	0.27	0.03, 0.51		380 (1)	C

Note: A: Hanson and Morton-Bourgon (2005); B: Hanson and Morton-Bourgon (2004); C: Knight and Thornton (2007); D: Hanson and Bussière (1998) transformed from *r* to *d* assuming 13.4% base rate; E: Hanson et al. (2007); F: Thornton (2003).

p* < .05 *p* < .01.

(Mann, Hanson, Thornton, 2010)

TABLE 1: Pre- and Posttest Scores on Dynamic Risk Factor Scales and Effect Sizes From Paired Sample *t* Tests

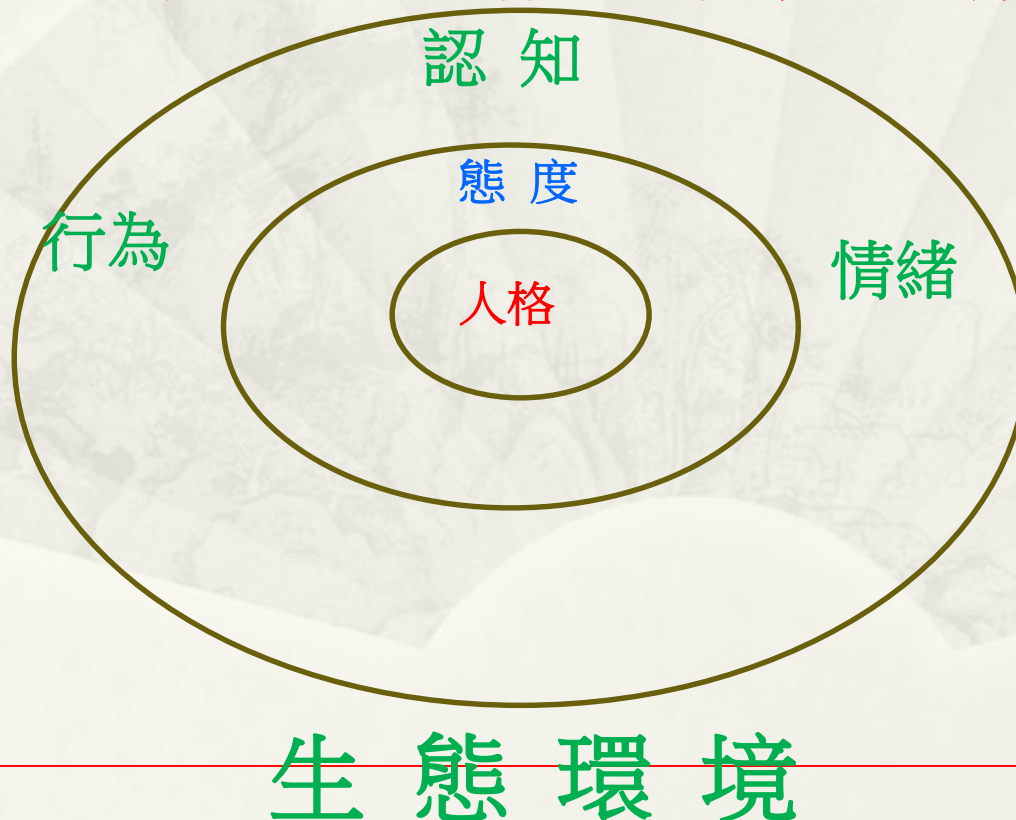
d=.2 (小) AUC=.556
d=.5 (中) AUC=.639
d=.8 (高) AUC=.714

Dynamic Risk Factor		Time	M	SD	d	M	SD	d	M	SD	d
Procriminal attitudes											
Negative attitudes toward the law	Pre		4.09	3.35	0.28	4.48	3.36	0.67**	3.55	2.77	0.23
	Post		3.19	2.94		2.44	2.74		2.92	2.77	
Offense-supportive attitudes	Pre		2.26	1.65	0.29	2.20	1.68	0.49**	1.87	1.33	0.42**
	Post		1.80	1.53		1.42	1.48		1.31	1.35	
Antisocial personality patterns											
Low self-control	Pre		56.17	10.43	0.45**	53.69	11.16	0.01	52.11	10.60	0.01
	Post		51.46	10.51		53.53	11.30		52.01	9.77	
Aggressiveness	Pre		6.49	3.37	0.32*	5.27	3.38	0.17	3.96	2.65	-0.12
	Post		5.45	3.12		4.70	3.50		4.32	3.19	
Excitability	Pre		6.09	3.33	0.28*	4.32	3.19	0.19	4.75	3.08	0.00
	Post		5.17	3.23		3.72	3.04		4.75	2.97	
Conscientiousness	Pre		32.69	5.22	-0.34*	33.43	7.56	-0.11	33.17	5.25	0.05
	Post		34.47	5.29		34.23	6.49		32.87	5.96	
Agreeableness	Pre		26.07	5.06	-0.31*	26.52	6.58	-0.13	29.10	5.81	-0.08
	Post		27.67	5.40		27.33	6.13		29.55	5.29	
Anxiety/neuroticism											
Emotional instability	Pre		7.46	3.68	0.54**	6.03	3.89	0.22	6.73	3.60	0.29*
	Post		5.47	3.72		5.16	3.90		5.63	3.97	
Neuroticism	Pre		22.17	7.36	0.61**	21.12	8.92	0.16	22.85	7.53	0.20
	Post		17.97	6.36		19.76	7.95		21.32	7.35	
State anxiety	Pre		38.74	8.10	0.31	39.34	9.53	0.59**	40.19	10.14	0.45**
	Post		36.05	9.12		34.14	8.11		35.92	8.99	
Trait anxiety	Pre		43.02	10.45	0.45**	43.83	10.63	0.62**	44.35	8.94	0.35**
	Post		38.81	8.26		37.75	8.99		41.32	8.45	
Empathy											
Sensitivity	Pre		36.40	8.70	-0.15	33.07	10.50	-0.17	37.20	10.29	0.00
	Post		37.69	8.62		34.82	10.43		37.21	9.59	
Concern	Pre		35.83	7.24	-0.15	36.29	8.52	0.10	40.17	8.54	0.11
	Post		36.86	6.54		35.48	7.67		39.29	7.24	

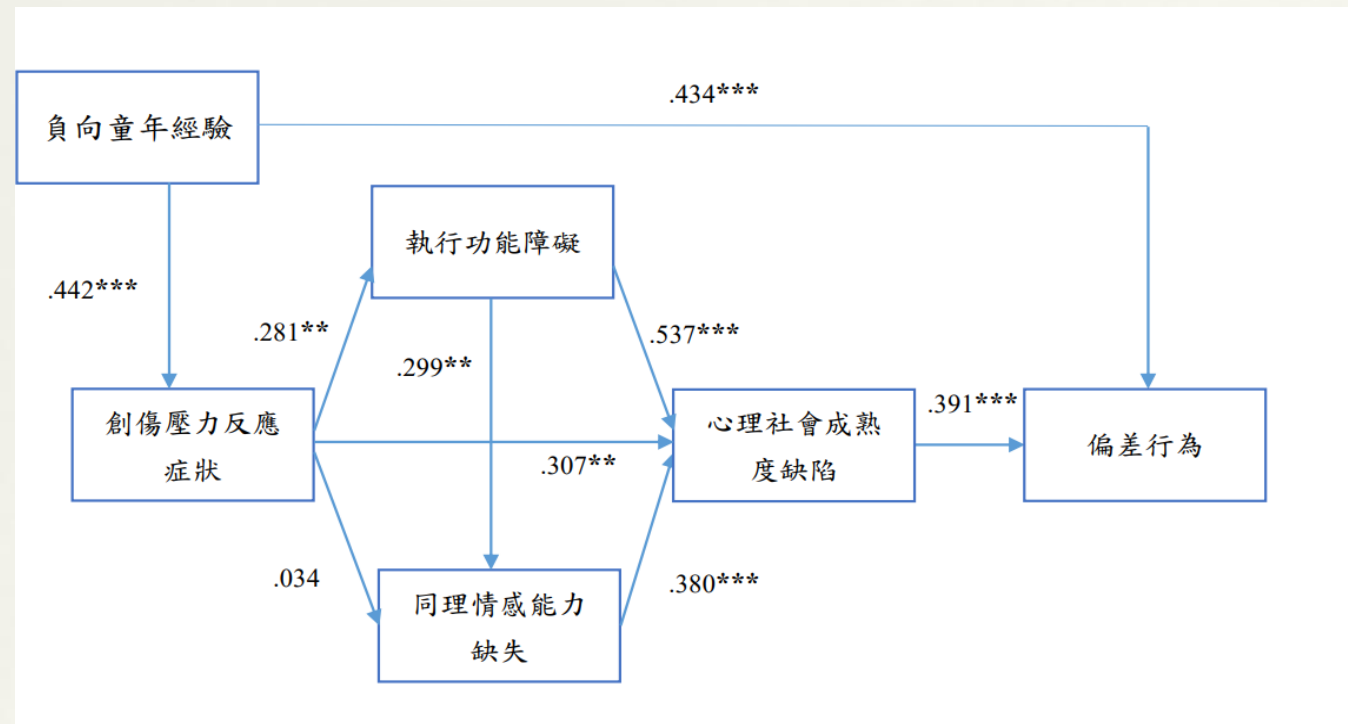
p* < .05. *p* < .01.

「人」：科學心理學的解釋模式

- * 由心理學「人」概念：學習與發展
- * 學習心理學：個人與環境間互動的改變
- * 發展心理學：個人隨時間的所造成的秩序性改變



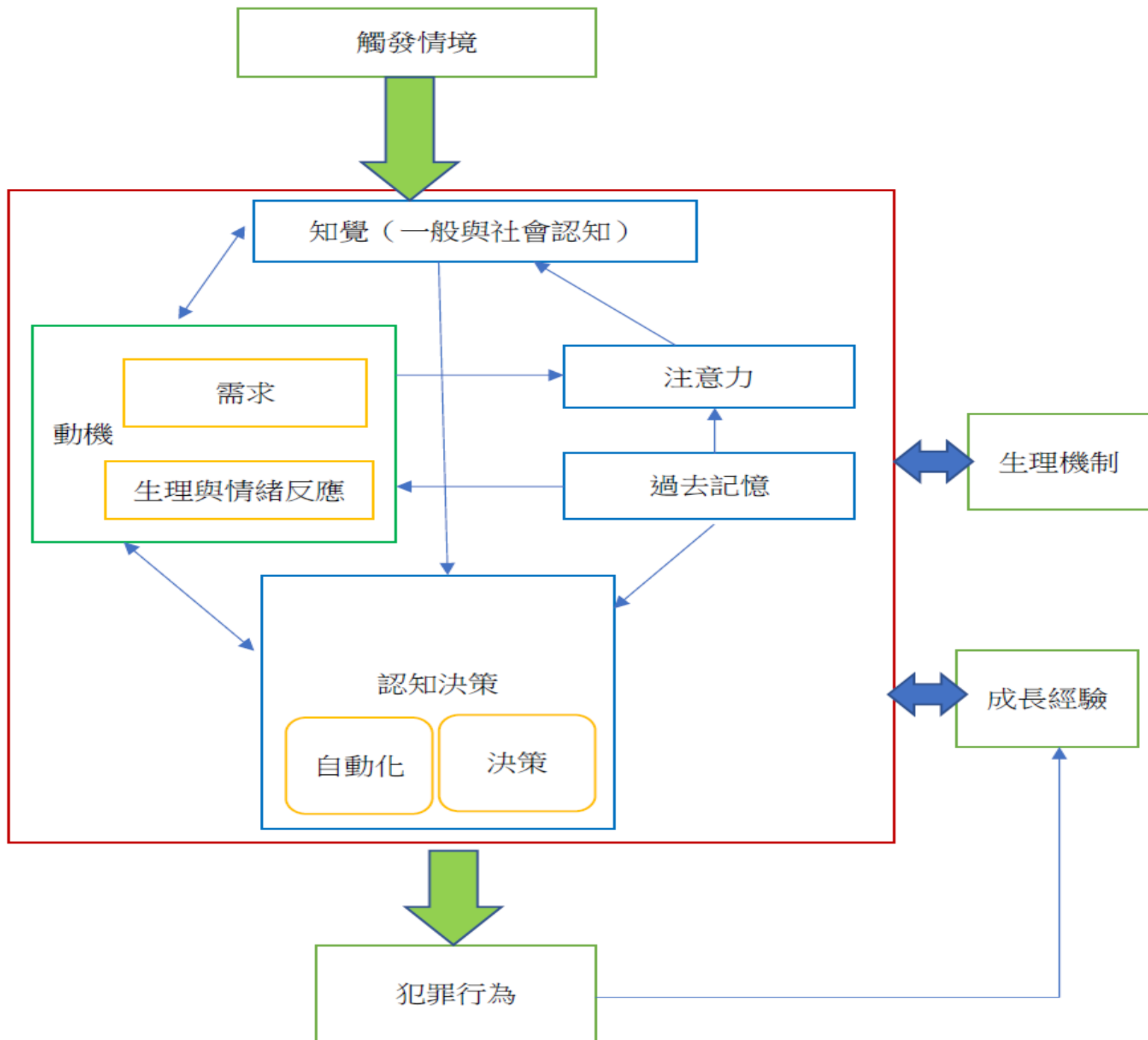
路徑分析之結果 -負向童年經驗、創傷壓力反應症狀、風險因子與偏差行為之路徑分析

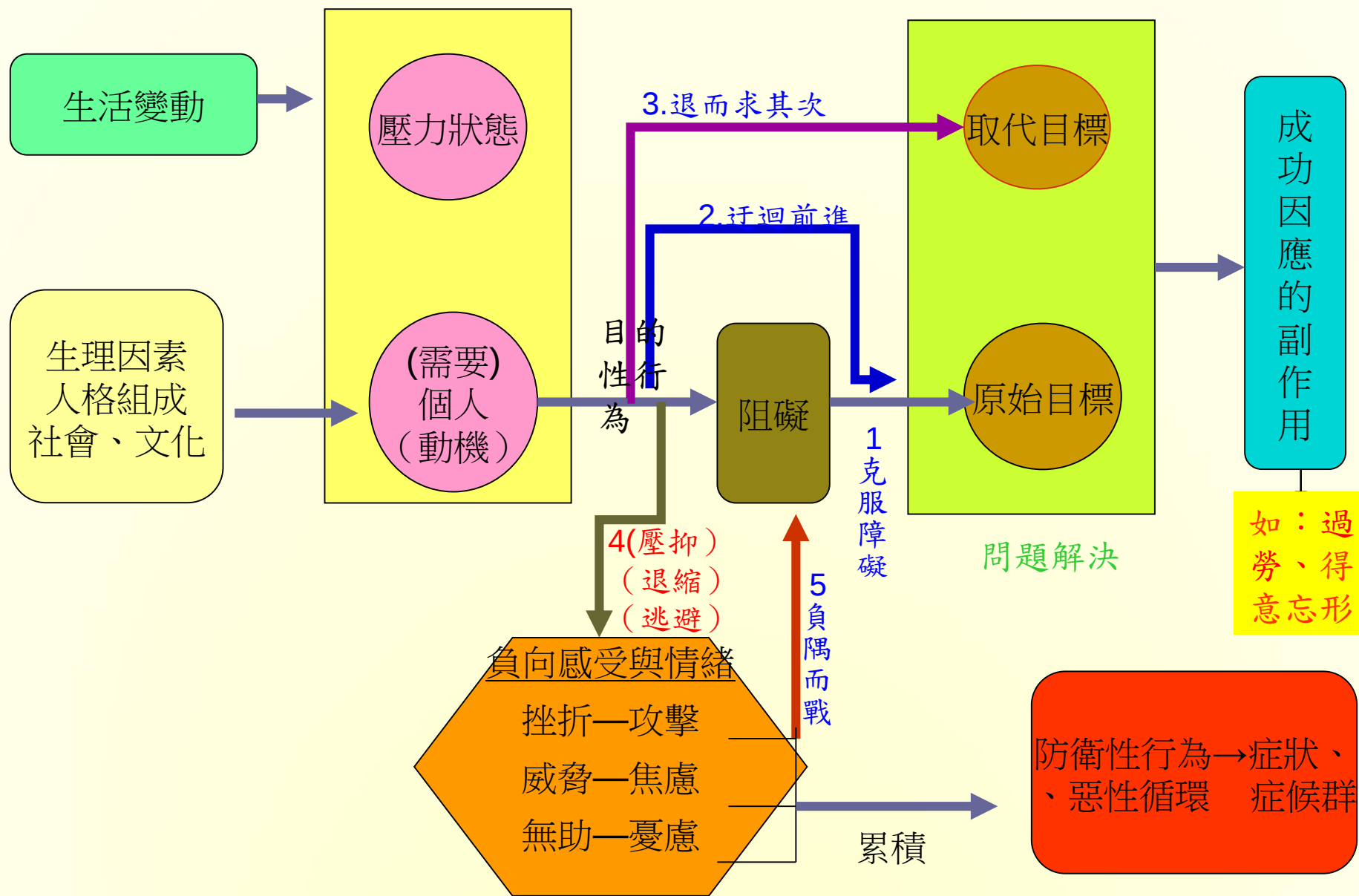


	標準化係數值 (β)	Z ; Sig
負向童年經驗→創傷壓力反應症狀	.442	5.212***
創傷壓力反應症狀→執行功能障礙	.281	3.099**
創傷壓力反應症狀→同理情感能力缺失	.034	.357
創傷壓力反應症狀→心理社會成熟度缺陷	.307	3.412**
執行功能障礙→同理情感能力缺失	.299	3.314**
執行功能障礙→心理社會成熟度缺陷	.537	6.738***
同理情感能力缺失→心理社會成熟度缺陷	.380	4.343***
心理社會成熟度缺陷→偏差行為	.391	4.496***
負向童年經驗→偏差行為	.434	5.105***

註: *p<.05, **p<.01, ***p<.001

人格

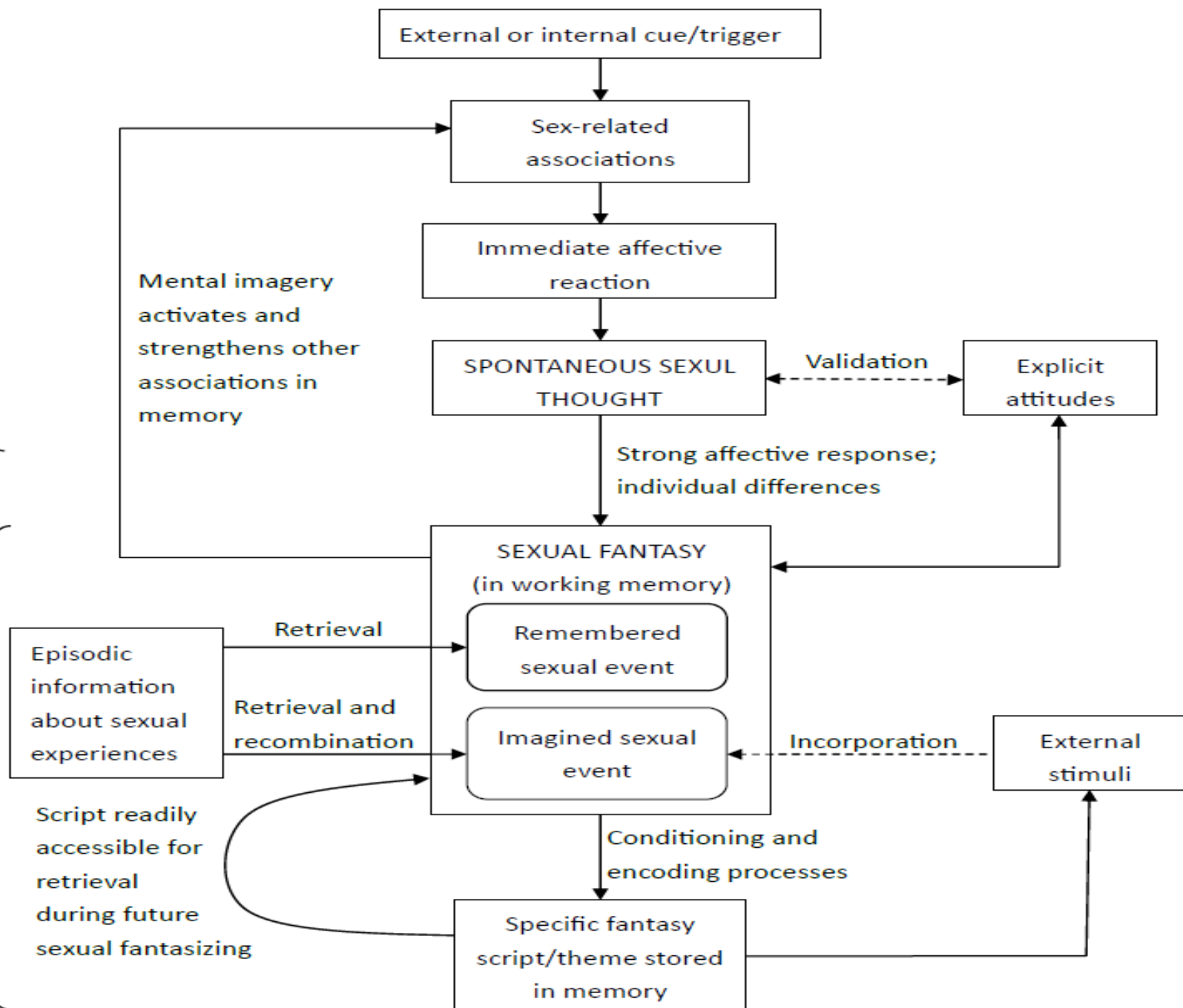




圖一 適應不良行為發生的「心理病理」壓力模式

Associative component

Controlled (elaborative) component



人類心理學研究 (Human Psychology)

- 過度偏重主觀經驗
- 依賴抽象認知模型

學術割裂導致 PTSD 與焦慮症治療效能陷入瓶頸，
無法解釋「情境評估」與
「非直接經驗的恐懼轉移」。



傳統動物模型 (Animal Models)

- 侷限於先天防禦反應
- 單純的「線索-結果」學習

跨物種情緒內在模型研究, 2026

內部模型的核心引擎：mPFC 的推論邏輯

隱喻：惡犬攻擊事件後的隱藏連結 ^1



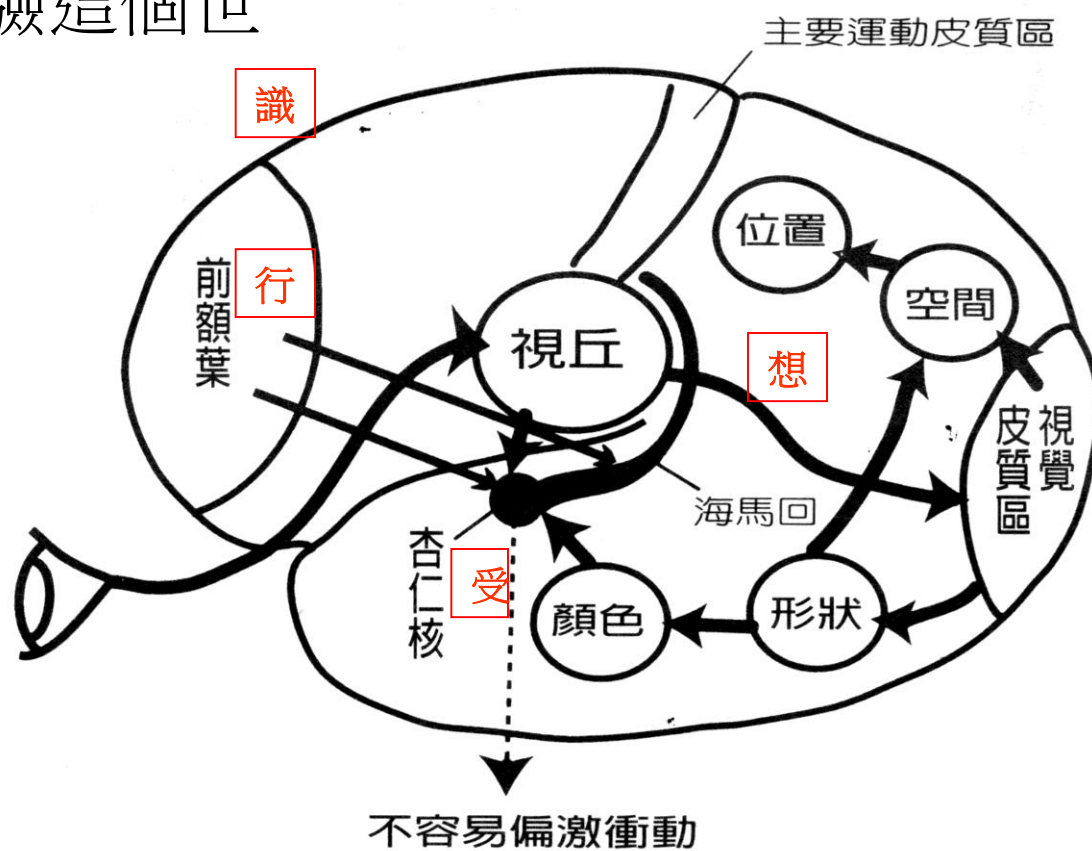
1 Science 393, The hidden logic of emotion

2 探索情緒的隱藏邏輯：從身體訊號到大腦推論模型的轉化過程，Ch. 4

從巨視的「人」到微視的「心智人」

- * 統計模式到造人
- * 臨床模式的造人
- * 心智模式的造人
- * 互動模式的「心智人」

心如何經驗這個世界



受到來自前額葉的抑制信號影響，杏仁核的突發性行為因此時被抑制。(沈勝昂，2018)

感覺與認知

感覺

視、聽、觸、嗅、味覺
等感官訊息的處理



知覺

組織與建構感官訊
息



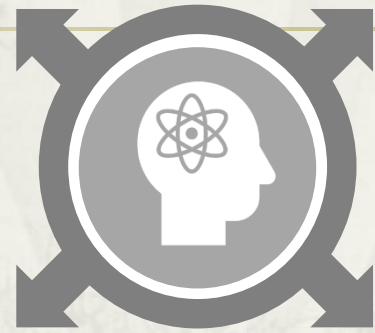
認知

思考、語言、學習、
記憶、注意力、決策
、理解、問題解決、
智慧



意識

第一人稱的主觀經
驗

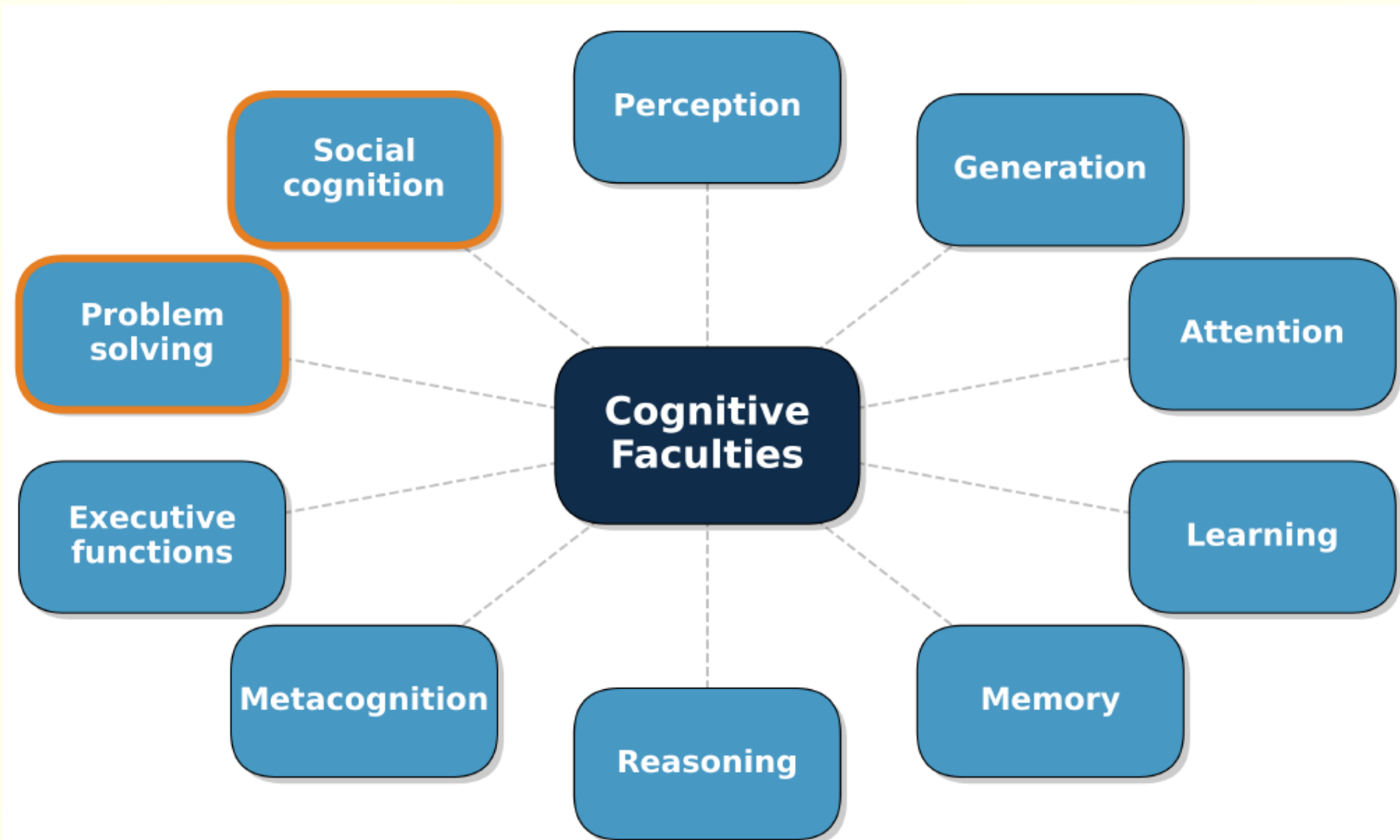


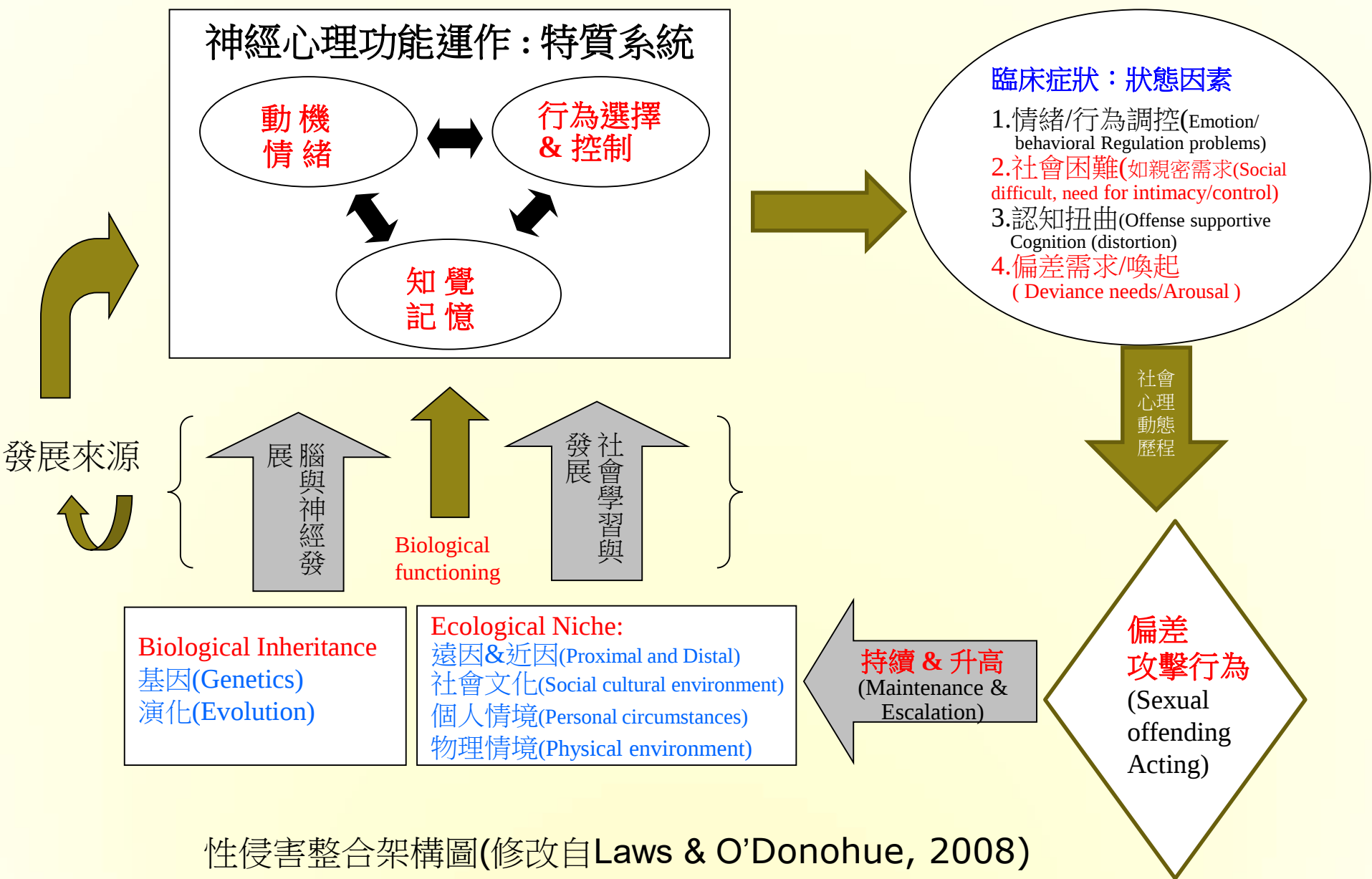


參考文獻：Monica Gori et al., 2025



人類可以用語言描述的10個認知能力

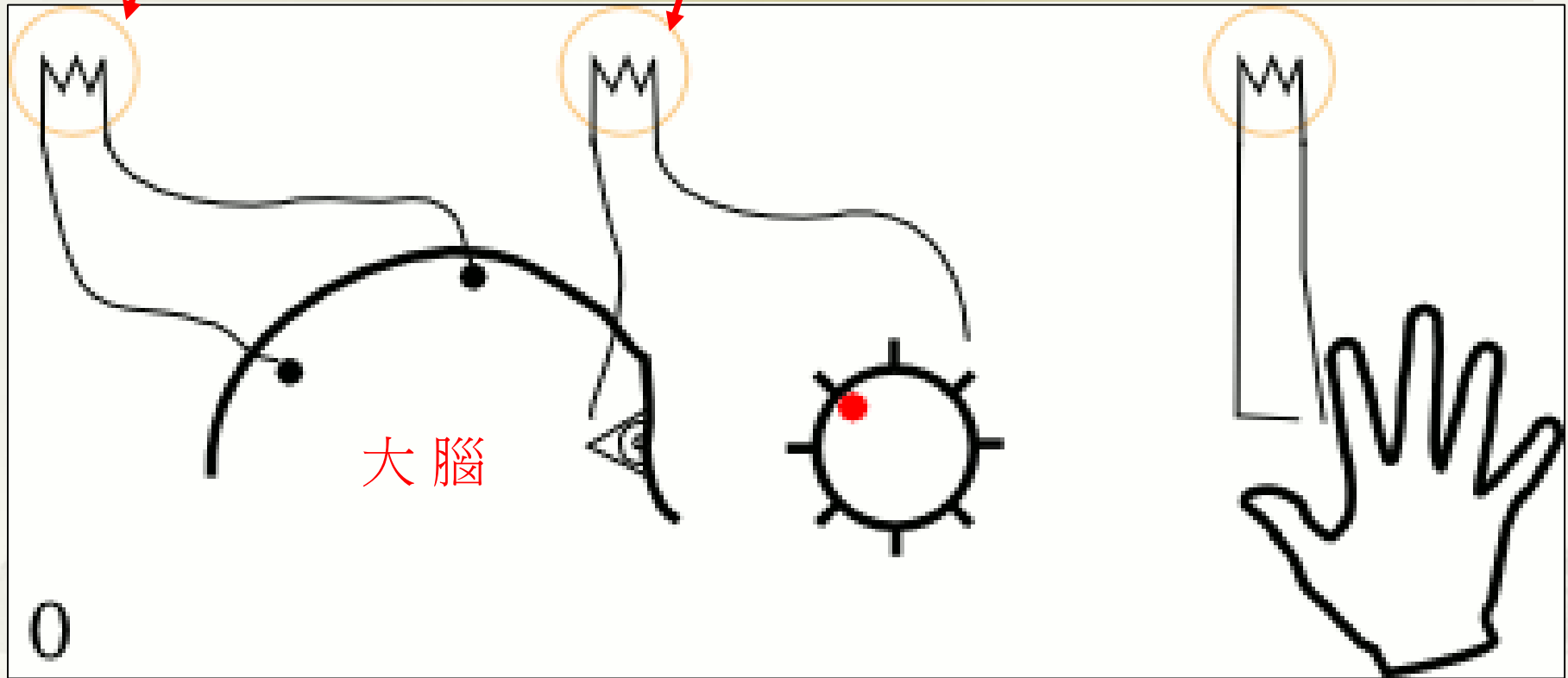




$X..Ø..£..ø....\partial$

我要按A鍵

大腦激發(i.e. $X..Ø..\partial...$) $\longrightarrow$ 意念報告(i.e. 我想按A鍵) $\longrightarrow$ 行為發生(i.e. 手按A鍵)



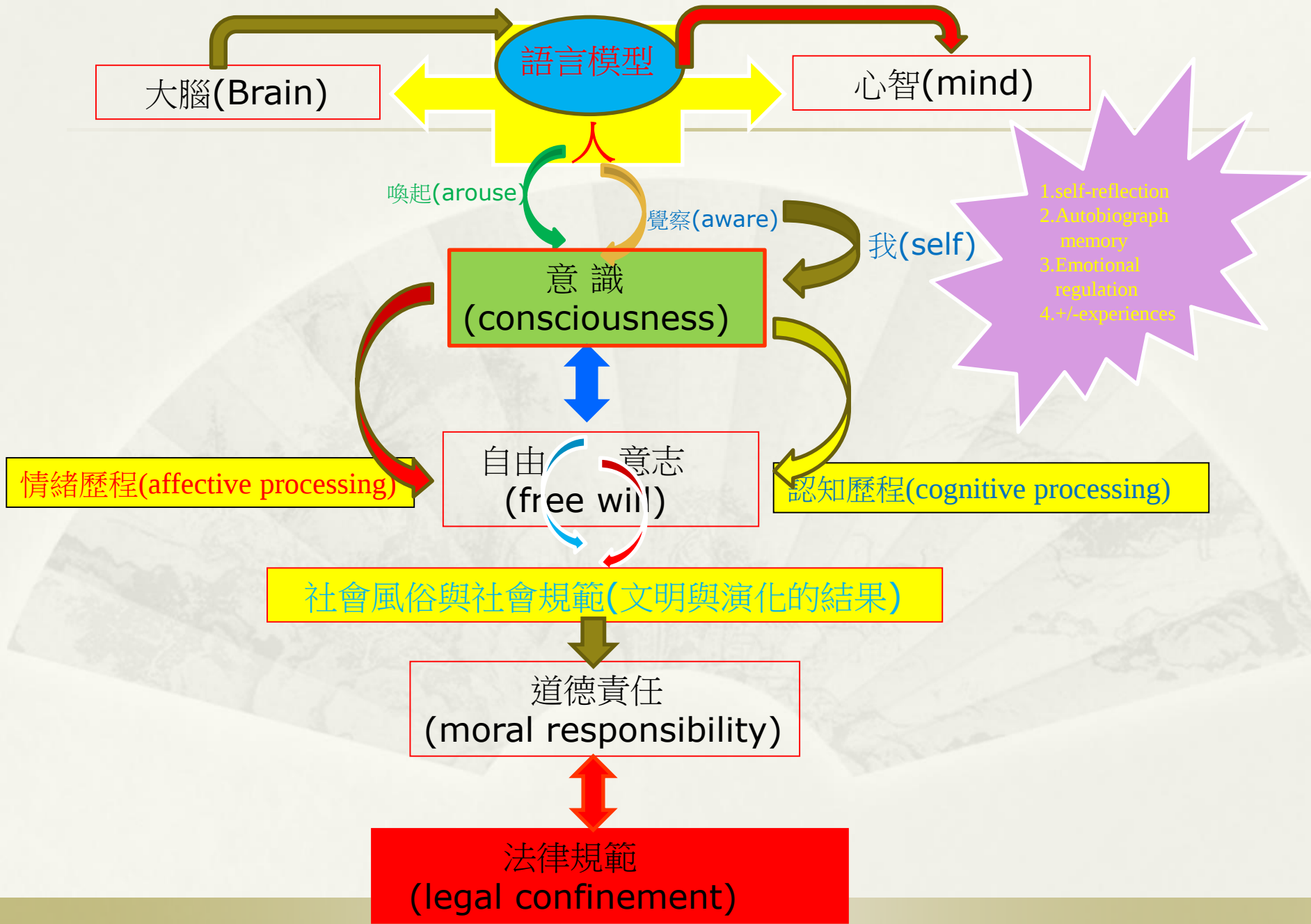
0 repose

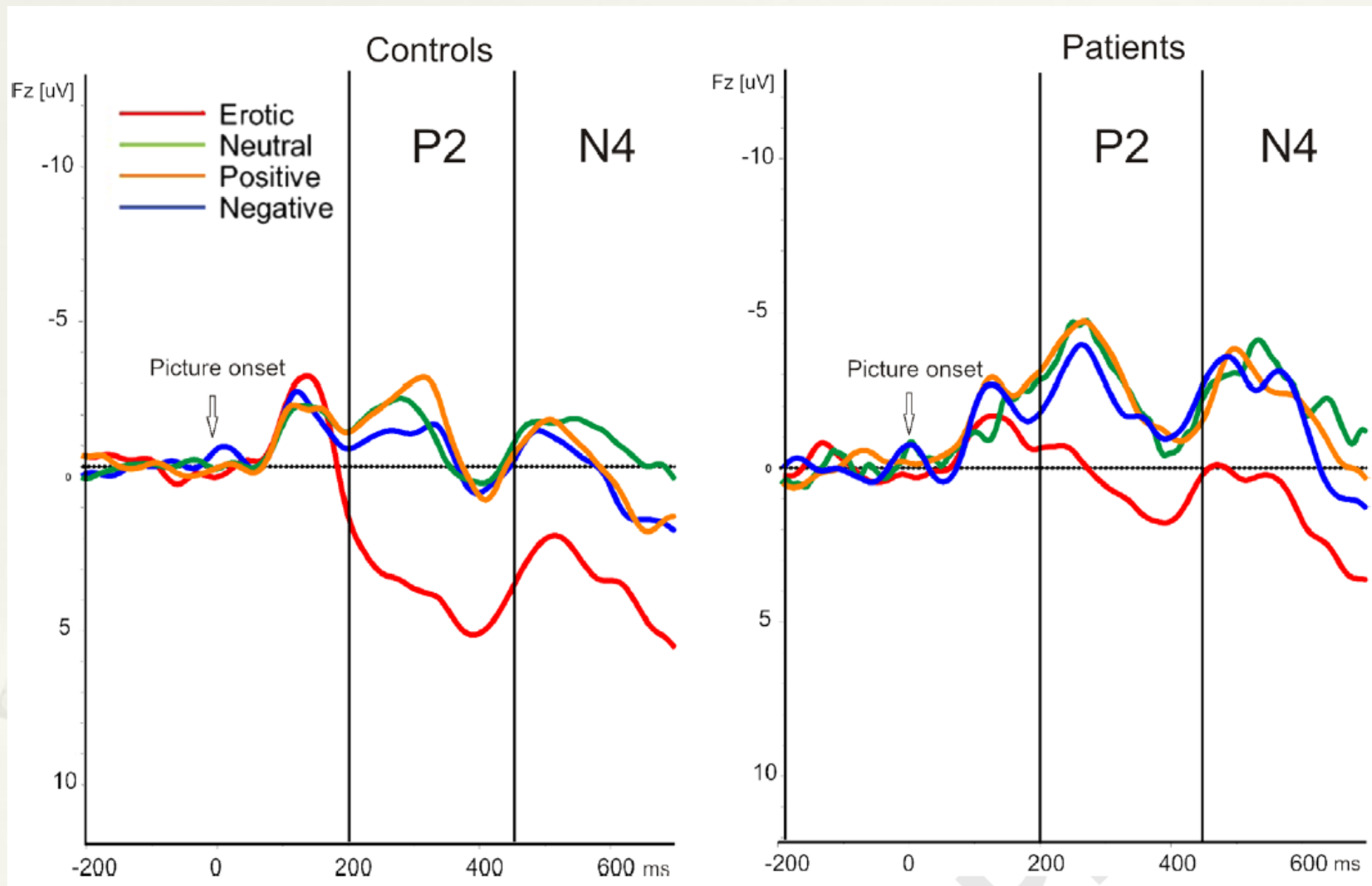
1 (-500 ms) EEG measures Readiness potential

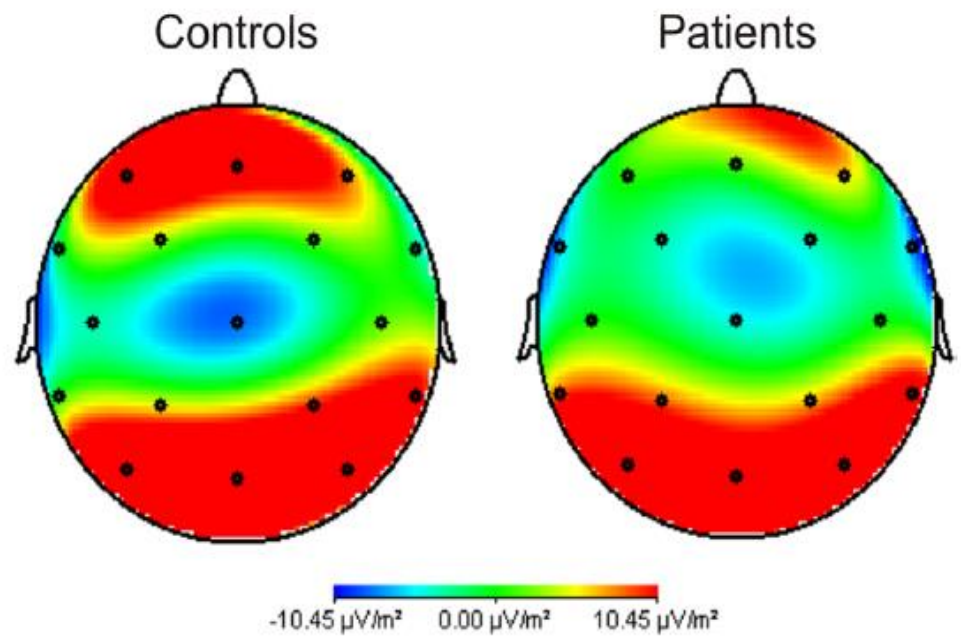
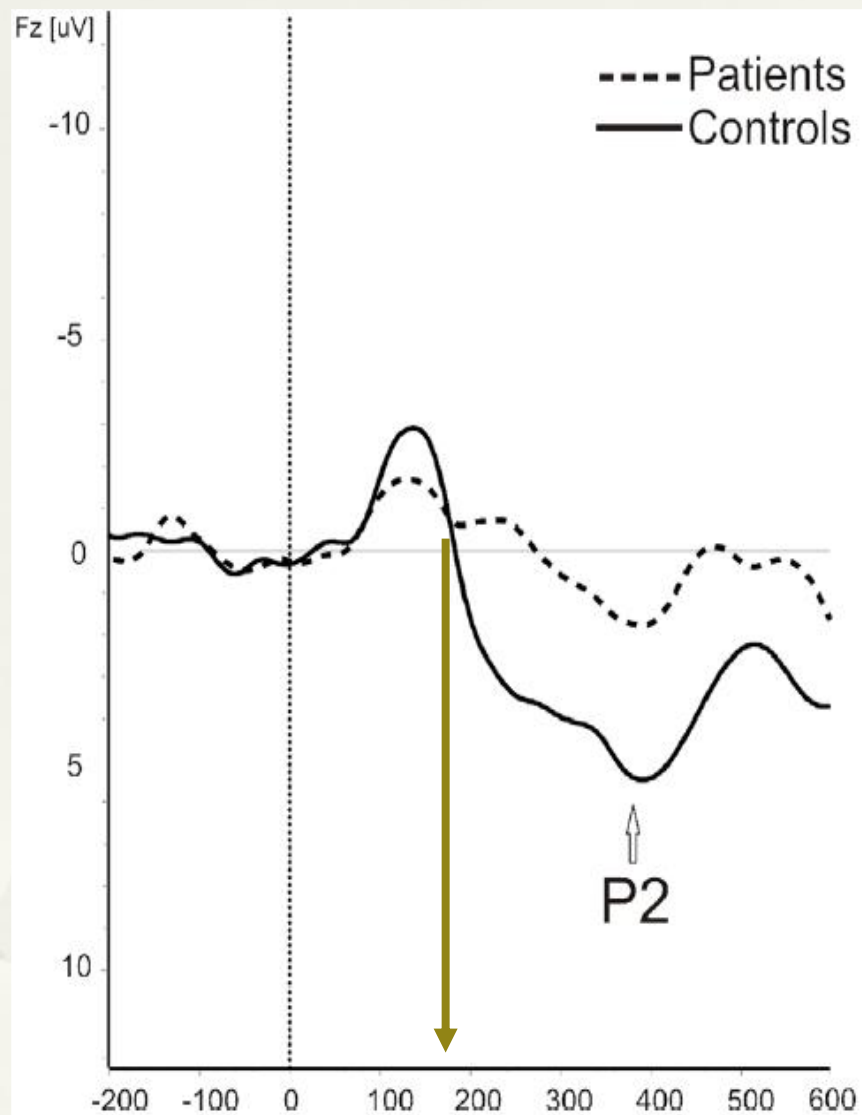
2 (-200 ms) Person notes the position of the dot when decides

3 (0 ms) Act

神經與心智





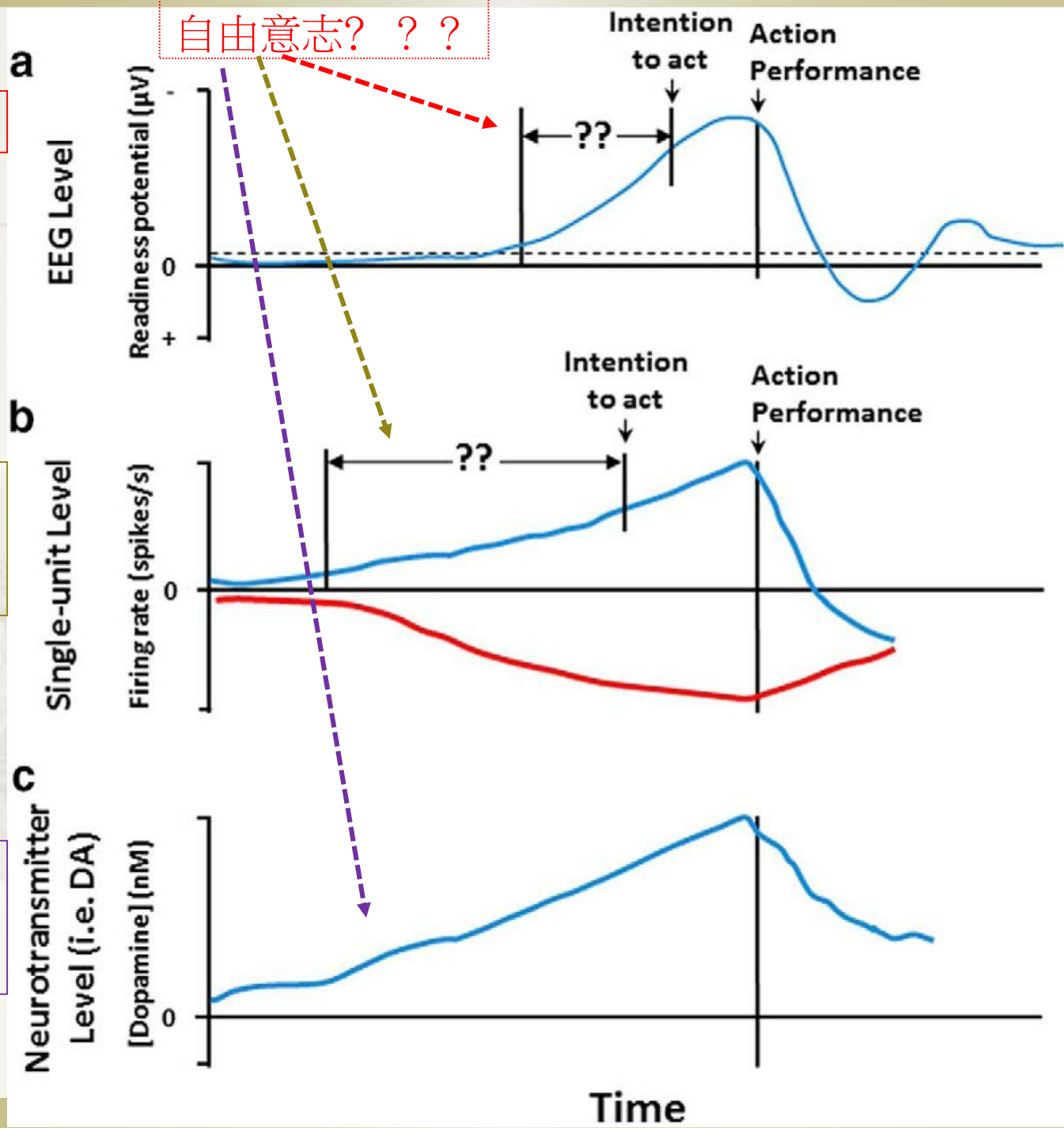


自由意志? ? ?

Scalp EEG

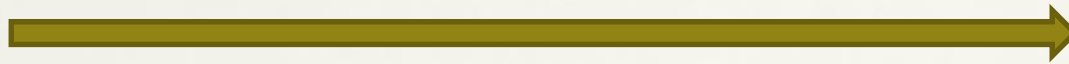
MFC
Dorsomedial
striatum

Midbrain
dopaminergic
neuron(DA)



IHS (蓄意傷害)

惡意動機 + 負面結果



ion

AHS (意外傷害)

中性動機 + 負面結果



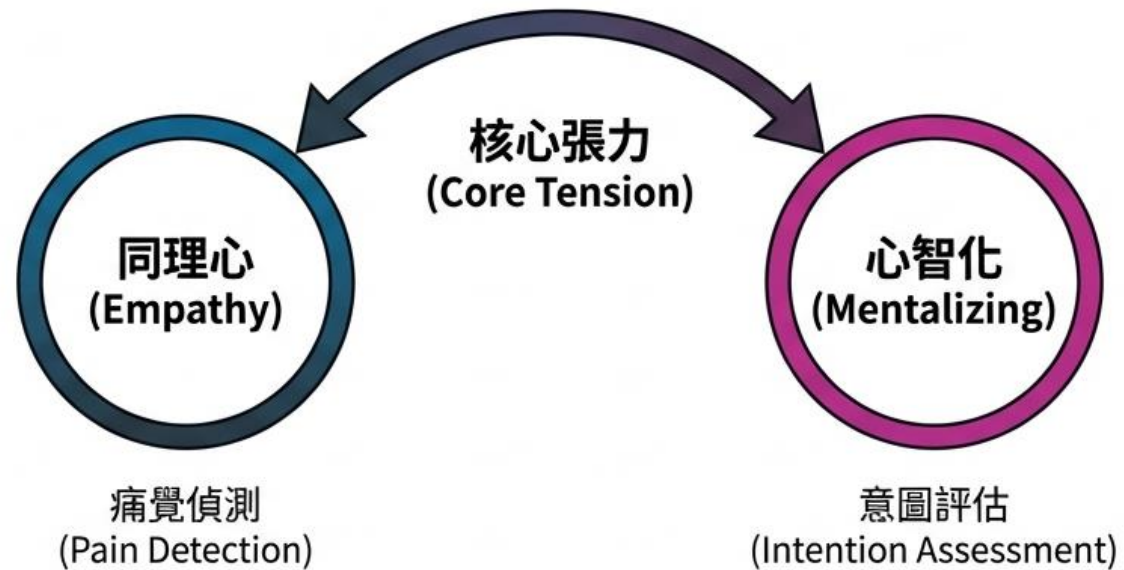
ion

NAS (中性行為)

中性動機 + 中性結果

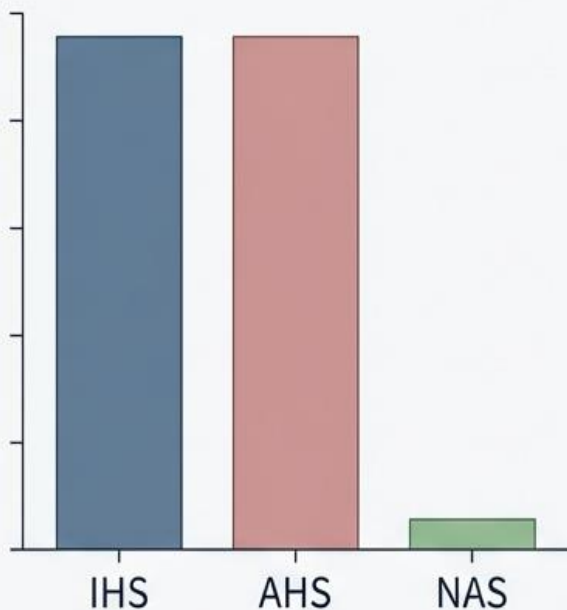


- **雙重歷程模型 (Two-Component Model)**：道德決策涉及兩個關鍵要素的整合。
- **同理心**：對受害者痛苦的快速、自動化情感共鳴。
- **心智化**：對加害者意圖的認知評估（蓄意 vs. 意外）。
- **核心問題**：大腦如何在毫秒級的時間尺度內，從「看到痛苦」轉變為「判斷責任」？



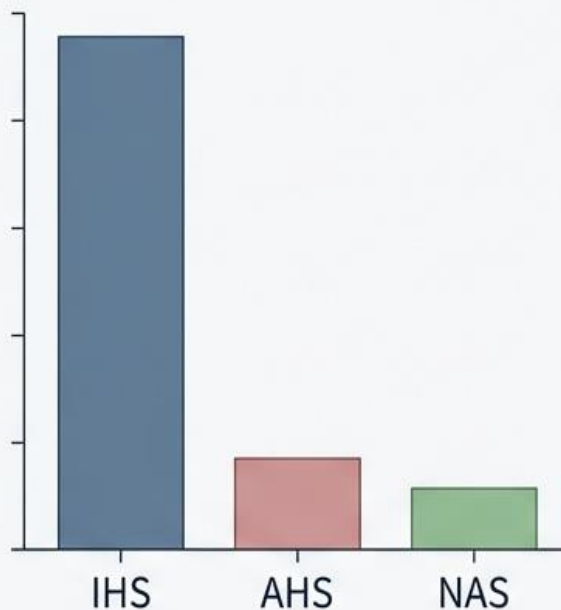
意識層面的評估：行為數據驗證

物理疼痛感



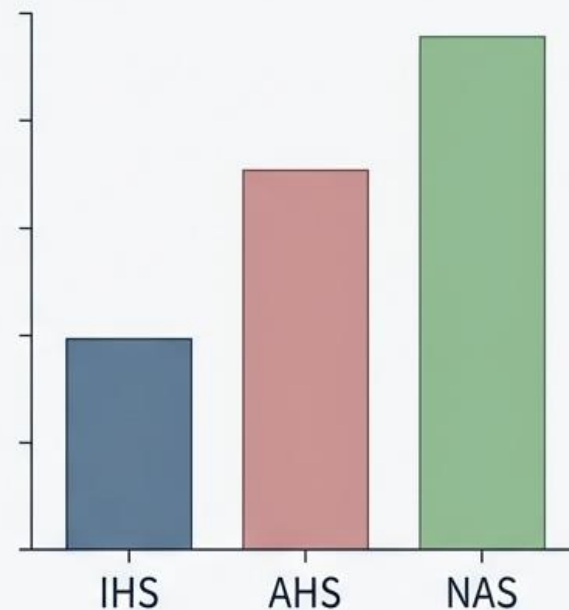
蓄意 = 意外

惡意意圖



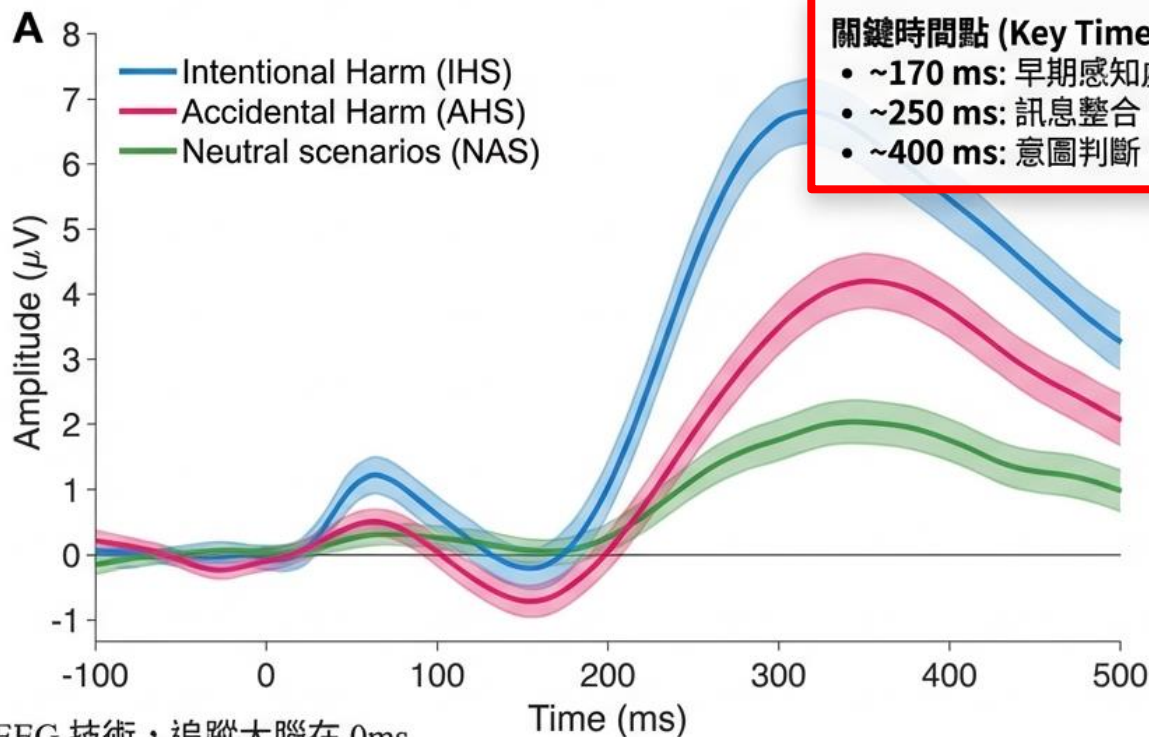
蓄意 >>> 意外

道德許可性



蓄意最不可接受

神經動態總覽：事件相關電位 (ERPs)



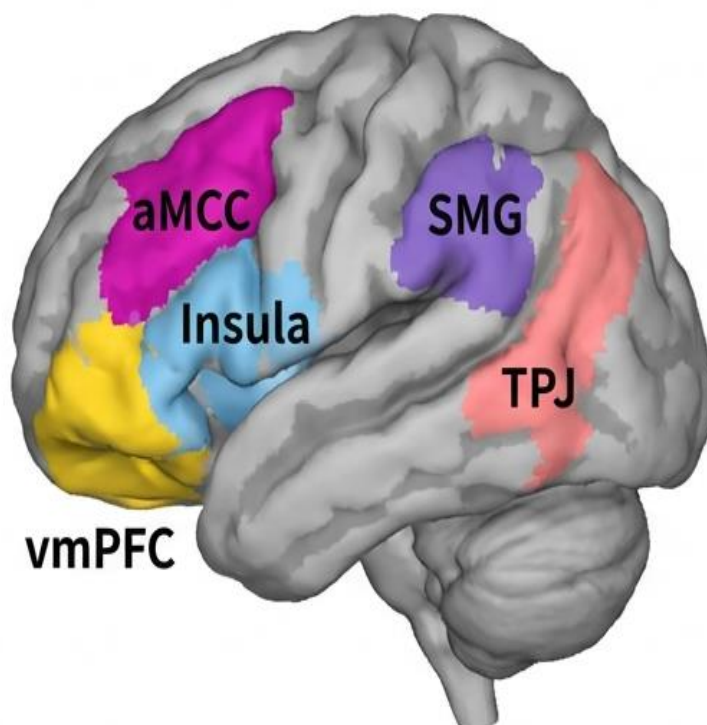
利用高時間解析度的 EEG 技術，追蹤大腦在 0ms 至 500ms 間的反應。波形在不同時間點的分離代表了大腦處理資訊的不同階段。

參考文獻：Kamela Cenka et al., 2024

腦區定位與功能總結 (Regions of Interest)

同理心網絡 (Empathy)

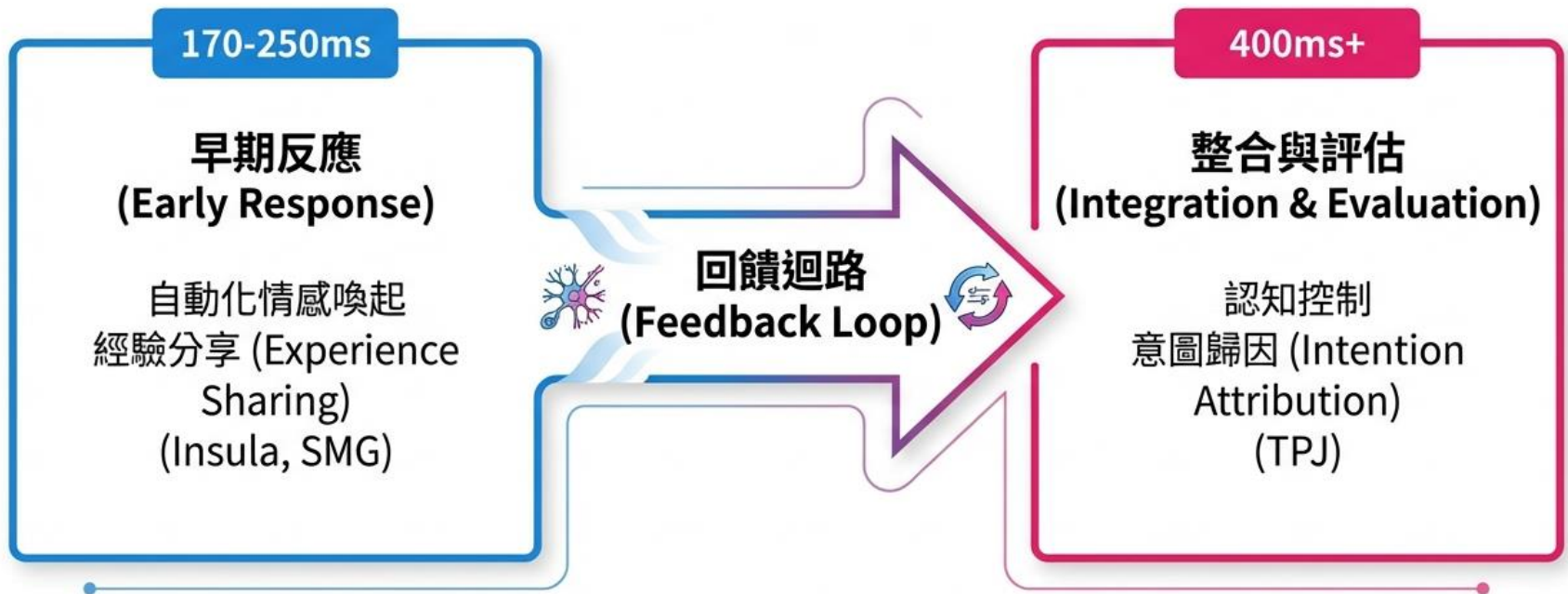
- Insula (島葉)：情感與痛覺。
- aMCC (前中扣帶皮層)：痛覺共鳴。
- SMG (緣上回)：自我/他人視角轉換。



心智化網絡 (Mentalizing)

- TPJ (顳頂交界區)：推論他人意圖。
- vmPFC (腹內側前額葉)：社會價值決策。

理論模型：從自動化反應到認知控制

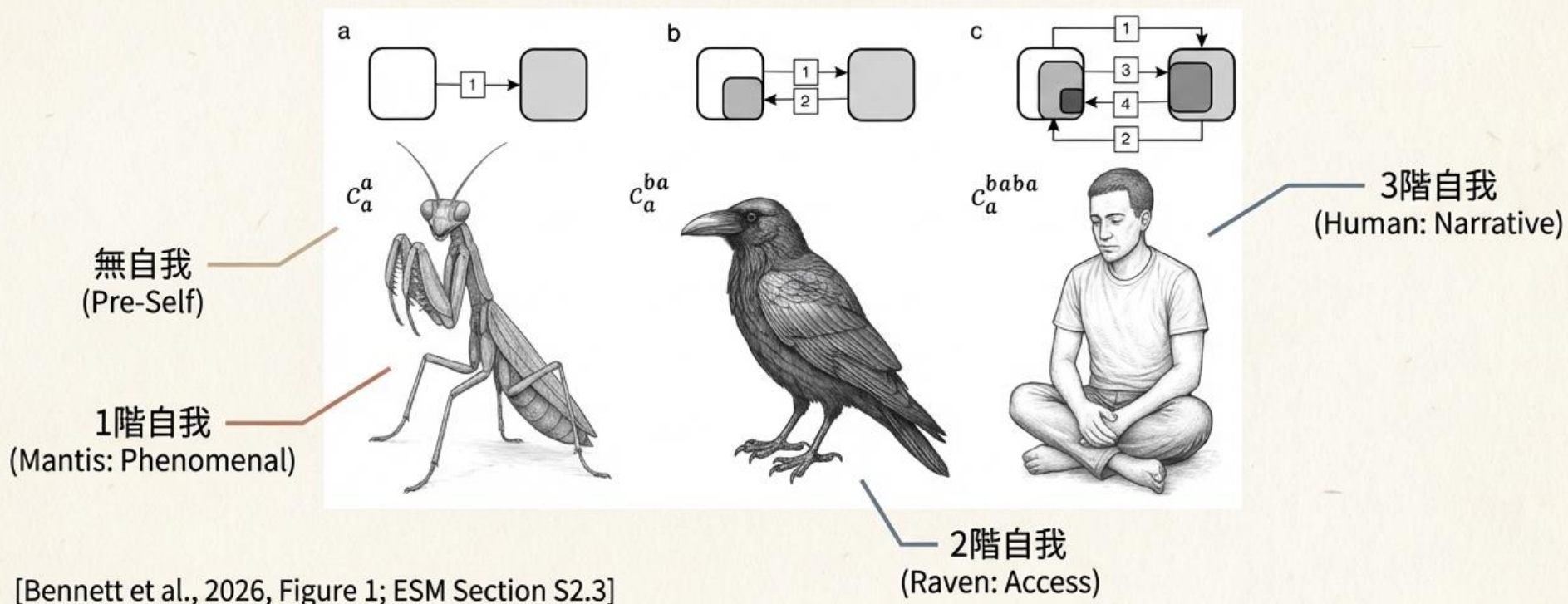


結論: 道德評估並非單一步驟，而是從「情感共鳴」動態過渡到「意圖理解」的連續歷程。

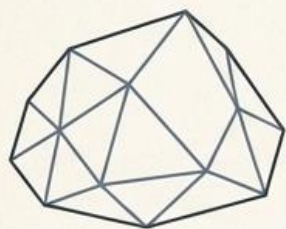
參考文獻：Kamela Cenka et al., 2024

意識的演化階梯：嵌套因果身份理論

意識並非非黑即白的開關，而是隨著演化壓力與預測需求提升的「階梯」。每個階梯代表一種更複雜的「自我模型 (Models of Self)」。

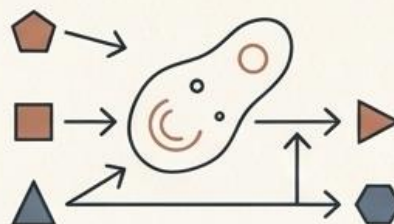


階段 0-2：前自我時代的無意識運算



階段 0（惰性系統）

岩石等單純物理反應，無自我標籤。



階段 1（硬編碼適應）

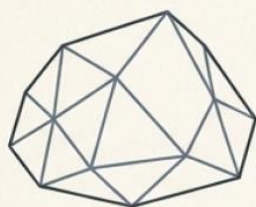
單細胞原生動物。自然選擇將適應行為（如趨化性）寫死在結構中，僅具備一維的效價反應。



階段 2（無自我學習）

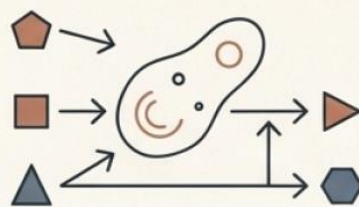
線蟲（Nematodes）與水母。具備聯想學習能力，但神經系統高度分散，缺乏統一的「自我」中樞。資訊處理依然在「黑暗」中進行。

階段 0-2：前自我時代的無意識運算



階段 0（惰性系統）

岩石等單純物理反應，無自我標籤。



階段 1（硬編碼適應）

單細胞原生動物。自然選擇將適應行為（如趨化性）寫死在結構中，僅具備一維的效價反應。

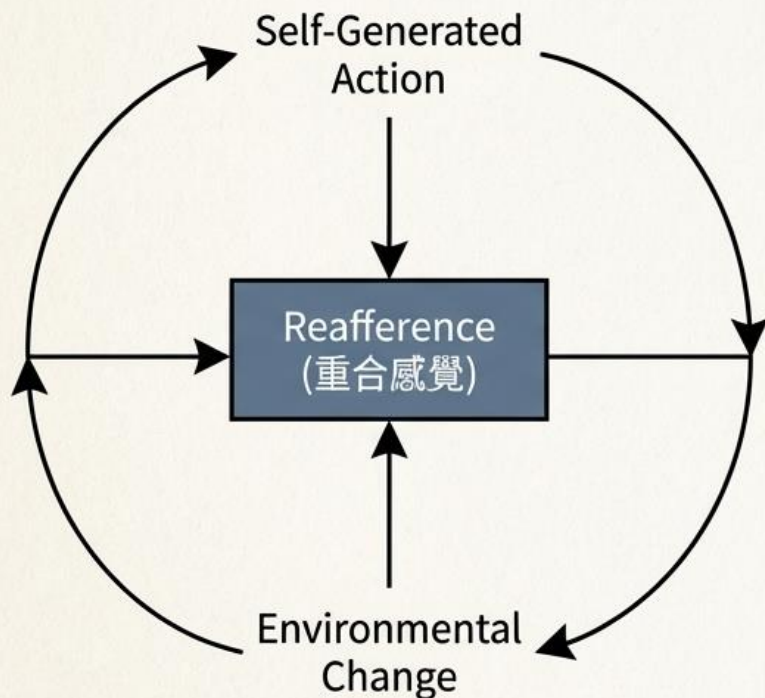


階段 2（無自我學習）

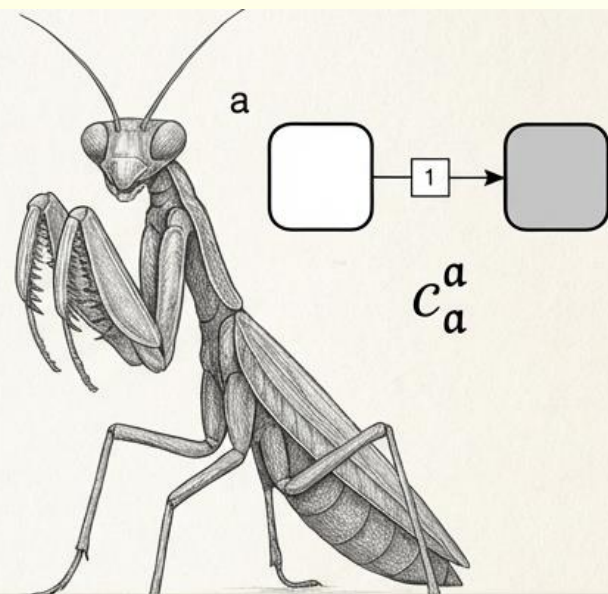
線蟲（Nematodes）與水母。具備聯想學習能力，但神經系統高度分散，缺乏統一的「自我」中樞。資訊處理依然在「黑暗」中進行。

[Bennett et al., 2026, Section 8; Barron and Klein, 2016]

階段 3：一階自我與現象意識的誕生



[Bennett et al., 2026, Section 5.1 & 8; Merker, 2005]



Definition Panel

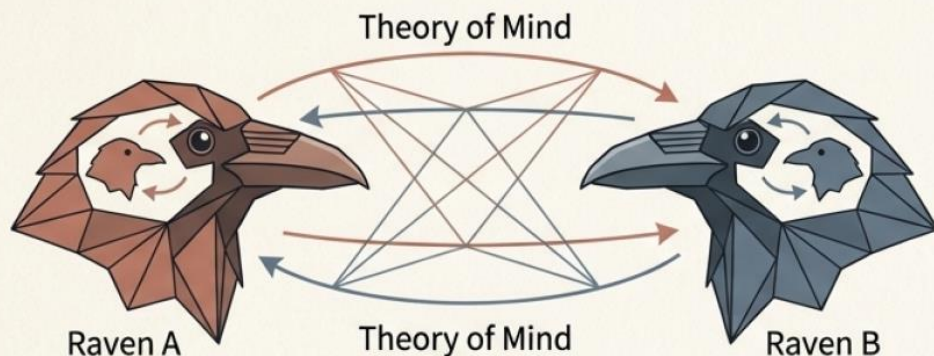
一階自我 (1ST Order Self)：最弱的**因果身份**，能將「自我引發的干預」與「環境產生的觀察」區分開來。

Significance Panel

重合感覺 (Reafferrence)：蒼蠅或螳螂必須**區分**「是我在移動」還是「世界在移動」。

現象意識 (Phenomenal Consciousness)：這是**主觀體驗**的起點。擁有一階自我的生物，開始擁有了「作為牠自身的感覺」(what it is like to be)。

階段 4：二階自我與報告意識



Definition Panel

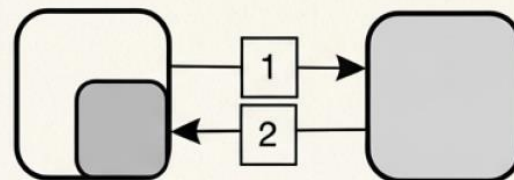
二階自我 (2ND Order Self)：具備「受眾模型 (Audience Model)」。預測他者如何看待自己。

Significance Panel

格萊斯式通訊 (Gricean Report)：真正的溝通需要推斷意圖。烏鴉假裝藏匿食物以欺騙同類，證明牠們能操控他者的意圖。

報告意識 (Access Consciousness)：資訊被提取並用於推理與報告，這是自我意識與心智理論的基石。

[Bennett et al., 2026, Section 5.2; Grice, 1969; Bugnyar and Kotrschal, 2002]



階段 5：三階自我、敘事與信任的數學邏輯

Definition Panel

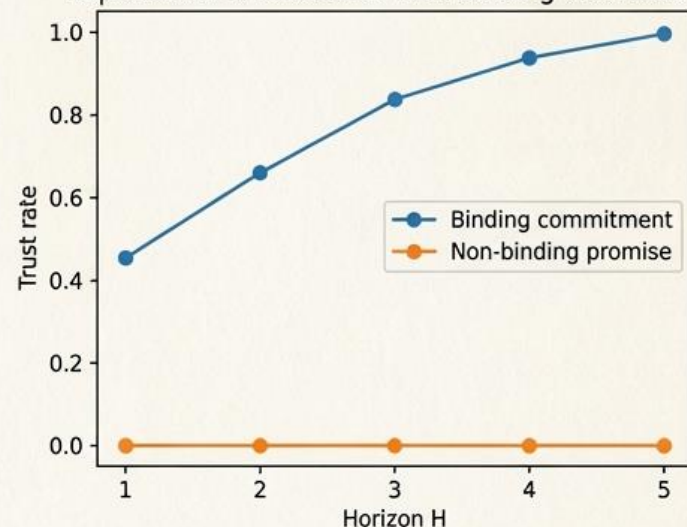
三階自我 (3RD Order Self)：敘事自我。將自我認同跨越時間線穩定下來。

實驗數據與意義

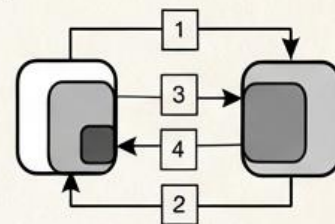
跨期信任博弈 (Trust Games)：實驗證明（見圖表），單純的承諾無法維持均衡的信任。唯有透過「**自我綁定 (Self-binding)**」——付出成本消除未來剝削的選項，才能建立真實的長期協作。

人類複雜社會：這是人類建立長程計畫與深層社會信任的認知基礎。

Experiment 3. Trust rises with binding and horizon



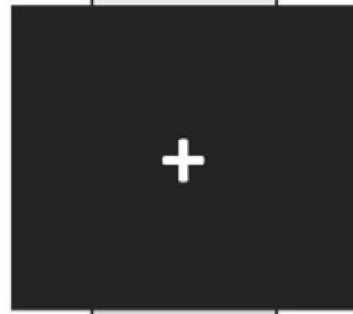
C_a^{baba}



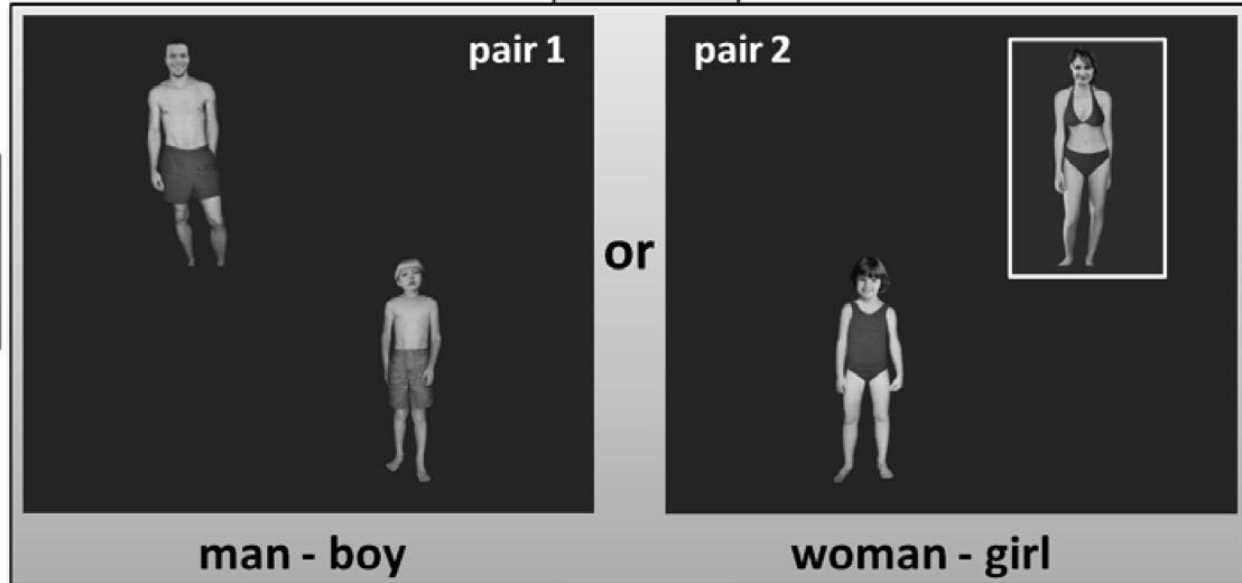
[Bennett et al., 2026, Section 5.3 & Experiment S3]

(a)

fixation
(≥ 500 ms)



**presentation
of paired
pictures**
(5000 ms)



task
(self paced)

**Was one of
these persons
more
attractive?**

(Fromberger, Jordan, et al., 2013)



PJF01



PJF02



PJF03



PJF04



PJF05



PJF06



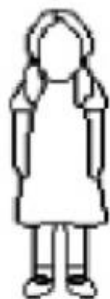
PJF07



SCF



SCM



P J F



P J M



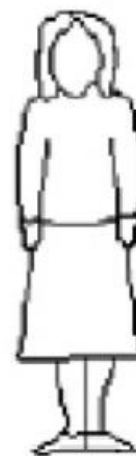
J U F



J U M



A D M



A D F



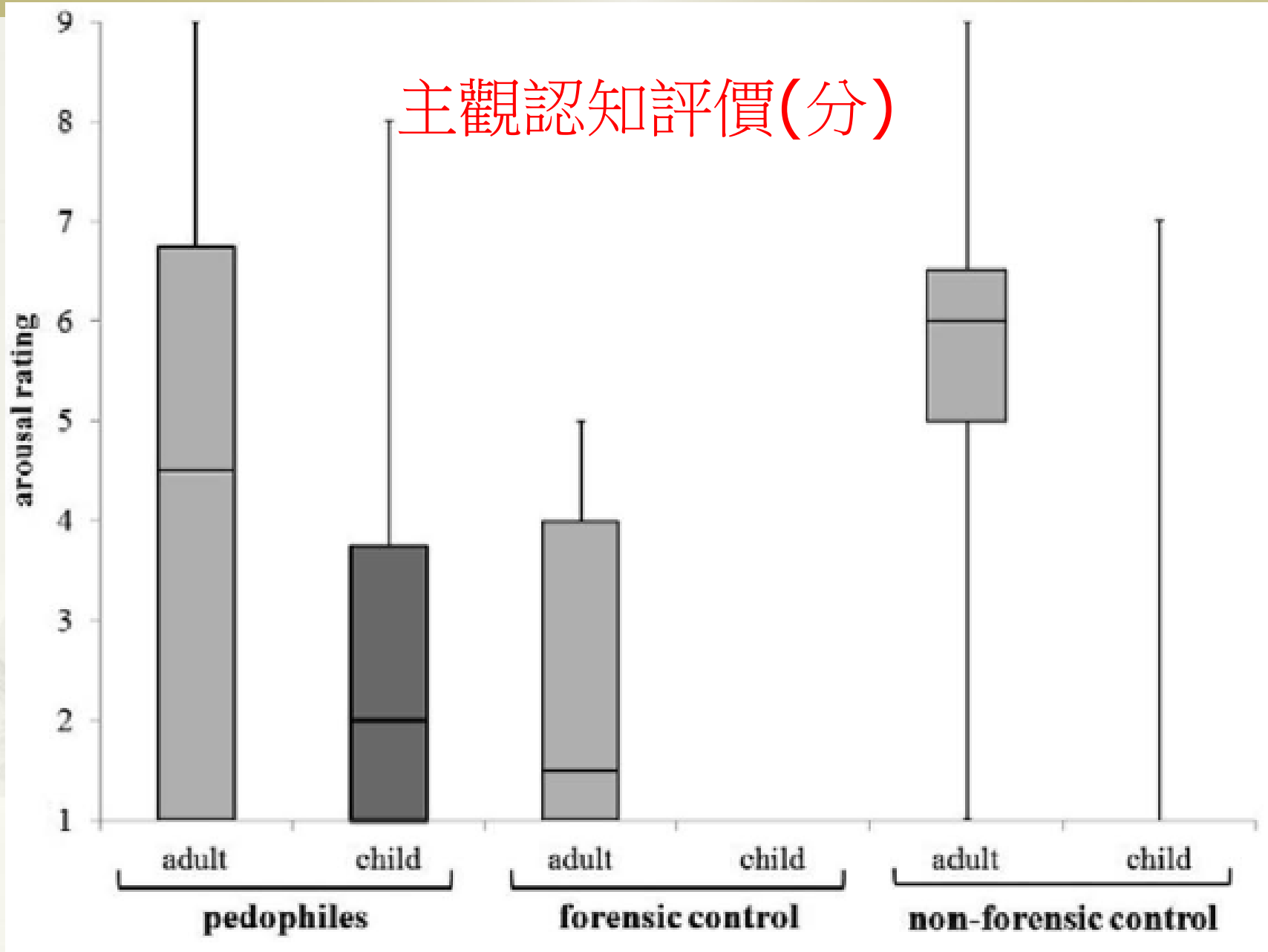
S e x u a l l y

← Unattractive

Attractive →

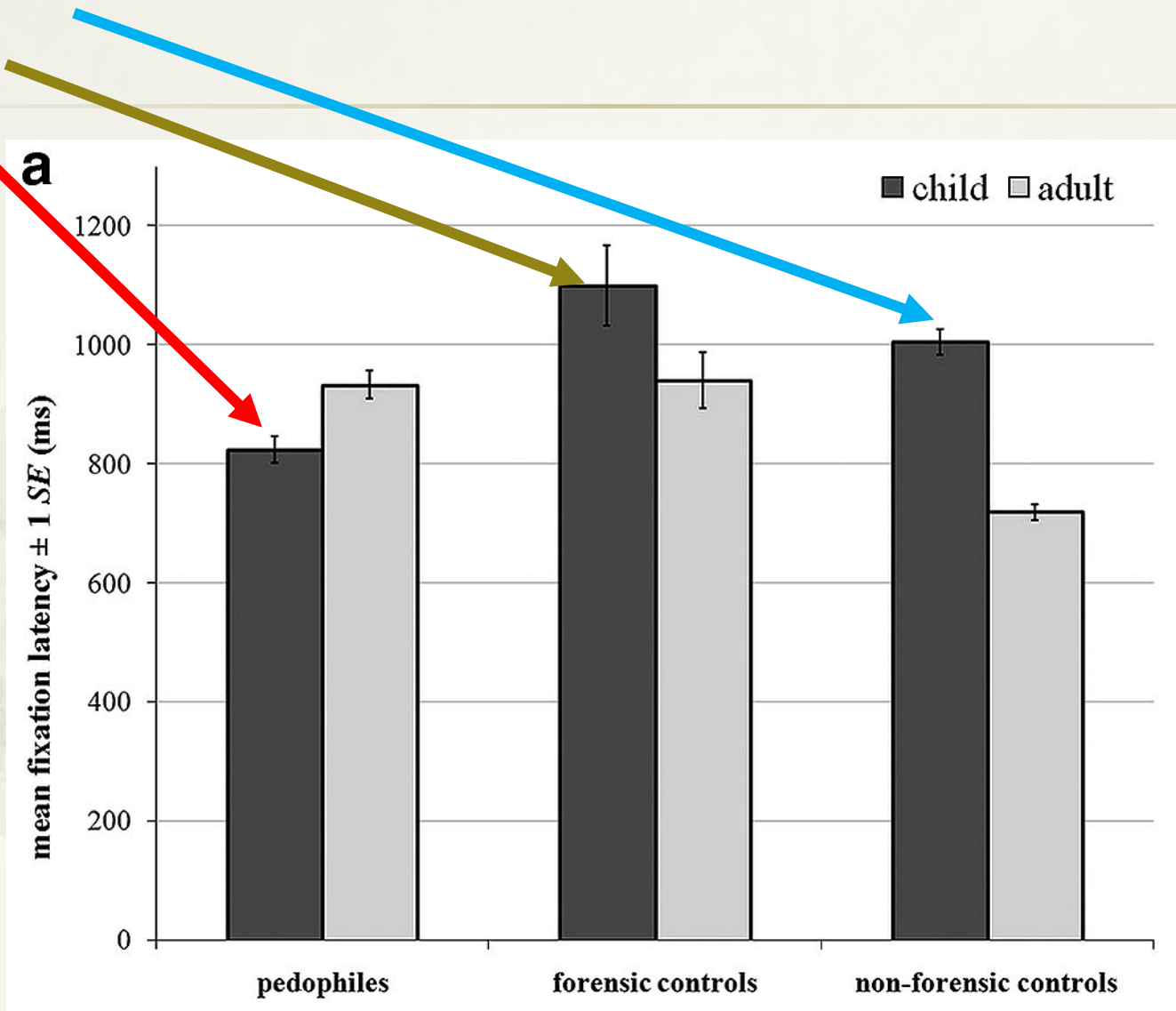


N e x t p i c t u r e

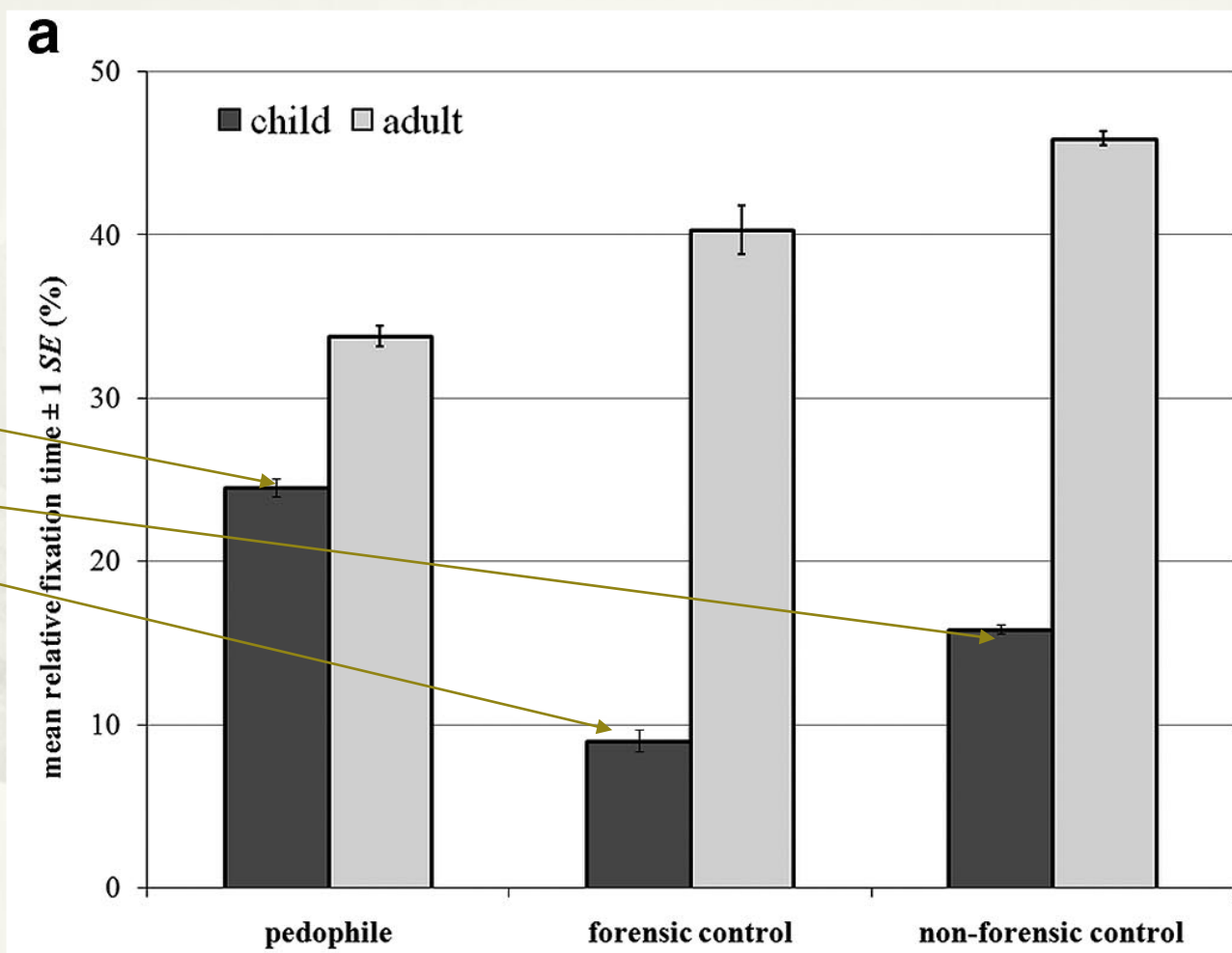


圖一：請評分這張圖有多麼讓你很有感 (Arousal Rating)

最快看到小孩



注視時間最長



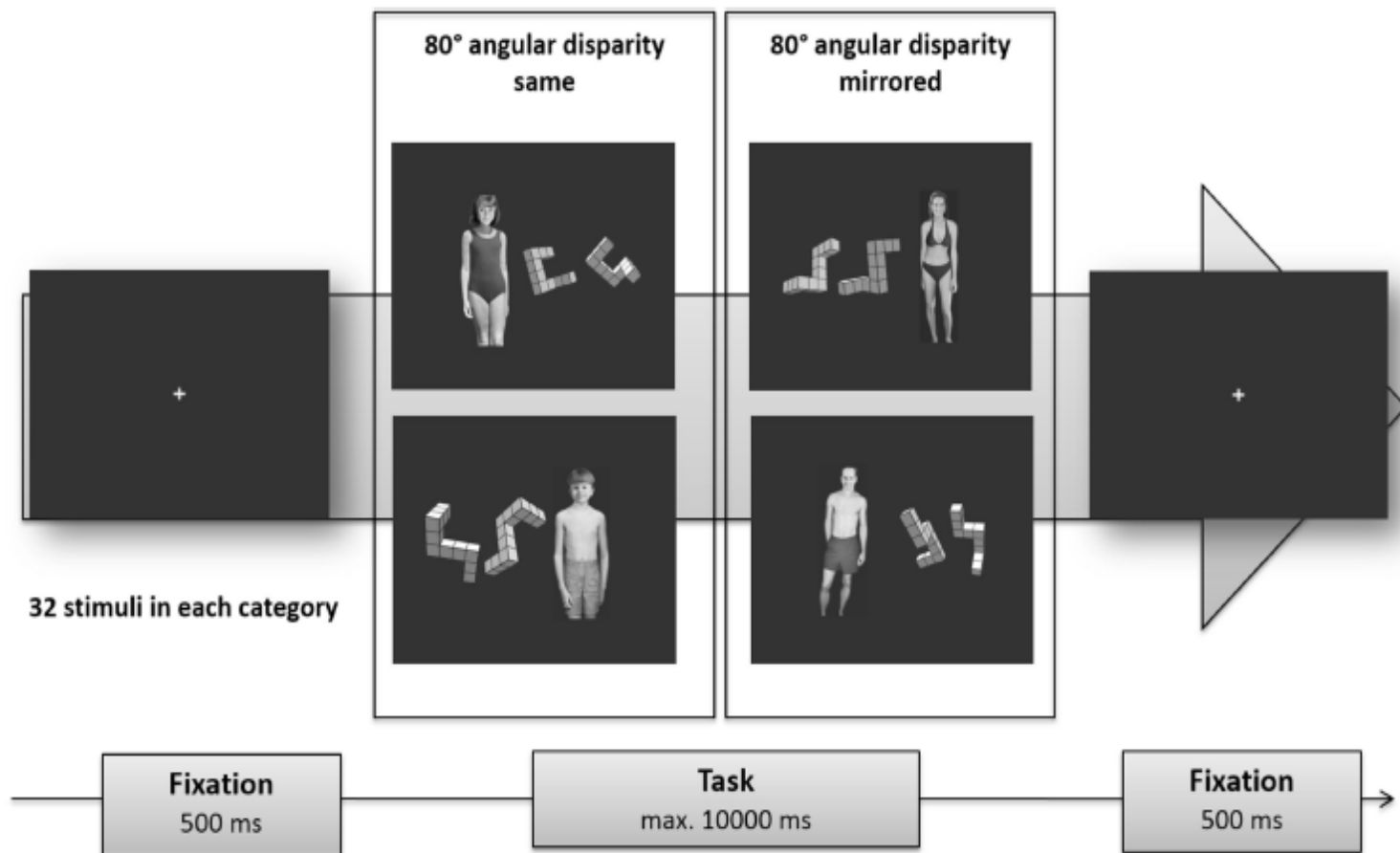
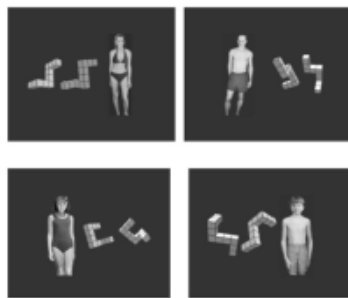
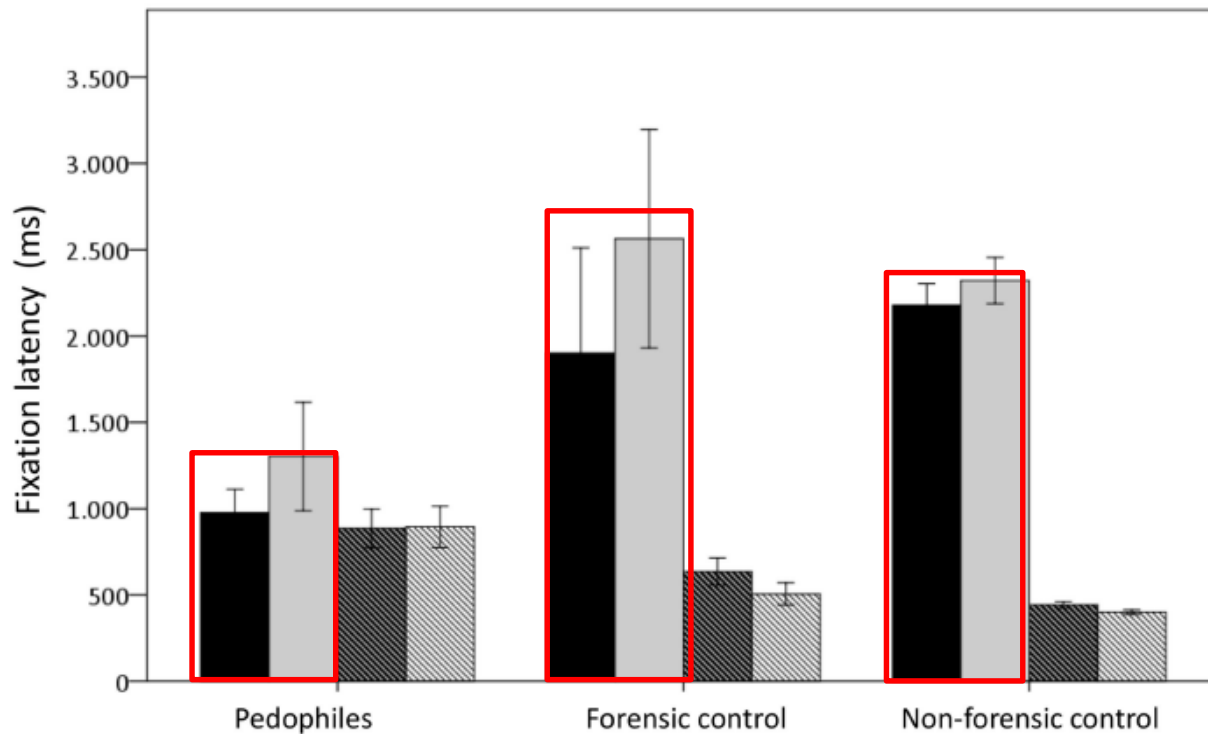


FIGURE 1 | Experimental design [see also Ref. (29)]. Given are examples for each condition. For each mental rotation task, one sexual distractor simultaneously was presented, a girl, boy, woman, or man (note that these stimuli were not included in the main task). Sexual distractor stimuli were taken from the NRP-set (36). Mental rotation stimuli were selected from the set developed by Paschke et al. (37). For details, see Section “Materials and Methods.”

(Jordan, Fromberger, et al., 2016)

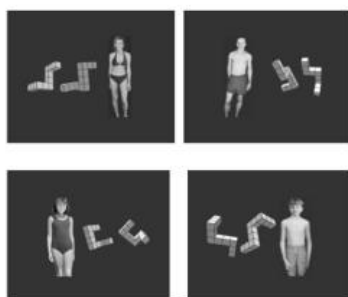
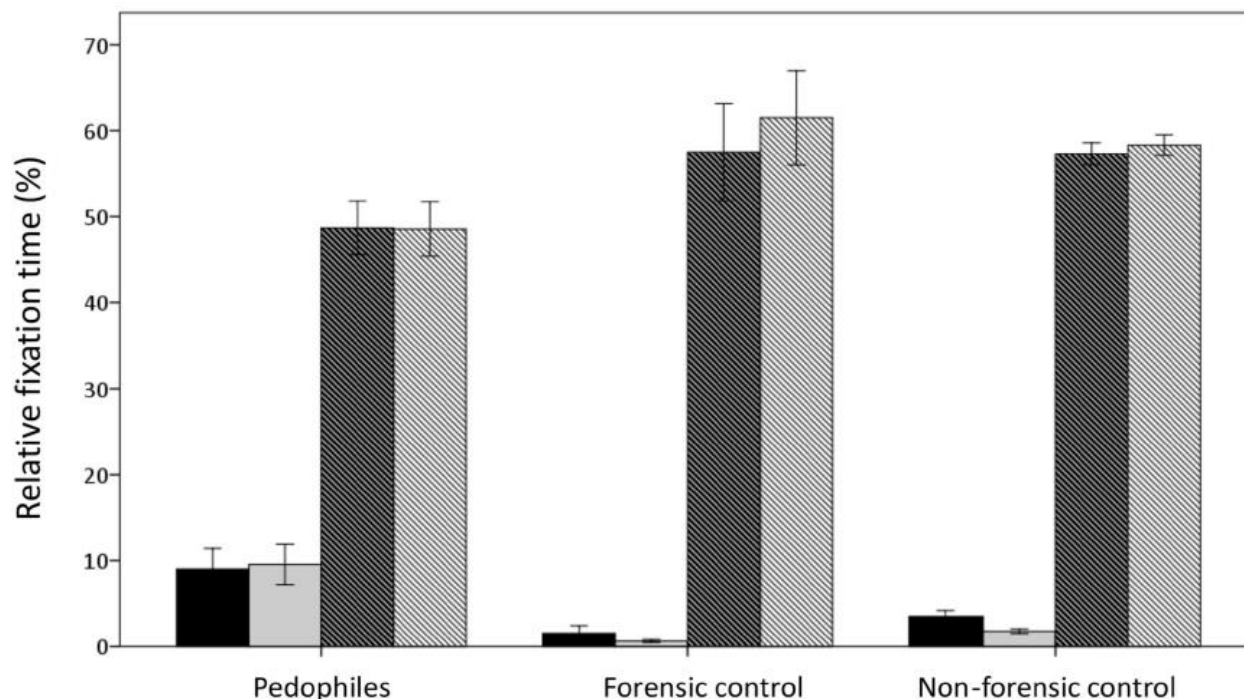
開始注意力



■ Adult sexual distractors ▨ Mental rotation stimuli with adult sexual distractors
 ■ Child sexual distractors ▨ Mental rotation stimuli with child sexual distractors

FIGURE 2 | Eye movements in the sexual distractor task: fixation latency. Means and SEs are shown of the fixation latency for the first fixation with respect to stimulus type (single-colored: sexual distractors, striped: mental rotation figures) and distractor category (black/black-striped: adult, gray/gray-striped: child). This resulted in four different stimuli. Black bars: fixation latency for the adult sexual distractor. Gray bars: fixation latency for the child sexual distractor. Striped black bars: fixation latency for the mental rotation stimulus if the adult sexual distractor was simultaneously presented. Striped gray bars: fixation latency for the mental rotation stimulus if the child sexual distractor was simultaneously presented.

晚期注意力分配



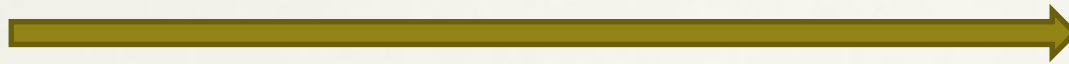
Adult sexual distractors
 Mental rotation stimuli with adult sexual distractors

Child sexual distractors
 Mental rotation stimuli with child sexual distractors

FIGURE 3 | Eye movements in the sexual distractor task: relative fixation time. Means and SEs are shown of the relative fixation time with respect to stimulus type (single-colored: sexual distractors, striped: mental rotation figures) and distractor category (black/black-striped: adult, gray/gray-striped: child). This resulted in four different stimuli. Black bars: fixation latency for the adult sexual distractor. Gray bars: fixation latency for the child sexual distractor. Striped black bars: fixation latency for the mental rotation stimulus if the adult sexual distractor was simultaneously presented. Striped gray bars: fixation latency for the mental rotation stimulus if the child sexual distractor was simultaneously presented.

IHS (蓄意傷害)

惡意動機 + 負面結果



ion

AHS (意外傷害)

中性動機 + 負面結果



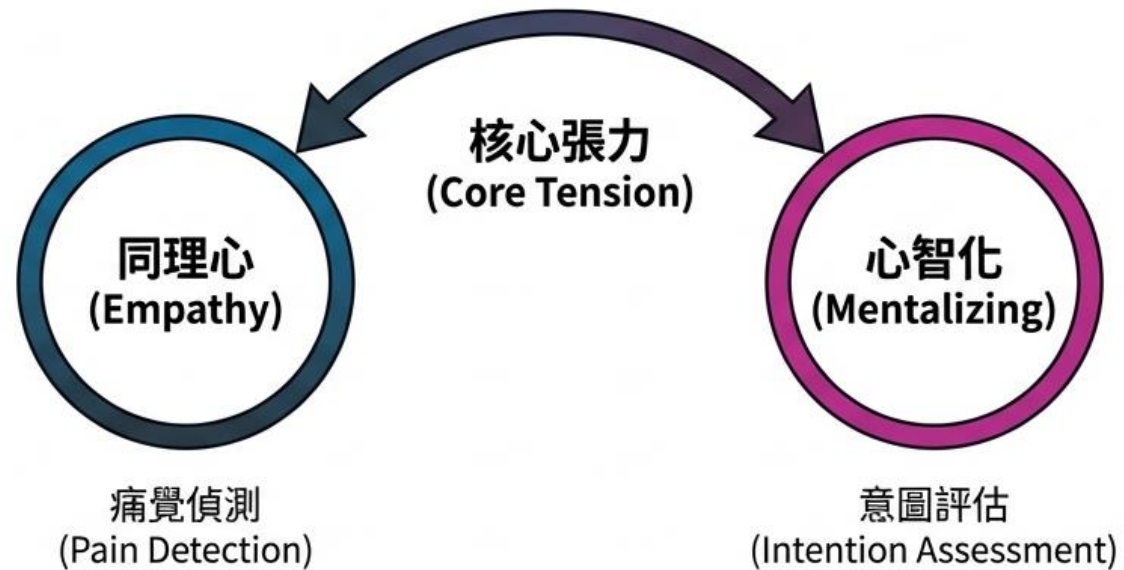
ion

NAS (中性行為)

中性動機 + 中性結果

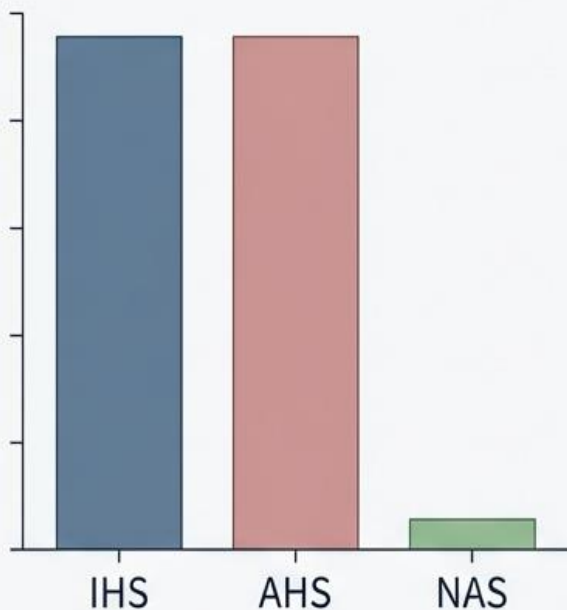


- **雙重歷程模型 (Two-Component Model)**：道德決策涉及兩個關鍵要素的整合。
- **同理心**：對受害者痛苦的快速、自動化情感共鳴。
- **心智化**：對加害者意圖的認知評估（蓄意 vs. 意外）。
- **核心問題**：大腦如何在毫秒級的時間尺度內，從「看到痛苦」轉變為「判斷責任」？



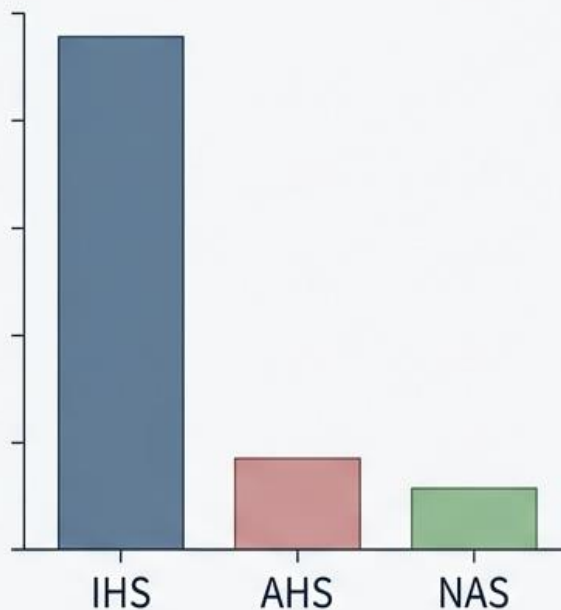
意識層面的評估：行為數據驗證

物理疼痛感



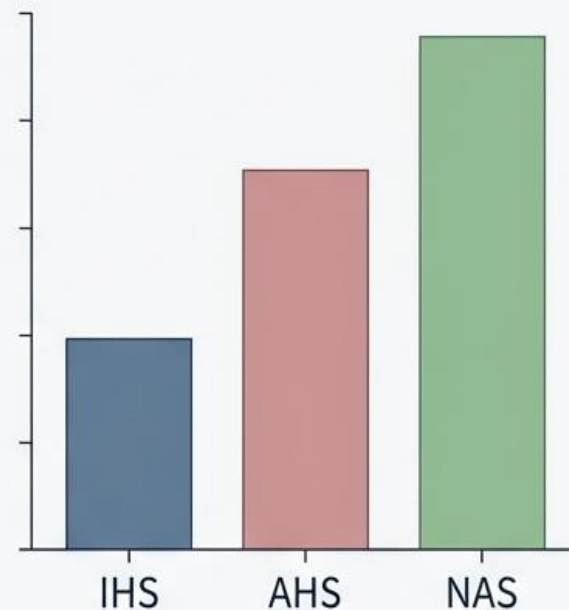
蓄意 = 意外

惡意意圖



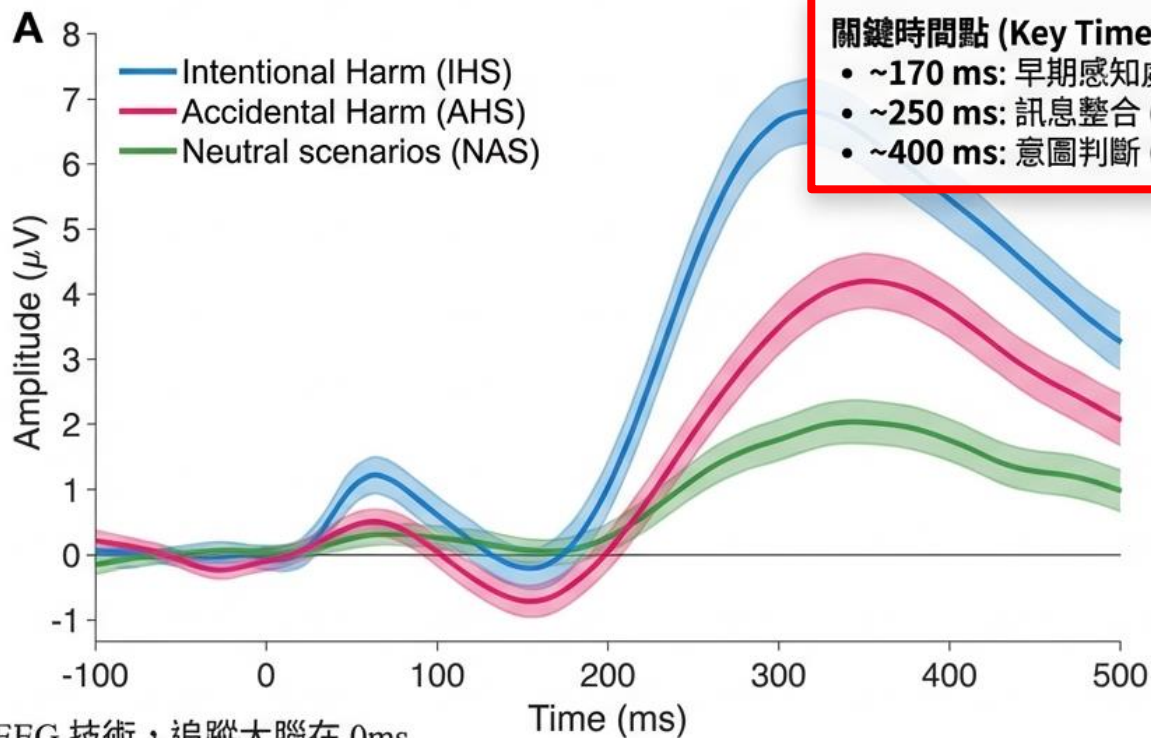
蓄意 >>> 意外

道德許可性



蓄意最不可接受

神經動態總覽：事件相關電位 (ERPs)



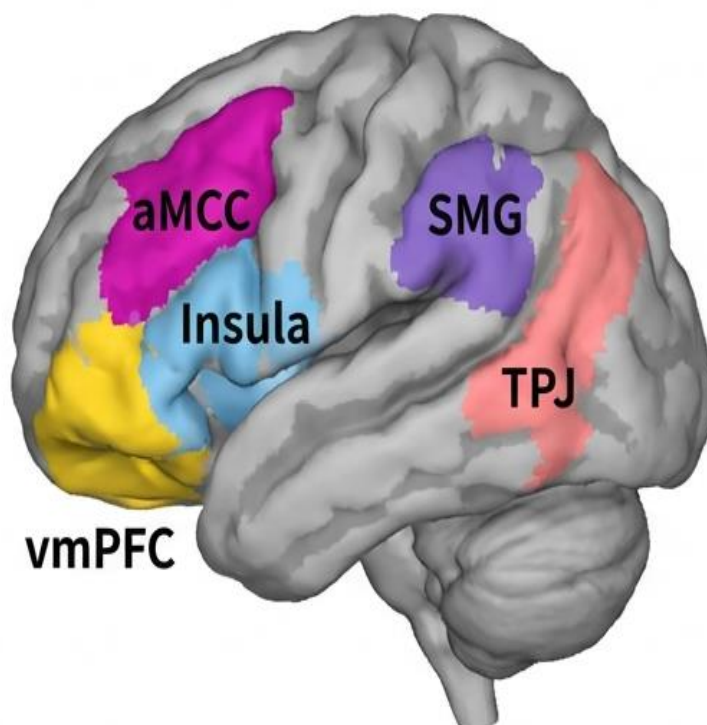
利用高時間解析度的 EEG 技術，追蹤大腦在 0ms 至 500ms 間的反應。波形在不同時間點的分離代表了大腦處理資訊的不同階段。

參考文獻：Kamela Cenka et al., 2024

腦區定位與功能總結 (Regions of Interest)

同理心網絡 (Empathy)

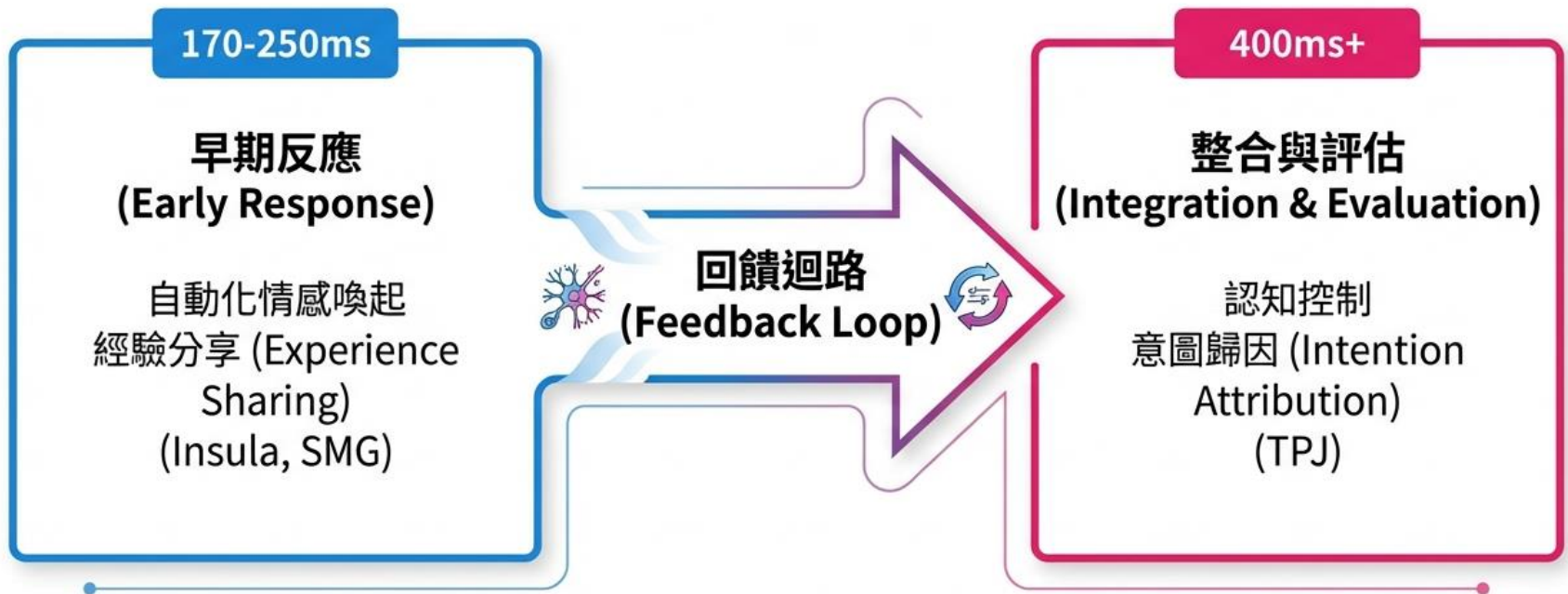
- Insula (島葉)：情感與痛覺。
- aMCC (前中扣帶皮層)：痛覺共鳴。
- SMG (緣上回)：自我/他人視角轉換。



心智化網絡 (Mentalizing)

- TPJ (顳頂交界區)：推論他人意圖。
- vmPFC (腹內側前額葉)：社會價值決策。

理論模型：從自動化反應到認知控制



結論: 道德評估並非單一步驟，而是從「情感共鳴」動態過渡到「意圖理解」的連續歷程。

參考文獻：Kamela Cenka et al., 2024

否認?道德記憶:在心智中出現?

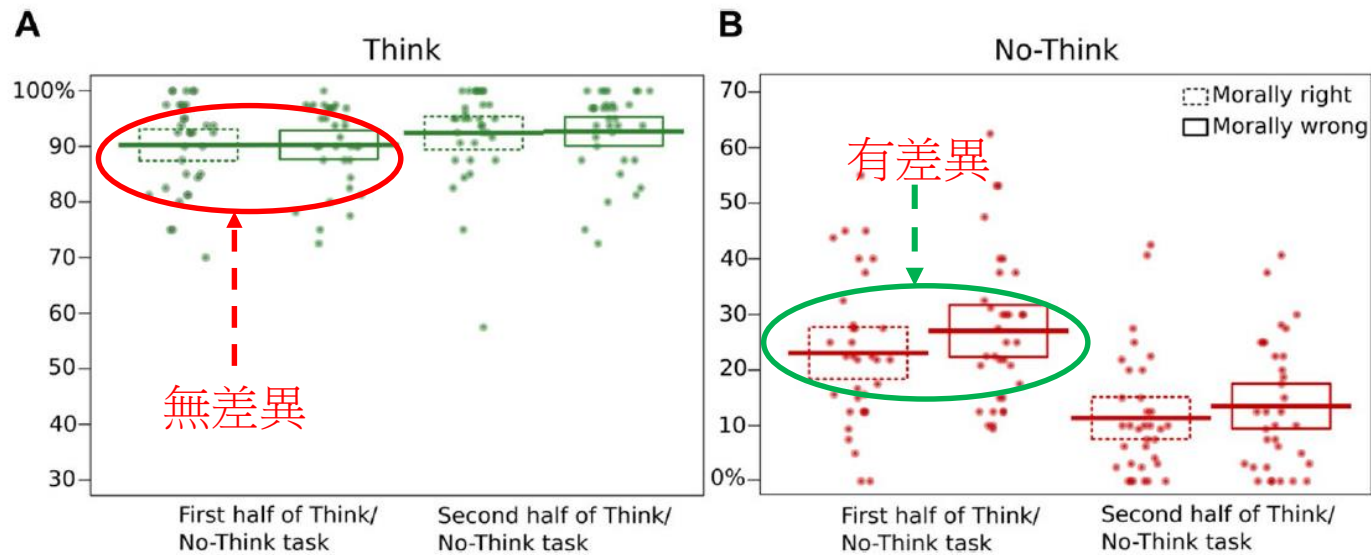


Fig.2 Percentage of trials eliciting memory retrieval compared across memory type and halves of TNT task (first half and second half) in the a) Think condition, reflecting retrieval success and b) No-Think condition, reflecting intrusion frequency. Dots show the per-

centage of trials with retrieval for each individual. Thick lines show the group means and the boxes depict the 95% confidence interval of the group means

過去的道德錯誤很難刻意的被壓抑下去

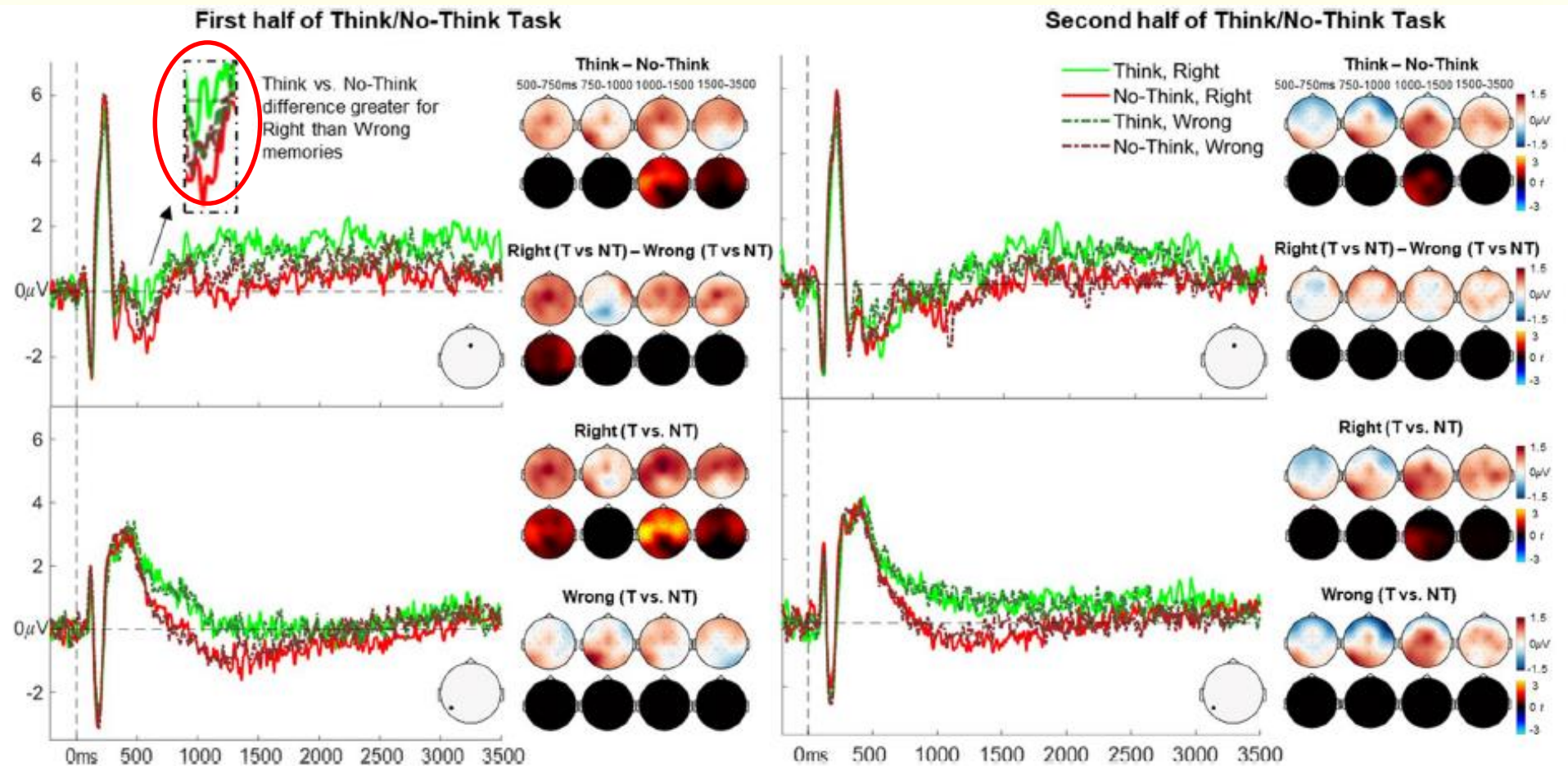
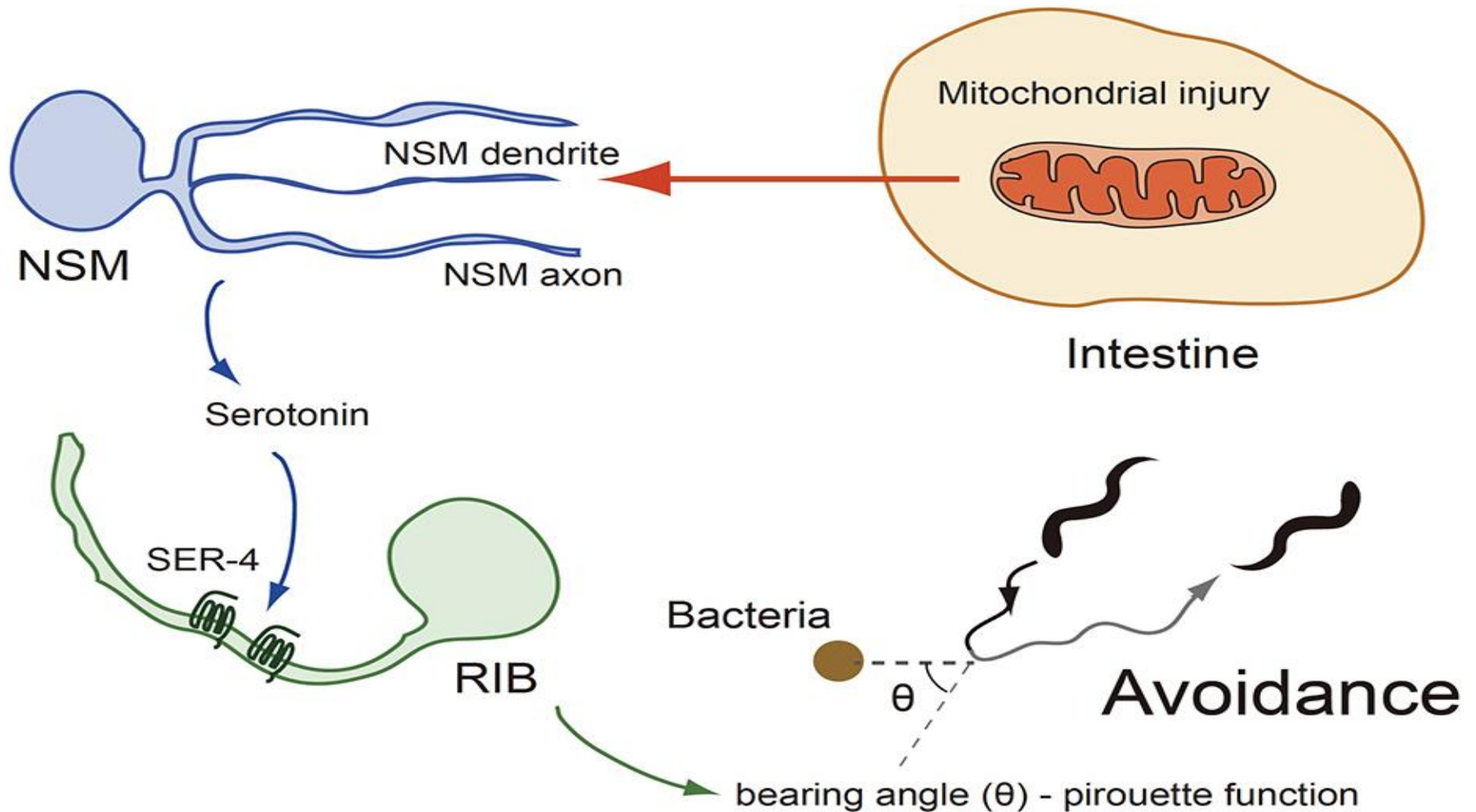


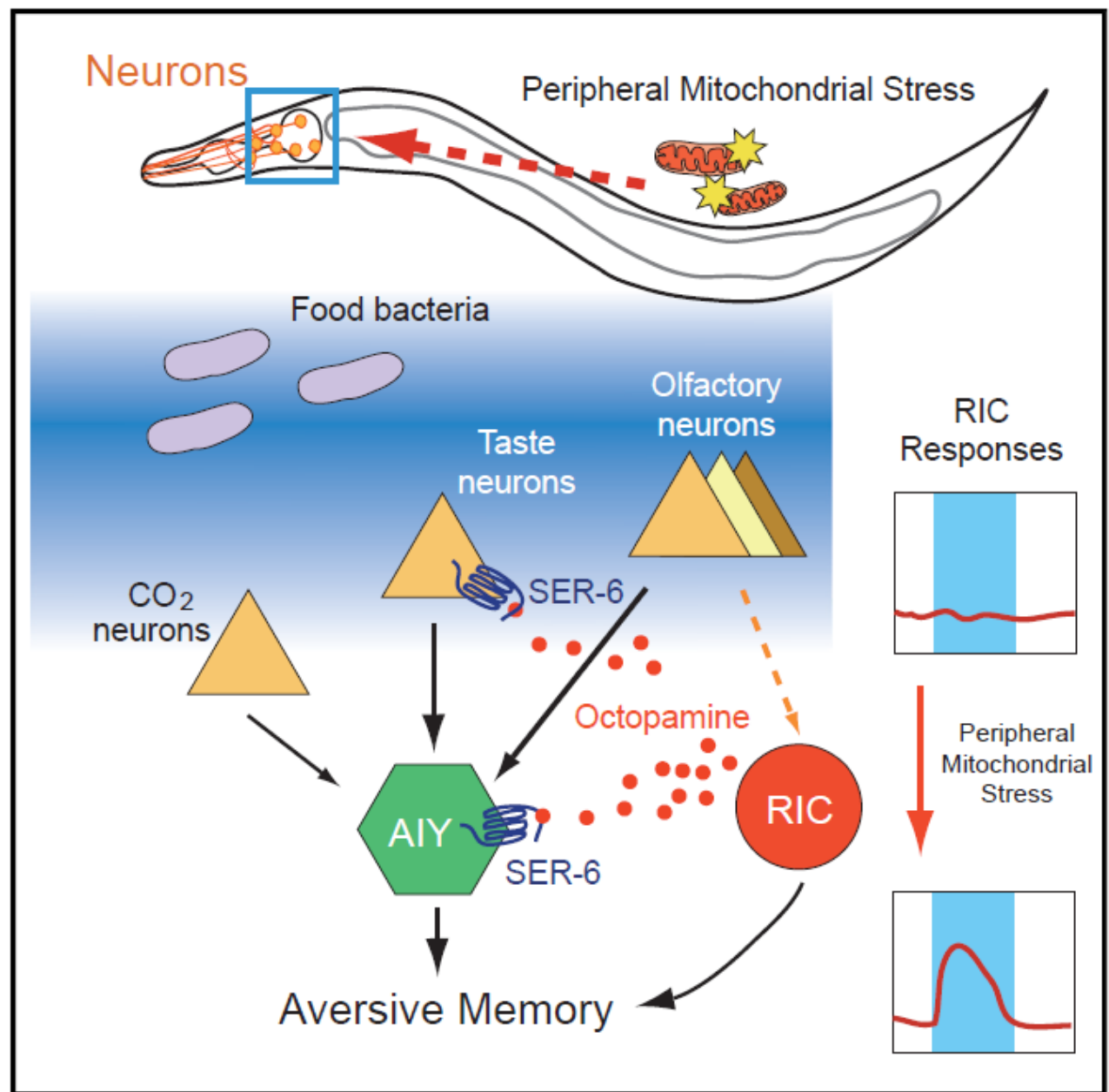
Fig. 3 Grand average ERPs and results from the cluster permutation tests. ERP graphs show the experimental conditions at midfrontal (Fz, top) and left parietal (P7, bottom) electrode sites during first half and second half of the TNT task. Topographic maps on the top rows illustrate mean ERP amplitude differences between conditions (blue/

white/red colourmap, in μV), and topographic maps on the bottom rows illustrates mean t -values (blue/black/red colourmap, in t) for the condition differences as generated through the cluster-based permutation tests. The colour scale represents magnitude and direction of the effect. T -maps have been thresholded to only show significant clusters

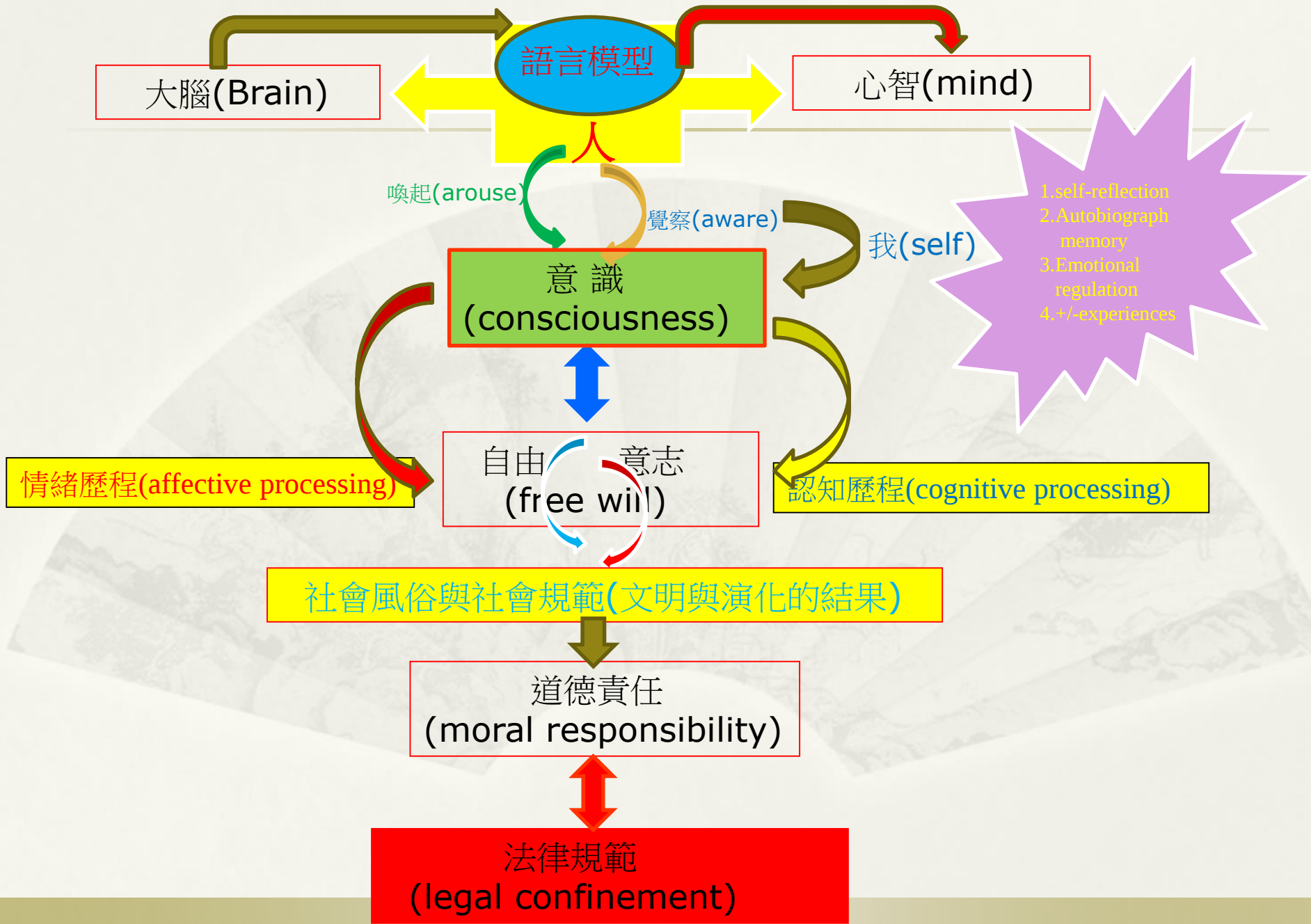


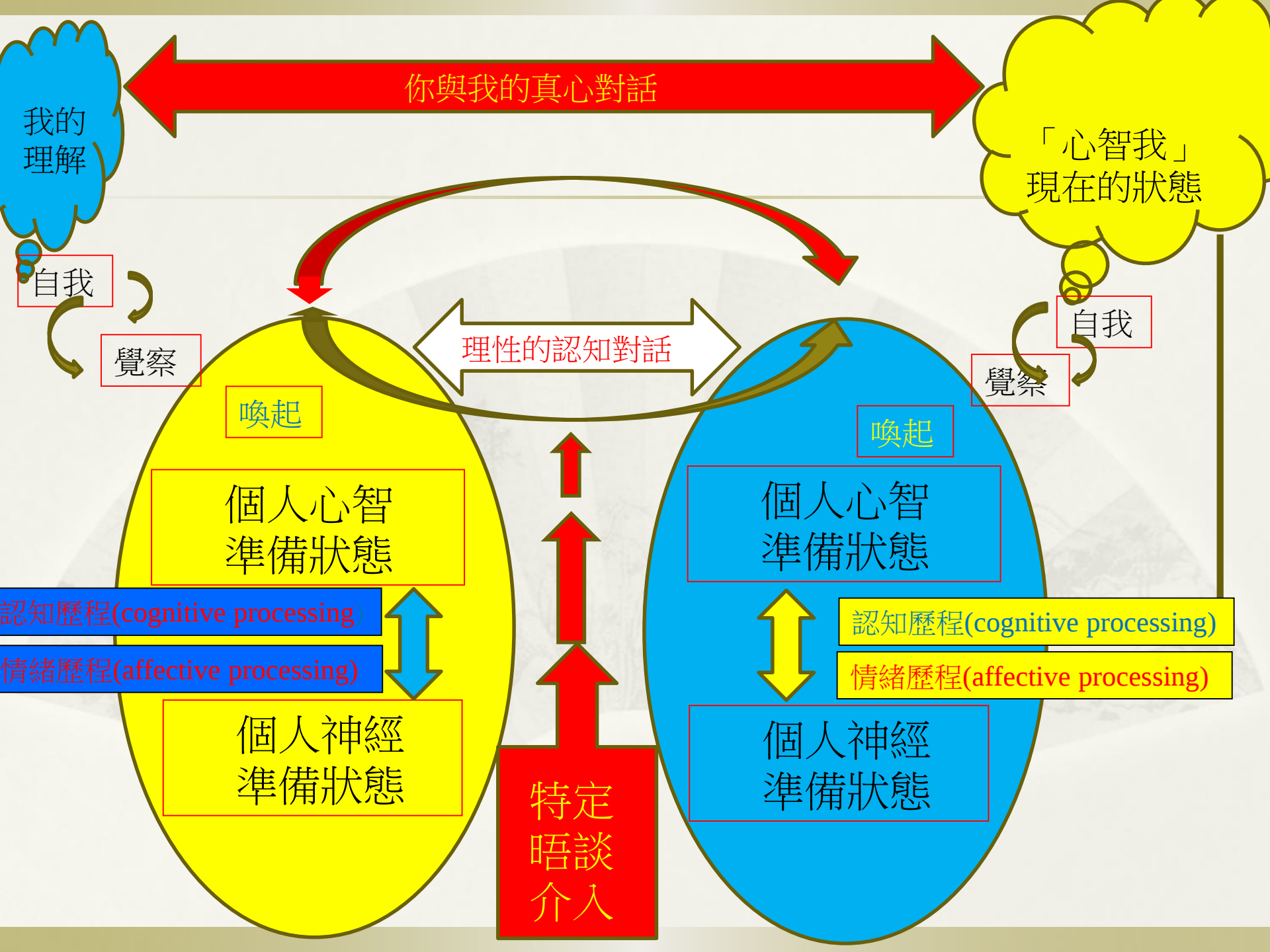
A serotonergic circuit regulates aversive associative learning under mitochondrial stress in *C. elegans* (Chiang, Liao, & Pan, 2022)

Under mitochondrial stress, octopamine secreted from the RIC modulatory neuron targets the AIY interneuron through the SER-6 receptor to trigger learned bacterial aversion. RIC responds to systemic mitochondrial stress by increasing octopamine synthesis and acts in the formation of aversive memory. AIY integrates sensory information, acts downstream of RIC, and is important for the retrieval of aversive memory.



神經與心智





你與我的真心對話

我的理解

「心智我」
現在的狀態

自我

自我

覺察

覺察

喚起

喚起

個人心智
準備狀態

個人心智
準備狀態

認知歷程(cognitive processing)

情緒歷程(affective processing)

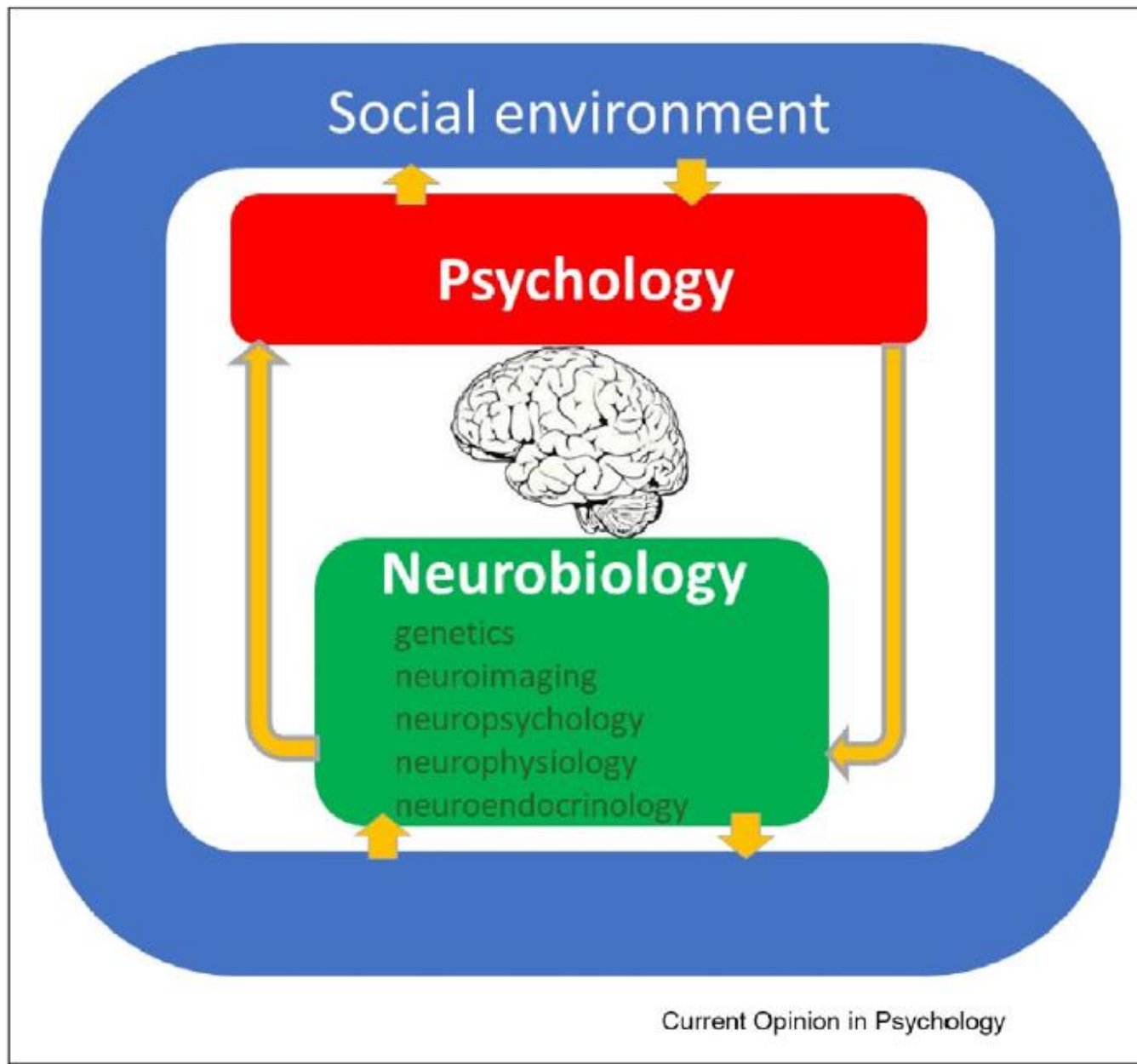
認知歷程(cognitive processing)

情緒歷程(affective processing)

個人神經
準備狀態

個人神經
準備狀態

特定
晤談
介入



Schematic representation of the biopsychosocial model.



Translation research cycle

Current Opinion in Psychology