神經與行為模型建構
(Neural & Behavioral Modeling)

課號：Psy7277　　教室： 北 206
識別碼：227M9280　　時間： 五 234

Quantitative psychology:
- psychometrics
- mathematatic psychology

認知心理學是什麼，那 computational cognitive psychology?

Overview

Parameter Estimation

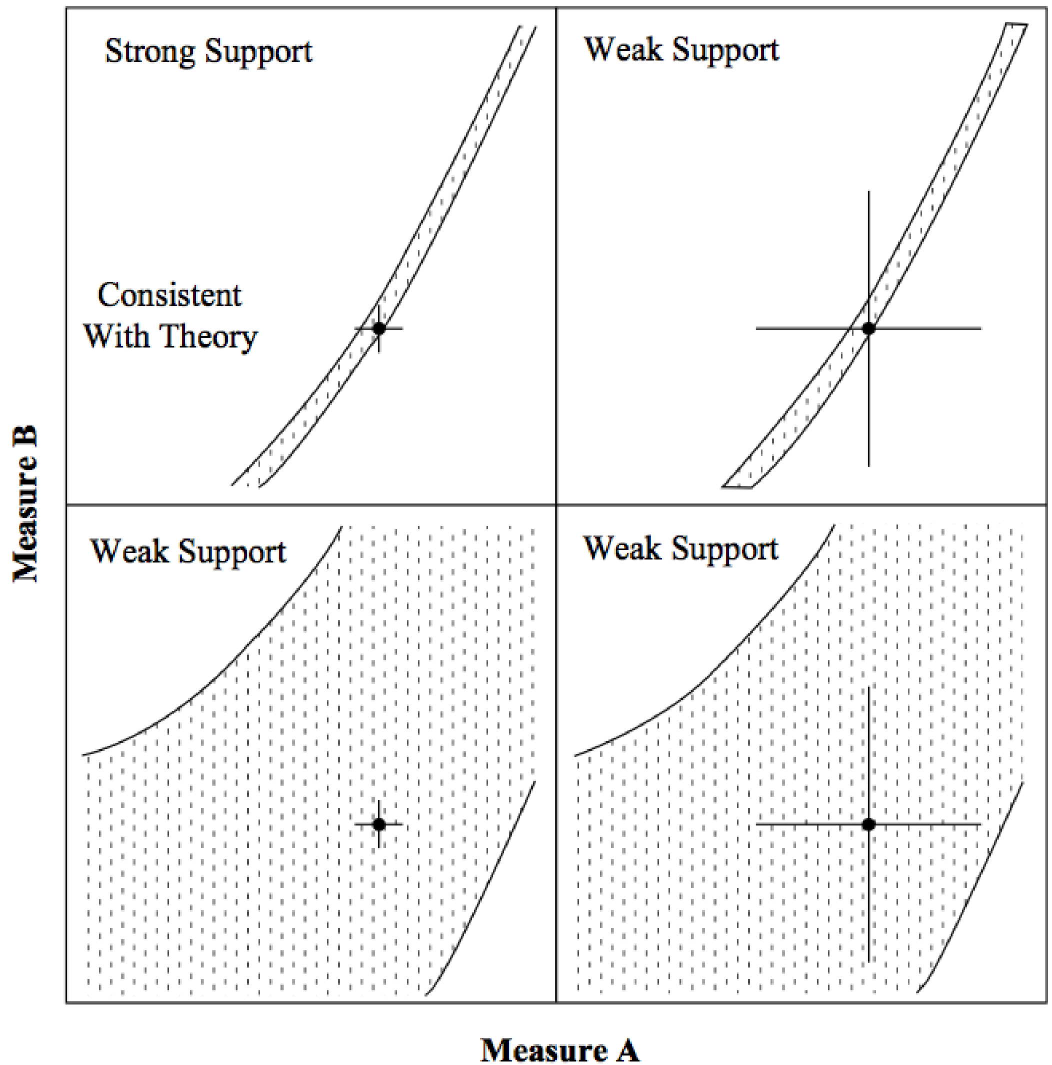
Model Comparison

Reaction Time

Choice Accuracy

理論 / 模型 vs. 現實 / 資料

若模型對資料 (A,B) 預測的範圍 (點點區) 太廣
則模型變成不可否証 (unfalsifiable)



心理 / 行為 / 認知模型建構：特色

simulation model, 對 model "可行性"的證明，看是否能描述某些現象，尚未被否決掉，但不代表是對的

目標是了解 / 解釋認知歷程

生成模型：夠了解系統到可製造特定現象

一種機制的可行性證明（不代表一定是此機制）

使用精確的語言描述

沒有口語描述的含糊空間（如 A 比 B 大，怎樣大 ?)

避免邏輯思考的各種陷阱（如雙向分離≠兩個系統）

符合認知歷程的原理原則

不做違反認知歷程的假設（如 RT 不為常態分佈）

同樣的模型亦可解釋其他相關的資料 (cf. 統計模型）

心理 / 行為 / 認知模型建構：對象 (1/2)

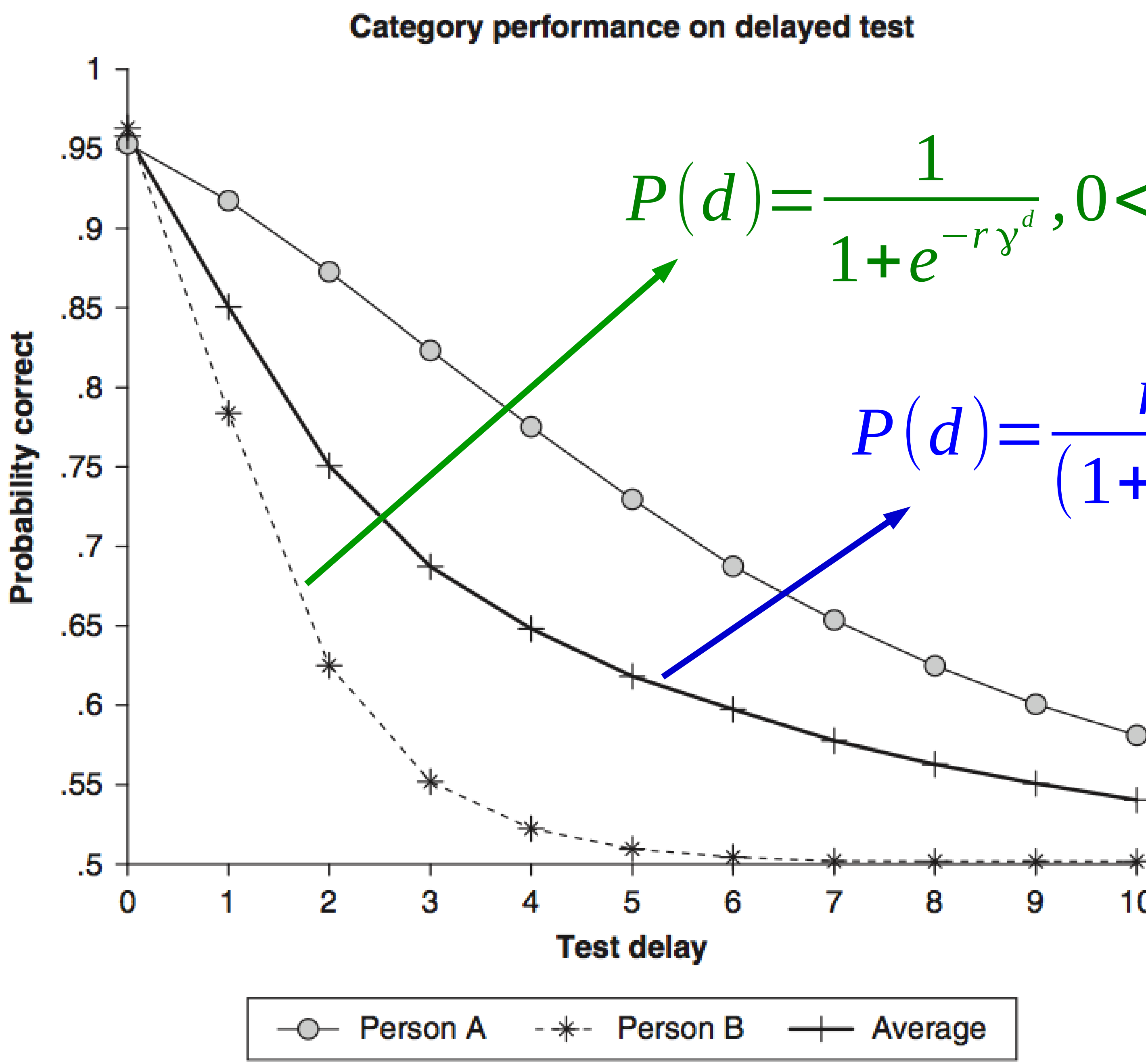
Individual modeling (IM)
一個受試者一個模型
可透過模型參數的不同討論個別差異

Aggregate modeling (cf. fixed effects)
全部受試者一個模型
無法討論個別差異且最適模型可能改變

Hierarchical modeling (cf. random effects)
假設受試者在某些變量上的分佈模型
需要許多受試者才有辦法確定分佈模型的參數

通常認知實驗 N_subj 小 N_trials 大所以使用 IM

心理 / 行為 / 認知模型建構：對象 (2/2)



$$P(d)=\frac{1}{1+e^{-r\gamma^{d}}},0<\gamma<1$$

$$P(d)=\frac{r}{(1+d)^{\gamma}},\gamma>0$$

心理 / 行為 / 認知模型建構：步驟

將概念 / 理論轉化成具體的模型

無法完全確定時需要進一步的假設

儘量減少這些免不了的假設數目

利用實驗觀察到的資料來估計模型參數

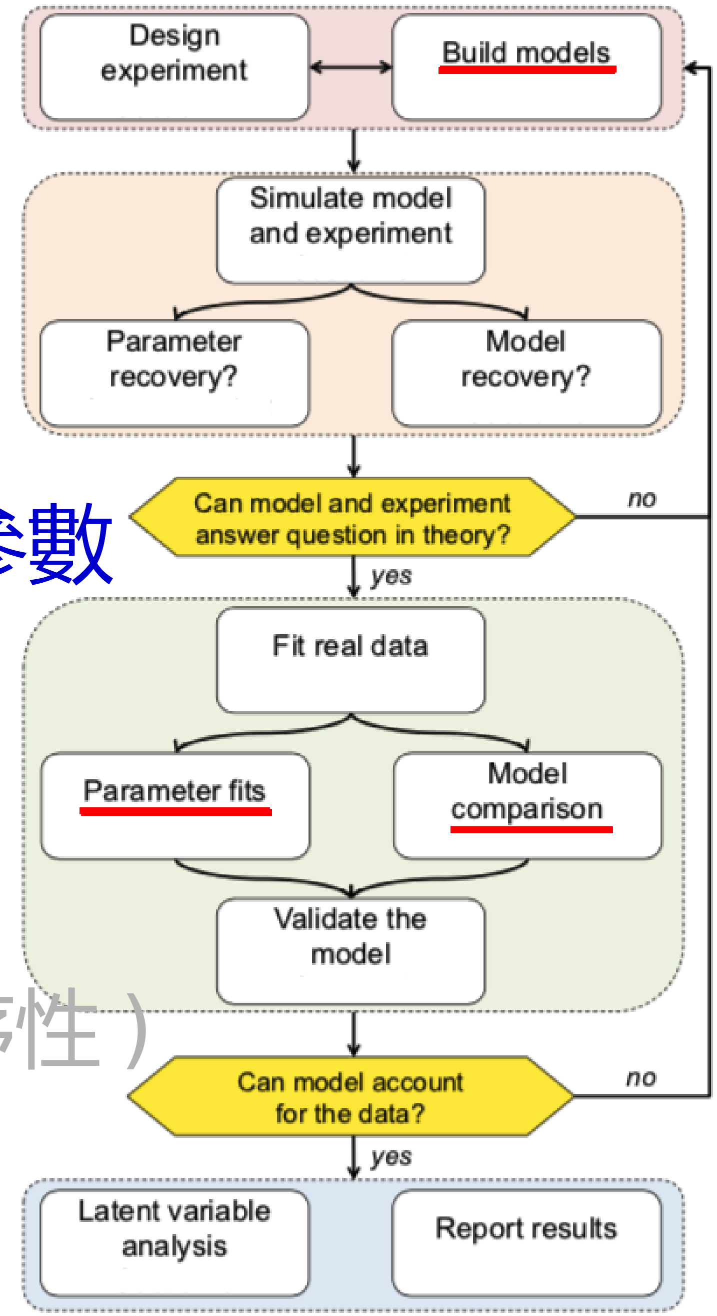
定義目標函數

最小 / 大化此函數是個最佳化問題

模型好壞的比較

定性的比較 (不同條件下結果的次序性)

定量的比較 (e.g. AIC & BIC)



Overview

Parameter Estimation

Model Comparison

Reaction Time

Choice Accuracy

模型裡面的參數

減少可互相取代的自由參數

若兩個參數對模型結果有多對一關係則一者為冗餘

如 $y(t)=e\{-a*(b-c)*d^t\}$ 可簡化為 $y(t)=e\{-ad^t\}$

自由參數反應無法直接量測的心理維度

如 $y(0)=e\{-a\}$ 反應初始的"記憶容量"

如調整"遺忘速度" d 的大小可以改變 y 的陡降程度

模型是借由調整自由參數來逼近資料

實驗資料點的數目要≥模型自由參數的數目

可手動調整參數的大小 (如設為 0) 來產生預測

需要最小化的目標函數

Objective/cost function

Sum of Squared Error (SSE):

$$SSE_{obs,prd} = \sum_{i=1}^{N} (obs_i - prd_i)^2$$

Weighted Sum of Squared Error (WSSE)/
Pearson's Chi-square for goodness-of-fit:

$$\chi^2 = \sum_i \frac{(obs_i - prd_i)^2}{prd_i}$$

需要最大化的目標函數 (1/2)

Pearson Correlation (r):

$$r_{obs,prd} = \frac{\sum_i (obs_i - \mu_{obs})(prd_i - \mu_{prd})}{\sqrt{\sum_i (obs_i - \mu_{obs})^2 \sum_i (prd_i - \mu_{prd})^2}}$$

% of variance explained:

$$\%Var = \frac{SSE_{\text{null}} - SSE_{\text{model}}}{SSE_{\text{null}}}$$

$$SSE_{\text{null}} = \sum_i (obs_i - \mu_{obs})^2$$

$$SSE_{\text{model}} = \sum_i (obs_i - prd_i)^2$$

需要最大化的目標函數 (2/2)

Likelihood: L($\boldsymbol{\theta}_{\mathbf{model}}$|$\mathbf{x}_{\mathbf{data}}$)= 常數 *P($\mathbf{x}_{\mathbf{data}}$|$\boldsymbol{\Theta}$=$\boldsymbol{\theta}_{\mathbf{model}}$)

$$L(\mathrm{p}|\mathrm{x}) = \mathrm{Prob}(x|\,p) = \binom{N}{x} p^{x}(1-p)^{N-x} = \frac{N!}{x!(N-x)!} p^{x}(1-p)^{N-x}$$

$$L(\alpha,\beta \mid x_1,\, x_2,\, \ldots\, x_n) = \prod_i \frac{1}{\beta\sqrt{2\pi}} \exp\left(-\frac{(x_i-\alpha)^2}{2\beta^2}\right)$$

$$L(\lambda|t_1,\, t_2,\, \ldots\, t_n) = \prod_{i=1}^{n} \lambda \exp(-\lambda \cdot t_i)$$

……

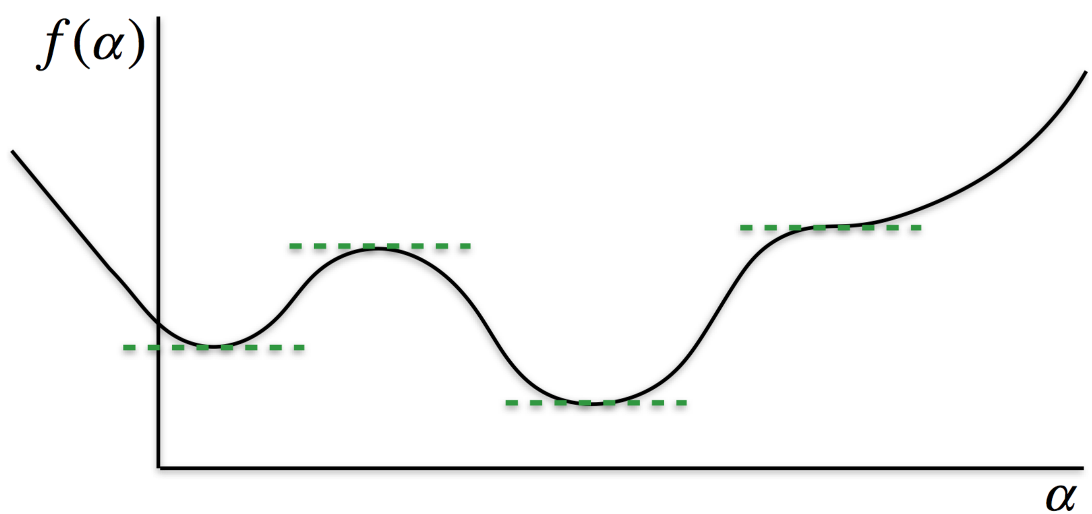
最佳化問題的解析解

若目標函數清楚簡單則可以用微分 =0 求解

$$L(x_1, x_2, \ldots, x_n \mid \alpha, \beta) = \prod_i \frac{1}{\beta\sqrt{2\pi}} \exp\left(-\frac{(x_i - \alpha)^2}{2\beta^2} \right)$$

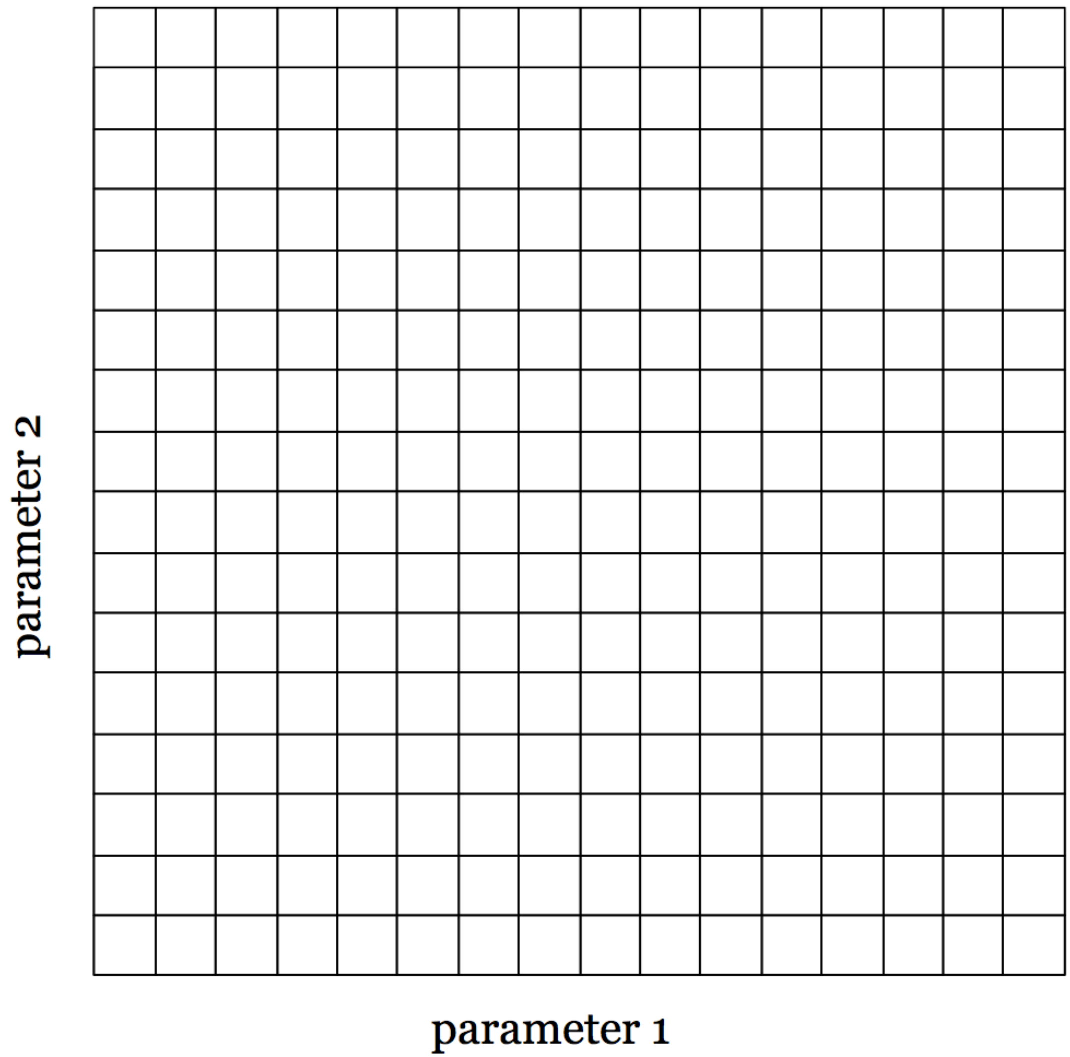
$$logL = \sum_{i=1}^{n} [-\log\beta - \frac{1}{2}\log(2\pi) - \frac{(x_i - \alpha)^2}{2\beta^2}]$$

$$\frac{\partial\, logL}{\partial\alpha} = 0 \Rightarrow \alpha = \frac{\sum_{i=1}^{n} x_i}{n}$$



暴力全局最佳化：Grid Search

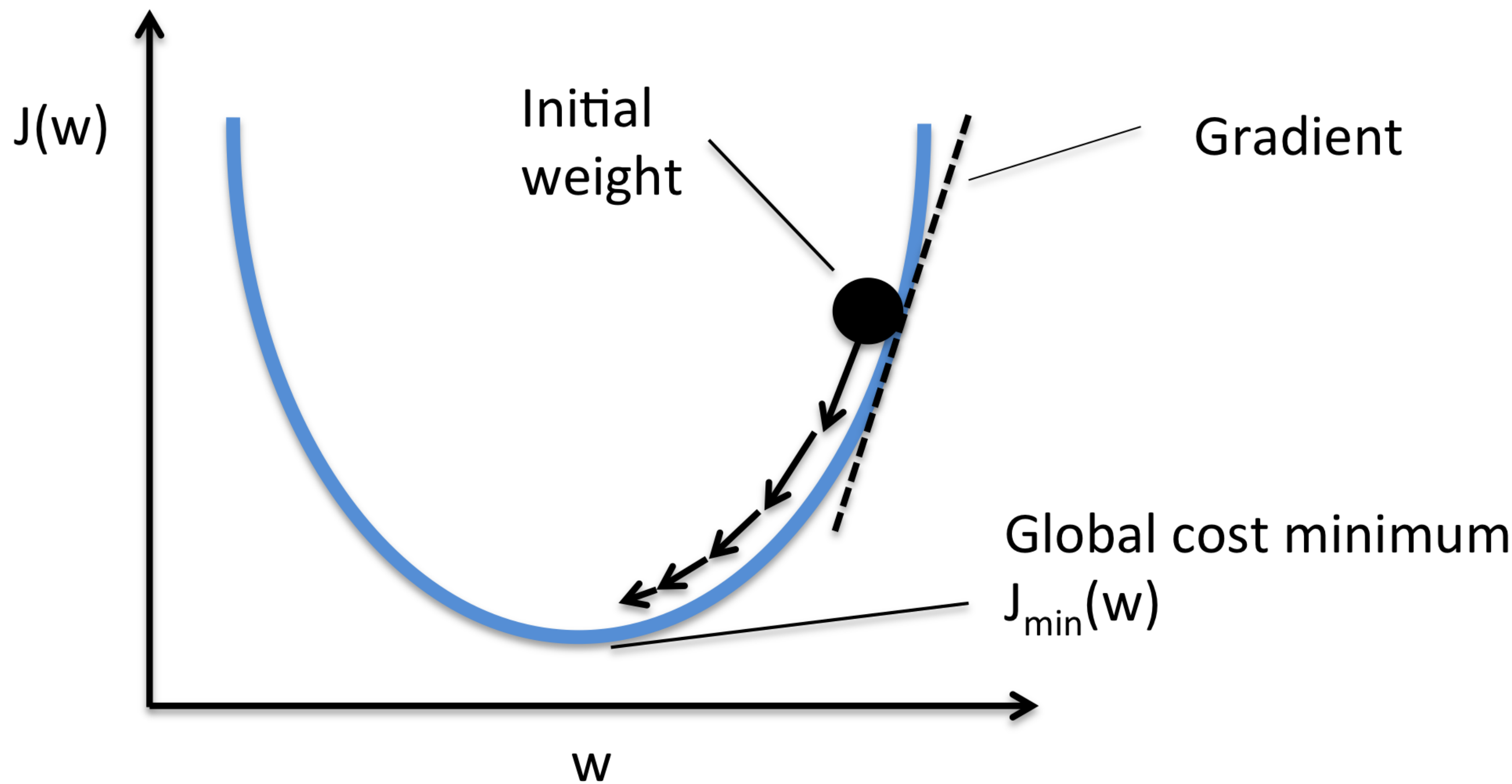
對參數空間離散化後，有 N 個參數就用 N 個套疊的迴圈來測試所有參數組合所產生的函數值



局部最佳化：Gradient Descent

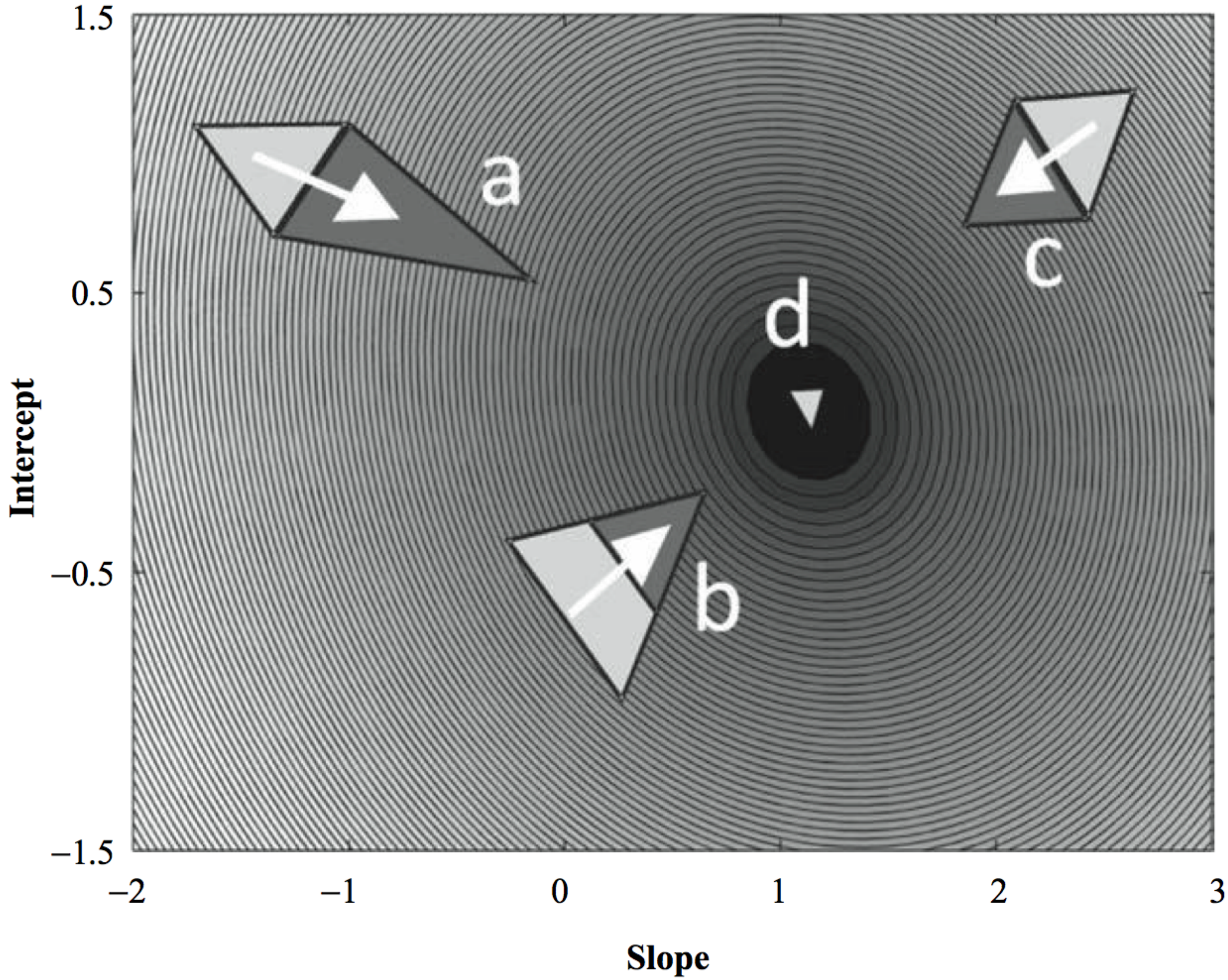
本質上就是有可能卡在區域極值的 greedy algorithm

$$x_{t+1} = x_t - \gamma_t \nabla f(x_t) \Rightarrow f(x_0) \geq f(x_1) \geq f(x_2) \geq \ldots$$



最佳化：Nelder-Mead Simplex

就是常在 Matlab 中用的 fminsearch



Overview

Parameter Estimation

Model Comparison

Reaction Time

Choice Accuracy

模型建構的三個層次

Data: A(53) > B (29)

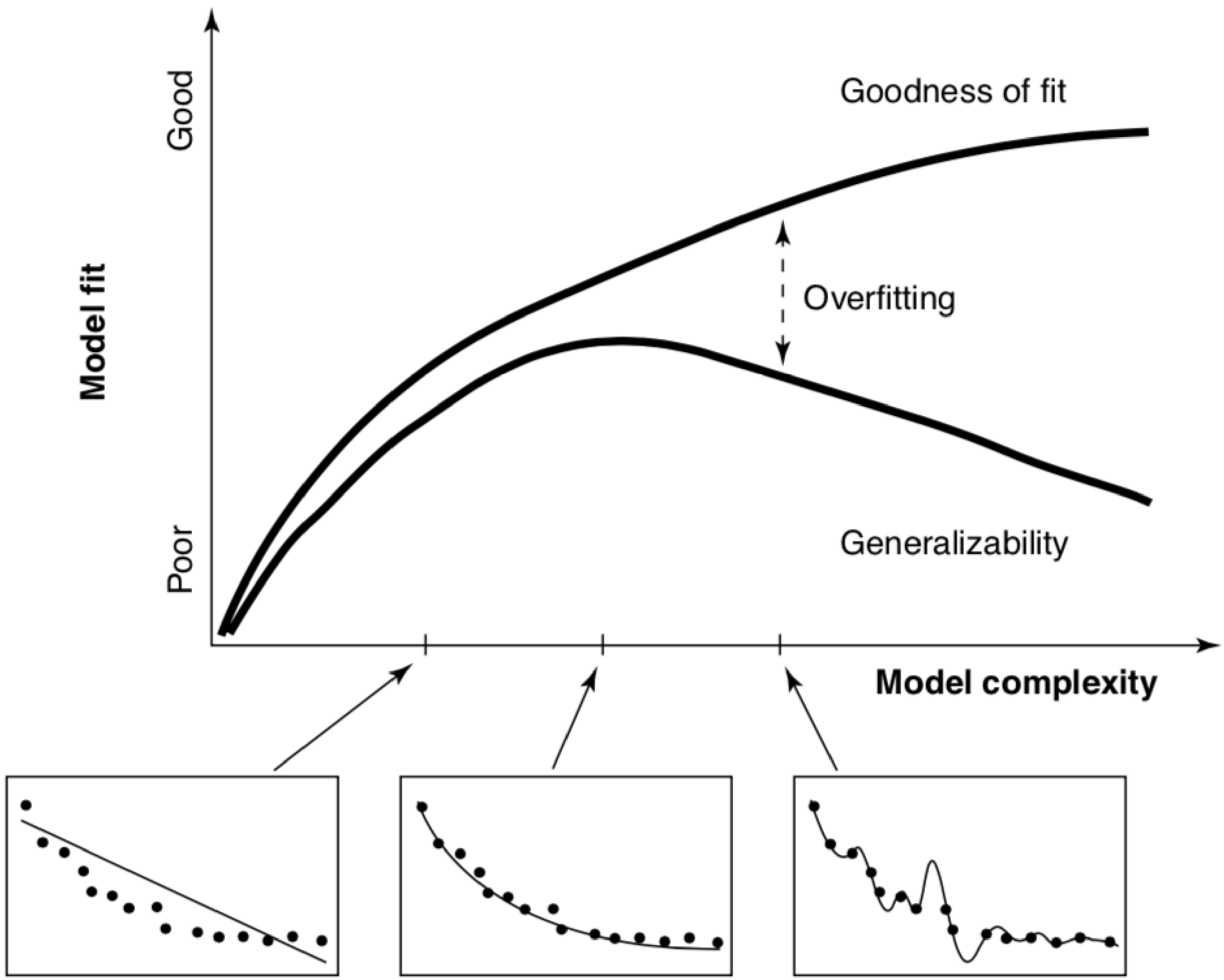
機制可行性的證明 (Existence Proof)
Model 1: A > B

定性比較 (Qualitative Comparison)
Model 1: A > B ← Good!
Model 2: A < B ← Bad!

定量比較 (Quantitative Comparison)
Model 1: A (53) > B (27) ← Good?
Model 2: A (55) > B (29) ← Bad?

Occam's Razor

選擇最簡單但能解釋最多資料的模型



模型好壞有不同的指標 (1/2)

Selection method	Criterion equation	Dimensions of complexity considered
Root Mean Squared Error	$RMSE = (SSE/N)^{1/2}$	None
Percent Variance Accounted For	$PVAF=100(1-SSE/SST)$	None
Akaike Information Criterion	$AIC = -2\ ln(f(y\|\theta_0)) + 2k$	Number of parameters
Bayesian Information Criterion	$BIC = -2\ ln(f(y\|\theta_0)) + k \cdot ln(n)$	Number of parameters, sample size
Bayesian Model Selection	$BMS=-ln \int f(y\|\theta)\pi(\theta)d\theta$	Number of parameters, sample size, functional form
Minimum Description Length	$MDL=-ln\ (f(y\|\theta_0)) + (k/2)ln(n/2\pi)+ln \int \sqrt{\det(I(\theta))}d\theta$	Number of parameters, sample size, functional form

In the equations above, *y* denotes observed data, θ is the model's parameter, θ_0 is the parameter value that maximizes the likelihood function $f(y|\theta)$, *k* is the number of parameters, n is the sample size, N is the number of data points fitted, SSE is the minimized sum of the squared errors between observations and predictions, SST is the sum of the squares total, $\pi(\theta)$ is the parameter prior density, $I(\theta)$ is the Fisher information matrix in mathematical statistics [a], *det* denotes the determinant of a matrix, and *ln* denotes the natural logarithm of base e.

AIC 在資料點 N>7 時對參數數目 K 懲罰比 BIC 小

Model	K	-2lnL	AIC	BIC
Wiener	10	8084.68	8104.68	**8162.10**
OU	10	8091.43	8111.43	8168.86
Rect acc	12	8167.69	8191.69	8260.60
Exp acc	12	8079.09	**8103.09**	8172.00

BIC 明顯偏好較簡單 (即小 K) 的模型

模型好壞有不同的指標 (2/2)

Selection method	Model fitted	Model the data were generated from		
		M_1	M_2	M_3
PVAF	M_1	0	0	0
	M_2	38	97	30
	M_3	62	3	70
AIC	M_1	79	0	0
	M_2	9	97	30
	M_3	12	3	70
MDL	M_1	86	0	0
	M_2	1	92	8
	M_3	13	8	92

Models M_1, M_2 and M_3 were defined as follows: M_1: $y = (1+t)^{-a}$; M_2: $y = (b+t)^{-a}$; M_3: $y = (1+bt)^{-a}$. In the model equations, a, b and c are parameters that were adjusted to fit each model to the data, which were generated using the same five points, t = 0.1, 2.1, 4.1 6.1, 8.1. Each sample of five observations was sampled from the binomially probability distribution of size n = 50. One thousand samples were generated from each model and served as the data to fit. Each selection method was then used to determine which model generated each of the samples.
The percentage of time each model was chosen for each dataset is shown.

Overview

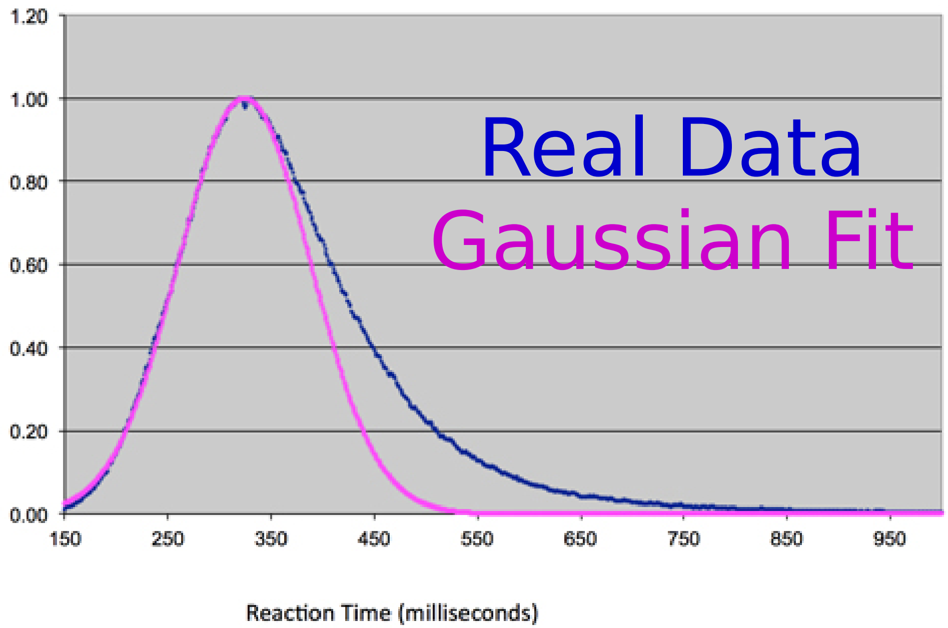
Parameter Estimation

Model Comparison

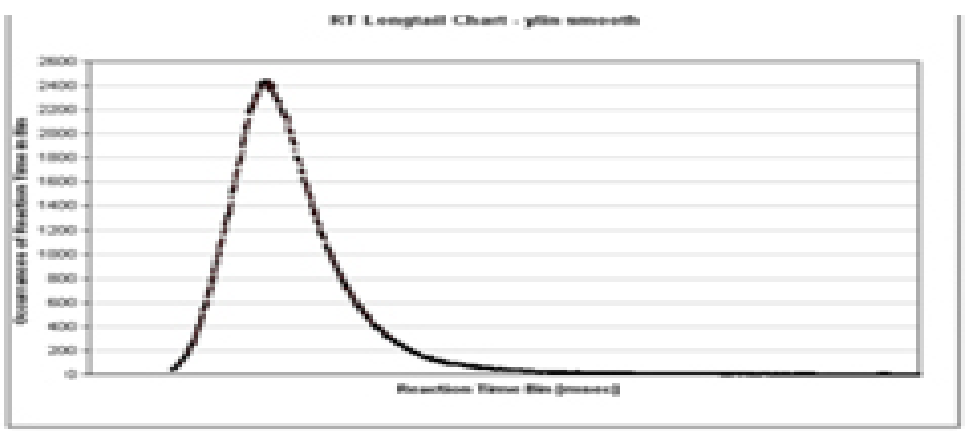
Reaction Time

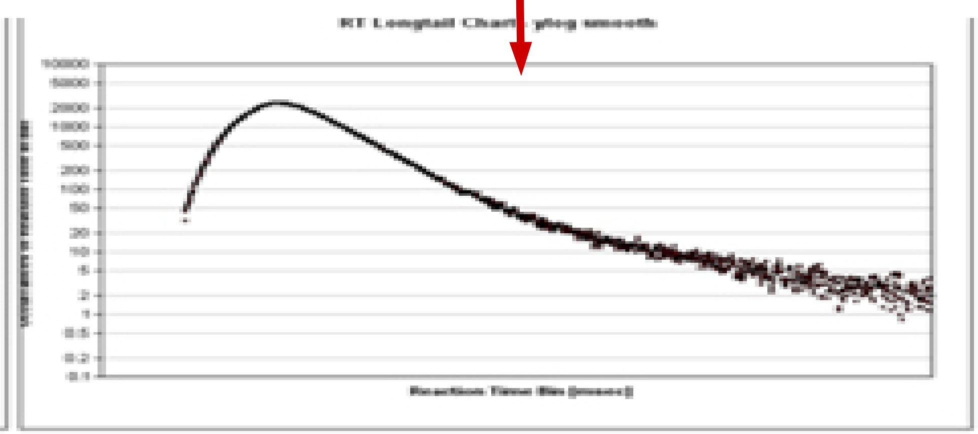
Choice Accuracy

Data: Reaction Times



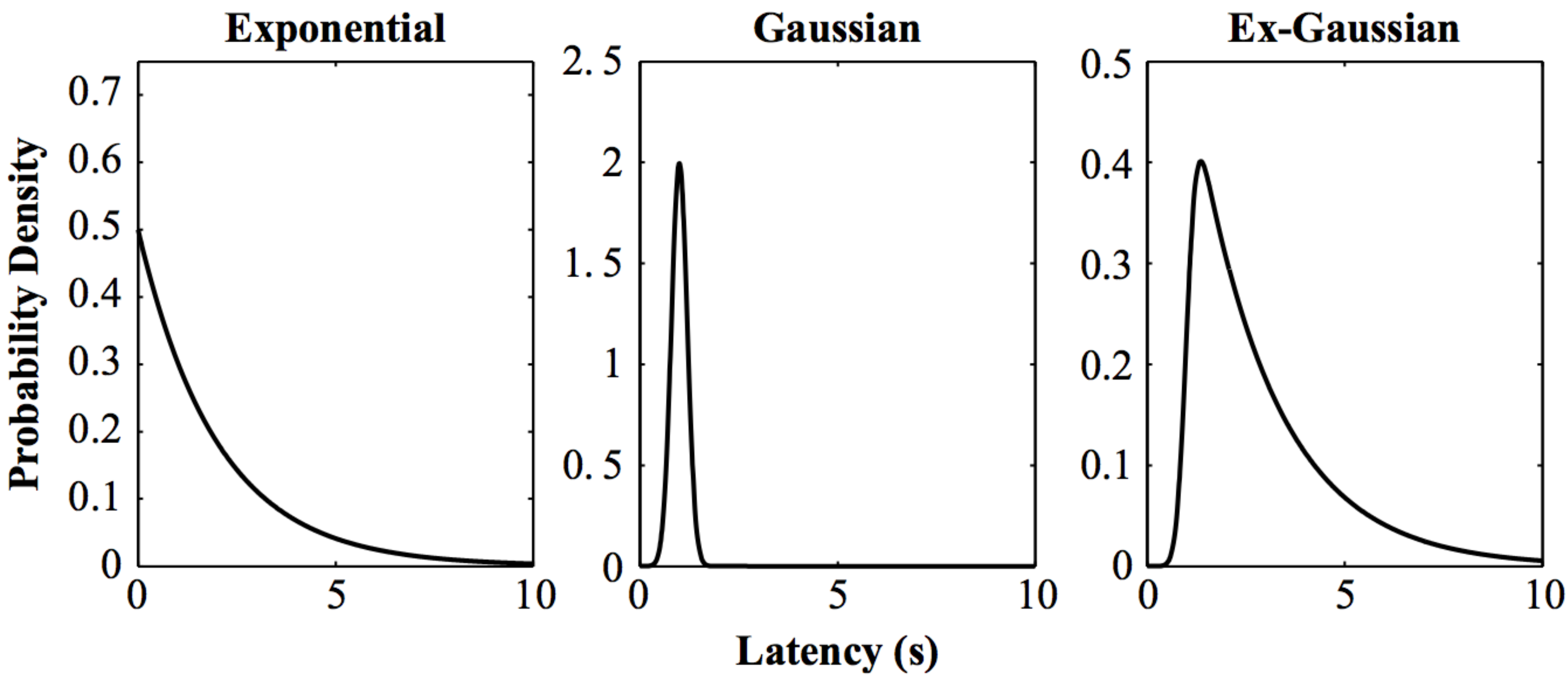
對資料取 log 後發現
前段是二次曲線
後段是一次直線





WHAT model: <u>ex-Gaussian</u>

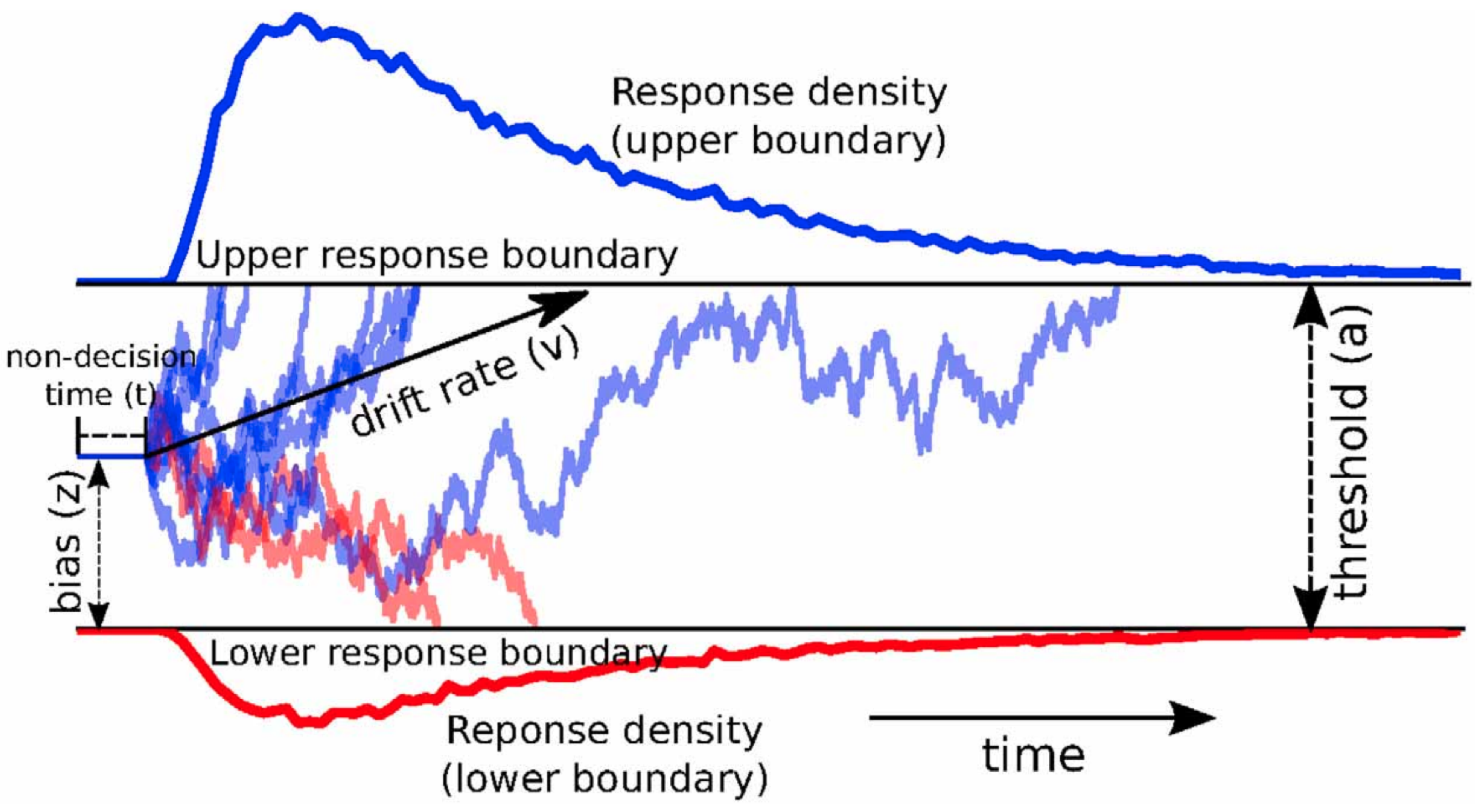
Gaussian= 認知作業時間
Exponential= 決策時間



主要是用 exponential 來描述長尾分布

HOW model: Diffusion

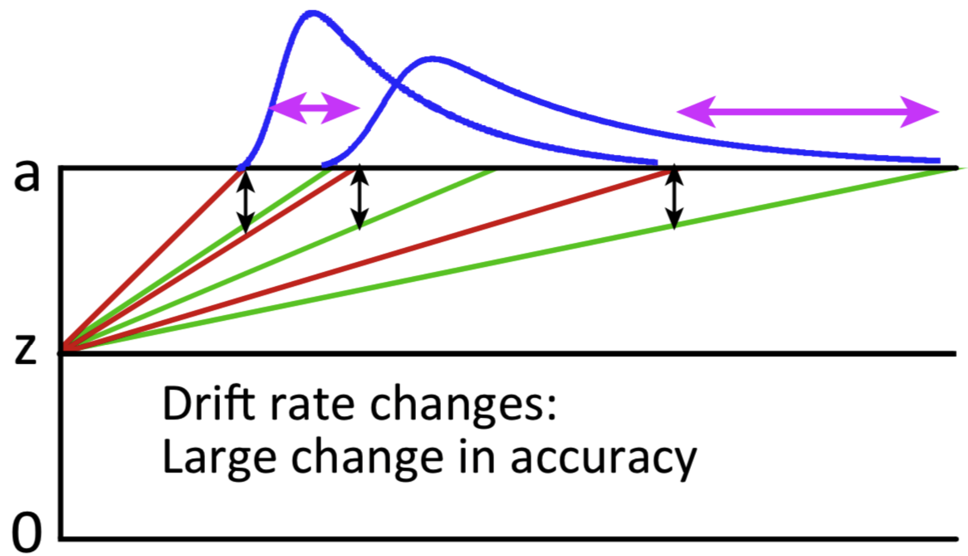
RT=Asymmetric random walk 多久會到終點

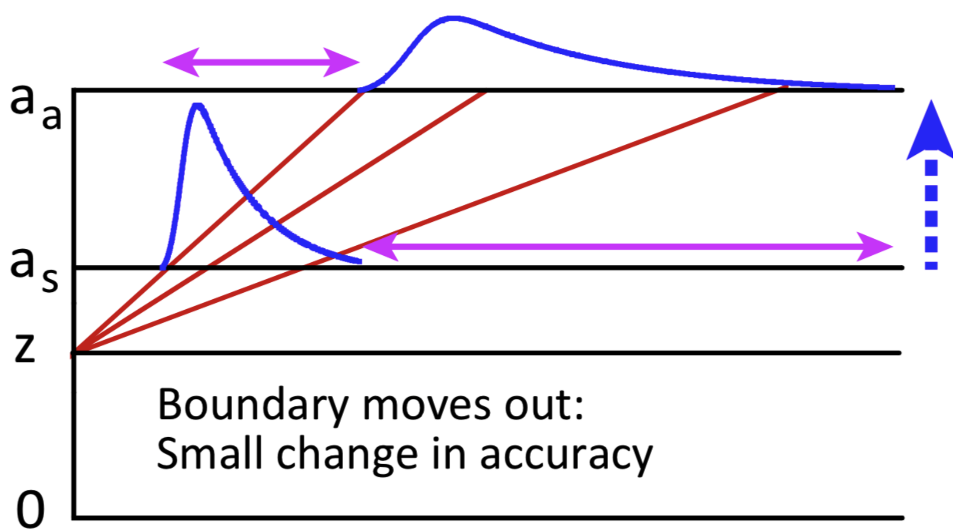


可有另外完全獨立的 racer/evidence accumulator

模型參數如何影響反應時間

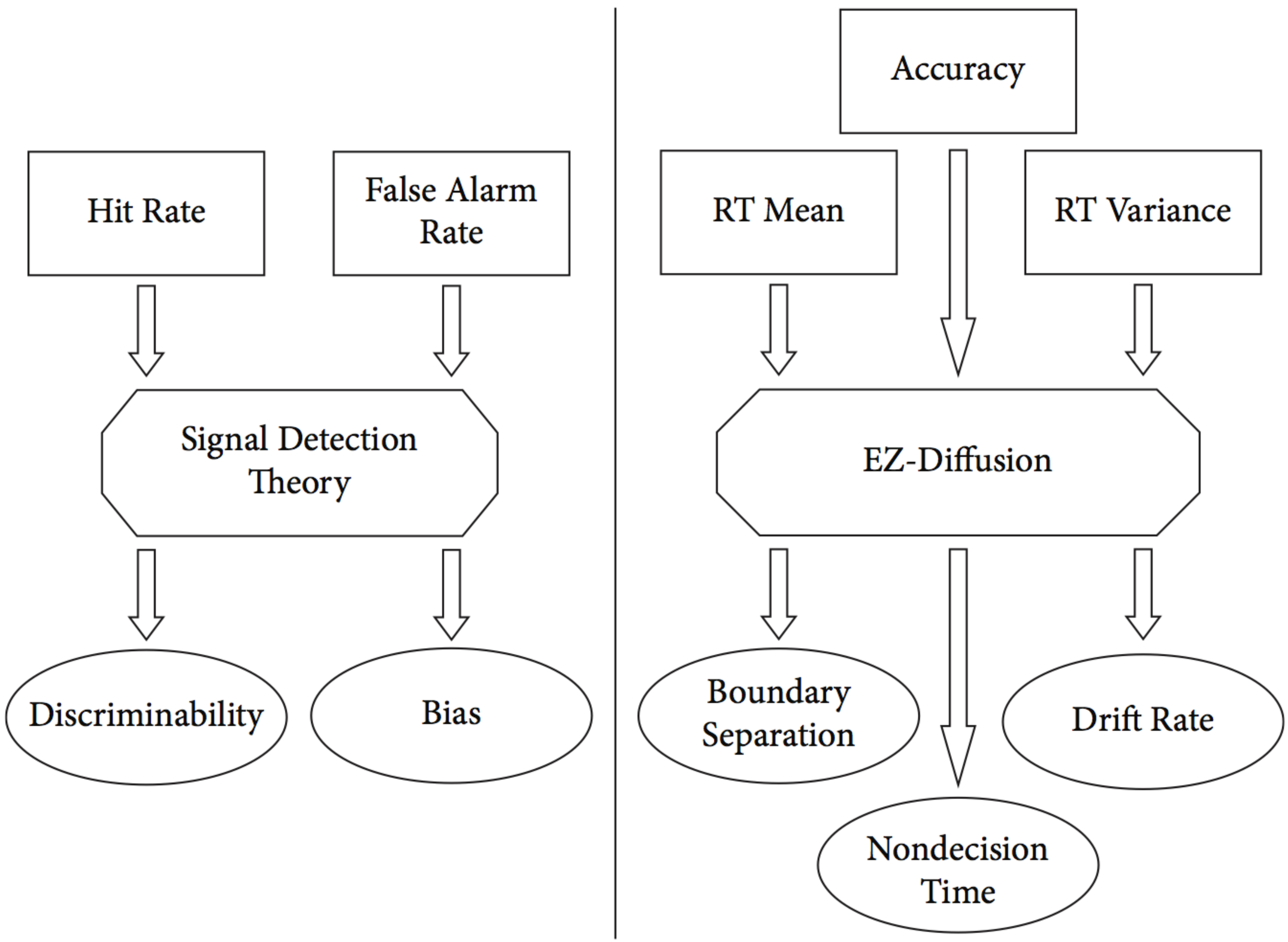
注意 RT 的長尾





3 個參數的 EZ-Diffusion Model

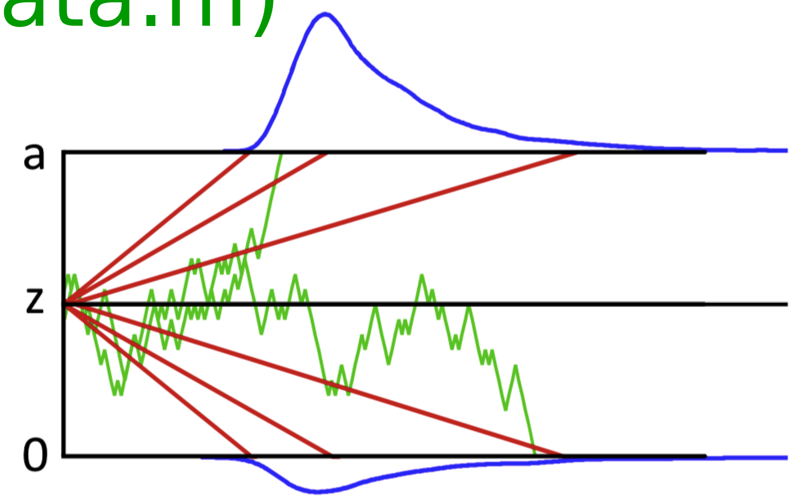
是簡化有 7-9 個參數的 Ratcliff Diffusion Model 來的



其實沒有很 EZ

Model→Data (EZdata.m)

$$P_c = \frac{1}{1 + \exp\left(-av/s^2\right)}$$



$$MRT = MDT + T_{er} \qquad MDT = \left(\frac{a}{2v}\right)\frac{1-\exp(y)}{1+\exp(y)} \qquad y = -va/s^2$$

$$VRT = z_* \left\{\tanh\left(z_* v_*\right) - z_* v_* \left[\operatorname{sech}\left(z_* v_*\right)\right]^2\right\} \Big/ v_*^3 \qquad v_* = v/s \text{ and } z_* = z/s$$

Data→Model (dataEZ.m)

Diffusion Model 的應用

Perceptual Decision (A vs. B)

刺激清晰度愈高 drift rate 愈高
個體速度與準度的取捨反應在 boundary separation

Recognition Memory (new vs. old)

Drift rate 不隨年紀增加
Nondecision time 隨年紀增加
Boundary separation 隨年紀增加

Lexical Decision (word vs. non-word)

高頻真字有較高且正方向的 drift rate
混亂的非字有高且反方向的 drift rate
似真的非字有低且反方向的 drift rate

Overview

Parameter Estimation

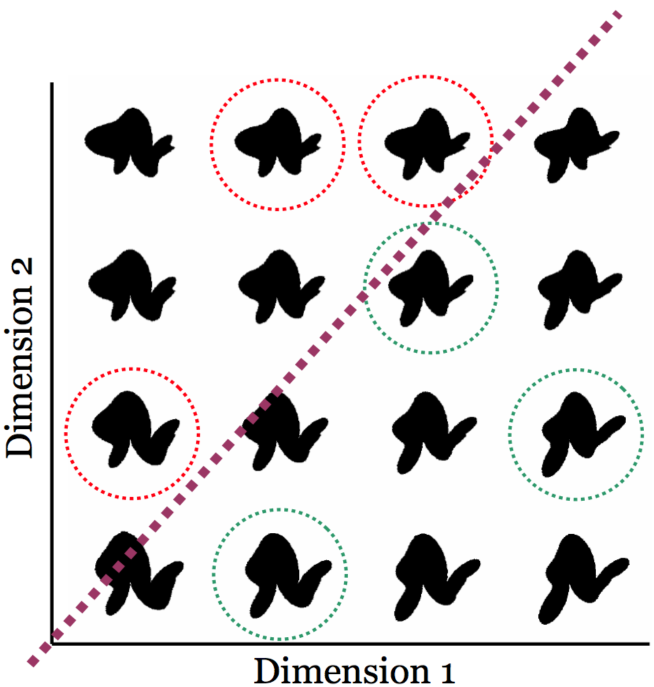
Model Comparison

Reaction Time

Choice Accuracy

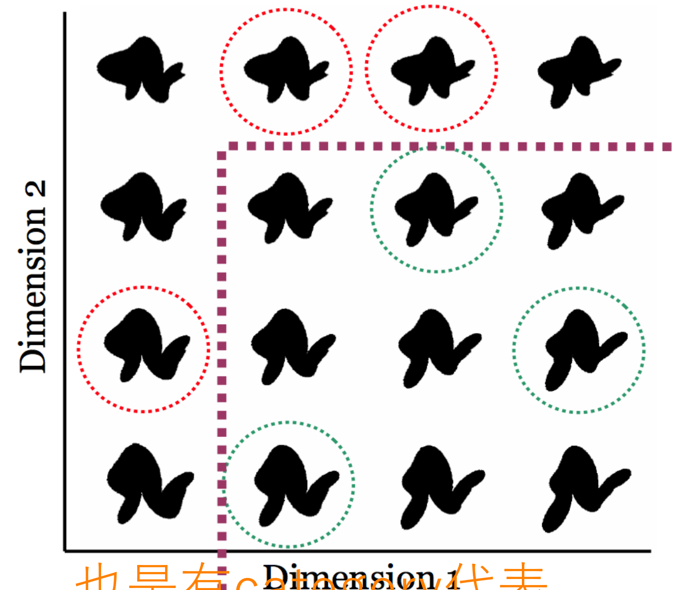
Categorization: Theories

y = a + bx

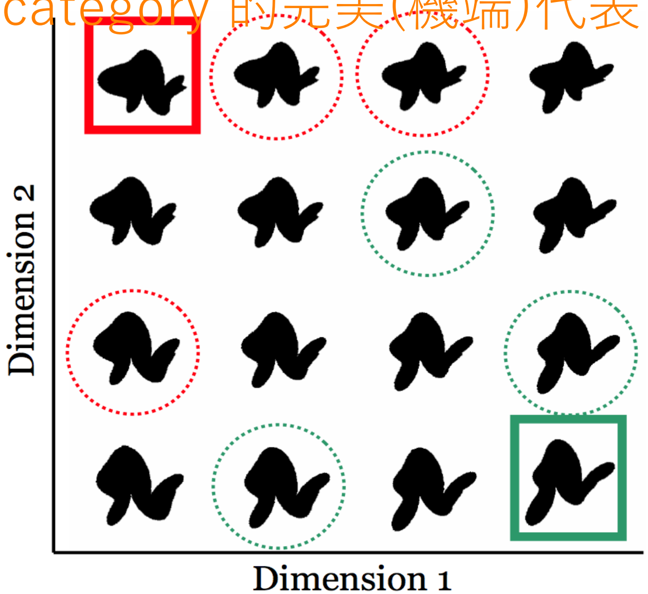


x \in A, if x < 1 & y > 3
x \in B, ow

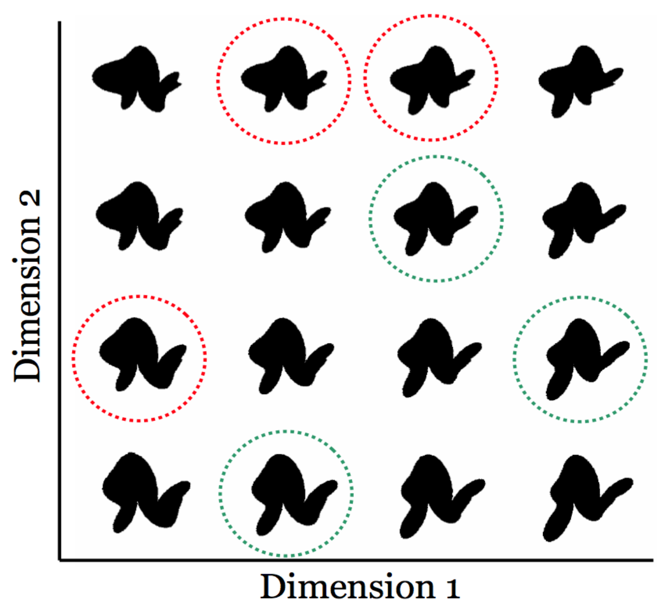
由左上到右下為：
Decision Boundaries
Rules
Ideals
Exemplars
Prototypes



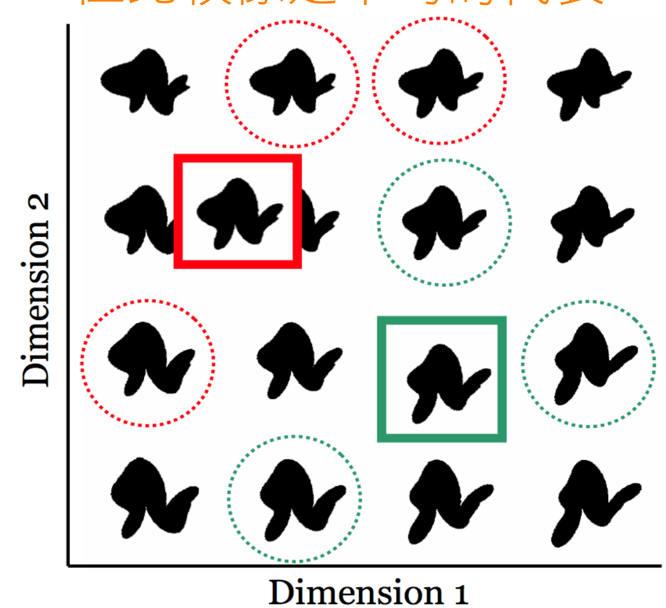
心中有完人存在
category 的完美(機端)代表



根據過去經驗的案例 (似 kNN)

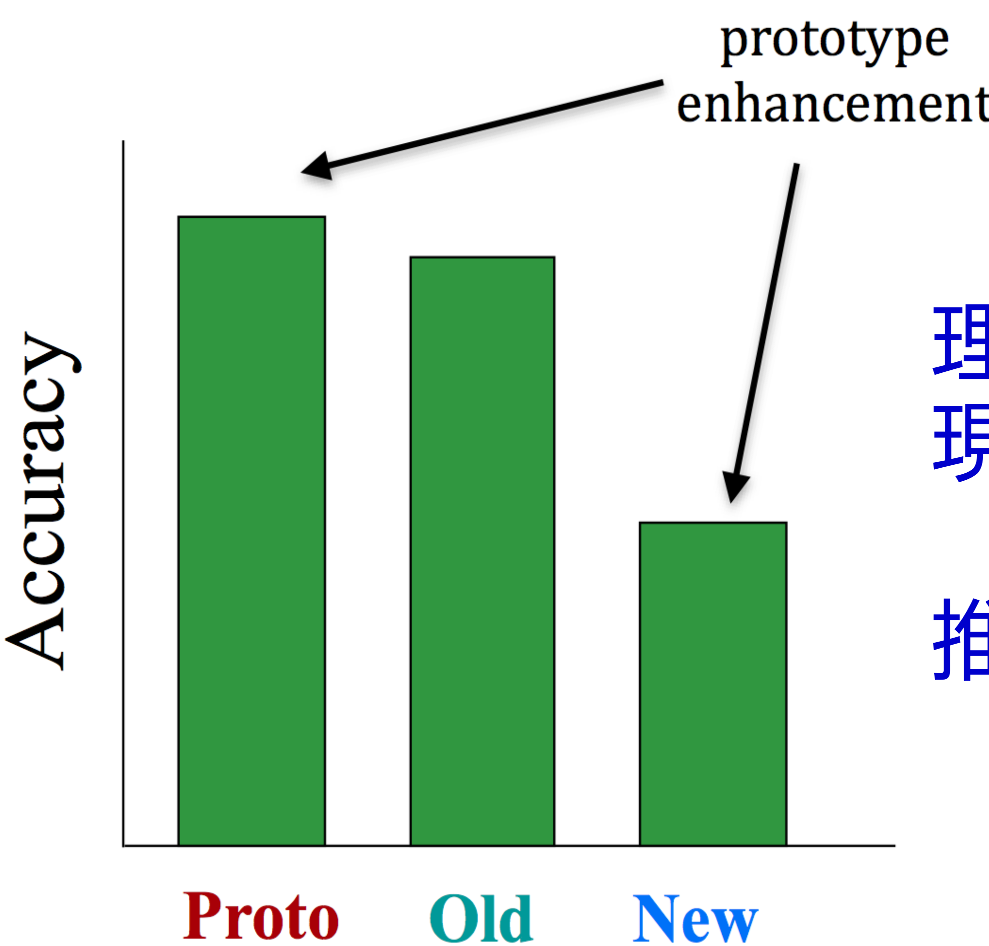


也是有category代表
但比較像是平均的代表



Categorization: Results

過去沒看過 prototypes 但分類很正確



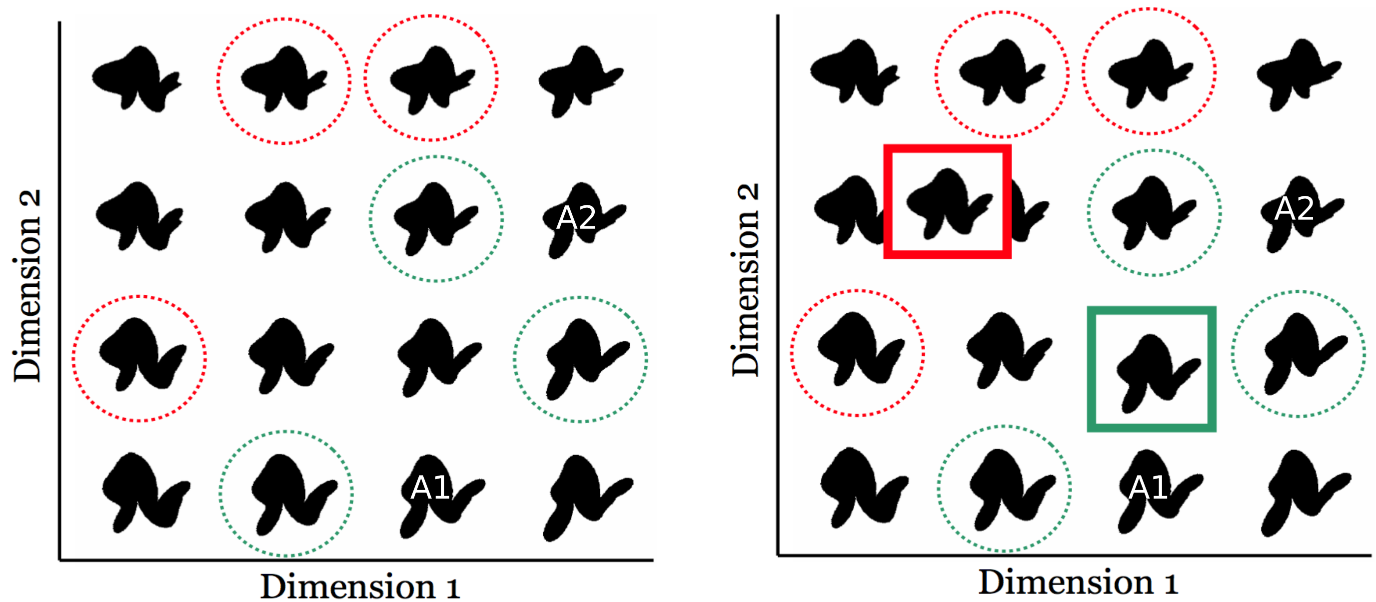
只有看到喜歡的理論產生這樣的結果
但沒有檢驗其他理論會不會產生一樣的結果

理論：如果我中樂透就辦休學
現象：等下我要去辦休學
⇒
推論：我中樂透了？
（我其實可能是失戀了）

所以 Prototype Theory 是對的？

Categorization: Model Comparison

可做質性比較



以分類正確率或是否更容易被視為綠色類別而言：

該點劃出個 neighborhood

Exemplar theory: A2>A1

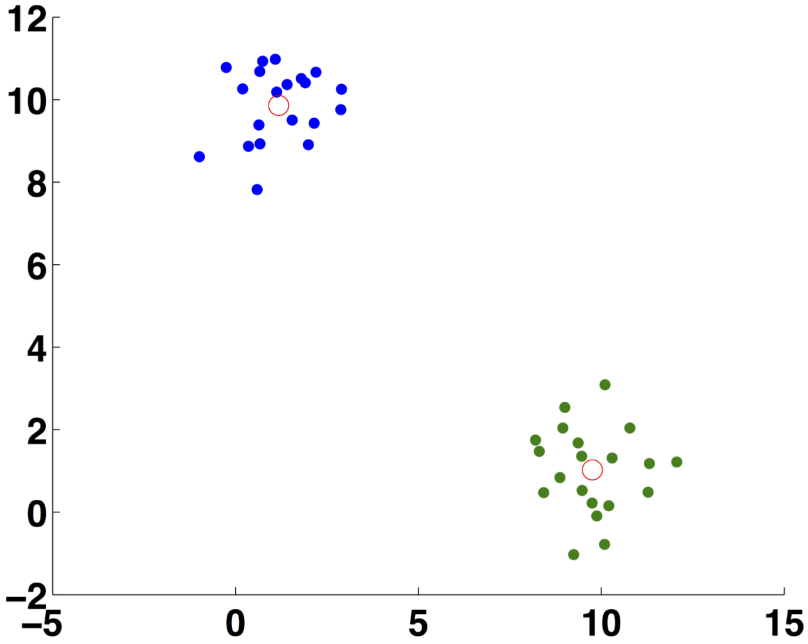
Prototype theory: A1>A2

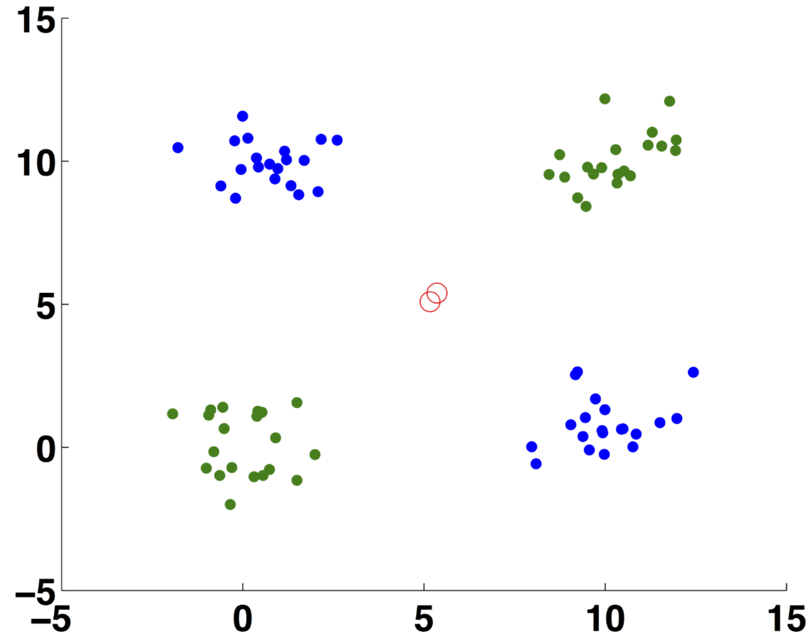
Fact: A2>A1

所以 Exemplar Theory 才是對的！

Categorization: Model Comparison

可做對 XOR problem 做質性比較





Exemplar theory: accuracy∼100%

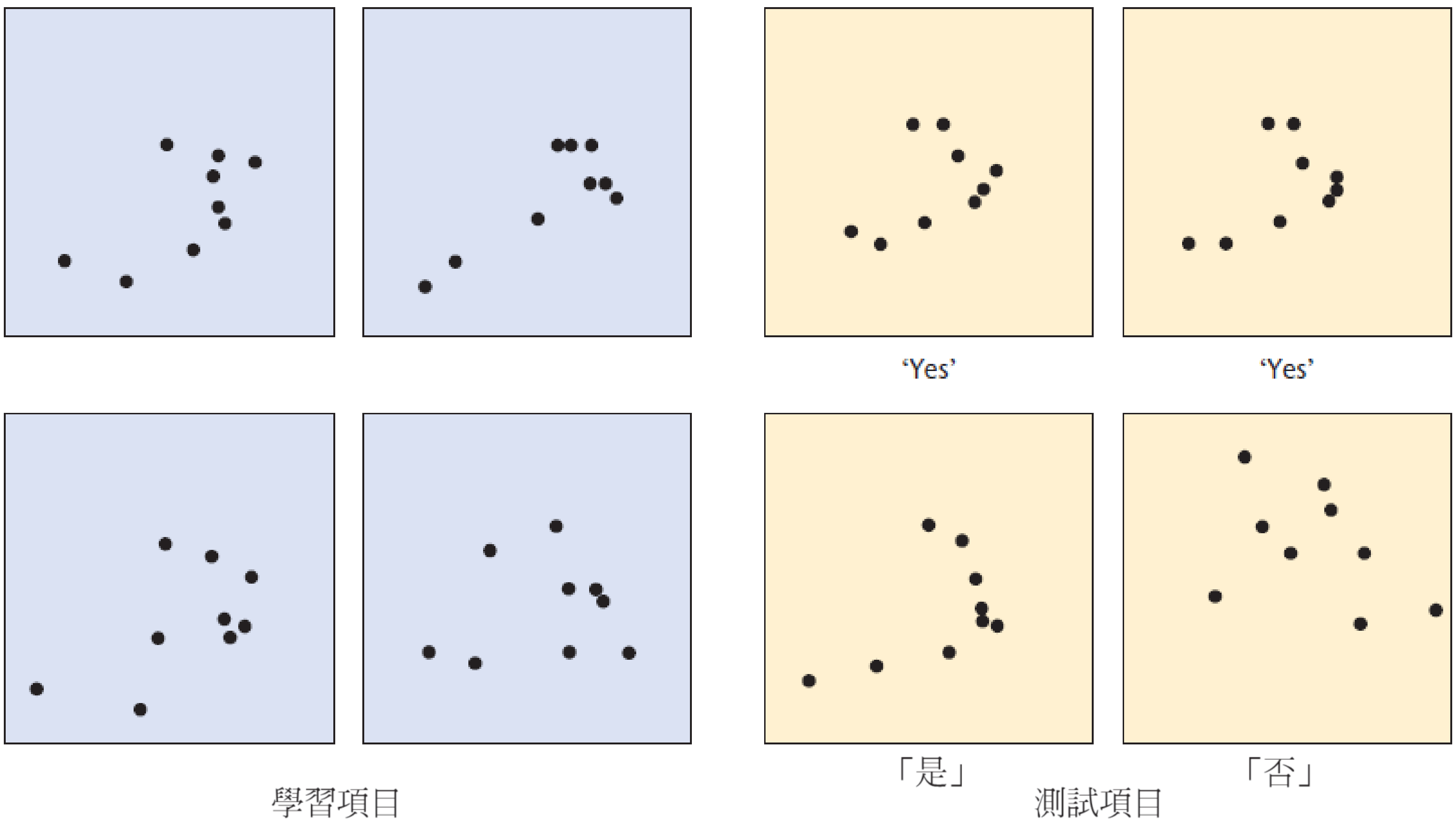
Prototype theory: accuracy∼0%

Fact: accuracy≫0

所以 Exemplar Theory 才是對的！

Dot-pattern Task: Stimuli

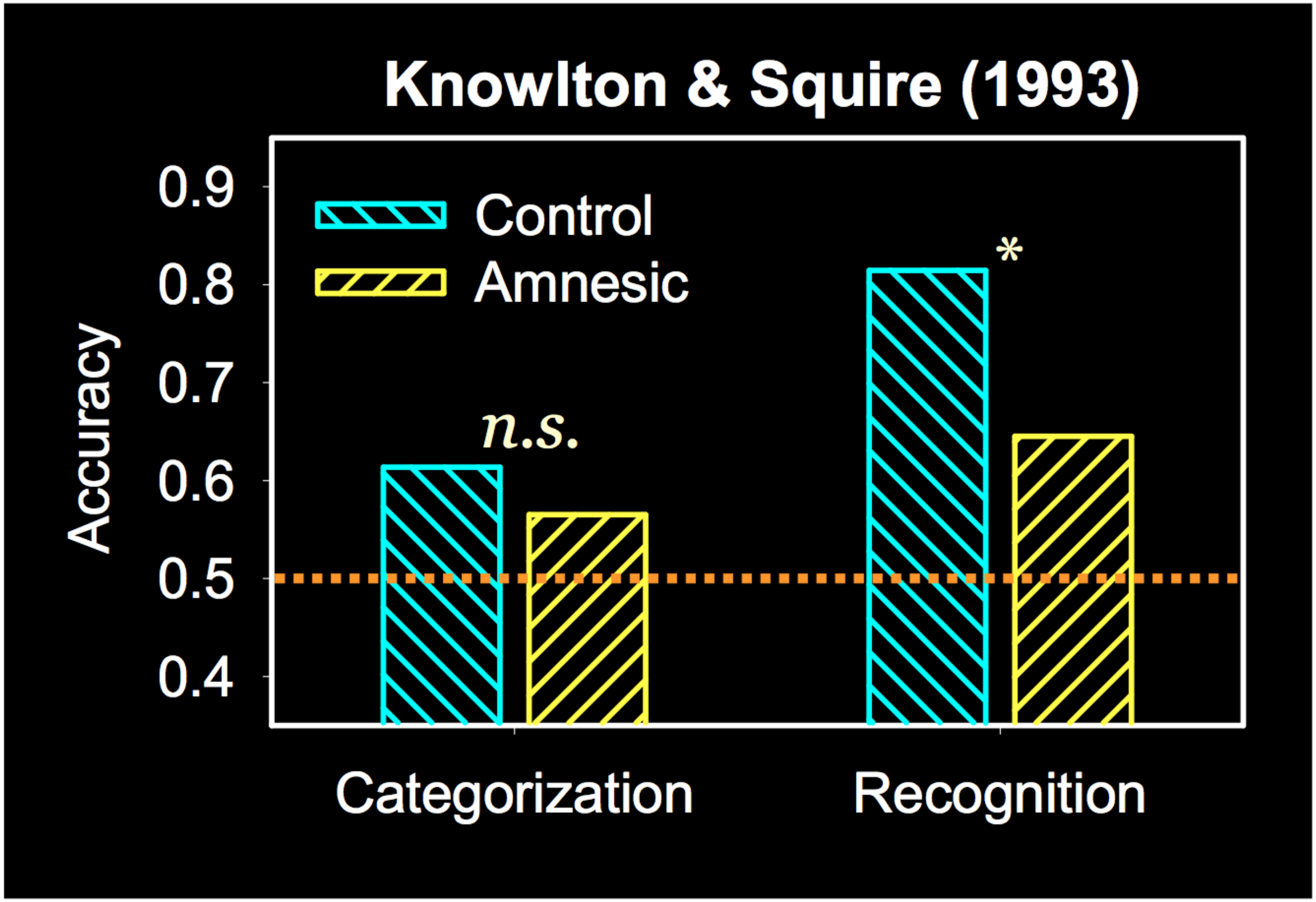
Recognition: 之前有沒有出現過？
Categorization: 和之前的是同類的嗎？
Identification: 這是之前的哪一個 ？



用來研究失憶症者之分類行為的點圖範例

Dot-pattern Task: Data

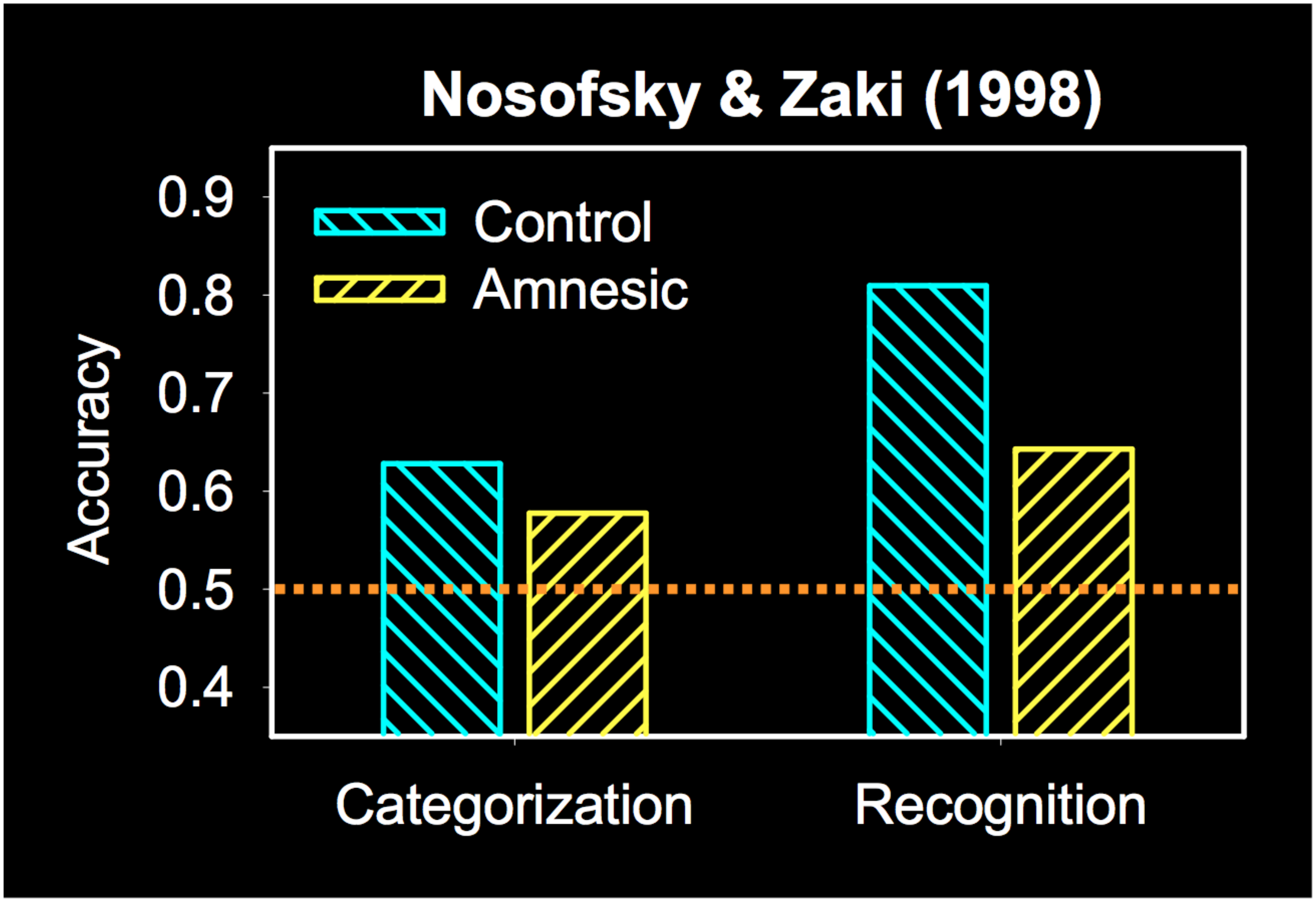
失憶者再認能力嚴重受損但分類能力仍與普通人差不多



所以再認和分類是兩個獨立的系統？

Dot-pattern Task: Model

Exemplar model 就可模擬所有分類和再認現象



兩類人的差別在記憶清晰度 / 敏感度這個模型參數

Exemplar Model: GCM

都是與記憶中的各種i物體

Identification: 這是之前的哪一個？

$$P(R_j|S_i) = \frac{b_j \eta_{ij}}{\sum_{k=1}^{n} b_k \eta_{ik}} \qquad \eta_{ij} = e^{-d_{ij}}$$

Luce's Choice Axiom

eta_ij: 表 ij 相似度 (d:距離)

Categorization: 和之前的是同類的嗎？

$$P(R_J|S_i) = \frac{b_J \sum_{j \in C_J} \eta_{ij}}{\sum_{K=1}^{m} (b_K \sum_{k \in C_K} \eta_{ik})}$$

大寫J表第幾類

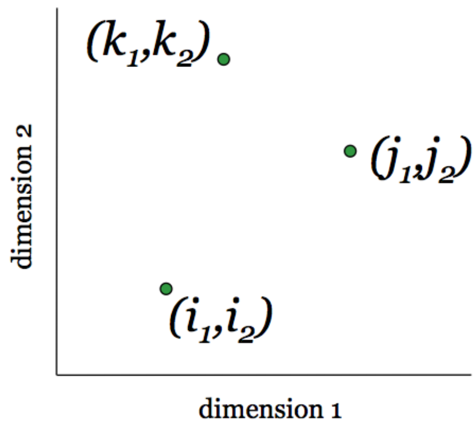
Recognition: 之前有沒有出現過？

$$F_i = \sum_{j \in C1} s_{ij} + \sum_{k \in C2} s_{ik}$$

算所有的 similarity 總和，看有沒有超過 threshold

3 Model Choices

1.



$$d_{ij} = \sum_{m=1}^{M} \left((i_m - j_m)^r \right)^{1/r}$$

2.

$$s_{ij} = \frac{1}{d_{ij}}$$

$$s_{ij} = \frac{1}{1 + d_{ij}}$$

$$* \, s_{ij} = e^{-d_{ij}} = \exp(-d_{ij})$$

$$s_{ij} = e^{-d_{ij}^2} = \exp(-d_{ij}^2)$$

Shepard's law of generalization

3. $$E_{A|i} = \sum_{j \in A} s_{ij}$$

- similarity to closest exemplar (nearest neighbor)

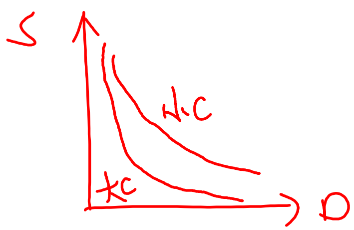
- average similarity to exemplars

- summed similarity to exemplars

Simulation of dot-pattern results

Nosofsky & Zeki (1998): 正常人 c 比較大

$$s(i,\mathrm{a}) = \exp(-c \bullet D)$$



大 c 相當於 kNN中的小 k
discrimination 愈好
generalization 愈差

$$P(\mathrm{C}) = [40^* s(i,h)]/[40^* s(i,h) + k_{\mathrm{C}}]$$ 是否和看過的同一類？

40項item

該類的平均 similarity

$P(C) \begin{matrix} > \\ = \\ < \end{matrix} 1/2$ if $40^*s \begin{matrix} > k_c \\ = k_c \\ < k_c \end{matrix}$

似threshold

k_C/k_R 決定

4 items

$$P_{\mathrm{old}}(\mathrm{R}) = [\delta + 4^* s(i,r)]/[\delta + 4^* s(i,r) + k_{\mathrm{R}}]$$

sim. sum 多大時 P≈0.5

δ=self-similarity

有無看過

注意 :N&Z (1998) 假設失憶者其實全部刺激都有記住

Game Over