

# 因果革命：人工智慧的大未來

## 第一章介紹

負責同學：謝承孝 104072138

指導教授：余朝恩 老師

1. 大家好，我是負責介紹第一章的同學，謝承孝。先跟大家講解一下，這份檔案的前半部是每張投影片的文字描述，每段文字最前面會有編號，對應到檔案後半部投影片右下角的編號。大家可以同時開兩個檔案，前後半部對照閱讀，較為方便。

在講解的部分，會先介紹作者的觀點，接著在下一段表達自身的觀點和理解，我的理解不一定正確。當然，有任何疑問也歡迎跟我討論。

2. 首先，這章節分為三個小主題，分別是：

(1) 因果的三個層級 (2) 迷你圖靈測驗 (3) 機率與因果

4. (1) 關聯

階梯的第一層需要依據被動的觀察結果進行預測，也就是蒐集和分析資料。典型關聯方法有：相關性(Correlation)或回歸(regression)，也就是找出與一組資料點擬合的一條線。在因果階梯第一層運作的系統，會缺乏彈性和適應性，因為只能依照資料進行分析，若沒有資料，則無法分析。

作者認為現在熱門的深度學習、機器學習，是打造具備超強能力但不具智慧的機器，依舊停留在第一層級；但是強 AI 的目標則是打造具備人類智慧的機器，能和人類交談、指導人類。

- (2) 介入

作者認為著手改變世界，才能登上因果的第二個層級，也就是在**細心控制的條件下進行實驗**。首先要釐清事件的**因果關係**，建立因果模型，才有辦法讓 AI 回答關於介入的問題，否則其無法以被動取得的資料，來回答問題。

如果覺得有點抽象的話，我舉個例子。這邊大家可以簡單回想一下，我們在個體經濟學內常遇到的問題，假設有 A 和 B 兩項產品，當 A 價格提高時，B 產品的銷量會如何改變？這邊會有兩種效果：替代和互補，首先我們必須先進行實驗，確定 A, B 兩產品間價格和銷量的因果關係（模型），才能進行推測，甚至運算。

- (3) 反事實推論

作者認為，資料就是事實，他無法告訴我們已知事實消失後，反事

實或想像世界可能發生的狀況；但是人類能穩定且重複執行這類尋求解釋的推論。例如：如果某個歷史事情不同的話，會怎麼樣？

作者引用且認同哈拉瑞的觀點，認為描繪虛構生物體現了人類的新能力，並把這種能力稱為「認知革命」。獅人雕像體現人類想像原本不存在事物的新能力，它是各種哲學理論、科學發現以及科技創新的前兆。



圖 1.3 史塔德洞穴的獅人。它是目前已知最古老的虛構生物（半獅半人），代表人類剛萌芽的認知能力，也就是推想反事實的能力。

資料來源：德國巴登符騰堡邦文化遺產局／烏爾姆博物館，Yvonne Mühleis 攝影。

其實我認為作者在這邊沒有闡述的很清楚，有些第三層級的舉例還停留在第二層級。以我的理解，在確認第二層級的因果關係正確後，開始思考虛構（與事實不同且不可逆）的事情發生，會產生什麼結果，此類的問題沒辦法透過實驗進行。舉個有名的**祖父悖論**當例子：假如我（有能力）回去殺掉我的祖父，那我就不會出生；如果我不出生，那我就不會殺掉我的祖父，如此一來我就會被生下來……無限循環。

6. 關於圖靈的模仿遊戲（the imitation game），可以參考泛科學的文章<sup>1</sup>，建議大家先簡單瞭解背景知識。**另外推薦大家，電影模仿遊戲超好看！！**

作者把「**機器（和人）應該如何呈現因果知識，才能像三歲小孩一樣快速取得需要的資訊，正確回答問題，並且毫不費力地實行？事實上，這就是本書要探討的主要問題。**」這個問題，稱為「**迷你圖靈測驗**」。這個測驗的構想是，選擇一個簡短的故事，以編碼輸入機器，測試機器是否能正確回答人類能回答的因果問題。作者以通過這個迷你測驗，為一生最重要的目標。迷你圖靈測驗有兩個條件：

- (1) 測驗僅限於因果推理，不包含視覺和自然語言等其他智慧
- (2) 允許參與者以任意方式編碼故事，讓機器不需由本身經驗擷取故事

在開始迷你圖靈測驗前，必須先處理**呈現問題**，才能處理**獲取問題**，而所謂的呈現，就是該如何**儲存資訊**供日後使用。作者認為，人工智慧對認知研究的重要貢獻是「先談呈現，再談獲取」的典範。追求良好的呈現方式可以讓我們了解，要如何從資料或程式設計者等各種來源獲取知識。

<sup>1</sup> 是人？還是機器？圖靈的模仿遊戲－《創新者們》。<https://pansci.asia/archives/85570>。

7. (這頁需要完整的前提和推論，所以跟書的內容會很像)

假設有個囚犯即將被行刑隊槍決。要完成這件事，一定會發生一連串事件。首先，法院必須下令執行。命令下達給隊長，隊長下令行刑隊的士兵(A和B)開槍。我們假設這兩人都遵守命令且槍法很好的槍手，因此他們一定會遵守命令開槍，而且只要其中有一人開槍，囚犯就會死亡。

〈圖 1.4〉呈現前面敘述的故事。每個未知數(CO、C、A、B、D)都是真/假變項。舉例來說，「D=真」代表囚犯死亡，「D=假」代表囚犯存活，依此類推。

這張圖可以回答階梯上每一層的因果問題。首先是關聯問題：如果囚犯死亡，是否表示法院已經下達命令？我們(或電腦)可以檢視圖形，追溯每個箭頭背後的規則，再運用標準邏輯，斷定如果隊長沒有下令那兩名士兵並不會開槍。同樣地，如果隊長沒有收到法院的命令，不會下令開槍，因此我們查詢後的答案是「是」。此外，假如發現A開了槍，可以得知B的哪些狀況？電腦沿箭頭回溯，斷定B一定也開了槍(如果隊長沒有下令，A不會開槍，所以B一定也不會開槍)。即使A不是B的原因，沒有因果關係(箭頭沒有從A指向B)，這點同樣為真。

〈圖 1.5〉是依據這個範例畫出的因果圖。這次介入必定會造成囚犯死亡。這個因果函數產生從A到D的箭頭。

在因果階梯向上爬一層，可以提出關於介入的問題。如果士兵A沒有等待隊長下令就自作主張開槍，那會怎麼樣？囚犯會死亡還是存活？其實這個問題有點矛盾。剛剛提過A一定要接受命令才會開槍，但現在又問如果他沒有接受命令就開槍的話會怎麼樣？如果你和電腦一樣只運用邏輯規則，會覺得這問題沒有意義，在這類狀況下會說「這樣無法計算」。但如果要通過迷你圖靈測試，電腦必須瞭解，A的決定不影響模型中非屬A開槍效果的變項。

最後，為了說明因果階梯的第三層，我們提出一個反事實問題。假設囚犯已經死亡。但如果當時A決定不開槍會怎麼樣？囚犯會活下來嗎？這個問題必須比較真實世界和A沒有開槍的虛構世界。在虛構世界中，畫向A的箭頭被擦去，讓A不用聽命於C。接著A設定為假，讓它的過往跟真實世界中相同，因此虛構世界看來是〈圖 1.6〉的樣子。要通過迷你圖靈測試，電腦必須斷定士兵B開槍會殺死囚犯，所以囚犯在虛構世界中也會死亡。

9. (這頁要先帶大家了解作者的想法，已經幫大家精簡過了)

作者認為，「機率提高應該可讓我們由因果階梯的第一層登上第二層」，這樣的直覺導致我們嘗試幾十年都宣告失敗。

$P(Y | X)$  表示式提出的條件機率，位於因果階梯的第一層，永遠無法靠本身回答第二層或第三層的查詢。試圖以看來比較簡單的第一層概念來「定義」因果關係，一定會失敗，因此這本書不會試圖定義因果關係，因為定義需要簡化，而簡化需要下到較低層。

哲學家努力地依據背景因素(干擾因子的另一種說法)來修正定義，產生準則  $P(Y | X, K=k) > P(Y | K=k)$ ，其中  $K$  是某些背景變項。但「哪些變項必須納入背景變項組  $K$ ，並當成修正依據？」一直沒有哲學家能提出可接受的通用答案，理由很簡單：共同創造也是因果概念，因此無法寫出機率公式。而作者認為，南西·卡特萊特(Nancy Cartwright)突破僵局，主張應該依據與結果「有因果相關」的因素來調整。作者說會在第四章提出  $k$  值的正確準則，但在前面不想破梗(第四章的朋朋加油~)。

如果  $P(Y | do(X)) > P(Y)$ ，則  $X$  是  $Y$  的原因(對  $X$  進行介入條件下，由機率提高來確定  $X$  是不是  $Y$  的原因)。這個定義能傳達「機率提高對因果的詮釋」，也可透過因果圖轉化成可以運算。有因果圖和資料，就能以代數清楚地回答這個問題，並由機率提高來確定  $X$  是不是  $Y$  的原因。

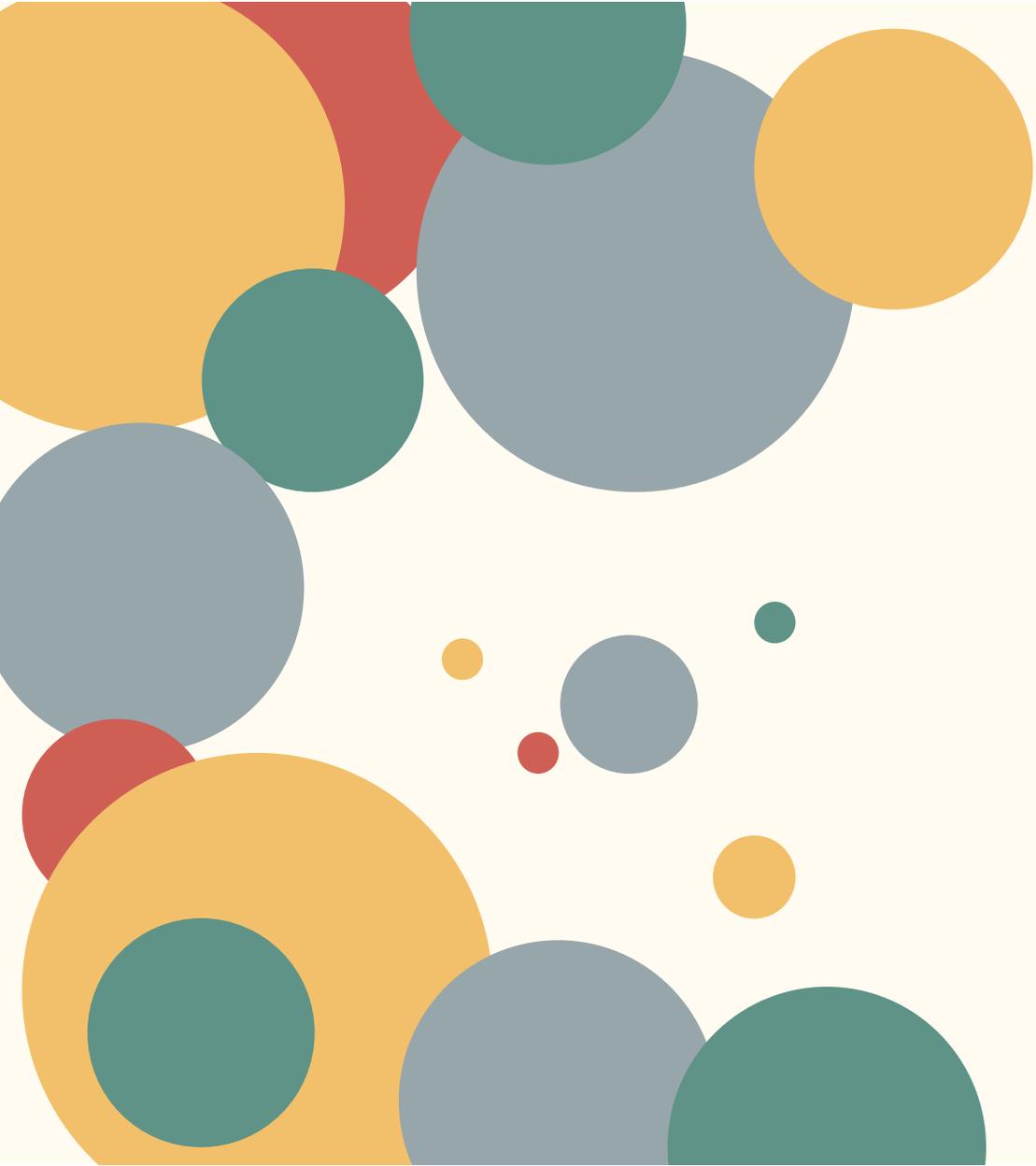
作者覺得主要重點是：機率代表我們心目中的靜態世界，因果性則告訴我們，當世界因為介入或想像而改變時，機率是否會改變以及會如何變化。

上面是作者的觀點，看得霧煞煞嗎？嗯沒關係我也是，作者的文章思路有點跳 tone，常寫不到重點，後面的章節應該也長這樣，大家保重。

經過我深思熟慮的理解後，認為作者想表達的是：機率確實可以讓我們通往第二甚至第三個層級，但是必須先確認條件機率間的因果關係是否正確，而所謂的因果就是  $P(Y | do(X)) > P(Y)$  的時候， $X$  為因， $Y$  為果。

以前面的士兵  $A$  為例：在圖 1.4 中，若  $A$  沒開槍時， $B$  也不會開槍，但是並不能寫成  $P(B|A)$ ，因為  $B$  會不會開槍是隊長  $C$  決定  $\Rightarrow P(B|C)=1$ 。

10. 謝謝大家！我的介紹到此結束，也感謝大家耐心的看完，有任何的疑惑，都可以來信詢問，也祝大家後面的章節順利。



# 因果革命： 人工智慧的大未來

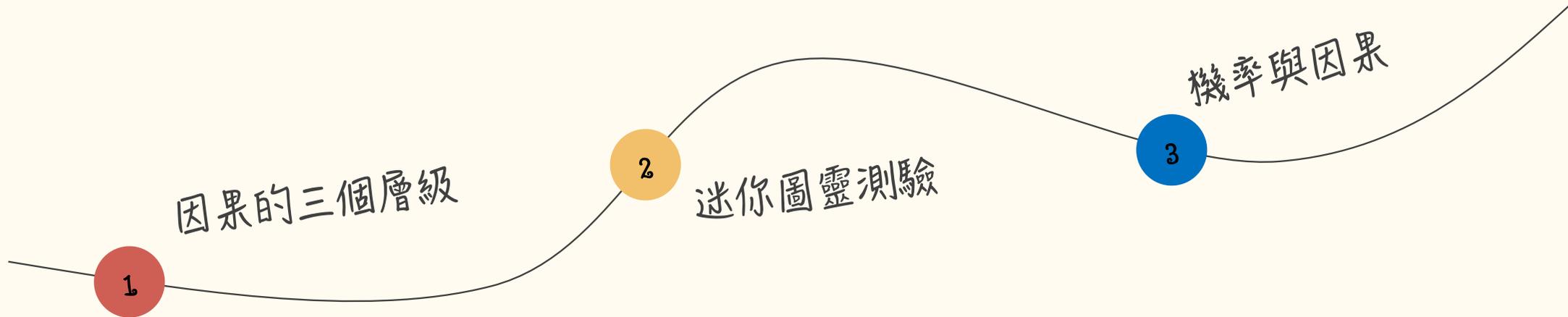
---

第一章節：因果階梯

負責同學：謝承孝 104072138

指導教授：余朝恩 老師

# 主題

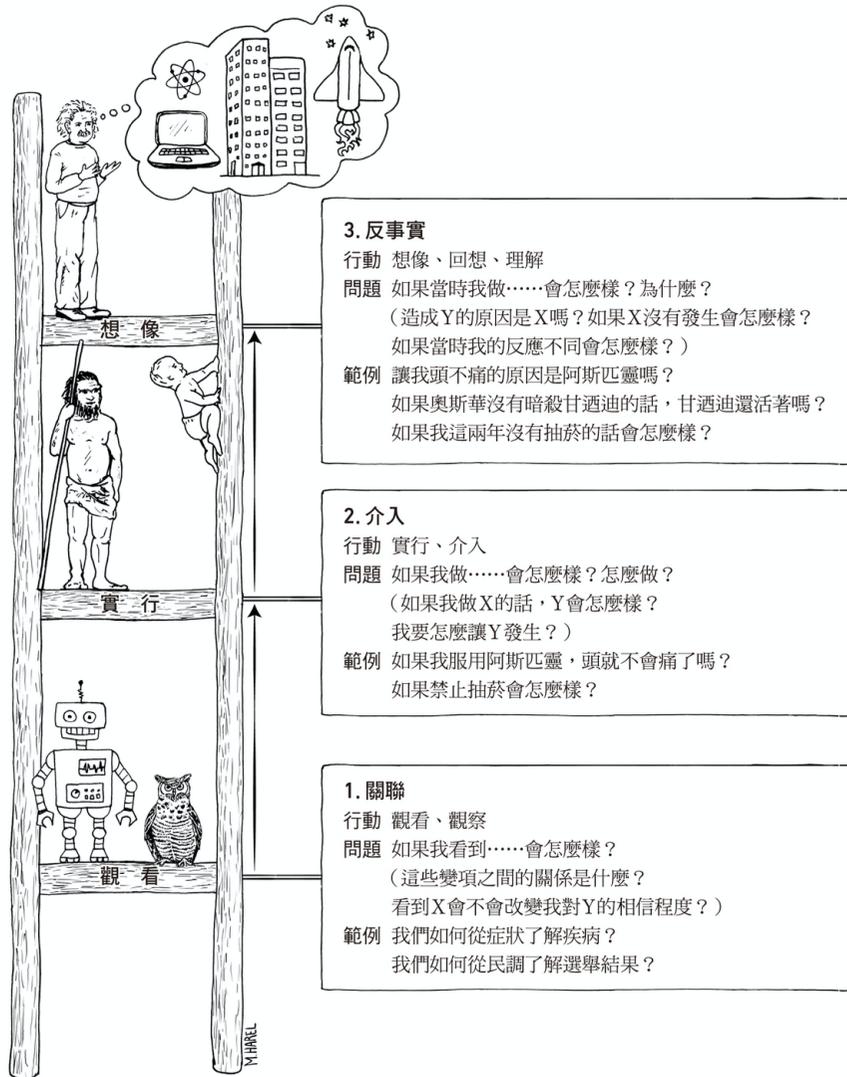




# 因果的三個層級

# 1

## 因果的三個層級



### 反事實推論

與事實不同且無法實驗

### 介入改變因素

細心控制的條件下進行實驗

### 蒐集和分析資料

依據被動的觀察結果進行預測



## 迷你圖靈測驗

2

## 迷你圖靈測驗

- 圖靈測驗

- 迷你圖靈測驗

- ✓ 因果問題
- ✓ 呈現與獲取



V.S.



## 2

# 迷你圖靈測驗

1st

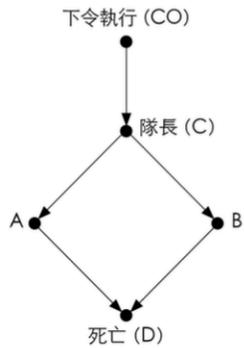


圖 1.4 行刑隊因果圖範例。A 和 B 代表士兵 A 和 B 的行動。

2nd

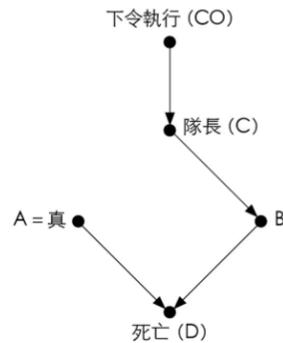


圖 1.5 關於介入的推理內容。士兵 A 決定開槍，刪除從 C 到 A 的箭頭，把 A 的值指定為真。

3rd

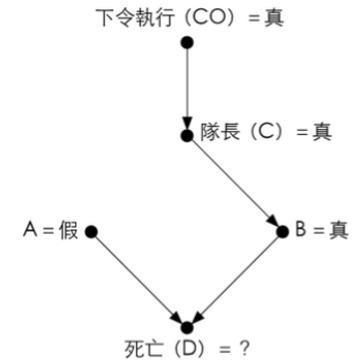


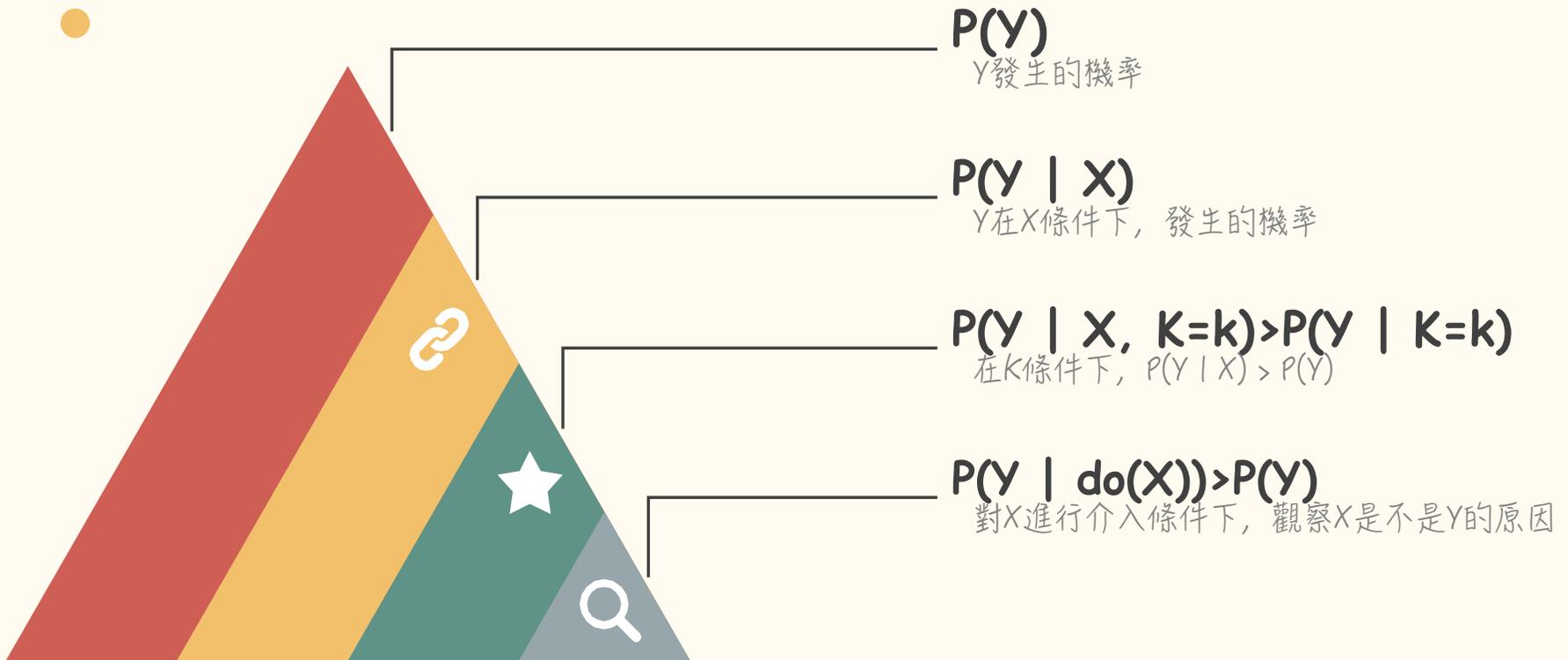
圖 1.6 反事實推理內容。我們觀察到囚犯已經死亡，並詢問如果士兵 A 決定不開槍的話，會出現什麼狀況。



# 機率與因果

### 3

## 機率與因果



THANK  
YOU

