

巨量資料與統計分析

政治大學統計系余清祥

2024年11月12日

第八週：定義問題(討論)

<http://csyue.nccu.edu.tw>

釐清問題的幾個相關疑問。。

□ 實際問題未必清楚定義，尤其是問題的陳述及其來源值得探究：

→ 疑問一：誰（最先）提出這個問題？

→ 疑問二：為什麼會以這種方式定義？

→ 疑問三：有助於理解問題的相關知識？

→ 疑問四：假設條件及邏輯推論合理嗎？

→ 疑問五：問題可從其他角度切入嗎？

範例一、化工廠排廢水造成死魚增多

Fogler and LeBlanc(1995)案例

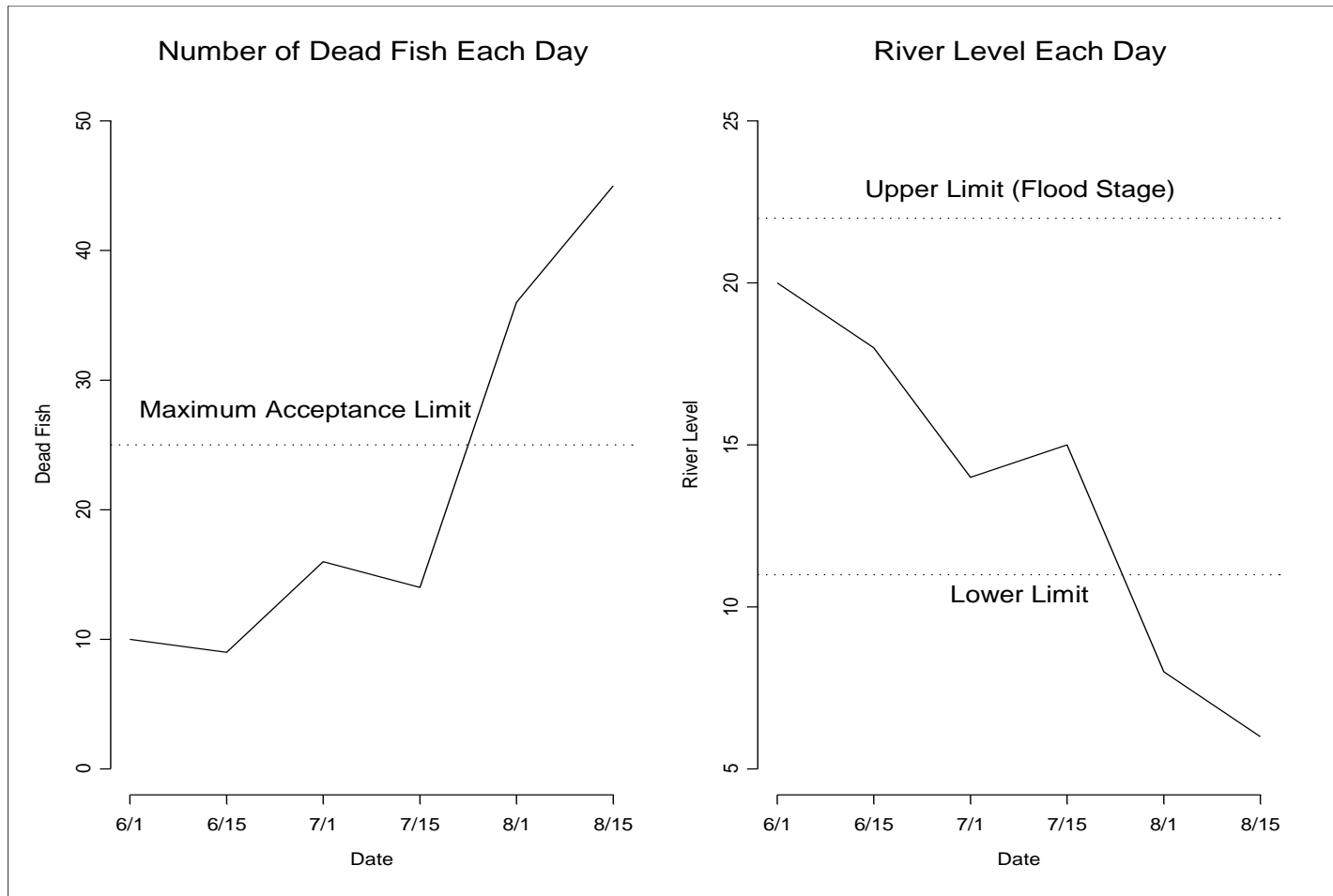
□ 問題敘述：旱象→水量少→化工廠廢水濃度高
→死魚多！！

→某地在夏天某段時間完全沒下雨，使得主要河流與湖泊水位較正常水位低許多，這段時間出現了比平常還多的死魚，且剛好都在某化工廠的下游。當地民眾認為是化工廠造成！



□ 某化學工廠將處理過(低毒性)的廢棄物，排放到一條流量相當大的河川，該工廠聘任專家觀察河川生態，以確保安全性。今年夏天專家發現以下現象：

Time Series Plot →



Control Chart ←

相關訊息：

□ 死魚總數在8月1日及15日超過可接受範圍。

→ 工廠在7月29日排放了一批較大量的廢水，但過去這樣的量並未造成任何問題。

→ 然而過去的經驗都是在水位較高時，無法確定水位低時是否一樣沒問題。

□ 如何解決死魚的問題？



定義問題範例一(續)

- 管理部門將死魚問題定義為「研發可降低廢棄物毒性的(化學)配方」
- 針對管理部門的建議，我們可以詢問
 - Who posed the problem?
 - Can reasons for arriving at the problem statement be explained?
 - Are the assumptions and reasoning valid?
 - Has sufficient data/information been collected?



當初的問題來源

- Where does the problem originate?
- Who did pose the problem statement in the first place?
- Can that person explain the reasoning as to how they arrived at that particular statement?
- Are the reasoning and assumptions valid?
- Has that person considered the situation from a number of different viewpoints before?
- Have you used the first four steps?



進一步探索問題

- Recall or learn the fundamental principles related to the problem
- Carry out an order-of-magnitude calculation
- Hypothesize what could be wrong
- Guess the result

定義問題範例一(續)

□ 進一步探索問題，步驟包括：

- → Identify available information

→ Learn Fundamental

(What would cause more dead fish?)

→ Missing information

(A fungus was found in near by areas.)

→ Hypothesis

(The fungus is the cause.)

→ More information (Proof!)

→ Define the real problem (Cure of the fungus!)

範例一(續)、化工廠的因應策略

- 生物專家化驗死魚，並到鄰近集水區看看是否有魚類或水生動物大量死亡，發現死魚身上有某種黴菌，且鄰近地區死亡生物上有相同黴菌。
- 微生物學家判斷夏天乾燥使得黴菌大量繁殖，導致死魚變多，證實工廠排放廢水與死魚無直接關係！
- 最後化工廠施放可殺死黴菌藥物，魚類死亡回復先前水準！



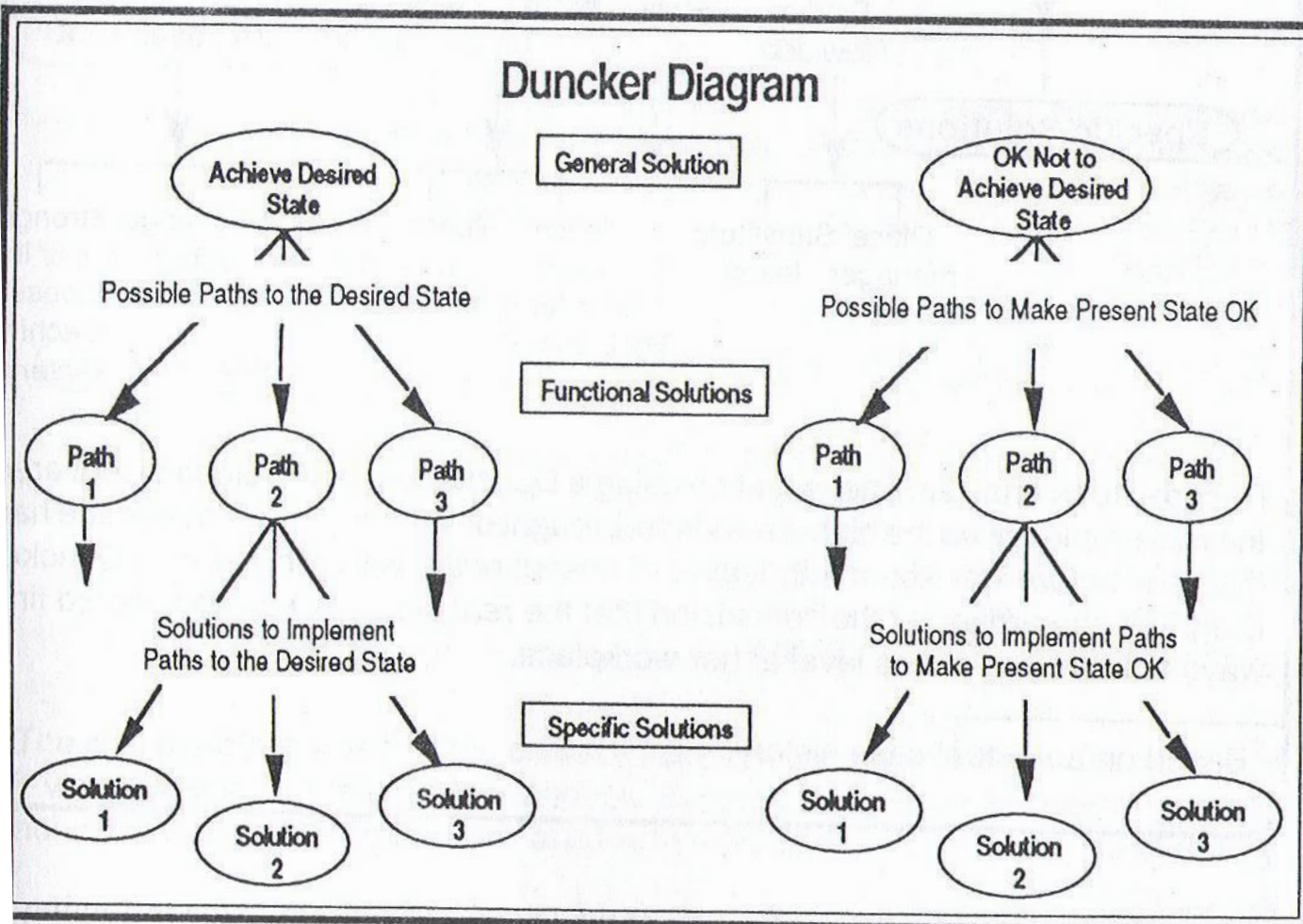
- Present state/Desire state

→ Write a statement of where you are and a statement of what you want to achieve.

- 在設計問卷時經常遭遇這種問題，如果無法獲得某些問項的答案時，應該如何因應？(例如：收入)
- 定義問題時盡量避免模糊或是模擬兩可的措辭，例如：「非常」、「相當」等。(英文中則有best, minimal, within a reasonable time, more efficient等)

■ Duncker diagram

→ Devise a pathway that make it OK to solve the problem posed to you.



Functional Solutions:
What to do.

Specific Solutions:
How to do it.

■ Statement-Restatement

→ Use the six triggers to restate the problem in a number of different ways.

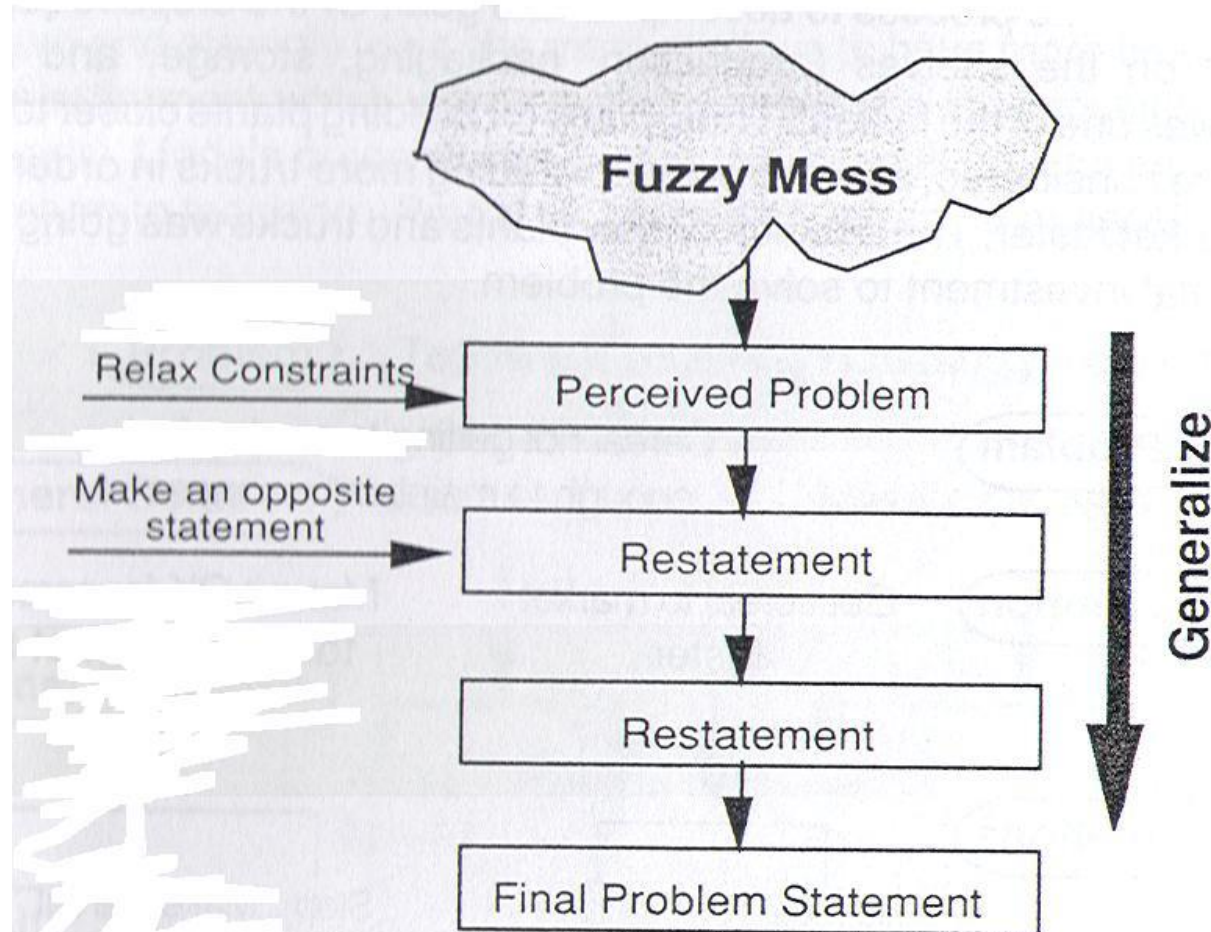
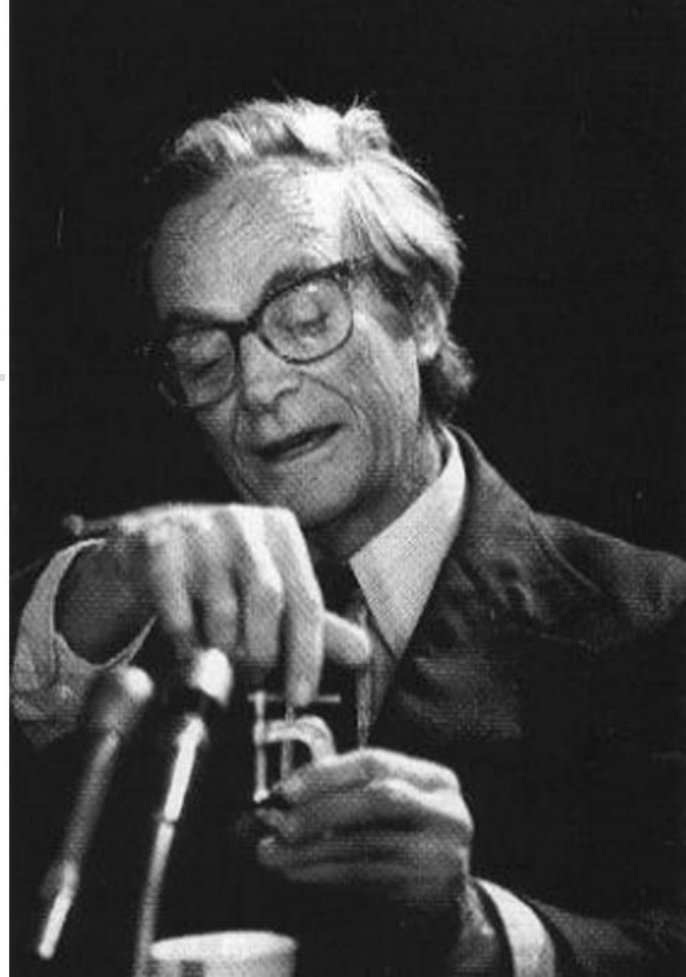


Figure 3-2. Stating the Real Problem

費曼的太空梭調查

- 1986年美國挑戰者號太空梭爆炸失事，物理學家費曼獲邀加入總統調查委員會。
- 費曼來回奔波美國首府華盛頓特區、休斯頓太空指揮中心、佛羅里達發射中心三地。
- 費曼在記者會實驗示範釀成災難主因為火箭助推器的O型橡皮環遇到低溫失去彈性，導致無法回復原狀而喪失密閉功能。



定義問題案例一



□ 「美教師連吃3個月麥當勞、

甩16公斤肥肉膽固醇下降」

→吃速食也會瘦？美國愛荷華州一名自然科老師約翰(John Cisna)想測試連吃3個月麥當勞(McDonald's)結果會如何，於是親自一項實驗，請他的學生每天照三餐幫忙買麥當勞速食，沒想到90天後他竟然沒變胖，反而甩肉37磅(約16公斤)。

案例一討論(續)



□ 問題在哪裡？

→ 為什麼速食給人的印象總是不健康，熱量太高、膽固醇太高、或是營養不均衡？

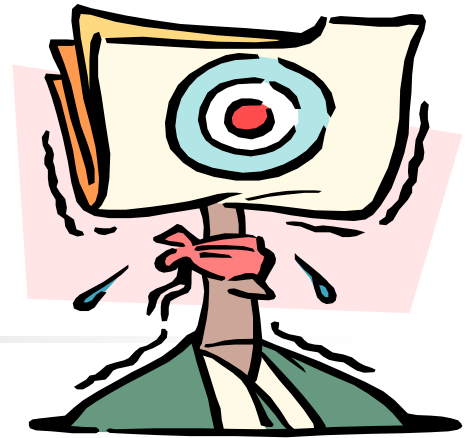
□ 這個報導顯示經過飲食控制，速食仍可達到減重、降低膽固醇的效果。

→ 為什麼多數吃速食的人無法看到這個效果？

→ 速食餐廳提供的餐點特色是什麼？

→ 何謂健康、如何測量健康？

定義問題案例二



□ 「多雨區學校、學童明顯變胖」

→ 據多年數據統計，基隆、宜蘭等東北角、北海岸等多雨地區，國小、國中學生明顯「又胖又矮」。（多雨使得學生無法運動）

□ 「因果關係」vs.「關連性」

→ 統計等數量化分析工具，可提供數值之間的關係程度，但單靠相關係數等統計量無法判定因果關係。（驟下結論！？）



案例二討論(續)

- 調查小學生的拼字能力，發現腳愈大的拼字能力也較強。
- 腳的大小影響拼字能力？
- 問題：哪些因素可能與學童變胖有關？
- 偏鄉運動設施不足（選項不多）、山區學校的空間較小。
- 鄉村地區的小孩在田間、野外活動機會不是較多嗎？（飲食？）

觀念移轉(Paradigm Shift)

- 有時因為過於熟悉問題背景，反而陷入不自覺的迷思，逆向思考、觀念移轉等技巧或許可協助我們找出關鍵。
- 例如：老狗學不了新把戲、先苦後甜、欲速則不達、別把雞蛋放在一個籃子裡
- 否定原先的命題，列出否定敘述的優點及特色，找出類似案例支持新論證，並規劃配合新想法的相關措施。

觀念移轉的練習

- 2003年SARS據說加速了臺灣宅配等物流業的興盛，如果未來再次爆發傳染病，會有哪些行業受到衝擊、哪些行業因此獲利？
- 如果賈伯斯還活著，蘋果電腦會推出什麼新產品，對人類電腦科技發展會產生什麼影響？
- 臺灣開始發展的地區是台南附近，現在臺灣重心卻在台北，重心轉移從何時開始、相關原因又是什麼？

問題與框架

□ 在台灣，從小大多假設凡事皆有標準答案，考試也多採選擇、填空等格式。

→ 相對於歐美學生，台灣學生的數字計算能力很強、反應速度也快；

（異口同聲的回答：「會」、「對」）

□ 比較不喜歡（或熟悉）的作業類型？

→ 申論題如何切入？

註：開放式、封閉式問卷的特色？

標準答案 vs. 思考空間

- 標準答案可以節省許多時間及成本，但也可能限制及壓縮了思考空間。
 - 最有可能的答案(Maximum Likelihood)的優缺點各是什麼？
- 台灣考試題目及制度的僵化，將影響國人全球化的競爭力。
 - 問題：既然希望國人接受更好的十二年國民教育，為什麼焦點總在「免試」？



問題討論練習

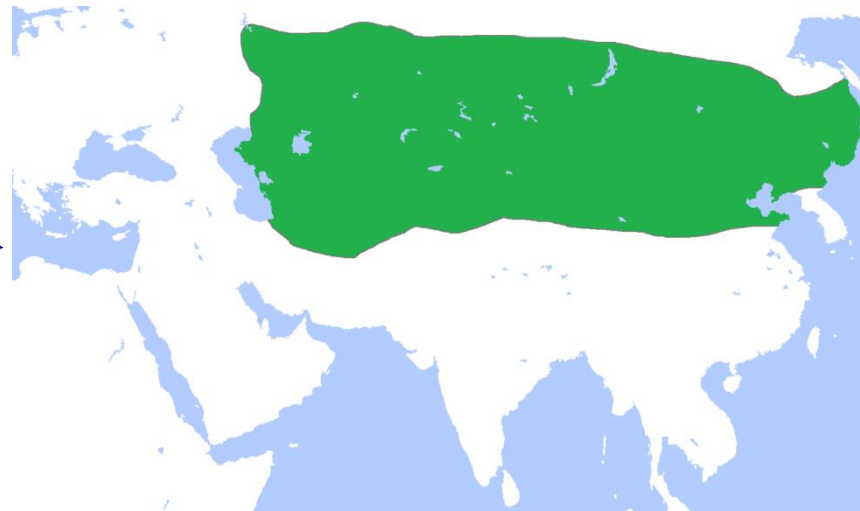
- 一位孩子在高二就讀的父親，有感於臺灣歷史考試題目的僵化，在網路上寫出對學校教育的看法（參考資料：「人家培養的是能力，而我們灌輸的是知識」）
 - 台灣高中的考題：「成吉思汗的繼承人窩闊臺，公元哪一年死？最遠打到哪裡？」
 - 美國中學生考題：「成吉思汗的繼承人窩闊臺，當初如果沒有死，歐洲會發生什麼變化？試從經濟、政治、社會三方面分析。」
- 註：窩闊台暴斃時，蒙古當時已進軍到維也納。

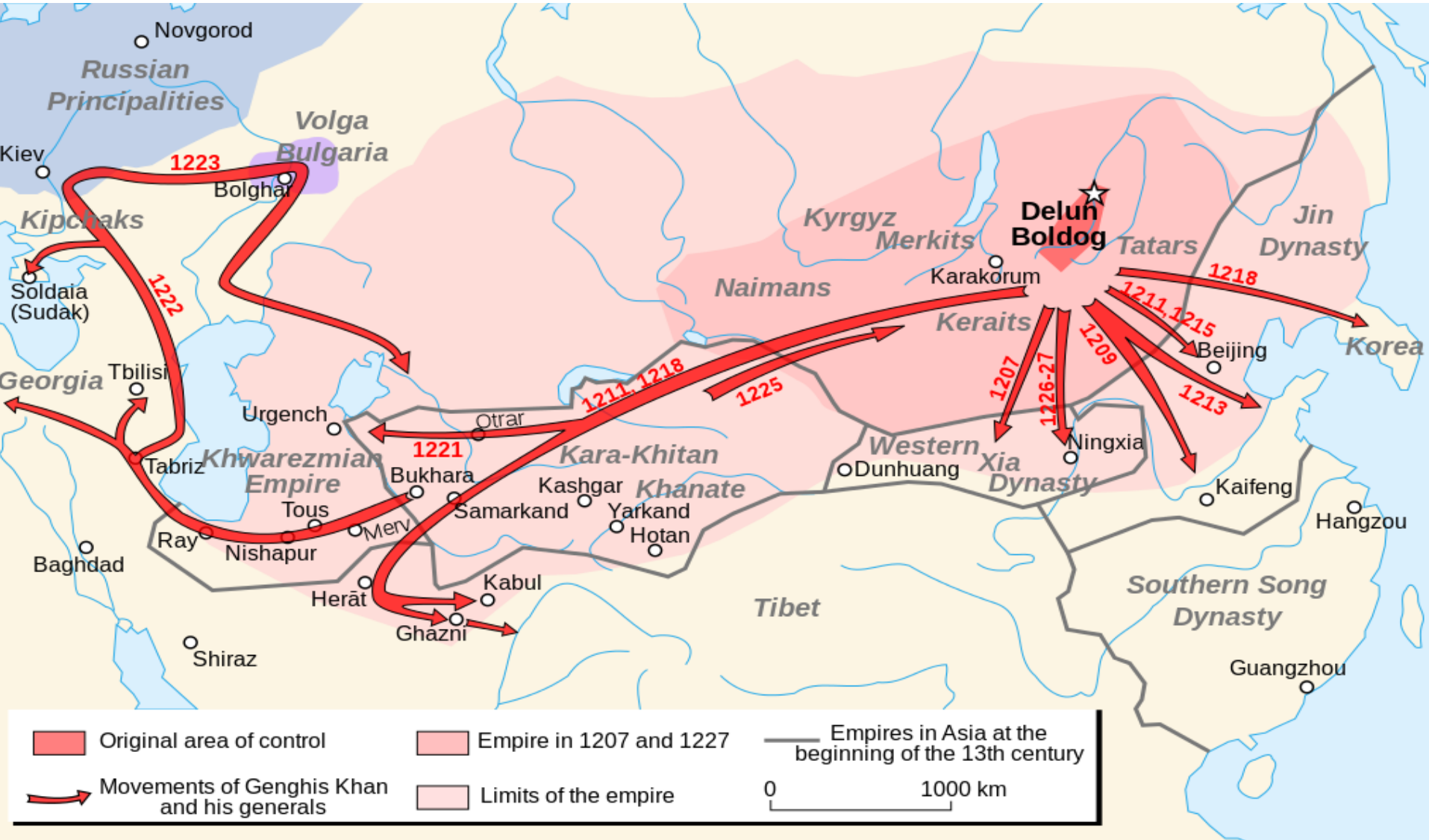
一些歷史相關背景

- 成吉思汗是大蒙古國皇帝，被尊稱為元太祖（元世祖忽必烈追封），其名「鐵木真」是由父親也速該在俘虜敵對部落

勇士鐵木真兀格，希望該勇士的勇氣會轉移至其身上而命名。

註：右圖為成吉思汗去世時的大蒙古帝國版圖。





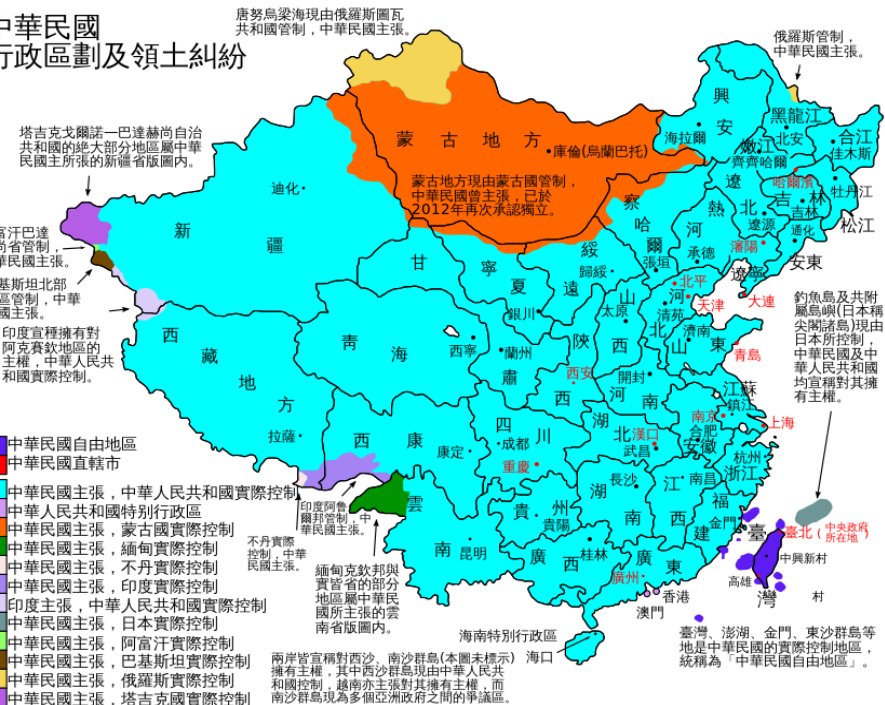
1206年~1227年成吉思汗在位時期，大蒙古國對金朝、西夏、西遼、花刺子模、高加索山一帶和東歐欽察草原的戰爭形勢圖，以及大蒙古國疆域的擴展。來源：維基百科。



漠北草原

成吉思汗生長於漠北草原，大致為中國現今的內蒙古，涵蓋中華民國疆域的興安、察哈爾、綏遠、寧夏四省。

中華民國行政區劃及領土糾紛



乞顏部的所在地

□ 也速該為乞顏部的酋長，部落大概位於現今的呼倫貝爾市（因為呼倫湖、貝爾湖得名）、海拉爾一帶。

→ 有別於一般大漠草原，乞顏部附近有額爾古納河，往北匯入黑龍江。



不同考題的回答範例

- 台灣考題的答案可從歷史課本、維基百科等文獻找到（1241年、匈牙利附近）。
- 美國考題的答案見仁見智，某個學生回答：
 - 這位蒙古領導人如果當初沒有死，那麼可怕的黑死病就不會被帶到歐洲去，後來才知道那個東西是老鼠身上的跳蚤引起的鼠疫。但是六百多年前，黑死病在歐洲猖獗的時候，誰曉得這個叫做鼠疫。如果沒有黑死病，神父跟修女就不會死亡。神父跟修女如果沒有死亡，就不會懷疑上帝的存在。如果沒有懷疑上帝的存在，就不會有意大利弗羅倫斯的文藝復興。如果沒有文藝復興，西班牙、南歐就不會強大，西班牙無敵艦隊就不可能建立。如果西班牙不夠強大，意大利不夠強大，盎格魯-撒克遜會提早 200 年強大，日耳曼會控制中歐，奧匈帝國就不可能存在。

美國學生答案的幾個關鍵詞

- 姑且不論美國學生答案是否合理，先檢視其中幾個重要關鍵詞：
 - 黑死病（鼠疫）；
 - 天主教、基督教的發展與傳遞；
 - 文藝復興；
 - 近代歐洲強權的興起歷史。
- 黑死病對歐洲人口造成了嚴重影響，改變社會結構、動搖了當時支配歐洲的羅馬天主教會的地位，並因此使得一些少數族群受到迫害。生存與否的不確定性，使得人們產生了「活在當下」的一種情緒，如同薄伽丘在《十日談》所描繪。（來源：維基百科）

幾個可能的討論方向

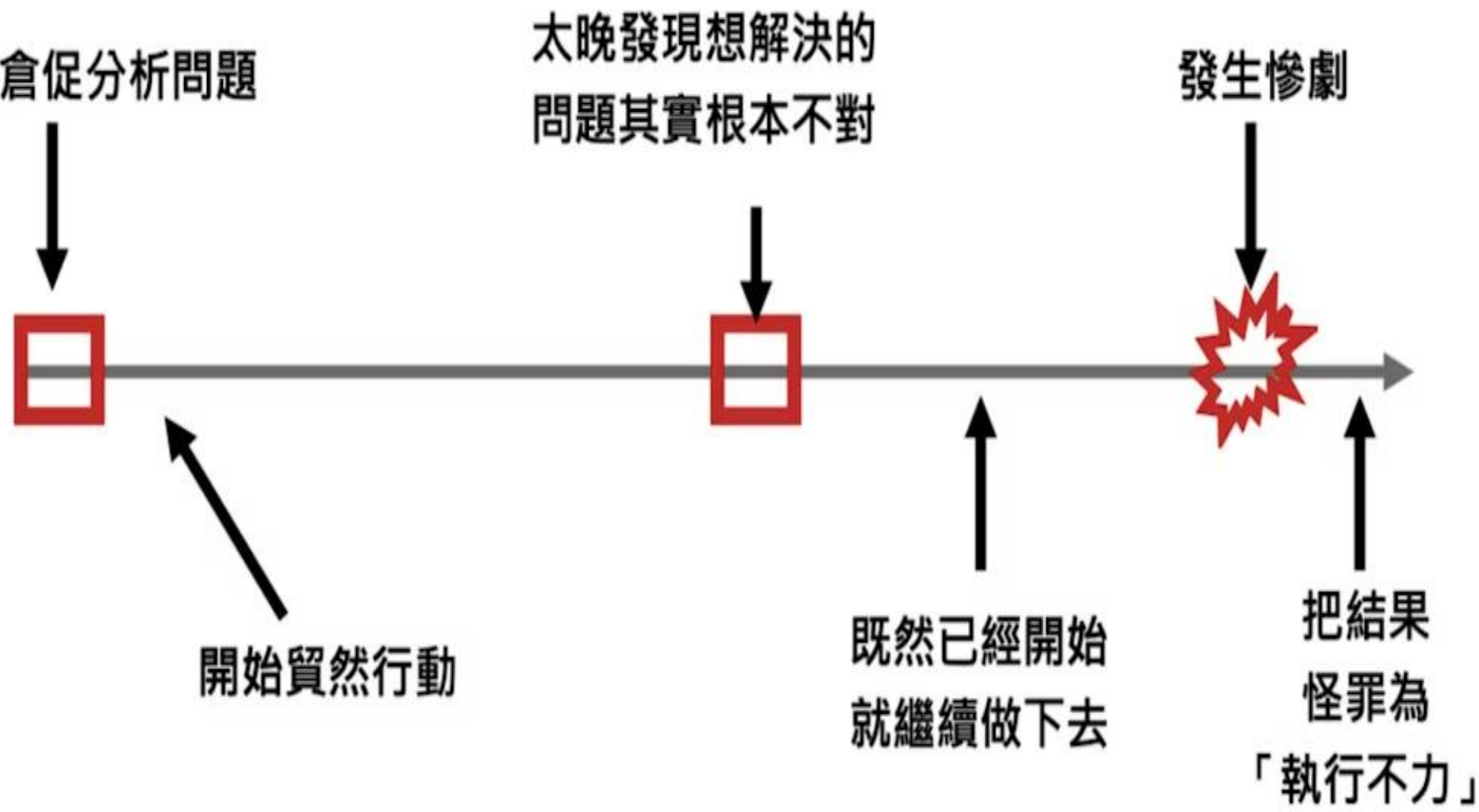
- 1999年12月的美國時代周刊評選出世界千年偉大人物，成吉思汗被列為千年人物中亞洲人的次位，僅次於印度聖雄甘地。
 - 不只亞洲人、西方人對成吉思汗也有很高的歷史定位評價，原因是什麼？
- 成吉思汗及其子嗣的西征，將中國的三大發明羅盤針、火藥、印刷術（現代化工具和知識的傳播）傳到西方，間接影響歐洲的文藝復興。（來源：劉學鈔《蒙古論叢》）

歐亞非帝國的比較



- 蒙古帝國並非第一個橫跨歐亞大陸的帝國，歐洲較為知名的馬其頓亞歷山大大帝，由希臘發跡，之後往東征戰波斯，最遠甚至到達阿富汗、印度河。
- 亞歷山大遠征帶來希臘文明在中東和中亞的繁盛，以及大夏—犍陀羅藝術在印度次大陸的發展。
- 成吉思汗的成功恰逢千年來最溫和、最潮濕的天氣，為以騎兵為主的蒙古大軍的戰馬提供豐富的飼料。

一、急著展開行動

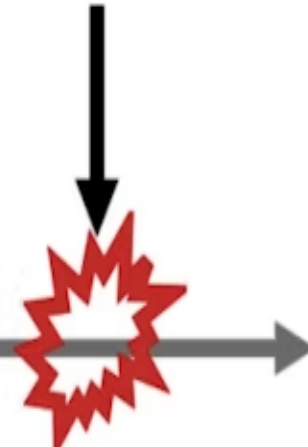


二、過猶不及，想太多而不斷鑽牛角尖，導致最後「分析癱瘓」

投入90%以上的
時間做分析



發生慘劇



必須有行動的壓力
「隨便做點什麼都好！」



《你問對問題了嗎？》一書作者建議比較好的辦法是：「繞一圈」。

1. 建立框架

我們要解決
什麼問題？

觸發「繞一圈」

3. 前進

如何維持動力？

2. 重組問題框架

對於這個問題，
有沒有不同的觀點？

- 跳出框架
- 重新思考目標
- 檢視亮點
- 照照鏡子
- 以他人觀點思考

