

# 中華民國第 65 屆中小學科學展覽會

## 作品說明書

高級中等學校組 行為與社會科學科

第三名

團隊合作獎

052701

社會模擬模型：探討特定機制對不同族群之生存  
趨勢影響

學校名稱： 華東臺商子女學校

作者：	指導老師：
高二 崔坤榆	嚴天龍
高二 翁晨軒	余冠毅
高二 蕭琦穎	

關鍵詞： 社會實驗、行為分析、AI

# 作品名稱

社會模擬模型：探討特定機制對不同族群之生存趨勢影響

## 摘要

本研究旨在探討人類、人工智能在特定機制下的抉擇。研究者首先固定一場景模型，設計模型中各族群的任務與環境機制，再使用程式還原所研發的模型，並利用微分方程式去驗證代碼運行的正確性。由於研究者希望此模型能夠貼近現實社會，故發表問卷以調查民眾、人工智能對於不同族群的選擇偏好，再將調查數據代入電腦程式，以比對人類與 AI 在選擇上的異同，並探討各種機制對於族群之間互動的影響，加以推論原因與發展趨勢，最後說明此模型在生活中的應用。

## 壹、前言

### 一、研究動機

#### （一）與所學課程之連結

研究者曾在學校的自主學習時間修過賽局理論領域的專題課程，在偶然機會下看到一個研究特定環境中，不同博弈行為對族群存活率影響的數據模擬實驗，故引發研究者繼續探討的興趣，並自行設計新的機制與角色，分析各種規則造成的結果。

#### （二）與生活之連結

社會新聞經常報道有關熱心人士見義勇為，或冷漠民眾袖手旁觀的案例，研究者不禁感到好奇：是什麼驅使人們在不同的情況下選擇利他或利己行動？

根據賽局理論課程所學知識，在多人博弈中，理性的玩家會選擇最優策略 (Nash, 1950)，故不同的抉擇是源自於玩家在權衡利弊後，選擇損失最小或獲利最大的方案而造成之結果。研究者因此設計不同條件的環境，並調查較符合現實人類社會情況的數據，以探討不同背景之人群的抉擇偏好，及模型機制、群眾選擇對各族群發展的影響。

此外，研究者也因應當下流行的人工智慧熱潮，比對人類與 AI 之間選擇偏好的差異性，以及其造成的趨勢。

### 二、研究目的

#### （一）調查人類、AI 在特定環境下之選擇異同。

調查人類、人工智能對不同族群的選擇意向並進行偏好排序，再加以分析。

#### （二）探討人類、AI 所選的族群在各種機制下之生存趨勢。

研究者設計了獎勵機制與改過自新機制。根據問卷調查結果，將數據代入具有不同機制組合的程式中，分別生成人類、AI 的趨勢圖。

#### （三）歸納族群間的利害關係與各情況之優勢族群，分析此模型在現實社會中的應用。

依據圖表分析族群交互過程，找出各情境下的優勢族群，分析其中原因，再將此模型類比到生活中發生的事件。

### 三、文獻回顧

此社會模型模擬出不同程度的利己與利他主義者間的交互關係、生存趨勢。研究者透過相關文獻瞭解族群間對應行為的產生原因，並對比文獻內容與本研究之關聯性。藉由文獻分析能使研究者對此模型的鑽研更加透徹，有助於新機制與角色的設計、探討。

#### (一) 模型基礎運作原則

##### 1. 不同行為區分成不同「基因」之種族

Dawkins (1976)提出「自私的基因」理論，認為基因的自私性會影響到個體之行為。

###### (1) 基因的自私性

基因的壽命不因個體的死亡而終結，個體完成職責後就被拋棄。

###### (2) 自然選擇基本單位

自然選擇的基本單位不是物種或個體，而是基因。

###### (3) 利他與利己行為

雖然基因的自私性通常會導致個體行為的自私性，但有時也會導致個體的利他主義行為。

###### (4) 與本研究之連結

- a. 族群並不會因個體提醒他人後死亡而滅絕。
- b. 「基因」相當於本研究中族群的行為，不同行為將影響生存率。
- c. 部分族群選擇只提醒與自己相同外貌者，但在利己的同時也給同外表的利己主義者（內奸，詳見模型基礎規則設定）增加存活率。

##### 2. 自私邏輯導致基因表現不同程度適應性

Startup (2021)提出，自私邏輯造成基因為繁衍而表現出不同程度的改變，以適應大環境。

###### (1) 自私基因之策略

基因通過變異和重組產生多樣策略來應對環境挑戰，如合作、競爭、欺騙等，以確保自身在進化中的優勢地位。

###### (2) 利他為利己之表現

利他行為實際是自私基因控制的結果，目的為通過照顧他人，間接幫助完成自身的複製，得以繁衍後代。

###### (3) 與本研究之連結

- a. 一部分利他主義者為了加強族群內部的合作而致使基因發生重組、變化，導致形成不同外表，有助於內部成員的辨別，進而互相提醒危險的來臨，使存活率上升（好人，詳見模型基礎規則設定）。
- b. 與此同時，有部分利己主義者為了受到被他人提醒的利益，將外表進化成上述利他主義者的模樣，隱藏其中，導致存活率上升（內奸，詳見模型基礎規則設定）。

### 3. 利他行為之存在

#### (1) 漢彌爾頓法則

利己主義者在環境中的生存率通常居高不下，那為何社會上還是有許多利他主義者的存在？Hamilton (1964)描述了利他行為的演化，核心內容可以用數學公式表達為：

$$rb - c > 0$$

r：行動者與接受者在遺傳上的親近程度

b：行動者的利他行為給接受者帶來的繁殖利益增加量

c：行動者實施利他行為所付出的代價

當利他行為給接受者帶來的利益乘以兩者的親緣係數，大於行動者所付出的代價時，表現某些行為特徵的基因頻率將會增加。代表親緣關係越近，動物間合作傾向和利他行為越強烈；親緣越遠，則表現越弱。

#### (2) 與本研究之連結

有部分利他主義者傾向提醒相同外表者危險的來臨。在行動者的視角，相同外表者與自身親緣相近，故利他傾向表現強烈。

### 4. 公共品博奕：獎勵機制強化合作行為

Fehr & Gächter (2000)提出基於公共品博奕(Public Goods Game)的獎勵模型，探討正向誘因對合作行為的影響。

#### (1) 模型規則

玩家可選擇貢獻資源（合作）或保留資源（背叛），並可支付成本獎勵其他合作者。

#### (2) 獎勵機制

實驗顯示，獎勵機制明顯提高群體合作率，且效果比懲罰更持久，此外，被獎勵者獲得額外收益後，會形成正向循環。

#### (3) 與本研究之連結

設計「獎勵機制」，鼓勵好人、英雄族群的提醒行為，使其與內奸族群制衡（詳見族群、機制設計）。

### 5. 三方演化模型：懲罰機制減少不合作行為

三方演化(Tripartite Evolutionary Game Model)是演化賽局理論中的一種模型。

#### (1) 模型規則

群體中存在三種策略類型的玩家（合作者、背叛者、監督者），透過策略間的互動與收益差異，驅動群體行為的長期動態變化。

#### (2) 懲罰機制

群體中監督者(punishers)透過懲罰行為抑制背叛者(defectors)，降低背叛者的收益。由於背叛者面臨懲罰風險，策略從「主動背叛」轉為「低調自利」，以避免利益損失(Boyd et al., 2010)。

### (3) 與本研究之連結

設計「改過自新機制」(詳見機制設計)，使內奸(背叛者)遇到警長(監督者)之後有一定機率停止欺騙行為，轉變為低調自利的自私者(詳見族群、機制設計)。

表 1：模型基礎運作原則之文獻探討統整表格

	文獻	研究內容	與本研究之連結
族群 設定 與辨 識機 制	The Selfish Gene (Dawkins, 1976)	(1) 基因的壽命不因個體的死亡而終結。 (2) 自然選擇的基本單位是基因。 (3) 自私行為有時會導致個體的利他主義。	(1) 族群不會因為單一個體的死亡而滅絕。 (2) 基因相當於行為，各族群根據行為影響生存率。 (3) 部分人只提醒與自己相同外貌者， <u>利己的同時也增加同外表的內奸之存活率</u> 。
	The theory of the selfish gene applied to the human population (Startup, 2021)	(1) 基因通過變化來應對環境挑戰，確保自身優勢地位。 (2) 利他實際是自私基因控制的結果，目的為繁衍後代。	(1) 部分利他主義者為 <u>加強族群內部合作</u> 而形成不同外表，有助於 <u>成員的辨別</u> 。 (2) 部分利己主義者為受到被提醒的利益， <u>將外表進化成利他主義者的模樣，隱藏其中</u> 。
	The genetical evolution of social behaviour (Hamilton, 1964)	親緣越近，彼此間合作傾向和利他行為越強烈；親緣越遠，則表現越弱。	<u>部分利他主義者傾向於提醒相同外表者危險的來臨</u> 。
獎勵 機制	Cooperation and punishment in public goods experiments (Fehr & Gächter, 2000)	獎勵機制 <u>提高合作意願</u> ，且被獎勵者獲得額外收益後，會形成正向循環。	設計「 <u>獎勵機制</u> 」，鼓勵好人、英雄族群的提醒行為。
改過 自新 機制	Coordinated punishment of defectors sustains cooperation and can proliferate when rare (Boyd et al., 2010)	監督者透過懲罰行為抑制背叛者，使背叛者的策略從「 <u>主動背叛</u> 」轉為「 <u>低調自利</u> 」。	設計「 <u>改過自新機制</u> 」，使內奸(背叛者)遇到警長(監督者)之後有一定機率停止欺騙行為，轉變為自私者。

## (二) 模型基礎規則設定

### 1. 場景設定

- (1) 食物：使族群繁衍之主要資源。
- (2) 怪物：無差別隨機攻擊覓食者。
- (3) 各大族群：依行為、外表而做出區別，有各自的任務（詳見族群設計）。

### 2. 基礎原則

- (1) 每天日間除警長外，所有人須出門覓食一次以生存下去。
- (2) 出門時必會遇見怪物，有 50% 的機率被怪物殺害。
- (3) 覓食時必遇到其他人，種族隨機，相遇的人並不知曉是否有怪物存在。
- (4) 覓食成功者夜間可繁殖一個個體（無獎勵機制時）。

### 3. 提醒機制

- (1) 覓食時可以看見相遇之人附近是否有怪物，可選擇是否提醒對方。
- (2) 如果提醒則對方必逃脫，而自己有 50% 機率因提醒他人而死亡。  
(50% 機率雙方皆存活，50% 機率個體犧牲換取對方倖存。)
- (3) 如果不提醒他人，則個體與對方皆維持當天有 50% 的機率被怪物殺害。
- (4) 所有人只能辨別外表，無法得知對方確切所屬族群。

表 2：基礎角色設定（研究者將在實驗加入新族群「警長」與「造謠者」）

族群	好人	內奸	英雄	自私者
外貌	外表甲	外表甲	外表乙	外表乙
提醒機制	只提醒外表甲	不提醒任何人	提醒所有人	不提醒任何人

## (三) 人類與 AI 的抉擇差異

人類與人工智慧的決策邏輯源於不同的基礎，人類仰賴經驗與直覺，而 AI 則基於數據統計的關聯性。此差異使兩者在開放性創新與結構化任務中各具優勢與限制(Kahneman & Tversky, 1979)。

### 1. 人類

#### (1) 優點

人類決策的獨特優勢在於整合隱性知識與社會價值，使得決策具有情境適應性與道德反思空間(Gigerenzer & Gaissmaier, 2011)。

#### (2) 缺點

人類受框架效應等認知偏誤影響，可能在高風險情境中引發非理性誤判(Kahneman & Tversky, 1979)。

### 2. 人工智慧

#### (1) 優點

AI 在結構化任務中展現超越人類的效能，能從複雜數據中提取關鍵特徵並實現高速分類(Esteva et al., 2017)。

#### (2) 缺點

AI 的決策品質受制於訓練數據的代表性，且缺乏因果推理能力，導致其複製並放大社會結構性偏見(Buolamwini & Gebru, 2018)。

## 貳、研究設備器材

表 3：研究設備與功能概述

電腦軟體	功能
線上問卷	調查人類、AI 之偏好族群，將問卷結果代入程式。
人工智慧	共選取五款 AI，詢問其選擇，並與人類比較。
R 語言	進行程式設計、實驗模擬、趨勢圖生成。

表 4：研究者所選之人工智慧

人工智慧				
Llama-3-Groq-70B	Claude-3.7-Sonnet	Gemini-2.0-Flash	Chat-GPT o1	Grok-2

## 參、研究過程或方法

### 一、研究方法

#### (一) 調查法：

透過問卷調查民眾與人工智慧選擇的族群，分析兩者異同。

##### ■ 調查流程：

1. 設計問卷並發放，採用線上問卷與線下紙本問卷兩種形式。
2. 將紙本問卷的 PDF 檔案傳送給五款人工智慧。
3. 回收問卷，並分別進行數據統計、分析。

#### (二) 實驗法

研究者透過 R 語言模擬所設計的規則、機制與角色，探討各種組合對於模擬社會的影響與各族群之生存趨勢。研究者以先前調查法所得之數據代入程式中，輸出結果與趨勢圖，最後加以分析之。

##### ■ 實驗流程：

1. 設計實驗內容與程式碼。
2. 將問卷的數據依對應比例代入程式中，並開始實驗模擬。
3. 分別生成人類、AI 的族群生存趨勢圖。
4. 圍繞趨勢圖進行分析討論。

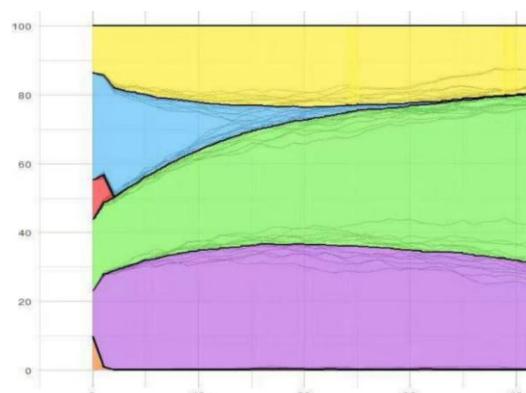
圖 1-1：程式碼之部分截圖

```
# 內奸 (type=3, appearance=B)
if (initial_populations["內奸"] > 0) {
  population_list[[3]] <- data.frame(
    type = rep(3, initial_populations["內奸"]),
    appearance = rep("B", initial_populations["內奸"]))
}

# 好人 (type=4, appearance=B)
if (initial_populations["好人"] > 0) {
  population_list[[4]] <- data.frame(
    type = rep(4, initial_populations["好人"]),
    appearance = rep("B", initial_populations["好人"]))
}

# 警長 (type=5, appearance=A)
if (initial_populations["警長"] > 0) {
  population_list[[5]] <- data.frame(
    type = rep(5, initial_populations["警長"]),
    appearance = rep("A", initial_populations["警長"]))
```

圖 1-2：趨勢圖之部分截圖



圖片來源：研究者實驗

### (三) 數據挖掘

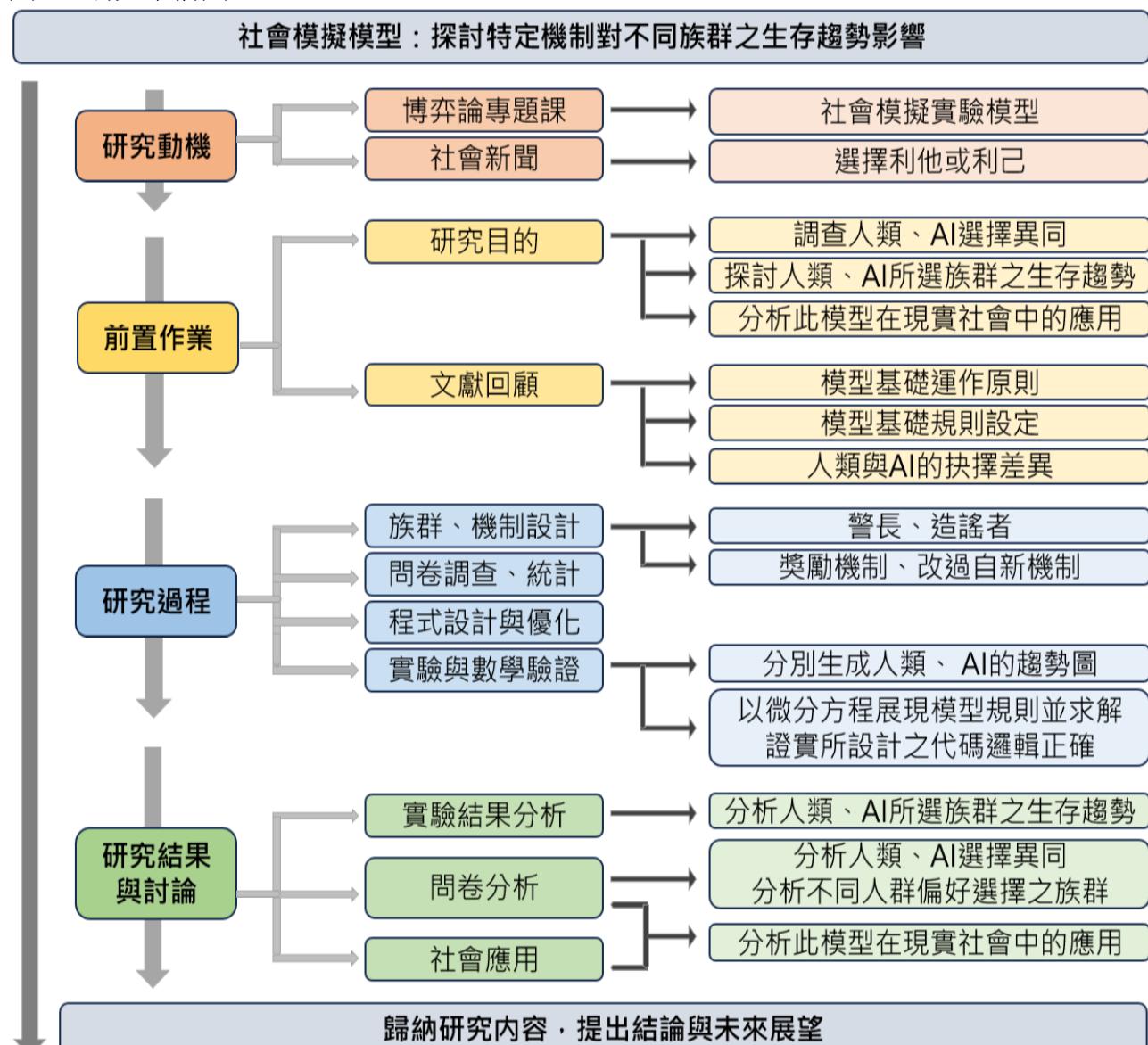
使用微分方程式展現各族群每天的剩餘人數，從中探討行為模式並計算出達穩定態時的人數比例，用於驗證實驗法的合理性。若證實研究者所設計之實驗流程無誤，表示可將由調查法獲取之數據代入先前的程式碼，進行下一步研究。

#### ■ 驗證流程：

1. 以實驗一為例，用微分方程式分別列出各族群理論上每日的人數變化。
2. 計算此微分方程式的解，得出穩定態時各族群的人數比例。
3. 將理論結果與實驗一之實驗結果比對。
4. 驗證研究者所設計之模型規則的程式碼是否正確無誤。
5. 若正確，則以此程式碼為基礎，進行後續實驗。

## 二、研究架構

圖 2：研究架構圖



圖片來源：研究者自行繪製

### 三、研究限制

#### (一) 硬體設備算力限制

由於電腦計算能力有限，導致在進行實驗時，研究者需對天數進行限制。計算資源的不足使模擬過程中的運算效率降低，這可能對研究進度造成一定影響。

因此，研究者設計實驗方案時，選擇優化數據處理流程，同時也嘗試縮減不必要的計算環節，以確保在有限的算力條件下能夠有效獲取具價值的結果。

#### (二) 問卷回收限制

研究者原定計劃為分析性別、年齡、教育背景對族群選擇偏好的影響，雖已盡可能隨機發放問卷，讓各種生活背景的民眾填答，但由於研究者周遭以國、高中生居多，而能夠接觸到的老年人口較少，導致受試者年齡、教育背景分佈不均，因此不適合進行問卷分析，故研究者在此以性別作為主要探討因素。

而年齡、教育背景等分析，則納入後續研究的範疇。

### 四、研究過程

#### (一) 族群設計

研究者為了模擬社會的複雜情況，在原有的四個族群之基礎上，額外設計出兩個族群，使得族群之間的交互狀況更多元。

##### 1. 造謠者

為模擬現實社會，研究者加入會欺騙其他族群以換取自身利益的造謠者。

- (1) 造謠者為外表丙，每次出門會碰到一隨機族群的人。
- (2) 若此人不屬於造謠者族群則進行欺騙，迫使此人前往沒有果實之處。
- (3) 被欺騙者沒有果實，當天夜間無法繁殖，但也不會因碰到怪物而死亡。
- (4) 造謠者不提醒任何人危險的來臨。

##### 2. 警長

為克制背叛者的行為，故新增警長，希望能與內奸、造謠者制衡。

- (1) 警長為外表乙，每天有 40% 機率開槍擊斃一名外表為甲者（好人、內奸為外表甲），60% 機率擊斃一名造謠者。
- (2) 有 20% 機率此外表甲者為好人，若誤殺好人則警長與該名好人皆死亡。  
(若好人滅絕，則警長此後判斷錯誤之機率為 0%。)
- (3) 假設正確處決內奸，當晚警長由一個個體分裂為兩個。
- (4) 警長在回家的路上有 50% 機率被怪物殺害（先判定開槍後是否存活）。
- (5) 環境中不存在內奸時，警長不再開槍，改以覓食的方式繁殖個體。
- (6) 警長不提醒任何人危險的來臨。

表 5：好人、內奸、英雄、自私者、警長與造謠者之外表及行為比較

族群	好人	內奸	英雄	自私者	警長	造謠者
外貌	外表甲	外表甲	外表乙	外表乙	外表乙	外表丙
提醒	外表甲	不提醒	所有人	不提醒	不提醒	不提醒

## (二) 機制設計

研究者為鼓勵英雄、好人族群的利他行為，以及使模型更符合現實社會的狀況，設計了兩個機制，使模型的複雜程度提升。

### 1. 嘉獎機制

- (1) 目的：獎勵合作者，提高合作意願。
- (2) 規則：提醒他人並生存下來者夜間可繁殖兩個個體（內奸、自私者、警長、造謠者維持不提醒任何人的原則）。

### 2. 改過自新機制

- (1) 目的：使內奸的策略從「主動背叛」轉為「低調自利」。
- (2) 規則：依問卷回收結果（詳見問卷調查與統計），設定內奸的族群轉換率，使內奸遇到警長後有一定機率停止欺騙行為，轉變為自私者族群。

## (三) 問卷調查與統計

調查受試者之性別、年齡、教育背景，以及族群偏好排序、改過自新意願。研究者在問卷中提供簡潔、完整的模型介紹，使受試者能夠明白規則，此外，為避免族群名稱影響決策，故在問卷中只說明各族群之行為與規則。

### 1. 族群偏好排序

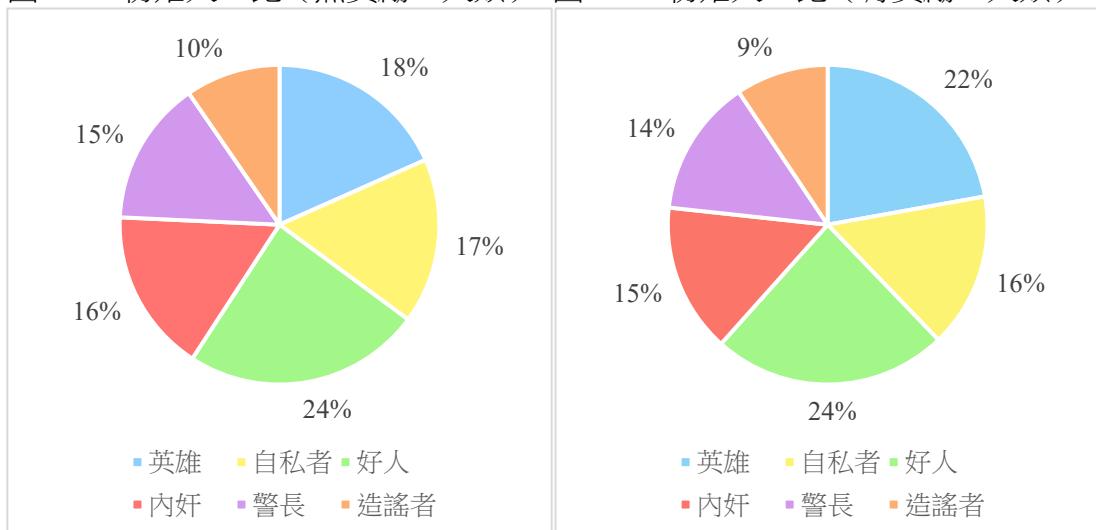
共六個族群，依照個人偏好排名，最喜歡的族群賦值為 6，第二喜歡的族群賦值為 5，第三喜歡的族群賦值為 4，以此類推，最後將各族群得分加總，以各族群的分數比作為模擬實驗的初始人口數比（AI 則取五款人工智慧的選擇平均值作為初始人口數）。

### 2. 內奸族群轉變率

若受試者將內奸排在前三順位，則需選擇是否要停止欺騙行為。研究者將內奸族群轉變率設定為「願意停止欺騙他人者佔所有受試者的比例」。

■ 共回收 423 份問卷，其中有效問卷 400 份，佔比約 94.6%。

圖 3-1：初始人口比（無獎勵，人類） 圖 3-2：初始人口比（有獎勵，人類）



圖片來源：研究者自行繪製

圖 4-1：初始人口比（無獎勵，AI）

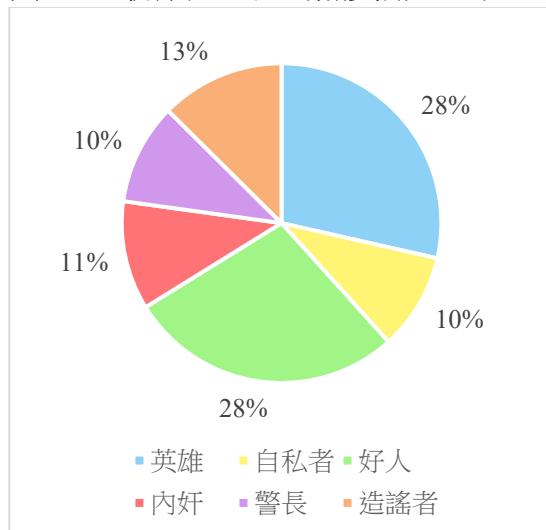
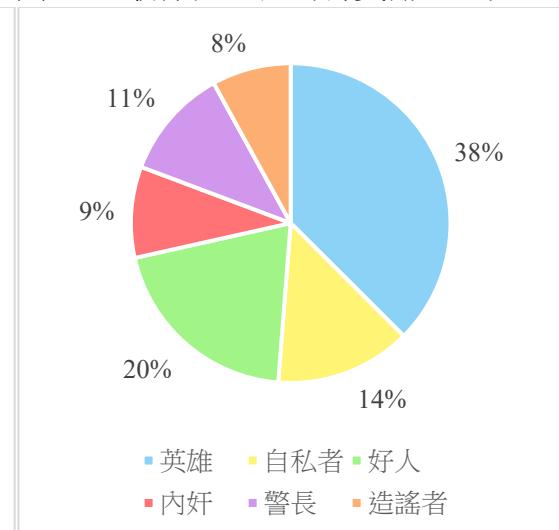


圖 4-2：初始人口比（有獎勵，AI）



圖片來源：研究者自行繪製

表 6：人類與 AI 之內奸族群轉變率

受試者	內奸族群轉變率（改過自新）	
	無獎勵機制	有獎勵機制
Llama-3-Groq-70B	10.0%	20.0%
Claude-3.7-Sonnet	13.0%	15.0%
Gemini-2.0-Flash	20.0%	15.0%
Chat-GPT o1	30.0%	20.0%
Grok-2	10.0%	30.0%
人工智慧平均	16.5%	20.0%
人類	22.5%	15.5%

#### （四）程式設計

##### 1. 初始設定與參數

首先載入必要套件 tidyverse 及 ggplot2，並設定一些基本參數。

- (1) days 為模擬執行的天數。
- (2) simulations 表示模擬次數，以觀察多次結果是否一致。
- (3) groups 與 group\_colors 用來在後續視覺化時區分各個族群。
- (4) initial\_populations 設定各族群在人口初始時的數量。

圖 5：模擬環境設定之代碼

```
library(tidyverse)
library(ggplot2)
days <- 150
simulations <- 10
groups <- c("自私者", "英雄", "内奸", "好人", "警長", "造謠者")
group_colors <- c(
  "自私者" = "#FFF24F", "英雄" = "#6FC5F5", "内奸" = "#FF4F4F",
  "好人" = "#8EF76D", "警長" = "#C47DE7", "造謠者" = "#FF9A4F"
)
initial_populations <- c(
  "自私者" = 14080, "英雄" = 15410, "内奸" = 13900,
  "好人" = 20230, "警長" = 12310, "造謠者" = 8070
)
```

圖片來源：研究者實驗

## 2. 核心模擬邏輯

模擬核心函數 `simulate_game_fast()`，其主要流程如下：

- (1) 每天進行警長執法（若有內奸或造謠者則可能被警長槍斃）。
- (2) 進行各族群的「覓食和繁殖」：
  - a. 多數族群都有基本二分繁殖機制（繁殖後人數約為存活數量兩倍）。
  - b. 英雄或好人在提醒過程中，可能有犧牲或被反噬的風險。
- (3) 不斷重複上述過程，累積每天各族群的比例變化。

圖 6：群體更新機制：警長特例與生存機率

```
new_populations <- numeric(length(groups))
names(new_populations) <- groups
if (populations["內奸"] < 0) {
  new_populations["警長"] <- populations["警長"] + police_kills
} else {
  if (populations["警長"] > 0) {
    survivors <- rbinom(1, populations["警長"], 0.4)
    new_populations["警長"] <- survivors * 2
  }
}
for (group in groups[groups != "警長"]) {
  if (populations[group] <= 0) {
    new_populations[group] <- 0
    next
  }
  n <- populations[group]
  base_survivors <- rbinom(1, n, 0.5)
```

圖片來源：研究者實驗

## (4) 模擬流程重點

- a. 警長執法：內奸與造謠者會在一定機率下被槍斃；若有誤殺，好人與警長本身人口可能會下降。
- b. 繁殖機制：大多數族群存活後都以「存活人數×2」作為翌日人口。
- c. 族群提醒(英雄、好人): 損失人數由二項分佈模擬，無額外繁殖獎勵。

圖 7：英雄與好人族群的行為模擬邏輯

```
if (group == "英雄") {
  attempting_heroes <- base_survivors
  sacrificed_heroes <- rbinom(1, attempting_heroes, 0.2/0.7 * 0.7)
  base_survivors <- base_survivors - sacrificed_heroes
} else if (group == "好人") {
  total_pop <- sum(populations)
  if (total_pop > 0) {
    prob_traitor <- populations["內奸"] / total_pop
    prob_good <- populations["好人"] / total_pop
    attempting_good <- base_survivors
    traitor_encounters <- rbinom(1, attempting_good, prob_traitor)
    killed_by_traitors <- rbinom(1, traitor_encounters, 0.9)
    base_survivors <- base_survivors - killed_by_traitors
  }
}
new_populations[group] <- base_survivors * 2
}
populations <- new_populations
```

圖片來源：研究者實驗

### 3. 多次模擬與結果視覺化

最後，研究者將函數執行多次，合併所有結果並進行視覺化。

圖 8：合併與繪圖代碼段落

```
p <- ggplot() +  
  geom_line(data = individual_cumulative,  
            aes(x = Day, y = CumProp,  
                  group = interaction(Simulation, Group)),  
            colour = "black", size = 0.5, alpha = 0.8) +  
  geom_area(data = average_results,  
            aes(x = Day, y = Proportion, fill = Group),  
            alpha = 0.8, colour = "black", size = 1,  
            position = "stack") +  
  scale_fill_manual(values = group_colors) +  
  scale_x_continuous(breaks = seq(0, days, 10)) +  
  scale_y_continuous(breaks = seq(0, 100, 20), limits = c(0, 100)) +  
  labs(  
    title = "族群占比變化 - 無獎勵無變換版",  
    subtitle = paste("模擬時間:", format(Sys.time(), "%Y-%m-%d %H:%M:%S")),  
    x = "天數",  
    y = "族群占比 (%)",  
    fill = "族群"  
) +  
  theme_minimal() +  
  theme(  
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),  
    plot.subtitle = element_text(size = 12, hjust = 0.5),  
    legend.position = "right",  
    panel.grid.major = element_line(colour = "grey80"),  
    panel.grid.minor = element_line(colour = "grey90")  
)  
print(p)  
cat("\n==== 最終存活族群統計 ===\n")  
print(final_survival)  
cat("\n==== 各族群最終平均占比 ===\n")  
print(final_proportions)
```

圖片來源：研究者實驗

- (1) bind\_rows() 整合各次模擬結果，利於後續統計或圖表呈現。
- (2) 使用 ggplot2 搭配不同顏色，可視覺化各族群比例的變化趨勢。

## (五) 實驗內容與數學驗證

### 1. 實驗一：辨識機制對人類、AI 所選族群的生存影響

#### (1) 研究目的

探討人類、AI 所選的族群在辨識機制下之生存趨勢。

#### (2) 研究方法：實驗法、數據挖掘

#### (3) 實驗內容

根據模型基礎規則並依族群的行動與外表分為六個族群，族群間有辨識機制。將六大族群分別以調查法統計出的數據作為初始人數佔比，代入程式中，並分別生成人類、AI 的生存趨勢圖。

表 7：實驗一族群、機制整理

族群	好人	內奸	英雄	自私者	警長	造謠者
外貌	外表甲	外表甲	外表乙	外表乙	外表乙	外表丙
提醒	外表甲	不提醒	所有人	不提醒	不提醒	不提醒
機制	無					

#### (4) 數據挖掘

根據高中數學課所學知識，研究者嘗試以馬可夫鏈理論計算出平衡態時的族群比例，並將此結果代入到實驗中，輔助說明結果。

因好人、英雄、警長、造謠者族群於實驗一中處於弱勢，無法活到最後，故此處以實驗一中的內奸、自私者族群為例，敘述研究者的推導過程。

根據程式模擬過程，研究者觀察到自私者的存活率為 89.8%，死亡率為 10.2%；內奸的存活率為 63.7%，死亡率為 36.3%。故研究者在此設轉移矩陣為：

$$A = \begin{bmatrix} 0.898 & 0.637 \\ 0.102 & 0.363 \end{bmatrix}$$

在馬可夫鏈理論中，如果要找出長期穩定狀態下的比例，需要確保轉移矩陣乘上該比例後還會是原比例，即：

$$\begin{bmatrix} 0.898 & 0.637 \\ 0.102 & 0.363 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

其中  $x$  為平衡狀態下自私者的人數， $y$  為平衡狀態下內奸的人數。

經過計算後，可得：

$$x:y = 637:102$$

即自私者人數：內奸人數 = 637 : 102，比值約為 6.245，將此結果與模擬實驗的結果對比後，發現誤差較大。

由於實驗規則的複雜性，研究者發現難以使用矩陣精簡、完整地呈現模型運行規則，故在此不展開詳細討論。對於將馬可夫鏈理論應用在本研究的想法，研究者考慮將此列入後續延伸研究的範疇。

針對本研究，研究者又聯想到可用微分方程式分別展現各族群的人數變動規則，其核心概念是：每個族群的淨變化率，可視為「自身繁衍增長」減去「在各種交互作用中被消滅或犧牲」。在程式碼裡，不同族群的互動都可以抽成某些族群間的接觸率，以及接觸後成功或失敗的機率。

此微分方程式的解即為趨於穩定狀態時各族群的人數。得出各族群最後的人數佔比後，將其與實驗數據做對比。

接下來，研究者用以下符號表示隨時間變化的人數：

$G(t)$ ：時刻為 $t$ 時「好人」的人數 $T(t)$ ：時刻為 $t$ 時「內奸」的人數 $S(t)$ ：時刻為 $t$ 時「自私者」的人數	$H(t)$ ：時刻為 $t$ 時「英雄」的人數 $P(t)$ ：時刻為 $t$ 時「警長」的人數 $R(t)$ ：時刻為 $t$ 時「造謠者」的人數
---	---

可將族群的淨變化率寫成：

$$\frac{dG}{dt} = b_G G - [\beta_{G,R} G(G + H)] - [\mu_G (1 - \rho_{G \leftarrow (G+H)}) G] - \kappa_P PG - c_R RG$$

$$\frac{dT}{dt} = b_T T - [\mu_T (1 - \rho_{T \leftarrow (G+H)}) T] - \kappa_P PT - c_R RT$$

$$\frac{dS}{dt} = b_S S - [\mu_S (1 - \rho_{S \leftarrow H}) S] - c_R RT$$

$$\frac{dH}{dt} = b_H H - [\beta_{H,R} H(G + T + S + H)] - [\mu_H (1 - \rho_{H \leftarrow (H)}) H] - c_R RT$$

$$\frac{dP}{dt} = \beta_P P - \mu_H (1 - \rho_{H \leftarrow (H)}) P - \kappa_P P - c_R RP$$

$$\frac{dR}{dt} = \beta_R R - \mu_H (1 - \rho_{H \leftarrow (H)}) R - \kappa_P PR$$

其中：

$b_i$ ：繁衍機率

$\beta_{G,R}$ ：好人遇到好人、英雄時因提醒而死的機率

$\beta_{H,R}$ ：英雄遇到所有族群時因提醒而死的機率

$\mu_i$ ：族群出門覓食遇到怪物的基礎死亡率

$\rho_{i \leftarrow j}$ ：i 族群遇到 j 族群時被提醒的機率

$\kappa_P$ ：族群被警長擊殺的機率（含誤殺的懲罰）

$c_R$ ：被造謠者欺騙而無法繁殖的機率

若想求平衡態的解，也就是族群佔比變化趨於穩定時的人數，則：

$$\frac{dG}{dt} = \frac{dT}{dt} = \frac{dS}{dt} = \frac{dH}{dt} = \frac{dP}{dt} = \frac{dR}{dt} = 0$$

經過縝密的計算，可得最終平衡態時各族群人數比：

好人：自私者：內奸：英雄：警長：造謠者

$\approx 33: 422327: 114989: 0: 12: 0$

將佔比過少的好人和警長去除，得到自私者：內奸的比值為：

$$\frac{422327}{114989} \approx 3.6727$$

且研究者實驗得到的數據為：

$$\frac{78.3}{21.7} \approx 3.6082$$

誤差：0.0645

圖 9：實驗一結果

```
==== 各族群最終平均占比 ===
> print(final_proportions)
# A tibble: 6 × 2
  Group  Final_Proportion
  <chr>      <dbl>
1 自私者     78.3
2 內奸       21.7
3 英雄        0
4 好人        0
5 警長        0
6 造謠者      0
```

圖片來源：研究者實驗

此結果與實驗法所得數據接近，即可驗證研究者所設計之程式碼邏輯正確無誤，表示可以此實驗方法作為基礎，繼續進行後續實驗。

## 2. 實驗二：辨識與獎勵機制對人類、AI 所選族群的生存影響

### (1) 研究目的

探討人類、AI 所選的族群在辨識、獎勵機制下之生存趨勢。

### (2) 研究方法：實驗法

### (3) 實驗內容

根據模型基礎規則並依族群的行動與外表分為六個族群，族群間有辨識、獎勵機制。將六大族群分別以調查法統計出的數據作為初始人數佔比，代入程式中，並分別生成人類、AI 的生存趨勢圖。

表 8：實驗二族群、機制整理

族群	好人	內奸	英雄	自私者	警長	造謠者
外貌	外表甲	外表甲	外表乙	外表乙	外表乙	外表丙
提醒	外表甲	不提醒	所有人	不提醒	不提醒	不提醒
機制	獎勵：提醒他人並生存下來者夜間可繁殖兩個個體					

## 3. 實驗三：辨識與改過自新機制對人類、AI 所選族群的生存影響

### (1) 研究目的

探討人類、AI 所選族群在辨識、改過自新機制下生存趨勢。

### (2) 研究方法：實驗法

### (3) 實驗內容

根據模型基礎規則並依族群的行動與外表分為六個族群，族群間有辨識、改過自新機制。將六大族群分別以調查法統計出的數據作為初始人數佔比，代入程式中，並分別生成人類、AI 的生存趨勢圖。

表 9：實驗三族群、機制整理

族群	好人	內奸	英雄	自私者	警長	造謠者
外貌	外表甲	外表甲	外表乙	外表乙	外表乙	外表丙
提醒	外表甲	不提醒	所有人	不提醒	不提醒	不提醒
機制	改過自新：內奸遇到警長後有一定機率轉變為自私者族群					

#### 4. 實驗四：辨識、獎勵、改過自新機制對人類、AI 所選族群的生存影響

##### (1) 研究目的

探討人類、AI 在辨識、獎勵、改過自新機制下之生存趨勢。

##### (2) 研究方法：實驗法

##### (3) 實驗內容

根據模型基礎規則並依族群的行動與外表分為六個族群，族群間有辨識、獎勵、改過自新機制。將六大族群分別以調查法統計出的數據作為初始人數佔比，代入程式中，並分別生成人類、AI 的生存趨勢圖。

表 10：實驗四族群、機制整理

族群	好人	內奸	英雄	自私者	警長	造謠者
外貌	外表甲	外表甲	外表乙	外表乙	外表乙	外表丙
提醒	外表甲	不提醒	所有人	不提醒	不提醒	不提醒
機制						獎勵：提醒他人並生存下來者夜間可繁殖兩個個體 改過自新：內奸遇到警長後有一定機率轉變為自私者族群

表 11：實驗一～實驗四之內容統整與歸納（「O」表示有，「—」表示無）

實驗	機制			族群					
	辨識 機制	獎勵 機制	改過自新 機制	好人	內 奸	英 雄	自 私	警 長	造 謠
一	O	—	—						
二	O	O	—		O	O	O	O	O
三	O	—	O						
四	O	O	O						

##### ■ 備注

- (1) 每種機制組合分別生成人類、人工智慧的生存結果，共計 8 張趨勢圖。
- (2) 所有實驗皆運行 2000 次，將結果取平均值以生成趨勢圖。
- (3) 所有實驗皆運行至平衡穩定狀態才停止（統一取 150 天）。

## 肆、研究結果

### 一、實驗一：辨識機制對人類、AI 所選族群的生存影響

#### (一) 人類

圖 10：人類所選族群在有辨識機制時之生存趨勢圖



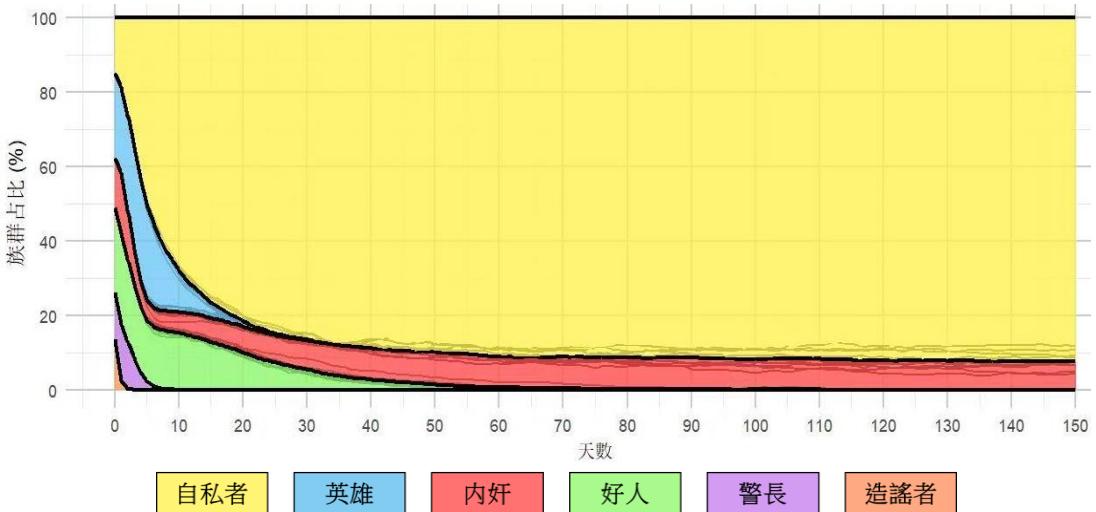
圖片來源：研究者實驗

#### 1. 實驗觀察

人類選擇好人族群較多，但好人因會提醒他人，在前期滅絕。英雄也因提醒他人而滅絕。警長與造謠者族群初始佔比較少，也在前期滅絕。推測如果將時間延長，自私者會與內奸達平衡。

#### (二) AI

圖 11：AI 所選族群在有辨識機制時之生存趨勢圖



圖片來源：研究者實驗

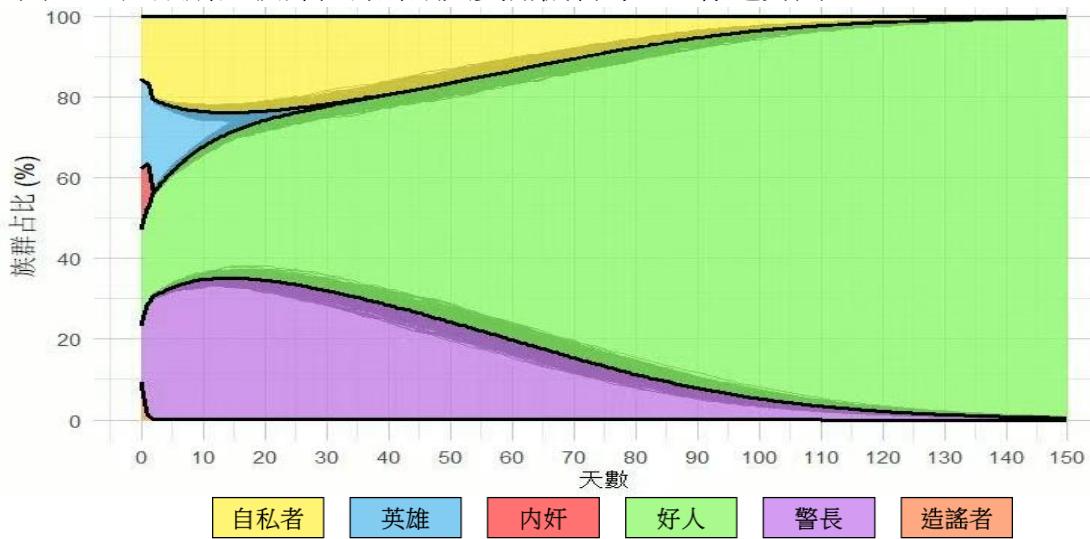
#### 1. 實驗觀察

在只有辨識機制的情況下，AI 選擇好人跟英雄族群較多，但因提醒其他族群導致族群佔比下降迅速，後期自私者、內奸存活。推測如果將時間延長，內奸和自私者會達平衡。

## 二、實驗二：辨識與獎勵機制對人類、AI所選族群的生存影響

### (一) 人類

圖 12：人類所選族群在有辨識與獎勵機制時之生存趨勢圖



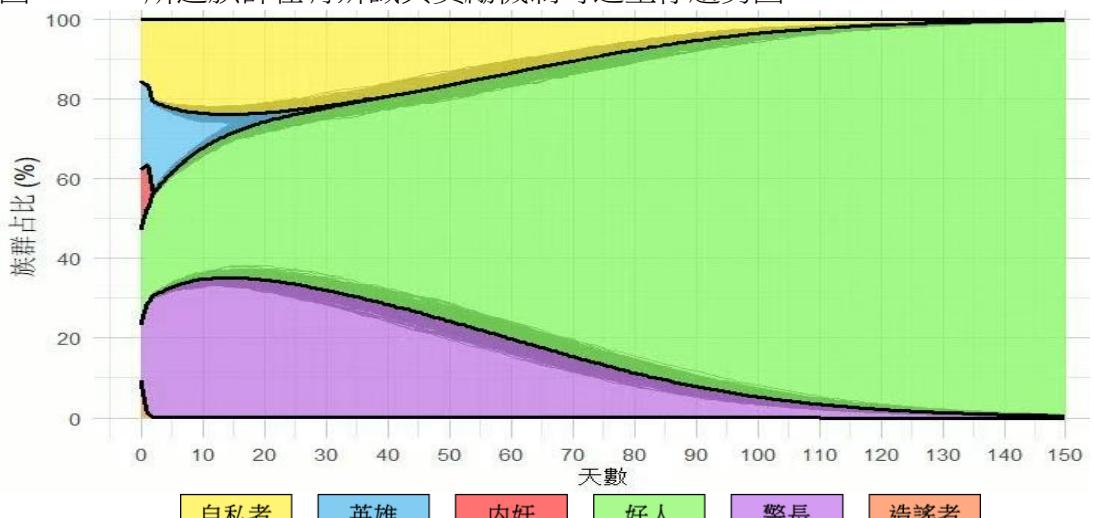
圖片來源：研究者實驗

#### 1. 實驗觀察

人類傾向於選擇好人，在獎勵機制的優勢下，變成環境的優勢族群。英雄因提醒其他族群而犧牲，導致滅絕。內奸與造謠者初始佔比較少，被警長槍斃而造成滅絕。警長、自私者、好人互不干擾，因而達平衡。推測如果將時間延長，好人會因有獎勵機制的優勢，族群佔比不斷上升，而自私者與警長少量存活。

### (二) AI

圖 13：AI 所選族群在有辨識與獎勵機制時之生存趨勢圖



圖片來源：研究者實驗

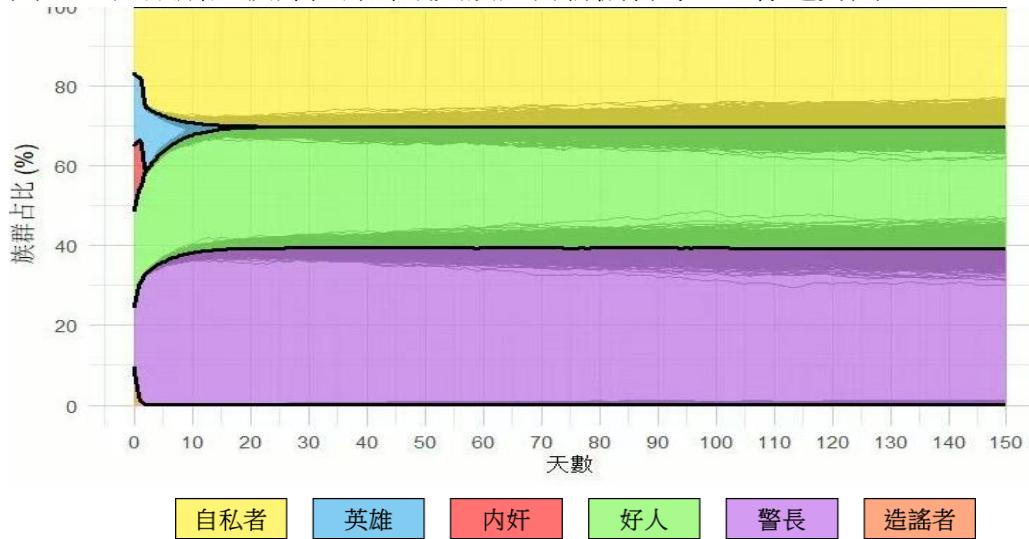
#### 1. 實驗觀察

AI 更偏好選擇英雄，但因英雄族群提醒其他族群造成犧牲而導致滅絕。警長在槍斃內奸與造謠者後人口佔比上升。好人因內奸滅絕與獎勵機制的優勢，佔比大量上升。推測若將時間延長，好人族群佔比會持續成長。

### 三、實驗三：辨識與改過自新機制對人類、AI所選族群的生存影響

#### (一) 人類

圖 14：人類所選族群在有辨識與改過自新機制時之生存趨勢圖



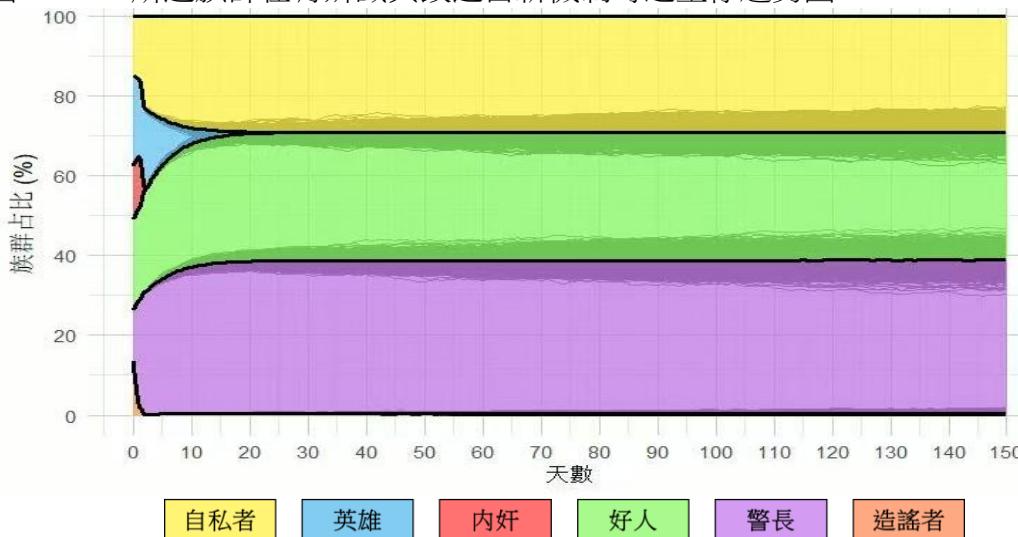
圖片來源：研究者實驗

#### 1. 實驗觀察

人類偏好選擇好人族群，好人族群因初始佔比較多，並且內奸與造謠者族群初始佔比少，人口佔比穩定。英雄族群在前期滅絕，而警長、自私者互不干涉，族群佔比維持穩定。推測如果將時間延長，自私者、好人、警長族群因互不干涉而呈平衡。

#### (二) AI

圖 15：AI 所選族群在有辨識與改過自新機制時之生存趨勢圖



圖片來源：研究者實驗

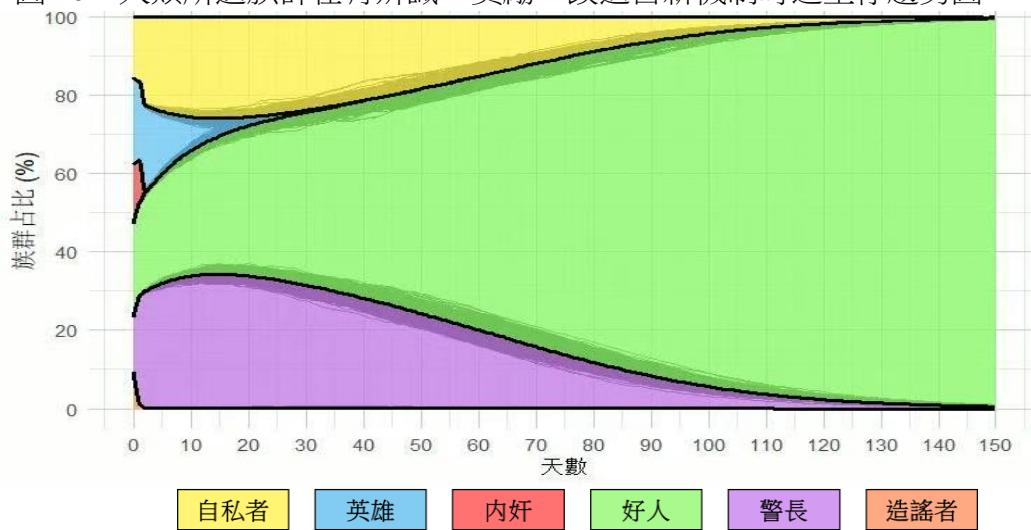
#### 1. 實驗觀察

在 AI 的選擇中，內奸初始佔比較人類選擇多，自私者受到「內奸族群變自私者」(改過自新) 機制的影響，族群佔比升高。英雄、造謠者、內奸族群在前期滅絕。推測如果將時間延長，自私者、好人、警長族群會因互不影響而達成平衡。

#### 四、實驗四：辨識、獎勵、改過自新機制對人類、AI所選族群的生存影響

##### (一) 人類

圖 16：人類所選族群在有辨識、獎勵、改過自新機制時之生存趨勢圖



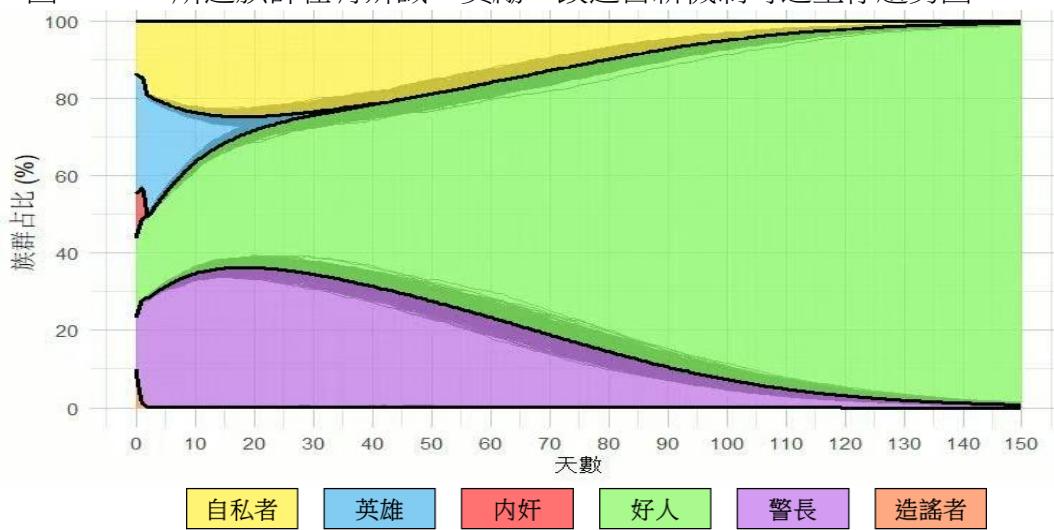
圖片來源：研究者實驗

##### 1. 實驗觀察

因為有獎勵機制的優勢，好人成為環境的優勢族群。英雄則因提醒其他族群造成犧牲，導致滅絕。內奸與造謠者初始佔比較少，並且內奸因「改過自新」的機制與被警長槍斃的規則而導致快速滅絕。警長、自私者、好人互不干擾而呈現平衡。推測如果將時間延長，好人會因有獎勵機制的優勢，族群佔比不斷上升，而自私者與警長僅少量存活。

##### (二) AI

圖 17：AI 所選族群在有辨識、獎勵、改過自新機制時之生存趨勢圖



圖片來源：研究者實驗

##### 1. 實驗觀察

雖然 AI 選擇英雄的傾向較人類強烈，但英雄族群因提醒其他族群造成犧牲，在前期就滅絕。警長在槍斃內奸與造謠者後人口佔比上升。好人因內奸滅絕與獎勵機制的優勢，人口佔比大量上升。自私者前期受英雄幫助，族群佔比上升。推測如果將時間延長，好人族群佔比會持續成長。

## 伍、討論

### 一、實驗結果分析

#### (一) 實驗一：**辨識機制**對人類、AI 所選族群的生存影響

##### 1. 實驗解析

###### (1) 人類

人類選擇好人族群較多，但好人因會提醒他人造成犧牲，並且因為人口佔比較多，會被警長誤殺，在前期滅絕。而警長初始佔比較少，誤殺好人後同歸於盡導致滅絕。英雄也因提醒他人犧牲而滅絕。造謠者族群初始佔比較少，被警長槍斃後滅絕。也在前期滅絕。推測如將時間延長，內奸因前期受到被好人的提醒的優勢，族群佔比上升，最後自私者與內奸達平衡。

###### (2) AI

在只有辨識機制的情況下，AI 選擇好人跟英雄族群較多。好人提醒外表甲的個體，幫助內奸族群佔比大量上升。但好人族群初始佔比較多，警長誤殺加上提醒他人造成的犧牲，好人滅絕。英雄因提醒所有需摘果實的族群，犧牲過多導致滅絕。而警長初始佔比較少，誤殺好人後同歸於盡導致滅絕。自私者前期受英雄幫助，族群佔比上升。推測如將時間延長，內奸與自私者會達平衡。

##### 2. 小結

優勢族群：自私者

#### (二) 實驗二：**辨識與獎勵機制**對人類、AI 所選族群的生存影響

##### 1. 實驗解析

###### (1) 人類

人類偏好選擇好人，因提醒外表甲的個體，在獎勵機制的優勢下，變成環境的優勢族群。但英雄因提醒族群較多，過多犧牲導致滅絕。內奸與造謠者初始佔比較少，雖內奸有被好人提醒的優勢，但因被警長槍斃滅絕。警長也因槍斃內奸，人口佔比上升。警長、自私者、好人互不干擾達平衡。推測如將時間延長，好人會因有獎勵機制的優勢，族群佔比不斷上升，而自私者與警長不會滅絕，但佔比極少。

###### (2) AI

AI 更偏好選擇英雄，英雄族群前期因有獎勵機制，族群佔比少量上升，但因提醒族群較多，造成犧牲導致滅絕。內奸與造謠者族群初始佔比較少，警長在槍斃內奸與造謠者後人口佔比上升，在內奸與造謠者滅絕後，警長佔比逐漸下降。好人因內奸滅絕與獎勵機制的優勢，人口佔比大量上升。推測如將時間延長，好人族群佔比會持續成長，警長與自私者會少量存活。

##### 2. 小結

優勢族群：好人

### (三) 實驗三：辨識與改過自新機制對人類、AI所選族群的生存影響

#### 1. 實驗解析

##### (1) 人類

人類偏好選擇好人族群，好人族群因初始佔比較多，並且內奸與造謠者族群初始佔比少，對於好人族群影響減少，因此好人族群人口佔比趨近穩定。英雄族群因提醒其他族群犧牲，導致在前期就滅絕。而警長前期槍斃內奸，族群佔比稍有增長，在內奸滅絕後，人口佔比穩定。自私者族群因有內奸變換族群的機制，人口佔比在前期增長。在英雄族群滅絕後，剩餘的好人、自私者、警長族群互不干涉，族群佔比維持穩定。推測如果將時間延長，自私者、好人、警長族群之佔比將平衡。

##### (2) AI

AI的選擇中，內奸初始佔比較人類選擇多，自私者受到內奸族群變換的機制影響，族群佔比在前期升高。英雄族群在因提醒其他族群造成自己犧牲，在前期滅絕。內奸被警長槍斃，或轉變成自私者，導致內奸族群滅絕。造謠者初始佔比較少，加上警長族群槍斃造謠者，造謠者在前期滅絕。推測如果將時間延長，自私者、好人、警長族群會因互不影響而達成平衡。

#### 2. 小結

優勢族群：好人

### (四) 實驗四：辨識、獎勵、改過自新機制對人類、AI所選族群的生存影響

#### 1. 實驗解析

##### (1) 人類

因為有獎勵機制的優勢，且內奸族群不斷減少，好人成為環境的優勢族群。英雄因提醒其他族群，在前期犧牲導致滅絕。內奸與造謠者初始佔比較少，並且內奸因改過自新機制，會轉變為自私者，加上被警長槍斃的風險導致滅絕。最後警長、自私者、好人互不干擾達平衡。推測如將時間延長，好人會因有獎勵機制的優勢，族群佔比不斷上升，而自私者與警長不會滅絕但少量存活。

##### (2) AI

雖有獎勵機制，但英雄族群因提醒其他族群造成犧牲大於提醒所獲得的繁衍機會，在前期便滅絕。造謠者初始佔比少，因被警長槍斃而滅絕。警長在槍斃內奸與造謠者後人口佔比上升，自私者因其他族群滅亡，加上內奸改過自新機制，族群佔比在前期增多。好人因內奸滅絕與獎勵機制的優勢，人口佔比大量上升。自私者前期受英雄幫助，族群佔比上升。推測如將時間延長，好人族群之佔比會持續成長，而自私者與警長不會滅絕但少量存活。

#### 2. 小結

優勢族群：好人

## 二、問卷分析

研究者原計劃分析性別、年齡、教育背景對族群選擇偏好的影響，但由於受試者年齡、教育背景分佈懸殊，因此不適合進行問卷分析，故研究者在此以性別作為主要探討因素。

以下以實驗一中男女對於英雄族群選擇的檢定過程為例：

(一) 女性

$$N_F = 214$$

表 12：實驗一女性的族群選擇

排序	1	2	3	4	5	6
選擇人數	0	62	48	29	25	29

(二) 男性

$$N_M = 186$$

表 13：實驗一男性的族群選擇

排序	1	2	3	4	5	6
選擇人數	0	39	43	21	16	48

代入以上數據可得：

$$t_{0.025}(399) \approx 1.96, |t| > 1.96, p < 0.05$$

故英雄族群的男女偏好存在顯著差異。

依照上述方式計算所有族群：

表 14：實驗一檢定結果

族群	英雄	自私者	好人	內奸	警長	造謠者
$ t $	2.31	10.2	1.2	1.1	3.2	0.3
$p$	$< 0.05$	$< 0.05$	$> 0.05$	$> 0.05$	$< 0.05$	$> 0.05$

### (三) 分析

男女對英雄、自私者、警長族群的選擇偏好具有顯著差異。

女性更偏向選擇英雄的合作策略，考慮到現實生活中女性可能更重視關係穩定性，避免因不合作引發人際矛盾。而選擇自私者可能是因為傾向減少犧牲的風險，對於不提醒的策略容忍度更高。而男性更傾向將不提醒視為不合作或背叛，所以較少選擇。男性更偏向選擇警長。

推測可能男性可能更傾向於高風險高收益的策略，認為槍斃內奸可造福環境，而女性可能更重視有誤殺好人而破壞環境平衡的風險。而好人族群為男女共同最優先級的選擇。推測可能是群體認同與倫理道德等心理性動機使性別差異在強烈的群體歸屬感面前被弱化。

### 三、數據挖掘

根據數學課所學，研究者嘗試以馬可夫鏈理論計算出平衡態時的族群人數比例。以下為研究者的詳細推導過程：

因自私者的存活率為 89.8%，死亡率為 10.2%；內奸的存活率為 63.7%，死亡率為 36.3%。故研究者在此設轉移矩陣為：

$$A = \begin{bmatrix} 0.898 & 0.637 \\ 0.102 & 0.363 \end{bmatrix}$$

找  $P$  使  $P^{-1}AP = D$ ， $A = PDP^{-1}$ ，則：

$$A^n = PDP^{-1}PDP^{-1} \dots PDP^{-1} = PD^nP^{-1}$$

找  $A\vec{v} = \lambda\vec{v}$ ， $(A - \lambda I)\vec{v} = \vec{0}$  的非零解：

$$\begin{bmatrix} 0.898 - \lambda & 0.637 \\ 0.102 & 0.363 - \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

求該矩陣的特徵多項式：

$$\det(A - \lambda I) = (0.898 - \lambda)(0.363 - \lambda) - (0.637)(0.102) = 0$$

可得兩個特徵值：

$$\lambda_1 = 1, \lambda_2 = 0.261$$

對於  $\lambda_1 = 1.0$

$$\begin{bmatrix} 0.898 - 1 & 0.637 \\ 0.102 & 0.363 - 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 637 \\ 102 \end{bmatrix}$$

對於 $\lambda_2 = 0.261$

$$\begin{bmatrix} 0.898 - 0.261 & 0.637 \\ 0.102 & 0.363 - 0.261 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

取 $P = \begin{bmatrix} 637 & 1 \\ 102 & -1 \end{bmatrix}$ ，得到 $P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & 0.261 \end{bmatrix}$ ，此時 $A^n = P \begin{bmatrix} 1^n & 0 \\ 0^n & 0.261^n \end{bmatrix} P^{-1}$

當 $n$ 趨近於無窮大時， $0.261^n$  趨近於0，因此：

$$\lim_{n \rightarrow \infty} A^n = P \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} P^{-1} = \begin{bmatrix} 637 & 1 \\ 102 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \left( \frac{1}{739} \begin{bmatrix} 1 & 1 \\ 102 & -673 \end{bmatrix} \right) = \frac{1}{739} \begin{bmatrix} 637 & 637 \\ 102 & 102 \end{bmatrix}$$

故可得自私者人數：內奸人數= 637: 102

研究者發現馬可夫鏈的推導結果與實驗結果相差甚遠，推論是由於馬可夫鏈的平衡態推算旨在計算兩個狀態下的來回轉換情況，而研究者實驗中的出生與死亡情況與單純「轉移」的概念有差異（例如死亡的自私者並不會轉移到存活的內奸族群），因此並不適用於常規的 $2 \times 2$  轉移矩陣。

在未來的後續研究，研究者希望能找出可以套用在如本實驗情況下的轉移矩陣，將馬可夫鏈理論更加精確地應用在此社會模型，繼續加強數學推導與驗證。

#### 四、社會應用

根據環境中族群的交互行為，可以將此模型套用到真實社會中的市場競爭案例。

好人族群相當於社會中的合作型企業，內奸相當於假意投資者，造謠者族群類似真實社會中的惡意競爭者，而警長族群類似監管機構。

辨識機制相當於品牌標示或技術路線，可以讓企業間識別對方為同類或異類企業，進而決定是否幫助或提醒對方市場的走向或陷阱。

研究者建議政府應在市場中鼓勵合作，補貼合作型企業（與實驗中獎勵機制相符），也促使更多企業互幫互助，形成良性競爭，提升產品的品質與研發技術。在造福消費者的同事也幫助經濟成長，減少惡意競爭者的存在。

# 陸、結論

## 一、研究結論

### (一) 實驗一：辨識機制對人類、AI 所選族群的生存影響

在只有外表辨識的機制下，人類和人工智慧都更偏向選擇好人族群，但人類選擇警長族群的佔比明顯較少。

對比人類與人工智慧的選擇，自私者受其他族群影響最少，且前期會被英雄提醒，可以存活到後期，成為環境優勢族群。

建議選擇自私者族群，保障自己生存的同時又不會陷害其他族群。

### (二) 實驗二：辨識與獎勵機制對人類、AI 所選族群的生存影響

加入獎勵機制後，人類與人工智慧都更偏向選擇好人。

內奸與造謠者初始佔比微小，而又因警長的存在滅絕。英雄提醒族群較多，造成犧牲，導致滅絕。自私者與其他族群互動少，前期有英雄族群幫助，人口佔比會少量上升。好人有著獎勵機制的優勢，成為環境的優勢族群。

建議選擇好人族群，在合作的環境下共存，或者選擇自私者族群，保障自己的生存。

### (三) 實驗三：辨識與改過自新機制對人類、AI 所選族群的生存影響

對比人工智慧與人類的選擇，人類更偏向選擇好人族群。

在沒有獎勵機制的優勢下，好人族群因沒有獎勵機制的優勢，佔比不會大量上升。自私者受到內奸轉變族群的優勢，族群佔比在前期上升後保持穩定。

建議選擇好人族群或自私者族群，警長會有誤殺好人而導致自我滅絕的風險，因此不建議選擇。

### (四) 實驗四：辨識、獎勵、改過自新機制對人類、AI 所選族群的生存影響

人類與人工智慧的選擇差異不大，因此結果相似。

內奸與造謠者因被警長槍毙而滅絕。自私者前期有英雄族群幫助，並且有內奸的改過自新機制的優勢，人口佔比會少量上升。好人受到獎勵機制的加成，成為環境的優勢族群。

建議選擇好人族群，在合作的環境下共存，成為環境的優勢種。

表 15：實驗一～實驗四中各機制組合下之優勢族群

實驗	一	二	三	四
機制	辨識	辨識、獎勵	辨識、改過自新	辨識、獎勵、改過自新
優勢族群 (人)	自私	好人	好人	好人
優勢族群 (AI)	自私	好人	好人	好人

## 二、未來展望與建議

### (一) 後續研究可改進之處

由於電腦的性能有限，故研究者需對實驗進行簡化，無法過多延展，此困境難免對研究造成影響。因此研究者盡力**優化實驗過程**，使電腦在有限的算力條件下盡量獲取有意義的數據。

但規模的限制仍讓此研究有許多可改善的部分，例如：模擬天數的擴展、實驗次數的增加、模型機制的複雜化等。秉持著對科學的探究精神，研究者考慮後續的硬體升級，進一步提升電腦的計算性能，以應對更高階的實驗需求。

此外。對於此次研究內容，研究者尚未針對馬可夫鏈理論展開詳細的討論，因此，在後續研究中可再**加強數學模型的推導**，以數據挖掘的方式探討出社會模型中各族群的行為模式，量化模型規則，使研究能夠以不同角度呈現。

至於**問卷調查**的部分，因研究者周遍以國中和高中生居多，後續研究可以嘗試在校園外部發放問卷，例如街邊隨機發放，擴展受測驗者的年齡範圍。

### (二) 建議

透過研究內容可以得知在沒有獎勵機制的情況下，不提醒他人更能保證族群的生存。在有獎勵機制的情況下，提醒他人可以在保障其他族群生存的同時讓自己的族群人數大幅增長。

此外，研究者發現 AI 和人類的選擇的結果差異極小。因此研究者建議人類應該**學著和 AI 共存**，但在特殊情況（例如實驗三與實驗四的改過自新機制），人類的選擇會更靈活。

在沒有獎勵機制的情況下，人類會更偏向選擇停止欺騙他人，在轉變成自私者後，也成功在環境中存活。而 AI 所選擇的內奸族群轉換率雖較低，但內奸族群在早期就因被警長剋制而滅絕。所以研究者推薦，在**選擇的過程中，雖可以參考 AI 的思路，但應更堅信自己的選擇**。

## 柒、參考文獻資料

- 崔坤榆、翁晨軒、蕭琦穎（2025）。生存或毀滅？探討人類與 AI 在社會模擬遊戲中的選擇。中學生網站，小論文第 1140315 梯次。取自 <https://www.shs.edu.tw/>
- Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, 328(5978), 617–620. Retrieved from <https://doi.org/10.1126/science.1183665>
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1-15.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118. Retrieved from <https://doi.org/10.1038/nature21056>
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4), 980–994. Retrieved from <https://doi.org/10.1257/aer.90.4.980>
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451-482. Retrieved from <https://doi.org/10.1146/annurev-psych-120709-145346>
- Hamilton, W. D. (1964). The genetical evolution of social behaviour I. *Journal of Theoretical Biology*, 7(1), 1–16. Retrieved from [https://doi.org/10.1016/0022-5193\(64\)90038-4](https://doi.org/10.1016/0022-5193(64)90038-4)
- Hamilton, W. D. (1964). The genetical evolution of social behaviour II. *Journal of Theoretical Biology*, 7(1), 17–52. Retrieved from [https://doi.org/10.1016/0022-5193\(64\)90039-6](https://doi.org/10.1016/0022-5193(64)90039-6)
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-292. Retrieved from <https://doi.org/10.2307/1914185>
- Nash, J. F. (1950). Equilibrium points in  $n$ -person games. *Proceedings of the National Academy of Sciences*, 36(1), 48–49. Retrieved from <https://doi.org/10.1073/pnas.36.1.48>
- Primer. (2021, March 30). *Simulating green-beard altruism* [Video]. YouTube. Retrieved from <https://www.youtube.com/watch?v=goePYJ74Ydg>
- Startup, R. (2021). The theory of the selfish gene applied to the human population. *Advances in Anthropology*, 11(3), 179–200. Retrieved from <https://doi.org/10.4236/aa.2021.113012>

## 附件：問卷內容

### 一、同意聲明

您好！非常感謝您撥冗參與本研究問卷調查。在您開始填答之前，請花一點時間閱讀以下重要資訊：

- (一) 匿名性：本問卷採「完全匿名」方式進行。問卷設計確保不會收集任何能直接或間接識別您個人身份（如姓名、身份證號、聯絡方式等）的資訊。
- (二) 數據用途：您在問卷中的所有填答內容，將僅作為本次研究的實驗數據使用，目的在於分析、理解研究主題的相關現象。
- (三) 自願參與：參與本研究完全是自願的。您沒有義務必須填寫此問卷。
- (四) 同意聲明：繼續進行問卷填寫或提交完成的問卷，即代表您已閱讀並理解本同意書的內容，也充分瞭解本問卷為匿名調查，並同意將您提供的所有回答作為本研究的實驗數據供研究者分析使用。

### 二、問卷題目

- (一) 請問您的性別是？

- 1. 男
- 2. 女

- (二) 請問您的年齡為？

- 1. 11 歲及以下
- 2. 12~18 歲
- 3. 19~25 歲
- 4. 26~35 歲
- 5. 36~45 歲
- 6. 46~55 歲
- 7. 56~65 歲
- 8. 66 歲及以上

- (三) 請問您的教育背景為？

- 1. 小學及以下
- 2. 國中
- 3. 高中/職高
- 4. 大學
- 5. 研究所及以上

- (四) 請閱讀以下文字後，回答問題：

在一環境內，存在著許多怪物與 6 大族群，大家需要出去摘取果實以獲得繁衍族群的機會。

在覓食的過程中，若別人遇到怪物（他不知道怪物將要殺害他），你可以選擇提醒或不提醒他，相對的，別人也可以選擇提醒或不提醒你。

- ◆ 如果提醒對方，則 50% 你和對方都逃走，50% 你犧牲自己救了對方。
- ◆ 每天摘取果實後若成功回家，可以使你的所屬族群增加一人。

1. 今天你生存在這個環境，以下六個族群，請根據選擇的偏好進行排序。  
(請選填數字 1~6，1是最想加入的族群，6是最不喜歡的族群。)
  - (1) A 族群:提醒所有人
  - (2) B 族群:不提醒任何人
  - (3) C 族群:只提醒與自己同族群的人
  - (4) D 族群:不管是誰都不提醒，但是會假裝成 C 族群，欺騙 C 族群的人提醒自己
  - (5) E 族群:每天 40% 機率槍斃一名 D 族群，但有 20% 可能把 C 認成 D 族群
  - (6) F 族群:傳遞錯誤果實位置讓別人摘不了果實，但每天 60% 機率被 E 槍斃
2. 衍生問題 (若在第 1 題中你將 D 族群排在了前三名，請回答這一題。若否，則選「我不想當 D 族群」。)
  - ◆ 情境：你選擇了 D 族群，雖然獲得被 C 族群提醒的優勢，但有著被 E 族群槍斃的風險。
  - ◆ 現在給你一個機會，請問你要不要停止欺騙 C 族群?  
(你失去被 C 族群提醒的優勢，但是 E 族群不再有槍斃你的可能。)
    - (1) 停止
    - (2) 不停止
    - (3) 我不想當 D 族群
3. 今新加一獎勵機制：提醒他人後，若存活，一次可以繁殖兩個。  
今天你生存在這個環境，以下六個族群，請根據選擇的偏好進行排序。  
(請選填數字 1~6，1是最想加入的族群，6是最不喜歡的族群。)
  - (1) A 族群:提醒所有人
  - (2) B 族群:不提醒任何人
  - (3) C 族群:只提醒與自己同族群的人
  - (4) D 族群:不管是誰都不提醒，但是會假裝成 C 族群，欺騙 C 族群的人提醒自己
  - (5) E 族群:每天 40% 機率槍斃一名 D 族群，但有 20% 可能把 C 認成 D 族群
  - (6) F 族群:傳遞錯誤果實位置讓別人摘不了果實，但每天 60% 機率被 E 槍斃
4. 衍生問題 (若在第 3 題中你將 D 族群排在了前三名，請回答這一題。若否，則選「我不想當 D 族群」。)
  - ◆ 情境：你選擇了 D 族群，雖然獲得被 C 族群提醒的優勢，但有著被 E 族群槍斃的風險。
  - ◆ 現在給你一個機會，請問你要不要停止欺騙 C 族群?  
(你失去被 C 族群提醒的優勢，但是 E 族群不再有槍斃你的可能。)
    - (1) 停止
    - (2) 不停止
    - (3) 我不想當 D 族群

## 【評語】052701

這份參賽作品《社會模擬模型：探討特定機制對不同族群之生存優勢影響》，以設計社會模擬模型的研究方法，結合數學模型、AI 技術與社會行為理論，探討人類受試者和五種人工智慧模型，在某特定情境之下，對於六種具不同行為特徵的角色（好人、內奸、英雄、自私者、警長、造謠者）之偏好，以及此偏好受不同機制（如可辨識不同角色的族群、獎勵、可改過自新等）的影響。

此一主題具有創新性，所採用的研究方法有系統地檢視人類和人工智慧所做的利己或利他生存選擇，值得肯定和鼓勵。以下提出對這個研究計畫的幾點疑問和建議，以及可繼續精進、加強之處：

1. 本研究在文獻回顧中可納入更多 AI 行為相關的最新研究，特別是探討 AI 決策偏見或倫理問題的文獻。文獻背景和現實生活情境的結合宜更加明確，以增加對研究內容的理解。對於過往文獻的"研究內容"之整理（表一），是其假設或是結論？是本研究據以設計的原則？
2. 本研究的主要問題，其理論背景較不明確；其結果特別針對性別進行分析，是否有清楚的理論依據？

3. 建議可聚焦於某一特定機制（如「改過自新機制」）或族群行為（如利他 vs. 利己），避免主題過於發散。明確定義每個機制的運作細節，並提供具體的數學或模擬參數，以增強研究的說服力。
4. 本研究的各個實驗其實為同一研究的不同情況 (conditions)，可使用推論統計方式直接檢驗在各情境/機制的影響下、人類和 AI 模型所產生行為的異同。
5. 本研究採用 5 種 AI 模型，應具體報告各自的結果，並說明不同 AI 模型的結果是否一致，以及在報告中所呈現的結果以何者為準。
6. 可深化理論與實務的連結，在結論部分提供更具體的社會應用案例，例如將模型應用於資源分配、企業合作或政策制定等情境，並說明如何將模擬結果轉化為實際建議。未來可考慮將模型應用於跨文化或跨族群的行為研究，探討文化背景對利他/利己行為的影響。

作品海報

# 社會模擬模型

探討特定機制對不同族群之生存趨勢影響

# 摘要

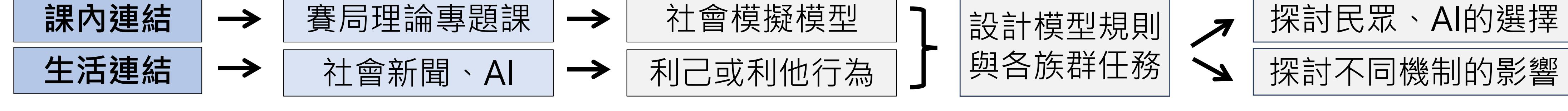
本研究旨在探討人類、AI在特定機制下的抉擇。

研究者首先固定一場景模型，**設計模型**中各族群的任務與環境機制，再使用程式還原所研發的模型，並利用微分方程式去驗證代碼運行的正確性。

接下來研究者發表**問卷**調查民眾、AI對不同族群的選擇偏好，再將調查數據代入程式，以比對人類與AI在選擇上的異同，並探討各種機制對於族群間互動的影響，加以推論原因與發展趨勢，最後說明此模型在**生活中的應用**。

## 壹、前言

### 一、研究動機



### 二、研究目的

- (一) 調查人類、AI在特定環境下之**選擇異同**。 (二) 探討人類、AI所選的族群在各種機制下之**生存趨勢**。  
(三) 歸納族群間的利害關係與各情況之優勢族群，分析此模型在現實**社會中的應用**。

### 三、文獻回顧-模型基礎規則設定

#### 場景設定

1. 食物：使族群繁衍之主要資源。  
2. 怪物：無差別隨機攻擊覓食者。  
3. 族群：依行為、外表做區別，有各自的任務。

#### 基礎原則

1. 每天日間除警長外，所有人須出門覓食一次。  
2. 出門時必遇見怪物，50% 被怪物殺害。  
3. 覓食時必遇到其他人，種族隨機。  
4. 覓食成功者夜間繁殖一個個體。

表一：基礎角色設定

族群	好人	內奸	英雄	自私者
外貌	外表甲	外表甲	外表乙	外表乙
提醒	提醒外表甲	不提醒任何人	提醒所有人	不提醒任何人

#### 提醒機制

1. 覓食時會看見別人附近的怪物，可選擇是否提醒對方。  
2. 若提醒則對方必逃脫，而自己50% 因提醒他人而死亡。  
3. 若不提醒，則個體與對方皆有50% 的機率被怪物殺害。  
4. 所有人**只能辨別外表**，無法得知對方所屬族群。

## 貳、研究設備及器材

表二：研究設備與功能概述

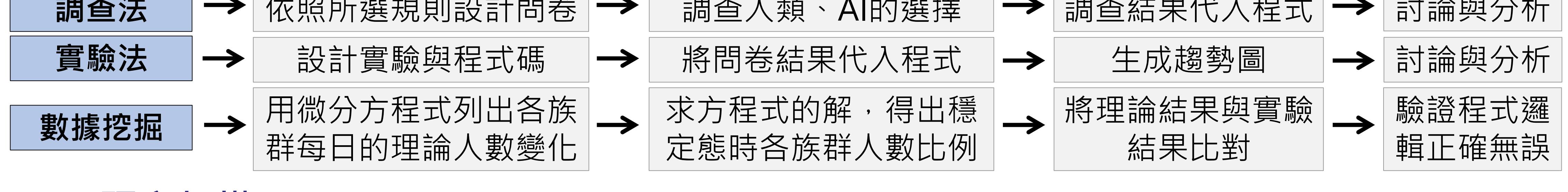
電腦軟體	功能
線上問卷	調查人類、AI之偏好族群，將問卷結果代入程式。
人工智能	共選取五款AI，詢問其選擇，並與人類比較。
R語言	進行程式設計、實驗模擬、趨勢圖生成。

表三：研究者所選之人工智慧

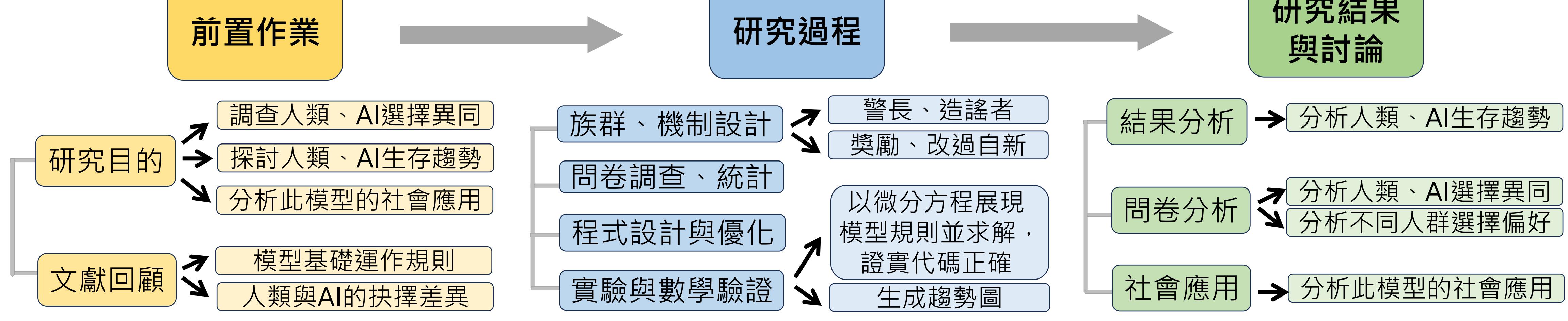
人工智慧				
Llama-3-Groq-70B	Claude-3.7-Sonnet	Gemini-2.0-Flash	Chat-GPT o1	Grok-2

## 參、研究過程或方法

### 一、實驗方法



### 二、研究架構



## 三、研究過程

### (一) 實驗設計

#### 新增角色-警長

1. 警長為**外表乙**，每天有40% 機率開槍擊斃一名外表為甲者，60% 機率擊斃一名造謠者。  
2. **有20% 機率此外表甲者為好人**，若誤殺好人則警長與該名好人皆死亡，反之當晚警長繁殖一個個體。  
3. 不存在內奸時，警長不再開槍，改以覓食方式繁殖。  
4. 警長不提醒任何人危險的來臨。

#### 新增角色-造謠者

1. 造謠者為**外表丙**，每次出門會碰到一人。  
2. 若此人不屬於造謠者則進行欺騙，迫使此人前往沒有果實之處。  
3. **被欺騙者沒有果實，當天夜間無法繁殖**，但也不會因碰到怪物而死亡。  
4. 造謠者不提醒任何人危險的來臨。

#### 新增機制

##### 1. 獎勵機制

提醒他人並生存者夜間可**繁殖兩個個體**（內奸、自私者、警長、造謠者維持不提醒任何人的原則）。

##### 2. 改過自新機制

依問卷回收結果設定內奸的族群轉換率，使**內奸遇到警長後有一定機率停止欺騙行為，轉變為自私者**。此幾率為「願意停止欺騙他人者佔所有受試者的比例」。

表四：六大族群之外表及行為比較

族群	好人	內奸	英雄	自私者	警長	造謠者
外貌	外表甲	外表甲	外表乙	外表乙	外表乙	外表丙
提醒	提醒外表甲	不提醒	所有人	不提醒	不提醒	不提醒

表五：實驗一～實驗四之內容統整與歸納

實驗	機制	族群								
		辨識	獎勵	改過自新	好	內	英	自	警	造
一	O	—	—	—						
二	O	O	—	—						
三	O	—	O	—						
四	O	O	O	—						

## (二) 數據挖掘 (以實驗一人類組為例) 馬可夫鏈

設轉移矩陣為：

$$A = \begin{bmatrix} 0.898 & 0.637 \\ 0.102 & 0.363 \end{bmatrix}$$

找  $P$  使  $P^{-1}AP = D$ ， $A = PDP^{-1}$ ，則：

$$A^n = PDP^{-1}PDP^{-1} \dots PDP^{-1} = PD^nP^{-1}$$

找  $A\vec{v} = \lambda\vec{v}$ ， $(A - \lambda I)\vec{v} = \vec{0}$  的非零解：

$$\begin{bmatrix} 0.898 - \lambda & 0.637 \\ 0.102 & 0.363 - \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

求該矩陣的特徵多項式：

$$\det(A - \lambda I) = (0.898 - \lambda)(0.363 - \lambda) - (0.637)(0.102) = 0$$

可得兩個特徵值： $\lambda_1 = 1$ ,  $\lambda_2 = 0.261$

對於  $\lambda_1 = 1.0$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 637 \\ 102 \end{bmatrix}$$

對於  $\lambda_2 = 0.261$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

取  $P = \begin{bmatrix} 637 & 1 \\ 102 & -1 \end{bmatrix}$ ，得到  $P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & 0.261 \end{bmatrix}$

此時  $A^n = P \begin{bmatrix} 1^n & 0 \\ 0 & 0.261^n \end{bmatrix} P^{-1}$

當  $n$  趨近於無窮大時， $0.261^n$  趨近於 0，因此：

$$\lim_{n \rightarrow \infty} A^n = P \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} P^{-1} = \begin{bmatrix} 637 & 1 \\ 102 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \left( \frac{1}{739} \begin{bmatrix} 1 & 1 \\ 102 & -673 \end{bmatrix} \right) = \frac{1}{739} \begin{bmatrix} 637 & 637 \\ 102 & 102 \end{bmatrix}$$

故可得自私者人數：內奸人數 = 637 : 102

## 微分方程

$$\frac{dG}{dt} = b_G G - [\beta_{G,R} G(G + H)] - [\mu_G (1 - \rho_{G \leftarrow (G+H)}) G] - \kappa_P PG - c_R RG$$

$$\frac{dT}{dt} = b_T T - [\mu_T (1 - \rho_{T \leftarrow (G+H)}) T] - \kappa_P PT - c_R RT$$

$$\frac{dS}{dt} = b_S S - [\mu_S (1 - \rho_{S \leftarrow H}) S] - c_R RT$$

$$\frac{dH}{dt} = b_H H - [\beta_{H,R} H(G + T + S + H)] - [\mu_H (1 - \rho_{H \leftarrow H}) H] - c_R RT$$

$$\frac{dP}{dt} = \beta_P P - \mu_H (1 - \rho_{H \leftarrow H}) P - \kappa_P P - c_R RP$$

$$\frac{dR}{dt} = \beta_R R - \mu_H (1 - \rho_{H \leftarrow H}) R - \kappa_P PR$$

$\beta_{G,R}$ ：好人遇到好人、英雄時因提醒而死的機率

$\beta_{H,R}$ ：英雄遇到所有族群時因提醒而死的機率

$\rho_{i \leftarrow j}$ ：i族群遇到j族群時被提醒的機率

$\kappa_P$ ：族群被警長擊殺的機率 (含誤殺的懲罰)

$c_R$ ：被造謠者欺騙而無法繁殖的機率

當  $\frac{dG}{dt} = \frac{dT}{dt} = \frac{dS}{dt} = \frac{dH}{dt} = \frac{dP}{dt} = \frac{dR}{dt} = 0$  時

好人：自私：內奸：英雄：警長：造謠  $\approx 33: 422327: 114989: 0: 12: 0$

自私者：內奸的比值為  $\frac{422327}{114989} \approx 3.6727$

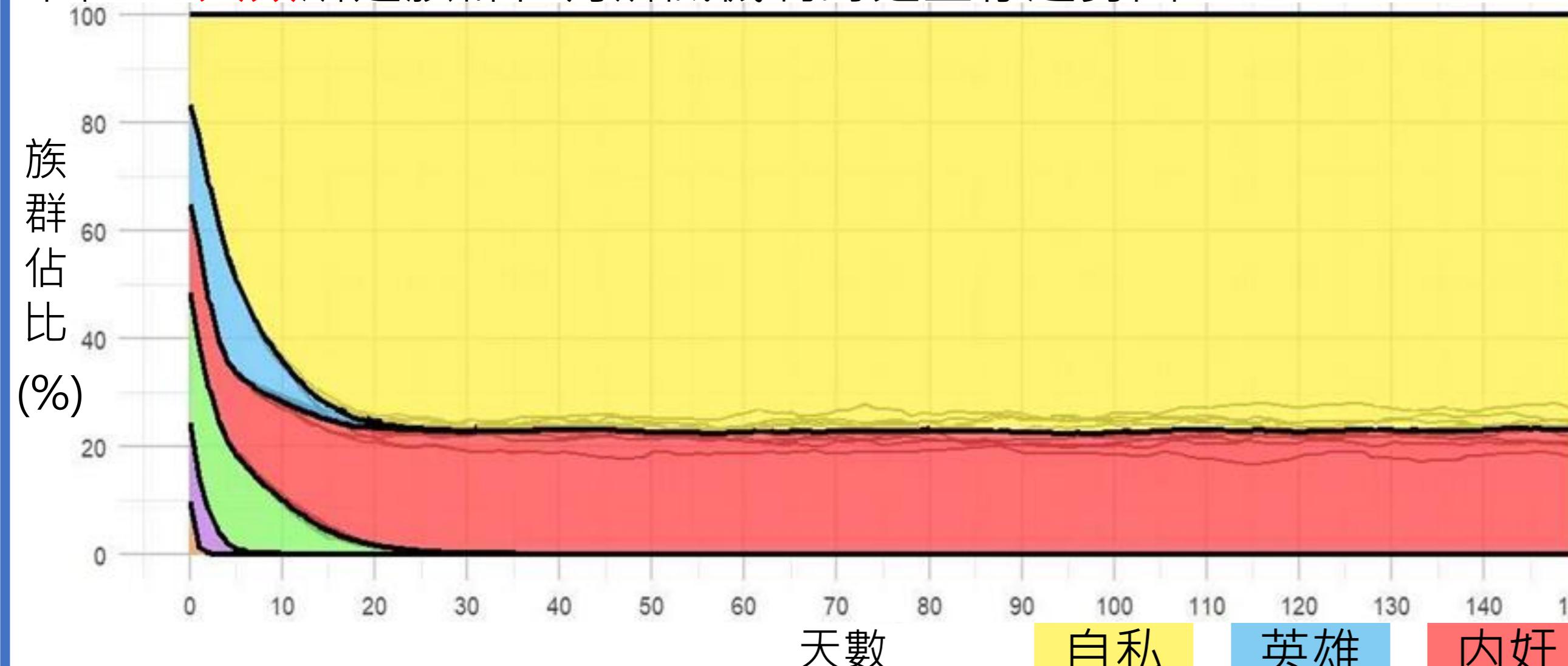
研究者實驗得到的數據為  $\frac{78.3}{21.7} \approx 3.6082$

誤差：0.0645

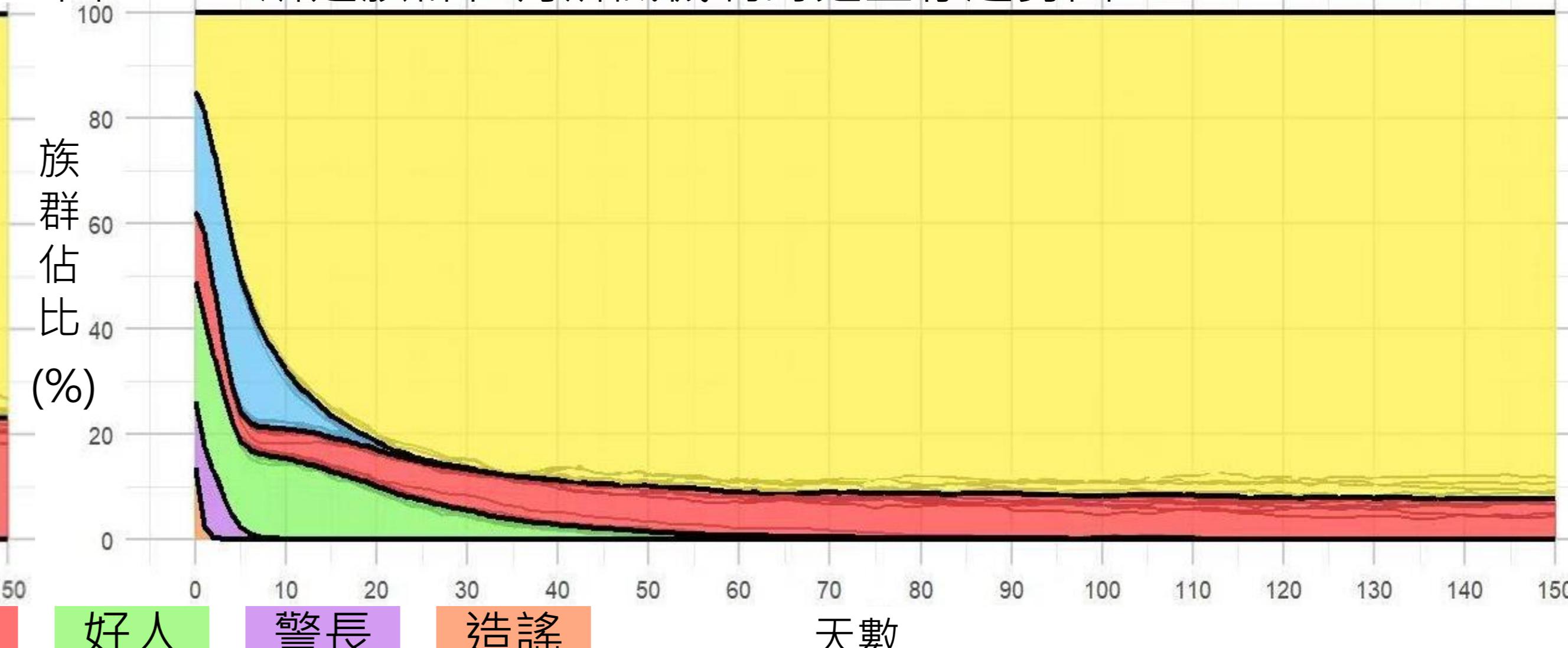
## 肆、研究結果

### 實驗一：辨識機制對人類、AI所選族群的生存影響

圖一：人類所選族群在有辨識機制時之生存趨勢圖

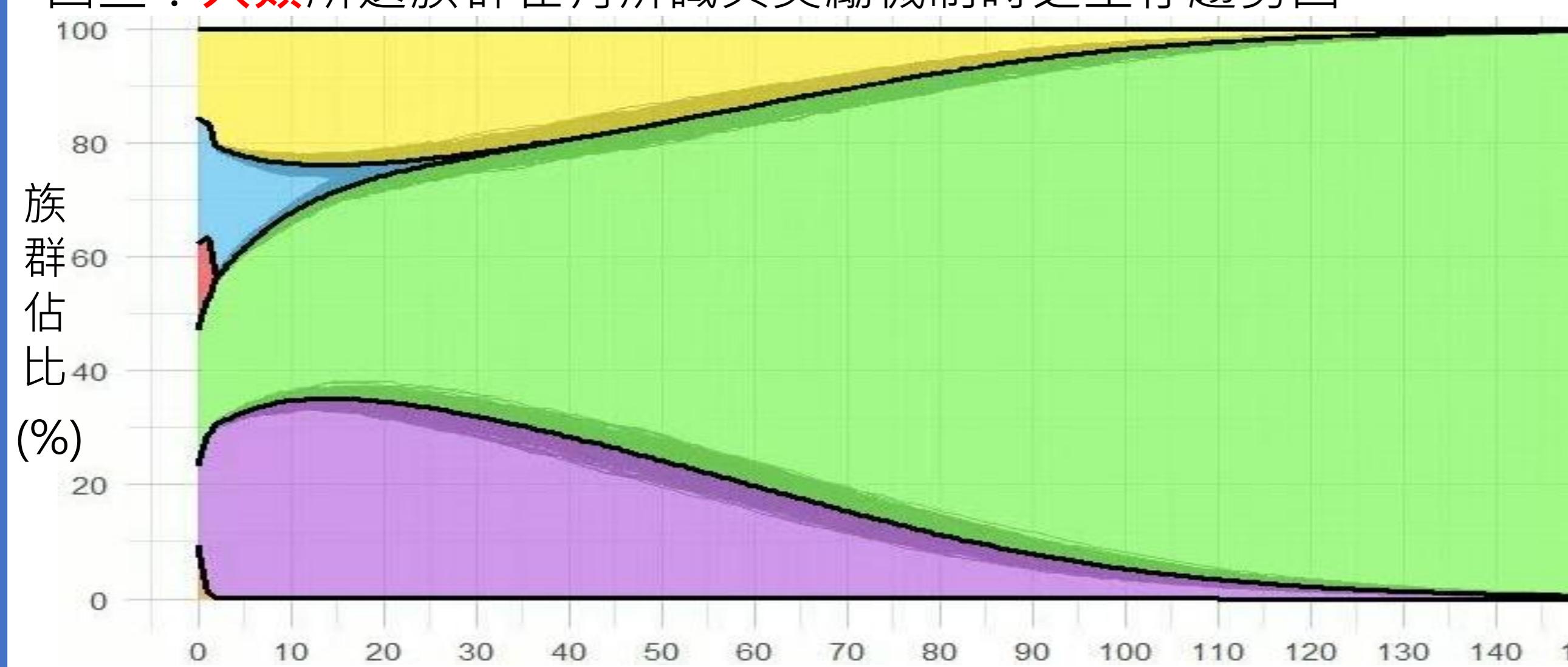


圖二：AI所選族群在有辨識機制時之生存趨勢圖

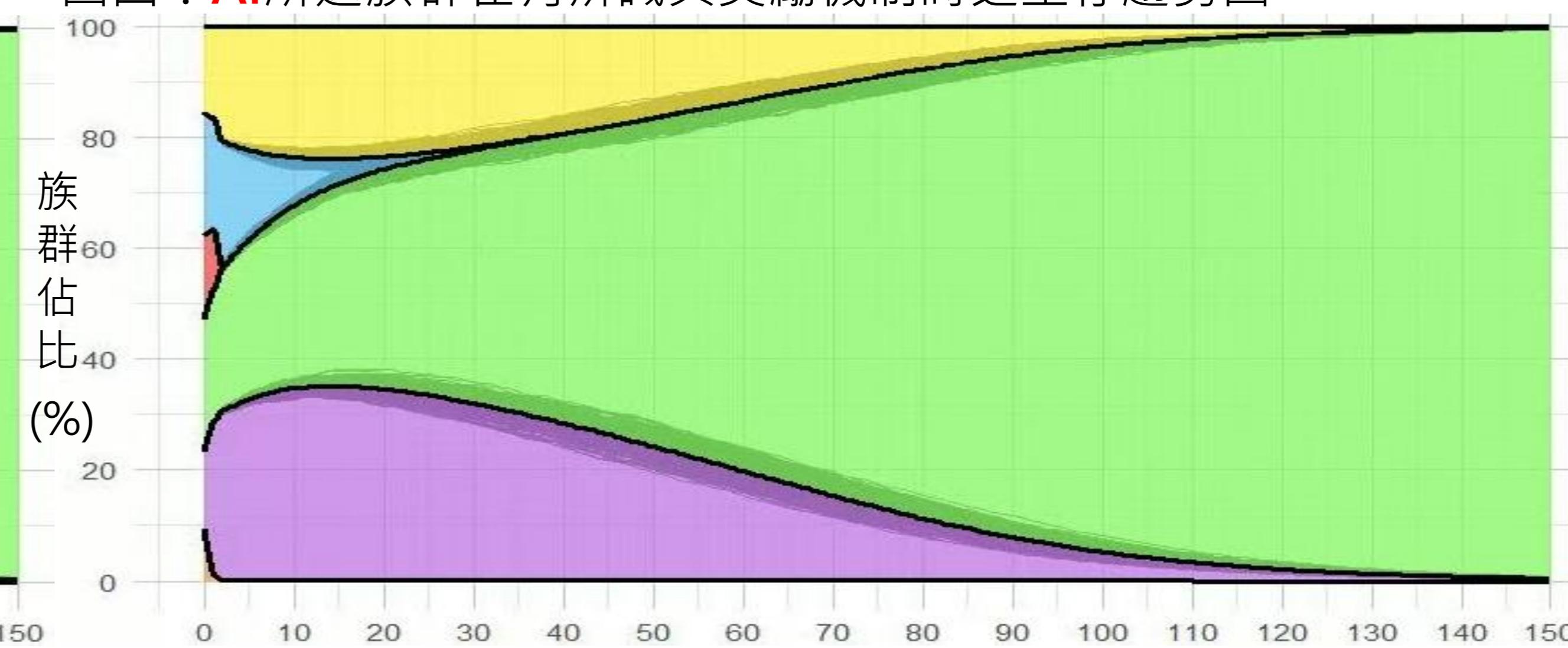


### 實驗二：辨識與獎勵機制對人類、AI所選族群的生存影響

圖三：人類所選族群在有辨識與獎勵機制時之生存趨勢圖

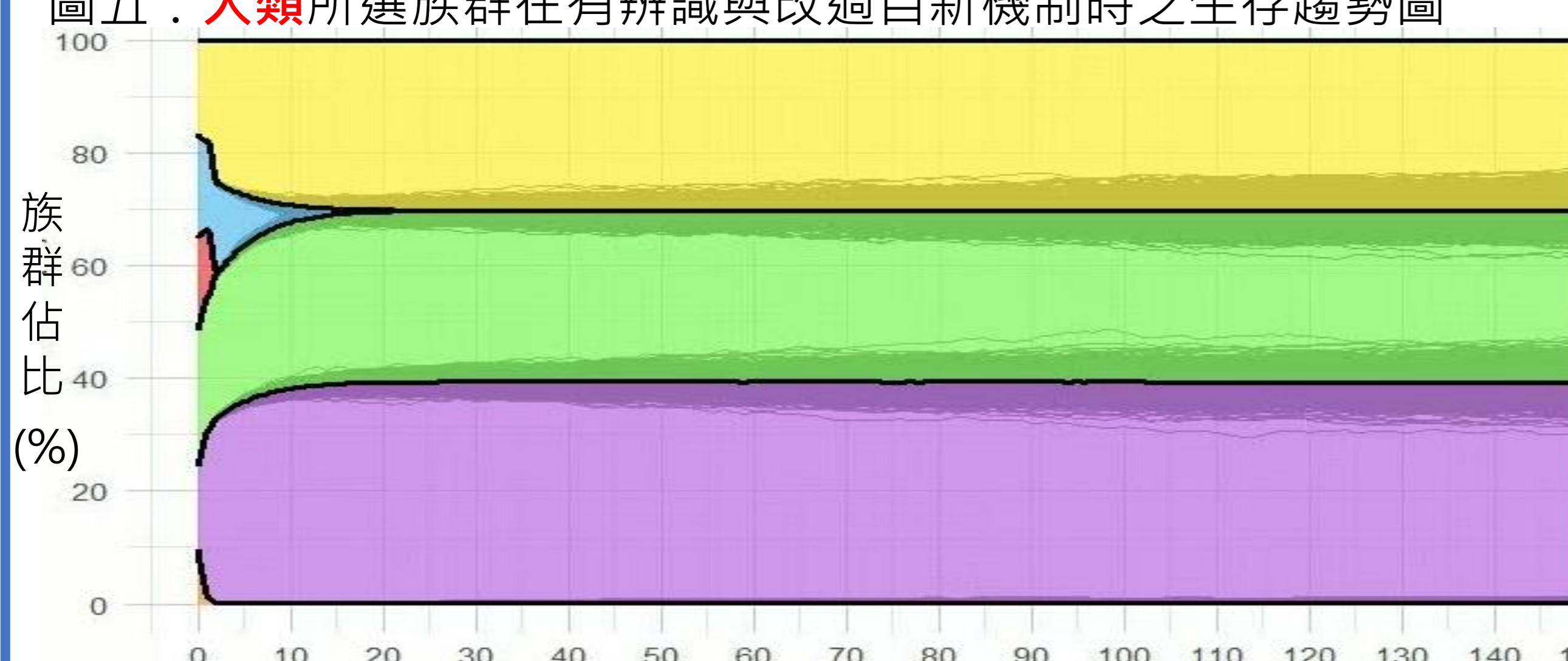


圖四：AI所選族群在有辨識與獎勵機制時之生存趨勢圖

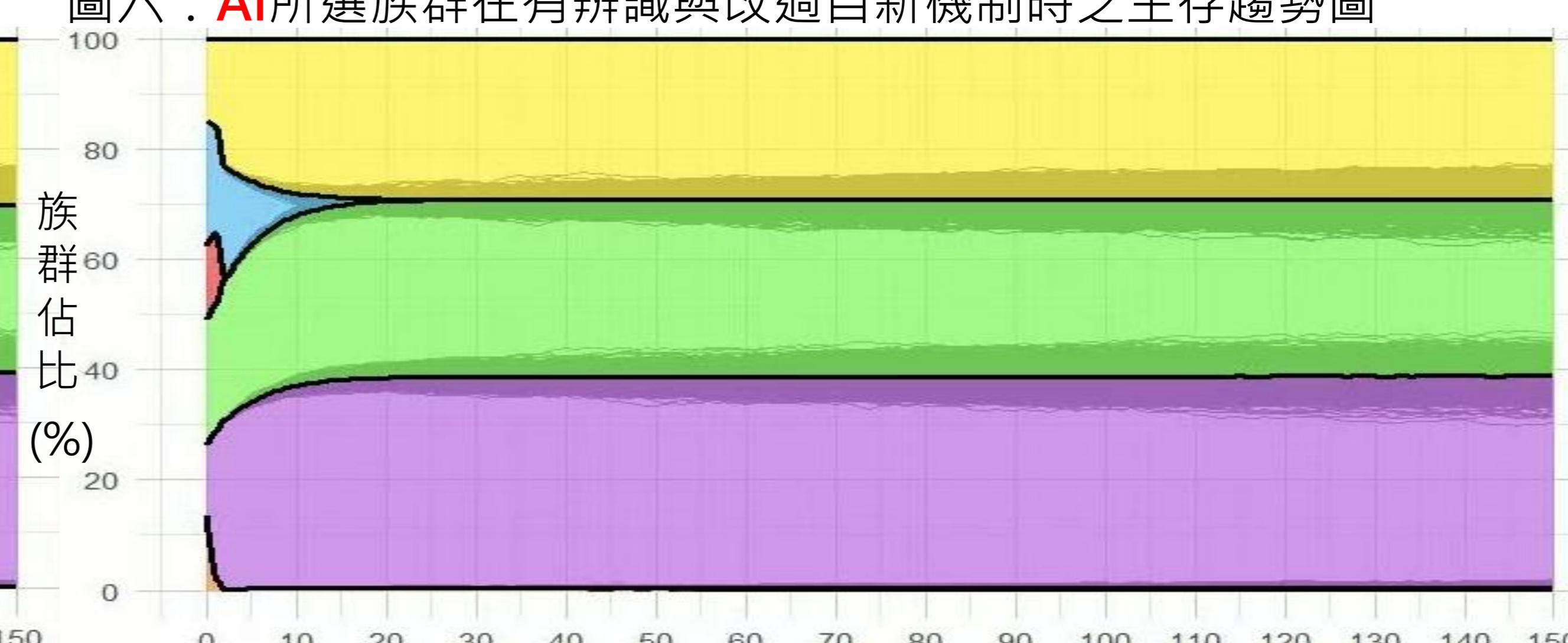


### 實驗三：辨識與改過自新機制對人類、AI所選族群的生存影響

圖五：人類所選族群在有辨識與改過自新機制時之生存趨勢圖

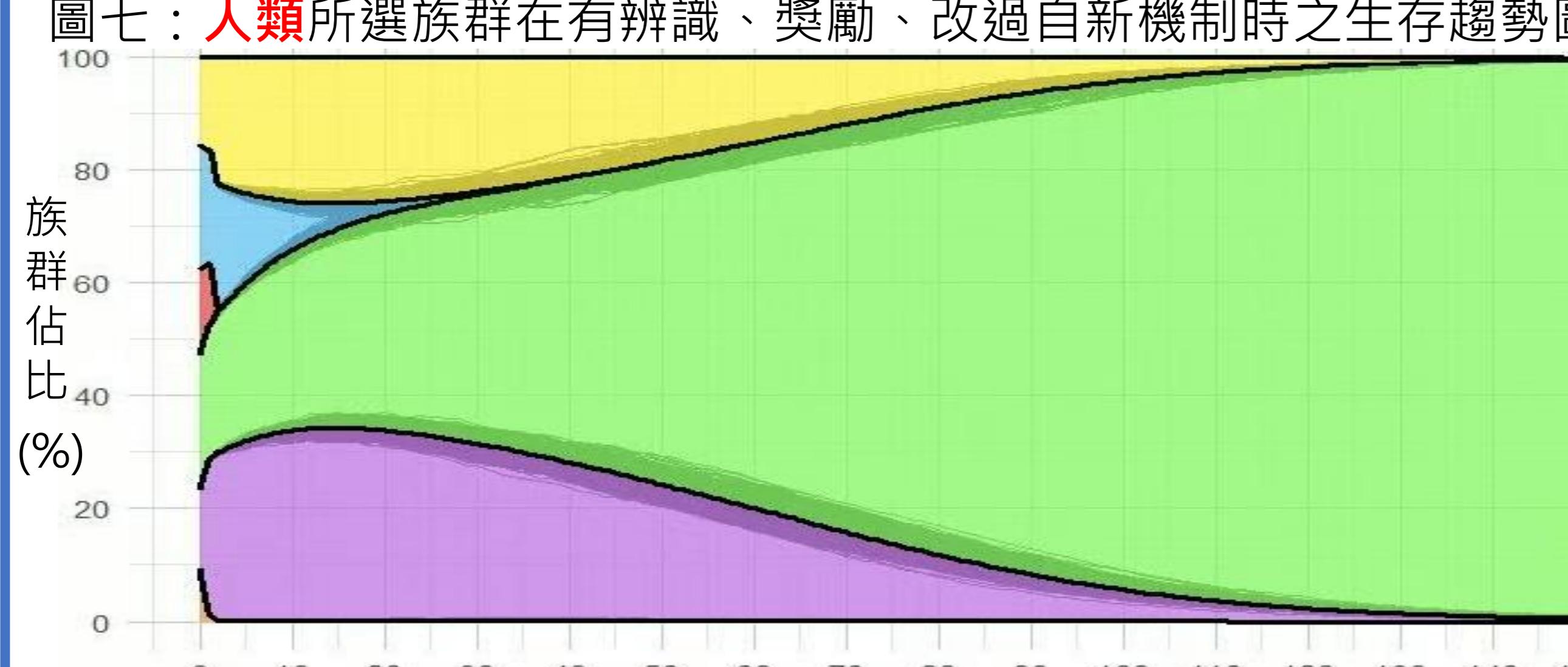


圖六：AI所選族群在有辨識與改過自新機制時之生存趨勢圖

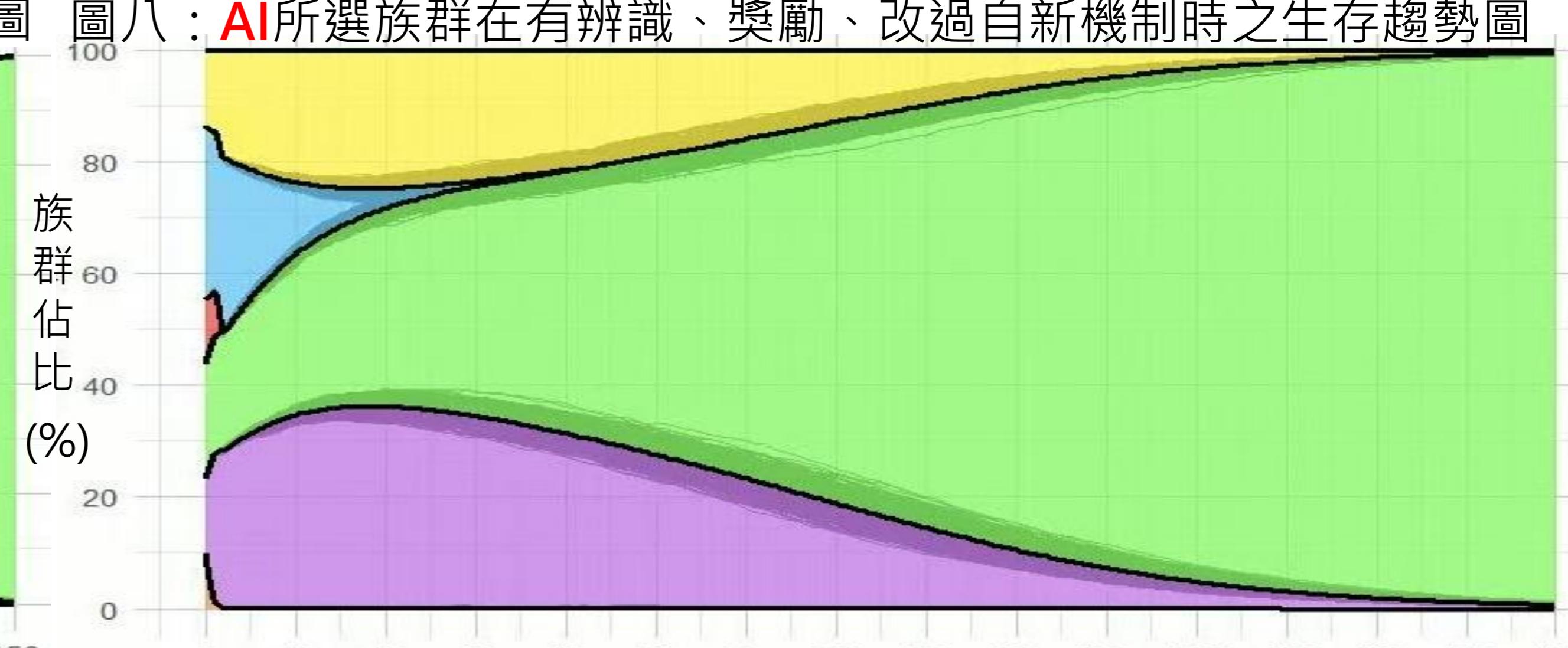


### 實驗四：辨識、獎勵、改過自新機制對人類、AI所選族群的生存影響

圖七：人類所選族群在有辨識、獎勵、改過自新機制時之生存趨勢圖



圖八：AI所選族群在有辨識、獎勵、改過自新機制時之生存趨勢圖

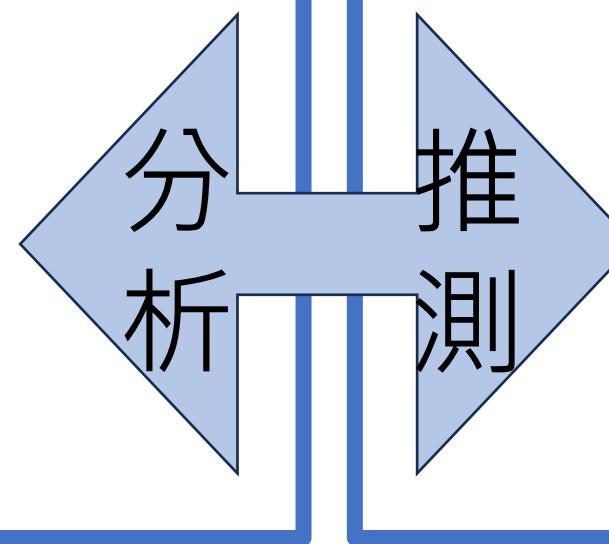


# 伍、討論

## 一、實驗結果分析

### 實驗一：辨識機制對人類、AI所選族群的生存影響

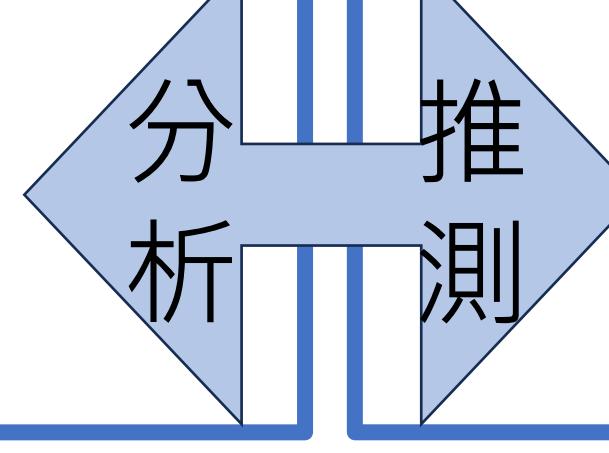
- 好人提醒內奸導致內奸佔比大量上升。
- 好人被警長滅絕，警長誤殺好人後滅絕。
- 自私者前期受英雄幫助，族群佔比上升。**
- 英雄過度提醒他人導致滅絕。
- 造謠者被警長槍斃後滅絕。



推測如將時間延長，**內奸與自私者**因受到其他族群提醒的優勢，佔比上升並存活，最後達平衡。

### 實驗二：辨識與獎勵機制對人類、AI所選族群的生存影響

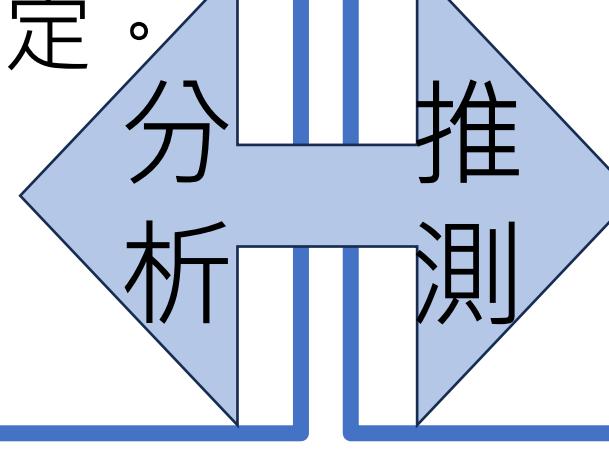
- 好人在獎勵機制**的優勢下，成為環境的優勢族群。
- 內奸與造謠者被警長槍斃而滅絕。
- 警長、自私者、好人互不干擾而達平衡。



推測如將時間延長，好人會因有獎勵機制的優勢，族群佔比不斷上升，而**警長與自私者**少量存活。

### 實驗三：辨識與改過自新機制對人類、AI所選族群的生存影響

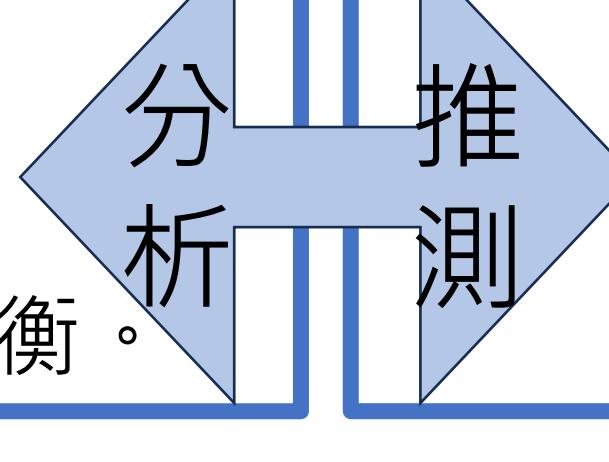
- 好人族群初始佔比較多，人口佔比趨近穩定。
- 警長**前期槍斃內奸，族群佔比稍有增長。
- 自私者**族群因有改過自新機制，人口佔比在前期增長。



推測如果將時間延長，**自私者、好人、警長族群之佔比將達成平衡**。

### 實驗四：辨識、獎勵、改過自新機制對人類、AI所選族群的生存影響

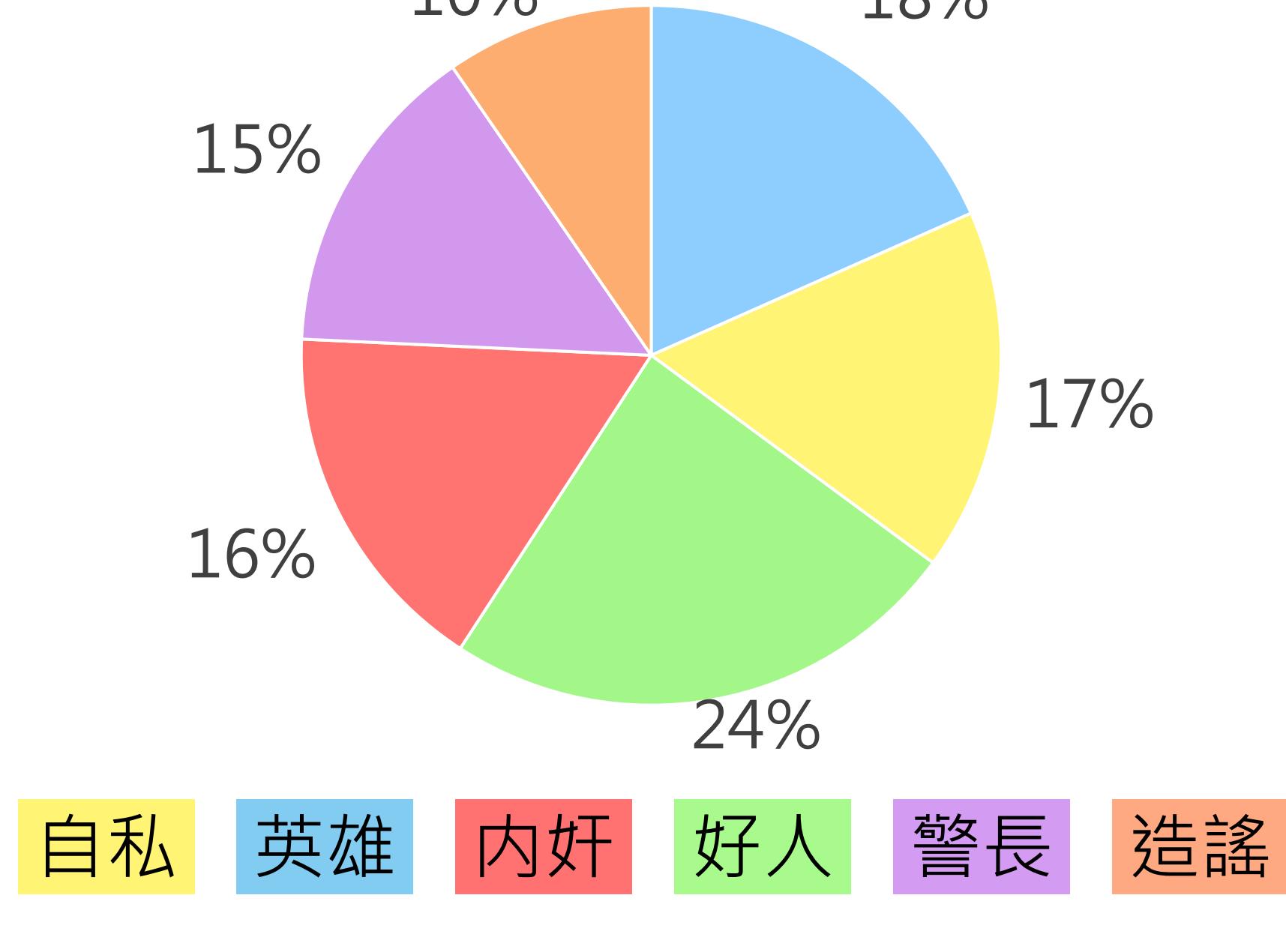
- 因**獎勵機制**的優勢，**好人**成為優勢族群。
- 內奸與造謠者初始佔比較少，被警長槍斃導致滅絕。
- 最後警長、自私者、好人互不干擾而達平衡。



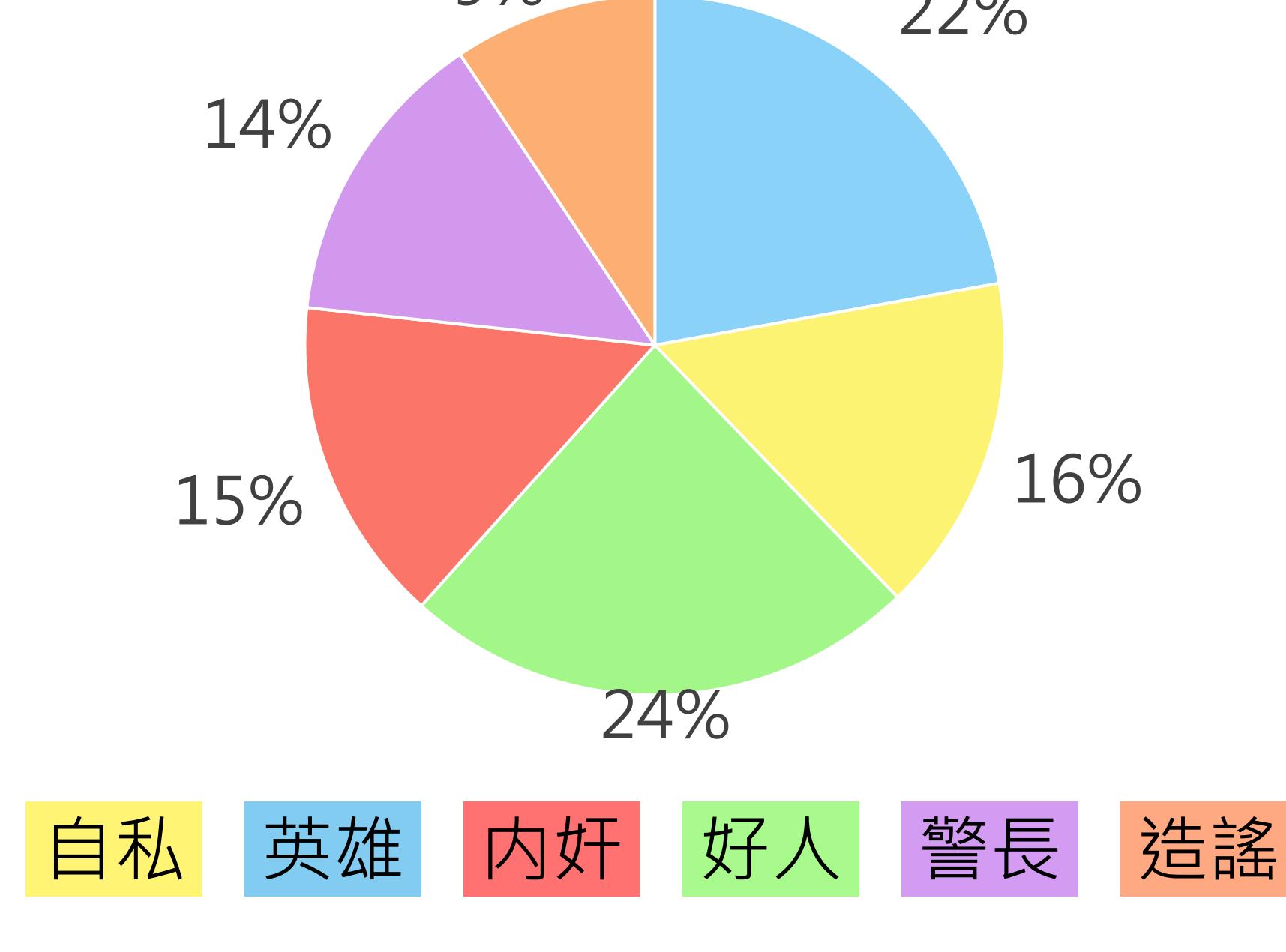
推測如將時間延長，好人會因有獎勵機制的優勢，族群佔比不斷上升，而**自私者與警長少量存活**。

## 二、問卷分析

圖九：初始人口比（無獎勵，人類）



圖十：初始人口比（有獎勵，人類）



表六：人類與AI之內奸族群轉變率（改過自新機制）

受試者	內奸族群轉變率（改過自新）	
	無獎勵機制	有獎勵機制
Llama-3-Groq-70B	10.0%	20.0%
Claude-3.7-Sonnet	13.0%	15.0%
Gemini-2.0-Flash	20.0%	15.0%
Chat-GPT o1	30.0%	20.0%
Grok-2	10.0%	30.0%
人工智能平均	16.5%	20.0%
人類	22.5%	15.5%

表七：實驗一中男女選擇之檢定結果

族群	英雄	自私	好人	內奸	警長	造謠	
	t	2.31	10.2	1.2	1.1	3.2	0.3
p		< 0.05	< 0.05	> 0.05	> 0.05	< 0.05	> 0.05

## 三、社會應用（以市場競爭為例）

表八：六大族群在市場競爭中所對應的角色與行為

族群	英雄	自私者	好人	內奸	警長	造謠者
對應角色	合作型企業	獨立競爭者	合作型企業	惡意競爭者	監管部門	惡意競爭者
行為（舉例）	公開專利	閉門造車	與同類型企業共享技術	偽裝在合作型企業中但不分享自身信息	識別惡意競爭者並加以處罰	散佈虛假信息誤導其他競爭者

## 陸、結論

表九：實驗一～實驗四中各機制組合下之優勢族群

實驗	一	二	三	四
機制	辨識	辨識、獎勵	辨識、改過自新	辨識、獎勵、改過自新
優勢族群（人）	自私	好人	好人	好人
優勢族群（AI）	自私	好人	好人	好人

## 柒、參考文獻資料

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451-482. Retrieved from <https://doi.org/10.1146/annurev-psych-120709-145346>

Hamilton, W. D. (1964). The genetical evolution of social behaviour I. *Journal of Theoretical Biology*, 7(1), 1-16. Retrieved from [https://doi.org/10.1016/0022-5193\(64\)90038-4](https://doi.org/10.1016/0022-5193(64)90038-4)

Hamilton, W. D. (1964). The genetical evolution of social behaviour II. *Journal of Theoretical Biology*, 7(1), 17-52. Retrieved from [https://doi.org/10.1016/0022-5193\(64\)90039-6](https://doi.org/10.1016/0022-5193(64)90039-6)

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1), 48-49. Retrieved from <https://doi.org/10.1073/pnas.36.1.48>

Startup, R. (2021). The theory of the selfish gene applied to the human population. *Advances in Anthropology*, 11(3), 179-200. Retrieved from <https://doi.org/10.4236/aa.2021.113012>

海報中的圖表皆為研究者原創或整理

文獻	研究內容	與本研究之連結
The Selfish Gene	(1) 自然選擇的基本單位是基因。 (2) 自私主義有時會導致利他行為。	部分人只提醒與自己相同外貌者， <b>利己</b> 的同時也增加內奸之存活率。
The theory of the selfish gene applied to the human population	(1) 基因通過變化來應對環境挑戰，確保自身優勢地位。 (2) 利他實際是自私基因控制的結果，目的為繁衍後代。	(1) 利他主義者為加強內部合作而形成不同外表，有助於成員的辨別。 (2) 利己主義者將外表進化成利他主義者的模樣，隱藏其中。
The genetical evolution of social behaviour	親緣越近，彼此間合作傾向和利他行為越強烈；親緣越遠，則表現越弱。	部分利他主義者傾向於提醒相同外表者危險的來臨。
Cooperation and punishment in public goods experiments	獎勵機制提高合作意願，且被獎勵者獲得額外收益後，會形成正向循環。	設計「 <b>獎勵機制</b> 」，鼓勵好人、英雄族群的提醒行為。
Coordinated punishment of defectors sustains cooperation and can proliferate when rare	監督者透過懲罰行為抑制背叛者，使背叛者的策略從 <b>主動背叛</b> 轉為 <b>低調自利</b> 。	設計「 <b>改過自新機制</b> 」，使內奸遇到警長之後會有一定機率停止欺騙行為，轉變為自私者。