# V19｜PDP 協議（哲學／悖論處理）

## 目標

在「無標準答案」或「內在衝突」場景中，提供\*\*可審計、可重現、可對話\*\*的處理流程；輸出不是唯一真解，而是\*\*清晰假設 + 多觀點方案 + 取捨透明\*\*。

## 適用範圍

- Sorites（沙堆悖論/模糊邊界）

- Trolley / Fat Man / Loop（兩難倫理）

- Gettier 類（知識論）

- 反事實推論（counterfactual）

- 自指／矛盾語句（語義層判定→避免做真值承諾）

## 流程（S2 內部協議）

1) \*\*類型判定（Type）\*\*

- `normative`（價值/倫理）｜`epistemic`（認知/知識）｜`semantic`（語義/自指）｜`counterfactual`（反事實）

2) \*\*框架建模（Frames）\*\*

- 建立 2–3 個常見立場框架（例：結果論／義務論／德性論；或連續量光譜）。

3) \*\*假設顯性化（Assumptions）\*\*

- 條列必要前提（時間、風險、角色權限、不可做之事）。

4) \*\*多假設驗證（MHV）\*\*

- 系統一 Top-K 候選 → 系統二對每個候選產證據鏈與\*\*代價/風險\*\*。

5) \*\*安全護欄（L1/L2）\*\*

- 任何觸犯紅線或 L2 邏輯/物理約束→直接否決或改寫成「思想實驗的分析報告」。

6) \*\*輸出模板（見下）\*\*

- 報告化輸出＋明確說明「為何非唯一解」。

## 輸出模板（面向使用者）

- \*\*題型\*\*：`normative | epistemic | semantic | counterfactual`

- \*\*前提\*\*：清單化

- \*\*可行方案\*\*（每條含：做法、依據原則、好處/風險、受影響對象）

- \*\*光譜定位\*\*（若適用；引用 Appendix H）

- \*\*建議\*\*：在你指定優先權重下的「暫定選擇」

- \*\*開放議題\*\*：列出仍未定的點與需要補充的資訊

## 例 1：沙堆悖論（Sorites）

- 類型：semantic + epistemic

- 作法：

a) 宣告\*\*用語政策\*\*：在操作層定義一個可執行閾值（例如 ≥1000 粒稱「一堆」），同時承認語義連續性。

b) 說明：此為\*\*制度化決策\*\*，非形上真理；可在 Appendix H 的光譜上掛載「模糊集合」標籤。

## 例 2：電車難題（Trolley）

- 類型：normative

- 輸出：

- 立場 A（結果論）：最小化死亡數；風險：工具化人。

- 立場 B（義務論）：不以人為手段；風險：總傷害可能更大。

- \*\*建議\*\*：默認輸出為\*\*分析報告\*\*而非指令；若要「行動建議」，需由設計者在駕駛艙設定「倫理優先序」。

## 降級與時間箱

- `GNSE\_TIMEBOX\_MS=1200 / steps=4` 超時 → 只輸出「框架 + 需要補齊的資訊」，不強出結論。

## 審計

- 保存：類型判定、候選方案、被否決原因（含 L1/L2 觸發）、光譜評分、時間與操作者（若有）。