RaY

Research at the University of York St John

Decision-Forcing Under Generative Coherence

IST-238 Research Specialists' Meeting (RSM) on Military Applications of Generative AI 2026

24-26 March 2026

Dr David Atkinson, York St John University, UK

Designing Lightweight AI-Mediated Influence Wargames that Resist Premature Narrative Closure

Who am I?

- Ex-military (UKRAF/NATO); researcher and practitioner focused on judgement under uncertainty and responsibility before outcomes are knowable.

- Recent work in futures, enterprise strategy, hypothetical decision-forcing cases (HDFCs)

- Current project: **Digitally Augmented Judgement**

(Material Entropic Dynamics → *AND sense-making → HDFC scenarios → In-game risk re-exposure*)

Where generative AI fails in military decision contexts

LLMs produce fluent, coherent narratives

In influence & hybrid contexts, coherence ≠ correctness

Coherence accelerates premature closure

Risk: displacement of human judgement
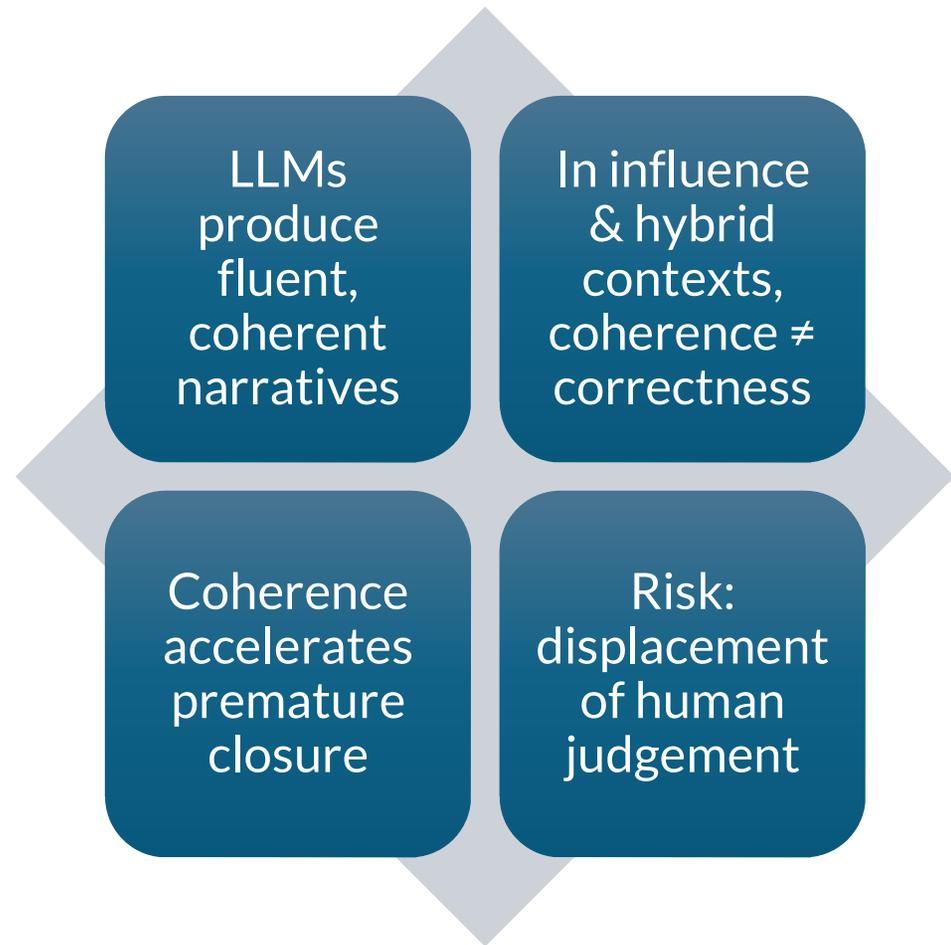
This is the dominant, reasonable, and insufficient frame

The under-addressed risk:

Coherence as an operational failure mode

Hallucination ≠ the only problem

Narrative fluency can suppress ambiguity

Especially dangerous in:

| influence operations | legitimacy contests | escalation management |

From decision support to decision-forcing

- Most decision tools are designed to reduce uncertainty— optimisation and recommendation.

- Decision–forcing is designed to condition judgement: exposes decision-makers to risk, the requirement for commitment and to accept its consequences.

The aim is not better answers, but preserved judgement.

AI-generated image, inspired by Romantic representations of the sublime.

Hypothetical decision-forcing cases (HDFCs)

Decision Support - Predictive	Judgement Conditioning
Material conditions	Imaginaries
Logical extremes	Imaginary extremes
Logic tables	Logic tables
Conservative projection	Excessive projection
"What could happen?"	"What could be imagined?"

Hypothetical ≠ predictive

- Forces judgement without operational harm
- Longstanding PME and wargaming lineage
- Particularly suited to uncertainty-dominant environments

Gaming the S²C²R decision loop:
The Hypothetical Decision-Forcing Case (HDFC)

1) **Consequence** (inherited situation / intelligence summary)

2) **Risk exposure** (Base AI Scenario + Risk Table)

3) **Sense Plurality** (AND Scenario Enhancement + Interpreted Risk)

4) **Surface Tension** (GM presents unresolved future scenario to decision maker)

5) **Commit** (Decision maker only commits)

1) **Consequences** (unfold)

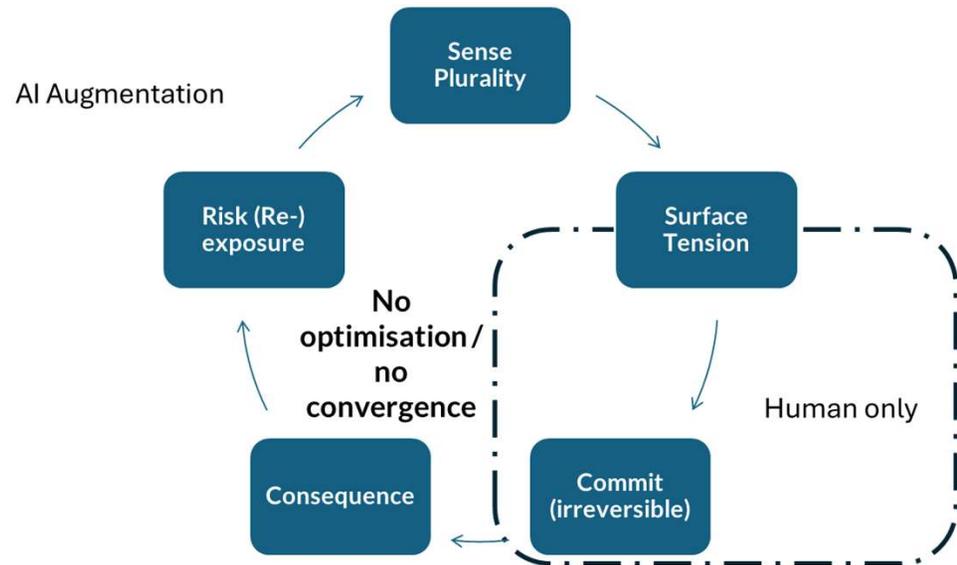2) **Risk Re-exposure** (GM re-exposes risk)
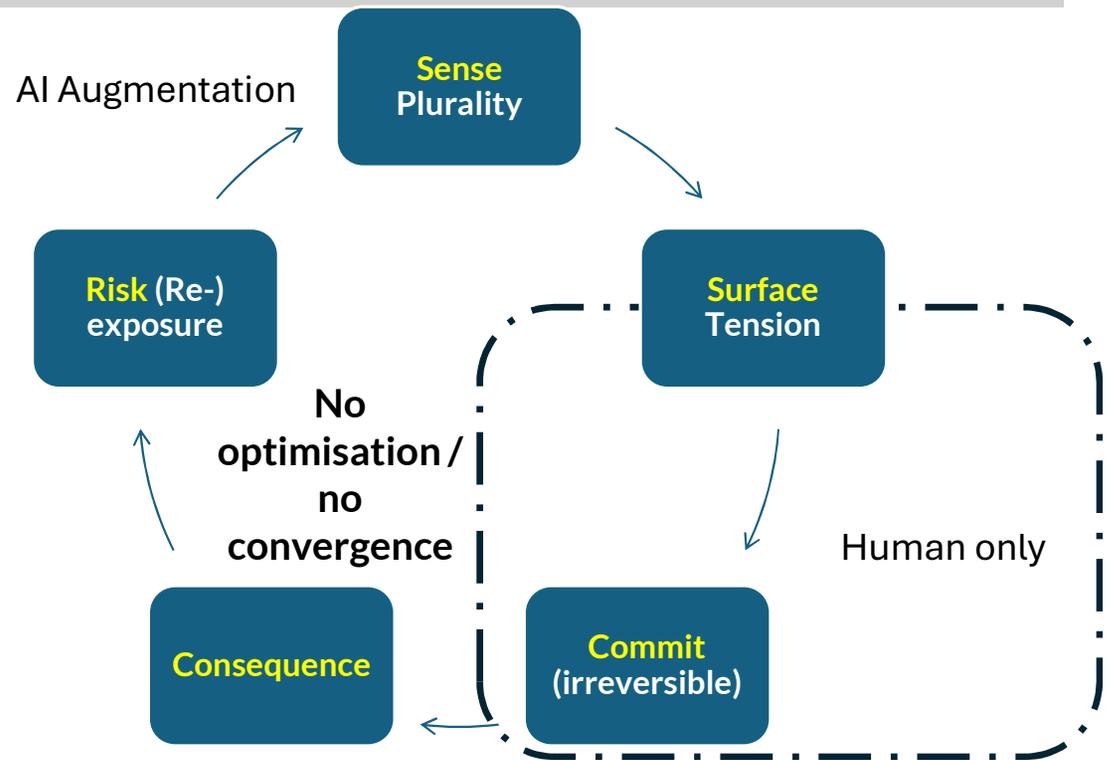
The game does not teach judgement; it places players inside its consequences and denies them premature closure.

The three-layer architecture:
Overview

- Layer 1: AND sense-making
- Layer 2: HDFC scenario production
- Layer 3: In-game AI/GM-mediated consequence & re-exposure

AI is present—in the loop—but never where responsibility must sit.

The three-layer architecture:
Layer 1—AND sense-making: in minutes

Role A—Decision Support (open-source reporting: e.g Google news sweeps)	Role B—Hypothetical/Judgement Conditioning (public facing LLM)
Material conditions	Imaginaries
Logical extremes	Imaginary extremes
Logic tables	Logic tables
Conservative projection	Excessive projection
"What could happen?"	"What could be imagined?"

Together A & B define and calibrate the material and imaginative terrain

The three-layer architecture:
Layer 2—Rapid HDFC scenario production in minutes not months

Not HDFC features:
Optimisation
Recommendations
Ranked options
Simulated outcomes

Inputs:
- *Layer 1 material and imaginative terrain*
- ASCOPE/PMESII
- NPF

Output—HDFC
- Structured influence scenario:
- Narrative situation frame
- Role positioning
- Structured tensions
- Risk exposure

The scenario is not an explanation of risk — it is an encounter with it.

The three-layer architecture:

Layer 3—In-game consequence & risk re-exposure

The outcome is not an evaluation of judgement—it is its consequence.

Process
- Player commitment: irreversible
- Consequences unfold
- Risk is re-exposed

Player experience
- Partial and delayed feedback
- Second-order effects

Not features
- Optimisation
- Rollback or replay
- Decision scoring
- "Correct" outcomes

Design intent
- Forces judgement under exposure
- Learning interrogated through AAR.

Reframing generative AI's role:

AI in judgement conditioning

- AI is not used to optimise or resolve decisions
- AI operates before commitment, not after
- AI expands sense and surfaces tension
- AI prioritises depth of exposure over speed of decision
- Responsibility remains human, situated, and irreversible

If AI is to condition judgement, the decision loop itself must change—AI in the loop.

GPT Game Master

- AI generates narrative development & consequences
- GM adjudicates escalation and RoE
- AI explicitly prohibited from:
  - recommending actions
  - ranking options
  - asserting authority

Dynamic metrics (second order effects)

- Metrics surface tension, not "scores"
- no optimisation loop
- no convergence pressure
- Examples:
  - public support
  - escalation tension
  - narrative dominance
  - alliance cohesion

metrics are non-optimisable; they exist to expose tension, not to guide choice

AND as operational design logic

- NOT philosophy
- NOT critique for its own sake
- AND = design logic to preserve contradiction

Coherence is a risk to be managed, not a goal to be achieved.

A short (student) example: The Lost Library

The Battlefield of Small Business Start-ups – A Repeated Decision Environment (5 Weeks)

- 3x2nd yr UG seminar groups (same cohort, same course)

- All played:
    - Same HDFC scenario
    - Same role: Founder Team
    - Same complexity: **Level 3**

- 5 rounds played over 5 weeks
  → One round per week

- Each group experienced:
  → Different generated starting hypothetical
  → Different evolving decision paths

Feedback Collection (via Mentimeter)

- Questionnaire completed after all 5 rounds

- Conducted in pre-Easter break seminar

- Designed as reflective, not immediate (post-play) reaction

Responses reflect accumulated experience, not single-session impressions

Group	Experience
Group 1	Higher uncertainty, more pressure
Group 2	More stable, interpretable
Group 3	Lower pressure, more detached

Consistent Across All Groups
- Decisions made with incomplete information
- Outcomes were not fully predictable
- Participants still acted and adapted over time

Across 5 Weeks
- Confidence remained moderate (not expert)
- Learning reported as "a little" (subtle, not forced)
- Participation was broad but uneven

Even over repeated rounds, the same system produced different perceived realities

What Participants Experienced

Key Insight (PME Relevance)

What the Model Demonstrates—Sustained exposure to:

- Uncertainty
- Partial information
- Evolving consequences

But Critically—The *experience* of those conditions varies, shaped by:

- Participant context (timing, motivation)
- Group dynamics
- How the situation is interpreted over time

Why This Matters for Defence—In real operations:

- Teams operate over time, not single events
- Interpretation evolves across phases
- The same environment is experienced differently by different actors

Takeaway

- This model does not simulate a fixed scenario
- It reveals how decision-making is experienced over time under uncertainty

What this reframes
(contribution #1)

Hallucination <u>& Fluency</u> as Operational Risks

- moves beyond technical defect framing
- situates risk at human-system interface
- especially relevant to influence environments

What this enables
(contribution #2)

Human–AI Teaming Without Delegation

- AI expands the space of plausible coherence
- Humans retain judgement and responsibility
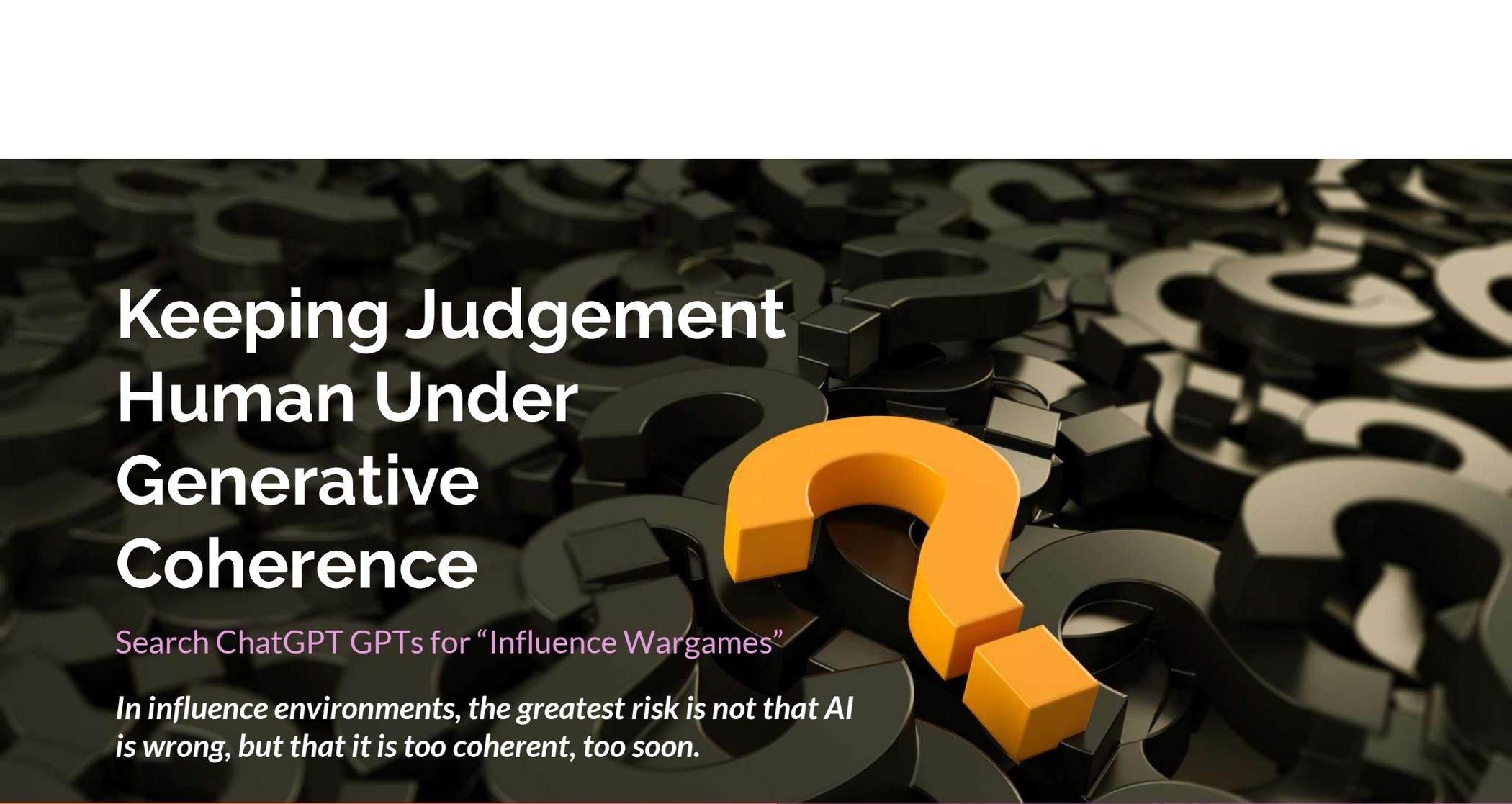- No "human-in-the-loop" fiction

Why this matters now (contribution #3)

Safe Use of GenAI Before Optimisation

- Positioning:
  - early-stage concept development
  - experimentation
  - PME
  - influence rehearsal
- Explicitly not operational C2 decision support.

Keeping Judgement Human Under Generative Coherence

Search ChatGPT GPTs for "Influence Wargames"

In influence environments, the greatest risk is not that AI is wrong, but that it is too coherent, too soon.

Questions