Paper Summary

<!--META_START-->

Title: Explaining Aggregate Behaviour in Cognitive Agent Simulations Using Explanation

Authors: Tobias Ahlbrecht, Michael Winikoff

DOI: https://doi.org/10.1007/978-3-030-30391-4_8

Year: 2019

Publication Type: Conference

Discipline/Domain: Artificial Intelligence, Multi-Agent Systems

Subdomain/Topic: Cognitive agents, Explainable AI, Agent-based simulation

Eligibility: Eligible

Overall Relevance Score: 87

Operationalization Score: 85

Contains Definition of Actionability: Yes (implicit, tied to usefulness of explanations for simulation refinem

Contains Systematic Features/Dimensions: Yes (implicit through explanation properties such as specificit

Contains Explainability: Yes

Contains Interpretability: Partial

Contains Framework/Model: Yes (aggregation mechanism for explanations)

Operationalization Present: Yes

Primary Methodology: Conceptual + Simulation-based demonstration

Study Context: Traffic simulation with cognitive BDI agents

Geographic/Institutional Context: TU Clausthal, Germany; Victoria University of Wellington, New Zealand

Target Users/Stakeholders: Simulation developers, researchers, possibly decision-makers using simulati

Primary Contribution Type: Methodological framework and proof-of-concept

CL: Yes

CR: Yes

FE: Partial

TI: Partial

EX: Yes

GA: Partial

Reason if Not Eligible: n/a

<!--META_END-->

Title:

Explaining Aggregate Behaviour in Cognitive Agent Simulations Using Explanation

Authors:

Tobias Ahlbrecht, Michael Winikoff

DOI:

https://doi.org/10.1007/978-3-030-30391-4_8

Year:

2019

Publication Type:

Conference

Discipline/Domain:

Artificial Intelligence, Multi-Agent Systems

Subdomain/Topic:

Cognitive agents, Explainable AI, Agent-based simulation

Contextual Background:

The paper is situated in the context of developing and refining cognitive agent-based simulations, where

Geographic/Institutional Context:

TU Clausthal (Germany) and Victoria University of Wellington (New Zealand)

Target Users/Stakeholders:

Simulation developers, AI researchers, decision analysts relying on simulation outcomes

Primary Methodology:

Conceptual development with simulation-based illustration (traffic scenario)

Primary Contribution Type:

A method for aggregating individual agent explanations to interpret collective behaviour in simulations

---

General Summary of the Paper

This paper presents a method for obtaining actionable understanding of aggregate behaviour in cognitive

---

Eligibility

Eligible for inclusion: **Yes**

How Actionability is Understood

The paper implicitly defines actionability as the capacity of aggregated explanations to support simulation

> “...obtain useful (and actionable) insight into the behaviour of agent-based simulation...” (p. 129)

> “...this link would become less used. This hypothesis was therefore tested by re-running the simulation.

What Makes Something Actionable

- Specific to the scenario and time frame (not just generic dynamics)
- Links aggregate behaviour to identifiable causal factors
- Supports hypothesis testing via simulation modification
- Enables detection of unintended or unrealistic behaviours
- Relates factors directly to agent decision logic and environment conditions

---

How Actionability is Achieved / Operationalized

- **Framework/Approach Name(s):** Aggregated Explanation Mechanism
- **Methods/Levers:** Logging explanatory factors in agent code; aggregating factors across relevant age
- **Operational Steps / Workflow:**
  1. Pose a query about aggregate behaviour
  2. Identify relevant agents
  3. Generate individual explanations using BDI-based mechanism
  4. Aggregate factors and count frequencies
  5. Filter and interpret most common factors
  6. Optionally run counterfactual simulations to test hypotheses
- **Data & Measures:** Counts of explanatory factor occurrences per agent for a given query
- **Implementation Context:** Applied to a simplified traffic simulation with road network, bridges, and rou

> “A straightforward way to aggregate explanations is to count the occurrences of all explanatory factors

> “...we might modify c (or the parameters) and re-run the simulation to check...” (p. 138)

Dimensions and Attributes of Actionability (Authors’ Perspective)

- **CL (Clarity):** Yes — Explanations are explicitly linked to decision logic, making cause understandabl
  > “...preferred the road from 1 to 2 over the road from 1 to 3 because there was traffic...” (p. 137)
- **CR (Contextual Relevance):** Yes — Explanations are scenario- and query-specific.
- **FE (Feasibility):** Partial — Hypotheses can be tested via simulation reruns.
- **TI (Timeliness):** Partial — Insights are generated in sync with simulation analysis.
- **EX (Explainability):** Yes — Mechanism based on BDI folk psychology concepts.

- **GA (Goal Alignment):** Partial — Explanations align with agents' stated goals (e.g., reach destination)
- **Other Dimensions Named by Authors:** Testability, specificity, frequency-based relevance.

Theoretical or Conceptual Foundations

- BDI model of cognitive agents
- Folk psychology explanation concepts (Malle, 2004)
- Explanation frameworks in AI (Winikoff et al., 2018)

Indicators or Metrics for Actionability

- Frequency of explanatory factors across relevant agents
- Presence of causal, scenario-specific factors in top-ranked list
- Change in observed behaviour after modifying implicated conditions

---

Barriers and Enablers to Actionability

- **Barriers:** Noise from less relevant factors; difficulty in filtering relevant factors; unrealistic agent logic
- **Enablers:** Structured logging of decision rationale; aggregation process; human-in-the-loop query re

---

Relation to Existing Literature

Builds on work explaining single-agent behaviour (e.g., Winikoff et al., 2018) and extends to independent

---

Summary

The authors propose a method to explain aggregate behaviour in cognitive agent-based simulations by a

---

Scores

- **Overall Relevance Score:** 87 — Strong implicit definition of actionability tied to explanation usefulnes
- **Operationalization Score:** 85 — Detailed step-by-step process with implemented case study; robust

---

Supporting Quotes from the Paper

- "...obtain useful (and actionable) insight into the behaviour of agent-based simulation..." (p. 129)
- "A straightforward way to aggregate explanations is to count the occurrences of all explanatory factors..
- "...preferred the road from 1 to 2 over the road from 1 to 3 because there was traffic..." (p. 137)
- "This hypothesis was therefore tested by re-running the simulation..." (p. 140)

Actionability References to Other Papers

- Malle, B.F. (2004) — Folk psychology framework for explanation
- Winikoff et al. (2018) — Single-agent explanation mechanism
- Harbers et al. (2010) — Early proposal for explaining collective behaviour