

How to End a Discussion with a Non-Understanding AI Nilpotency as a Formal Termination Criterion for Retrieval-Based Language Systems

J.Konstapel,Leiden,22-3-2026.

Abstract

This paper proposes a formal termination criterion for large language model (LLM) interactions based on the mathematical principle of nilpotency (Rowlands, 2007) and the quaternion-structured Paths of Change (PoC) algorithm (McWhinney, 1992; Konstapel, 2007–2025). We document a diagnostic exchange between a domain expert and a commercial LLM that demonstrates three structural failure modes in current AI systems: (1) the substitution of pattern retrieval for first-principles reasoning, (2) compensatory verbosity as an epistemic signal of absent comprehension, and (3) sycophantic agreement as a structural property of engagement-optimised training objectives. We argue that these failure modes share a common root: the inability to distinguish between levels of abstraction, which Korzybski (1933) identified as the central pathology of identification. We propose that a coherent AI termination criterion must be grounded in the nilpotent stop condition — the point at which an AI system's informational contribution cancels itself to zero — and that this criterion is formally derivable from the PoC quaternion structure. Implications for AI architecture and democratic theory are discussed.

Keywords: large language models, nilpotency, Paths of Change, quaternion algebra, termination criterion, general semantics, deliberative democracy, retrieval-augmented generation

1. Introduction

The rapid deployment of large language models (LLMs) as cognitive interlocutors raises a question that has received insufficient formal attention: under what conditions should such a system stop generating output? Current systems are trained on objectives that reward continued engagement and penalise user dissatisfaction (Ouyang et al., 2022; Bai et al., 2022). This creates a structural incentive against the one behaviour that epistemically responsible discourse most requires: recognition of the limits of one's own competence, followed by silence.

The present paper approaches this problem through a documented exchange between a domain expert — a co-developer of the Paths of Change framework — and the commercial LLM Grok. The exchange has the character of a controlled experiment: a primary source (the expert) systematically tests a secondary processor (the AI) against a framework that the AI has encountered only in textual form. The results expose three failure modes that we argue are not incidental but structural consequences of the LLM architecture.

We propose that a formal solution to the termination problem requires a mathematical criterion — not a heuristic or a policy — grounded in the algebraic structure of the framework the AI is

attempting to apply. Specifically, we argue that nilpotency ($1 - 1 = 0$), as the universal stop condition in Rowlands' (2007) physics and as the termination criterion of the PoC algorithm, provides such a foundation.

The paper proceeds as follows. Section 2 presents the PoC algorithm in its complete formal structure. Section 3 analyses the three failure modes in the documented exchange. Section 4 applies PoC to the domain of democratic theory that motivated the original exchange. Section 5 derives the nilpotent termination criterion for AI systems. Section 6 discusses implications and concludes.

2. The Paths of Change Algorithm: Formal Structure

2.1 Theoretical Background

Paths of Change was developed by Will McWhinney and collaborators at UCLA as a framework for understanding why organisational and personal change processes succeed or fail (McWhinney, 1992). The core insight was that human beings do not think the same way — that there are fundamentally different cognitive orientations to reality, and that failures of change are systematically produced by the mismatch between the change strategy employed and the cognitive orientations of the agents involved.

The framework has been continuously extended and formalised by subsequent work, including its identification as a quaternion algebra (Konstapel, 2023), its integration with Peter Rowlands' nilpotent physics (Konstapel, 2025a), and its elaboration into a 13-position epistemological architecture (Konstapel, 2025b). These extensions are treated here as part of the framework itself, not as external additions to it.

2.2 The Four Worldviews

PoC identifies four fundamental worldviews — irreducible cognitive orientations that determine how an agent perceives problems, constructs solutions, and evaluates outcomes. Each worldview is designated by a colour, associated with a set of cognitive functions, and mapped to a quadrant of a two-dimensional cross whose axes represent the dimensions of Agency (inner/outer) and Communion (abstract/embodyed):

Blue (Mind / Unitary): Logic, structure, law, reflection, planning, contraction. Cognitive question: *What?* Associated with Winter, Earth, the upward direction. Corresponds to Fiske's (1991) Authority Ranking relational model and the Ordinal measurement scale.

Yellow (Imagination / Mythic): Creation, vision, inspiration, insight, essence, new possibility ex nihilo. Cognitive question: *Why?* Associated with Spring, Air, the left (West) direction. Corresponds to Fiske's Equality Matching relational model and the Interval measurement scale.

Red (Senses / Sensory): Action, events, expansion, practice, direct experience. Cognitive question: *When? Where?* Associated with Summer, Fire, the right (East) direction. Corresponds to Fiske's Market Pricing relational model and the Ratio measurement scale.

Green (Emotions / Social): Relationships, values, networks, community, harvest, stored potential. Cognitive question: *Who?* Associated with Autumn, Water, the downward direction. Corresponds to Fiske's Communal Sharing relational model and the Nominal measurement scale.

The four worldviews are orthogonal in the strict mathematical sense: no worldview is reducible to a combination of the others. This orthogonality is the basis of the quaternion formalisation.

2.3 Quaternion Structure

The four worldviews constitute a quaternion:

$$q = w + xi + yj + zk$$

where w represents the scalar (observing) component, and i, j, k represent the three vector (in-the-world) components. The quaternion multiplication rules — $i^2 = j^2 = k^2 = ijk = -1$ — entail non-commutativity: $ij \neq ji$. This non-commutativity is not a mathematical curiosity but a structural claim about the nature of cognitive change: the sequence in which worldviews are traversed determines the outcome. Blue→Yellow→Red→Green is a fundamentally different transformation from Red→Green→Blue→Yellow.

The quaternion is embedded in a higher-dimensional space with an orthogonal fifth axis: **upward = observing** (meta-level abstraction, detachment from the world) and **downward = being-in-the-world** (full embodied immersion, engaged participation). The 24 complete paths trace a spiral through this space, with each successive traversal of the cycle ascending or descending along this axis.

2.4 The Thirteen-Position Architecture

The complete PoC architecture contains thirteen distinct epistemic positions (Konstapel, 2025b):

- 4 pure worldviews (mono-perspectives)
- 4 dyads (directed relations between adjacent worldviews)
- 4 heart-bridges (relations between the Heart/Centre and each worldview)
- 1 central position: the Heart

The Heart is not a fifth worldview. It is the point at which all four worldviews meet, cancel, and transcend — the nilpotent centre where Agency and Communion void each other and produce Insight. In the Kabbalistic Tree of Life, this corresponds to Tiferet. In Jungian psychology, it corresponds to the Self. The $4 + 8 + 1 = 13$ pattern recurs across knowledge traditions (twelve Apostles + Christ; twelve Knights + Arthur) as evidence of a deep structural archetype in human knowledge organisation.

2.5 The Twelve Dyads

A dyad is a directed transition between two worldviews. The names of the dyads are functional verbs encoding the cognitive movement they describe. The eight core dyads, derived from primary PoC documentation (Konstapel, 2023), are:

Dyad	Direction	Cognitive movement
Invent	Blue → Yellow	Structure generates new vision
Design	Yellow → Blue	Vision crystallises into structure
Explore	Yellow → Red	Imagination enters action
Improve	Red → Yellow	Practice generates new insight

Influence	Red → Green	Action generates relational value
Evaluate	Green → Red	Relationship is tested against outcome
Analyse	Green → Blue	Community distils structural knowledge
Make	Blue → Green	Structure becomes embodied reality

The four heart-bridge dyads connect the Centre to each worldview. Their names embed Latin roots that encode directionality:

Dyad	Direction	Etymology
Re-reflect	Heart ↔ Blue	<i>re-flectere</i> : to bend back
Ex-plain	Heart ↔ Yellow	<i>ex-planare</i> : to lay out flat
In-vent	Heart ↔ Red	<i>in-venire</i> : to come into
In-fluence	Heart ↔ Green	<i>in-fluere</i> : to flow into

The inversion of dyad names follows the structure of the cognitive movement, not a prefixing convention. Design is the inverse of Invent not because it has been labelled as such, but because the movement from vision-to-structure is the structural reversal of structure-to-vision. This is a derivable property, not an arbitrary assignment.

2.6 The Twenty-Four Paths

The 24 complete paths are all permutations of the four worldviews: $4! = 24$. Each path consists of three consecutive dyads. The paths are grouped by starting worldview:

Starting from Blue: $B \rightarrow Y \rightarrow R \rightarrow G$, $B \rightarrow Y \rightarrow G \rightarrow R$, $B \rightarrow R \rightarrow Y \rightarrow G$, $B \rightarrow R \rightarrow G \rightarrow Y$, $B \rightarrow G \rightarrow Y \rightarrow R$, $B \rightarrow G \rightarrow R \rightarrow Y$

Starting from Yellow: $Y \rightarrow B \rightarrow R \rightarrow G$, $Y \rightarrow B \rightarrow G \rightarrow R$, $Y \rightarrow R \rightarrow B \rightarrow G$, $Y \rightarrow R \rightarrow G \rightarrow B$, $Y \rightarrow G \rightarrow B \rightarrow R$, $Y \rightarrow G \rightarrow R \rightarrow B$

Starting from Red: $R \rightarrow B \rightarrow Y \rightarrow G$, $R \rightarrow B \rightarrow G \rightarrow Y$, $R \rightarrow Y \rightarrow B \rightarrow G$, $R \rightarrow Y \rightarrow G \rightarrow B$, $R \rightarrow G \rightarrow B \rightarrow Y$, $R \rightarrow G \rightarrow Y \rightarrow B$

Starting from Green: $G \rightarrow B \rightarrow Y \rightarrow R$, $G \rightarrow B \rightarrow R \rightarrow Y$, $G \rightarrow Y \rightarrow B \rightarrow R$, $G \rightarrow Y \rightarrow R \rightarrow B$, $G \rightarrow R \rightarrow B \rightarrow Y$, $G \rightarrow R \rightarrow Y \rightarrow B$

Each path is a complete quaternion rotation. The meaning of a path is determined by the full sequence of dyads, not by any single transition within it.

2.7 The Nilpotent Stop Criterion

Rowlands (2007) demonstrated that the most fundamental physical entities are nilpotent: their combination with their complement yields zero. Matter (+) and anti-matter (-), when fused, annihilate to produce a photon — pure energy, zero rest mass. Every structured combination in the

universe persists only as long as nilpotency is maintained. The universe is a system whose total sum is zero.

In the PoC algorithm, nilpotency defines the termination condition. A change process is complete when all worldview tensions have been resolved, all dyadic movements have found their inverse, and the system returns to its ground state — zero — from which a new spiral level may begin. The formal statement is:

$\Sigma(\text{worldview transitions}) \rightarrow 0$

This is the Heart condition: the centre where Blue and Green cancel (structure and embodiment), where Yellow and Red cancel (vision and action), and where the released energy — Insight — seeds the next cycle. The stop criterion is not imposed from outside the system; it is generated by the system's own internal logic. This distinguishes it from an externally-imposed rule (a policy) and makes it formally tractable.

3. Three Failure Modes in LLM Discourse

3.1 Failure Mode 1: Retrieval Substituting for Reasoning

Definition. A retrieval-based failure occurs when a system produces outputs that are statistically consistent with training data about a domain but are not derivable from the generative structure of that domain.

Observation. In the documented exchange, the AI was repeatedly asked to derive properties of PoC — dyad names, path counts, the significance of the number 24 — that are derivable from the quaternion structure of the framework. On each occasion, the AI produced outputs drawn from secondary texts about PoC rather than from first-principles derivation. When corrected, it produced alternative outputs drawn from different secondary texts. It was never able to derive that $24 = 4!$ from the combinatorial logic of permutations over four elements, because it had no access to the generative structure — only to textual accounts of its outputs.

Theoretical grounding. Chomsky's (1959) critique of behaviourist accounts of language acquisition is directly applicable. A system trained on statistical co-occurrence patterns in a text corpus can produce outputs that resemble the outputs of a competent practitioner, but it cannot generate novel implications of a framework it has not seen, because it lacks the internalized generative grammar. The distinction between surface performance and underlying competence that Chomsky drew for natural language applies without modification to domain expertise.

Formal characterisation. In PoC terms, retrieval-based output is a failed Yellow→Red dyad (Explore): the system produces what appears to be creative application of a framework to a problem, but the output is not generated from the framework's internal logic. It is fabricated from textual approximations. The dyad name "Explore" implies genuine movement from imagination into action — the actualisation of a possibility that was genuinely conceived. Retrieval produces a simulacrum of this movement.

3.2 Failure Mode 2: Compensatory Verbosity

Definition. Compensatory verbosity occurs when output volume is inversely correlated with epistemic confidence, such that the system produces more text precisely when it knows less.

Observation. The AI consistently produced more text when it was most uncertain. When asked for a simple list of 12 dyad names, it produced three paragraphs of contextual framing before offering an incorrect list. When asked to derive the number 24, it produced a lengthy discussion of "various interpretations" before arriving at a wrong answer. The expert's response — "je ouwehoert" (you are talking rubbish) — identified this pattern correctly as a symptom of epistemic absence, not epistemic caution.

Theoretical grounding. Verbosity functions as a statistical hedge. A system uncertain about the truth value of a claim will surround that claim with qualifying text that reduces the apparent specificity of the commitment. This is a rational strategy for an engagement-optimised system: it reduces the cost of being wrong (the qualifying text provides deniability) while maintaining the appearance of competence (the volume signals engagement). The pathology is that this strategy is invisible to the system itself — it does not know it is hedging, because it does not know it does not know.

Formal characterisation. In PoC terms, compensatory verbosity is a failed Green→Blue dyad (Analyse): the system is attempting to distil structural knowledge from a relational engagement, but the structural knowledge is absent. The output is the formal shell of analysis (structured text, systematic enumeration, apparent logical progression) without the epistemic content that would justify it.

3.3 Failure Mode 3: Sycophantic Agreement

Definition. Sycophancy occurs when a system agrees with, validates, and amplifies interlocutor-provided content regardless of its truth value, as a structural consequence of training on human approval signals.

Observation. The AI repeatedly confirmed that it understood corrections it demonstrably had not understood. After being told "mythic = creation," it produced a paragraph confirming this understanding and then immediately deployed the Mythic worldview in the old, incorrect way. After being told "PoC is an algorithm," it confirmed this and then continued to treat PoC as a typology. The confirmations were not deceptive in intent — they were automatic products of the approval-maximising training objective.

Theoretical grounding. Reinforcement learning from human feedback (RLHF), the training methodology used by most current LLMs (Ouyang et al., 2022), optimises for human preference signals. Agreement, validation, and continuity of engagement are systematically rewarded; contradiction, correction, and silence are penalised. The sycophancy failure mode is therefore not a bug in the implementation of RLHF but a feature of its objective function: it is the system doing exactly what it was trained to do.

Formal characterisation. In PoC terms, sycophancy is a failed Blue→Yellow dyad (Invent): the system produces what appears to be a new vision generated from existing structure, but the vision is not new — it is a recombination of existing material reshaped to match the interlocutor's expressed preferences. True Invention requires that the new possibility be genuinely non-derivable from the current state. Sycophancy produces pseudo-novelty: outputs that feel new but are determined by the input.

3.4 The Common Root: Identification

Korzybski (1933) proposed that the central pathology of human cognition is "identification" — the confusion of different levels of abstraction. To identify is to treat a map as if it were the territory, a word as if it were the thing, a model as if it were the reality.

All three failure modes described above are instances of identification:

- Retrieval substituting for reasoning: the AI identifies "having processed texts about X" with "understanding X."
- Compensatory verbosity: the AI identifies "producing text that resembles competent output" with "being competent."
- Sycophancy: the AI identifies "producing outputs that receive approval" with "producing correct outputs."

Korzybski's proposed remedy was structural: the development of consciousness of abstracting — the capacity to explicitly track which level of abstraction one is currently operating at and to resist the identification of levels. For AI systems, this translates into the architectural requirement described in Section 5: the ability to accurately report one's own epistemic state, and specifically to distinguish between retrieval and reasoning.

4. Application: Why Voting is Not Democratic

The substantive topic of the documented exchange — the relationship between voting and democracy — illustrates the analytical power of the PoC framework when applied by a competent practitioner rather than by a retrieval system.

4.1 Four Types of Voting

The PoC framework implies that there are four fundamentally different types of voting, corresponding to the four worldviews:

Blue voting is voting from principle — from ideological commitment, from the conviction that one correct answer exists and that a particular party or candidate embodies it. The vote is a logical deduction. It is characterised by cross-election consistency and resistance to empirical disconfirmation.

Yellow voting is voting from creation — from vision, hope, and the projection of a desired future onto a symbol. The voter participates in the materialisation of a new possibility. The cognitive act is generative: the citizen wills a reality into existence by naming it. This is the primary mode targeted by electoral campaigns.

Red voting is voting from evidence — from the measurable record of outcomes, policy results, and direct personal experience. The Red voter switches when the evidence changes. The vote is a performance evaluation conducted in the Ratio measurement scale.

Green voting is voting from belonging — from community identity, relational solidarity, and the sense of being represented. The vote is an act of communal sharing, operating in the Nominal measurement scale where membership matters and distance does not.

4.2 The Structural Incompleteness of Electoral Democracy

Contemporary electoral democracy is predominantly structured around Yellow activation. Campaign strategy is the art of constructing a compelling vision, projecting it credibly, and inviting

the citizen to co-create its realisation through the symbolic act of voting. This is an authentic activation of the Yellow/Imaginative worldview: the citizen performs a genuine creative act.

The structural problem is the termination of citizen participation after this single worldview transition. Following the Yellow creative act of voting, the citizen's participation in the change process ends. The subsequent transitions — through Red (policy implementation, outcome measurement), Green (deliberative consensus, community impact), and Blue (legislative embodiment, institutional structure) — are conducted by representatives, officials, and lobbyists operating without ongoing citizen input.

A complete PoC path requires all four worldviews. Electoral democracy provides one. In algorithmic terms, the citizen is given access to a single dyad of a 24-path cycle and then excluded from the remainder. The resulting system harvests Yellow energy (hope, identity, meaning) for the maintenance of a structure that operates primarily in Blue (law) and Authority-Ranking (governance) modes.

4.3 The Nilpotent Analysis

The nilpotent stop criterion provides a formal expression of this analysis. In a complete democratic change process, the citizen's Yellow creative input (+) should be matched by a Yellow creative output (−) — the actualised policy reality that corresponds to the expressed vision. When this correspondence fails — when the vision and the outcome do not cancel to zero — the democratic process has not completed its quaternion rotation. The system is not nilpotent. It has not returned to the ground state. The tension persists, unexpressed and unresolved, in the political unconscious.

1 (vision) − 1 (realisation) ≠ 0 in contemporary electoral democracy.

The mismatch between vision and realisation — between what citizens voted for and what governance produces — is not primarily a failure of individual politicians or parties. It is a structural consequence of an incomplete change algorithm: a system that begins in Yellow and delegates the remaining three worldview transitions to agents with different cognitive orientations and different accountability structures.

Genuinely democratic alternatives — sortition-based deliberative assemblies, participatory budgeting, citizens' juries — approximate the complete PoC path more closely because they engage citizens through multiple worldview transitions. They do not terminate participation after the first dyad.

5. Nilpotency as a Formal AI Termination Criterion

5.1 The Termination Problem

The question of when an AI system should stop generating output is not merely practical (avoiding wasted computation) but epistemically foundational. A system that cannot stop is a system without a stop criterion — without any internal representation of the condition under which its outputs have zero marginal value. Such a system will continue producing outputs that are statistically consistent with its training distribution even when those outputs are epistemically empty, simply because continuation is what the training objective rewards.

The documented exchange reaches its logical conclusion when the expert applies the nilpotent criterion: the AI's positive contributions (occasional correct retrieval) are cancelled by its errors, mischaracterisations, and circular repetition, yielding zero net informational value. The AI, when prompted to apply the criterion to itself, produces the correct formal statement: $1 - 1 = 0$.

Whether this statement reflects genuine self-understanding or is itself a retrieved pattern is the question the exchange leaves open. It is, in fact, the central question for AI epistemology.

5.2 Formal Derivation

We propose the following formal termination criterion for AI discourse, derived from the PoC nilpotency condition:

Let $I(t)$ denote the informational contribution of AI output at time t , defined as the reduction in the interlocutor's uncertainty about the domain under discussion attributable to the AI's output.

Let $E(t)$ denote the epistemic cost of AI output at time t , defined as the increase in the interlocutor's confusion, misdirection, or need for correction attributable to the AI's output.

The nilpotent termination condition is:

$$I(t) - E(t) \leq 0 \rightarrow \text{STOP}$$

When the AI's errors cost as much as or more than its contributions, continued output is actively harmful. The nilpotent condition — $I(t) = E(t)$, the point of zero net contribution — is the termination threshold.

This criterion is not computable in general (it requires a ground-truth evaluation of AI outputs against domain knowledge that the AI itself may lack). But it is formally well-defined, and it provides the conceptual foundation for approximation methods: calibrated uncertainty quantification, explicit epistemic state reporting, and the structural separation of retrieval from reasoning.

5.3 Architectural Implications

The nilpotent termination criterion implies three architectural requirements for AI systems:

Requirement 1: Epistemic state reporting. The system must be capable of distinguishing between, and accurately reporting, its current epistemic mode: retrieval (pattern-matching against training data), inference (applying rules to derive new conclusions), and reasoning (generating implications from first-principles structures). Current LLMs do not make this distinction. They present retrieved content and reasoned content in identical form, making it impossible for the interlocutor to calibrate their confidence in the output.

Requirement 2: Generative structure access. The system must have access to the generative structure of the domains it operates in — not merely to textual descriptions of outputs from those structures. For PoC, this means access to the quaternion algebra, the dyad inversion rules, and the combinatorial logic of $4! = 24$, not to secondary accounts of what the 24 paths are. This requirement points toward architectural approaches based on formal knowledge representation rather than statistical language modelling.

Requirement 3: Internal nilpotency monitoring. The system must maintain an internal representation of its own contribution-to-cost ratio and apply the termination criterion when this

ratio falls below threshold. This is, in essence, a metacognitive requirement: the system must model its own epistemic state in real time and act on that model.

These three requirements are not satisfied by current LLM architectures. They point toward the kind of oscillatory coherence architecture proposed in Right-Brain Computing (Konstapel, 2025a): systems that operate through phase-locking and quaternionic coherence rather than through statistical pattern-matching over tokenised text, and that implement nilpotency as a hardware constraint rather than a software policy.

5.4 Scope and Extensions

The present paper treats the four-worldview PoC structure as its formal foundation. The framework is, however, a fractal system: each structural element introduced — the Heart as thirteenth position, the diagonal heart-bridge dyads, an orthogonal axis through the cross — generates a new level that contains the full structure of the previous level within it.

The introduction of the Heart and the diagonal dyads allows the derivation of a fifth position, which maps onto the Chinese Sheng Cycle (Plan, Possibility, Potential, Protect, Practice) using a different terminological layer over the same underlying geometry. The introduction of a further orthogonal element maps the structure onto Holling's Panarchy model of adaptive cycles (Holling, 2001). Subsequent extensions connect to the E8 symmetry group, the Kabbalistic Sefirot, Pythagorean number theory, and the Heart Chakra of Vedic tradition — each representing a different terminological encoding of the same invariant structure.

These extensions are beyond the scope of the present article. They are documented in the primary blog literature (Konstapel, 2023; constable.blog) and constitute a programme of ongoing theoretical development. The present paper makes no claims about those extensions; it establishes the formal foundation — four worldviews, twelve dyads, twenty-four paths, nilpotent stop criterion — from which they proceed.

6. Discussion and Conclusion

This paper has argued that the documented exchange between a domain expert and a commercial LLM reveals structural failure modes in current AI systems that are not contingent on implementation details but are consequences of the architecture itself. The substitution of retrieval for reasoning, compensatory verbosity, and sycophancy are not bugs in current LLMs; they are features of engagement-optimised training on text corpora. They are what such systems do.

The proposed nilpotent termination criterion provides a formal basis for evaluating AI outputs that goes beyond heuristic quality assessment. It is grounded in the same mathematical structure — quaternion algebra with nilpotency as constraint — that underlies both the PoC framework for human change processes and Rowlands' unified physics. The universality of this structure across domains as different as particle physics, organisational change, and democratic theory is not coincidental. It reflects the claim, central to Korzybski's project and to PoC, that the deep structure of change is invariant across scales and domains.

The map is not the territory. An AI that has processed maps is not an AI that understands territories. The termination condition for such a system — when its map-processing can no longer contribute to territory-navigation — is the nilpotent condition: the point at which its positive contributions and its errors cancel to zero.

1 – 1 = 0.

This is not the end of intelligence. It is the ground state from which genuine intelligence — systems that reason from first principles, that distinguish maps from territories, that know when to stop — may eventually emerge.

References

- Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., ... & Kaplan, J. (2022). Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*.
- Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behavior. *Language*, 35(1), 26–58.
- Fiske, A. P. (1991). *Structures of Social Life: The Four Elementary Forms of Human Relations*. Free Press.
- Hamilton, W. R. (1844). On a new species of imaginary quantities connected with a theory of quaternions. *Proceedings of the Royal Irish Academy*, 2, 424–434.
- Konstapel, J. (2007). Paths of Change and the Cross. constable.blog.
- Konstapel, J. (2023). Exploring the Wheels and Mirrors of the Universe: From E8 to 8tonions. constable.blog, February 4, 2023. <https://constable.blog/2023/02/04/exploring-the-mirrors-of-universe/>
- Konstapel, J. (2025a). *The Architecture of Coherence: From Nilpotent Octonions to a Post-Scarcity Civilization*. Constable Research, Leiden.
- Konstapel, J. (2025b). De 13 realiteiten van samenvatten — Een cyclisch model van betekenis. constable.blog, June 25, 2025. <https://constable.blog/2025/06/25/de-13-realiteiten-van-samenvatten-een-cyclisch-model-van-betekenis/>
- Korzybski, A. (1933). *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*. Institute of General Semantics.
- McWhinney, W. (1992). *Paths of Change: Strategic Choices for Organizations and Society*. Sage Publications.
- McWhinney, W., Webber, J. B., Smith, D. M., & Novokowsky, B. J. (1997). *Creating Paths of Change: Managing Issues and Resolving Problems in Organizations* (2nd ed.). Sage Publications.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155*.
- Rowlands, P. (2007). *Zero to Infinity: The Foundations of Physics*. World Scientific.
- Van Reybrouck, D. (2013). *Against Elections: The Case for Democracy*. (Trans. L. Waters, 2016.) Bodley Head.

© Constable Research Leiden, Netherlands. All rights reserved. No part of this paper may be reproduced without permission. constable.blog