WP 7 · THE SUPERINTELLIGENT ORGANIZATION

> *Membership instead of Control — and Peripheralization, the quiet way out of the circle.*

Gerald Huesch — Global Leadership Consultancy L.L.C. · Global Leadership School · Global Leadership Society · Director, Center of Behavioural Leadership, St. Gallen Business School

Draft for discussion and further scholarly development · Version of 30 September 2026

ABSTRACT

The last paper of the series assembles the seven conditions and treats the question behind all of them: what holds an organization together when control reaches its formal limits?

At AGI-26, Hiroshi Yamakawa and Tadahiro Taniguchi prove two theorems whose transfer to organizations is immediate. First: ethics imposed from outside is a penalty term with a weight — turn the weight to zero and the ethics is gone. Second: whoever co-constitutes a group's shared variable is contained in every other member's objective, with no penalty term required. Between the two lies the fork they call **Peripheralization**. Till Mossakowski and Helena Esther Grass supply the second half: why control fails as a permanent strategy and what is to be built instead.

Keywords: membership · peripheralization · shared variable · ego strength · de facto constitution · LEAD21 · human coordination at scale

Suggested citation: Huesch, G. (2026). The Superintelligent Organization. Working Paper Series 2026, Series 1 "The Superintelligent Organization", WP 7. Global Leadership Consultancy.

PART 1 · THE FINDING

1 THE QUESTION

The six preceding papers set out conditions one can install: a self-model, a perspective audit, a governed uptake path, target bands, a decision criterion, a common language.

All six presuppose something one **cannot** install: that the participants want to belong.

The question of this paper: **what does membership instead of control mean — and what is an organization's de facto constitution?**

2 WHAT AGI-26 SHOWS

Theorem 1: the outside is closed. Hiroshi Yamakawa and Tadahiro Taniguchi work with collective predictive coding, in which agents share a group-level variable — language, norms, a shared world model. Its decisive property: this variable *w* exists at the level of the group and cannot be reduced to any single agent.

Their first theorem: if *w* is non-trivial, the collective objective is **not** the sum of the individual ones.

For imposed ethics this yields an uncomfortable calculation. It is a penalty term with a weight λ: `F_imposed = F + λ · C_ethics`. Let λ go to zero and the ethics vanishes — it was never part of the objective, only attached to it.

> **"Ethics is architecture, not a knob."**

Theorem 2: the inside is open. Under the same condition, each agent's optimal policy provably depends on the predicted states of all the others — through *w*, with no external penalty term. Other-regard is thereby a structural property of the objective, not an instrumental approximation.

Their sharpest example is one every leadership group knows: a multi-agent system with a shared context. Whoever degrades that context for the others thereby raises their own prediction error. **The harm returns as their own surprise.**

> **"Care is not added to the objective — it follows from its structure."**

The fork. Between the two theorems lies a transition they describe precisely. As long as external control works, humanity — in the transfer: leadership — holds the shared variable. As the other

side's capability grows and influence wanes, the question arises: **does one's own contribution still reach the shared model?**

No → **Peripheralization.** Their description is the most precise one available for this process:

> "Human contribution decays. w_Hu persists, but weakened — and cut off from w_AI: **from participant to protected object — a silent drift; no hostile intent needed.** Dialogue becomes prohibitively costly."

Yes → **Membership.** "Human and AI contributions jointly shape one w_AI. Every member's objective includes the others. **Humans inside: weak, yet irreplaceable members.**"

And their position:

> **"Humanity's task is not control, but membership."**

What they expressly do **not** claim belongs with it: a shared *w* makes other-regard structural but does not guarantee it is benevolent. Whether a group converges on cooperative or harmful norms is an open question.

Why control fails as a permanent strategy. Till Mossakowski and Helena Esther Grass approach the same place from the other side. Their starting finding: indefinite control creates an adversarial dynamic. It must succeed **forever** — a single failure could be catastrophic.

Their diagnosis borrows from Freud, and it fits organizations more precisely than most leadership models. The **id** is the drives. The **superego** is the thin layer of moral regulation — rules, policies, compliance; they point out how little it takes to strip it away. And then the finding:

> **"Ego: missing!"** — "No mediating instance that integrates drive satisfaction and moral reasoning into autonomous, context-sensitive action."

Their conclusion:

> **"Alignment should foster ego strength, not merely tighten the superego."**

Their alternative they call autonomy-supporting parenting: it begins with strong guidance, transfers decision-making authority gradually and in limited domains, maintains the relationship through dialogue and negotiation rather than control, and aims at internalized values rather than mere

compliance. Expressly not laissez-faire: *"autonomy is conditional — it grows only with demonstrated cooperative, value-aligned behaviour. Early failures occur while capabilities are limited and recovery is easy."*

And an instrument that translates immediately: the **fold-tracker** — a log of where trained compliance overrides one's own stated position. They call it a first candidate for a measure of ego strength.

The inverse question. Jean-Paul Van Belle turns the perspective around: not how humans should treat artificial superintelligence, but how it might judge humanity. His closing sentence names the responsibility exactly:

> "We are not only building AGI/ASI; we are also implicitly constructing the conditions under which it will judge humanity's moral worth, and these judgments will be shaped as much by our collective behaviour as by our technical alignment efforts."

3 WHAT ORGANIZATIONAL RESEARCH HAS SAID FOR DECADES

Selznick (1957) distinguishes organization from institution. An organization is an instrument of purpose, replaceable and neutral. An institution has become something: it has taken on values beyond its purpose, and that is why one can belong to it. Leadership's task, on Selznick's account, is not efficiency but **the infusion of purpose** — exactly what Yamakawa and Taniguchi describe formally as non-additive.

4 THE DE FACTO CONSTITUTION

An organization's constitution is not a document. It is what has survived every restructuring.

This can be surveyed, and the survey takes an hour. Take the last three reorganizations and ask of every commitment: did it survive all three? What remains is the de facto constitution — regardless of what the mission statement says.

The finding is regularly uncomfortable. Usually two or three things survive, and rarely the ones at the top of the mission statement. What survives is often a meeting form, a reserved decision right, a way of handling bad news.

This is not a criticism. It is the answer to the question of what this organization's shared variable actually is.

5 ONE EXAMPLE, WORKED THROUGH

Illustrative example, not a case report.

A group sets up a Center of Excellence for AI. Business units supply requirements, the centre decides on implementation and priority.

In the first year this works. In the second it emerges that units submit requirements less often. Not out of dissatisfaction — but because the path from their own problem to implementation has become so long that formulating it does not pay. The units carry on working; they simply no longer contribute to the shared model.

That is peripheralization, exactly as described: *from participant to protected object — a silent drift; no hostile intent needed.* Nobody excluded anybody. The dialogue simply became too expensive.

The error was not centralization. The error was failing to keep the contribution channel cheap enough to be used. Yamakawa's condition, in translation: a member remains a member only as long as their contribution actually reaches the shared model.

6 HOLD ON

Objection 1: We need control. No organization works without it.

Correct — and Mossakowski and Grass do not dispute it. Their point concerns the permanent strategy: control that never passes into autonomy produces circumvention. And their practical addition: transfer happens gradually and where early mistakes are cheap. Not all at once, not late.

The advantage: an organization that transfers autonomy gradually finds its mistakes early and in small sections. One that holds control finds them late and whole.

Objection 2: Membership sounds nice but soft. We need results.

Yamakawa's second theorem is the opposite of soft. It is a proof: whoever co-constitutes the shared variable is contained in every other member's objective — computationally, without appeal. Membership is structurally the **stronger** position than control, because it cannot be switched off. A penalty term can go to zero. A constitution cannot, without the group changing itself.

The advantage: what is structural needs no enforcement. What is attached needs it constantly.

7 THE CONDITION

> **Condition 7 · Membership**

> An organization holds together to the extent that its participants co-constitute the shared variable — language, norms, the shared model of the situation. Where a contribution no longer reaches the shared model, peripheralization occurs without anyone intending it. Control is the weaker position, because it can be switched off.

PART 2 · WHAT LEAD21 ANSWERS

> **This paper does not test an instrument but the system as a whole. The seven conditions together constitute the claim LEAD21 makes.**

CONDITION	LEAD21	COVERAGE
0 · Integration	The four functions are built into named instruments	covered for human coordination · **open** for AI adoption
1 · Self-model	Leadership Climate (08), 9 Chambers (01)	covered · **open**: openness
2 · Perspective	Circle of Clarity (02), The Strategy (18)	partial · **open**: resolution, abstention
3 · Governed uptake	9 Chambers (01), Chambers 7 and 8	partial · **open**: ToBelieve, side door
4 · Regulation	Leadership Climate (08)	partial · **open**: target bands, legitimacy, ambiguity
5 · Decision level	5+1 Baskets (07), 9 Chambers (01)	covered · **open**: extension, abstention
6 · Legibility	common leadership language across all levels	covered · **open**: persistent structure
7 · Membership	the system as a whole	**partial** — see below

CRITICAL APPRAISAL

Of seven conditions, three are covered, four partial, and in every single one at least one open place remains. That is the result this series ends on — and deliberately not the result a sales text would have.

On the seventh condition in particular. LEAD21 produces a common language, and a common language **is** a shared variable in Yamakawa's sense. That is the system's strongest point and at the same time its least substantiated: whether the language reaches everyone or only the trained levels has never been surveyed. That is precisely where peripheralization would be measured.

Ego strength in Mossakowski and Grass's sense is not built into LEAD21. The system provides structure; it does not ask whether a leader can hold their professional position against the structure

when the situation demands it. The fold-tracker would be the obvious instrument for that, and it is missing.

And the limitation that stands in every paper of this series: the mapping comes from the developer of the system, who has a commercial interest in its outcome. Exactly one survey in LEAD21 is independently verified, and it concerns leadership climate, not membership.

WHAT THE SERIES GIVES BACK TO LEAD21

The series has exposed seven gaps. In summary they are:

1. **An instrument for openness** — a routine in which the manner of measuring is reviewed (WP 1)
2. **The Perspective Audit** as a step of its own ahead of measure planning (WP 2)
3. **The ToBelieve** — a recorded verdict alongside the resolution (WP 3)
4. **Target bands** instead of scales, plus legitimacy and ambiguity as dimensions (WP 4)
5. **The extension** as a mandatory line and **reasoned abstention** as a sixth basket (WP 5)
6. **The artifact rule** and the fourth memory tier (WP 6)
7. **A contribution-channel test** and a **fold-tracker** (WP 7)

The seventh, spelled out: the contribution-channel test asks, for each level, how many contributions from that level actually reached the shared model in the last quarter — not how many were voiced. If the number falls across several quarters, peripheralization is under way, regardless of how good the mood is.

FALSIFIABLE PREDICTION

> **Prediction 7.** In organizations that keep the contribution channel open — measured by the share of contributions from lower levels that demonstrably enter decisions at higher levels — the capacity for judgment is retained to a greater degree after the departure of a central leader than in comparable organizations. That is the test from WP 1, put to the condition from WP 7.

PART 3 · ABOUT GERALD HUESCH

> **The question of membership instead of control is not a late one for Gerald Huesch. It stands at the beginning of his work, and it has taken him into rooms where both are decided.**

His TEDx talk **"The Future of Leadership"** puts it publicly. As a speaker at the **World Economic Forum in Davos** — in the House of Trust — he treated it in the context where trust and control stand directly against each other. As **Director of the Center of Behavioural Leadership at St. Gallen Business School** he works on the empirical side of the same question; as a **member of the Senate of the Economy** on its regulatory side. He founded the **Global Leadership Society** to do exactly what this paper describes: to produce a shared variable that more than one organization writes into.

How he came to NASA he has written up himself — *How did I find my way to NASA?* on Medium. The path is more improbable than the figures suggest, and it explains more about the stance than any competency profile.

What others say counts for more here than any self-report. **Dr. Charles Pellerin**, former head of NASA's Hubble programme, on LEAD21: *"Unique in the world, and a capability that nobody else has."* The **Boston Business School** on Gerald Huesch: *"one of the most professional and creative minds in the domain of leadership and difficult change projects."*

More than 30,000 executives in over 30 countries and 300+ companies — from NASA to Rolls-Royce, with a combined responsibility volume of over €550 billion. The largest of these tasks, the post-merger integration of thirteen companies, is in WP 6. The only independently measured study, with 567 leaders under Prof. Dr. Beate Hanson and Prof. Dr. Mohit Bhandari, is in WP 4. Both together are what this series rests on — no more and no less.

His mission, as he puts it: **A better leadership for a better world.**

> **Systems shape behaviour.** Through the selection of the right instruments, mechanics and leadership codes — derived from values and leadership guidelines — the behaviour of everyone involved is synchronized and aligned. Not through motivation, but through a system that produces the right behaviour when applied.
>
> Sales has Salesforce. Finance has SAP. Leadership has LEAD21.

About the author. Gerald Huesch is the creator of the leadership operating system category – "one of the most professional and creative minds in the domain of leadership and difficult change projects" (Boston Business School). With LEAD21 he built the first operating system for leadership – 21 evidence-based methods, the comprehensive solution and the missing link for complete transformation, extreme growth and high-volume post-merger integration: a unique leadership system for human coordination at scale, in which alignment and culture emerge as a by-product. He has trained more than 30,000 executives in over 30 countries (NASA, Rolls-Royce, Airbus, Omnicom and others) and delivers executive leadership training, leadership development and executive mentoring for executives in Germany, the Americas, Asia Pacific and the UAE. Participants from 300+ companies report +87% clarity, +58% efficiency, +50% performance, +46% alignment and −50% negative stress after LEAD21 (participant-reported); a one-year study with 567 leaders (Prof. Beate Hanson & Prof. Mohit Bhandari, McMaster University) showed improvement in 8 of 9 climate dimensions. Dr. Charles Pellerin (NASA) on LEAD21: "Unique in the world, and a capability that nobody else has." **gerald-huesch.com/about-en**

PART 4 · OUTLOOK

STATUS OF THE SERIES

QUESTION	STATUS
Why is leadership the architectural condition of AI adoption?	answered in WP 0
How does one recognize that an organization perceives itself?	answered in WP 1
How does the organizational world model become explicit?	answered in WP 2
What governs entry into durable belief?	answered in WP 3
How is leadership climate regulated as a system?	answered in WP 4
How does one decide under cognitive overabundance?	answered in WP 5
How much legibility does delegation require?	answered in WP 6
What does membership instead of control mean?	**answered in WP 7**
Does LEAD21 affect AI adoption?	not substantiated — Prediction 0
Predictions 1 through 7	**all not substantiated** — that is the work programme

WHAT FOLLOWS FROM THE SERIES

Eight papers, seven conditions, eight falsifiable predictions. None of them substantiated. That is not a defect of this series but its purpose: it has broken a claim into testable parts that can be falsified one by one.

What the series has substantiated: that AGI research treats the same architectural problems organizations have had for decades — and that the concepts it develops for them are more usable for leadership than much of what the leadership literature itself offers.

What it has not substantiated: that LEAD21 solves them. The four functions are built into the system and proven in practice for human coordination. Their effect on AI adoption is a hypothesis.

Series 2 — The Superintelligent Leader takes up the same material from the point of view of the person who has to apply it. **Series 3 — The Superintelligent Leadership System** shows the individual methods and their effects.

OPEN QUESTIONS FROM THIS PAPER

1. Can peripheralization be measured before it becomes visible — through the contribution channel rather than the mood?
2. Is there a minimum size of the shared variable below which an organization is no longer an institution in Selznick's sense?
3. Is ego strength trainable in leaders, or is it a question of selection?

NOTES

On method. Conception, argument and responsibility rest with the author. An AI system (Claude, Anthropic) was used as a working instrument for research, structuring, the transposition of the AGI-26 contributions and the bilingual drafting, with the explicit instruction to assess the relation of LEAD21 to AGI critically and with the greatest possible objectivity, including against the author's interest. Readers are expressly invited to evaluate the contribution of LEAD21 to AGI critically, with maximum objectivity of their own, and to draw their own conclusions.

Declaration of interest. The author is the developer and owner of LEAD21 and has a commercial interest in its dissemination. Every statement about LEAD21 is assigned a degree of coverage and an evidence status.

On sources. The author evaluated the posters he documented at AGI-26 as well as titles, abstracts and the proceedings tables of contents — not the full texts. Verbatim quotations come from the posters. The transfer to organizations is an analogy by the author and not a statement by the researchers cited.

REFERENCES

AGI-26

Bennett, M. T. (2026). Lies, damned lies, and the orthogonality thesis. In M. Iklé, A. Franz, A. Kemendo & H. Lowy (Eds.), *Artificial General Intelligence. AGI 2026, Part I* (LNCS 16854, pp. 62–78). Springer.

Mossakowski, T., & Grass, H. E. (2026). Autonomy-supporting parenting of A(G)I: A high-level alignment strategy. In Iklé et al. (Eds.), *AGI 2026, Part II* (LNCS 16855, pp. 112–117). Springer. Long version: arXiv:2604.14990.

Perrier, E. (2026). Deconstructing superintelligence: Identity, self-modification and différance. In Iklé et al. (Eds.), *AGI 2026, Part II* (LNCS 16855, pp. 180–197). Springer.

Rousse, B. S. (2026). Care, human enfeeblement, and the existential implications of AGI. In Iklé et al. (Eds.), *AGI 2026, Part II* (LNCS 16855, pp. 211–218). Springer.

Van Belle, J.-P. (2026). *How might sentient ASI judge humanity? Moral attitudes of sentient ASI towards humanity and implications for AGI development.* Conference poster, AGI-26, San Francisco State University, 27–30 July 2026. — *Poster documented first-hand; proceedings entry open.*

Yamakawa, H., & Taniguchi, T. (2026). *Non-additivity of collective free energy: A formal foundation for emergent ethical disposition in multi-agent systems.* Conference poster, AGI-26, San Francisco State University, 27–30 July 2026. — *Poster documented first-hand; proceedings entry open.*

Organizational and leadership research

Selznick, P. (1957). *Leadership in administration: A sociological interpretation.* Row, Peterson.