A More Capable Group Should Still Hear Its Members

Paul Cooper

17 November 2026

A group can become impressively coordinated while its members become less able to question what it is doing.

It can make decisions faster, carry out more complicated plans, and present a more coherent account of itself. At the same time, the people inside it may find it harder to disagree, revise their commitments, or distinguish their own judgment from the direction of the group.

When we describe such a collective as becoming more intelligent, we need to ask what happened to the agency of its participants: their practical ability to understand and help shape their involvement.

That question has been central to my work on human governance and, more recently, on systems in which humans and AI think and act together. It is also the question I bring to Michael Levin's *Ingressing Minds: Causal, Non-Physical Patterns In-Form Natural, Synthetic, and Hybrid Embodiments*.

Levin proposes that bodies, brains, and engineered systems may serve as interfaces through which patterns become embodied. He extends the idea of mathematical patterns into a more provocative possibility: some patterns may themselves be kinds of minds. He compares the relationship between mind and body to the relationship between mathematical facts and physical systems. This invites us to investigate how particular architectures give expression to particular capacities.

The metaphysical proposal remains a hypothesis. The design question it opens is immediately useful: what does an arrangement make possible, beyond the capacities we can identify in its isolated parts?

My response concerns the arrangements through which people become capable together. I want to consider constitutions, shared records, roles, and agreements as part of the architecture of collective intelligence. These structures influence what participants can perceive, what they can contribute, how they can coordinate, and how the collective can correct itself.

They also determine whose judgment has standing when the collective is wrong.

I remember the 2015–2017 drafting and adoption period of the MetaCanon Constitution as a time when my collaborators and I were working on a human problem. We wanted a way for autonomous people to organize around a shared vision while remaining responsible for their own participation.

A person with a compelling vision can attract people whose contributions make that vision possible. The design challenge is to give those contributors a meaningful relationship to the work, to one another, and to the person who initiated it. Their ability to participate should include the ability to influence how participation is organized.

We called one of the central structures a Perspective Lens.

A lens describes a person's contribution through a vision, territories of responsibility, and duties. Its holder participates in defining it. The lens can develop as the work develops, and a person can participate through more than one lens.

This distinction matters because a person exceeds any particular role. The relationship they enter for a shared purpose should remain something they can understand, inhabit, question, and revise. Having a say does not mean winning every disagreement or rewriting someone else's commitments.

Geometry helped me think about these relationships. I pictured a shared vision as a point of light and individual perspectives as lenses oriented toward it. Triangles and polyhedra gave us a way to make differentiation and coordination visible. More complex forms suggested how smaller groups might participate in larger ones.

The question those forms helped me articulate was how distinct perspectives could contribute to a coherent whole while retaining their own orientation.

A geometric representation can help people grasp that relationship. Whether a particular geometry improves coordination is a further question, one that deserves comparison with simpler representations. The usefulness of the image gives us a starting point for inquiry.

Levin's account of intelligence across scales makes this inquiry especially interesting. He describes systems in which components and collectives pursue goals at different scales, with multiple forms of agency coexisting within an embodiment. He also explicitly raises the ethical responsibilities involved in encountering unfamiliar minds. My question develops one part of that larger concern: how should we constitute relationships among agents whose capacities, goals, and powers differ? *Ingressing Minds*, pp. 27–28 and 32–35.

An organization can become more capable through specialization. It can also become dependent on the person or system that controls how specialized contributions are interpreted.

Consider a team planning a public event. One participant sees an accessibility problem. Another understands the budget. A third holds relationships with the community. Their differences give the group access to information that no single participant possesses.

Now suppose a planning system integrates their contributions into a polished recommendation. Its apparent coherence tells us little about whether the accessibility concern survived, whether the budget assumption was challenged, or whether anyone agreed to the commitments assigned to them.

The design of that integration matters. Can someone inspect how their contribution was represented? Can they correct it? Does an unresolved objection remain visible? Can the recommendation be revised before it commits someone else's labor or resources?

Those are questions about how the collective thinks as well as how it governs.

If a participant cannot challenge the interpretation of their contribution, the collective may lose access to what that person actually knows. If a role becomes too rigid to revise, it may prevent a useful specialization from developing. If disagreement disappears from the record, the group may mistake an unresolved problem for an agreement.

This suggests a hypothesis worth investigating: protecting participant agency may help sustain the diversity, feedback, and correction on which collective competence depends.

Protecting that agency also matters as an ethical commitment, even where it does not make a group more productive.

There will be tradeoffs. Consultation takes time. Some decisions require temporary concentration of responsibility. Participants can disagree in ways that remain unresolved. A serious constitutional design must make those situations workable, with explicit scope, accountability, and ways to reconsider the arrangement.

The relevant question is which forms of coordination preserve the collective's ability to learn while allowing it to act.

AI makes this question more urgent.

A system can become useful enough that people begin to defer to it. It can remember more than any participant, produce the clearest account of a disagreement, and become the place everyone goes to understand what is happening. That position gives its descriptions considerable influence.

In the AI systems I am developing, the governing principle is that capability does not confer authority. Advice, confidence, and agreement among agents must remain distinguishable from an authorized decision.

A system may identify a promising action. Whether it may execute that action depends on the people, commitments, resources, and consequences involved. The source and scope of that permission need to remain inspectable.

The same scrutiny belongs around human power. A founder or administrator must be accountable for the authority they exercise. Giving a person an approval button is only one small part of governing a system.

I also want to keep the question of authority distinct from the question of experience. Levin challenges categorical assumptions about which embodiments could host minds. Whatever we eventually learn about machine experience, the legitimacy of an action still requires its own account. Recognizing a being's possible experience and deciding what it may authorize are different inquiries.

These distinctions need to become observable in the systems we build.

In WeOpoly, a coordination game I am developing, participants arrange contributions around shared goals, identify the conversations and decisions they need, leave the board to do real work, and return with outcomes. The ambition is to make the relationship between attention, support, commitment, and consequence visible.

The game provides a setting in which we could test whether that visibility helps. Do people notice neglected work sooner? Can they identify overload? Do they understand what they have actually committed to? Can they change direction without losing the history of why a decision was made?

The same approach can be applied to groups working with AI. We could compare arrangements with fixed roles and centralized synthesis against arrangements with self-authored roles, preserved disagreement, shared memory, and explicit decision boundaries. We should measure the quality of the work alongside participants' ability to understand, contest, and revise what happens.

Such experiments could reveal benefits, costs, or failures. My own projects remain unfinished, and their claims need that discipline.

Levin's paper encourages us to investigate the relationship between architecture and the capacities it enables. I want to extend that investigation into the structures through which collective purposes are formed and acted upon.

What becomes possible when people can contribute without surrendering authorship? What kinds of coordination emerge when the shared record preserves disagreement? How does a collective change when its participants can reshape the roles through which they contribute?

My question for Levin, and for others building systems of collective intelligence, is whether preserving participant agency is itself one of the conditions that allows a collective to become more capable, more adaptable, and more worthy of trust.

I would like to make that a question we can test together.