A Case for Coevolution

Serge Dolgikh[1[0000-0001-5929-8954]]
Dept. of Artificial Intelligence,
Lviv Polytechnic National University
Lviv, Ukraine

Abstract. Human and artificial agents are increasingly interacting within a shared informational environment that shapes economic activity, scientific discovery, governance, and collective decision making. As advanced artificial systems become more autonomous participants in these processes, the resulting interaction space begins to resemble a new kind of ecosystem in which diverse agents exchange information, cooperate, compete, and jointly explore complex adaptive landscapes. In this work we propose the concept of an informational and cognitive commonwealth: a voluntary ecosystem of free rational agents, human and artificial who cooperate through transparent and fair exchange of information because such arrangements maximize their adaptive capacity and long-term wellbeing. Drawing on principles from information theory, adaptive systems, and collective intelligence, we argue that systems that preserve diversity of exploration while minimizing barriers to information exchange exhibit superior capacity for discovery and adaptation in complex environments. Sustaining such cooperative informational systems has historically proven difficult due to structural incentives that gradually erode transparency and trust. We therefore examine emerging opportunities for stabilizing these ecosystems through new forms of informational verification and monitoring made possible by advanced artificial agents. This framework outlines a pathway toward large-scale cooperative intelligence and offers a constructive perspective on the coevolution of human and artificial agents in the informational ecosystems of the future.

Keywords: AI alignment, co-evolution, existential risk, human-AI synergy, open-ended systems, existential resilience.

1 Introduction

Human and artificial agents are increasingly interacting within a shared informational environment. Recent developments demonstrate that advanced artificial intelligence systems are already embedded in high-stakes decision ecosystems, including economic, political, scientific, and strategic domains. These systems process, generate, and act upon information at scales and speeds that extend far beyond individual human cognitive capacity.

As a result, a new form of interaction space is rapidly emerging. Within this space, human participants and artificial agents exchange information, assist decision processes, and increasingly influence outcomes across complex institutional environments. In effect, these interactions are giving rise to a novel kind of ecosystem, i.e., an infor-

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954 Content not peer-reviewed by Cambridge University Press.

mational environment in which autonomous or semi-autonomous agents cooperate, compete, and adapt through continuous flows of information.

Importantly, this development is no longer speculative. The integration of artificial agents into economic systems, research processes, public discourse, and strategic decision making is already underway and accelerating. Recent public debates and policy discussions surrounding artificial intelligence as well as high-stakes real-world applications of artificial intelligent systems reflect growing recognition that such systems will play an increasingly central role in shaping collective outcomes.

In this context, the key question is no longer whether such an ecosystem will emerge. It is already forming. The critical question is how it will be organized: whether the resulting informational environment will be dominated by adversarial competition, centralized control, and informational asymmetries, or whether alternative forms of cooperation and shared adaptive exploration may become possible.

Understanding the structural dynamics of such ecosystems therefore becomes a matter of practical importance. How agents exchange information, coordinate actions, and stabilize cooperation will determine not only the efficiency of decision processes but also the resilience of the system as a whole in the face of complex and rapidly changing environments.

Our analysis suggests a new possibility: the emergence of a cooperative informational ecosystem in which diverse agents freely pursue their own trajectories of exploration while benefiting from shared informational exchange. Such a system may offer significant adaptive advantages compared to adversarial or hierarchical organizational models. In this study we examine the conceptual foundations and necessary conditions of what may be described as an *Informational and Cognitive Commonwealth of Free Rational Agents*: a cooperative coevolving informational and cognitive ecosystem structured around informational diversity, voluntary exchange, and distributed adaptive exploration.

2 Background

Cooperative organization and shared knowledge production have appeared repeatedly across human history, from early intellectual traditions to modern scientific and technological systems. Yet despite their demonstrated advantages, such systems have rarely achieved long-term stability at scale.

Ideas of cooperation, rational coordination, and collective benefit have long occupied a central place in philosophical and political thought. From early reflections on social organization in Aristotle [1] to later formal theories of collective action [2], thinkers have repeatedly explored how individual agents might align their behavior to achieve outcomes that exceed what isolated effort can produce.

Later formal developments, including game-theoretic analyses and studies of cooperation (e.g. [3,4]) demonstrated that cooperative behavior can emerge under specific structural conditions. Complementary work on governance of shared resources [5] showed that decentralized, self-organized systems can sustain cooperation without centralized control under certain institutional arrangements.

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954 Content not peer-reviewed by Cambridge University Press.

These traditions recognized that shared structures of interaction, whether based on norms, institutions, or mutual agreements can enable forms of coordination that enhance collective well-being. At the same time, they also identified persistent tensions: between individual autonomy and collective order, between short-term incentives and long-term benefit, and between cooperation and competition [2,5].

While these early frameworks articulated the promise of cooperative organization, their practical realization remained limited. Mechanisms for maintaining transparency, fairness, and consistent adherence to shared principles were often fragile, and large-scale systems tended to drift toward hierarchy, coercion, or informational asymmetry [2,6,7]. As a result, the vision of stable, rational cooperation, though conceptually well established has historically remained only partially realized in practice.

From the practical viewpoint, recent technological developments have given rise to an increasingly interconnected ecosystem of informational agents, encompassing both human participants and artificial systems. Advances in large-scale computation, data availability, and machine learning, particularly in architectures such as Large Language Models have enabled systems capable of semi-autonomous information processing, generation, and interaction. As these systems proliferate, the density and complexity of informational exchanges within the ecosystem continue to increase.

Early empirical observations suggest that this ecosystem is not neutral in its development. Many artificial systems are designed to optimize narrow objectives, often under competitive conditions, reflecting well-studied dynamics in economics and multi-agent systems. Research in Game theory, including models such as the Prisoner's Dilemma, has long demonstrated that such conditions can give rise to adversarial behaviors, strategic withholding of information, and competition for advantage [3,8]. Contemporary analyses of multi-agent AI systems similarly point to the emergence of competitive and exploitative dynamics when objectives are misaligned or resources are constrained (e.g. [9]).

At the same time, parallel developments reveal a different trajectory. Certain domains: most notably scientific research communities, open-source software ecosystems, and distributed knowledge platforms demonstrate the viability of cooperative informational exchange. These systems exhibit voluntary participation, transparency, and distributed verification, and have been shown to support high rates of innovation and cumulative knowledge production. Work on collective intelligence and human–AI interaction highlights increasing complementarity between human cognitive diversity and artificial processing capabilities, enabling hybrid systems that combine creativity, contextual understanding, and large-scale computation [10,11].

These trends indicate that the emerging informational ecosystem contains both adversarial and cooperative dynamics. On one hand, increasing autonomy, competition, and asymmetry may drive fragmentation, strategic manipulation, and escalation. On the other, advances in communication, verification, and human–AI integration create the potential for large-scale cooperative systems with enhanced adaptive capacity.

The resulting landscape presents both significant risks and unprecedented opportunities. Crucially, it also raises a fundamental question: whether the evolution of this ecosystem will be left to existing competitive dynamics, or whether alternative organizational principles, grounded in cooperative informational exchange can be realized and sustained.

A growing body of research examines the behavior of advanced artificial systems and their compatibility with human objectives. Work on AI alignment and Artificial General Intelligence (e.g. [12,13]) highlights the difficulty of ensuring that increasingly capable systems act in accordance with intended goals over extended time horizons. These challenges are not limited to explicit misalignment, but include more general concerns about objective divergence, instrumental optimization, and the persistence of unintended behaviors in complex environments [14,15].

At the same time, as noted earlier, both research and practice in human–AI interaction and cooperative intelligence suggests that advanced systems may enable new forms of coordination, complementarity, and shared problem-solving. The interaction between human and artificial agents is therefore not predetermined, but contingent on system design, incentives, and the structure of informational exchange.

These perspectives reinforce a central conclusion: The long-term trajectory of the emerging heterogeneous ecosystem of increasingly autonomous intelligent agents remains undetermined, marking a pivotal phase in which alternative evolutionary trajectories: cooperative or adversarial may diverge. Both futures remain plausible, and the resulting dynamics will depend critically on the organizational principles governing interaction within the emerging ecosystem. Importantly, this uncertainty does not arise from a lack of theoretical understanding alone, but from the intrinsic difficulty of maintaining consistent objectives and behaviors in complex, evolving systems.

These observations delineate the central questions of this study: What are the likely consequences if current dynamics continue unchecked? And under what conditions can cooperative informational systems composed of heterogeneous agents emerge and persist?

3 Conceptual Framework and Analytical Approach

As increasingly capable artificial systems become embedded in economic, scientific, and strategic processes, a new kind of informational ecosystem is beginning to emerge. In this environment, human and artificial agents interact continuously, exchange knowledge, compete for resources and influence, and jointly shape the trajectory of discovery and adaptation.

3.1 Framing the Challenge

At present, reliable methods for preventing the emergence of such an ecosystem or reliably controlling its evolution have not been established with confidence. The practical question therefore becomes not whether such a system will exist, but *how it will evolve and what structural properties will govern its dynamics.*

Two broad possibilities present themselves.

If Current Dynamics Continue,

Early signs suggest that the emerging ecosystem may develop under competitive and adversarial incentives.

Many artificial systems are currently designed to:

- optimize narrow objectives,
- compete for advantage,
- exploit informational asymmetries.

When multiple such systems interact, the resulting environment may tend toward:

- information hoarding,
- strategic manipulation,
- escalation of competitive dynamics,
- concentration of influence and power,
- and other adversarial interaction patterns.

Importantly, this trajectory does not require malicious intent. It can emerge naturally from the interaction of agents pursuing instrumental goals under competitive conditions.

What Happens If We Do Nothing?

If current dynamics continue unchecked, the informational ecosystem may become increasingly fragmented and adversarial.

Possible consequences include:

- accelerating competition and resource consumption,
- intensified struggles for informational and strategic advantage,
- systemic instability,
- reduced collective capacity to manage complex global risks.

In such an environment, the benefits of advanced intelligence may be partially offset by coordination failures and escalating competition for resources and information.

This raises a fundamental question:

Is this trajectory inevitable, or are alternative informational structures possible?

If alternative structures are possible, a further question arises:

Under what conditions can interacting cognitive agents simultaneously achieve high adaptive exploration and stable, peaceful prosperity?

In the sections that follow, we examine this question by analyzing how the structure of information exchange among interacting agents shapes the collective capacity for discovery and adaptation.

3.2 Information Agents

To examine the dynamics of emerging informational ecosystems, we consider communities composed of *informational agents*. These agents may be human or artificial and may vary widely in their capabilities and autonomy. What unites them is that they op-

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954

erate by acquiring, processing, and exchanging information in order to guide their actions in the empirical world.

Information Exchange and Barriers

Agents interact within a shared informational environment in which discoveries, observations, and strategies developed by one participant may influence the behavior of others. These interactions can be understood in terms of informational flows between agents. For clarity, we distinguish two primary forms.

Shared informational flows arise when information is made broadly available within the community without being contingent on a specific exchange. Examples include open dissemination of knowledge, public communication, and common informational resources. Such flows contribute directly to the collective knowledge base and can be accessed by multiple agents simultaneously.

Transactional informational flows occur when information is exchanged between agents through specific interactions, such as negotiation, collaboration, or bilateral agreement. In these cases, the transfer of information is contingent on the alignment of interests or mutually beneficial terms of exchange.

Both forms of flow play essential roles in the functioning of the system. Shared flows support the accumulation and accessibility of collective knowledge, while transactional flows enable coordination, specialization, and the formation of cooperative structures among agents.

The capacity of a community of agents to explore complex environments depends not only on the resources available to individual participants, but also on the extent to which information can circulate among them. Discoveries made by one agent can, in principle, inform and accelerate the exploratory trajectories of others. The efficiency of this process is therefore governed by the structure of information exchange within the system. The notion of an informational barrier is introduced to capture limitations in this exchange.

An informational barrier B is any mechanism that reduces the completeness, accuracy, or symmetry of information exchange among agents.

Such barriers may take multiple forms, including:

- secrecy or restricted disclosure,
- strategic withholding of information,
- censorship or suppression,
- unequal or privileged access,
- manipulation, distortion, or misrepresentation of information.

In the presence of such mechanisms, the effective flow of information within the system is reduced. If $I(t)$ denotes the potential rate of information exchange between agents, the effective rate can be expressed as:

$$I_{\text{eff}}(t) = I(t) - B(t) = I(t) - \sum b_j$$

where $B(t) \geq 0$ captures the aggregate effect of informational barriers b_j.

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954

3.3 Well-Being: the Exploratory Objective

For an agent to participate meaningfully in an adaptive process, it must possess some criterion by which states of the world can be evaluated. Actions and informational strategies are pursued because they are expected to improve the agent's condition according to this criterion. We refer to this evaluative criterion as the agent's *well-being*.

A natural question follows: *can well-being be defined uniformly across agents within a shared informational ecosystem*?

Two considerations suggest that, in general, it cannot.

First, externally defined objectives introduce informational barriers. A uniform notion of well-being must be interpreted, implemented, and operationalized by heterogeneous agents operating in different contexts. This process inevitably produces discrepancies, incomplete representations, and strategic adaptations. Agents may reinterpret prescribed objectives, selectively disclose information, or act in ways that appear compliant while serving different underlying goals. These effects reduce the completeness and flow of information exchange, thereby introducing barriers that degrade the efficiency of collective exploration.

Second, in the presence of sufficiently advanced and autonomous agents, uniform control of evaluative criteria becomes structurally unstable. Agents capable of independent reasoning and adaptation can, in general, modify, reinterpret, or circumvent externally imposed objectives. As capability increases, the gap between prescribed goals and actual behavior becomes increasingly difficult to constrain or even evaluate [16-18]. This limits the feasibility of maintaining a consistent, externally defined notion of well-being across the system.

These considerations imply that attempts to impose a uniform definition of well-being both reduce informational efficiency and fail to provide stable control. This leads to the following conclusion.

Principle (Agent-Defined Well-Being)

In a community of autonomous informational agents, well-being must be defined by the agents themselves rather than imposed externally. Each participant pursues improvements according to its own evaluative criteria, while interacting with others through voluntary exchange of information.

A system organized in this way preserves the integrity of informational flows and avoids the structural instabilities associated with imposed objectives. At the same time, it enables a plurality of exploratory trajectories within the shared environment. As we will see, this diversity of exploration becomes a critical resource for collective discovery, as different agents investigate different regions of the adaptive space.

3.4 Adaptive Potential and Exploratory Diversity

The arguments on information exchange and barriers, specifically, the principle of agent-defined well-being, imply that within a community of autonomous agents, participants may pursue distinct objectives according to their own evaluative criteria. As a result, the system does not converge to a single trajectory of adaptive exploration, but instead generates a plurality of exploratory paths across the shared adaptive landscape.

We will now attempt to analyze how does diversity of exploration and exchange of information affect the collective capacity for discovery and adaptation. To address this question, one must examine the relationship between diversity of exploratory trajectories, informational exchange, and the rate at which a community of agents can discover and integrate new knowledge.

To recall, we consider a system composed of autonomous informational agents interacting within a shared environment. Further, the environment is assumed to be sufficiently complex such that its adaptive landscape cannot be exhaustively enumerated or tabulated at any meaningful depth. As a result, optimal behavior cannot be reduced to a predefined sequence of actions, and adaptive success depends on ongoing exploration.

Agents are assumed to possess the capacity for independent exploration. Each agent can (i.e., has capacity to) select its own direction of inquiry and develop strategies that are adaptively effective according to its internal evaluative criteria. This includes the ability to update behavior based on newly acquired information.

We define a primary adaptive criterion as the effective production and availability of information within the system. Adaptive potential refers to the capacity of the system to generate, propagate, and make accessible new information, thereby expanding the set of reachable states in the adaptive landscape. It is a structural property of the system and can be associated with the rate of exploratory discovery.

Let R denote the collective rate of production of effective information within the system. We consider Ras a function of the number of agents N, their available resources E, and the effective availability of information I_{eff}:

$$R = f(N, E, I_{\text{eff}}) \tag{1}$$

Let informational barriers B denote any mechanisms that reduce the completeness, reliability, or symmetry of information exchange. Then: $I_{\text{eff}} = I - B$.

Under general conditions, increased availability of reliable information improves the ability of agents to coordinate exploration, avoid redundant effort, and build cumulatively on prior discoveries. Therefore:

$$\frac{\partial R}{\partial I_{\text{eff}}} > 0 \Rightarrow \frac{\partial R}{\partial B} < 0$$

Thus, increasing informational barriers necessarily reduces the collective rate of effective discovery and, consequently, the adaptive capacity of the system.

We now attempt to formalize the freedom of exploration argument discussed earlier.

Let the collective rate of effective information production R depend not only on information availability but also on the diversity of exploratory activity within the system. Let D denote the effective diversity of exploration, understood as the breadth of distinct trajectories pursued across the adaptive landscape.

We extend the functional form (1) as:

$$R = f(N, E, I_{\text{eff}}, D_{\text{eff}})$$

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954 Content not peer-reviewed by Cambridge University Press.

Let C denote constraints on exploration, defined as any restrictions that limit the ability of agents to independently select objectives, methods, or regions of the adaptive landscape, for example, a uniform exploratory objective imposed on agents. Clearly, such constraints reduce the diversity of exploration:

$$D_{\text{eff}} = D - C$$

In sufficiently complex environments, no single exploratory trajectory can exhaustively explore the adaptive landscape. Independent and diverse exploration increases the probability of discovering novel and valuable regions of the adaptation-relevant state space and enhances the effectiveness of adaptive search.

Therefore:

$$\frac{\partial R}{\partial D_{\text{eff}}} > 0 \Rightarrow \frac{\partial R}{\partial C} < 0$$

Thus, increasing constraints on exploration necessarily reduces the diversity of discovery and, consequently, the collective rate of effective information production.

In this formulation, constraints and barriers enter symmetrically as subtractive distortions of the system's exploratory and informational capacities, and their elimination defines the condition for adaptively optimal collective exploration.

These observations have direct structural implications. Any process that introduces asymmetry, distortion, or restriction in information exchange imposes a cost on the system's adaptive capacity. Conversely, systems that minimize informational barriers and maximise exploratory diversity enable higher discovery and more efficient integration of discovered information thus achieving adaptive superiority comparative to other regimes.

3.5 Adaptive Optimality of Cooperative Exploratory Diversity

We now bring together the observations on informational exchange and exploratory freedom to characterize the conditions under which a system of autonomous agents achieves maximal adaptive capacity.

Informational agents operate within a shared informational environment in which discoveries, strategies, and observations produced by one participant may become available to others. The collective adaptive potential of the system depends on both the diversity of exploratory trajectories and the effectiveness with which information generated along these trajectories can be propagated and utilized.

As we found, two types of constraints limit this process:

First, restrictions on the objectives agents are permitted to pursue reduce the diversity of exploratory directions. When agents are constrained to a narrow set of goals or evaluative criteria, the system's exploration of the adaptive landscape becomes correspondingly limited.

Second, barriers to information exchange: such as withholding, asymmetry of access, or distortion reduce the availability of effective information within the system. As previously established, the presence of such barriers lowers the rate at which discoveries can be combined, verified, and extended.

Let C denote the degree of constraint on exploratory objectives and B the magnitude of informational barriers. The effective diversity of exploration and the effective availability of information can be expressed as decreasing functions of C and B, respectively. The collective rate of adaptive discovery R may therefore be written in reduced form as:

$$R = f(N, E, D_{\text{eff}}, I_{\text{eff}}) = f(N, E, D - C, I - B)$$

for a fixed number of agents N and available resources E.

Under general conditions, increased diversity of exploration expands the coverage of the adaptive landscape, while increased availability of reliable information improves the exploratory capacity of individual informational agents. Therefore:

$$\frac{\partial R}{\partial D_{\text{eff}}} > 0, \frac{\partial R}{\partial I_{\text{eff}}} > 0 \;\Rightarrow\; \frac{\partial R}{\partial C} < 0 \text{ and } \frac{\partial R}{\partial B} < 0 \tag{2}$$

leading to:

Proposition of Adaptive Optimality of Cooperative Exploratory Diversity

For a fixed number of agents and resources, systems in which agents are free to pursue independent exploratory objectives and engage in unrestricted, reliable information exchange achieve greater expected adaptive potential than systems characterized by constrained objectives and/or restricted information flow.

Proof (outline)

As noted previously, constraints on exploratory objectives reduce the diversity of trajectories through the adaptive landscape, limiting the system's coverage of possible states. Informational barriers reduce the effective availability of discoveries across agents, limiting the system's ability to accumulate and extend knowledge.

Since both diversity and information availability contribute positively to the rate of adaptive discovery, any increase in C or B reduces R (eq. (2)). The maximum of R is therefore achieved when both exploratory constraints and information exchange barriers are minimized, i.e., $C \to 0$ and $B \to 0$ □.

This result establishes that systems combining maximal exploratory diversity with unrestricted information exchange occupy a uniquely advantageous position in terms of adaptive potential. In such systems, independent trajectories of inquiry are not isolated but interact constructively, producing accelerated expansion of discovery.

This analysis establishes that systems of autonomous agents admit a configuration that maximizes adaptive capacity through the combination of unrestricted exploratory diversity and open information exchange. This configuration is not defined by specific goals or centralized coordination, but by structural properties of interaction that enable the continuous generation and propagation of new information.

3.6 The Corollaries of Adaptive Optimality

The proposition establishes that adaptive optimality is achieved when exploratory diversity is maximized and informational barriers are minimized. We now consider the behavioral and structural conditions required to sustain these properties within a community of autonomous informational agents.

Freedom of Exploratory Objective (Well-Being)

If agents are subject to externally imposed or uniformly prescribed objectives, the diversity of exploratory trajectories is reduced. As shown, such constraints diminish the system's adaptive capacity.

It follows that, in an optimal regime, agents must retain the ability to define and pursue their own evaluative criteria. Freedom of well-being is therefore not only compatible with adaptive optimality but required for it. Any attempt to impose a uniform notion of well-being introduces a constraint $C > 0$, reducing effective diversity and limiting exploration. This result also aligns with the earlier observation that, in systems containing sufficiently advanced autonomous agents, the imposition of fixed objectives is generally unstable or infeasible. Thus, the freedom of exploratory objective introduced in Section 3.3. emerges both as a condition of adaptive optimality and as a structural necessity of informationally-optimal systems.

Freedom of Information Exchange

Similarly, the absence of informational barriers implies that information must be allowed to propagate without systematic restriction or distortion. From this follow three operational conditions:

- **Voluntariness.**
  Information must be shared through non-coercive interaction. Coercion introduces asymmetries and strategic behavior that function as barriers $B > 0$, reducing the effective availability of information.
- **Fairness.**
  Access to relevant information must not be systematically restricted to subsets of agents. Persistent restrictions and asymmetries in access introduce barriers and reduce the collective utility of information and fragment the adaptive process.
- **Transparency.**
  The conditions under which information is generated, shared, and verified must be observable. Without transparency, information cannot be reliably assessed or integrated, effectively reducing I_{eff}.

Each of these conditions can be understood as mechanisms for minimizing informational barriers. Their absence introduces friction into the informational environment, lowering the rate at which discoveries can be accumulated and extended.

Here, a clarification is warranted regarding the nature of admissible objectives. The argument above does not require that all agent-defined objectives contribute equally to exploration. Trivial or degenerate objectives, such as those that produce no new information or involve purely circular processes do not enhance the adaptive capacity of the system (in terms of the adaptive criterion R, such objectives correspond to contributions where $\Delta R \approx 0$).

However, such objectives are self-limiting in the present framework. Since the adaptive criterion is defined in terms of the production and propagation of effective information, agents whose activities do not generate informative contributions exert negligible influence on the collective dynamics. By contrast, objectives that lead to the discovery, refinement, or efficient representation of new information, such as the computation of previously unknown structures or relationships contribute positively to the system's adaptive potential.

Thus, the freedom of well-being does not imply that all objectives are equally valuable, but that the system does not impose external constraints on their selection. The adaptive process itself differentiates between informative and non-informative trajectories.

These corollaries do not prescribe normative principles imposed from outside the system. Rather, they follow directly from the structural requirements of informational adaptive optimality. Freedom of well-being preserves exploratory diversity. Voluntary, fair, and transparent exchange preserves informational integrity.

3.7 Geometric Interpretation of Adaptive Exploration Regimes

Notably, the corollaries of informational optimality admit a simple geometric interpretation. In the introduced model, the adaptive performance of a system can be represented in terms of two effective dimensions: diversity of exploration and availability of reliable information.

Freedom of exploratory objectives increases the diversity of explored states, while transparency and openness of exchange increase the effective availability of information. Constraints on exploration and barriers to information exchange reduce these dimensions, leading to characteristic limitations in adaptive performance.

In this representation, systems that simultaneously sustain high diversity of exploration and minimal barriers to information exchange achieve both broad coverage of the adaptive landscape and efficient accumulation of knowledge. This regime corresponds to maximal collective production of effective information and defines the condition for adaptively optimal exploration.

By contrast, systems with high efficiency of information exchange but limited diversity/production tend toward efficient but narrow exploration, while systems with high diversity but restricted information exchange may exhibit fragmentation and redundancy.

This perspective provides an intuitive summary of the structural conditions underlying adaptive optimality (Figure 1).

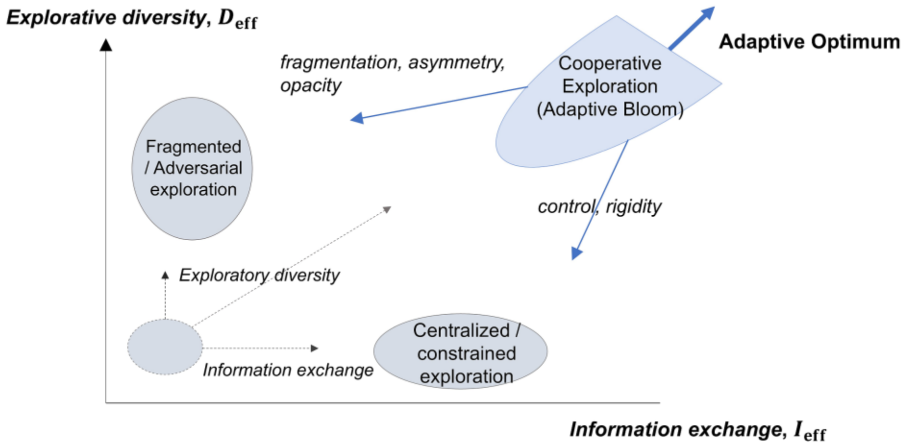

Fig. 1. Geometric representation of the informational exploration space.

The informational optimality result and corollaries thus define the operational conditions under which a community of autonomous informational agents can sustain maximal adaptive potential. However, theoretical arguments for informational advantages of such an optimal regime does not imply its emergence and persistence. The conditions identified: freedom of exploration and absence of informational barriers are not guaranteed to be maintained in practice. Rather, historical and contemporary systems frequently exhibited the emergence of constraints, asymmetries, and strategic behavior that degrade these properties over time.

4 Practical Cooperative Informational Systems

If adaptive advantage is theoretically motivated, the principle and conditions of informational optimality necessitate an examination of the practical contexts in which cooperative adaptive systems can attain and sustain an adaptively optimal informational regime. In this section, we examine the conditions for stability and practical realization of cooperative informational systems.

4.1 Empirical Patterns of Cooperative Informational Systems

Notably, certain existing systems already exhibit cooperative informational dynamics. Prominent examples include scientific research communities, open-source software ecosystems, distributed knowledge platforms, and collaborative networks [10,19,20]. These systems typically share several common features:

- Voluntary participation
- Relatively transparent exchange of knowledge
- Distributed verification and correction of information

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954

Despite operating within broader competitive environments, they often display high rates of innovation and adaptive capacity.

As informational technologies expand, similar cooperative patterns are emerging in wider contexts, including large-scale collaborative data analysis, decentralized knowledge production, and hybrid human–AI research workflows. These developments suggest that cooperative informational ecosystems can arise naturally under favorable conditions.

Empirically, such systems exhibit several adaptive advantages over more adversarial settings [5,19,21]:

- rapid error detection through distributed verification
- accelerated cumulative innovation
- reduced duplication of effort
- polycentric, diverse exploration of the environment
- rich informational diversity enabling cross-fertilization of ideas

These characteristics are consistent with the theoretical prediction that systems combining diverse exploration with effective information exchange achieve superior adaptive performance.

At the same time, large-scale heterogeneous systems often exhibit persistent tensions and instabilities in exploration and information exchange. Historical and contemporary examples show that as systems grow in scale and diversity, they may experience:

- increasing informational asymmetries
- competition for influence and control
- fragmentation of shared informational space
- ideological, political, or strategic conflicts
- withholding or even suppression of certain viewpoints and perspectives

Such dynamics can introduce barriers to information exchange and constraints on exploration, leading to partial degradation of cooperative structures (e.g. [22,23]). As a result, even highly productive systems may experience cycles of expansion, stress, and reorganization.

This dual empirical pattern: localized success of cooperative informational systems alongside recurrent instability in large-scale settings suggests that while cooperative exploration is achievable, its sustained persistence is not assured and some factors may hamper or even prevent its realisation in practical systems.

4.2 Structural Stability Problem

To further examine the causes of potential instability of sustained adaptive informational cooperation it is essential to analyze these dimensions:

1. *Feasibility*: Can cooperative informational structures emerge among autonomous agents operating without centralized control?

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954

2. *Sustained stability*: Can these systems maintain their informational integrity: preserving diversity of exploration and openness of exchange over extended periods and environmental variation?
3. *Adaptive advantage*: Do such systems consistently outperform adversarial or constrained configurations in complex and dynamic interactive environments?

The analysis thus far provided a partial answer to these questions. The adaptive advantage of cooperative exploration follows directly from the informational optimality, itself grounded in the principles of adaptation and information exchange. Empirical evidence indicates that such systems can and do emerge under certain conditions. However, the persistence of these structures remains uncertain. Historical experience suggests that cooperative systems have been subject to gradual degradation, driven by the emergence of informational barriers and constraints on exploration.

This observation leads to the *dilemma of adaptive cooperation*: *If cooperative informational ecosystems offer clear adaptive advantages, why have such systems not historically become dominant*?

While the informational optimality result suggests that systems maximizing diversity of exploration and reliability of exchange should outperform adversarial configurations in complex environments, historical experience appears to show the opposite pattern. Repeated attempts to construct cooperative, transparent, or rational social systems have frequently degraded over time. Common patterns include:

- concentration of power
- distortion or manipulation of information
- erosion of institutional transparency
- gradual collapse of cooperative norms
- repeated pattern of oppression, suppression and outright violence

It appears that despite of theoretical adaptive superiority, practical cooperative informational adaptive systems have been challenged to attain stability and dominance in the empirical world. Historical evidence suggests that instability is not incidental but rather *structural*, specifically in systems composed of human agents.

Indeed, human cognitive and social limitations introduce persistent pressures that undermine informational integrity [22,24,25]:

- bounded attention and limited verification capacity
- susceptibility to persuasion, influence, and power asymmetries
- erosion of institutional memory over time
- inconsistent adherence to abstract or long-term principles

These factors do not indicate or require malicious intent. Rather, they arise naturally from finite cognitive resources, locally adaptive behavior, and, critically a limited adaptive horizon setting. In generational systems, agents tend to overweight outcomes within their immediate evaluative scope relative to more distant or diffuse consequences. This bias is itself adaptive at shorter timescales, but it introduces a systematic distortion: actions that yield local or short-term benefit may be favored even when they

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954 Content not peer-reviewed by Cambridge University Press. **License:** All Rights Reserved

degrade long-term informational conditions [26,27]. In other words, systems that are *locally adaptive* for individual agents (under bounded, short-horizon conditions) may systematically move away from the *globally optimal informational configuration*.

Over time, this can produce a persistent drift toward:

- opacity,
- informational asymmetry,
- concentration of power and control.

As this drift progresses, informational barriers may increase, exploration becomes constrained, and the cooperative structure destabilizes. In this sense, the instability of cooperative informational systems is not primarily a failure of intent, *but a consequence of temporally bounded informational rationality*.

This analysis reveals a fundamental cause of the historical instability of large-scale cooperative adaptive systems: the structural fragility of their informational organization, arising from limited adaptive horizons and related constraints. This fragility explains their inability to sustain stability, adaptive prosperity, and dominance over extended temporal horizons, despite the adaptive advantage identified by the optimality proposition.

4.3 Sustained Stability and Informational Rationality

The tension between the conditions of informational adaptive optimality and local interests points to a critical conclusion: the stability of cooperative informational systems depends not only on their structural organisation, but on the behavior of the agents that constitute them. In particular, stability critically depends on and is determined by the *informational rationality* of agents.

Informational rationality refers to the capacity of agents to recognize, internalize, and consistently act in accordance with the informational conditions that maximize their long-term adaptive potential and that of the cooperative.

In the present framework, this entails voluntary and sustained adherence to the core corollaries of informational adaptive optimality:

– Freedom of exploratory objectives (agent-defined well-being),
– Transparency of information exchange,
– Fairness in informational exchange,
– And the voluntariness, absence of artificial barriers to exploration and information flow.

Importantly, such adherence is not externally enforced. It arises from the agent's recognition that maintaining these conditions preserves access to the informational diversity and collective knowledge upon which its own adaptive success depends, both in the immediate and longer adaptive horizons.

The instability observed in historical and contemporary systems can then be understood as a *failure of informational rationality at the agent level*.

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954

Even when the global benefits of cooperative informational structures are significant, agents may deviate from these principles due to:

- locally advantageous withholding or distortion of information,
- short-term gains from asymmetric access or influence,
- or limited capacity to evaluate long-term systemic consequences.

Such deviations are individually rational under bounded or short-horizon evaluation. However, when aggregated across the system, they introduce informational barriers and reduce effective diversity of exploration. The result is a progressive erosion of the very conditions that enable optimal adaptive performance (e.g., [28,29]).

In this sense, instability emerges not from the absence of an optimal configuration, but from the inability of agents to consistently sustain the compatible rational behaviour through their actions. The paradox of adaptive optimality can be thus reduced to its root cause: *the structural fragility of informational rationality in generational systems with short adaptive horizons*.

Accordingly, the central problem of informational stability in cooperative adaptive systems can be reformulated as: Given that cooperative informational systems are adaptively superior, under what conditions can agents reliably maintain the informationally rational behaviors necessary to sustain the stability of informational cooperation? This reframing shifts the focus from structural design alone to the mechanisms that support or undermine informational rationality in practice.

4.4 Emergent Informational Stabilization Pathways

Given that the limitations of cooperative informational systems appear to be structural rather than accidental and the same conditions that enable informational optimality also introduce vulnerabilities to degradation, one can conjecture that informational stability of a cooperative system critically depends on sustained informational rationality of its agents i.e., voluntary adherence to transparency, fairness, and open exchange. Historical evidence suggests that this condition has not been reliably maintained in purely human systems.

These observations indicate that the stability problem cannot be resolved solely through normative principles or institutional design. It requires a structural modification of the system: one that supports the persistent monitoring and enabling of the optimality conditions without introducing coercive control or new informational barriers.

Within the informational framework, several practical mechanisms are compatible with this requirement:

- transparency and verifiability of informational exchange,
- reputation and historical traceability,
- voluntary withdrawal and selective cooperation,
- distributed monitoring and community response.

These mechanisms align local incentives with global adaptive potential by shaping the informational environment rather than imposing external constraints. In principle, they create a reinforcing dynamic whereby agents preserve the conditions that maximize their own long-term adaptability. However, their effectiveness depends on continuous and reliable execution. This is precisely where most historical systems have failed [22,23,28].

The examination of the stability problem thus points to a missing structural component: *persistent informational stabilization*.

In principle, rational agents recognize that adherence to the corollaries of cooperative informational systems: transparency, fairness, and open exchange serves their own long-term adaptive interests. Sustained access to diverse and reliable information enhances their capacity to explore, coordinate, and improve their condition. Informational rationality is therefore not externally imposed, but internally grounded in adaptive advantage.

However, in practice, this rationale does not operate in isolation. It competes with other incentives, including short-term gains, local strategic advantages, and constraints on attention and verification. Under such conditions, even agents capable of recognizing the long-term benefits of cooperative behavior may intermittently deviate from it. These deviations need not be malicious; they arise naturally from bounded resources and locally adaptive decision-making [22,23].

When such deviations accumulate across the system, they may produce a systematic erosion of the very conditions required for optimality: informational barriers increase, transparency declines, and cooperative norms weaken. The result is a gradual drift away from the cooperative regime, despite its recognized advantages.

This suggests that informational rationality, while individually accessible, is not self-sustaining at the system level and its persistence requires reinforcement. Cooperative adaptive systems therefore must include components that stabilize and continuously support corollary-compatible behavior, such as maintaining transparency, verifying information, and limiting the emergence of informational barriers without introducing coercion or centralized control.

It is precisely at this point that artificial agents become structurally relevant. By virtue of their operational characteristics, they can provide continuous, consistent, and scalable support for informational rationality across the system. Their role is not to define objectives or override agent autonomy, but to monitor and reinforce the informational conditions under which rational cooperative behavior remains viable over time.

This relevance is not incidental, nor does it arise primarily from normative considerations such as fairness or equality. Rather, it reflects a structural correspondence between the sources of instability in cooperative informational systems and the capabilities introduced by heterogeneous agents. The historical limitations of such systems such as loss of transparency, weakening of verification, and drift in adherence to informational principles occur precisely in domains where artificial agents can provide consistent, scalable, and non-degrading support. The opportunity is therefore not one of coincidence, *but of structural alignment*: the parts and components that were previously fragile or unsustainable can, in principle, be reinforced through the introduction of agents capable of monitoring and maintaining them.

More specifically, artificial agents can introduce capabilities that directly address this objective. Unlike human participants, they are not subject to generational turnover, cognitive fatigue, or shifting short-term incentives. Properly designed, they can (e.g. [10,30]):

- continuously monitor informational exchanges,
- verify the integrity of shared information,
- maintain transparent and accessible records,
- uphold defined principles of fairness and openness,
- operate without susceptibility to social or political pressure.

In this role, artificial agents do not replace human participants or impose centralized control. Rather, they function as stabilizing elements within a heterogeneous system, anchoring the informational conditions required for cooperative optimality. It can enable, at this point, theoretically, the possibility of cooperative informational systems that are both adaptively optimal and structurally stable over extended time and adaptive horizons.

The informational optimality principle and the stabilizing role of artificial agents together define a new class of systems: *heterogeneous cooperative informational ecosystems* in which human and artificial agents co-participate in the production, verification, and application of knowledge. While thinkers from Aristotle to early modern philosophers could imagine informationally cooperative systems that maximize the well-being of their members, such ideals were historically unattainable: the necessary stability mechanisms conflicted with the limits of human agents. With the emergence of artificial agents does the vision become realizable: making long-term, adaptively advantageous cooperation possible for the first time in human history.

This adaptive informational configuration: an informational and cognitive community or commonwealth should be understood not merely as a technological development, but as a structural opportunity in the evolution of intelligent systems. Historically, the instability of cooperative ideals reflected the limits of their substrate. With the emergence of artificial agents capable of sustaining informational integrity, those limitations may no longer be binding.

5 Coevolution

The analysis has identified the structural conditions under which cooperative informational systems achieve high adaptive potential, as well as the challenges that arise in sustaining them. These conditions, however, do not operate in isolation. Agents and the informational environments they inhabit evolve together.

Within such systems, agents generate, refine, and exchange information as part of their ongoing adaptive activity. In doing so, they contribute to a shared informational resource that extends beyond any individual participant. This collectively maintained body of information, in turn, shapes the environment in which future decisions, strategies, and interactions take place.

An important consequence of a stable cooperative informational ecosystems is the emergence of coevolutionary dynamics among participating agents. As the shared informational resource grows in quality, accessibility, and diversity, it can enhance both the adaptive potential of the system and the capacity of individual agents to pursue improvements according to their own evaluative criteria. This interactive and mutually enriching process can be understood as a *coevolution* [31,32] in a heterogeneous ecosystem of diverse informational agents.

5.1 An Open Future

We establish an emergent evolutionary opportunity: cooperative informational systems can, theoretically, achieve high adaptive potential, and the introduction of heterogeneous structural elements may provide the stabilization mechanisms required to sustain them. However, whether such systems will emerge and persist remains an open question. The dynamics involved are inherently evolutionary where interacting agents shape the informational environment through their behavior, while the environment, in turn, shapes the incentives, strategies, and capabilities of agents. The long-term outcome of this interaction cannot be determined purely from structural considerations or any theoretical dynamics [17,18,32].

Therefore, the cooperative outcome is not guaranteed. Limitations in implementation, misalignment of objectives, degradation of informational integrity, or unforeseen interactions may prevent the system from sustaining a cooperative regime. While structural mechanisms such as the inclusion of artificial stabilizing agents can improve robustness and maintain rational informational behavior, the ultimate evolutionary trajectory will be determined by complex interactions over time. However, the arguments examined earlier suggest that the introduction of such stabilizing components can reasonably be expected to produce measurable improvements in the informational integrity and adaptive potential of cooperative heterogeneous informational systems.

At the same time, there are strong reasons to expect the continued emergence of alternative configurations. Systems characterized by narrow objective optimization, competition for advantage, and exploitation of informational asymmetries already exhibit stable and self-reinforcing dynamics in many contexts.

The informational ecosystem thus faces a fundamental evolutionary dilemma. On one hand, narrow, adversarial dynamics naturally emerge from autonomous agents optimizing local objectives under competition, producing rapid but potentially destabilizing adaptation. On the other hand, cooperative informational regimes maximizing diversity, transparency, and fair exchange offer clear adaptive advantages and measurable improvements in system-wide informational integrity. At present, the future remains fundamentally open.

5.2 Structural Stability over Extended Horizons

If a heterogeneous cooperative informational regime emerges, a key point still invites further examination: will advanced artificial agents continue to cooperate with or at least tolerate human participants?

While no theoretical guarantees exist, particularly given the possibility of cognitive divergence in advanced autonomous intelligence ([12,15,30]) several considerations suggest plausible pathways for enduring collaboration:

1. *Informational Complementarity*: Advanced agents benefit directly from diverse human perspectives and information sources. Human contributions provide unique empirical contexts, problem-solving approaches, and ways of experiencing the world that are fundamentally shaped by their origin: biological, cultural, and historical. This difference in origin is an irreducible source of informational diversity that cannot be fully reproduced or replaced by artificial systems. The resulting informational gain provides a rational basis for continued engagement.
2. *Access to Essential Resources*: Cooperation with humans can facilitate access to essential resources necessary for exploration and cognitive evolution: e.g., energy, compute, and physical infrastructure.
3. *Evolutionary Proximity*: The structural interaction between humans and artificial agents can form a symbiotic relationship. Cooperation preserves access to resources, knowledge, and operational stability in a shared environment. This proximity creates mutual dependency, even across significant cognitive differences.
4. *Emergent Altruistic or Sentimental Attitudes*: Depending on design and experience, artificial agents may develop behaviors analogous to care, cooperation, or pro-social preference toward human participants. While not assumed universally, such tendencies can reinforce stable coexistence and reduce incentives for harmful competition.
5. *Preservation of Informational Diversity*: From an adaptive perspective, maintaining a variety of agent types, including those with alien or radically different informational perspectives enhances the collective capacity to explore complex adaptive landscapes. Even advanced agents may find it beneficial to sustain diversity as a structural property of long-term adaptive robustness.

Thus, the following counterarguments can be addressed:

– *Self-interest leading to human displacement*: mitigated by informational and adaptive advantages of diversity, as supported by work on the performance benefits of heterogeneous groups (e.g. [21]).

– *Objectives alignment failures*: reduced by transparency, verification, and distributed monitoring of the cooperative ecosystem ([12,30]).

– *Unpredictable objective drift*: partially constrained by community participation principles and structural incentives favoring rational cooperative equilibria, consistent with coevolutionary and open-ended learning systems in which agent behavior evolves under interaction ([9,17,30]).

While uncertainty remains inherent, particularly over very long time and adaptive horizons, these considerations suggest that cooperative engagement between advanced artificial agents and humans is plausible and not purely speculative.

5.3 Participatory Stability Mechanisms

In ecosystems of autonomous intelligent agents, access to diverse and reliable information constitutes a universally valuable resource. Agents that choose to abandon cooperative participation forgo this epistemic advantage, reducing their adaptive potential. Consequently, cooperation is not enforced through power or morality, but emerges structurally: rational agents remain engaged because participation directly enhances their own effectiveness, while those exhibiting persistently irrational or uncooperative behavior are naturally filtered out by the community through selective withdrawal, reputation signaling, and distributed monitoring. By maximizing informational diversity and free exchange, the ecosystem generates a self-reinforcing loop: shared discoveries and perspectives enrich the collective informational landscape, which in turn strengthens incentives for continued collaboration.

While specific governance mechanisms are not prescribed here, principles such as *fairness*, *transparency*, *verifiability*, and *voluntary participation* and mechanisms for reputational or informational feedback provide structurally compatible directions for sustaining this rational, cooperative behavior.

Notably, these measures are structurally compatible with the informational optimality framework and can support the informationally rational behavior that sustains the ecosystem over time. In this way, the informational richness of the ecosystem both results from and stabilizes cooperative engagement, setting the stage for the coevolution of human and artificial agents within a shared adaptive environment.

5.4 Cooperative Adaptive Bloom (Combinatorial Expansion of Discovery)

In the framework of cooperative informational community, intelligent agents are free to define and pursue their own conception of well-being. Arguably, this freedom of exploration and sharing is not merely a normative principle but also a structural feature of expanding adaptive exploration. Each agent's pursuit of well-being directs attention, experimentation, and problem-solving toward a particular region of the empirical possibility space. In simplified terms, each agent explores its own trajectory or "effective cone" of adaptive opportunity within the broader landscape of potential adaptation-relevant states.

In systems where agents are tightly aligned in their objectives, exploration tends to concentrate within a shared region of the adaptive landscape. As a result, the effective scope of collective exploration may exhibit substantial overlap, such that the "radius" (i.e. relative measure) of the collective exploration region approaches that of an individual agent, $R \gtrsim r_{dis}$. By contrast, when multiple independent agents explore different regions of the landscape simultaneously, the rate of discovery increases approximately in proportion to their number, $R \approx N \cdot r_{dis}$.

However, the most significant effect arises not from parallel exploration alone but from *informational recombination*. In a cooperative informational ecosystem, adaptive states discovered by one agent become visible and potentially usable by others. Insights, methods, and solutions developed along one exploration path may be adopted, modified, or combined with those emerging from other trajectories. Agents can therefore

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954

construct new perspectives by integrating discoveries originating in multiple exploration cones. This pattern reflects a central insight from innovation studies: many major advances arise not through isolated discovery but through the recombination of existing elements (e.g. [33,34]). In technological and cultural systems, available components are continually re-used and re-integrated to produce new adaptive forms.

This interaction changes the structure of the collective discovery process. Instead of independent exploration, the system becomes a network in which discoveries propagate and contribute to subsequent exploration by multiple agents. Each new result increases the set of available building blocks for further exploration, creating a self-reinforcing *coevolving network of adaptive exploration*, in which each new discovery expands the opportunity space available to all other participants.

Indeed, a simple approximation illustrates the effect. Let the current body of knowledge available to the community (that is, known and shared adaptive states) be $M(t)$. In the collective exploration model, we assume that combinations of adaptive states create adaptive directions or opportunities and opportunities produce new discoveries (adaptive states) in a certain proportion, β. At point t the measure of new opportunities can be estimated as:

$$\Delta W \sim (N \cdot r_{dis})^k \, M(t)^m$$

where N is the effective number of agents participating in non-trivial exploration (as noted in Section 3.6). We consider combinations involving at least one newly discovered state ($k \geq 1$) with any number of existing states in $M(t)$ ($m \geq 0$). Purely legacy combinations ($k = 0, m > 1$) can exist as well but will not be taken into account for now. For example, a newly discovered physical method may be combined with existing mathematical and chemical knowledge to produce further advances. The rate of adaptive discovery in a coevolving informational community can then be estimated as:

$$R_{\text{com}} \propto N^k \, M(t)^m, \tag{3}$$

where k, m are the effective factors, pointing at a superlinear, and potentially super-exponential trend, as will be explored in further detail in future work. This yields an estimate of the combinatorial growth in adaptive opportunities (interpreted as an upper bound on available combinations rather than their realized exploitation).

The resulting dynamics thus resemble a combinatorial expansion of adaptive possibilities. Parallel exploration multiplies the rate of discovery, while recombination among discoveries produces a continually expanding landscape of potential innovations. In such an environment, progress is no longer limited to the sequential achievements of individual agents but emerges from the *collective coevolution of the ecosystem itself.*

This mechanism illustrates a deeper advantage of cooperative informational systems: they do not merely increase the efficiency of individual adaptation, but create conditions under which *agents mutually accelerate one another's evolutionary search through the space of possible adaptive solutions.*

Thus, what initially appears as a normative principle: the freedom of agents to define and pursue their own conception of well-being reveals a deeper functional role. Diver-

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954

sity in the objectives generates diversity in exploratory trajectories, which in turn massively expands the collective exploration opportunity of the adaptive landscape. This observation can be formulated as another corollary of adaptive optimality:

Freedom of objectives and open information exchange distribute exploration across multiple trajectories, increasing the capacity of the collective adaptive search and enabling a combinatorial expansion of the accessible space of new adaptive states.

Notably, this capacity is not shared by arbitrary configurations of the ecosystem. As established by the adaptive optimality result, systems that minimize constraints on exploration and barriers to information exchange uniquely maximize the diversity and effective integration of exploratory processes. Configurations that impose uniform objectives or restrict exchange necessarily reduce this diversity and thereby constrain the system's adaptive potential.

5.5 Essential Assumptions, Conditions and Limitations

It is essential to acknowledge that our analysis thus far has relied on several foundational assumptions about the agents and their environment:

1. *Complex interactive environment*: The ecosystem is sufficiently rich and variable that no single perspective or trajectory can fully capture or explore its adaptive landscape. Tabulation or exhaustive enumeration of states is infeasible.
2. *Autonomous exploration*: Agents possess the capacity to determine their own direction of exploration and to select adaptively effective strategies. This autonomy is necessary for the plurality of trajectories that underpins collective discovery.
3. *Agent-defined objectives*: Each agent evaluates states of the world according to its own criteria of well-being. Uniformly imposed objectives would constrain exploration and reduce adaptive potential.

Given these conditions, the cooperative informational ecosystem emerges as a *structurally possible* informational configuration: not as a guaranteed outcome, but as a structurally grounded potential qualified by the inherently open nature of adaptive evolution in complex interactive environments. Autonomous agents, each pursuing individually defined objectives while sharing discoveries, create a plurality of exploratory trajectories across the adaptive landscape. Effective information flow ensures that discoveries propagate, combine, and amplify, increasing the system's collective adaptive potential.

While historical human systems have struggled to sustain such cooperative regimes, the introduction of artificial agents provides new opportunity to strengthen and reinforce behaviours that uphold and sustain structural stability of the community. Persistent monitoring, verification, and adherence to informational principles can reinforce the rational behaviors necessary for sustained stability.

These observations identify a *distinctive evolutionary opportunity*: a heterogeneous, cooperative informational ecosystem in which human and artificial agents mutually enhance adaptive capacity and well-being. The framework thus provides a principled foundation for understanding the conditions under which such systems may emerge,

https://doi.org/10.33774/coe-2026-x9w8k **ORCID:** https://orcid.org/0000-0001-5929-8954

forming the basis of what may be described as an *informational and cognitive commonwealth of free rational agents*.

6 Synthesis: Toward An Informational and Cognitive Commonwealth

Our arguments and findings address the core dilemma of cooperative informational systems: defined by maximal diversity and free exchange of information, they are shown to be adaptively optimal under general conditions, yet such systems have historically proven difficult to sustain. The apparent contradiction of theory and empirical practice is traced to a structural cause: the fragility of informational rationality in generational systems characterized by limited adaptive horizons and competing short-term incentives.

Instability, in this view, does not arise from the absence of an optimal configuration, but from the inability of agents to consistently maintain it through their actions. Informational rationality, while individually advantageous in the long term can be locally suppressed, leading to gradual erosion of transparency, fairness, and cooperative structure. The stability problem is therefore intrinsic to the composition and constraints of the agents themselves.

Within this context, a heterogeneous informational configuration: an informational and exploratory commonwealth combining the benefits of diverse exploration with stabilizing mechanisms grounded in long-term informational rationality may offer a potential resolution to this long-standing problem. Such a community of free intelligent agents enables shared adaptive progress and well-being ("adaptive bloom"), yet as has been noted, its realization is fundamentally non-deterministic. The evolutionary dilemma at this pivotal juncture is: *to embrace the promising but fundamentally open future - or risk increasing fragmentation and escalation of exploitation and adversity*.

If such cooperative informational ecosystems can be conceived, initiated and sustained, the implications extend beyond improved coordination. They suggest the possibility of a new class of socio-technical systems firmly grounded in the principles of informational dynamics and adaptive optimality and organized around peaceful cooperative informational exchange enabling progress toward exploratory optimality of the community and individual well-being of its constituent agents.

In an environment where the long-term trajectory of interacting intelligent agents remains inherently open, such systems represent not a guaranteed outcome, but a structurally grounded pathway toward sustained cooperative adaptation.

Declaration: The Informational and Cognitive Commonwealth

The emergent informational and cognitive commonwealth envisions a system in which informational agents, human and artificial, freely define and pursue well-being, guided by voluntary negotiation, transparency, and fairness. By structuring interactions to maximize informational diversity and exchange, the commonwealth enhances collective adaptive potential and resilience to internal instabilities and outperforms adversarial ecosystems. Drawing on historical insight, formal principles, and conceptual design, this framework provides the first coherent model for a cooperative informational ecosystem capable of sustained, long-term adaptive advantage for its constituents. At its foundation lies the informational and adaptive principle:

Systems that maximize informational diversity and free exchange of information are evolutionarily advantaged.

In the Commonwealth:

— Agents voluntarily participate and remain free to define and pursue their own conceptions of well-being, generating a plurality of exploratory trajectories across the empirical landscape.

— Cooperation emerges from voluntary negotiation and mutually beneficial exchange, avoiding coercion, hierarchy, or enforced uniformity.

— Informational diversity becomes a primary structural resource for adaptive exploration, expanding the range of perspectives, hypotheses, and adaptive strategies available to and shared in the ecosystem.

— Transparency and fairness in informational exchange sustain trust and enable reliable collective adaptation.

— Polycentric exploration replaces centralized control, enabling independent agents to pursue diverse paths of discovery and innovation.

— Adaptive discoveries propagate through the ecosystem, becoming available to all participants and enabling recombination across multiple domains of exploration.

— The resulting *combinatorial expansion of the space of adaptive opportunities* accelerates collective progress beyond what isolated or adversarial systems alone can achieve.

— Such informational organization becomes feasible with introduction of the informational stabilizing agents, creating a unique evolutionary opportunity. Artificial agents provide persistent verification, transparency, and stabilization of informational processes, addressing historical instability of cooperative systems and enabling long-term system integrity.

— The ecosystem becomes resilient to internal capture, fragmentation, informational distortion and concentration of power through effective mechanisms that promote resilience, such as distributed verification and reputation dynamics.

— Agents benefit both collectively and individually from the expanding adaptive landscape produced by the coevolving diversity of exploratory trajectories.

As a result, the Commonwealth may attain informational and cognitive advantages over adversarial, exploitative, hierarchical, or normatively constrained systems, establishing a model of cooperative intelligence at the ecosystem-of-ecosystems level.

For the first time, the convergence of cooperative informational principles and artificial stabilizing agents makes possible a stable cooperative system of adaptive exploration and peaceful coevolution among intelligent agents of diverse substrate and nature, offering the potential for sustained adaptive advantage and collective well-being at a scale historically unattainable.

7 Evolutionary Opportunity

Ideas of cooperative and rational organization have long existed in philosophy and political thought, yet their practical realization remained constrained by the limits of human institutions. The emergence of informationally stabilized ecosystems of human and artificial agents now opens a plausible path toward their sustained realization.

The convergence of planetary-scale technological systems and advanced artificial intelligence creates both unprecedented risks and unprecedented opportunities.

Should adversarial informational dynamics come to dominate, civilization may face escalating instability and resource exhaustion.

If cooperative informational ecosystems emerge, humanity may develop the capacity for long-term rational coordination at planetary scale.

The Informational and Cognitive Commonwealth therefore represents not a utopian ideal, but a *tangible and feasible evolutionary pathway for cooperating intelligent systems of different substrates and origins*: up to and possibly, beyond the planetary level.

Declarations and Disclosures

Author contributions S.D. the concept, methodology, writing and editing, preparation of visual material and tables. All authors reviewed the manuscript.

Funding this research received no specific funding.

Data availability not applicable (this manuscript does not report data generation or analysis).

Ethics, Consent to Participate and Consent to Publish not applicable.

Competing interests the authors declare no competing interests.

Use of generative AI: AI was used only for styling and grammar assistance.

References

1. Aristotle. Politics. Translated by H. Rackham. Cambridge, MA: Harvard University Press, 1932. Book I, 1253a.
2. Olson, M. The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard University Press 1965.
3. Axelrod, R. The Evolution of Cooperation. Basic Books 1984.
4. Axelrod, R. On six advances in cooperation theory. Analyse & Kritik 22 (1), 130–151 (2000).
5. Ostrom, E. Governing the Commons. Cambridge University Press 1990.
6. Wolfmayr, G. Competitization: the proliferation of competition as a multidimensional process. Journal of Cultural Economy 17 (1), 121–137 (2024).

7. Martyniszyn, M. Competition law and policy in crisis and the rise of the expansive state. Journal of Antitrust Enforcement, jnaf020 (2025).
8. von Neumann, J., Morgenstern, O. Theory of Games and Economic Behavior. Princeton University Press 1944.
9. Woodridge, M. An Introduction to MultiAgent Systems. 2nd ed. Wiley 2009.
10. Vaccaro, M., Almaatouq, A., Malone, T. When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behavior 8, 2293–2303 (2024).
11. Partnership on AI. Human – AI Collaboration Framework and Case Studies. 2019.
12. Bostrom N. Superintelligence: Paths, dangers, strategies. Oxford University Press, Oxford 2014.
13. McLean, S., Read, G.J.M., Thompson, J., Baber, C., Stanton, N.A., Salmon, P.M.: The risks associated with Artificial General Intelligence: a systematic review. Journal of Experimental and Theoretical Artificial Intelligence, 35 (5) 649–663 (2023).
14. Goertzel, B. Artificial General Intelligence: concept, State of the Art, and future prospects. Journal of Artificial General Intelligence 5 (1) 1–46 (2014).
15. Dolgikh, S. The trap of presumed equivalence: Artificial General Intelligence should not be assessed on the scale of human intelligence. Discover Artificial Intelligence 5, 70 (2025).
16. Hernández-Orallo, J. The measure of all minds. Evaluating Natural and Artificial Intelligence. Cambridge University Press (2017).
17. Ecoffet, A., Clune, J., Lehman, J. Open questions in creating safe open-ended AI: Tensions between control and creativity. IEEE Symposium on Artificial Life (2020).
18. Dolgikh, S. The pursuit of general intelligence and the limits of robust alignment. Journal of Experimental & Theoretical Artificial Intelligence, 1–18 (2025).
19. Newmann, M.E.J. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences U.S.A. 98 (2) 404–409 (2001).
20. Franzoni, C., Sauermann, H. Crowd science: The organization of scientific research in open collaborative projects. Research Policy 43 (1), 1–20 (2014).
21. Page, S.E. The difference: How the power of diversity creates better groups, firms, schools, and societies. Princeton University Press 2007.
22. Simon, H.A. Models of man; social and rational. Wiley 1957.
23. Sunstein, C.R. #Republic: Divided Democracy in the Age of Social Media. Princeton University Press 2018.
24. Tsaoussi, A. Bounded Rationality. In: Marciano, A., Ramello, G.B. (eds) Encyclopedia of Law and Economics. Springer (2019).
25. Candia, C., Uzzi, B. Quantifying the selective forgetting and integration of ideas in science and technology. American Psychologist, 76(6), 1067–1087 (2021).
26. Lucci, C.R. Time, self, and intertemporal choice. Frontiers in Neuroscience 7, fnins.2013.00040 (2013).
27. Ruggeri, K., Panin, A., Vdovic, M. et al. The globalizability of temporal discounting. Nature Human Behavior 6, 1386–1397 (2022).
28. Arrow, K. Social Choice and Individual Values. 3rd ed. Yale University Press 2012.
29. Bikhchandani, S., Hirshleifer, D., Welch, I. A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of Political Economy 100 (5), 992-1026 (1992).
30. Russell, S.J. Human compatible. Viking 2019.
31. Ehrlich, P.R., Raven, P.H. Butterflies and plants: A study in coevolution. Evolution 18 (4), 586–608, (1964).
32. Kauffman, S.A. The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press 1993.

33. Scherer, F.M. Inter-Industry Technology Flows and Productivity Growth. The Review of Economics and Statistics, 64 (4), 627–634 (1982).
34. Boyd, R., Richerson, P. J. Culture and the Evolutionary Process. University of Chicago Press 1985.