Bio-CPC: Collective Predictive Computation in Biological Organization

Daichi G. Suzuki[1]*, Shigeto Dobata[2], Koryu Kin[3], Reina Kobayashi[4], Yusuke Hayashi[5]

[1]Institute of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan
[2]Department of General Systems Studies, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
[3]Department of Biochemistry and Systems Biomedicine, Juntendo University, Tokyo, Japan
[4]Department of Biomedical Engineering, Duke University, Durham, NC, USA
[5]AI Alignment Network, Tokyo, Japan
*Corresponding author: suzuki.daichi.gp@u.tsukuba.ac.jp

Abstract

Biological organization is characterized by hierarchically nested levels in which relatively autonomous units become integrated into higher-order systems. Major evolutionary transitions such as multicellularity and eusociality therefore raise a common problem: how can locally informed, goal-directed units coordinate heterogeneous responses so that they collectively support a larger whole? Here, building on collective predictive coding (CPC), originally developed to explain the emergence of shared symbolic representations through decentralized Bayesian inference, we propose Bio-CPC, a broader framework of collective predictive computation for biological organization. In this framework, information originating in multiple constituent units can be integrated into a collectively informed latent variable that carries information about states relevant to the higher-order system. Unit-specific mappings translate this collective variable into potentially heterogeneous but mutually compatible responses whose joint effects regulate the collectively relevant state toward a preferred or reference state. Bio-CPC thus extends the original CPC architecture by coupling collective inference to an additional biological alignment problem: the establishment, calibration, or maintenance of compatible mappings from collectively informed states to unit-specific responses. Importantly, Bio-CPC is not restricted to decentralized architectures. The same computational organization can be realized along a continuum from predominantly local inference among constituent units to increasingly centralized integration by specialized subsystems. Such centralization may strengthen functional integration within multicellular organisms, as illustrated by the evolution of nervous systems in some animal lineages. The framework also has an evolutionary implication: as higher-order regulation becomes increasingly dependent on information that is available only through multi-unit integration, natural selection should favor communication and integration architectures that improve how such information is transmitted and combined across units. Developmental interactions may coordinate inherited response capacities and communication mechanisms, with recurrently effective mappings subsequently stabilized through genetic accommodation. Bio-CPC is specifically supported when information originating in multiple units contributes causally to a higher-order state estimate and when compatible response mappings are established, calibrated, or maintained. Bio-CPC predicts distinguishable perturbation effects where these two processes can be selectively targeted; conflict-suppression mechanisms help maintain response-mapping alignment. This provides a testable information-processing framework for investigating both the emergence and progressive integration of higher-order biological organization.

Keywords: predictive coding; collective intelligence; active inference; major transitions in evolution; multicellularity; individuality; sociality

Word count: 9,768
Figures: 3
Tables: 3 (incl. Glossary)

Introduction

The biological world is characterized by hierarchically nested levels of organization, including biomolecules, organelles, cells, organs, organ systems, individuals, social groups, populations, communities, and ecosystems. Understanding how new levels of biological organization emerge is one of the most fundamental questions in evolutionary biology. Such emergence is closely related to several of the major transitions in evolution proposed by Maynard Smith and Szathmáry (1995), especially those involving higher levels of organization: multicellularity, eusociality, and symbolic communication. In particular, animal multicellularity and insect eusociality exemplify the fraternal transitions discussed by Queller (2000). Despite the diversity of these transitions, they share a common organizational problem: how relatively autonomous lower-level units become integrated into a coherent higher-level whole.

One important aspect of this problem is associated with *goal-directed regulation*. Living systems characteristically regulate their states and behavior in ways that maintain viability and achieve context-dependent outcomes. Here, we use "goal-directed" in a minimal operational sense to refer to regulation around biologically relevant preferred states; the term does not imply conscious intention or explicit cognitive representation. Depending on internal and external conditions, such systems may prioritize feeding, escape, repair, growth, or reproduction. We refer to systems exhibiting this form of flexible state-dependent regulation as agents. This operational definition is distinct from the developmental, learning, and evolutionary processes through which regulatory capacities arise or change. In the formalization below, this regulatory relation is represented as the deviation between a current biologically relevant state and a preferred state.

Across levels of biological organization, the emergence of a higher-order organization may require the coordination of goal-directed regulation among its constituent units. In multicellular organisms, for example, the activities of individual cells must be coordinated with the maintenance and functioning of the organism as a whole; similarly, in social groups, the behavior of individual members must be coordinated with group-level outcomes. Such coordination can be realized through different architectures. In relatively decentralized systems, information may be integrated through local reciprocal interactions among constituent units, while more strongly integrated systems may evolve specialized structures that concentrate information integration and control. Increases in organizational complexity can therefore involve not only changes in the alignment of goal-directed regulation across levels, but also changes in where and how the relevant information is integrated.

Such coordination creates an information-processing problem. Constituent units must use information about their external and internal environments to regulate their actions in ways that are appropriate not only to their local conditions but also to states relevant to the higher-level organization. Predictive coding provides one account of inference from sensory observations, and active inference extends this perspective to action (Friston, 2010, 2013). When multiple goal-directed units form a higher-order organization, however, the problem is no longer only how each unit

performs such inference, but also how information distributed among units becomes coordinated so that their actions collectively support goal-directed regulation at the higher organizational level.

From this perspective, we take collective predictive coding (CPC; Taniguchi, 2024), a framework for modeling the emergence of symbolic communication, as a starting point. We ask whether its computational architecture can be generalized beyond symbolic communication to a broader form of *collective predictive computation in biological systems*. We refer to this extension as Bio-CPC. Here, computation includes both the construction of collectively informed estimates from information distributed across constituent units and the mapping of those estimates onto physiological, developmental, or behavioral responses. We further ask whether this framework can clarify changes in communication and integration architectures associated with major evolutionary transitions, as well as subsequent changes in the degree of centralization within higher-order systems.

Research on self-organization and collective cognition examines how local interactions generate coordinated biological outcomes (Camazine et al., 2001; Couzin, 2009). Collective sensing, distributed inference, and adaptive control already provide accounts of how information distributed across units can support coordinated regulation (Berdahl et al., 2013; Bandyopadhyay and Chung, 2014; Friston et al., 2024). Bio-CPC focuses on how collective state estimation is coupled to the establishment, calibration, or maintenance of compatible response mappings through interactions among constituent units. These mappings may differ across units, allowing a collectively informed estimate to guide distinct responses that jointly support higher-order regulation. Key terms and their use in this paper are summarized in the Glossary.

Glossary. This glossary explains how key terms are used in this paper.

Term	Usage in this paper
Active inference	A framework coupling inference about hidden states with action under a generative model. Actions can seek preferred outcomes and acquire information.
Agent	A unit or system that exhibits flexible, state-dependent regulation around biologically relevant preferred states. This operational usage does not imply conscious intention.
Codebook	The correspondences between a repertoire of signs and agents' private representations. A codebook includes these correspondences, rather than only the set of available signs.
Collective inference	Inference in which information originating in multiple constituent units contributes causally to an estimate implemented within the collective system. Integration may be distributed or centralized.
Collective predictive coding (CPC)	A framework interpreting symbol emergence as collective inference over a shared latent variable, based on agents' distributed observations and communication. Relations between shared signs and private representations can be learned through interaction.

Computation	Construction of (collectively) informed estimates and their transformation into physiological, developmental, or behavioral responses.
Co-option	The evolutionary recruitment of an existing trait or mechanism for a new function or context.
Darwinian fitness	Reproductive success, or expected reproductive contribution, in a specified ecological and social context. Here, fitness consequences are distinguished from the proximate states that a biological system regulates.
Division of labor	The specialization of different constituent units in different tasks or functions.
Evolutionary individuality	The condition in which a collective functions as an evolutionary individual in its own right. In the sense used here, this involves mutual dependence among its components and sufficiently limited internal fitness conflict.
Generative model	A probabilistic model specifying how observations depend on latent states, together with assumptions about how those latent states are distributed.
Genetic accommodation and assimilation	Genetic accommodation is evolutionary modification of a novel phenotype's regulation or expression through selection on heritable variation. Genetic assimilation is a special case in which an initially environmentally induced phenotype becomes expressed without the inducing condition.
Goal-directedness	Regulation that tends to approach or maintain biologically relevant preferred states or ranges. These states need not be explicitly represented by the system.
Inference model	A model specifying probabilities for latent states conditional on observations or intermediate representations. The joint inference distributions used here also include the empirical distribution of observations.
Latent state	A state treated as unobserved within a model and inferred from observations. Its biological implementation may nevertheless be measurable in an experiment.
Major evolutionary transitions	Major evolutionary changes involving new ways of storing or transmitting heritable information or the emergence of higher-level evolutionary individuals. Transitions in individuality form a subset of these transitions.
Predictive coding (PC)	A framework of inference in which a generative model predicts observations and discrepancies between predicted and actual observations guide updates of inferred states or model parameters. Here, predictive does not necessarily mean forecasting future events.
Representation	A state, pattern, or sign that carries information used in inference or regulation. The term is used functionally here and does not require conscious awareness or an explicitly symbolic format.

Response mapping	A relation connecting an inferred state to a unit's response. In Bio-CPC, the response depends on both collective information and the unit's local state.
Response-mapping alignment	Functional compatibility among unit-specific response mappings, such that their joint effects support an independently specified collective regulatory outcome. In Bio-CPC, inter-unit interactions contribute to establishing, calibrating, or maintaining this compatibility.
System and organization	A system is a set of interacting entities considered together. Organization refers to the arrangement of relations and processes through which those entities function together.
Unit	A constituent component identified relative to the system and organizational level being analyzed. A unit at one level may itself be a system composed of further units.
Variational free energy	An objective for approximate probabilistic inference and model learning. In the formulation used here, it measures the discrepancy between the joint inference and generative distributions through their Kullback–Leibler divergence.

Predictive Coding and Active Inference

Predictive coding was originally proposed as a theory of neural information processing in which the nervous system continuously generates predictions about its sensory inputs and uses prediction errors, the discrepancies between predicted and actual inputs, to update its internal model of the world. In hierarchical formulations, higher levels of a neural hierarchy generate predictions about activity at lower levels, while residual prediction errors are propagated upward to revise those predictions (Rao and Ballard, 1999). Perception can thus be understood not simply as the bottom-up reconstruction of the external world from sensory signals, but as an iterative process of inference in which internal predictions are continuously adjusted to reduce prediction error.

Predictive coding is often situated within the broader framework of the free-energy principle (FEP), which casts perception as approximate inference about the hidden causes of sensory observations (Friston, 2010, 2013). Within this framework, predictive coding provides a possible process theory for perceptual inference, in which predictions are iteratively updated in response to prediction errors. Active inference extends this logic from perception to action. Rather than reducing prediction errors solely by revising internal states, agents can also act on the environment so that subsequent sensory observations become more consistent with preferred outcomes (Friston et al., 2009; Friston, 2010).

Throughout this paper, "predictive" is used primarily in the generative-inference sense: internal or collective latent states support predictions about observations under partially observable conditions. It does not require explicit forecasting of future states. Nor does Bio-CPC require a neural implementation for prediction. Although predictive coding was initially developed primarily as a theory of nervous-system function, the more general logic of predictive regulation need not be restricted to organisms with nervous systems. Living systems must continuously regulate their states in response to uncertain and changing environments, and in many cases successful regulation benefits

from exploiting regularities that predict future conditions rather than merely reacting to present conditions. Indeed, predictive responses have been demonstrated at the cellular level. For example, *Escherichia coli* can evolve regulatory responses that exploit reliable temporal correlations between environmental signals, allowing one environmental cue to trigger physiological changes appropriate for a condition that normally follows it (Tagkopoulos et al., 2008). Such findings have motivated a "predictive-dynamic" view of cellular regulation, in which cellular responses reflect not only immediate homeostatic requirements but also the temporal structure of the environments in which organisms evolved (Freddolino and Tavazoie, 2012). Importantly, such predictive regulation does not presuppose a nervous system or an explicit cognitive representation of future conditions.

Bacterial chemotaxis provides an especially clear example of the coupling between environmental inference and action. *E. coli* searches for favorable chemical environments by alternating relatively straight runs with stochastic tumbles that change its swimming direction. Because of its small size, an individual bacterium cannot reliably determine the direction of a chemical gradient by simultaneously comparing concentrations at spatially separated receptors. Instead, it compares chemical signals across time; classical experiments have shown that its chemotactic response depends on a weighted history of sensory stimulation extending several seconds into the past (Segall et al., 1986). The bacterium must therefore use temporally varying and noisy sensory information to regulate its probability of changing direction.

The computational relation between inference and action can be summarized by a minimal single-agent model (Figure 1). Let s denote a partially observable state of the agent's internal and external environments, o a local sensory observation generated by that state, and s^* a preferred or reference state. The agent forms a latent or internal state z from o. For present purposes, the perceptual component can be represented by an agent-specific generative model

$$p_\theta(z, o) = p_\theta(z)p_\theta(o \mid z),$$

together with an inference model

$$q_\phi(z, o) = q_\phi(z \mid o)q(o).$$

Here, θ and ϕ respectively parameterize the generative and inference models. $q_\phi(z, o)$ denotes the joint inference model, while $q_\phi(z \mid o)$ is the conditional inference distribution over the latent state given an observation. $q(o)$ denotes the empirical distribution of observations. Following the joint-distribution formulation used below for CPC, variational free energy is defined as the Kullback–Leibler divergence between the inference and generative models:

$$\begin{aligned} \mathcal{F}(p_\theta, q_\phi) &\overset{\text{def}}{=} \mathcal{D}_{\text{KL}}\left[q_\phi(z, o) \| p_\theta(z, o)\right] \\ &= \underbrace{\mathbb{E}_{q_\phi}\left[\ln \frac{q_\phi(z \mid o)}{p_\theta(z)}\right]}_{\text{Regularization}} + \underbrace{\mathbb{E}_{q_\phi}\left[\ln \frac{q(o)}{p_\theta(o \mid z)}\right]}_{\text{Reconstruction error}}. \end{aligned}$$

The first term regularizes the inferred latent representation relative to its prior distribution, while the second corresponds to the reconstruction term relating observations to latent states. The inferred state is then coupled to action. At this level of abstraction, action selection can be represented schematically as

$$a \sim \pi(a \mid z),$$

where π denotes the state-dependent response policy. In *E. coli*, the relevant actions include running and tumbling. Actions in turn change the bacterium's relation to its environment and thereby alter subsequent sensory observations. Goal-directedness is then specified at the regulatory level. We use s^* to denote a preferred or reference state of the environment, or more generally a preferred region of state space. We write $d(s, s^*)$ for a nonnegative discrepancy function that is minimized when the current state s coincides with the preferred state or lies within the preferred region. The action policy is goal-directed insofar as its effects tend to reduce this discrepancy. The preferred state or region need not be explicitly represented as an additional input to the policy; it may instead be implicit in the regulatory architecture that determines how inferred states are mapped onto actions.

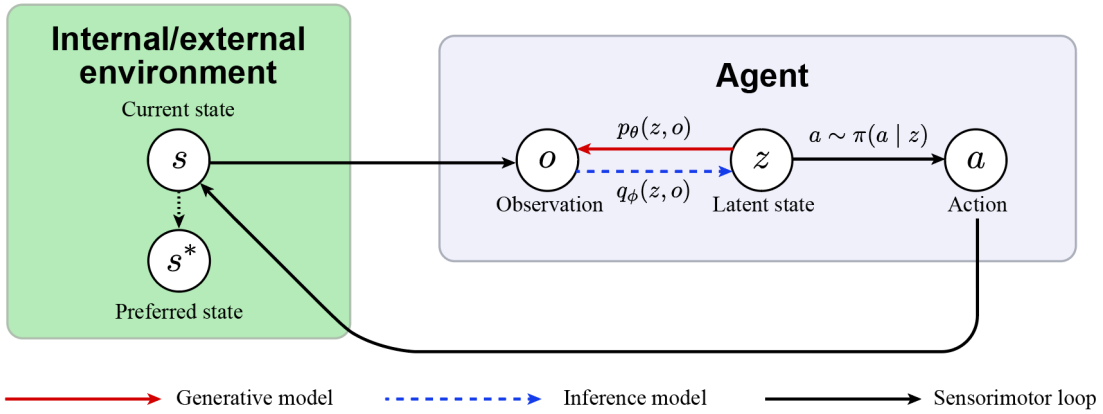


Figure 1. Single-agent predictive computation. The current agent–environment state s gives rise to an observation o, which is mapped onto a latent/internal state z that guides action a. A preferred state s^* provides the reference for goal-directed regulation, such that the effects of action tend to reduce $d(s, s^*)$. The red and blue pathways represent the joint generative and inference models, $p_\theta(z, o)$ and $q_\phi(z, o)$, respectively.

More detailed models instantiate this general architecture in different ways. Tschantz et al. (2020) formulated bacterial chemotaxis within active inference, coupling beliefs about environmental states, model parameters, and control states to action selection. Nakamura and Kobayashi (2021) related the biochemical network underlying *E. coli* chemotaxis to an optimal filter that estimates whether the cell is moving up or down a ligand gradient from noisy sensory signals. They subsequently formulated run-and-tumble chemotaxis as a partially observable Markov decision process combined with Kullback–Leibler control, thereby coupling sensory estimation to motor control (Nakamura and Kobayashi, 2022). These models differ in implementation, but all illustrate the single-agent computational baseline relevant here: partially observable states are inferred from local

sensory information, and inferred states are coupled to actions that alter the agent–environment relation. None of these models by itself establishes that bacteria literally implement neural predictive coding, nor is such an assumption required here.

Collective Predictive Coding (CPC)

The predictive-computation architecture described above is formulated at the level of a single agent: observations inform an internal estimate of partially observable states, and that estimate contributes to the selection of actions. The central question for collective organization is what changes when the information required for effective regulation is distributed across multiple agents or units. When multiple agents interact, however, no single agent normally has direct access to the observations or internal states of all the others. This raises a further inferential problem: how can information distributed across multiple agents be integrated without a central processor that has access to the entire system?

The CPC hypothesis was proposed as a framework for understanding such distributed inference, originally in the context of symbol emergence (Taniguchi, 2024). CPC regards a symbol system not simply as a channel through which already established meanings are transmitted, but as a shared representation that emerges through interactions among agents. Each agent performs inference on the basis of its own observations and internal generative model, while communication among agents enables the system as a whole to approximate inference over a representation shared among them.

Consider N agents jointly attending to an object. Each agent i receives a local observation o_i of the object and forms a private latent representation z_i. CPC introduces an additional latent variable m shared across agents via messages about the jointly attended object. Writing $\mathbf{z} = (z_1, \dots, z_N)$ and $\mathbf{o} = (o_1, \dots, o_N)$, a simplified generative model can be represented as

$$p_\theta(m, \mathbf{z}, \mathbf{o}) = p(m) \prod_{i=1}^{N} p_{\theta_i}(z_i \mid m) p_{\theta_i}(o_i \mid z_i).$$

The corresponding joint inference model can be factorized as

$$q_{\phi,\psi}(m, \mathbf{z}, \mathbf{o}) = q_\psi(m \mid \mathbf{z}) \prod_{i=1}^{N} q_{\phi_i}(z_i \mid o_i) q_i(o_i).$$

Here, $q_i(o_i)$ specifies the marginal distribution over local observations, $q_{\phi_i}(z_i \mid o_i)$ specifies local conditional inference from observations to private latent states, and $q_\psi(m \mid \mathbf{z})$ specifies collective inference over the shared latent variable. The factorization assumes that m is conditionally independent of the observations $\mathbf{o}$ given the private latent states $\mathbf{z}$.

The collective variational free energy is then defined as the Kullback–Leibler divergence between the joint inference and generative models:

$$\mathcal{F}_{\mathrm{CPC}}(p_\theta, q_{\phi,\psi}) \overset{\mathrm{def}}{=} \mathcal{D}_{\mathrm{KL}}\left[q_{\phi,\psi}(m, \mathbf{z}, \mathbf{o}) \| p_\theta(m, \mathbf{z}, \mathbf{o})\right]$$

$$= \underbrace{\mathbb{E}_q\left[\ln \frac{q_\psi(m \mid \mathbf{z})}{p(m)}\right]}_{\text{Collective regularization}} + \underbrace{\sum_{i=1}^{N} \mathbb{E}_q\left[\ln \frac{q_{\phi_i}(z_i \mid o_i)}{p_{\theta_i}(z_i \mid m)}\right]}_{\text{Agent-specific regularization}} + \underbrace{\sum_{i=1}^{N} \mathbb{E}_q\left[\ln \frac{q_i(o_i)}{p_{\theta_i}(o_i \mid z_i)}\right]}_{\text{Prediction error}}.$$

The first term is intrinsically collective, whereas the latter two decompose across agents. The second term regularizes agent-specific latent states relative to the collective variable, while the third is a local reconstruction term relating the observation distribution to the generative likelihood. Accordingly, θ_i, ϕ_i, and ψ respectively parameterize the generative model, local inference, and collective inference summarized in Figure 2.

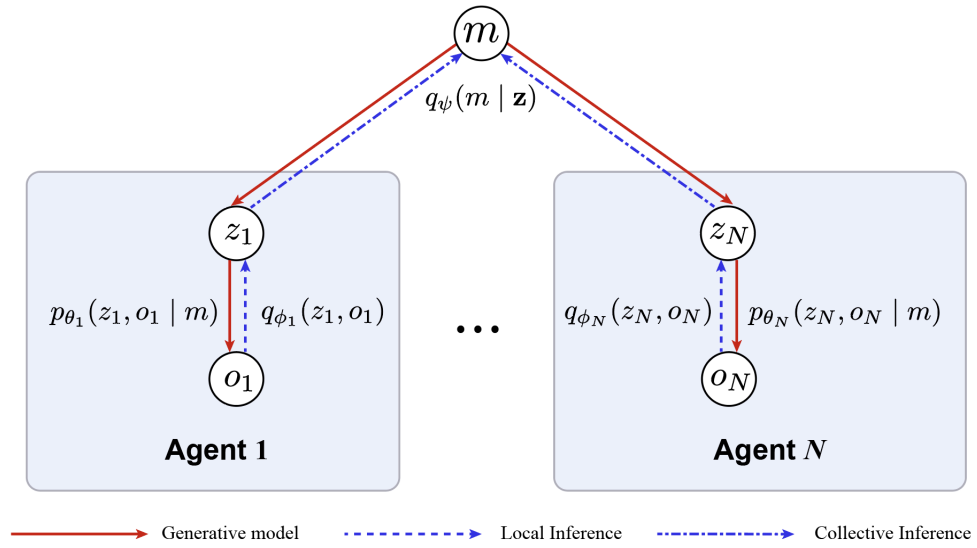


Figure 2. Simplified computational architecture of CPC. Each agent $i \in \{1, \ldots, N\}$ receives a local observation o_i and maintains a private latent representation z_i. The shared symbolic latent variable m couples the otherwise cognitively closed agents. The local generative relation is written compactly as $p_{\theta_i}(z_i, o_i \mid m) = p_{\theta_i}(z_i \mid m)p_{\theta_i}(o_i \mid z_i)$, while the local inference relation is written as $q_{\phi_i}(z_i, o_i) = q_{\phi_i}(z_i \mid o_i)q_i(o_i)$. Collective inference over the shared latent variable is represented by $q_\psi(m \mid \mathbf{z})$. Together, these factors define the joint generative and inference models used in the collective free-energy formulation. The graphical representation of q_ψ does not imply a centralized encoder with direct access to all agent-specific latent states; in CPC, the corresponding collective inference can be approximated through decentralized communication, as in the Metropolis–Hastings naming game.

The Metropolis–Hastings naming game (MHNG) provides one example of decentralized inference over a shared latent variable m (Taniguchi et al., 2023). Although m is latent with respect to sensory observations in the generative model, proposed values of m in MHNG are exchanged as observable signs during communication. The speaker proposes a sign according to its own generative model, and the listener evaluates the proposal against its own private representation, without directly

accessing the speaker's internal state. Under the model's assumptions, repeated exchanges approximate posterior sampling over m conditioned on information from both agents; the recursive extension generalizes this procedure to multiple agents (Inukai et al., 2023). These algorithms illustrate how the collective inference represented by $q_\psi(m \mid \mathbf{z})$ can be approximated through local communication. Nevertheless, Bio-CPC does not require this particular algorithm or interaction protocol.

An important consequence is that the codebook used by communicating agents can itself be shaped through communication. Here, a codebook refers to the correspondences between a repertoire of signs and the agents' private perceptual representations. Conventional models of communication typically assume that sender and receiver already share, at least approximately, a mapping between signals and what those signals represent. CPC instead addresses how such mappings can become coordinated through interactions. In the model above, the agent-specific generative relation $p_{\theta_i}(z_i \mid m)$ describes how a shared sign m is related to the private latent representation of agent i. These mappings need not initially be identical across agents. When sign exchange is coupled to learning of each agent's generative-model parameters, these correspondences can become coordinated through interaction. In this sense, a shared codebook can emerge through communication itself, rather than being fully specified before communication begins. More precisely, implementations of MHNG may assume a predefined repertoire of possible signs, but the correspondences between those signs and the agents' perceptual categories, the semantics of the code, are learned and aligned through interaction (Inukai et al., 2023; Hayashi, 2026).

CPC therefore adds a level of inference that is absent from ordinary individual predictive coding. Individual agents infer latent states from their own observations, while interactions among them support inference over a shared latent variable that is not reducible to the information available to any one agent. In its original formulation, this shared variable is given a symbolic interpretation; concrete samples of m function as externally exchanged signs, while repeated interaction aligns the relations between those signs and the agents' internal representations.

It is important, however, to distinguish this formal role of m as a shared latent variable from its interpretation as a symbolic sign. The mathematical architecture of CPC identifies a variable whose state is collectively constrained by information distributed among multiple agents; its interpretation as a linguistic or symbolic sign follows from the domain in which CPC was originally developed. This distinction raises a broader question: could the same basic architecture of distributed inference be applied when the collectively constrained variable has a biological rather than symbolic interpretation? We consider this possibility in the following section.

In making this generalization, we retain the computational organization of CPC while extending its interpretation beyond symbolic communication and allowing different biological implementations. We therefore reserve the term *collective predictive coding* for the original framework and use *collective predictive computation* for the broader biological architecture proposed here. This change in terminology is deliberate. In Bio-CPC, collectively inferred states need not be expressed through a shared repertoire of symbolic signs. They may instead become functionally effective by being transformed into physiological, developmental, or behavioral responses. Bio-CPC thus concerns two coupled information-processing transformations: the construction of a higher-order estimate from information originating in multiple units, and the translation of that estimate through unit-specific mappings into mutually compatible responses. We use "computation" here in this limited functional sense; the term does not imply a digital, neural, or symbolic implementation.

Biological Extension for Collective Predictive Computation

We now extend this architecture to biological systems in which collective inference is coupled to physiological, developmental, or behavioral regulation. Here, m retains its role as a latent variable inferred from distributed observations, while a biological goal specifies a preferred or reference state toward which responses are directed. In the notation introduced above, inference concerns the current state s, whereas goal-directed regulation is evaluated relative to the preferred state s^*.

Specifically, we distinguish the state relevant to collective regulation from the latent variable through which distributed information about that state is integrated. Figure 3A summarizes the proposed architecture, and Table 1 describes its principal variables and relations. Let s_C denote an independently measurable state of the higher-order system or its environment that is causally or predictively relevant to collective regulation. Each constituent unit i receives a local observation o_i and forms a unit-specific latent or internal state z_i. We denote by m_C a collectively informed latent variable whose posterior is constrained by information originating in multiple units. The variable m_C is not identified with s_C by definition, nor is it required to be a point estimate of s_C. Their relationship must be established empirically. The distinction concerns their roles in the model: s_C is the state whose regulation is evaluated, m_C integrates information originating in multiple units, and s_C^* specifies the regulatory reference. Calling m_C latent does not imply that its biological implementation is inaccessible to experimental measurement. Conversely, the independent measurability of s_C refers to the experimenter's ability to assess it, not to direct access by every constituent unit. Depending on the system, s_C may correspond to the physiological condition of an organism, the developmental state or composition of a tissue, or an external condition such as the location of a resource or predator. A candidate s_C should therefore not be an arbitrary macroscopic descriptor selected by the observer.

Table 1. Notation and formalization of Bio-CPC.

Symbol / relation	Interpretation in Bio-CPC
s_C	Independently measurable state of the higher-order system's internal and external environments that is causally or predictively relevant to collective regulation.
s_C^*	Preferred or reference state corresponding to s_C; it specifies the proximate collective-level biological goal.
$d(s_C, s_C^*)$	System-specific measure of deviation between the current collectively relevant state and its preferred or reference state.
o_i	Local observation available to constituent unit i, including signals originating from the external environment or from other constituent units.
z_i	Unit-specific latent or internal state inferred from local information.

$\mathbf{o} = (o_1, \dots, o_N)$	Local observations distributed across the constituent units.
$\mathbf{z} = (z_1, \dots, z_N)$	Unit-specific latent or internal states.
m_C	Collectively informed latent variable constrained by information originating in multiple units. It is not identified by definition with either s_C or a biological goal.
$p_\omega(s_C \mid m_C)$	Observer-level decoding model used to evaluate how inferred values of m_C relate to independently measured s_C.
a_i	Response generated by constituent unit i.
$\pi_i(a_i \mid m_C, z_i)$	Unit-specific biological response mapping through which the collective latent variable and the unit-specific state jointly influence response a_i.

At the computational level, the simplest Bio-CPC model retains the factorization of the original CPC architecture. Writing $\mathbf{z} = (z_1, \dots, z_N)$ and $\mathbf{o} = (o_1, \dots, o_N)$, the generative model can be expressed as

$$p_\theta(m_C, \mathbf{z}, \mathbf{o}) = p_\theta(m_C) \prod_{i=1}^{N} p_{\theta_i}(z_i \mid m_C) p_{\theta_i}(o_i \mid z_i).$$

The corresponding joint inference model can be factorized as

$$q_{\phi,\psi}(m_C, \mathbf{z}, \mathbf{o}) = q_\psi(m_C \mid \mathbf{z}) \prod_{i=1}^{N} q_{\phi_i}(z_i \mid o_i) q_i(o_i).$$

Here, $q_i(o_i)$ specifies the marginal distribution over the local observations available to unit i, $q_{\phi_i}(z_i \mid o_i)$ specifies local conditional inference, and $q_\psi(m_C \mid \mathbf{z})$ specifies collective inference supported by inter-unit signaling and communication. As in the CPC formulation, these equations describe computational dependencies without prescribing their biological implementation, including the network structure and temporal dynamics of inter-unit interactions.

The corresponding Bio-CPC variational free energy is defined as the Kullback–Leibler divergence between the joint inference and generative models:

$$\begin{aligned}\mathcal{F}_{\text{BioCPC}}(p_\theta, q_{\phi,\psi}) &\overset{\text{def}}{=} \mathcal{D}_{\text{KL}}\left[q_{\phi,\psi}(m_C, \mathbf{z}, \mathbf{o}) \| p_\theta(m_C, \mathbf{z}, \mathbf{o})\right] \\ &= \underbrace{\mathbb{E}_q\left[\ln \frac{q_\psi(m_C \mid \mathbf{z})}{p_\theta(m_C)}\right]}_{\text{Collective regularization}} + \underbrace{\sum_{i=1}^{N} \mathbb{E}_q\left[\ln \frac{q_{\phi_i}(z_i \mid o_i)}{p_{\theta_i}(z_i \mid m_C)}\right]}_{\text{Unit-level regularization}} + \underbrace{\sum_{i=1}^{N} \mathbb{E}_q\left[\ln \frac{q_i(o_i)}{p_{\theta_i}(o_i \mid z_i)}\right]}_{\text{Prediction error}}.\end{aligned}$$

The first term is intrinsically collective, whereas the latter two decompose across constituent units. The second term regularizes unit-specific latent states relative to the collectively informed variable m_C, while the third is a local reconstruction term relating local observation distributions to the generative likelihoods. This decomposition makes explicit the distinction between inference performed locally by individual units and the additional constraint introduced by the collectively informed variable m_C.

The independently measurable state s_C does not itself enter this variational objective. Instead, its relation to m_C provides an empirical criterion for evaluating whether the inferred collective variable carries information about the state of interest. This relation can be quantified by an externally estimated readout model,

$$p_\omega(s_C \mid m_C),$$

fitted from inferred values of m_C paired with measurements of s_C obtained independently of the local observations used to infer m_C. Where a point estimate is useful, one may define

$$\hat{s}_C = \mathbb{E}_{p_\omega(s_C|m_C)}[s_C].$$

This readout is an observer-level validation model, not a biological response mapping assumed to be implemented by the system. Its role is to test whether m_C tracks the independently measurable state s_C, for example through cross-validated decoding accuracy, mutual information, or posterior uncertainty.

The observer-level readout $p_\omega(s_C \mid m_C)$ should be distinguished from the biological mappings through which the collective variable becomes causally effective. The response of unit i can be represented as

$$a_i \sim \pi_i(a_i \mid m_C, z_i).$$

To connect collective inference explicitly to goal-directed regulation, we denote by s_C^* a preferred or reference state of s_C, and by $d(s_C, s_C^*)$ the deviation of the current collectively relevant state from that preferred state. Minimization of variational free energy concerns the accuracy and regularization of collective inference and does not by itself imply regulation toward s_C^*. Goal-directed regulation enters through the response mappings π_i, whose joint effects are evaluated by whether they reduce $d(s_C, s_C^*)$. The heterogeneous responses a_i are functionally compatible insofar as their joint effect tends to reduce this deviation. Thus, m_C contributes to inference about the current collectively relevant state, while s_C^* specifies the state toward which collective regulation is directed. The

collective reference s_C^* need not be explicitly represented by every constituent unit. It may instead be implemented through regulatory architectures in which different local response capacities jointly maintain the relevant collective state. How interactions establish or preserve this compatibility is the biological alignment problem introduced below.

The same m_C need not generate the same response a_i in every unit, because the appropriate response depends also on the unit-specific state z_i. Higher-order biological organization may therefore depend on heterogeneous but mutually compatible responses to the collectively informed variable m_C. The dependence of a_i on m_C is specified at the computational level; it does not require every unit to access the same physically localized state. In a distributed realization, effects summarized by m_C may be mediated through signals that become part of a unit's subsequent observations o_i. Responses a_i can therefore alter other units' observations and contribute to later updates of z_i and m_C. In a centralized realization, a specialized integrator may implement a more localized estimate and transmit signals that influence downstream responses. The factorizations of $p_\theta(m_C, \mathbf{z}, \mathbf{o})$ and $q_{\phi,\psi}(m_C, \mathbf{z}, \mathbf{o})$ specify computational dependencies while abstracting from these temporal feedback loops.

This introduces a biological alignment problem that is structurally related to, but not identical with, the codebook problem in CPC. In original CPC, interaction aligns the mappings between a shared sign and agent-specific internal representations. Bio-CPC retains the inferential relation between the collective variable and unit-specific latent states, but additionally asks how a collectively constrained estimate becomes causally effective through heterogeneous unit-specific responses. We therefore focus on the family of unit-specific response mappings $\{\pi_i(a_i \mid m_C, z_i)\}_{i=1}^N$. These mappings are aligned not when they are identical across units, but when their heterogeneous outputs are jointly compatible with regulation of s_C toward s_C^*, that is, with reducing or maintaining a low value of $d(s_C, s_C^*)$. The relevant preferred state and discrepancy measure must be specified independently of the mappings being evaluated. In multicellular organisms, for example, such a criterion may be independently measurable as a regulated range of tissue composition, maintenance of organism-level morphology, or successful resource acquisition. Inherited or previously learned architectures may constrain these mappings, but interaction among units must contribute to establishing, calibrating, or maintaining their functional compatibility. Interaction-dependent alignment need not involve changes to the mappings themselves during each regulatory episode. It may also involve the context-dependent recruitment of established response capacities, provided that interaction contributes to the functional compatibility of their joint deployment.

In the proposed biological extension, a biological system approximates the collective-inference architecture of CPC when the following four conditions are satisfied. First, information about a collectively relevant state is distributed across multiple units, and no single unit, using information available independently of the others, can recover an estimate equivalent to the collective estimate in all relevant respects, such as accuracy, precision, robustness, or scope. Second, information originating in multiple units contributes causally to a higher-order estimate; this integration may be realized through distributed interactions, convergence onto specialized integrators, or combinations of the two. Third, the resulting collectively informed estimate influences heterogeneous unit-specific responses through distinct estimate-to-response mappings. Fourth, interaction establishes, calibrates, or maintains sufficient compatibility among these mappings for the resulting heterogeneous unit responses to regulate s_C toward s_C^*, thereby supporting a coherent

collective-level outcome. These criteria are therefore agnostic about whether the physical implementation of integration is decentralized or centralized. Parallel responses to a common input without inter-unit information exchange, local feedback regulation without multi-unit state estimation, and distributed sensing without system-internal integration do not by themselves establish Bio-CPC. Conversely, centralization does not disqualify a system if the central estimate is itself causally constructed from information distributed among constituent units and participates in coordinating their heterogeneous response mappings. The present formalization specifies inferential dependencies and a functional criterion for response compatibility. Still, it does not specify the dynamics by which inter-unit interactions establish, calibrate, or maintain this compatibility, or by which the resulting responses change s_C. Whether this fourth condition is satisfied must therefore be evaluated in each biological system.

The relative contribution of unit-specific and collectively integrated information can vary continuously across biological systems. Figure 3B illustrates a condition in which responses depend primarily on local states z_i, while m_C makes a limited but nonzero contribution through the response mappings $\pi_i(a_i \mid m_C, z_i)$. In Figure 3C, responses depend more strongly on m_C, although local states continue to provide information for collective inference and to shape unit-specific responses. These panels illustrate a possible evolutionary transition from predominantly local regulation in early multicellular organization toward greater reliance on collective estimates formed by, for example, specialized neural integrators. Nevertheless, functional dependence on m_C remains distinct from anatomical or physical centralization: strong reliance on collective estimates can also be realized through distributed interactions.

Here, s_C^* specifies a proximate regulatory outcome. Its evolutionary relevance depends on the consequences of the underlying regulatory architecture for survival and reproduction, without requiring organisms to represent or directly estimate their own Darwinian fitness. Across generations, evolution shapes receptors, regulatory mechanisms, preferences, and communication architectures. Developmental processes construct and stabilize higher-order organization through interactions among constituent units, while physiological or behavioral processes use established mechanisms to infer states and generate context-dependent responses. These processes can influence one another but should not be conflated.

Response-mapping alignment and biological integrity

The formation of a higher-order biological organization creates the possibility of tension between levels (Maynard Smith and Szathmáry, 1995). A constituent unit may respond in ways favorable to its own persistence or reproduction but detrimental to the collective. Conversely, higher-level organization may constrain the behavior or reproduction of its components. Whether and how this tension is mitigated during evolution is a central question in studies of major transitions in evolution. Research has focused on mechanisms that reduce within-group evolutionary conflict, promote interdependence and division of labor, and coordinate the activities of constituent units (West et al., 2015). The resolution of lower-level evolutionary conflicts is one of the central conditions under which a higher-level organization can itself acquire properties of an evolutionary individual.

Response-mapping alignment concerns the compatibility of unit-specific responses with a collective regulatory outcome. Constituent units may produce different responses while contributing to the same collective goal; whether they also have distinct local regulatory objectives is a separate question. Biological integration typically entails functional differentiation, so aligned mappings need

not generate similar responses in different constituent units. High integration is achieved when these heterogeneous local responses collectively maintain the relevant higher-level organization. Such compatibility need not eliminate trade-offs among regulated variables. The relevant criterion is whether the joint response supports an independently specified collective regulatory outcome.

This formulation is complementary to, but narrower than, general accounts of distributed control, collective intelligence (including swarm intelligence), active inference, and major transitions in evolution. Bio-CPC is not intended as a general label for system-level regulation emerging from local interactions. Its specific explanatory contribution is to decompose such regulation into two coupled processes: (i) multi-unit integration of information into an estimate of a collectively relevant state, and (ii) heterogeneous unit-specific response mappings whose functional compatibility is established, calibrated, or maintained through interaction. Biological systems may additionally require mechanisms that preserve this alignment when components become noisy, unresponsive, or non-cooperative. This decomposition distinguishes failures of collective estimation from failures of response mapping, while conflict-suppression mechanisms determine the robustness of the latter against internally generated disruption.

The resulting proposal can be summarized in Figure 3 and Table 2.

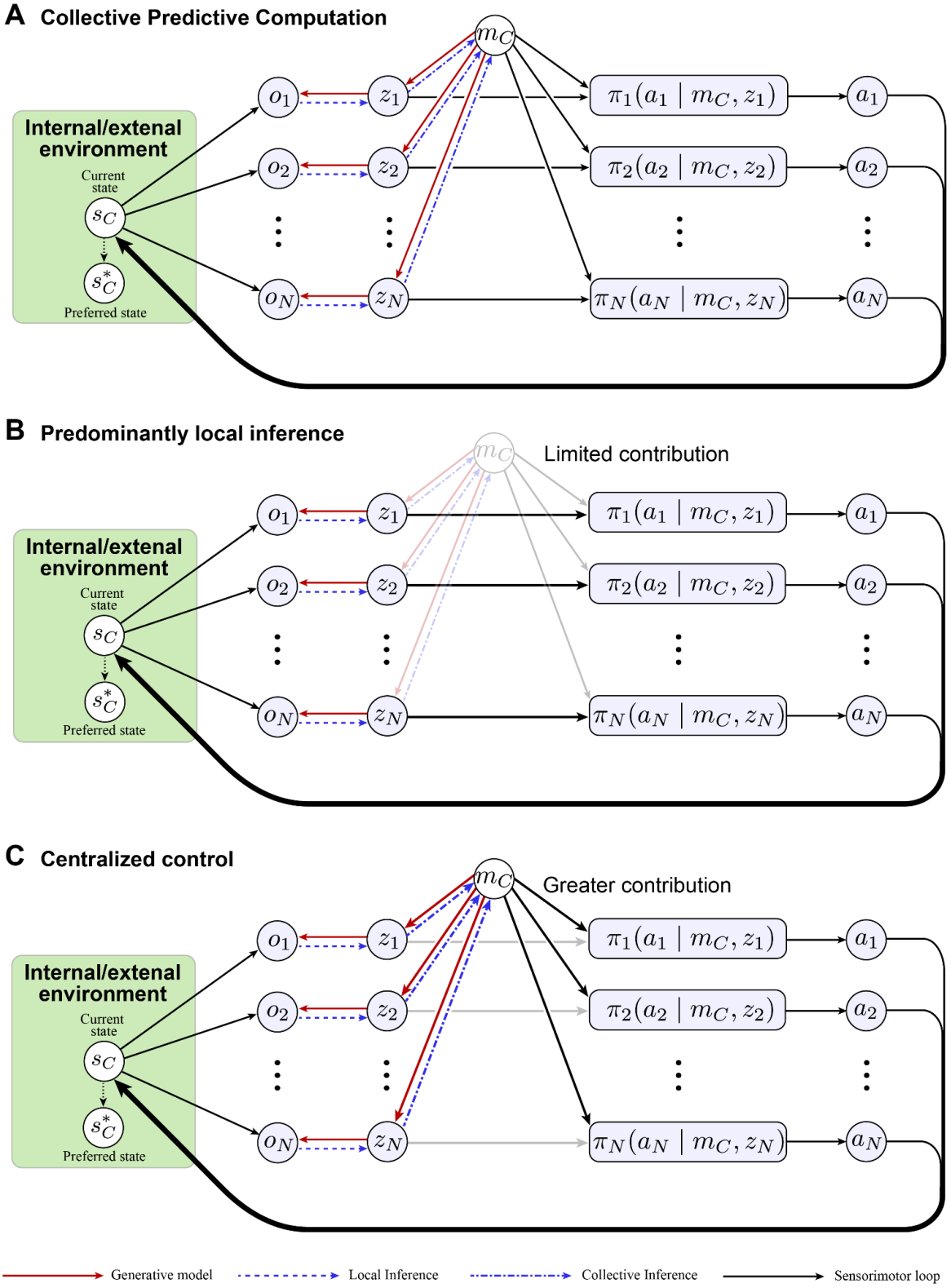


Figure 3. Collective inference architecture of Bio-CPC. (A) Information about a collectively relevant state s_C is distributed across local observations o_i available to constituent units. Each unit forms a unit-specific latent or internal state z_i, and information originating in multiple units is integrated into a collectively informed latent variable m_C. The relation between m_C and the independently measurable state s_C is empirical rather than definitional and may be evaluated by an external readout model. Together with the

unit-specific latent state z_i, m_C influences heterogeneous responses a_i through unit-specific response mappings $\pi_i(a_i \mid m_C, z_i)$. The mappings are functionally aligned when their heterogeneous outputs jointly support a measurable collective-level outcome. (B, C) Illustrative cases with different degrees of reliance on collective inference. In B, unit responses depend mainly on local states z_i, with a limited contribution from m_C. In C, collective inference is concentrated in a specialized integrator, and m_C plays a greater role in coordinating unit-specific responses. Fainter nodes and arrows indicate smaller functional contributions to regulation. The black feedback loop represents the effects of unit responses on s_C; explicit temporal updates are not shown.

Table 2. Comparison between original CPC and Bio-CPC.

	Original CPC	**Bio-CPC**
Higher-level latent variable	Shared symbolic latent variable m	Collectively informed latent variable m_C, empirically related to an independently measurable state s_C
Mapping	Sign-to-representation mapping	Collective-variable-to-unit-specific response mapping
Alignment problem	Codebook alignment	Response-mapping alignment
Integration architecture	Decentralized inference through interactions among agents; no central integrator	Distributed, centralized, or intermediate integration
Higher-order outcome	Symbol-system formation	Higher-order biological integration

The following sections examine this proposal across different scales of biological organization. These cases should not be interpreted as a single obligatory evolutionary sequence. Nevertheless, the evolution of functional integration within multicellular organisms provides an important case in which the biological implementation of collective inference may shift from predominantly decentralized interactions toward increasingly centralized integration. This transition is not necessarily monotonic or universal, and distributed and centralized processes can coexist within the same organism. Moreover, the hierarchy is recursive: a highly integrated multicellular individual can itself become a constituent agent in a new, relatively decentralized collective at the social level. We therefore consider a single-celled organism as a single-agent baseline, multicellularity as a domain in which collective inference first arises among cells, organismal individuality as a domain in which this inference may become increasingly centralized, and sociality as a higher-order domain in which already integrated individuals again interact as relatively autonomous agents.

Single-celled organism

A single-celled organism provides a single-agent baseline at the cellular scale considered here. As illustrated above by bacterial chemotaxis, a single cell can infer partially observable environmental states and map them onto context-dependent actions (Segall et al., 1986; Nakamura and Kobayashi, 2021, 2022). Predictive coding and active inference provide good descriptions of its inference and regulation. Still, because the cell is treated as one agent, this description does not require alignment among separate cells. Multicellularity introduces the additional problem central to Bio-CPC: how multiple units with different local information and potentially differentiated response capacities coordinate their mappings so that heterogeneous actions jointly support a higher-order multicellular system.

Multicellularity

The transition to multicellularity, which occurred multiple times on the tree of life (Lamża, 2023; Bozdag et al., 2026), fundamentally changes the alignment problem. In a multicellular system, each cell receives spatiotemporally localized signals and generates a distinct response accordingly, yet the responses of multiple cells must collectively produce and maintain a coherent higher-order organization. Importantly, many molecular components required for such coordination did not originate de novo with multicellularity. Signaling, adhesion, and regulatory systems already present in unicellular ancestors were repeatedly co-opted and elaborated for cell–cell communication and developmental regulation (Grosberg and Strathmann, 2007; Rokas, 2008; Sebé-Pedrós et al., 2017; Ruiz-Trillo et al., 2023). Multicellularity can thus be viewed partly as a transition in which mappings originally linking individual cells to their environment became incorporated into coordinated mappings among cells within a higher-order system.

What, then, is the relevant collective-level variable in such a system? At the evolutionary timescale, division of labor can increase the Darwinian fitness of the multicellular collective by allowing different cells to specialize in functions that cannot be performed simultaneously with equal efficiency by a single cell. A fundamental example is the trade-off between viability-enhancing and reproductive functions. Depending on the structure of this trade-off, specialization can be favored, potentially leading to germ–soma differentiation and a shift in fitness from the cellular to the collective level (Michod, 2006; Ispolatov et al., 2012; West et al., 2015). At the developmental timescale, Bio-CPC concerns proximate states of the developing system, such as cell-type composition, tissue morphology, or spatial patterning. Natural selection can shape the regulatory architectures and preferred states associated with these processes, but the framework does not require cells to estimate Darwinian fitness itself.

Dictyostelia provide a particularly instructive example in which the evolutionary co-option of cellular signaling can be connected to such proximate collective regulation. Components of cyclic adenosine monophosphate (cAMP) signaling used ancestrally in unicellular encystation were recruited for extracellular signaling controlling aggregation, migration, and differentiation during multicellular development (Du et al., 2015; Kin and Schaap, 2021). In *Dictyostelium discoideum*, aggregated amoebae differentiate principally into reproductive spores and non-reproductive stalk cells. Remarkably, the proportions of prestalk and prespore cells are regulated over a broad range of aggregate sizes, and experimentally perturbed proportions can return toward their characteristic values. Secreted signals produced by differentiating cells contribute to this proportion regulation (Oyama et al., 1983; Inouye, 1989). The relevant collective state is therefore not simply the fate of any particular cell, but the composition of the multicellular aggregate. Individual cells respond to

locally available signals by changing their differentiation state, while their combined responses maintain the relative abundance of cell types within an appropriate range.

This system provides a concrete candidate mapping between Bio-CPC and experimentally accessible quantities. The collective state s_C can be defined as the prestalk–prespore composition of the aggregate; o_i corresponds to locally available cues; m_C would represent a collectively informed estimate of that composition, or of its deviation from the regulated range; and z_i and a_i correspond, respectively, to cell-specific internal states or fate tendencies and differentiation responses. This interpretation remains a working hypothesis; in particular, proportion regulation by itself does not demonstrate collective inference. A stronger test would require showing that individual cells possess only partial information about aggregate composition and that intercellular interaction improves the effective estimation of deviations from the appropriate collective state.

The distinction between collective estimation and downstream response mapping also suggests an experimental test. Here, intercellular coupling supports the integration of information from different cells summarized by $q_\psi(m_C \mid \mathbf{z})$. Selectively disrupting this coupling, while preserving the cells' capacity to detect the cues that remain available and to execute downstream differentiation, should reduce the contribution of multi-cell information to the collective estimate. Conversely, selectively perturbing fate-response pathways downstream of the inferred collective state would alter the response mappings $\pi_i(a_i \mid m_C, z_i)$. Such perturbations should initially preserve information about aggregate composition in m_C while disrupting differentiation responses that would restore cell-type proportions to their regulated range. Demonstrating this double dissociation, within a temporal window that distinguishes primary effects from subsequent feedback, would provide substantially stronger evidence for Bio-CPC than proportion regulation alone.

The connection with predictive coding becomes particularly explicit in theoretical work on morphogenesis. Friston et al. (2015) modeled development as active inference by an ensemble of cells that share a generative model of a target morphology. Each cell receives local chemical signals from other cells and uses them to infer its own position and identity within the developing ensemble. It then changes its position and differentiation state so that the signals it encounters and produces become increasingly consistent with those predicted by the shared model. Crucially, no cell has direct access to the complete configuration of the organism. The target morphology emerges only when the locally inferred states and responses of all cells become mutually consistent.

This model is particularly close to the collective-estimation component of Bio-CPC. Reciprocal signaling allows locally informed cells to coordinate around a higher-order morphology even though no individual cell has access to the complete configuration. However, active-inference models of morphogenesis do not by themselves establish the second, Bio-CPC-specific requirement proposed here: that interaction calibrates the heterogeneous mappings through which collectively relevant states are translated into cell-specific responses. A stronger Bio-CPC interpretation would therefore require evidence that intercellular interaction not only improves estimation of a higher-order state but also establishes or maintains compatibility among these response mappings.

Multicellularity therefore provides a natural domain in which an individual state-to-response mapping can be transformed into a collectively coordinated set of mappings. In relatively simple multicellular systems, the corresponding collective estimate may remain distributed across reciprocal cell–cell interactions, with no specialized structure uniquely responsible for integration. Such systems represent the relatively decentralized pole of Bio-CPC. Further functional integration within

multicellular organisms can then involve not only stronger alignment among cellular mappings but also changes in how collective information is integrated.

Individuality

Evolutionary individuality involves mutual dependence among constituent units and sufficiently limited within-group conflict for the collective to be treated as an individual in its own right (West et al., 2015). Here, we focus on functional integration within multicellular individuals and on evolutionary changes in how information from multiple constituent units is integrated. In some animal lineages, the evolution of *nervous systems* provided increasingly specialized mechanisms for combining information from multiple tissues and coordinating heterogeneous responses. Such centralization is a possible mode of functional integration rather than a defining condition of evolutionary individuality. From the Bio-CPC perspective, the emergence of a specialized neural integrator can constitute an evolutionarily derived implementation in which information originating across many constituent units is increasingly concentrated into higher-order neural states that coordinate responses throughout the multicellular organism.

What matters is therefore not whether a central integrator exists, but how its informational state is constructed. A centralized architecture remains compatible with Bio-CPC when the integrator's estimate depends causally on information originating in multiple lower-level units and when that estimate contributes to the coordination of heterogeneous response mappings. By contrast, a controller that possesses the relevant information independently of such multi-unit integration and simply issues commands to otherwise passive effectors would lack the collective-inference component of Bio-CPC.

Organismal individuality is also maintained by distributed regulatory systems beyond the nervous system. The *immune system* provides a particularly clear example. Individual immune cells have access only to locally available and often ambiguous information, including molecular signatures of pathogens, antigen presentation, cytokines released by other cells, and signals associated with tissue damage. Yet their interactions generate coordinated organism-level responses involving activation, suppression, migration, cytokine secretion, antibody production, and cytotoxicity. For example, experimental and modeling work shows that feedback involving antigen and interleukin-2 (IL-2) signaling can scale the collective cytokine output of T-cell populations to antigen abundance (Tkach et al., 2014). The immune response has accordingly been interpreted within an active-inference framework as an inferential process distinguishing such states as self from non-self and threatening from non-threatening conditions (Bhat et al., 2021). From a Bio-CPC perspective, the more specific problem is how incomplete information distributed among heterogeneous immune cells can be integrated into a sufficiently coherent estimate of an organism-level condition such as infection or tissue damage, and how that estimate is translated into mutually compatible response patterns. Although cytokine signaling provides a plausible biological substrate for this process, coordinated cytokine-dependent responses alone do not establish collective inference or interaction-dependent alignment of response mappings. Autoimmunity or excessive inflammation could potentially reflect inaccurate estimation of an immunologically relevant condition, inappropriate response mappings despite adequate estimation, or a combination of both. Distinguishing these possibilities would provide a potential application of the Bio-CPC framework.

Regeneration provides another candidate domain of Bio-CPC. Following injury, cells at different locations encounter distinct local signals, yet their coordinated proliferation, migration,

differentiation, and physiological responses can restore organism-level morphology (Levin, 2009; McLaughlin and Levin, 2018). A specifically Bio-CPC interpretation would require showing that interactions among these cells improve estimation of an independently measurable organism-level pattern variable and that this inferential contribution can be dissociated experimentally from downstream morphogenetic response mappings. Without such evidence, regeneration is more conservatively described as distributed regulation.

Cancer illustrates the converse situation—a breakdown of organism-level integration. Normal multicellularity depends on cooperation in proliferation control, regulated cell death, resource use, differentiation, and maintenance of the extracellular environment, while cancer involves recurrent violations of these cooperative constraints (Aktipis et al., 2015). In the present framework, cancer can therefore be interpreted not simply as increased prediction error but as a failure of alignment between local state-to-response mappings and higher-level organization. A cancer cell may possess adequate information about its local environment while nevertheless responding in ways that increase cell-lineage proliferation at the expense of organismal integrity.

This distinction reveals a feature of biological collective organization that is not captured by standard CPC models. MHNG and related formulations generally assume agents participating in a common inferential process, while biological collectives may contain components whose local dynamics become systematically incompatible with collective requirements. If a cell responds inappropriately because its estimate of a collectively relevant condition is inaccurate, restoring the availability or reliability of the intercellular signals that contribute to its local observations may in principle improve that estimate and restore coordination. If its state-to-response mapping itself has changed so that locally favored responses conflict with organism-level integrity, additional information alone will not suffice. Evolved mechanisms such as growth suppression, apoptosis, immune surveillance, and tissue compartmentalization can help maintain organism-level integration by constraining or eliminating non-cooperative components (Ågren et al., 2019). Their role cannot be reduced to providing those components with additional information.

From the present information-processing perspective, organismal individuality need not correspond to either purely distributed or purely centralized regulation. Rather, integrated organisms may combine multiple architectures operating on different spatial and temporal scales. Nervous systems can provide relatively rapid and centralized integration, while immune, regenerative, metabolic, and developmental processes may remain substantially distributed. The evolution of individuality can therefore involve shifts in the relative contribution of these architectures to collective inference rather than the replacement of one by another. Bio-CPC must accommodate this coexistence while retaining the distinction between collective state estimation, response-mapping alignment, and mechanisms that maintain alignment against noisy, unresponsive, or non-cooperative components.

Sociality

Social organization provides perhaps the closest biological analogue to the original formulation of CPC. Individual animals are relatively autonomous agents, each possessing its own sensory information, internal states, and behavioral repertoire, yet social interactions allow information distributed across individuals to influence group-level behavior. Animal groups can thereby exhibit forms of collective sensing, decision-making, and information processing that cannot be straightforwardly attributed to any single individual (Couzin, 2009). The central problem is therefore

similar to that addressed by CPC: how can multiple agents with only partial and locally available information collectively estimate relevant states of the world and coordinate their responses?

At evolutionary timescales, social behavior is shaped by its fitness consequences, including inclusive-fitness effects in kin-structured societies (Boomsma, 2022). These evolutionary criteria should be distinguished from the proximate environmental or social states estimated during collective behavior, such as resource location, predator threat, nest-site quality, or brood needs. Highly integrated societies can emerge when responses to such states are sufficiently cooperative and conflicts among individuals are sufficiently suppressed (Queller and Strassmann, 2009; West et al., 2015).

One potential advantage of social organization is that pooling information across individuals can support more accurate or robust estimates of environmental states than those available to individuals acting alone. In fish shoals, for example, quorum-like interactions can improve the accuracy and speed of collective decisions (Ward et al., 2008), and both decision speed and accuracy under predation threat can increase with group size through distributed vigilance and social information transfer (Ward et al., 2011). Such benefits are not automatic, however. Social information can also amplify errors or produce maladaptive information cascades when individuals rely excessively on correlated or unreliable information. The benefits of collective estimation therefore depend on the structure and dynamics of interactions, rather than following automatically from group living.

The second Bio-CPC component, response-mapping alignment, provides a direct connection to the codebook problem in CPC. Social organization depends not only on information exchange but also on compatible, not necessarily identical, behavioral mappings through which environmental and social signals influence response patterns in different individuals. Such mappings partition the world into behaviorally relevant states; some differences are ignored, while others lead to distinct responses. Quorum rules in fish, recruitment signals in social insects, and socially learned associations provide examples of mappings through which information acquired by one individual can systematically affect the responses of others (Ward et al., 2008; Gordon, 2016). From a Bio-CPC perspective, the emergence or stabilization of such mappings can be interpreted as response-mapping alignment: a coordination of how socially relevant distinctions are translated into behavior across agents. Division of labor in social insects likewise illustrates how different response tendencies can contribute to the same colony-level outcome (Oster and Wilson, 1978; Gordon, 2016).

Eusocial insect colonies provide especially informative cases of collective regulation. Colonies achieve complex regulation of foraging, brood care, defense, and other collective activities typically through distributed interactions rather than centralized control, and task allocation can change dynamically with colony and environmental conditions (Oster and Wilson, 1978; Gordon, 2016). Some eusocial lineages have undergone transitions to *superorganismality* involving irreversible reproductive division of labor and differentiated castes (Boomsma and Gawne, 2018). This is more specific than the broader notion of *organismality* based on high cooperation and relatively low internal conflict (Queller and Strassmann, 2009). In the Bio-CPC framework, such societies provide an important case in which individual state-to-response mappings become strongly embedded within a higher-order system of collective regulation.

The honeybee waggle dance provides an especially close connection to original CPC because it involves an explicit communication code. Features of the waggle run encode the direction and distance of a resource and enable followers to locate the advertised destination (Dyer, 2002).

Disrupting directional information in the dance can reduce recruitment efficiency and, under some ecological conditions, colony-level food acquisition (Sherman and Visscher, 2002). Moreover, the code is not entirely independent of social experience; bees deprived of opportunities to follow dances before performing their own first dances show larger directional errors and abnormal distance encoding, demonstrating a role for social learning in the accurate production of the signal (Dong et al., 2023). The waggle dance therefore demonstrates socially calibrated signal production within an evolutionarily structured communication code. This makes it a particularly promising candidate for Bio-CPC, but does not by itself establish the proposed coupling between collective estimation and response-mapping alignment; a stronger test would ask whether social interaction both improves collective estimation of the relevant environmental state and calibrates the mappings through which that estimate is translated into individual and colony-level responses.

Behavioral ecology and theories of collective behavior already provide sophisticated accounts of social information use, quorum responses, collective decision-making, cooperation, and division of labor. Bio-CPC would therefore add little if it merely renamed these phenomena as prediction-error minimization. Its potential advantage is more specific: Bio-CPC offers a common generative framework relating individual inference, communication, collective state estimation, and the alignment of state-to-response mappings. A strong empirical application should consequently demonstrate not only that a group performs better than isolated individuals, but also how distributed information gives rise to a collective latent state and how interaction establishes, calibrates, or maintains compatibility among the mappings through which that state is translated into individual responses. Sociality is therefore the level of biological organization at which the connection between collective inference and the original problem of codebook emergence becomes most explicit.

Evolutionary and developmental implications

Bio-CPC also reframes the relationship between genetic specification and developmental organization. Inherited architecture can specify receptors, signaling pathways, response repertoires, and capacities for plasticity without specifying the final higher-order configuration element by element. Developmental interactions can then determine which of these response capacities are recruited and how they are coordinated in context. In this respect, Bio-CPC is consistent with accounts emphasizing the flexible deployment of conserved developmental regulatory mechanisms, but places particular emphasis on the collective calibration of locally specified response capacities (Gerhart and Kirschner, 2007; Davidson and Erwin, 2010).

The developmental response-mapping architecture is therefore neither purely genetically specified nor purely emergent. Inherited regulatory architectures constrain the repertoire of possible state-to-response mappings, while developmental interactions determine how these mappings are deployed and coordinated in particular contexts. The explanatory focus consequently shifts from asking only which genes specify a phenotype to asking how genetically inherited response capacities are collectively activated and coordinated during development. This perspective is especially relevant to complex biological organization, where the global outcome need not be explicitly represented in any individual component.

Such developmental flexibility may also provide a route for evolutionary stabilization. A novel phenotype induced by environmental change or mutation may initially be accommodated through developmental or behavioral plasticity, including learning where applicable. Subsequent selection on heritable variation in the underlying regulatory architecture may modify the expression or

effects of that phenotype. Such evolution constitutes *genetic accommodation*; *genetic assimilation* is the special case in which an initially environmentally induced phenotype becomes expressed without the inducing condition (Waddington, 1953; West-Eberhard, 2003; Moczek et al., 2011). From a Bio-CPC perspective, evolution therefore need not specify an entire collective solution in advance; it can modify the local mappings and communication mechanisms through which coordinated organization is repeatedly reconstructed.

The broader implication is therefore more specific than the claim that development contributes to phenotype construction. Bio-CPC distinguishes two separable functional targets of evolutionary stabilization: the reliability with which information originating in multiple units is integrated into estimates of collectively relevant states, and the reliability with which those estimates are translated into compatible heterogeneous responses.

The first target yields a comparative evolutionary prediction. As higher-order regulation becomes increasingly dependent on information that is unavailable to individual constituent units but recoverable through multi-unit integration, natural selection should favor communication and integration architectures that improve the fidelity, range, speed, specificity, or robustness with which such information is transmitted and combined. Across phylogenetically related systems, greater dependence on genuinely multi-unit information should therefore be associated, other things being equal, with stronger integrative coupling, more elaborate communication channels, or greater specialization of integrative structures. This prediction concerns informational dependence rather than organizational complexity per se: where higher-order regulation can be maintained without increased reliance on multi-unit information, Bio-CPC does not predict a corresponding elaboration or centralization of integrative architecture. Evolution may therefore modify not only the mappings through which units respond to m_C, but also the signaling channels and network topology through which m_C is collectively realized. Increasingly specialized integrative structures provide one possible, rather than necessary, outcome of this process, with the evolution of nervous systems representing an extreme case.

The second target concerns the stabilization and differentiation of unit-specific response mappings, allowing distinct local responses to remain compatible with higher-order organization. Evolution can thus stabilize both the architecture through which collective estimates are constructed and the mappings through which those estimates become causally effective, without genetically specifying the global outcome element by element.

Operationalization and testable predictions

A Bio-CPC hypothesis should be evaluated against explicit models of the same system. Relevant alternatives include parallel responses to a common input without inter-unit information exchange, local feedback-control models without multi-unit state estimation, and collective-estimation models in which response compatibility does not depend on inter-unit interaction. Feedback regulation alone does not establish the coupling of collective inference and interaction-dependent response-mapping alignment proposed here. The degree of centralization and the locus of integration are empirical properties of how collective inference is implemented. A candidate collectively relevant state s_C should first be defined and measured independently of the local observations o_i used to fit the Bio-CPC model. The model then yields a posterior over the collectively informed latent variable m_C. The empirical relation between the two can be evaluated by fitting a readout model $p_\omega(s_C \mid m_C)$ on one subset of observations and testing it on held-out data. This collective readout should be compared

with corresponding local readouts $p(s_C \mid o_i)$ or $p(s_C \mid z_i)$. The single-unit benchmark should exclude information contributed by other units through communication; an estimate already incorporating such information is not based on a single source alone. Evidence for collective inference would require that m_C predicts s_C more accurately, precisely, or robustly than information available to individual units alone, and that perturbing inter-unit communication degrades this relation while preserving local sensing as far as possible. Importantly, successful decoding by an external observer is not by itself sufficient to establish Bio-CPC. The latent variable must be implemented within the biological system and causally connected to downstream responses.

The spatial organization of information integration should therefore be characterized independently of whether collective inference occurs. Perturbations of candidate integrative hubs, together with perturbations of distributed inter-unit coupling, can test whether the collective estimate is realized broadly across the system or concentrated in a specialized subsystem. If disruption of a central integrator abolishes the collective estimate despite preserved local sensing, this would indicate a relatively centralized realization of Bio-CPC rather than falsify the hypothesis. Bio-CPC would instead be weakened if the putative integrator could generate an equivalent estimate from information available to it independently of the other constituent units. Where an explicit generative model is available, model-based or information-theoretic measures can quantify the contribution of multi-unit information to the collective estimate, while network or causal-intervention measures can separately quantify the spatial centralization of its implementation.

Response-mapping alignment should be assessed separately. The criterion of compatibility should be defined before, and independently of, fitting the unit-specific response mappings, so that successful collective performance is not used post hoc to define alignment. Where goal-directed regulation is evaluated, this requires specifying s_C^* and the discrepancy $d(s_C, s_C^*)$ independently of the unit-specific response mappings. Evidence for alignment should then show that the joint responses tend to reduce, or maintain a low value of, this deviation. Perturbations of state-to-response pathways should preserve information about s_C in m_C while disrupting the compatibility of unit-specific responses and thereby impairing the collective-level outcome. To establish the Bio=CPC-specific alignment process, rather than mere pre-existing compatibility, one should additionally show that inter-unit interaction establishes, recalibrates, or maintains these mappings. For example, this could be demonstrated by showing that altered interaction history produces a corresponding change in estimate-to-response relations and in their cross-unit compatibility. The strongest evidence for Bio-CPC would therefore combine this interaction dependence with a double dissociation: communication perturbations impair collective estimation while sparing downstream response capacity, while response-mapping perturbations spare collective estimation while impairing coordinated correction. Such tests require sufficiently selective interventions and a temporal window in which primary effects can be distinguished from subsequent feedback. Conversely, the hypothesis would be weakened if individual units already have sufficient local information or if communication has no causal effect on the collective estimate. Failure to obtain this dissociation would be inconclusive unless a specified model predicts separable effects under the tested conditions. Testing the evolutionary prediction outlined above additionally requires comparisons across phylogenetically related systems to assess whether greater dependence on multi-unit information is associated with more elaborate communication and integration architectures. Demonstrating the proposed coupling within a single system would not by itself establish this evolutionary relationship.

Evidence for either component alone would not establish their proposed coupling. Evidence for that coupling would support the applicability of Bio-CPC, whereas comparative support over existing estimation-and-control approaches would require explicit model-level alternatives.

Discussion

Bio-CPC overlaps with theories of active inference, collective sensing, distributed control, collective intelligence (including swarm intelligence), and major transitions in evolution, but it is not intended as an umbrella term for these phenomena. Consensus and distributed-control frameworks have shown how local information exchange can generate coordinated dynamics or agreement without centralized control (Olfati-Saber et al., 2007), while distributed Bayesian filtering explicitly addresses how locally estimated probability distributions can be combined into a consensus estimate across networked agents (Bandyopadhyay and Chung, 2014). In biological systems, collective-sensing studies likewise show how interactions among individuals or cells can support estimates or responses that are unavailable or less reliable at the level of isolated units (Berdahl et al., 2013; Camley, 2018). Multi-agent active-inference frameworks provide an especially close comparison. Friston et al. (2024), for example, modeled federated inference through belief sharing among agents with complementary observations and further showed how the likelihood mappings linking agents' beliefs to communicative expressions can be acquired and emerge through interaction. Waade et al. (2025) likewise examined how interacting active-inference agents can constitute an emergent group-level agent with a generative model that is not trivially reducible to those of its constituent agents. These approaches therefore overlap substantially with the collective-estimation component of Bio-CPC and, in the case of federated inference, also with the problem of interaction-dependent mapping alignment.

Fujiwara and Naoki (2026) modeled anticipatory regulation of multiple physiological variables through reinforcement learning. A Bio-CPC interpretation of such regulation would additionally require evidence that information originating in multiple units contributes to a collectively relevant state estimate and that interaction establishes, calibrates, or maintains compatible mappings from that estimate to heterogeneous responses.

The distinctive claim of Bio-CPC is therefore not that distributed estimation, belief sharing, interaction-dependent mapping learning, or shared adaptive control is itself novel. Related inference and control frameworks already provide accounts of these processes. Bio-CPC instead isolates a different biological problem: how a collectively informed estimate becomes causally effective through heterogeneous unit-specific responses, and how interaction establishes or maintains the compatibility of those different estimate-to-response mappings. This differs from the alignment of beliefs or communicative likelihood mappings because the relevant outputs need not converge on a common representation or expression. Different units may instead produce systematically different physiological, developmental, or behavioral responses whose joint consequences support a higher-order organization. Bio-CPC therefore treats collective estimation and heterogeneous response-mapping alignment as distinct but coupled processes and predicts distinguishable perturbation effects where the processes can be selectively targeted.

These two processes can fail independently. A collective may fail because its estimate of the relevant state is inaccurate despite appropriate response mappings, or because an adequately estimated state is translated into responses incompatible with higher-level organization. The spatial organization of information integration is orthogonal to this distinction: a centralized integrator can instantiate Bio-CPC when its state is causally constructed from information originating in multiple constituent

units and contributes to the coordination of heterogeneous response mappings. Systems in which information is merely pooled by an external observer, components produce parallel responses to a common input without inter-unit information exchange, or a central controller acts on information available independently of multi-unit integration fall outside the framework.

This extension also yields an evolutionary hypothesis that is absent from the original formulation of CPC. As higher-order regulation becomes increasingly dependent on information that cannot be recovered by individual constituent units alone, natural selection should act not only on local response mappings but also on the communication and integration architectures through which such multi-unit information is combined. Higher-order individuality may therefore involve both a redistribution of inferential load from relatively autonomous constituent units toward collectively informed variables and an evolutionary elaboration of the channels and structures through which those variables are realized. The emergence and elaboration of nervous systems provide a candidate transition from predominantly decentralized collective inference toward increasingly centralized integration. Centralization is nevertheless scale-relative: a brain may function as a central integrator relative to peripheral cells and tissues while itself realizing inference through distributed interactions among neurons. An organism integrated in this way can, in turn, act as a relatively autonomous unit in a new decentralized Bio-CPC process at the social level. Bio-CPC is therefore best understood as a recursive framework for transitions in biological integration, rather than as a theory committed to decentralization at every organizational level.

The present formulation provides a criterion for evaluating response-mapping alignment relative to s^*_C, but leaves open how conflicts between collective and unit-specific regulatory objectives are managed. Where a constituent unit has an identifiable local preferred state, which we may denote by s^*_i, responses favoring regulation toward this state may compromise collective regulation, while collective regulation may impose local costs. Developing explicit models of how interactions and regulatory constraints accommodate these competing demands, including how priorities are established or modified when the objectives cannot be jointly satisfied, remains a task for future work.

The proposal developed here remains a hypothesis about hierarchical biological organization. Its value will therefore depend on whether CPC-inspired models generate predictions that go beyond existing accounts of developmental regulation, collective behavior, cooperation, and major transitions in evolution. A particularly important empirical challenge is to identify measurable collective states, quantify how information about those states is distributed among constituent units, and determine whether interaction improves their estimation and aligns the resulting response patterns. If successful, such an approach could provide a common information-processing perspective on how new levels of biological individuality and organization emerge from interactions among simpler goal-directed units.

Acknowledgements

We thank Dr. Mana Fujiwara for her valuable comments on the manuscript. The authors acknowledge the use of ChatGPT (GPT-5.6 Sol; OpenAI) to assist with drafting portions of the manuscript, restructuring the text, and language editing. All AI-assisted content was critically reviewed, revised, and validated by the authors, who take full responsibility for the scientific arguments, accuracy, citations, and final content of the manuscript.

Funding

This work was supported by JSPS KAKENHI (grant numbers JP24H01538 to DGS and JP26K02091 to SD) and by The Nippon Foundation HUMAI Program (grant number 26GS-02 2026-06 to RK and DGS).

References

Ågren JA, Davies NG, Foster KR (2019) Enforcement is central to the evolution of cooperation. Nature Ecology & Evolution 3: 1018–1029. doi:10.1038/s41559-019-0907-1.

Aktipis CA, Boddy AM, Jansen G, Hibner U, Hochberg ME, Maley CC, Wilkinson GS (2015) Cancer across the tree of life: cooperation and cheating in multicellularity. Philosophical Transactions of the Royal Society B 370: 20140219. doi:10.1098/rstb.2014.0219.

Bandyopadhyay S, Chung S-J (2014) Distributed estimation using Bayesian consensus filtering. Proceedings of the 2014 American Control Conference: 634–641. doi:10.1109/ACC.2014.6858896.

Berdahl A, Torney CJ, Ioannou CC, Faria JJ, Couzin ID (2013) Emergent sensing of complex environments by mobile animal groups. Science 339: 574–576. doi:10.1126/science.1225883.

Bhat A, Parr T, Ramstead M, Friston K (2021) Immunoceptive inference: why are psychiatric disorders and immune responses intertwined? Biology & Philosophy 36: 27. doi:10.1007/s10539-021-09801-6.

Boomsma JJ (2022) Domains and Major Transitions of Social Evolution. Oxford University Press, Oxford.

Boomsma JJ, Gawne R (2018) Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation. Biological Reviews 93: 28–54. doi:10.1111/brv.12330.

Bozdag GO, Tong K, Yunker PJ, Herron MD, Ratcliff WC (2026) Overcoming key challenges in the evolutionary transition to multicellularity. Nature 658: 327–336. doi:10.1038/s41586-026-11038-w.

Camazine S, Deneubourg JL, Franks NR, Sneyd J, Theraulaz G, Bonabeau E (2001) Self-Organization in Biological Systems. Princeton University Press, Princeton, NJ.

Camley BA (2018) Collective gradient sensing and chemotaxis: modeling and recent developments. Journal of Physics: Condensed Matter 30: 223001. doi:10.1088/1361-648X/aabd9f.

Couzin ID (2009) Collective cognition in animal groups. Trends in Cognitive Sciences 13: 36–43. doi:10.1016/j.tics.2008.10.002.

Davidson EH, Erwin DH (2010) Evolutionary innovation and stability in animal gene networks. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 314: 182–186. doi:10.1002/jez.b.21329.

Dong S, Lin T, Nieh JC, Tan K (2023) Social signal learning of the waggle dance in honey bees. Science 379: 1015–1018. doi:10.1126/science.ade1702.

Du Q, Kawabe Y, Schilde C, Chen Z-H, Schaap P (2015) The evolution of aggregative multicellularity and cell–cell communication in the Dictyostelia. Journal of Molecular Biology 427: 3722–3733. doi:10.1016/j.jmb.2015.08.008.

Dyer FC (2002) The biology of the dance language. Annual Review of Entomology 47: 917–949. doi:10.1146/annurev.ento.47.091201.145306.

Freddolino PL, Tavazoie S (2012) Beyond homeostasis: a predictive-dynamic framework for understanding cellular behavior. Annual Review of Cell and Developmental Biology 28: 363–384. doi:10.1146/annurev-cellbio-092910-154129.

Friston KJ (2010) The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11: 127–138. doi:10.1038/nrn2787.

Friston KJ (2013) Life as we know it. Journal of the Royal Society Interface 10: 20130475. doi:10.1098/rsif.2013.0475.

Friston KJ, Daunizeau J, Kiebel SJ (2009) Reinforcement learning or active inference? PLoS ONE 4: e6421. doi:10.1371/journal.pone.0006421.

Friston K, Levin M, Sengupta B, Pezzulo G (2015) Knowing one's place: a free-energy approach to pattern regulation. Journal of the Royal Society Interface 12: 20141383. doi:10.1098/rsif.2014.1383.

Friston KJ, Parr T, Heins C, Constant A, Friedman D, Isomura T, Fields C, Verbelen T, Ramstead M, Clippinger J, Frith CD (2024) Federated inference and belief sharing. Neuroscience & Biobehavioral Reviews 156: 105500. doi:10.1016/j.neubiorev.2023.105500.

Fujiwara M, Naoki H (2026) Gradual proactive regulation of body state by reinforcement learning of homeostasis. Neuroscience Research 223: 105021. doi:10.1016/j.neures.2026.105021.

Gerhart J, Kirschner M (2007) The theory of facilitated variation. Proceedings of the National Academy of Sciences USA 104: 8582–8589. doi:10.1073/pnas.0701035104.

Gordon DM (2016) From division of labor to the collective behavior of social insects. Behavioral Ecology and Sociobiology 70: 1101–1108. doi:10.1007/s00265-015-2045-3.

Grosberg RK, Strathmann RR (2007) The evolution of multicellularity: a minor major transition? Annual Review of Ecology, Evolution, and Systematics 38: 621–654. doi:10.1146/annurev.ecolsys.36.102403.114735.

Hayashi Y (2026) CPCの数理的理解：Shannon理論から集合的予測符号化へ [Mathematical understanding of collective predictive coding: from Shannon theory to CPC]. 日本ロボット学会誌 [Journal of the Robotics Society of Japan] 44: 451–455 (in Japanese). doi:10.7210/jrsj.44.451.

Inouye K (1989) Control of cell type proportions by a secreted factor in *Dictyostelium discoideum*. Development 107: 605–609. doi:10.1242/dev.107.3.605.

Inukai J, Taniguchi T, Taniguchi A, Hagiwara Y (2023) Recursive Metropolis–Hastings naming game: symbol emergence in a multi-agent system based on probabilistic generative models. Frontiers in Artificial Intelligence 6: 1229127. doi:10.3389/frai.2023.1229127.

Ispolatov I, Ackermann M, Doebeli M (2012) Division of labour and the evolution of multicellularity. Proceedings of the Royal Society B 279: 1768–1776. doi:10.1098/rspb.2011.1999.

Kin K, Schaap P (2021) Evolution of multicellular complexity in the dictyostelid social amoebas. Genes 12: 487. doi:10.3390/genes12040487.

Lamża Ł (2023) Diversity of 'simple' multicellular eukaryotes: 45 independent cases and six types of multicellularity. Biological Reviews 98: 2188–2209. doi:10.1111/brv.13001.

Levin M (2009) Bioelectric mechanisms in regeneration: unique aspects and future perspectives. Seminars in Cell & Developmental Biology 20: 543–556. doi:10.1016/j.semcdb.2009.04.013.

Maynard Smith J, Szathmáry E (1995) The Major Transitions in Evolution. W. H. Freeman/Spektrum, Oxford.

McLaughlin KA, Levin M (2018) Bioelectric signaling in regeneration: mechanisms of ionic controls of growth and form. Developmental Biology 433: 177–189. doi:10.1016/j.ydbio.2017.08.032.

Michod RE (2006) The group covariance effect and fitness trade-offs during evolutionary transitions in individuality. Proceedings of the National Academy of Sciences USA 103: 9113–9117. doi:10.1073/pnas.0601080103.

Moczek AP, Sultan S, Foster S, Ledón-Rettig C, Dworkin I, Nijhout HF, Abouheif E, Pfennig DW (2011) The role of developmental plasticity in evolutionary innovation. Proceedings of the Royal Society B 278: 2705–2713. doi:10.1098/rspb.2011.0971.

Nakamura K, Kobayashi TJ (2021) Connection between the bacterial chemotactic network and optimal filtering. Physical Review Letters 126: 128102. doi:10.1103/PhysRevLett.126.128102.

Nakamura K, Kobayashi TJ (2022) Optimal sensing and control of run-and-tumble chemotaxis. Physical Review Research 4: 013120. doi:10.1103/PhysRevResearch.4.013120.

Olfati-Saber R, Fax JA, Murray RM (2007) Consensus and cooperation in networked multi-agent systems. Proceedings of the IEEE 95: 215–233. doi:10.1109/JPROC.2006.887293.

Oster GF, Wilson EO (1978) Caste and Ecology in the Social Insects. Monographs in Population Biology 12. Princeton University Press, Princeton, NJ. doi:10.2307/j.ctvx5wb34.

Oyama M, Okamoto K, Takeuchi I (1983) Proportion regulation without pattern formation in *Dictyostelium discoideum*. Journal of Embryology and Experimental Morphology 75: 293–301.

Queller DC (2000) Relatedness and the fraternal major transitions. Philosophical Transactions of the Royal Society B 355: 1647–1655. doi:10.1098/rstb.2000.0727.

Queller DC, Strassmann JE (2009) Beyond society: the evolution of organismality. Philosophical Transactions of the Royal Society B 364: 3143–3155. doi:10.1098/rstb.2009.0095.

Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience 2: 79–87. doi:10.1038/4580.

Rokas A (2008) The origins of multicellularity and the early history of the genetic toolkit for animal development. Annual Review of Genetics 42: 235–251. doi:10.1146/annurev.genet.42.110807.091513.

Ruiz-Trillo I, Kin K, Casacuberta E (2023) The origin of metazoan multicellularity: a potential microbial black swan event. Annual Review of Microbiology 77: 499–516. doi:10.1146/annurev-micro-032421-120023.

Sebé-Pedrós A, Degnan BM, Ruiz-Trillo I (2017) The origin of Metazoa: a unicellular perspective. Nature Reviews Genetics 18: 498–512. doi:10.1038/nrg.2017.21.

Segall JE, Block SM, Berg HC (1986) Temporal comparisons in bacterial chemotaxis. Proceedings of the National Academy of Sciences USA 83: 8987–8991. doi:10.1073/pnas.83.23.8987.

Sherman G, Visscher PK (2002) Honeybee colonies achieve fitness through dancing. Nature 419: 920–922. doi:10.1038/nature01127.

Tagkopoulos I, Liu Y-C, Tavazoie S (2008) Predictive behavior within microbial genetic networks. Science 320: 1313–1317. doi:10.1126/science.1154456.

Taniguchi T (2024) Collective predictive coding hypothesis: symbol emergence as decentralized Bayesian inference. Frontiers in Robotics and AI 11: 1353870. doi:10.3389/frobt.2024.1353870.

Taniguchi T, Yoshida Y, Matsui Y, Hoang NL, Taniguchi A, Hagiwara Y (2023) Emergent communication through Metropolis–Hastings naming game with deep generative models. Advanced Robotics 37: 1266–1282. doi:10.1080/01691864.2023.2260856.

Tkach KE, Barik D, Voisinne G, Malandro N, Hathorn MM, Cotari JW, Vogel R, Merghoub T, Wolchok J, Krichevsky O, Altan-Bonnet G (2014) T cells translate individual, quantal activation into collective, analog cytokine responses via time-integrated feedbacks. eLife 3: e01944. doi:10.7554/eLife.01944.

Tschantz A, Seth AK, Buckley CL (2020) Learning action-oriented models through active inference. PLOS Computational Biology 16(4): e1007805. doi:10.1371/journal.pcbi.1007805.

Waade PT, Olesen CL, Laursen JE, Nehrer SW, Heins C, Friston K, Mathys C (2025) As one and many: relating individual and emergent group-level generative models in active inference. Entropy 27: 143. doi:10.3390/e27020143.

Waddington CH (1953) Genetic assimilation of an acquired character. Evolution 7: 118–126. doi:10.1111/j.1558-5646.1953.tb00070.x.

Ward AJW, Sumpter DJT, Couzin ID, Hart PJB, Krause J (2008) Quorum decision-making facilitates information transfer in fish shoals. Proceedings of the National Academy of Sciences USA 105: 6948–6953. doi:10.1073/pnas.0710344105.

Ward AJW, Herbert-Read JE, Sumpter DJT, Krause J (2011) Fast and accurate decisions through collective vigilance in fish shoals. Proceedings of the National Academy of Sciences USA 108: 2312–2315. doi:10.1073/pnas.1007102108.

West SA, Fisher RM, Gardner A, Kiers ET (2015) Major evolutionary transitions in individuality. Proceedings of the National Academy of Sciences USA 112: 10112–10119. doi:10.1073/pnas.1421402112.

West-Eberhard MJ (2003) Developmental Plasticity and Evolution. Oxford University Press, Oxford.