Between Individual Brains and Collective Behavior: Multi-level Emergence in a Group Formation Task

Ekaterina Sangati[1], Federico Sangati[2], Yi-Shan Cheng[1] and Acer Yu-Chan Chang[3]

[1]Neural Computation Unit, Okinawa Institute of Science and Technology, Onna, Okinawa 904-0495 Japan
[2]Cognitive Neurorobotics Research Unit, Okinawa Institute of Science and Technology, Onna, Okinawa 904-0495 Japan
[3]Department of Psychology, Rikkyo University, Saitama 352-8558, Japan
kat.sangati@gmail.com

Abstract

Emergence is a property often claimed to apply to complex systems on multiple levels of organization: individual behavior emerges from underlying neural activity and social patterns – from constituent behaviors of the individuals. Furthermore, the emergent level is typically characterized as possessing autonomy from the lower-level phenomena and as exerting downward causation on them. In this study, we investigate such a multi-level emergence in the context of a single simple task. We evolve agents controlled by a small neural network to travel in formation. We then compute measures of emergence stemming from an approach known as Integrated Information Decomposition. Results are presented for both the final behavior and the evolutionary changes that led to it.

Introduction

Social behaviors are often characterized as emergent collective phenomena, where individual actions give rise to complex group-level patterns that cannot be predicted solely by examining the actions of individual members. Examples of such behavior can be seen in the synchronized movements of bird flocks, intricate functioning of ant colonies or seemingly sudden appearances of traffic jams. In these systems, the group behavior is said to be an *emergent* property with its own higher-level dynamics, not reducible to and *causally decoupled* from the properties and processes of the individual components. Moreover, it is also considered to affect individual members in a process called *downward causation*. Consider a football match: each player independently perceives the ball's location and other players, then selects and executes actions to help win the game (the middle level in Figure 1), which together constitute team movements on the field (top level). However, emergent passing patterns and collective coordination, along with each player's role, ultimately determine the unfolding of the game (top level's causal decoupling). This, in turn, affects each player's perception and action and their neural activity (downward causation), thereby completing the inter-level cycle of influence.

The study of emergent collective behaviors often involves attempts to uncover simple local "rules" whose interaction will produce an observed global pattern. Famously, for ex-
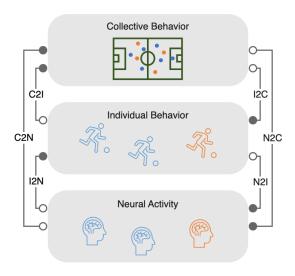

Figure 1: Levels of organization and interactions between them. Note that interactions are not necessarily causal but can include other types of relationships (constitution, constraint, realization, statistical dependence etc.). The filled compared to empty circles indicate the relative importance given to the respective levels in analysis.

ample, Reynolds (1987) showed that the flocking behavior can be explained by assuming that each individual is following three rules: maintain a safe distance from your neighbors, move in the same direction as your neighbors, and move towards the center of the group. Implementing these rules in a computational agent-based model produces a realistic-looking flocking behavior. However, the model does not claim that the real-life individual birds are actually explicitly following these rules nor does it explain what exact individual cognitive mechanisms underlie behavior that looks as if it was produced by them (cf., work by Couzin et al. 2003; Couzin 2009). That is, such a model is missing an explication of the bottom level of Figure 1.

What we propose in this paper is that a full understanding of collective behavior requires an investigation into the complex interplay between individual neural activity, individ-

ual behavior, and group dynamics. As can be intuited from the schematic connections in Figure 1, this investigation can take two explanatory approaches with respect to the relative focus on different levels. On the one hand, we could set up an inquiry that takes lower-level components and processes as a primary target of analysis, such as looking at how cognitive or neural mechanisms underlie individual behaviors (N2I) that then constitute collective patterns (I2C)[1], consistent with a mechanistic approach to explanation in cognitive science (Bechtel and Abrahamsen, 2005). For example, understanding ant foraging might involve the study of their individual capacities to produce and sense pheromones (Steck, 2012) and navigate to the nest (Wehner et al., 2006), and how these capacities in multiple ants contribute to collective food search behavior (Feinerman and Korman, 2017).

In this paper, we focus on the complementary focus of analysis, one that emphasizes the higher compared to the lower level. Such an approach consists in looking at whether and how emergent higher-level patterns acquire a certain degree of autonomy from the lower level and then constrain individual behaviors (C2I) or their underlying cognitive and neural activity - directly (C2N) or indirectly through behavior (I2N). In the ant case, for example, selection pressures on the level of the colony can be analyzed independently of particular genetic and morphological specialization of the individuals. However, they can also be seen as contributing to such changes (Campbell, 1990), including, for instance, a reduction of brain complexity (Traniello et al., 2022).

The importance of selection forces in collective phenomena highlights the need to explicitly consider the relationship between what is being optimized and how behavioral performance relates to emergent effects across different levels of organization. In other words, is it the case, for instance, that putting a selection pressure on collective task *performance* leads to a greater independence of global patterns (i.e., higher estimated degree of *emergence* and *causal decoupling*) and their stronger constraint over individual activity (*downward causation*)? Furthermore, do any such effects apply only across "contiguous" levels, e.g., from collective behavior to individual behavior? Or is there a more direct shaping of individual cognitive and neural activity by the collective patterns? The aim of this paper is to start answering these questions with an example scenario, which is easy to artificially evolve and analyze on all the levels of organization discussed so far and across all the possible interactions between them.

Specifically, we evolve a group of three agents to solve a well-studied multi-agent formation task. We adopt a very

liberal understanding of formation: the agents are required to move together while staying cohesive within a specified radius. Each agent can sense the distance to other agents in their vicinity (within a certain visual field) but has no understanding of a global coordinate system or every agent position in it (cf. distance-based formation control in Oh et al. 2015). Moreover, no specific formation configuration or common movement heading is imposed. We then analyze emergence on three scales:

- **C2I**: the relationship between a formation center of mass (CoM) position and positions of individual agents[2]

- **I2N**: the relationship between individual position of each agent and its brain activity

- **C2N**: the relationship between a formation CoM position and each of the individual agents' brain activity

In order to quantify emergent effects, we utilize a recently developed causal emergence framework proposed by Rosas et al. (2020) and apply ΦID or Integrated Information Decomposition (Mediano et al., 2019, 2022c) to our analysis. This framework is based on information theory approach to causality (Granger, 1969; Pearl, 1995), recent advances in Partial Information Decomposition (Williams and Beer, 2010) and a weak interpretation of Integrated Information Theory (Oizumi et al., 2014; Mediano et al., 2022b). Specifically, it views emergence as "the capability of some supervenient [higher level] feature to provide predictive power that cannot be reduced to underlying [partial] microscale phenomena" (Mediano et al., 2022c, p.5). That is, whereas in a strong view on emergence, a macro-level phenomenon can be said to be impossible to derive from micro-level phenomena even with perfect knowledge of all the micro details, in a weaker view (1) such derivation is non-trivial and (2) macro-level patterns capture regularities beyond idiosyncratic micro-level details – both of which to the extent that a macro-level can be seen as irreducible in an explanatory sense (Bedau, 1997, 2002). Adopting an information theoretic approach to emergence allows one to formulate specific quantities for *causal emergence*, *downward causation*, and *causal decoupling*. While one can expect that these quantities would be applicable to observations generated by strongly emergent phenomena (should such phenomena exist), the ΦID approach presents a weak, epistemic stance on emergence, which is also the stance we adopt in the current paper (cf. Dewhurst, 2021; Hulswit, 2005; Hoel et al., 2013).

Related work

Previous research Rosas et al. (2020) has shown that ΦID effectively identifies relevant features of complex systems

---

[1]Another possibility is that neural activity might have direct effects on the collective level (N2C), as explored in recent studies of inter-brain synchrony which correlates with better social coordination (Czeszumski et al., 2020). However, the precise nature of these effects, for instance, their causal contribution to this phenomenon, is still controversial (Holroyd, 2022).

[2]Note again, that these labels are meant to indicate our interest in emergent properties of the higher level, without implying direct causal influence from higher to lower levels.

associated with emergent phenomena. For instance, ΦID applied to a model of flocking behavior, revealed that the flock's center of mass is a better predictor of its temporal changes than individual bird movements. However, this holds true only when optimal model parameters are chosen to simulate flocking behavior instead of dispersion or regular trajectories. Similarly, in an experiment involving macaques' reaching movements, ΦID highlighted that the neural activity most predictive of reaching motion is an emergent property of the underlying neural activity, raising questions about the brain-behavior relationship.

The novelty of our study with respect to this work is in 1) analyzing both types of emergence (C2I and I2N) in a single behavioral context, 2) exploring a third possible inter-level effect (C2N) and 3) investigating the changes in emergence over the evolution of target behavior, that is, in relation to behavioral fitness.

The last point brings us to another body of related work: examination of changes in integrated information over the course of evolution of cognitively interesting behavior. For example, Yaeger (2009) investigated the evolution of agents in a simple ecosystem and computed their neural complexity according to an early version of integrated information measure known as TSE complexity (Tononi et al., 1994). This measure is said to reflect a balance between functional integration and separation in a system or what makes it "more than the sum of its parts" and is proposed to be a correlate of consciousness but can also be viewed as a measure of dynamical complexity (Mediano et al., 2022a). It was shown that subjecting the agents to natural selection operating at an individual level led to a rapid increase in neural complexity, its stabilization at a "good enough" level and its correlation with behavioral adaptation.

While we use a measure related to TSE complexity and similarly follow an evolutionary approach, studies such as this focus on information dynamics within agents' neural systems and the emergence relation between neural units and a neural system as a whole. In this sense, they are confined to the lowest level depicted in Figure 1 while we aim to capture emergence between this and higher levels.

Research questions

The aim of this study is to investigate whether imposing a collective-level fitness function on a multi-agent system results in observable emergent behaviors, downward causation, and causal decoupling across different levels of analysis. Specifically, we propose three research questions:

1. How is the collective level of adaptive behavior related to emergence, downward causation, and causal decoupling at the three inter-level comparisons C2I, I2N, and C2N?

2. What evolutionary trends are exhibited by emergence, downward causation, and causal decoupling?

3. Given what is found in analysis (1) and (2), what are the implications for understanding complex social systems?

These questions will be answered in the context of the evolution of the formation task as described next.

Methods

Formation task

In the simulation, the three agents (circles of 10 units in radius) are set in an unbounded $2D$ environment. They undergo four trials, where in each trial, the initial angle of the heading directions of the agents is initialized randomly, whereas their positions are fixed. Specifically, the agents are placed on a circumference of radius 50 centered on origin at angles $0°$, $120°$, $240°$ respectively. Each trial is set to last 500 steps. The **performance** score of the agents for each trial is computed as the *displacement D* which is defined as the sum of the linear displacements *away* from the starting point of the agents' center of mass (CoM) when agents are *in-formation*: when they are at most 100 units away from the center of mass. More formally, this is defined as:

$$D(t) = \begin{cases} D(t-1) + \Delta r(t), & \begin{array}{l} \text{if "moving away"} \\ + \text{"in-formation"} \\ + \text{"not already covered"}^3 \end{array} \\ \\ 0, & \text{otherwise} \end{cases}$$

(1)

where $\Delta r(t)$ is the radial incremental distance of the CoM between two consecutive time steps. In simple words, agents are rewarded when they move in-formation as far away as possible from the initial position. To better clarify this, Figure 2 shows three different plots of agents' positions at subsequent stages of the evolutionary process. In these plots, D is the sum of the lengths of the green segments.[4]

Agent architecture

The network controlling the behavior of each agent is shown in Figure 3. Each of the 4 *sensors s* can perceive other agents in its 90-degree quadrant, relative to its body axis (based on its heading direction). At each time step, s receives a

---

[3]The idea here is that segments in which the CoM moves away from the starting point should not be counted twice (or more times) if the specific radial range of the displacement has already been "covered" in the previous movement history. This is to discourage agents from moving back and forth (e.g., making circles).

[4]Note that in the center frame, the green line (displacement) goes beyond the end point of the CoM at the agents' end positions (indicated as ●), because after reaching the further distance away from the initial position, the agents go backward. Also, there is a gap in the displacement line, because in that segment, the CoM was traveling for a small part away from the initial position but not-in-formation and later on in-formation but backward. Readers may benefit from viewing the three videos of the agents' simulation at `https://gitlab.com/kercos/formation`
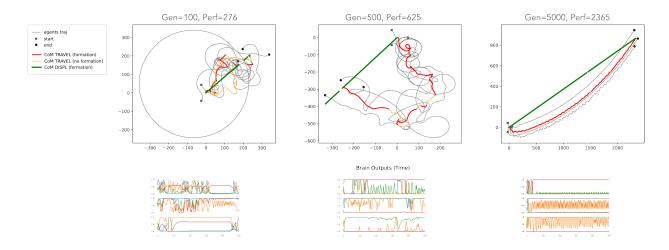
Figure 2: Traces of agents' path (in grey) in one example simulation run (run 19, trial 1) along with the trace of their center of mass (CoM) when agents are in-formation (thick red line) and not-in-formation (thin orange line). The green line represents the overall traveled displacement of CoM during formation. Bottom panel shows output of the 4 internal neurons of each agent.

signal $I_s \in [0,1]$ in input. The signal I_s decreases linearly in proportion to the distance to the closest agent: 1 if the agent is adjacent (2 body radius away, equivalent to 20 units), and 0 if it is beyond the maximum perceived distance (which is set to 200 units) or absent. That is, only one agent at a time can be perceived within each quadrant. The sensor output $O_s^{\mathcal{S}}$ is defined as:

$$O_s^{\mathcal{S}} = G_s\, \sigma(I_s + \theta_s) \qquad (2)$$

where G_s and θ_s are the *sensory gain* and *bias*, while σ is the standard activation sigmoid function $1/(1 + e^{-x})$.

The sensor's output is propagated to each *neural node* n_i in layer $\mathcal{N}$ (depicted in blue in Figure 3), which consists of 4 nodes. Each node in this layer is a continuous-time recurrent neural network (CTRNN) as described by Beer (1995).

All neural nodes are fully connected, both to all other neurons and to themselves. The delta of the neuron state Δy_n between two consecutive time steps is calculated via the Euler method integration of the differential equation governing the state change, as follows:

$$\Delta y_{n_i} = \frac{\Delta t}{\tau_{n_i}} \left(-y_{n_i} + W_{s,n_i} O_s^{\mathcal{S}} + \sum_{j=1}^{|\mathcal{N}|} W_{n_j,n_i} O_{n_j}^{\mathcal{N}} \right) \quad (3)$$

Here, τ_n is the *neural time constant*, Δt is the *step size constant* for the integration (set to 0.1), and $O_n^{\mathcal{N}}$ is the output of neuron $n \in \mathcal{N}$ calculated as:

$$O_n^{\mathcal{N}} = G_n\, \sigma(y_n + \theta_n) \qquad (4)$$

with G_n and θ_n being the *neural gain* and *bias*.

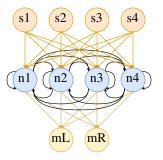Next, the output of each motor $m \in \mathcal{M}$ is calculated as:

Figure 3: Architecture of agent's network in the implementation. The entire network consists of three layers: the sensory layer (top) with 4 nodes, the neural layer (middle) with 4 nodes, and the motor layer (bottom) with 2 nodes.

$$O_m^{\mathcal{M}} = G_m\, \sigma \left(\sum_{n=1}^{|\mathcal{N}|} W_{n,m} O_n^{\mathcal{N}} + \theta_m \right) \qquad (5)$$

where G_m and θ_m are the *motor gain* and *bias*.

The *displacement* of the agent Δx at every step is computed as the difference between the two motors:

$$\Delta x = O_R^{\mathcal{M}} - O_L^{\mathcal{M}} \qquad (6)$$

Finally, the *angular variation* $\Delta \alpha$ is calculated as the ratio of displacement by the body radius r (10 units):

$$\Delta \alpha = \Delta x / r \qquad (7)$$

When agents collide, they are not able to move (in terms of displacement), but they rotate in place according to the angular variation.

Evolutionary Algorithm

Three distinct[5] populations of 48 agents each, are evolved independently in 20 separate evolution runs (each initialized with a unique random seed).

In the first generation, all agent genotypes are set to arrays of zeros and agents are combined into triplets based on their (still unranked) position in respective populations. They perform four task trials and their overall fitness is computed as the mean of their trial performances. The same fitness score is assigned to each member of the triplet. In each subsequent generation, agents are paired for the task based on the same rank in the three populations. That is, the n-th best agent of the first population is grouped with the n-th best agent of the second and third population. This means that the triplet of agents is kept together across all generations, even though they may undergo independent mutations, change their behavior in the simulation and change their resulting rank in the population across generations.

All network parameters (weights, gains, biases, taus)[6] are evolved for $5,000$ *generations*, using a real-valued genetic algorithm in the fixed range $[-1, 1]$. At each generation, the network parameters encoded in each agent's evolved genotype are converted to the correct ranges via linear interpolation.[7]

New generations are created by keeping an elite population of the top 5% of the existing solutions and by mutating (with a 5% variance) and crossing over the remaining 95% of the individuals to get the rest of the new solutions. Mutation consists in adding zero-mean Gaussian mutation noise with a variance of 0.05 to the solutions and crossover consists in swapping each parameter between a pair of solutions with a probability of 0.1.

Measures of causal emergence, downward causation, and causal decoupling

In this research, we employ the causal emergence framework by Rosas et al. (2020) to quantify emergence. This framework characterizes emergent behavior as macro-level information synergy, detailing quantitative measures for causal emergence, downward causation, and causal decoupling. To ensure computational feasibility, we use Rosas et al. (2020) "practical criteria", providing sufficient conditions for the aforementioned concepts and allowing application to large systems without extensive calculations. This is crucial as the computational intensity of partial information decomposition, a fundamental method in this framework, increases super-exponentially with system size, rendering traditional computation impractical.

Causal emergence

To compute the practical measure of causal emergence, Rosas et al. (2020) define the quantity $\Psi_{t,t'}^{(1)}(V)$ as:

$$\Psi_{t,t'}^{(1)}(V) := I(V_t; V_{t'}) - \sum_j I(X_t^j; V_{t'}) \qquad (8)$$

Here, j is the index of the micro variables X. The first part on the right-hand side is the mutual information between two time steps t and t' of a macro feature V, and the second part is the sum of the mutual information between each micro variable X^j at time t and the macro feature V at next time step t'. The value of Ψ greater than 0 is sufficient to indicate that there is synergistic information present in macro dynamics that cannot be explained by any individual micro variable, suggesting causal emergence. It is worth noting that the superscript (1) in $\Psi_{t,t'}^{(1)}(V)$ indicates that we consider only individual variables and not variable sets of size greater than one.

Downward causation

Rosas et al. (2020) define the second quantity $\Delta_{t,t'}^{(1)}(V)$ as:

$$\Delta_{t,t'}^{(1)}(V) := max_j(I(V_t; X_{t'}^j) - \sum_i I(X_t^i; X_{t'}^j)) \qquad (9)$$

Downward causation can be sufficiently claimed when $\Delta^{(1)}t, t'(V) > 0$. If a macro variable V_t, holds more information about the future state of at least one micro variable $X_{t'}^j$, than the sum of the information held by current states of all individual micro variables X_t^i, then the system can be said to exhibit downward causation. This is because the information content of the macro variable regarding the future state of the individual micro variable exceeds the combined information of the present states of all micro variables.

Causal decoupling

For causal decoupling Rosas et al. (2020) define the third quantity $\Gamma_{t,t'}^{(1)}(V)$ as:

$$\Gamma_{t,t'}^{(1)}(V) := max_j I(V_t; X_{t'}^j) \qquad (10)$$

Where $\Psi_{t,t'}^{(1)}(V) > 0$ and $\Gamma_{t,t'}^{(1)}(V) = 0$ are the sufficient criteria for causal decoupling. This criterion indicates that when causal emergence is present ($\Psi_{t,t'}^{(1)}(V) > 0$) and the macro variable has no predictive power on the future state of every individual micro variable ($\Gamma_{t,t'}^{(1)}(V) = 0$), the macro process can be considered "decoupled" from its micro processes.

Data analysis

For every measure listed above, we computed three quantities corresponding to our target inter-level comparisons. That is, we computed Ψ_{C2I}, Ψ_{C2N}, Ψ_{I2N} for causal emergence and analogous quantities for Δ and Γ. The macro and

micro variables were entered into computation[8] as presented in Table 1.

Comparison	V	X
C2I	d(CoM)	d(Agents)
C2N	d(CoM)	Agents brain output
I2N	d(Agents)	Agents brain output

Table 1: Variables used for emergence calculation. d(.) represents Euclidean distance from the origin (start point), so that d(Agents) is a triplet of distances. *CoM* stands for center of mass.

In the context of our study, we selected d(CoM) as a measure of collective behavior because it directly reflects the collective performance. We chose to focus on d(Agents) as a simplified and computationally tractable representation of individual behavior of agents. This choice was made due to the direct relationship between d(Agents) and the collective performance measure d(CoM). It is important to emphasize that d(CoM) is a function of d(Agents), and therefore d(Agents) can be viewed as an individual agent's contribution towards the collective goal. In this framework, the output of the neurons plays a crucial role. The displacement of the agent at every step, which is a key component of d(Agents), is computed as the difference between the two motors. These motors are directly influenced by the output of the neurons. Thus, the output of the neurons collectively influences the agent's behavior, and by extension, the collective behavior represented by d(CoM). This highlights the importance of the 'Agents brain output' at the neuron level in our study, as it provides a bridge between individual neural activity and collective performance.

Note that there are infinite ways to define relevant micro and macro variables. For instance, regarding the states of the individual agents, we could consider a wide array of parameters including full pose, velocity, acceleration, and many more. However, the decision to use d(Agents) as the representative measure of agent states in this study was motivated by the balance between the complexity of the model and its interpretability.

Each quantity was computed separately for every trial, using the time series of a complete trial run (500 time steps). Next, the values were averaged across trials and across agents to obtain a single value per evolution run and per generation.

In order to calculate information-theoretical measures, we utilized the code provided by Rosas et al. (2020). This code

was adapted to our specific needs. Note that the implementation of mutual information calculation in this code assumes that the variables being analyzed are Gaussian. This assumption was made in order to simplify the computation of mutual information. While this assumption may not always hold in real-world scenarios, it allowed us to obtain accurate and efficient results for our specific analysis.

Results

Emergence in evolved solutions

First we analyzed the three ΦID measures in the best performing group of agents of the final generation of all the evolution runs (Figure 4a-4c). We found positive values for Ψ_{C2N} and Ψ_{I2N}, suggesting that behavior of the group (CoM displacement) and behavior of the individual agents (agent displacement) is an emergent property of the agents' individual neural activity. We also found that for these measures, Γ is close to 0, which together with positive Ψ suggests the existence of causal decoupling of the higher-level. At the same time, surprisingly, we found no evidence for either emergence or decoupling of the collective level from the individual behavior, as indicated by negative Ψ_{C2I} and positive Γ_{C2I}. This result is interesting in that it shows that emergence of the social collective level from the individual neural activity can be detected independently of its emergence from individual behavior.

Having seen that Ψ_{C2N} and Ψ_{I2N} values are very similar we wondered whether there is nonetheless a difference between them in the final generation. The paired t-test showed that Ψ_{I2N} was smaller than Ψ_{C2N} and the difference was large and statistically significant ($t(19) = -5.72, p < .001$; Cohen's $d = -1.28$, see also Figure 4d). This result is consistent with an intuition that the higher one goes in the hierarchy of the system's organization, the more emergence from the underlying neural level can be found.

Finally, we saw no evidence of downward causation at any of the levels considered, as indicated by negative Δ values. This confirms that the measures distinguished by the ΦID framework are independent components of the overall concept of emergence.

Considering behavioral fitness, we found that all three varieties of Ψ are highly (and significantly, all $p < .001$) correlated with performance: negatively for C2I ($r = -.8$) and positively for C2N ($r = .82$) and I2N ($r = .87$), Figure 5. This indicates that better collective performance is related to higher emergence of the social and behavioral level from the neural implementation but lower emergence of the social level from the individual behavior. We elaborate on possible reasons for this below.

We also found a high negative correlation of performance with Δ_{C2I} ($r = -.8$) and a high positive correlation with Γ_{C2I} ($r = .79$). However, given that we found no downward causation or decoupling for C2I, these correlations are difficult to interpret.

---

[8]Note that in the current analysis macro-variables V are one-dimensional time series while micro-variables X are multidimensional. This means that when d(Agents) is entered into computation as a V, each agent's d is paired with activity of its own 4 neurons. However, when d(Agents) is entered as X, the triplet of values is paired jointly with a single d(CoM).
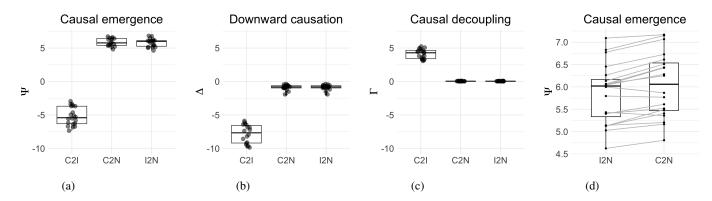
Figure 4: Values of ΦID measures in the best performing agents across all the evolution runs. (a-c) Distribution of ΦID values in final generation of all evolution runs. (d) Comparison of Ψ values in the final generation.
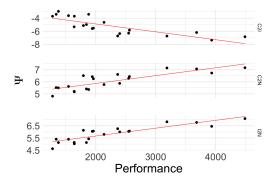

Figure 5: Correlation between Ψ and performance.

Evolution of emergence

Next, we looked at the changes in ΦID measures over evolution. Figure 6a shows improvement in behavioral performance (distance traveled) over 5000 generations in all the 20 evolution runs. The evolution of Ψ values is consistent with the correlations we found in the last generation. That is, as performance increases over time, Ψ_{C2N} and Ψ_{I2N} also increase while Ψ_{C2I} decreases (Figure 6b). Interestingly, however, collective and behavioral level shows emergence from the neural level immediately from the start, suggesting that this type of emergence is inherent to these inter-level relationships regardless of behavioral adaptation – at least with respect to specific choices of the neural architecture and macro variables of interest.

On the other hand, the Ψ_{C2I} for some of the evolution runs starts out as positive but quickly becomes negative as performance increases (example evolution runs in Figure 7). To get further intuition about the relationship between this measure and type of behavior, we can again look at Figure 2, which shows the evolution of run 19. In generation 100, the agents exhibit complex trajectories but stay in formation for only about half of the trial and travel a relatively short distance together. However, their Ψ_{C2I} is positive and high (3.3 in the trial depicted). By generation 500, they are able to

optimize their trajectories and travel more efficiently while staying in formation. They still exhibit relatively complex behavior and their Ψ_{C2I} is still small although nearing 0 (0.22). Finally, at the end of the evolution run, particular trajectories are almost linear and very similar to the CoM trajectory. This allows the agents to travel very far while in formation but the Ψ_{C2I} is low (-7.3).

These results are consistent with what has been reported by Rosas et al. (2020) in their flocking simulation. Also there the Ψ_{C2I} was negative for the behavior in which individual boids were moving in an ordered pattern. This was explained as self-predictability of the CoM being high but accompanied by low mutual information of individuals, indicating high redundancy in the system.

There are no notable evolutionary patterns for the other measures: Δ and Γ.

Discussion

We have analyzed three types of emergence in the context of an evolution of a collective task: collective from individual behavior (C2I), collective from individual neural activity (C2N) and individual behavior from individual neural activity (I2N). We used a simple formation task, and the agents we evolved to solve it, had minimally complex brain and body anatomy. This allowed us to easily compute measures of emergence proposed by the ΦID framework: causal emergence, causal decoupling and downward causation.

We found evidence for causal emergence and causal decoupling for the relation between behavioral level and neural activity (both C2N and I2N), which were present from the beginning of the evolution and increased as behavior became optimized for collective performance. This indicates that using neural processes to explain individual and collective behaviors may not be always suitable and it is true even for a toy, small-scale model in which all the relevant variables can be directly observed and measured.

Regarding the relationship between C2N and I2N, we found that collective level is relatively more emergent than
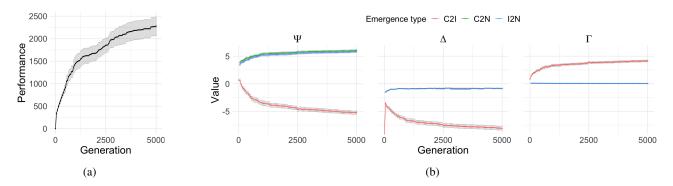
Figure 6: Changes in performance and ΦID measures over generations. Points represent mean values and error bars standard error of the mean.
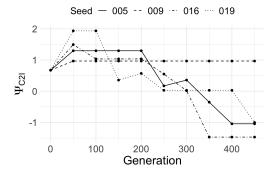

Figure 7: The changes of Ψ_{C2I} in early generations.

individual level from the underlying neural activity. However, we observed no evolutionary changes in this relationship and the values were similar and correlated with performance to a similar degree. It is an open question whether in other task settings the two values could diverge more. That is, in our example neural activity was related to CoM and individual displacements as macro variables and both are relatively similar. It would be interesting to see whether different results could be obtained when collective and individual variables pick out different patterns: for instance, collective variable capturing the type of coordination between agents and behavioral variables tracking their particular movements (i.e., the way coordination is realized).

Our negative results for the emergence at C2I which accompanies clear behavioral success in the simulated task imply that there exists strong information redundancy across our agents. In our task, to achieve higher performance for the collective goal, agents need to synchronize their actions and move in the same direction. The synchronization leads to agents sharing the same information, which in turn produces a negative result. The broader implication of this is that success in a collective task can be achieved in a variety of ways, which do not always require emergence understood as information synergy (according to the framework

we adopt).

Similarly, a lack of evidence for downward causation across different comparisons imply that this process is not necessary for individuals to optimize collective performance. This goes contrary to some intuitions in the complexity science community, in which downward causation is seen as the key to the irreducibility of higher levels. For instance, in the enactive approach to social cognition, interaction is said to have an autonomy of its own (emergence and causal decoupling) and also constrain or enslave individual processes (De Jaegher and Di Paolo, 2007; Di Paolo and De Jaegher, 2012). Again, it might be that this particular task is not appropriate for demonstrating this phenomenon. Future work could both explore other tasks or, alternatively, try to explicitly maximize the Δ measure over evolution and observe what type of behaviors and neural activity emerge as a result.

In summary, our simulation shows that reductionist approaches with partial access to information of the micro-level processes, may never lend us a full understanding of collective behavior.

Acknowledgements

References

Bechtel, W. and Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in history and philosophy of biological and biomedical sciences*, 36(2):421–441.

Bedau, M. (2002). Downward causation and the autonomy of weak emergence. *Principia an international journal of epistemology*, 6(1):5–50.

Bedau, M. A. (1997). Weak emergence. *Philosophical Perspectives. A Supplement to Nous*, 11:375–399.

Beer, R. D. (1995). On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, 3(4):469–509.

Campbell, D. T. (1990). Levels of organization, downward causation, and the selection-theory approach to evolutionary epistemology. In Greenberg, G. and Tobach, E., editors, *Theories of the evolution of knowing*, volume 4 of *The T. C. Schneirla Conference*, pages 1–17. Lawrence Erlbaum Associates, Inc, Hillsdale.

Couzin, I. D. (2009). Collective cognition in animal groups. *Trends in cognitive sciences*, 13(1):36–43.

Couzin, I. D., Krause, J., and Others (2003). Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior*, 32(1):10–1016.

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., and König, P. (2020). Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. *Frontiers in human neuroscience*, 14:39.

De Jaegher, H. and Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, 6(4):485–507.

Dewhurst, J. (2021). Causal emergence from effective information. *Thought A Journal of Philosophy*, 10(3):158–168.

Di Paolo, E. and De Jaegher, H. (2012). The interactive brain hypothesis. *Frontiers in Human Neuroscience*, 6:163.

Feinerman, O. and Korman, A. (2017). Individual versus collective cognition in social insects. *The Journal of experimental biology*, 220(Pt 1):73–82.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 37(3):424–438.

Hoel, E. P., Albantakis, L., and Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49):19790–19795.

Holroyd, C. B. (2022). Interbrain synchrony: on wavy ground. *Trends in neurosciences*, 45(5):346–357.

Hulswit, M. (2005). How causal is downward causation? *Journal for General Philosophy of Science. Zeitschrift fur Allgemeine Wissenschaftstheorie*, 36(2):261–287.

Mediano, P. A., Rosas, F. E., Farah, J. C., Shanahan, M., Bor, D., and Barrett, A. B. (2022a). Integrated information as a common signature of dynamical and information-processing complexity. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(1):013115.

Mediano, P. A. M., Rosas, F., Carhart-Harris, R. L., Seth, A. K., and Barrett, A. B. (2019). Beyond integrated information: A taxonomy of information dynamics phenomena.

Mediano, P. A. M., Rosas, F. E., Bor, D., Seth, A. K., and Barrett, A. B. (2022b). The strength of weak integrated information theory. *Trends in cognitive sciences*, 26(8):646–655.

Mediano, P. A. M., Rosas, F. E., Luppi, A. I., Jensen, H. J., Seth, A. K., Barrett, A. B., Carhart-Harris, R. L., and Bor, D. (2022c). Greater than the parts: a review of the information decomposition approach to causal emergence. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 380(2227):20210246.

Oh, K.-K., Park, M.-C., and Ahn, H.-S. (2015). A survey of multi-agent formation control. *Automatica: the journal of IFAC, the International Federation of Automatic Control*, 53:424–440.

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS computational biology*, 10(5):e1003588.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '87, pages 25–34, New York, NY, USA. Association for Computing Machinery.

Rosas, F. E., Mediano, P. A. M., Jensen, H. J., Seth, A. K., Barrett, A. B., Carhart-Harris, R. L., and Bor, D. (2020). Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. *PLoS computational biology*, 16(12):e1008289.

Steck, K. (2012). Just follow your nose: homing by olfactory cues in ants. *Current opinion in neurobiology*, 22(2):231–235.

Tononi, G., Sporns, O., and Edelman, G. M. (1994). A measure for brain complexity. *Proceedings of the National Academy of Sciences USA*, 91:5033–5037.

Traniello, J. F., Linksvayer, T. A., and Coto, Z. N. (2022). Social complexity and brain evolution: insights from ant neuroarchitecture and genomics. *Current opinion in insect science*, 53:100962.

Wehner, R., Boyer, M., Loertscher, F., Sommer, S., and Menzi, U. (2006). Ant navigation: one-way routes rather than maps. *Current biology: CB*, 16(1):75–79.

Williams, P. L. and Beer, R. D. (2010). Nonnegative decomposition of multivariate information.

Yaeger, L. S. (2009). How evolution guides complexity. *HFSP journal*, 3(5):328–339.