The Twenty-Seventh International Symposium on Artificial Life and Robotics 2022 (AROB 27th 2022),
The Seventh International Symposium on BioComplexity 2022 (ISBC 7th 2022),
The Fifth International Symposium on Swarm Behavior and Bio-Inspired Robotics 2022 (SWARM 5th 2022)
ONLINE, January 25-27, 2022

# Information Integration in Division of Labor: Validation of Earlier Findings

Soheil Keshmiri and Federico Sangati and Ekaterina Sangati†

Embodied Cognitive Science Unit, Okinawa Institute of Science and Technology Graduate School (OIST), Okinawa, Japan
({soheil.keshmiri,federico.sangati,ekaterina.sangati}@oist.jp)

**Abstract:** The present study focuses on evaluation of the underlying dynamics in two-agent, single-target pursuit within the context of division of labor paradigm. Specifically, it aims at clarifying which of the two tracking strategies viz. equipping agents with switching complementary roles versus designated specialization may result in increased information integration between agents. Although our previous findings on this topic hinted at significantly higher benefit of the first strategy, they suffered from the assumptions of linearity and independence of agents, imposed by our parametric Gaussian formulation of information integration. Here we address these shortcomings through adaptation of non-parametric multivariate information integration formalism. We show (1) that with this new formulation our previous results still hold, (2) that they also remain intact if target is taken into account for quantification of agents' information integration and (3) that the agents' relation with respect to the tracking task can be explained in terms of potential colinearity of their actions. We further discuss some of potential issues that require careful consideration while studying such dynamics and conclude by highlighting recent advances that can potentially help overcome such pitfalls.

**Keywords:** Division of labor; Multivariate mutual information; Conditional mutual information; Synergy.

## 1. INTRODUCTION

Division of labor is an important coordination strategy in a variety of species of animals and insects [1], as well as in robotic applications [2]. Recent research shows that in addition to a social function, it might also play a cognitive role by helping individual agents offload part of the processing onto others, thereby allowing for smaller and less energetically costly brains [3-5]. This idea can be called a collective intelligence hypothesis. It predicts that individuals that engage in division of labor exhibit less complex cognitive and neural processing. It has sometimes also been argued that joint action might lead to a formation of tight coupling between co-actors which could be better seen as a new emergent "supra-personal system". If this is the case, one could wonder whether the formation of such a system could be captured by a measure of joint neural information integration - that could also be called inter-brain synergy (but see Discussion). Should such a phenomenon occur, one would expect this measure to be higher the tighter the cooperation and the more individual cognitive offloading is taking place.[1]

We recently studied [6] this paradigm through a simulation setting in which artificial agents were evolved to jointly control a tracker in order to follow a target moving in a one-dimensional environment. One of the agents in a pair was responsible for moving the tracker in one direction while the other agent moved it in the opposite direction. In our study, we considered two division of labor strategies: "generalist" and "specialist" strategies. They differed in the amount of agents' task specialization. Specifically, generalist agents switched complementary roles between different trials, learning to perform both tasks. On the other hand, specialist agents evolved to always perform the same role. As a result, while the generalists could, in principle, perform their task independently, the specialists' performance was tightly dependent on cooperation with their partner. At the end of evolutionary runs we calculated the agents' individual and joint neural information integration.

In that study, we observed that the division of labor paradigm leads to a lower level of individual neural information integration in the specialists, compared to the generalists. To realize the joint information shared, we further treated the agents as two independent multivariate normal variables and computed their joint covariance matrix as the sum of their respective covariance matrices. We then utilized this joint covariance matrix to compute their joint neural information integration. Contrary to our prediction, we observed that the agents' joint shared information exactly mirrored the individuals' information integration.

Considering these observations, a crucial point that necessitates further scrutiny of our previous results is the fact that our former treatment of the topic opted for parametric formulation of information integration [7] where the agents' data were assumed to follow normal distribution. This imposed two simplifying assumptions: (1) that agents' sensory, neural, and motor data were Gaussian (2) that the two agents' actions were independent, given their common task.

Given these assumptions, it could be argued that we prevented ourselves from finding what we set out to investigate. That is, we intended to capture the integration between the two agents but in our measure assumed them to be independent systems. While this was an important preliminary step in formulating our measure of joint information integration, the effect of simplifying assumptions on the results we originally obtained cannot be underestimated.

To address these limitations, in the present study, we reevaluate our previous findings in two steps: (1) we eliminate the need for normality of agents' sensory, neural, and motor time series through the use of non-parametric, multivariate formulation of information integration measure (2)

† Ekaterina Sangati is the presenter of this paper.

[1]The first author does not support the ideas of "collective intelligence", "supra-personal systems" or "tight coupling between interacting agents".

we compute three quantities, namely, multivariate mutual information (MI), conditional MI (i.e., cMI; conditioned on agent-target distance), and the synergy, thereby eliminating the assumption of agents' independence.

Our contributions are threefold. First, we show that, after switching to non-parametric formulation of information integration, our previous results still hold. This clarifies that our previous observations were not due to some potential under/over-estimation imposed by normality and independence assumptions, but indeed stem from the dynamics of division of labor paradigm, as formulated in these types of simulations. Second, we further consolidate this observation by showing that our result hold after conditioning the agents' shared information on their task (i.e., target's movement). Interestingly, this conditioning also reveals that such shared information reduces once the target's behaviour is taken into consideration. In other words, it implies that agents' relation with respect to the tracking task can be explained in terms of potential colinearity of their actions. Third, we verify such correlation-based nature of agents' relation through the use of the measure, referred to as synergy [8].

## 2. METHODS

### 2.1. Task and simulation

Data analyzed in the present article comes from the simulation reported in our previous work [6]. Consult this reference for details about the model implementation. Here we reproduce the figures that show the task that agents performed (Figure 1) and their internal architecture (Figure 2).

### 2.2. Present measures

In [6], we simulated 100 seeds, per generalist and specialist settings. We observed that 41 out of 100 seeds for generalists and 99 out of 100 for specialists converged. Therefore, we use all converged seeds in the case of generalists along with the first 41 (i.e., out of 99) converged seeds for specialists. Furthermore, and similar to [6], we only consider the output of agents' sensory and neural nodes (i.e., $S_i$,$N_i$, $i = 1, 2$, in Figure 2).

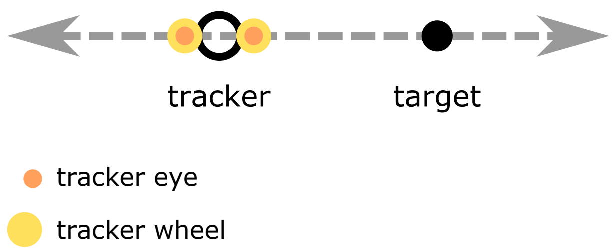


**Fig. 1**. **Experimental setup.** The target moves horizontally on a 1D line and reverses at fixed points. The tracker is controlled by two agents to move on the same line. The agents are evaluated for being able to keep the tracker on top of the target as much as possible throughout the trial. Both agents perceive the distance to the target with 'eyes' positioned on two sides of the tracker and propel the tracker by means of two wheels that output left and right velocity. This figure is adapted from [6].

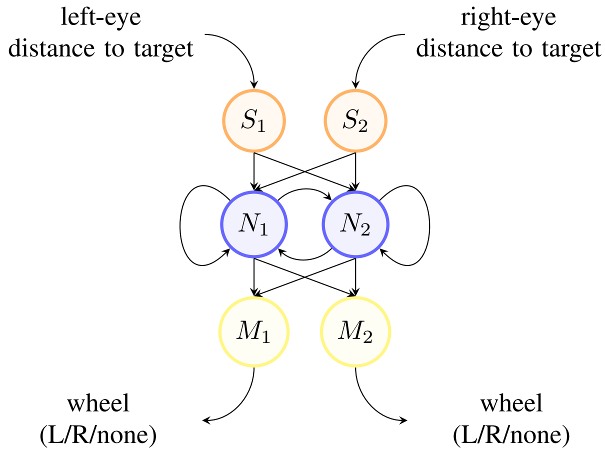


**Fig. 2**. **Network architecture**. $S_1$, $S_2$ are sensory nodes, $N_1$, $N_2$ are inner nodes (brain), $M_1$, $M_2$ are motor nodes. The brain is implemented as Continuous Time Recurrent Neural Network. The network parameters are evolved via a real-valued mixed genetic algorithm. The mapping from motor node activation to tracker wheels determines the type of agents. Generalists have their left motor output connected to the left wheel of the tracker on half of the trials while the right motor output is ignored and their right motor output connected to the right wheel on the other half of the trials while the left motor output is ignored. A compatible mapping is implemented for their co-acting agent. For the specialists the mapping is similar but they never experience a switch, thereby half of the agents in the population always controlling the right and the other half always the left wheel. This figure is adapted from [6].

We reevaluate our previous findings in two steps.

First, we eliminate the need for normality of agents' time series data. We achieve this through the use of non-parametric, multivariate formulation of this measure i.e., Kraskov-Stögbauer-Grassberger (KSG) [9]. KSG estimation builds on the non-linear and model-free capabilities of kernel estimation with bias correction, thereby resulting in a better data efficiency and accuracy as well as being effectively parameter-free. It is considered to provide best solution for (among other information-theoretic measures) mutual information (MI) and conditional mutual information (cMI) [11]. We use JIDT [10] implementation of KSG while computing MI and cMI.

Second, we eliminate the assumption of agents' independence through the use of two additional measures:

- conditional MI (cMI): in which we compute the agents' non-parametric, multivariate shared information, conditioned on agent-target distance.
- synergy (S): in which we examine whether agents indeed collectively (as hypothesized by "collective intelligence hypothesis") perform their task or their relation is purely correlational in nature.

For the case of synergy computation, we follow its original formulation by Schneidman et al. [8] i.e.,

$$S(A, B) = MI(A, B|\tau) - MI(A, B), \quad (1)$$

where $A$ and $B$ represent the two agents and $\tau$ is the target.

### 2.3. Analysis

For each computed measure (i.e., MI, cMI, and S), we perform Wilcoxon ranksum test between generalist and specialist groups. For each test, we report the test statistics and the p-values. It is worthy of note that although we report the original p-values, we only consider the p-values that survive the Bonferroni-correction (i.e., $\frac{0.05}{2} = 0.025$, where 2 indicates the number of groups) as significant. Additionally, we provide the respective Wilcoxon ranksum tests' effect-size [12]:

$$r = \frac{W}{\sqrt{N}} \quad (2)$$

with $W$ and $N$ denoting the Wilcoxon statistics and the sample size, respectively. $r$ is considered [13] small when $\leq 0.3$, medium when $0.3 < r < 0.5$, and large when $\geq 0.5$.

## 3. RESULTS

### 3.1. Mutual Information (MI)

We observed that (Figure 3) generalists were associated with significantly higher MI than specialist (W= 5.6989, $p < 1.21e^{-08}$) and that such a difference was marked with a large effect-size (r = 0.8900). Table 1 summarizes MI's descriptive statistics for generalist and specialist groups.

### 3.2. Conditional Mutual Information (cMI)

Similar to the case of MI, generalists showed (Figure 4) significantly higher cMI than specialists (W = 5.6061, $p < 2.07e^{-08}$) and that this difference exhibited a large effect-size (r = 0.8755). cMI's descriptive statistics for generalist and specialist groups are presented in Table 2.

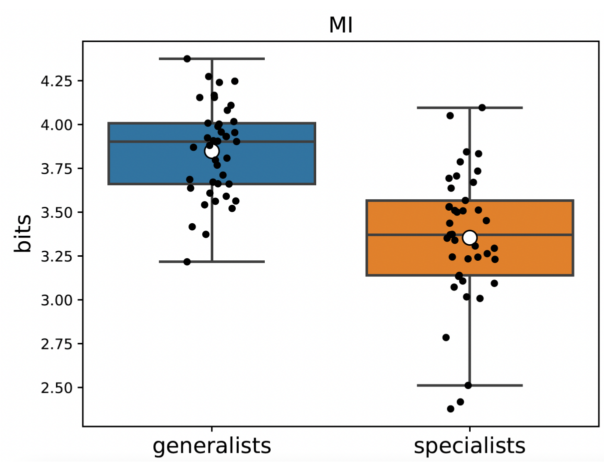


**Fig. 3.** Mutual Information (MI) between two agents in generalist and specialist groups. In this plot, each dot represents MI for a specific trial that we carried out for each group.

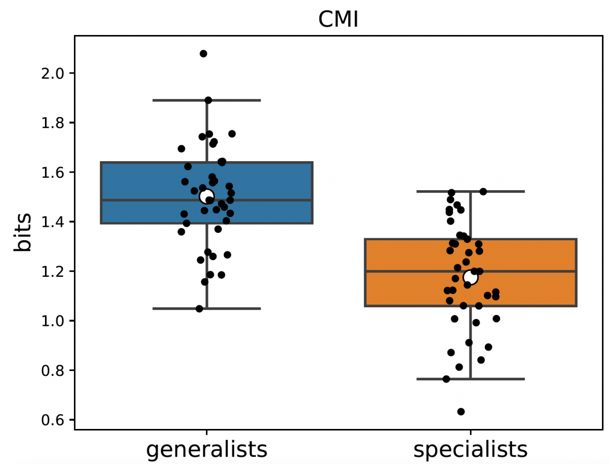


**Fig. 4.** Conditional Mutual Information (cMI) between two agents in generalist and specialist groups. In this plot, each dot represents cMI for a specific trial that we carried out for each group.

### 3.3. Synergy (S)

Generalists were associated (Figure 5) with significantly smaller (i.e., more negatively inclined) synergy than specialists (W = -2.9723, $p < 0.003$). This significant difference showed a medium effect size (r = 0.4642). Table 3 summarizes the synergy's descriptive statistics for these groups.

## 4. DISCUSSION

The central theme of the present study was to reevaluate our earlier findings [6] about the underlying dynamics in two-agent, single-target pursuit within the context of division of labor paradigm. In that study, we analyzed the nature of such a dynamics in terms of degree of information integration by the two agents while tracking the target. We achieved this objective through the use of a parametric formulation of information integration [7] in which agents' time

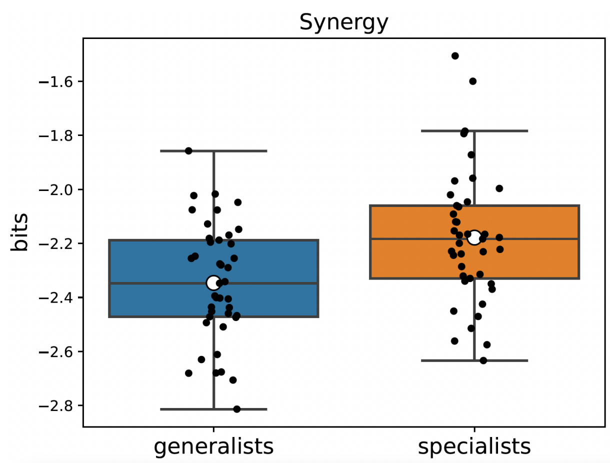


**Fig. 5.** Synergy (S) between two agents in generalist and specialist groups. In this plot, each dot represents S for a specific trial that we carried out for each group.

| Mutual Information (MI) | | | | |
|---|---|---|---|---|
| Group | Mean | Standard Deviation | Median | Confidence Interval |
| Generalists | 3.8479 | 0.2630 | 3.9025 | [3.3728, 4.2736] |
| Specialists | 3.3535 | 0.3809 | 3.3708 | [2.4168, 4.0502] |

**Table 1**. Mutual Information (MI). Descriptive statistics for generalist and specialist groups.

| Conditional Mutual Information (cMI) | | | | |
|---|---|---|---|---|
| Group | Mean | Standard Deviation | Median | Confidence Interval |
| Generalists | 1.5016 | 0.2053 | 1.4868 | [1.1563, 1.8902] |
| Specialists | 1.1749 | 0.2185 | 1.1994 | [0.7642, 1.5167] |

**Table 2**. Conditional Mutual Information (cMI). Descriptive statistics for generalist and specialist groups.

| Synergy (S) | | | | |
|---|---|---|---|---|
| Group | Mean | Standard Deviation | Median | Confidence Interval |
| Generalists | -2.3463 | 0.2162 | -2.3475 | [-2.7059, -2.0172] |
| Specialists | -2.1786 | 0.2429 | -2.1833 | [-2.5750, -1.5997] |

**Table 3**. Synergy (S). Descriptive statistics for generalist and specialist groups.

series data were assumed to be Gaussian and the agents' actions while tracking the target were considered to be independent from each other. These assumptions inevitably limited the scope and informativeness of our results as we were unable to derive an informed conclusion on whether our observations were indeed due (at least partially) to the nature of agents' interaction (given our specific implementation of the division of labor) or they were potentially confounded by our simplifying assumptions.

To this end, the MI analysis in the present study (i.e., Figure 3) that was based on non-parametric formulation of information integration [9] provided evidence for the validity of the finding in [6], thereby verifying that our previous observation was in fact a manifestation of the agents' dynamics (within the context of current implementation that is) and not a potential artefact due to linear and independence assumptions imposed by parametric formulation of information integration. We observed that the generalists were associated with significantly (with a large effect) higher information integration than their specialists counterpart. Collectively, these results challenge the "collective intelligence hypothesis" by demonstrating that the generalists with their complementary role-switching are the group with the capacity for higher information integration.

The aforementioned observation was further consolidated by our finding based on cMI (i.e., Figure 4). There, we observed that the generalists significantly higher information integration was preserved (again, with a large effect) after the target's behaviour (i.e., a common confounding variable [15, p. 138] affecting/influencing the two agents' actions) was accounted for. However, the more interesting implication of this result was its verification of our original agents' independent actions assumption in [6]. Concretely, cMI result indicated a decrease in information integration that was present in both, generalist and specialist groups. In other words, the introduction of target's behvaiour (i.e., its one-dimensional movement) substantially and significantly explained away (Appendix 4 along with Figures 6 and 7) the two agents' information integration (i.e., shared information). This, in turn, challenges the proposal, pertaining to the formation of a tight coupling between agents in division of labor in its present setting. Specifically, the explaining away of the agents' information integration by the target suggests that it was more grounded on their individuals' independent actions (although closely resembling due to the nature of task) than the outcome of any potential interaction. In other words, the observed agents' MI was not due to their coupling but coincidental and potentially due to collinearity induced by their task. This interpretation is further strengthened by the result of two groups' synergy (i.e., Figure 5) in which we observed that the generalists with whose cMI significantly higher than specialists were also the group with significantly lower synergy values. Furthermore, the negative values in both, generalists and specialists, highlight the collinearity of agents' actions by showing the greater redundancy (i.e., MI) in agents' informational space. In this regard, it is worthy of note that although there are a number of valid concerns regarding the interpretability of the information-based quantities with negative values (i.e., from the information theory perspective) [16], they are still highly useful for realizing the degree of misinformation (e.g., [11, p. 171] and [17, p. 40]) and therefore a valuable safeguard against derivation of spurious and/or detrimental conclusion where there is none.

In retrospect, and while reflecting on the notion of "synergy" as advocated in some of today's research, there are a couple of crucial points that are worth elaborating on. In what follows, we present them from two different groundings: one examining the "above-and-beyond" take on synergy, and the other from a simple and straightforward linear perspective.

First, let us consider the case of XOR logic gate i.e., the widely used example in the synergy literature (for a thorough treatment of the topic, see [14]). Let $x, y$ be the two i.i.d binary random variables that form the inputs to this gate. Let also $\tau$ represent this gate's output. Without loss of generality, let assume, $x = 1$ and $y = 0$. It is apparent that in such a setting $\tau = 1$. We know that MI$(x, y) = \sum_{\forall x,y} p(x, y) log(\frac{p(x,y)}{p(x)p(y)})$. Now, if we solely focus on the two inputs viz. $x$ and $y$, (i.e., discarding the value of $\tau$), it is apparent that the value of $x$ does not provide any information about what the value of $y$ could be. In other words, the two inputs remain independent (as they should be, by definition) and therefore $p(x, y) = p(x)p(y)$. As a result, the logarithmic term $log(\frac{p(x,y)}{p(x)p(y)}) = 0$ and hence MI$(x = 1, y = 0) = 0$ (it is straightforward to examine all four combination of binary values for $x$ and $y$ and verify that indeed MI$(x, y) = 0, \forall x, y \in \{0, 1\}$ for the case of XOR logic gate). However, the situation changes drastically once we take into account the information from the gate's output $\tau$, thereby computing the cMI$(x, y|\tau) = \sum_{\forall x,y} p(x, y, \tau) log(\frac{p(y|x,\tau)}{p(y|\tau)})$, where $p(y|x, \tau) = \frac{p(x,y|\tau)}{p(x|\tau)}$. Considering the case $x = 1, y = 0, \tau = 1$ above, it is apparent that $p(x = 1, y = 0|\tau = 1) = 0.5$ and $p(x = 1|\tau = 1) = p(x = 1) = 0.25$. ($x$ is an i.i.d and is, in principle, independent of $\tau$). This means that $log(\frac{p(y|x,\tau)}{p(y|\tau)}) = 1$ (assuming base 2 for the logarithm). We also know that $p(x = 1, y = 0, \tau = 1) = 0.5$, and hence we have cMI$(x = 1, y = 0|\tau = 1) = 0.5$ (i.e., 0.5 (from joint probability term) $\times$ 1 (from logarithm term)). Summing over all possible combination of $x, y$ and $\tau$ (i.e., XOR logic gate's table), we observe that cMI$(x, y|\tau) = 1$, as expected.

What this example signifies is that the (desirable) relation between $x$ and $y$ is indeed utterly defined by the output and has no tangible interpretation/utility unless and until the outcome of the process is explicitly and clearly specified. Put another way, this context-specificity requirement shows that the synergy is rather a bi-product than transcendental (e.g., above-and-beyond viewpoint) aspect of a given interaction. This, in turn, appears to resonate with Simon's note (while picturing an ant on a challenging beach terrain) [18, p. 52] that "the apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself."

Second, and potentially more pressing issue, that the example above underlines is the detrimental effect that any hidden relation between input-output can play in observing spurious relation between the inputs. To appreciate the issue, let us consider it from a pure linear perspective.[2] Precisely, let us define a new hypothetical measure (while following the Schneidman et al. [8] formalism in Section 2.2, equation (1)) that is purely based on correlation between $x, y$, and $\tau : C(x, y) = corr(x, y|\tau) - corr(x, y)$, where the first and second terms quantify the partial (i.e., with respect to $\tau$) and the ordinary correlations between $x$ and $y$. We know that the partial correlation between $x$ and $y$ is $\frac{\rho_{xy} - (\rho_{x\tau}\rho_{y\tau})}{\sqrt{1-\rho^2_{x\tau}}\sqrt{1-\rho^2_{y\tau}}}$, where $\rho_{xy} = corr(x, y)$. Discarding the denominator (i.e., the normalizing term), it is quite evident that all one needs to attain $C(x, y) > 0$ is to ensure that the product $\rho_{x\tau}\rho_{y\tau} < 0$ which is satisfied when one of $x$ or $y$ maintains an anti-correlation with $\tau$ (while the other correlates positively). In fact, the $x, y$ own relation (i.e., the second term $corr(x, y)$ and $\rho_{xy}$ in the first term's numerator) is completely irrelevant for this purpose: $corr(x, y) = \rho_{xy} = 0$ and one would still obtain a positive $C(x, y)$ as long as one of $x$ or $y$ is negatively correlated with $\tau$. [3] From the synergy and information integration perspective, the implication of this example is as follows. We know that for non-binary, non-Gaussian variables, MI can be high while correlation is low and vice versa [11, p. 39, footnote 2]. As a result, it is legitimate to consider a hypothetical but plausible scenario in which while the agents maintain a low MI, one of the two agent achieves a high (negative) correlation with the task, resulting in an increase in cMI (and consequentially S $> 0$ in equation (1)), thereby providing a misleading evidence for agents' "synergistic" interaction.

Although above examples serve as words of caution against light adaptation/interpretation of these measures and subsequently overstatement of their quantitative power, they should certainly not be taken as a testimony that all efforts for better understanding and quantification of the unfolding dynamics among interacting agents are in vain. In fact, there are a number of advances and progressive results [14, 16, 20] that aim at more refined decomposition of information in such interactions (for an engaging review of the topic with a comprehensive list of references, see [21]). Despite their current limitations (e.g., they are applicable for binary data with small number of variables in multivariate settings), these approaches can prove valuable in providing insights about the nature of interactions in such settings as division of labor. Therefore, their utilization for further analysis of the results presented in this study as well as its variations is a desirable ambition for the future research. This is in particular intriguing for the case of present setup of division of labor, given the agents' low data-dimensionality (i.e., six dimensions with two for each of sensory, neural, and motor nodes).

## REFERENCES

[1] J.M. Smith, E. Szathmary, E., *The major transitions in evolution*, Oxford University Press, 1997.

[2] P. Trueba, A. Prieto, F. Bellas, P. Caamaño, P., R.R. Duro, "Specialization analysis of embodied evolution for robotic collective tasks." *Robotics and Autonomous Systems*, Vol. 61, No. 7, 682–693, 2013.

[3] C. Anderson, D.W. McShea, "Individual versus social complexity, with particular reference to ant colonies", *Biological Reviews*, Vol. 76, No. 2, pp. 211–237, 2001.

[4] R.G. Bednarik, "Doing with less: hominin brain atrophy", *Homo*, Vol. 65, No. 6, pp. 433–449, 2014.

[5] J.M. DeSilva, J.F.A. Traniello, A.G. Claxton, L.D. Fannin, "When and why did human brains decrease in size? A new change-point analysis and insights from brain

[2] This transformation can be better appreciated once one realizes the close correspondence between mutual information and correlation. For instance, MI for two Gaussian random variables is $-\frac{1}{2}log(1 - \rho^2)$ [19] where $\rho$ is the correlation coefficient of these variables.

[3] It is important to note that in the present example, the significance of these correlations are of no concern. They could be of any (statistically) negligible value with $abs(\rho_{x\tau}) > 0, abs(\rho_{y\tau}) > 0, \rho_{xy} = 0$, (where "$abs$" returns the absolute value of its numerical argument), and $C(x, y) > 0$ still holds.

evolution in ants." *Frontiers in Ecology and Evolution*, Vol. 9, pp. 712, 2021.

[6] E. Sangati, S. Keshmiri, F. Sangati, "Evolution of Neural Complexity in Division of Labor Tasks", *Proceedings of the ALIFE 2021: The 2021 Conference on Artificial Life*, pp. 88, "https://doi.org/10.1162/isal_a_00422"

[7] A.K. Seth, G.M. Edelman, "Environment and behavior influence the complexity of evolved neural networks", *Adaptive Behavior*, Vol. 12, No. 1, pp. 5–20, 2004.

[8] E. Schneidman, W. Bialek, M.J. Berry, "Synergy, redundancy, and independence in population codes", *Journal of Neuroscience*, Vol. 23, No. 37, pp. 11539–11553, 2003.

[9] A. Kraskov, H. Stögbauer, P. Grassberger, P., "Estimating mutual information", *Physical Review E*, Vol. 69, No. 6, pp. 066138, 2004.

[10] J.T. Lizier, "JIDT: An information-theoretic toolkit for studying the dynamics of complex systems.", *Frontiers in Robotics and AI*, Vol. 1, pp. 11, 2014.

[11] M. Wibral, R. Vicente, M. Lindner, "Transfer entropy in neuroscience", *In Directed information measures in neuroscience, Springer, Berlin, Heidelberg*, pp. 3–36, 2014.

[12] R. Rosenthal, H. Cooper, L. Hedges, "Parametric measures of effect size", *The handbook of research synthesis*, Vol. 621, No. 2, pp. 231–244, 1994.

[13] M. Tomczak, E. Tomczak, "The need to report effect size estimates revisited. An overview of some recommended measures of effect size.", *Trends in Sport Sciences*, Vol. 1, No. 21, pp. 19–25, 2014.

[14] J.T. Lizier, N. Bertschinger, J. Jost, M. Wibral, "Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work.", *Entropy*, Vol. 20, No. 4, pp. 307, 2018.

[15] J. Pearl, D. Mackenzie, "The book of why: the new science of cause and effect.", *Allen Lane, an Imprint of Penguin Books*, 2018.

[16] P.L. Williams, R.D. Beer, "Nonnegative decomposition of multivariate information.", *arXiv preprint arXiv:1004.2515*, 2010.

[17] M. Wibral, V. Raul Vicente, J.T. Lizier, "An introduction to transfer entropy.", *Berlin: Springer*, 2016.

[18] H.A. Simon, "The sciences of the artificial.", *The MIT Press. Cambridge, Massachusetts, London, England*, 1996.

[19] G.A. Darbellay, I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space.", *IEEE Transactions on Information Theory*, Vol. 45, No. 4, pp. 1315–1321, 1999.

[20] V. Griffith, C. Koch, "Quantifying synergistic mutual information.", *Guided self-organization: inception*, pp. 159–190, 2014.

[21] P.A. Mediano, F.E Rosas, A.I. Luppi, H.J. Jensen, A.K. Seth, A.B. Barrett, R.L. Carhart-Harris, D. Bor, "Greater than the parts: A review of the information decomposition approach to causal emergence.", *arXiv preprint arXiv:2111.06518*, 2021.

# APPENDIX

## A. GENERALIST: MI VERSUS cMI

Conditioning on target resulted in significant reduction of integration information within generalist group (Figure 6, W= 7.7948, $p < 6.45e^{-15}$) with a large effect (r = 1.2173).

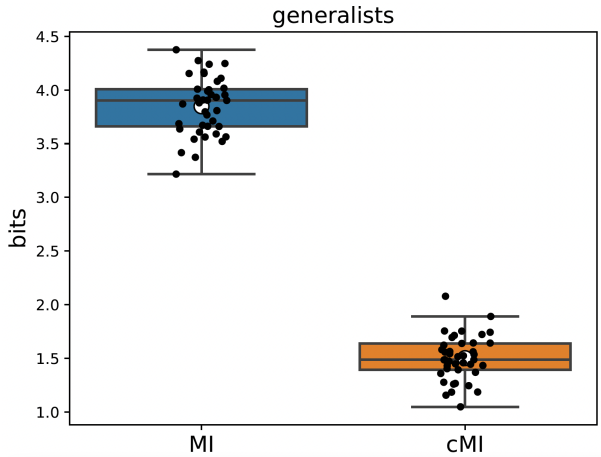


**Fig. 6**. Generalists. Difference between MI and cMI.

## B. SPECIALIST: MI VERSUS cMI

Similar to the case of generalists, Conditioning on target resulted in significant reduction of integration information within specialist group (Figure 7, W= 7.7948, $p < 6.45e^{-15}$) with a large effect (r = 1.2173).

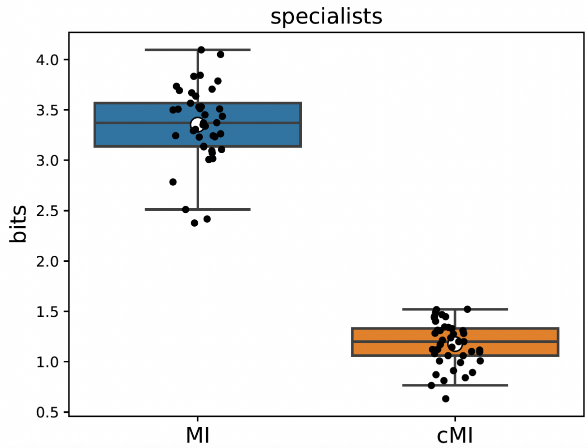


**Fig. 7**. Specialists. Difference between MI and cMI.