# An Information-Theoretic Analysis of the Emergence of Social Cohesion


Matthew Stevenson[1], Matthew Scott[1], Madalina Sas[1] and Jeremy Pitt[1]

[1]Imperial College London, London, UK
j.pitt@imperial.ac.uk



## Abstract

Social cohesion is a critical property of successful and sustainable social groups, yet remains elusive to define, quantify and metricate. This paper operationalises social cohesion, on the basis that there is some behavioural connection between voluntary association under different social arrangements and the emergence of social cohesion. It first defines a dual collective action and cooperative survival scenario, demanding informed choice of institutional rules (social arrangements) and institutional co-members. This is implemented and animated by a self-organising multi-agent system, where agents can be either cooperative or uncooperative (i.e., more or less compliant with the application of selected rules). An information-theoretic technique is then used to determine the emergence and strength of cohesion, and experimental parameters are varied to determine the conditions under which cohesion emerges more or less strongly. It is found that cohesion is most clearly distinguished at the scale of pairs and triplets, and that it emerges most strongly in uncertain environments where there is an approximately even split between cooperative and uncooperative agents.


## Introduction

In *The Dawn of Everything*, Graeber and Wengrow (2021) describe the ability of certain indigenous peoples (for example, the Amazonian Nambikwara peoples) to shift effectively between different *social arrangements* according to the seasons. For instance, during the dry season when resources were more scarce, chiefs in Nambikwara society acted as authoritarian leaders who could command small bands of foragers. In contrast, during the wet season, resources became more plentiful, and people were left to their own pursuits. Thus, this transition between seasons coincided with a transition among the Nambikwara people to an alternative social system, in which leaders became mediators, who did not issue orders but instead resolved issues in a more diplomatic manner. It can be concluded that the Nambikwara peoples displayed a sense of *social cohesion*, and that maintaining their associations with one another *irrespective* of the past, present or future social arrangements is a critical feature of that cohesion.

In general, *social cohesion* appears to be one of the most important determinants of successful and sustainable human communities and social systems, yet it still appears to be one of the hardest to define and metricate, cf. Nowak et al. (2019). Indeed, the phenomenon of social cohesion is difficult to directly observe and measure, not just because of a wide variance in definition, but also because it is abstract (like other socially constructed conceptual resource, for example norms, or social capital), and also because it is latent: it can only be indirectly inferred from the effects that it has.

However, from the evidence of the Nambikwara people, we speculate that there some behavioural connection between voluntary association under different social arrangements and the emergence of social cohesion. In making this connection, the process of operationalisation can infer the existence and extent of social cohesion, by determining some observable and measurable effects it has on individual decision-making, group formation and collective action. We therefore propose that social cohesion is an emergent phenomenon and can be measured and understood using an information theoretic framework, that has previously been used to detect and quantify emergence in complex systems by measuring non-linear interdependencies across groups of agents of varying sizes (Rosas et al., 2020).

Accordingly, this paper is structured as follows. The next section expands on the background to this paper, with a more detailed review of social cohesion and information theory. After that, a dual collective action and cooperative survival scenario is defined, demanding informed choice of institutional rules (social arrangements) and institutional co-members. This is implemented and animated by a self-organising multi-agent system, where agents can be either cooperative or uncooperative (i.e., more or less compliant with the application of selected rules). An information-theoretic technique is then used to determine the emergence and strength of cohesion, and experimental parameters are varied to determine the conditions under which cohesion emerges more or less strongly. Experimental results show that cohesion is most clearly distinguished at the scale of pairs and triplets, and that it emerges most strongly in uncertain environments where there is an approximately even split between cooperative and uncooperative agents.

# Background

This section provides further background to this paper, with a survey of some approaches to social cohesion and a brief introduction to information theory, and in particular the framework of Partial Information Decomposition (PID).

## Social Cohesion

There are numerous analyses of social cohesion in the social science literature, each taking its own approach towards metricating the concept, making direct comparison difficult. However, a survey of the various domains can inform a fresh perspective on what social cohesion 'is' – or rather 'does' – and how to measure it.

So, for example, Jenson (2011) proposed three such domains: social inclusion, cultural and ethnic homogeneity, and participation and belonging. Dragolov et al (2018) follow a similar model, also breaking down the concept into three domains: social relations, connectedness and focus on the common good, which are further broken down into subdomains. Schneifer et al. (2017) summarise further attempts to identify the different components: and identify six domains that feature commonly across the literature: social relations, attachment and belonging, orientation toward the common good, shared values, equality and inequality, and objective and subjective quality of life.

Other researchers choose to frame the concept in a different light. Chan et al (2006) propose a two-by-two framework, breaking it down into two components, subjective experience and objective manifestations, across two dimensions: a horizontal dimension, characterising social cohesion within civil society, and a vertical dimension, i.e., the cohesion that exists between the citizens and their governing institutions. This conceptualisation is more of a broad framework, which focuses less on capturing the social components of cohesion, instead opting to focus on the different scales at which can operate: this is an important factor identified by Abrams et al. (2023). Fonseca et al. (2019) identify three 'levels' that constitute social cohesion: individual, community, and institution. This too emphasises the importance of capturing social cohesion on different *scales*.

Given this variance in definition and 'units', for a software agent to represent and reason with social cohesion, we need to define and measure a property that has no general agreement on definition, and is not itself directly measurable. In this sense, social cohesion shares similarities with trust, norms and and social capital (Petruzzi et al., 2017). From this perspective, social cohesion appears to short-cut computations and coordinate expectations regarding constitutional choice in unbounded sequential interactions (Mertzani et al., 2024), and so is both a determinant for *and* a product of those interactions. As a kind of recursively emergent, abstract and latent social construction, we propose that social cohesion can be usefully analysed using information theory, cf. (Scott et al., 2024).

## Information Theory

Information theory was established as a field after the seminal paper *A Mathematical Theory Of Communication* (Shannon, 1948). Since, it has found applications in a wide variety of domains, from physics to communications, economics and neuroscience. It provides us with several quantities that characterise the amount of information within and between random variables, which the analysis of social cohesion in this project depends upon. The definitions and diagram below are taken from Cover and Thomas (2006).

Consider a random variable $X$ with alphabet $\mathcal{X}$, i.e. the set of outcomes. $X$ takes a specific outcome $x$ from the alphabet with probability $p(x)$. The *entropy* $H(X)$ is:

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{1}$$

This characterises the amount of *uncertainty* in the random variable. This definition can be extended to define the *conditional entropy* $H(Y|X)$ between two random variables $X$ and $Y$, which is defined by:

$$H(Y \mid X) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(y \mid x) \tag{2}$$

This allows us to define the *mutual information* $I(X;Y)$ between two discrete random variables:

$$I(X;Y) = H(Y) - H(Y|X) \tag{3}$$

This quantifies the amount of information that $X$ contains about $Y$. This is symmetric, i.e. $I(X;Y) = I(Y;X)$, and so it could equally be said that it quantifies the amount of information $Y$ contains about $X$. The relationship between all of these quantities is illustrated by the Venn diagram in Figure 1, where $H(X,Y)$ is the *joint entropy*:

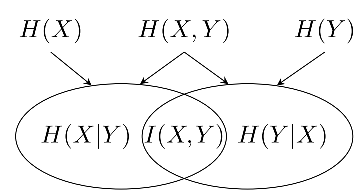



Figure 1: Illustrating the relationship between entropy and mutual information.

The mutual information between *multiple* random variables and a target variable can be defined using the chain rule for mutual information but it is difficult to directly calculate because of its bivariate nature:

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y \mid X_1, X_2, \ldots, X_{i-1}) \tag{4}$$

### Partial Information Decomposition

Partial Information Decomposition (PID) (Williams and Beer, 2010) allows a more exact expression of the mutual information for more than two variables. It does so by decomposing mutual information into indivisible information *atoms* which allow us to precisely track all the beyond-pairwise interactions (see Figure 2) through a formalism of sources and targets:

1. **Unique** information, which is the information that one source variable contains about the target, which no other source variable does.
2. **Redundant** information, which is the amount of information that source variables share about the target.
3. **Synergistic** information, which is the amount of information that is present in the *interaction* between the source variables.

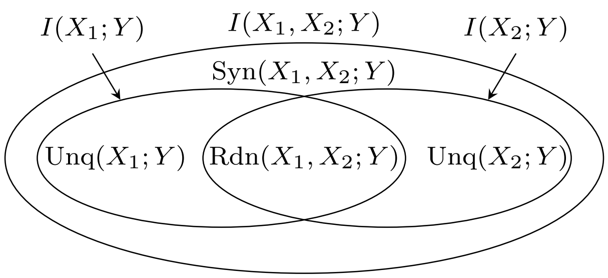



Figure 2: Partial Information Decomposition of two sources $X_1$,$X_2$ and one target $Y$ (Williams and Beer, 2010)

.

In the simplest case, with two sources and one target:

$$I(X_1,X_2;Y) = \mathrm{Unq}(X_1;Y) + \mathrm{Unq}(X_2;Y) + \mathrm{Rdn}(X_1,X_2;Y) + \mathrm{Syn}(X_1,X_2;Y) \tag{5}$$

For more than two sources, we have to keep track of all the combinations of two or more variables. Using PID also requires the choice of either a synergy or a redundancy function. Many have been proposed in the literature; for simplicity and computational efficiency, we use Minimum Mutual Information (MMI) to quantify *redundancy* (Barrett, 2015).

According to Rosas et al's (2020) theory of causal emergence, the *synergy* in a collective system variable can be used to quantify emergent behaviour, and we shall also use it to quantify social cohesion.

## Experimental Setting: *Megabike*

This section describes the experimental setting for measuring the emergence of social cohesion using the PID framework. Specifically, we use a dual collective action and cooperative survival scenario *Megabike* (Scott and Pitt, 2023).

### The *Megabike* Scenario

The *Megabike* scenario is inspired by party bikes for multiple riders. It involves a group of eight otherwise autonomous agents taking control of a single vehicle (a *megabike*), and navigating a typical AI/multi-agent gridworld in search of rewards (*lootboxes*). Each agent is individually capable of pedalling, braking, and steering the *megabike*; consequently, the agents must collectively agree on (and each agent implicitly agrees to voluntarily comply with) the social arrangements (also *regime*) that determine direction (steerage), effort (pedalling and braking), lootbox targeting, and loot allocation to replenish energy. This is illustrated in Figure 3.

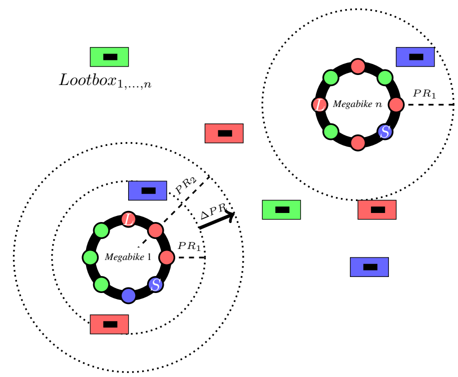



Figure 3: *Megabike*: agents voluntarily apply to bikes with specific political regimes, and work together (or not) in search of lootboxes. Energy exhaustion causes elimination.

### Core Elements

The *Megabike* scenario has several core constituent parts: agents, who occupy *megabikes*, which have regimes (social arrangements), to direct a search for lootboxes (to replenish the energy of agents). These are briefly described in turn.

**Agents** These are the 'players' of the simulation. They form associations with other agents on *megabikes*, and pedal around the simulation world, depleting energy as they pedal. They can regain some energy by colliding with a *lootbox*. When their energy meter reaches 0, they die and are removed from the simulation. Agents are characterised by a *Platonic Tendency*, which influences how cooperative or selfless they are. Each agent also maintains a trust network, recording its trust in other agents based on interactions with them (*Agent Trust)*, and their trust in different regimes according to their experience of them (*Regime Trust*).

**Megabikes** The vehicle that the agents operate. It is an 8-seater bike, and each agent riding on it can apply a steering angle, pedalling force, and braking force.

**Lootboxes** These are the simulated resources that can be collected in order to replenish energy to the agents, if they are of corresponding colours. Once collected, they are removed from the world.

**Regimes** Each *megabike* is categorised into one of the **three** possible regimes. These determine how many representative agents it has, and how certain decisions are made, specifically accepting or rejecting applications to join the *megabike*, deciding the direction and force to pedal, and lootbox allocation. The three regimes are:

- **One**: Executive decisions are made by a single agent.
- **Some**: Collective decisions are made by a majority vote of three representative agents.
- **Many**: Collective decisions are made a majority vote of all of the agents.

Some decisions require a small energetic cost. In some cases, agents also have the opportunity to not comply with the decision, but they risk being an unfavoured teammate in subsequent iterations due to a loss of *Agent Trust*.

## Scenario Flow

The scenario is composed of *iterations* and *rounds*. One full simulation consists of 100 iterations; for each iteration, 100 rounds are executed. Iterations are a 'low frequency' loop, where agents have the chance to form new associations with agents on a different *megabike*. Rounds are a 'high frequency' loop, where agents pedal through the world and collect lootboxes in a struggle for survival.

**Populating the world** When spawning agents, they are initialised with a *Platonic Tendency* between 0 and 1. This represents their likelihood to co-operate or act selflessly in the simulation, where 1 is most selfless and 0 is most selfish. It follows a bimodal distribution: a certain proportion of agents are spawned with the 'good' *Platonic Tendency* $p$ ($p > 0.5$), and the rest of the agents are spawned in with the 'bad' *Platonic Tendency* $1 - p$. They are also initialised with empty agent trust networks and regime trust networks, and are not initially assigned to any *megabike*.

Megabikes have a standard maximum capacity of 8 agents, while there are 3 regimes to consider. As such, we spawn a number $N = 8 \times 3n$ agents for $3n$ megabikes. *Megabikes* are hardcoded as one of the three regime types and this is immutable throughout the simulation.

The number of lootboxes that spawn in is governed by a ratio of $lootboxes : agents$ set at $2.5 : 1$. This ensures that the quantity of lootboxes scaled appropriately with the number of agents in the simulation. They are spawned with a random amount of 'loot' (i.e., energy), such that there is enough to distribute a 'reasonable' amount to each agent but not so much that even a full *megabike* of energy-depleted agents will be saturated by it. Finally, they are randomly assigned a fixed spawn location.

## Iterations

Each simulation consists of 100 iterations. Each iteration consists of a series of phases: disembarkation, representative selection, and association formation, as follows:

**Disembarkation** Agents disembark from the *megabikes* they rode in the previous iteration and form a pool of all surviving agents.

**Representative Selection** In order to perform the acceptance process properly for each regime, there needs to be representatives on the *one* and *some megabikes*. Therefore, a small group of agents are chosen randomly to be these initial representatives. These agents are then responsible for deciding on the admittance of agents to their *megabikes*.

**Forming Associations** The remaining majority of agents who were not selected as representatives now need to find a *megabike* to join: they form a queue in random order and apply to join a *megabike*. To decide on which bike to join, they assign a *Trust Score* $s$ to each *megabike* according to:

$$s = w_A T_A + w_R T_R \tag{6}$$

where $w_A$ and $w_R$ are experimentally adjustable weights.

$T_A$ is the *Agent Trust*, calculated by an agent at the front of the queue, is the average trust in all $n$ agents currently on a *megabike*, where $t_i$ is the trust of agent $i$ ($1 \leq i \leq n$):

$$T_A = \frac{1}{n} \sum_{i=1}^{n} t_i \tag{7}$$

$T_R$ is the *Regime Trust*, calculated as $1 - G$, where $G$ is the Gini coefficient (Sen, 1997) of the *megabike*'s regime. This is calculated by first constructing a vector of incomes for all agents riding under that regime in the previous iteration. This can then be used to calculate a Gini Coefficient for each regime, using the formula:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \bar{x}} \tag{8}$$

where $x_i$ is the income of agent $i$, $\bar{x}$ is the mean income, and $n$ is the number of incomes. This produces a value $G$ between 0 and 1 for each regime, where a higher value represents greater inequality. This is inverted by performing $1 - G$, such that a lower value represents greater inequality.

The Gini coefficient is used as a trust metric because it captures, within an appropriate range, an agent's willingness to expose itself to risk manifested explicitly as a probability. As the applying agents will not be representatives, they must assess regimes by how likely they are to survive under that regime. This is effectively asking 'how equitable does this regime tend to be with its lootbox energy distributions', and this is captured by the Gini index. However, this is not the only metric that can be used for evaluating distributive justice, as identified by (Rescher, 1966), and future work will consider other metrics (Pitt et al., 2014).

In this way, the agents give a weighting to both their trust in the other agents on the *megabikes*, and their trust in the regime. After scoring each *megabike*, the agent ranks them and applies to the *megabike* with the highest score. This prompts the agents on this *megabike* to make an acceptance decision, which is implemented according to the governance regime of the *megabike*. Note that if an agent applies to join a *megabike* with the **Many** regime and zero agents are currently assigned to it, then the agent applying is automatically assigned to that *megabike*.

An agent's application to a *megabike* is either accepted or rejected. If accepted, the agent is removed from the queue and assigned to the *megabike*; if rejected, it is moved to the back of the queue and gets the chance to re-apply again a fixed number of times (equal to the number of *megabikes*, in the event it wants to apply to a different *megabike* each time). At the end of the application process, agents that have not been assigned to a *megabike* are assigned to one at random. The associations are now formed and the *megabikes* are ready to begin pedalling around the world.

## Rounds

Each iteration consists of 100 rounds, during which agents pedal around the world on their *megabike*, build (or lose) trust in their fellow bikers, and try to stay alive. The sequence of events in each round is: determine and apply pedalling or braking force and steering direction, lootbox allocation, termination and gossip.

**Direction Decision and Movement** Each *megabike*'s primary task each round is to decide on a direction to travel. The *decision* is made according to the regime type **One**, **Some** or **Many**. The key factors for each agent involved in the decision-making are if its *Platonic Tendency* $> 0.5$, or if its average *Agent Trust*, $T_A > 0.5$. If so, then it will target the lootbox with highest gain (i.e. the agent is selfless or has high trust); otherwise, it will target the nearest lootbox of its own colour.

Note also for each agent participating in the decision there is a small energetic cost. This accounts for the effort in processing and messaging; but the cost could be reciprocated by increased trust following a 'good' decision.

Following this *institutional* decision, each agent decides for itself the pedalling and steering forces it would prefer to apply. An agent can either comply with the institutional decision and apply forces on that basis, or apply their own preference. By applying a force, they lose energy in proportion to their pedalling power.

The server then uses the physics engine to update the location of each *megabike* according to the collective forces that have been applied.

**Lootbox Check and Allocation** After making a movement transition, a *megabike collects* a lootbox (or lootboxes) if its trajectory includes a lootbox's locations. Collecting a lootbox initiates an allocation process on the *megabike* whose outcome again depends on its governance regime.

If the lootbox is not the agent's colour, it proposes an equal distribution. Otherwise, the key factors for each agent involved in the decision-making process are its *Platonic Tendency* and average *Agent Trust*. If either of these values is greater than 0.5, then the agent will propose an equal distribution; otherwise, the agent will propose half of an even split to other agents, and the remainder to itself. In **Some** and **Many** regimes, the allocations proposed are aggregated and divided by the number of decision-makers to produce the final allocation. Once the allocation process is completed, each agent increases its energy according to its allocation. Finally, collected lootboxes are despawned.

If several *megabikes* have collected the same lootbox(es), their combined resources is first split evenly, and then the allocation process is applied to each *megabike's* share.

Note that in the current implementation, there is no energetic cost for participating in deciding the allocation. Such a cost could be introduced, and that cost could be reciprocated by increased trust following a 'fair' decision, according to objective or subjective assessments of fairness (see, respectively, Pitt et al. (2014) and Pitt (2017)) – and whether the agents want to invest some of their resources in making such fairness judgements.

**Terminating Agents** Agents have their energy meters checked, and if they have zero energy remaining they are terminated and removed from the simulation. They do not respawn in the following iteration.

If a representative from a *megabikes* dies, then it is replaced by another non-representative agent on the *megabike*, provided there are enough agents on the *megabike* to accommodate this.

**Gossiping and Trust Update** At the end of the round, agents gossip about the events of the round, sending messages to their co-riders and processing messages received from them. Agents send messages which state which lootbox they targeted, what forces they applied, and whether or not they conformed with the institutional decision. An agent receiving such messages use this information to update trust values in other agents in its trust network. This value is clamped between [0.0, 1.0] and incremented or decremented by an impact factor $\delta$, i.e., the trust $t'_i$ of some agent $a$ in the $i$th agent in $a$'s trust network in one round depends on its trust $t_i$ in the previous round:

$$t'_i = \mathbf{clamp}(t_i + \delta, 0.0, 1.0)$$

where $\delta$ is defined for each reported event.

In the current implementation, these communications are assumed to be reliable signals because all actions are, effectively, fully monitored. A more advanced gossiping system would involve selective monitoring (with an energetic cost proportional to reliability), selective communication (i.e., only with some co-riders), and yet another trust decision on whether to trust the signaller or not.

## Experimental Results

This section reports the experimental results. First the experimental method is summarised, and then the specific results: firstly establishing the baseline behaviour (better than random), and then varying certain parameters specified in a config file, for example the proportion of 'good' agents, the number of *megabikes*, and the capacity of each *megabike*. The primary set of command-line parameters is shown in Table 1.

| Parameter | Range |
|---|---|
| `BikerAgentCount` | `int` |
| `LootBoxRatio` | `float64` |
| `GlobalRuleCount` | `int` |
| `GoodPlatonicTendency` | `float64` |
| `ProportionOfGoodAgents` | `float64` |
| `LootBoxCount` | `int` |
| `MegaBikeCount` | `int` |

Table 1: Experimental Parameters

The experimentally-determined values of $w_A$ and $w_R$ were set to 0.9 and 0.1 respectively, for all agents. The relative strength on the *Agent Trust* was needed because if the weight for the *Regime Trust* was too high then the agents wouldn't form semi-stable groups without excessive interference from the previous iteration's regime performance.

The impact factors $\delta$ for each reported event are shown in Table 2. Choosing the same lootbox is considered a less reliable indicator of trust than working together or conforming to pedal directions proposed by a regime, and so has a correspondingly smaller impact.

| Event | Impact ($\delta$) |
|---|---|
| Same lootbox target | $+0.005$ |
| Pedal power $< 20\%$ | $-0.1$ |
| Pedal power $> 80\%$ | $+0.1$ |
| Conformance with regime decision | $-0.1$ |
| Non-conformance with regime decision | $+0.1$ |

Table 2: Trust impact factors for reported events

### Experimental Method

To quantify social cohesion in the *Megabike* scenario as a phenomenon of repeated voluntary association using PID, we must first define the source and target variables. $X_1$ and $X_2$ as the identities of agents in the simulator, and the target $Y$ as their likelihood to associate on a bike:

$$Y = \begin{cases} 1 & \text{if } X_1 \text{ and } X_2 \text{ are on the same \textit{megabike}} \\ 0 & \text{if they are not on the same \textit{megabikes}} \end{cases} \tag{9}$$

This probability distribution can be estimated directly by counting co-occurences of agents in the same *megabike* over the iterations, to generate a joint probability distribution over the three variables. Same for marginal distributions. All probability distributions are then used to calculate the mutual information terms according to Eqs. (1-4), and then the cohesion metric, which is the *synergy* and we denote by $\Psi$:

$$\begin{aligned} \Psi &= \mathrm{Syn}(X_1, X_2; Y) \\ &= I(X_1, X_2; Y) - \textstyle\sum_{i=1}^{2} \mathrm{Unq}(X_i; Y) - \mathrm{Rdn}(X_1, X_2; Y) \\ &= I(X_1, X_2; Y) - \textstyle\sum_{i=1}^{2} I(X_i; Y) + \mathrm{Rdn}(X_1, X_2; Y) \end{aligned} \tag{10}$$

This aims to isolate the information that is only present in both source variables when taken together, but not apart. This quantifies how much more agent pairs, as a group, predict their associations versus their individual tendencies. It is indicative of cooperation in the system at the level of the source variables (the agents) to influence the target (whether they are on the same *megabike*). To better understand higher-order social relationships, we will also consider groups larger than two.

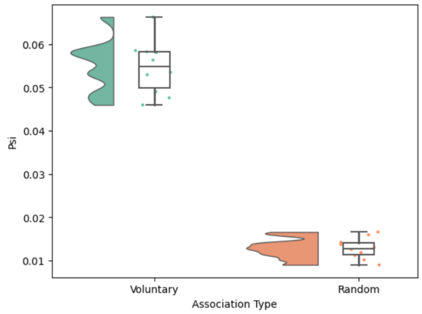
$\Psi > 0$ indicates that there is some amount of information or predictive ability that is only explained by the agents together but cannot be explained by their individual tendencies; there is *meaningful structure in the association patterns, owing to the interaction between pairs (or groups) of agents.* The more the synergy could explain their associations, the stronger the degree of emergence cohesion.

Each experiment was performed by analysing the simulation log in a Jupyter Notebook and producing a $\Psi$ for each run using the Python discrete information theory package `dit` (James et al., 2018). The results are plotted using a raincloud plot (Allen et al., 2021), which shows datapoints alongside a box and whisker and distribution overlay.

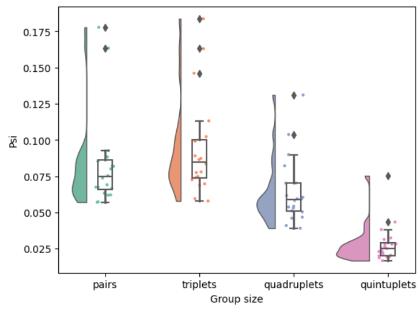
### Results

For each of the experiments, multiple (10, 20 or 30) runs of the simulator were performed for each experimental setup, to isolate anomalies and capture the real trends in the data.

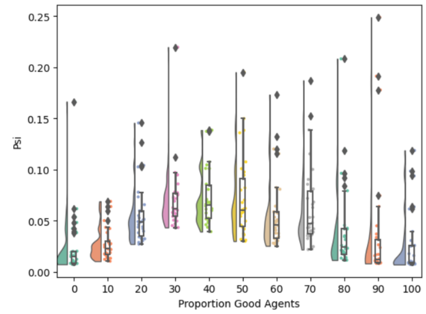
**Random vs Voluntary Association** In order to verify that this approach works as intended, the first experiment performed was a comparison between the agents undergoing voluntary re-association as normal, and an altered setup of the simulator where they are assigned to *megabikes* at random. Intuitively, if the method works, we should expect it



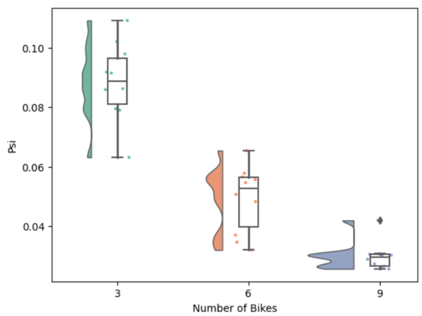

(a) *Voluntary vs. Random Association*



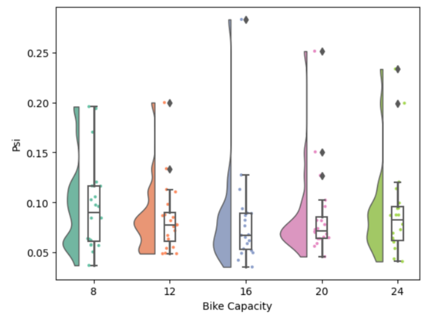

(b) *Scale Analysis of Cohesion*




(c) *Proportion of 'Good' Agents*




(d) *Number of Megabikes*




(e) *Increasing megabike Capacity*

Figure 4: Experimental Results

to produce non-zero values for voluntary self-organised association, and approximately zero for random association. Figure 4a presents the results for this experiment in both setups, with 24 agents and 3 *megabikes* (1 per regime) in each.

These results validate our approach: we indeed record the median $\Psi$ near zero for the random association setup, and non-zero values for the voluntary. This is a promising result and gives credence to the approach.

**Scale of Cohesion** We can now proceed to the first real line of inquiry - at what *scale* is the cohesion developing? This is of interest to analyses of cohesion (Abrams et al., 2023) and our method lends itself well to such an investigation. By increasing the number of source variables, we can detect the extent to which different sized groups are emerging. For example, in the case of triplets, we form all unique groups of 3 when building the probability distribution from the simulation log, (as opposed to all unique pairs), and use 3 source variables rather than two in our calculation of $\Psi$. The same can be done for 4, 5, or larger group sizes. We can then run the simulator to generate multiple logs, and analyse them at each scale. Figure 4b presents this analysis for the standard experimental setup (24 agents, 3 *megabikes*).

These results are highly informative, and seem to suggest that groups of size 2 and 3 are emerging as the most coherent units, with groups of 4 slightly less so and to an even less extent groups of 5. This fits appropriately with our intuition that smaller groups form more easily, whilst also interestingly suggesting that triplets are forming as coherent units as pairs. In the following section, wherein we investigate the conditions that most favourable to cohesion developing, we must fix the scale at which we are looking in order to make results comparable. Therefore, in all the experiments below the measurement of cohesion is fixed at the scale of pairs, as this is the most simple and informative scale to examine.

**Proportion of 'Good' Agents** The first parameter of interest is the proportion of 'Good' agents in the simulator. As explained above, there are a certain proportion of agents in the simulator with the 'good' *Platonic Tendency* $p$, and the remaining agents are instantiated with the 'bad' tendency, which is simply $1-p$. This provides two variables to adjust: the value of 'good' *Platonic Tendency*, and the proportion of agents with this 'good' tendency. The 'good' *Platonic Tendency* value is fixed at 1 in this experiment, with the 'bad' agents therefore having 0, to isolate the effect of the proportion variable.

The experiment was performed by adjusting this proportion from 0% to 100% in steps of 10 and plotting $\Psi$ for each case. The results are presented in Figure 4c and suggest that:

- When there are few good agents in the simulator, and predominantly bad agents, very minimal cohesion develops. This could be explained as the agents all developing a low trust in one another, and so repeatedly avoiding each other. Eventually, with such low trust universally prevailing, they begin to get rejected by the initial representatives, and rotate between *megabikes*.

- When there are predominantly good agents, and few bad agents, a similarly low level of cohesion develops. This could be interpreted as all agents in the simulator developing high trust of all the other members they meet, meaning that they are likely to accept anyone that applies to their *megabike*. Thus, the random queuing order they get placed in at the start is likely to have the biggest influence upon grouping, as whoever applies they will accept. In this way, they do not form coherent groups as they do not need to, and as such cohesion is low.
- In the mixed setups, where there are more similar numbers of good and bad agents, the cohesion seems to peak. This is an intriguing result, and seems to suggest something notable - cohesion peaks in mixed scenarios as a protection mechanism against uncertainty. It becomes more important in these scenarios, and thus develops more strongly.

This is perhaps counter-intuitive to our initial assumptions, as one may expect the fully 'good' scenario to have strong cohesion develop if you understand it as mutual 'friendship'. What these results suggest is that in this scenario, cohesion is essentially acting as a protection mechanism against uncertainty, as it becomes most critical to separate the good from the bad in uncertain scenarios with a mixed agent pool, It also allows more diversity to flourish in tandem with coherence, a balance which is often a marker of flexibility and adaptability reminiscent of many complex adaptive systems in nature.

**Number of *Megabikes*** The second parameter we varied was the number of *megabikes* in the simulator. This scales with the number of agents: 3, 6, and 9 *megabikes* were instantiated in the simulator, with 24, 48, and 72 agents respectively. The results are presented in Figure 4d.

These findings indicate that it is more difficult for cohesion to emerge within a larger player-*megabike* environment. This is consistent with our intuition - more players and *megabikes* available mean it is harder to find people to group up with repeatedly.

***Megabike* Capacity** The final parameter eplored was the *capacity* of the *megabikes*. For these experiments, we returned to the standard 3-*megabikes* setup, but increased the capacity of the *megabikes* in increments of 4, up to a maximum of 24 agents. The results produced are shown in Figure 4e. The relatively similar $\Psi$ recorded in each case seems to suggest that the *megabike* capacity has minimal effect on the level of cohesion (re-association) that emerges.

Further investigations could involve experimenting with capacities below 8 to create an excess of labour, to see if the trend (or lack thereof) changes in the reverse direction. Moreover, how well the method and cohesion metrics scales in much larger systems with sparser social networks, more sophisticated agent behaviours, movement and decision-making (and, for example, being capable of deception (Sarkadi, 2024)), and more complex environments, including negotiation between *megabikes*.

## Summary and Conclusions

In summary, this paper has addressed the behavioural phenomenon of social cohesion in human societies using agent-based modelling of self-organising systems. In particular, the information theoretic framework of Partial Information Decomposition (PID) was used to measure social cohesion as an emergent property of voluntary association under different types of social arrangements (or political regime).

The specific contributions of this paper are:

1. the design and specification of a dual collective action and cooperative survival scenario that demands agents make sequential decisions about voluntary associations with one another;
2. the implementation of this scenario by programming a generic multi-agent simulator in Go;
3. integrating an information theoretic method for identifying the emergence of social cohesion and quantifying its strength; and
4. experimental results which established the conditions under which social cohesion emerges more or less strongly.

Social cohesion has been deeply studied in the social sciences, and various definitions and metrics have been proposed. However, sometimes, it is better to define a socially-constructed concept, like norms, trust and, as here, social cohesion not by what it *is*, but what it *does*. We interpret these experimental results as indirect yet compelling evidence that as an *emergent* phenomenon what social cohesion *does* is to create the conditions for de-risking regime change, as observed by Graeber and Wengrow (2021) in the Amazonian Nambikwara and other indigenous peoples. It might also decrease the individual and collective processing costs in determining whether or not decisions are 'good' or 'fair'.

In conclusion, by abstractly modelling human behaviour in the form of agents in a simulator, this work has shed some light on the elusive concept of social cohesion as an emergent phenomenon. It also provides some insight into the conditions under which it develops, and the conditions under which it strengthens or weakens. Methodologically, following in the line of work initiated by Scott et al (2024), it has reinforced the capability of information-theoretic techniques to measure socially-constructed properties (like leadership and social cohesion) in self-organising multi-agent systems, pointing to a novel synthesis of multi-agent systems, information theory and cybernetics for understanding human behaviour and social system dynamics.

## Acknowledgements

We are particularly grateful for the very helpful comments of the two anonymous reviewers.

## Sources

The code repository for this version of *Megabike* can be found at `https://github.com/MattSScott/MegaBike`.

## References

Abrams, D., Davies, B., and Horsham, Z. (2023). Rapid review: Measuring social cohesion. University of Kent and Belong Network. DOI: 10.22024/UniKent/01.02.101469, `https://kar.kent.ac.uk/101469/`.

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., and Kievit, R. A. (2021). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*, 4:63.

Barrett, A. B. (2015). Exploration of synergistic and redundant information sharing in static and dynamical gaussian systems. *Physical Review E*, 91(5):052802.

Chan, J., To, H.-P., and Chan, E. (2006). Reconsidering social cohesion: Developing a definition and analytical framework for empirical research. *Social Indicators Research*, 75(2):273–302.

Cover, T. M. and Thomas, J. A. (2006). *Elements of information theory (second edition)*. Wiley-Interscience.

Dragolov, G., Larsen, M., and Koch, M. (2018). Level, trend and profiles of social cohesion in asia. In *What Holds Asian Societies Together? Insights from the Social Cohesion Radar*, pages 69–95. Bertelsmann Stiftung.

Graeber, D. and Wengrow, D. (2021). *The dawn of everything: a new history of humanity*. Penguin Books Ltd.

James, R. G., Ellison, C. J., and Crutchfield, J. P. (2018). dit: a Python package for discrete information theory. *The Journal of Open Source Software*, 3(25):738.

Jenson, J. (2011). Defining and measuring social cohesion. Commonwealth Secretariat: UN Research Institute for Social Development. United Nations Digital Library: Social Policies in Small States Series.

Mertzani, A., Pitt, J., Sarkadi, S., Sas, M., Scott, M., and Smit, C. (2024). Cohesion and the explanation of constitutional choice in self-governing systems. Workshop on Agent-Based Modelling of Human Behaviour (ABMHuB) Workshop at ALife.

Nowak, A., Vallacher, R., Rychwalska, A., Roszczynska, M., Ziembowicz, K., Biesaga, M., and Kacprzyk, M. (2019). *Target in control: Social influence as distributed information processing*. Cham, CH: Springer.

Petruzzi, P. E., Pitt, J., and Busquets, D. (2017). Electronic social capital for self-organising multi-agent systems. *ACM Trans. Auton. Adapt. Syst.*, 12(3):13:1–13:25.

Pitt, J. (2017). Interactional justice and self-governance of open self-organising systems. In *11th IEEE International Conference SASO*, pages 31–40.

Pitt, J., Busquets, D., and Macbeth, A. (2014). Distributive justice for self-organised common-pool resource management. *ACM TAAS*, 9(3):14:1–14:39.

Rescher, N. (1966). *Distributive Justice*. Indianapolis, IN: Bobbs-Merrill Company, Inc.

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.

Sarkadi, S. (2024). Self-governing hybrid societies and deception. *ACM Trans. Auton. Adapt. Syst.*, 19(2):9.

Schiefer, D. and Van der Noll, J. (2017). The essentials of social cohesion: A literature review. *Social Indicators Research*, 132:579–603.

Scott, M. and Pitt, J. (2023). Interdependent self-organizing mechanisms for cooperative survival. *Artif. Life*, 29(2):198–234.

Scott, M., Sas, M., and Pitt, J. (2024). An information-theoretic analysis of leadership in self-organised collective action. In *IEEE International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*, pages 101–110.

Sen, A. (1997). *On Economic Inequality*. Clarendon Press.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Williams, P. L. and Beer, R. D. (2010). Nonnegative decomposition of multivariate information. *CoRR*, abs/1004.2515.

Xavier Fonseca, S. L. and Brazier, F. (2019). Social cohesion revisited: a new definition and how to characterize it. *Innovation: The European Journal of Social Science Research*, 32(2):231–253.