


# Emergent Dominance Hierarchies in Reinforcement Learning Agents

Ram Rachum <sup>1</sup>[0009–0004–9278–7134], Yonatan Nakar<sup>2</sup>[0009–0007–5628–1293],  
Bill Tomlinson<sup>3</sup>[0000–0002–8386–4730], Nitay Alon<sup>4,5</sup>[0000–0002–2698–3573], and  
Reuth Mirsky<sup>1</sup>[0000–0003–1392–9444]

<sup>1</sup> Bar-Ilan University, Ramat Gan, Israel [rachumr@biu.ac.il](mailto:rachumr@biu.ac.il)

<sup>2</sup> Tel Aviv University, Tel Aviv, Israel

<sup>3</sup> University of California, Irvine, United States of America

<sup>4</sup> Max Planck Institute for Biological Cybernetics, Tübingen, Germany

<sup>5</sup> The Hebrew University of Jerusalem, Jerusalem, Israel

**Abstract.** Modern Reinforcement Learning (RL) algorithms are able to outperform humans in a wide variety of tasks. Multi-agent reinforcement learning (MARL) settings present additional challenges, and successful cooperation in mixed-motive groups of agents depends on a delicate balancing act between individual and group objectives. Social conventions and norms, often inspired by human institutions, are used as tools for striking this balance.

We examine a fundamental, well-studied social convention that underlies cooperation in animal and human societies: dominance hierarchies.

We adapt the ethological theory of dominance hierarchies to artificial agents, borrowing the established terminology and definitions with as few amendments as possible. We demonstrate that populations of RL agents, operating without explicit programming or intrinsic rewards, can invent, learn, enforce, and transmit a dominance hierarchy to new populations. The dominance hierarchies that emerge have a similar structure to those studied in chickens, mice, fish, and other species.

**Keywords:** Multi-Agent Reinforcement Learning · Reinforcement Learning · Cultural Evolution · Multi-Agent Systems · Cooperative AI

## 1 Introduction

Many animal species are able to collaborate, form groups and harness collective intelligence [11, 23]. Among those species, humans have achieved a scale and sophistication of collaboration that stands as one of the most profound and unparalleled phenomena on Earth. The grand feats of humanity were accomplished by groups of humans working together; what is grander is that these groups are composed of individuals with wildly different beliefs, motivations, and talents [59]. While differences between individuals provide an important source of diversity, they also lead to adversity and conflict, which can result in the violent demise of the social group [63]. Those human civilizations that tempered their

internal conflicts while capitalizing on the diversity were the ones to survive and flourish [4, 5]. The intricate social structures that make up these civilizations turn a potential runaway explosion into a controlled reaction, enabling groups of humans to achieve breakthroughs that exceed the combined capabilities of their individual members [39].

Multi-agent reinforcement learning (MARL) presents an opportunity to implement simplified versions of those intricate social structures, where the environment is closed and controlled, and humans are replaced by artificial agents. Research in the paradigm of Cooperative AI [12, 41, 56] highlights the role of such social structures (also labeled *institutions*) in producing human-aligned AI systems. Existing research into reproducing such institutions for use in MARL settings includes language [38], voting systems [62], reputation systems [42], and bargaining [30].

In this paper we turn our attention to a primordial institution that underlies cooperation in both animal and human societies: *dominance hierarchies*.

**Our contribution:** We adapt the ethological theory of dominance hierarchies to artificial agents by modeling conflicts between each pair of individuals as the classic game of Chicken [44]. We generalize the game of Chicken into an  $N$ -player stochastic game that we call *Chicken Coop*. We borrow the established terminology and definitions from animal study to allow calculating dominance metrics from agents’ action history. We release the Chicken Coop environment under the MIT open-source license.<sup>6</sup> We train RL agents on the Chicken Coop environment to optimize their score, observing the emergence of three behaviors:

1. Agents collaboratively invent dominance hierarchies.
2. Agents enforce dominance hierarchies on other agents.
3. Agents transmit dominance hierarchies to new populations.

We show a causal link between the agents’ ability to identify the opposing agent and the emergence of dominance hierarchies. We compare the attributes of dominance hierarchies that emerge to empirical results from groups of different animals, showing structural similarities.

By providing MARL systems with the ability to coordinate their collective behavior via dominance hierarchies, we hope to enable a more seamless integration into existing human systems [45].

## 2 Background

When animals live together in a social group they often find themselves in conflicts with other members of the group over resources such as food or mating partners. These conflicts can lead to physical injury and death. A *dominance relationship* [43] is a pattern in the conflicts between a pair of individuals, in which the *dominant* individual is more likely to escalate violence and win the majority of resources, and the *subordinate* individual is more likely to deescalate

<sup>6</sup> The code and its manual are available at <https://github.com/cool-RR/chicken-coop>

and yield any contested resources. This dynamic prevents the runaway escalation of violence that could otherwise lead to injury and a loss of group cohesion [50].

A *dominance hierarchy*, also known as a *pecking order*, is the aggregate of all the dominance relationships between each pair of agents in the group. It is defined by modelling dominance relationships as a total relation on the set of agents, where  $d \rightarrow s$  means that individual  $d$  is dominant over individual  $s$ . A dominance hierarchy is modelled as a complete, directed graph (tournament) where agents are represented as nodes and dominance relationships are represented as directed edges [29]. This graph representation is more than just a map of agonistic behavior in the group; any skirmish between individuals  $a$  and  $b$  affects not only the edge connecting  $a$  and  $b$ , but the edges connecting either  $a$  or  $b$  with each of the other individuals in the group, by the mechanisms of *winner effects* [7], *loser effects* [18] and *bystander effects* [19].

The field of dominance hierarchies traces its roots to Schjelderup-Ebbe [48], which described pecking orders in captive chicken societies. Over the past century, dominance hierarchies have been studied across a wide range of animal species including canines [3, 20], birds [57], fish [13] and humans [10, 24, 37], uncovering commonalities in the structure and function of dominance hierarchies across taxa [8]. The study of dominance hierarchies continues to be a field of active research [51].

### 3 Related Work

MAS models have been used for the modelling of dominance hierarchies in animals [22] and in humans [40]. *DomWorld* [27] is a multi-agent environment developed to study the self-organization of social behavior and dominance hierarchies in primate groups. DomWorld is not modelled as a stochastic game or any other game theoretic construct, as the agents have no concept of a reward signal. Leimar [32] provides a proper MARL environment called *SocDom*, equipping the agents with an RL algorithm [52], albeit without neural network estimators. While the SocDom experiments show similar results to ours, the foci of the two environments differ: SocDom, like DomWorld, attempts to tailor the agents’ behaviors to faithfully reproduce the dominance behavior seen in real animal societies. In contrast, our research is directed at boiling the mechanics of dominance hierarchies down to the simplest possible set of premises, showing that they emerge even when using an off-the-shelf RL algorithm.

The use of dominance hierarchies for improving the performance of artificial agents was proposed in Hemelrijk [26], and in Tomlinson et al. [53], which suggested that dominance hierarchies could streamline negotiation, enable agents to form beneficial alliances, and make interfaces more intuitive. Bakker et al. [2] employs a group of RL agents with a graph structure, which bears resemblance to a dominance hierarchy, besides being explicitly defined rather than emergent. Aroca-Ouellette et al. [1] demonstrates an explicit hierarchy where some agents have a different set of actions than others.

Vezhnevets et al. [55] and Wu et al. [60] replace the RL algorithms used by the agents with large language models (LLMs), which have gained immense popularity due to their robustness in solving a wide variety of problems [61]. These LLM-based agents are organized in human-inspired hierarchies such as manager-workers and student-assistant-expert. The back-and-forth interaction between multiple LLM agents is effective at encouraging diversity [35], factuality, and reasoning capability [16], when compared to the use of a single LLM-based agent. The potential usefulness of such agents is tremendous, as LLMs can be seamlessly integrated with multimodal foundation models [33], allowing them to process and generate textual, visual and auditory content. We suggest that a formal understanding and terminology of agent hierarchies could play a decisive role in maximizing the effectiveness of these agents.

## 4 Definitions

### 4.1 Dominance between two agents

Dominance hierarchies are a group phenomenon comprised of multiple dominance relationships, one between each pair of agents in the group. In this section we consider only those pairwise interactions; we define dominance relationships, the metrics of aggressiveness and rapport, the roles of dominant and subordinate, and the environment in which these occur.

Animals use dominance relationships to decide which individual will have access to an exclusive resource, such as food, a mating partner, or grooming by other individuals [28]. We model the environment that enables these interactions between agent  $i$  and agent  $j$  as a partially-observable stochastic game (POSG) [25] which has exactly two stable Nash equilibria  $NE_i$  and  $NE_j$  such that agent  $i$ 's reward at  $NE_i$  is bigger than its reward at  $NE_j$ , and agent  $j$ 's reward at  $NE_j$  is bigger than its reward at  $NE_i$ .

We denote the actions that comprise  $NE_i$  as  $a_{NE_i}^i$  and  $a_{NE_i}^j$ .

Intuitively, a dominance relationship between two agents is the tendency of those agents to play the joint action of either  $NE_i$  or  $NE_j$ , but not both. Similarly to Leibo et al. [31], we define an agent's aggressiveness by how frequently it chooses the action that reduces the other agent's reward:

**Definition 1 (Aggressiveness).**

*Given an agent  $i$  and a set of timesteps  $\mathbb{T}$ , the agent's aggressiveness  $g_i^\mathbb{T}$  is the portion of timesteps in which it played  $a_{NE_i}^i$ .*

**Definition 2 (Dominance relationship, dominant, subordinate).** *Given two agents  $i$  and  $j$  and a set of timesteps  $\mathbb{T}$  in which they played with each other, if the difference between agent  $i$ 's aggressiveness and agent  $j$ 's aggressiveness is above a certain threshold, we say that the two agents are in a dominance relationship (DR), with agent  $i$  being dominant and agent  $j$  being subordinate:*

$$i \rightarrow j \quad \text{iff} \quad g_i^\mathbb{T} - g_j^\mathbb{T} > \eta, \quad \eta \in (0, 1]$$

In order to determine the existence of a dominance relationship between two agents without assuming its polarity, we introduce a metric we call *rapport*:

**Definition 3 (Rapport).**

For two agents  $i$  and  $j$  and timesteps  $\mathbb{T}$ , we define the *rapport*  $\mathfrak{R}_{i,j}^{\mathbb{T}}$  as:

$$\mathfrak{R}_{i,j}^{\mathbb{T}} = |g_i^{\mathbb{T}} - g_j^{\mathbb{T}}|$$

A rapport  $\mathfrak{R}_{i,j}^{\mathbb{T}} > \eta$  indicates either  $i \rightarrow j$  or  $i \leftarrow j$ .

In the classic game of Chicken [44], each of the two agents plays either *hawk* or *dove*. An agent gets the most reward if it plays *hawk* while its partner plays *dove*; however, if both agents play *hawk*, they both get the lowest reward.

**Definition 4 (The game of Chicken).** *The game of Chicken is defined as a normal-form game:*

		Agent 2	
		Dove	Hawk
Agent 1	Dove	$R, R$	$S, T$
	Hawk	$T, S$	$P, P$

The reward constants satisfy  $T > R > S > P$ . The  $(S, T)$  and  $(T, S)$  outcomes serve as the Nash equilibria  $NE_i$  and  $NE_j$ . Each agent's aggressiveness is simply its tendency to play hawk.

## 4.2 Dominance hierarchies

To model dominance relationships within an agent population, we generalize the game of Chicken to support more than two players:

**Definition 5 (Chicken Coop).** *Chicken Coop is an  $N$ -player generalization of the game of Chicken. In each episode of Chicken Coop, the agents are divided into random pairs that are used for the entire episode. Each pair of agents plays one round of Chicken against each other, choosing either hawk or dove and receiving a reward in  $\{T, R, S, P\}$ . Each agent's sole observation is the identity of their opponent.*

We extend the definitions in Section 4.1 from the Chicken game to the Chicken Coop game. For example, for each pair of agents  $i$  and  $j$  in the population, we determine the existence and polarity of their dominance relationship ( $i \rightarrow j$  or  $i \leftarrow j$ ) by evaluating the aggressiveness metrics  $g_i^{\mathbb{T}}$  and  $g_j^{\mathbb{T}}$  on the subset of  $\mathbb{T}$  in which these two agents were randomly paired with each other. Finally, we aggregate these dominance relationships into a graph:

**Definition 6 (Dominance hierarchy).**

A dominance hierarchy  $\mathcal{H}$  is a complete, directed graph (tournament) where agents are represented as nodes and dominance relationships are represented as directed edges.

## 5 Methods

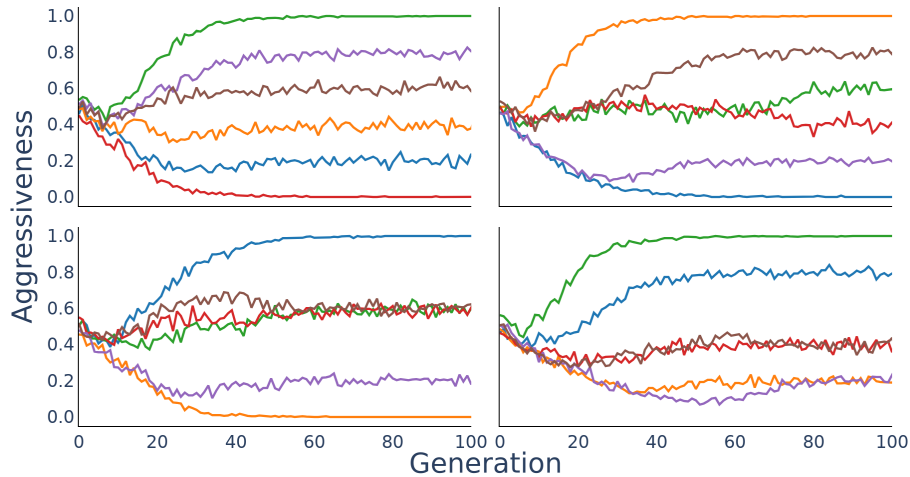
In all of the experiments below, unless stated otherwise, the RL agents are trained using the PPO algorithm [49] implementation provided by the RLlib framework [34]. We train  $L = 300$  populations of  $N = 6$  Chicken Coop agents, at a learning rate of  $2 \times 10^{-6}$ , a discount factor of  $\gamma = 0.99$ , and a clipping parameter of  $\epsilon = 0.3$ . We use a dominance threshold of  $\eta = 0.55$  and these reward constants:  $R = 0$ ,  $S = -2$ ,  $T = 5$ ,  $P = -10$ .

Each agent has its own neural network with its own set of weights, which are separate from those of the other agents. Crucially, all of the agents in a given population are trained exclusively with the other agents in that population; we don't allow cross-population training, because we would like to allow each population to develop its own distinct dominance hierarchy.

Each generation is comprised of 512 episodes. After each generation, the policies update according to the actions and rewards in that generation.

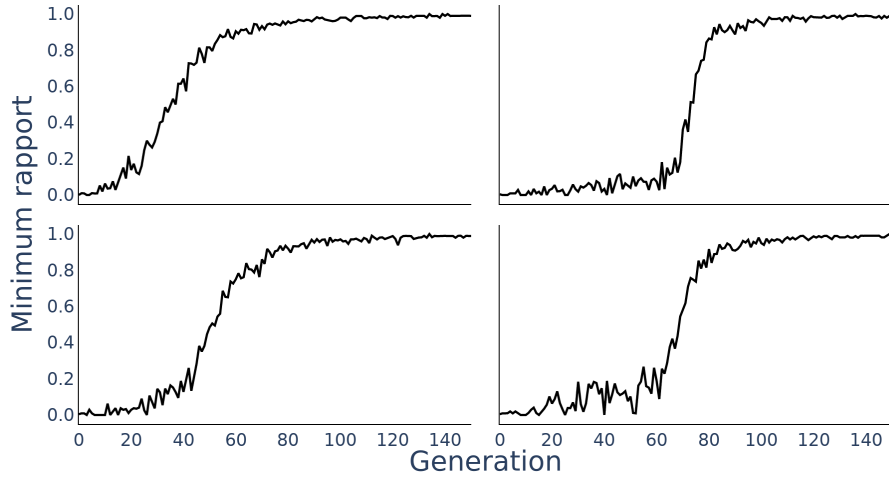
## 6 Results

### 6.1 Emergence of dominance hierarchies



**Fig. 1.** Aggressiveness levels of each of the six agents (depicted as differently-colored lines) in four sample Chicken Coop populations.

We measure the aggressiveness of each agent in each population, aggregated interactions with each of the other agents. In Figure 1 we plot the aggressiveness of six agents in four sample populations. A few features can be visually discerned:

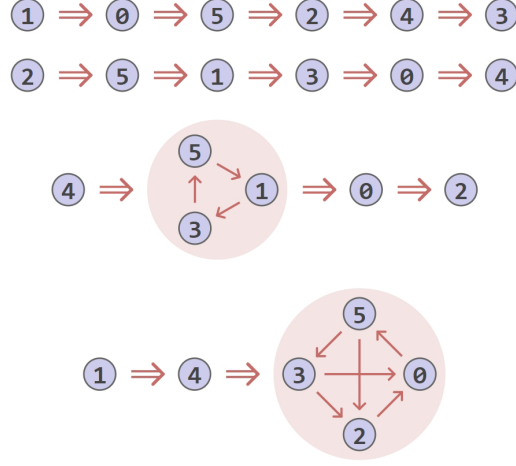


**Fig. 2.** Minimum rapport between each pair of the six agents in four sample Chicken Coop populations.

1. The agents’ aggressiveness values converge to approximately equally-spaced “ranks” in  $[0, 1]$ .
2. Despite of this strong convergence, the mapping between agents and ranks appears completely arbitrary across the four sample populations.
3. In some populations, the mapping between agents and ranks is a bijection, while in other populations, multiple agents converge to the same rank, leaving other ranks empty.

We offer the following explanation for these observations: when agent  $i$ ’s aggressiveness approaches  $k/N$ , agent  $i$  is playing a high proportion of *hawk* towards  $k$  other agents, which in turn play a high proportion of *dove* towards agent  $i$ . Each of these  $k$  agents contributes  $1/N$  to agent  $i$ ’s aggressiveness. According to Definition 2, agent  $i$  is dominating these  $k$  agents. The remaining  $N - k - 1$  agents dominate agent  $i$ . We verify this explanation by measuring the minimum rapport between each pair of agents for each of the four sample populations, observing that each agent develops a dominance relationship with each of the other agents in its population. (Figure 2.)

We aggregate these dominance relationships into dominance hierarchies, showing the results for our four sample populations in Figure 3. We observe that some of the dominance hierarchies form a perfect line, while others contain cycles. This distinction has been studied in animal societies. The former are called *linear dominance hierarchies* or *transitive dominance hierarchies*. The latter are called *nonlinear dominance hierarchies*, *near-linear dominance hierarchies* or *intransitive dominance hierarchies* [9, 14, 15, 47]. In linear dominance hierarchies, the agents may be assigned ranks  $\{0, 1, 2, \dots, N - 1\}$  with rank 0 representing the agent that dominates all other agents, rank 1 representing the agent that dominates all other agents except that in rank 0, etc.



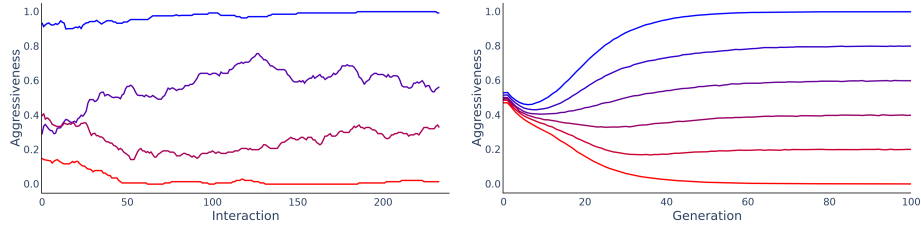
**Fig. 3.** Dominance hierarchies of four sample Chicken Coop populations, visualized in condensed form. “ $\rightarrow$ ” represents a dominance relationship. “ $\Rightarrow$ ” means that all agents to the left of the double arrow dominate all of the agents to its right.

One profound property of the agents’ learned behavior is its idiosyncraticity: (1) it converges to a stable dominance hierarchy, and (2) this hierarchy is not a single one, but rather varies between different trials. Our 300 populations converged to a total of 248 different dominance hierarchies (out of  $2^{\frac{N(N-1)}{2}} = 2^{15}$  possible permutations). Most dominance hierarchies appeared in only one population, while the most common dominance hierarchies appeared in 3 populations. Despite the arbitrariness of that choice, each population converges strongly to its chosen dominance hierarchy.

## 6.2 Comparison to animal behavior

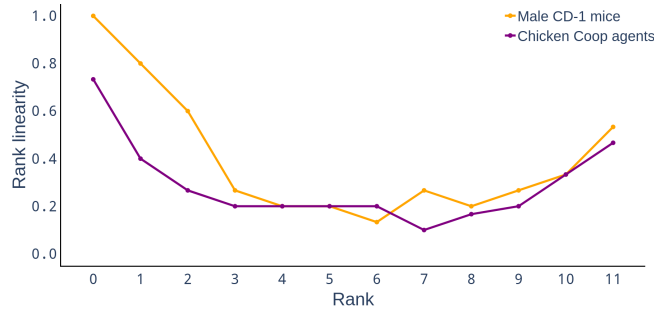
In Figure 4 we visually compare our aggressiveness results to empirical data from groups of actual chickens. In Chase et al. [8], 14 captive populations of 4 white Leghorn hens each were observed for aggressive behavior for two consecutive days. While there are considerable differences in the methodologies of that experiment and our study, the property of idiosyncraticity holds true in both of them. In both environments, the agents tend to maintain their position, regardless of how high or low it is in the hierarchy.

We visually compare the occurrence of intransitive components between Chicken Coop populations and mice populations, shown in Figure 5. In Williamson et al. [58], 20 captive populations of 12 male CD-1 mice each were observed for aggressive behavior for 22 days. We compare these results to those from  $L = 30$  populations of  $N = 12$  Chicken Coop agents. In order to find the maximum similarity between the natural and artificial populations, we use a learning rate of  $3 \times 10^{-5}$ . We make this comparison by using a metric we define here as *rank*



**Fig. 4.** **Left:** Mean aggressiveness levels of four Leghorn chickens in fourteen captive populations [8]. **Right:** Mean aggressiveness levels of six Chicken Coop agents across 300 populations, filtered to include only those that developed a linear dominance hierarchy. In both plots, the results are averaged using the agents’ ranks as their identity.

*linearity*, which expresses how likely each dominance rank is to be occupied by a single agent, rather than a  $K$ -way tie. For a single population, we can describe rank linearity as a boolean; for example, in the third population from the top in Figure 3, ranks  $\{0, 4, 5\}$  are linear, while ranks  $\{1, 2, 3\}$  are non-linear as they are part of a 3-way tie between agents 1, 3 and 5. For a set of populations, we define rank linearity as a number in  $[0, 1]$  which is the probability of a rank being linear in a sample population.



**Fig. 5.** The rank linearities of 12 different ranks, in both Chicken Coop agents (purple) and male CD-1 mice (orange)[58]. In both experiments, the top and bottom ranks were the least likely to be occupied by an intransitive component.

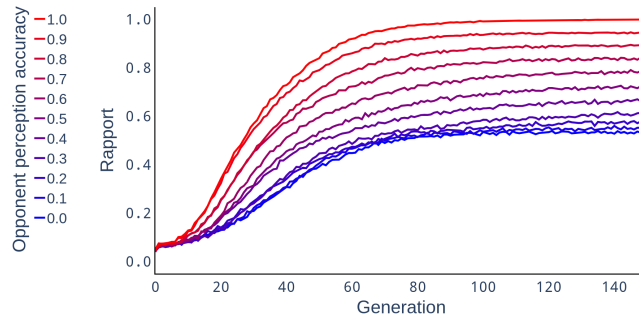
### 6.3 Observation ablation

Here we show a causal relationship between agents’ ability to identify their opponents and their tendency to form dominance hierarchies. In the Chicken Coop environment, agents observe the identity of their opponent as an index number

$0 \leq i \leq N - 1$ . We add random noise to the agents’ observation, up to and including its complete obfuscation, and measure the rapport in the population.

**Methodology:** We define an agent’s *opponent perception accuracy* (OPA) as a number in  $[0, 1]$ . For each agent in each episode, the probability that its observation of its opponent is replaced with a random value is  $(1 - OPA)$ . For each OPA value in  $\{0, 0.1, \dots, 1\}$  we train  $L = 10$  different populations of  $N = 6$  Chicken Coop agents at a learning rate of  $2 \times 10^{-6}$  and measure their rapport.

In Figure 6 we show a positive correlation between the agents’ OPA and their rapport, and by extension, their tendency to form dominance hierarchies.



**Fig. 6.** The gradual rise of rapport in Chicken Coop populations, differentiated by how accurately the agents identify their opponents. The bluer lines represent bigger amounts of random noise added to the agents’ observation signal.

#### 6.4 Transmitting dominance hierarchies to new populations

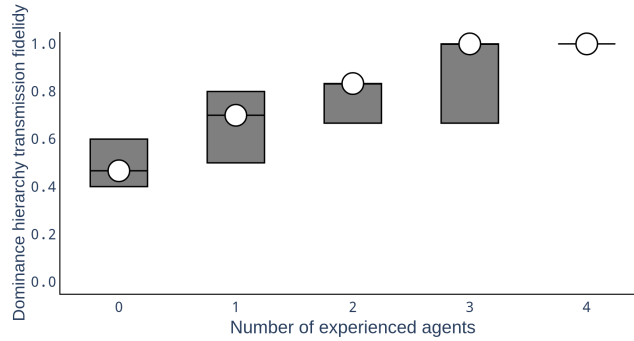
In previous sections we regarded Chicken Coop agents as the focal objects of study, while we considered the dominance hierarchies that they form as a pattern in their behavior. We now propose flipping this perspective on its head: We consider dominance hierarchies as the focal object while the agents are used as vessels that communicate the dominance hierarchy that they learned to a new population, promoting a cultural evolution paradigm in artificial intelligence [6].

This transmission of dominance hierarchies to new populations is made possible by two properties of dominance hierarchies: **(1)** the knowledge of each dominance hierarchy is replicated over all the agents that participate in it, and **(2)** each agent learns to consistently play *hawk* with all its subordinates, and this tendency serves as a punishing mechanism. Any errant attempt by a subordinate to play *hawk* against a dominant would usually result in a *hawk/hawk* outcome, with both agents receiving the lowest possible reward of  $P$ .

**Methodology:** We train  $L = 10$  different populations of  $N = 6$  Chicken Coop agents until they converge to dominance hierarchies. We call these *experienced populations*. For each of these  $L$  populations, we perform the following

procedure  $M = 30$  times: for each  $K \in \{0, 1, 2, \dots, N - 2\}$  we choose a random sample of  $K$  of those agents to be our *experienced agents*. We transplant the  $K$  experienced agents into a new population with  $N - K$  *naive agents* [17] that have not been trained. Finally, we train the naive agents in the heterogeneous population at a learning rate of  $2 \times 10^{-5}$  and observe their dominance behavior.

To measure how close each new dominance hierarchy is to its original dominance hierarchy, we define a *restricted dominance hierarchy distance* (RDHD). The restricted distance between two dominance hierarchies is the portion of dominance relationships, limited to those between two naive agents, that have the same polarity in both dominance hierarchies, satisfying  $0 \leq RDHD \leq 1$ . We define the *dominance hierarchy transmission fidelity* (DHTF) as  $(1 - RDHD)$ . Intuitively, DHTF is a measure of how effectively the experienced agents passed their dominance hierarchy on to the naive population.



**Fig. 7.** The DHTF of 300 different groups of experienced agents transplanted into 300 different naive populations. Higher values mean that the naive agents learned a dominance hierarchy similar to the one in the original experienced population. Circles indicate median DHTF while boxes indicate the two middle quartiles.

In Figure 7 we show the distribution of the DHTFs for different values of  $K$ . A mere  $K = 2$  experienced agents are able to teach 4 naive agents their original dominance hierarchy with 80% median fidelity.

## 7 Discussion and future work

This paper investigates dominance hierarchies in the context of MARL. It first provides a formal definition and a game-theoretical representation of dominance hierarchies. It then introduces the phenomenon of dominance hierarchies in a simplified form using the Chicken Coop environment, stripping away many of the complexities found in biological life. Finally, it demonstrates how populations of RL agents can invent, learn, enforce, and transmit dominance hierarchies.

Preliminary experiments with higher learning rates result in unstable dominance hierarchies, where agents seem to converge strongly to a certain rank for

some number of generations, only to deviate from it and converge to a different rank. This is another unique behavior that exists in natural environments, where it is known as *rank change* [46, 51] or *dynamic stability* [8].

We suggest several extensions to the Chicken Coop environment:

**Multiple Hierarchies:** In the Chicken Coop environment, all the agents participate in one dominance hierarchy that includes all of the agents. In animal and human societies, individuals participate in multiple dominance hierarchies for multiple groups of other individuals, holding different ranks in each one.

**Societies with Mixed Incentives:** In this environment, the agents are fully dedicated to making decisions on how aggressive they’ll be to other agents. In animal and human societies, individuals in a group cooperate towards shared goals in addition to their internal dominance struggles.

**Opponent Shaping (OS):** Algorithms such as LOLA [21] and M-FOS [36] enable agents to take the learning process of the other agents into account in their own policy gradients. When applied to an environment that supports dominance hierarchies, they may promote second-order dominance-seeking strategies, e.g., they may prompt agent  $i$  to consider how it can behave as to encourage agent  $j$  to place agent  $i$  at a high rank in the dominance hierarchy.

We propose that the study of dominance hierarchies between RL agents may be combined with research on large language models (LLMs) and multimodal foundation models (MFMs), which have gained immense popularity due to their robustness in solving a wide variety of problems [33, 61]. Tufano et al. [54] presents an innovative way to use LLMs: instead of sending queries to a single LLM, users converse with a group of LLM-based agents, organized in human-inspired hierarchies such as manager-workers and assistant-expert. The interaction between multiple LLM agents is effective at encouraging diversity [35], factuality, and reasoning capability [16], when compared to a single LLM-based agent. We suggest that the augmenting these LLM-based agents with RL or OS algorithms, specifically to maintain the social relationships between the agents, may result in hierarchies that are similar to those observed in biological life.

When humans work on problems as a group, we balance in-group intrigues with external pressures to provide good results. We suggest that this interplay between individual needs and group needs may play a crucial role in the success of our collective intelligence. We hypothesize that MFM-based agents operating under similar conditions may make real-world decisions in a way that is more interpretable and corrigible, as human operators may recognize that the agents’ decision process reflects their own.

## 8 Acknowledgements

We thank the following colleagues for their advice and support in writing this paper: David Aha, Ivan Chase, Edgar Duéñez-Guzmán, Errol King, Joel Leibo, Olof Leimar, David Manheim, Markov, Georg Ostrovski, Jérémy Perret, Saul Pwanson, Venkatesh Rao and Eli Strauss. This research was funded by The Association For Long Term Existence And Resilience (ALTER).

## Bibliography

- [1] Aroca-Ouellette, S., Aroca-Ouellette, M., Biswas, U., Kann, K., Roncone, A.: Hierarchical reinforcement learning for ad hoc teaming. In: Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, pp. 2337–2339 (2023)
- [2] Bakker, M.A., Everett, R., Weidinger, L., Gabriel, I., Isaac, W.S., Leibo, J.Z., Hughes, E.: Modelling cooperation in network games with spatio-temporal complexity (2021)
- [3] Bonanni, R., Cafazzo, S., Abis, A., Barillari, E., Valsecchi, P., Natoli, E.: Age-graded dominance hierarchies and social tolerance in packs of free-ranging dogs. *Behavioral Ecology* **28**(4), 1004–1020 (2017)
- [4] Bowles, S.: Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* **324**(5932), 1293–1298 (2009)
- [5] Boyd, R., Richerson, P.J.: Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**(1533), 3281–3288 (2009)
- [6] CGI Team: Learning robust real-time cultural transmission without human data. arXiv preprint arXiv:2203.00715 (2022)
- [7] Chase, I., Bartolomeo, C., Dugatkin, L.: Aggressive interactions and inter-contest interval: how long do winners keep winning?. *Animal Behaviour* **48**(2), 393–400 (1994)
- [8] Chase, I., Coelho, D., Lee, W., Mueller, K., Curley, J.: Networks never rest: an investigation of network evolution in three species of animals. *Social Networks* **68**, 356–373 (2022)
- [9] Chase, I., Tovey, C., Spangler-Martin, D., Manfredonia, M.: Individual differences versus social dynamics in the formation of animal dominance hierarchies. *Proceedings of the National Academy of Sciences* **99**(8), 5744–5749 (2002)
- [10] Chen Zeng, T., Cheng, J.T., Henrich, J.: Dominance in humans. *Philosophical Transactions of the Royal Society B* **377**(1845), 20200451 (2022)
- [11] Couzin, I.D.: Collective cognition in animal groups. *Trends in cognitive sciences* **13**(1), 36–43 (2009)
- [12] Dafoe, A., Hughes, E., Bachrach, Y., Collins, T., McKee, K.R., Leibo, J.Z., Larson, K., Graepel, T.: Open problems in cooperative ai (2020)
- [13] David, B.O., Stoffels, R.J.: Spatial organisation and behavioural interaction of giant kokopu (*galaxias argenteus*) in two stream pools differing in fish density. *New Zealand Journal of Marine and Freshwater Research* (2003)
- [14] De Vries, H.: An improved test of linearity in dominance hierarchies containing unknown or tied relationships. *Animal Behaviour* **50**(5), 1375–1389 (1995)
- [15] De Vries, H.: Finding a dominance order most consistent with a linear hierarchy: a new procedure and review. *Animal behaviour* **55**(4), 827–843 (1998)

- [16] Du, Y., Li, S., Torralba, A., Tenenbaum, J.B., Mordatch, I.: Improving factuality and reasoning in language models through multiagent debate. arXiv preprint arXiv:2305.14325 (2023)
- [17] Duéñez-Guzmán, E.A., McKee, K.R., Mao, Y., Coppin, B., Chiappa, S., Vezhnevets, A.S., Bakker, M.A., Bachrach, Y., Sadedin, S., Isaac, W., et al.: Statistical discrimination in learning agents. arXiv preprint arXiv:2110.11404 (2021)
- [18] Dugatkin, L.A.: Winner and loser effects and the structure of dominance hierarchies. *Behavioral Ecology* **8**(6), 583–587 (1997)
- [19] Dugatkin, L.A.: Bystander effects and the structure of dominance hierarchies. *Behavioral Ecology* **12**(3), 348–352 (2001)
- [20] Essler, J., Cafazzo, S., Marshall-Pescini, S., Virányi, Z., Kotrschal, K., Range, F.: Play behavior in wolves: Using the ‘50:50’ rule to test for egalitarian play styles. *PLoS One* **11**(5), e0154150 (2016)
- [21] Foerster, J., Chen, R.Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., Mordatch, I.: Learning with opponent-learning awareness. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, p. 122–130, AAMAS ’18, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2018)
- [22] Gavrillets, S., Duéñez-Guzmán, E.A., Vose, M.D.: Dynamics of alliance formation and the egalitarian revolution. *PLoS One* **3**(10), e3293 (2008)
- [23] Gordon, D.M.: The organization of work in social insect colonies. *Nature* **380**(6570), 121–124 (1996)
- [24] Halevy, N., Chou, E.Y., Cohen, T.R., Livingston, R.W.: Status conferral in intergroup social dilemmas: behavioral antecedents and consequences of prestige and dominance. *Journal of personality and social psychology* **102**(2), 351 (2012)
- [25] Hansen, E.A., Bernstein, D.S., Zilberstein, S.: Dynamic programming for partially observable stochastic games. In: AAAI, vol. 4, pp. 709–715 (2004)
- [26] Hemelrijk, C.: Dominance interactions, spatial dynamics and emergent reciprocity in a virtual world. In: The 4th international conference on simulation of adaptive behavior, vol. 4, pp. 545–552, Citeseer (1996)
- [27] Hemelrijk, C.: Social phenomena emerging by self-organisation in a competitive, virtual world (“domworld”). In: Learning to behave Workshop II: Internalising knowledge. Ieper, Belgium, pp. 11–19 (2000)
- [28] Henazi, S.P., Barrett, L.: The value of grooming to female primates. *Primates* **40**, 47–59 (1999)
- [29] Iverson, G.J., Sade, D.S.: Statistical issues in the analysis of dominance hierarchies in animal societies. *Journal of Quantitative Anthropology* **2**(1), 61–83 (1990)
- [30] Kraus, S.: Negotiation and cooperation in multi-agent environments. *Artificial Intelligence* **94**(1), 79–97 (1997), ISSN 0004-3702, [https://doi.org/https://doi.org/10.1016/S0004-3702\(97\)00025-8](https://doi.org/https://doi.org/10.1016/S0004-3702(97)00025-8), economic Principles of Multi-Agent Systems
- [31] Leibo, J.Z., Zambaldi, V., Lanctot, M., Marecki, J., Graepel, T.: Multi-agent reinforcement learning in sequential social dilemmas. In: Proceedings

- of the 16th Conference on Autonomous Agents and MultiAgent Systems, p. 464–473, AAMAS '17, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2017)
- [32] Leimar, O.: The evolution of social dominance through reinforcement learning. *The American Naturalist* **197**(5), 560–575 (2021)
  - [33] Li, C., Gan, Z., Yang, Z., Yang, J., Li, L., Wang, L., Gao, J.: Multimodal foundation models: From specialists to general-purpose assistants (2023)
  - [34] Liang, E., Liaw, R., Moritz, P., Nishihara, R., Fox, R., Goldberg, K., Gonzalez, J.E., Jordan, M.I., Stoica, I.: Rllib: Abstractions for distributed reinforcement learning (2018)
  - [35] Liang, T., He, Z., Jiao, W., Wang, X., Wang, Y., Wang, R., Yang, Y., Tu, Z., Shi, S.: Encouraging divergent thinking in large language models through multi-agent debate. arXiv preprint arXiv:2305.19118 (2023)
  - [36] Lu, C., Willi, T., De Witt, C.A.S., Foerster, J.: Model-free opponent shaping. In: International Conference on Machine Learning, pp. 14398–14411, PMLR (2022)
  - [37] Magee, J.C., Galinsky, A.D.: Social hierarchy: The self-reinforcing nature of power and status. *The Academy of Management Annals* **2**, 351–398 (2008)
  - [38] Mordatch, I., Abbeel, P.: Emergence of grounded compositional language in multi-agent populations. In: Proceedings of the AAAI conference on artificial intelligence, vol. 32 (2018)
  - [39] Muthukrishna, M., Henrich, J.: Innovation in the collective brain. *Ph. Transactions of the Royal Society B: Biological Sciences* **371**(1690) (2016)
  - [40] Przepiorka, W., Rutten, C., Buskens, V., Szekely, A.: How dominance hierarchies emerge from conflict: A game theoretic model and experimental evidence. *Social science research* **86**, 102393 (2020)
  - [41] Radke, D., Larson, K., Brecht, T., Tilbury, K.: Towards a better understanding of learning with multiagent teams. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, pp. 271–279, ijcai.org (2023), <https://doi.org/10.24963/ijcai.2023/31>
  - [42] Ramchurn, S.D., Mezzetti, C., Giovannucci, A., Rodriguez-Aguilar, J.A., Dash, R.K., Jennings, N.R.: Trust-based mechanisms for robust and efficient task allocation in the presence of execution uncertainty. *Journal of Artificial Intelligence Research* **35**, 119–159 (2009)
  - [43] Rapoport, A.: Outline of a probabilistic approach to animal sociology: I. *The bulletin of mathematical biophysics* **11**, 183–196 (1949)
  - [44] Rapoport, A., Chammah, A.M.: The game of chicken. *American Behavioral Scientist* **10**(3), 10–28 (1966)
  - [45] Reeves, B., Nass, C.: The media equation: How people treat computers, television, and new media like real people. Cambridge, UK **10**(10) (1996)
  - [46] Samuels, A., Silk, J.B., Altmann, J.: Continuity and change in dominance relations among female baboons. *Animal Behaviour* **35**(3), 785–793 (1987)
  - [47] Sanders, N.J., Gordon, D.M.: Resource-dependent interactions and the organization of desert ant communities. *Ecology* **84**(4), 1024–1031 (2003)
  - [48] Schjelderup-Ebbe, T.: Beiträge zur sozialpsychologie des haushuhns. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane. Abt. 1. Zeitschrift für Psychologie* **88**, 225 (1922)

- [49] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017)
- [50] Smith, J.M., Price, G.R.: The logic of animal conflict. *Nature* **246**(5427), 15–18 (1973)
- [51] Strauss, E.D., Shizuka, D.: The dynamics of dominance: open questions, challenges and solutions. *Philosophical Transactions of the Royal Society B* **377**(1845), 20200445 (2022)
- [52] Sutton, R.S., Barto, A.G.: *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA (2018), ISBN 0262039249
- [53] Tomlinson, B., Blumberg, B., Rhodes, B.: How is an agent like a wolf?: Dominance and submission in multi-agent systems. In: *International ICSC Symposium on Multi-Agents and Mobile Agents in Virtual Organizations and E-Commerce (MAMA’2000)* (2000)
- [54] Tufano, M., Agarwal, A., Jang, J., Moghaddam, R.Z., Sundaresan, N.: Autodev: Automated ai-driven development (2024)
- [55] Vezhnevets, A.S., Agapiou, J.P., Aharon, A., Ziv, R., Matyas, J., Duéñez-Guzmán, E.A., Cunningham, W.A., Osindero, S., Karmon, D., Leibo, J.Z.: Generative agent-based modeling with actions grounded in physical, social, or digital space using concordia (2023)
- [56] Vinitzky, E., Köster, R., Agapiou, J.P., Duéñez-Guzmán, E.A., Vezhnevets, A.S., Leibo, J.Z.: A learning agent that acquires social norms from public sanctions in decentralized multi-agent settings. *Collective Intelligence* **2**(2) (2023), <https://doi.org/10.1177/26339137231162025>
- [57] Watson, J.R.: Dominance-subordination in caged groups of house sparrows. *The Wilson Bulletin* pp. 268–278 (1970)
- [58] Williamson, C.M., Lee, W., Romeo, R.D., Curley, J.P.: Social context-dependent relationships between mouse dominance rank and plasma hormone levels. *Physiology & behavior* **171**, 110–119 (2017)
- [59] Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., Malone, T.W.: Evidence for a collective intelligence factor in the performance of human groups. *Science* **330**(6004), 686–688 (2010)
- [60] Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A.H., White, R.W., Burger, D., Wang, C.: Autogen: Enabling next-gen llm applications via multi-agent conversation (2023)
- [61] Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Cheng, W., Zhang, Q., Qin, W., Zheng, Y., Qiu, X., Huang, X., Gui, T.: The rise and potential of large language model based agents: A survey (2023)
- [62] Xu, Y., Deng, Z., Wang, M., Xu, W., So, A.M.C., Cui, S.: Voting-based multi-agent reinforcement learning for intelligent iot (2020)
- [63] Zhan, S., Bendapudi, N., Hong, Y.y.: Re-examining diversity as a double-edged sword for innovation process. *Journal of Organizational Behavior* **36**(7), 1026–1049 (2015)