JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 29, Number 0, 2022

Pp. 1–9
DOI: 10.1089/cmb.2021.0549

Research Article

Sharing Rewards Undermines Coordinated Hunting

MINGLU ZHAO,[1] NING TANG,[1] ANNYA L. DAHMANI,[2] YIXIN ZHU,[1]
FEDERICO ROSSANO,[3] and TAO GAO[1,4]

ABSTRACT

Coordinated hunting is widely observed in animals, and sharing rewards is often considered a major incentive for its success. While current theories about the role played by sharing in coordinated hunting are based on correlational evidence, we reveal the causal roles of sharing rewards through computational modeling with a state-of-the-art Multi-agent Reinforcement Learning (MARL) algorithm. We show that counterintuitively, while selfish agents reach robust coordination, sharing rewards undermines coordination. Hunting coordination modeled through sharing rewards (1) suffers from the free-rider problem, (2) plateaus at a small group size, and (3) is not a Nash equilibrium. Moreover, individually rewarded predators outperform predators that share rewards, especially when the hunting is difficult, the group size is large, and the action cost is high. Our results shed new light on the actual importance of prosocial motives for successful coordination in nonhuman animals and suggest that sharing rewards might simply be a byproduct of hunting, instead of a design strategy aimed at facilitating group coordination. This also highlights that current artificial intelligence modeling of human-like coordination in a group setting that assumes rewards sharing as a motivator (e.g., MARL) might not be adequately capturing what is truly necessary for successful coordination.

Keywords: animal coordination, collective hunting, free-rider problem, multi-agent reinforcement learning, sharing reward.

1. INTRODUCTION

Coordinated hunting has been widely observed in the animal kingdom in many different species, such as wolves (MacNulty et al., 2014), hyenas (Holekamp et al., 1997), dolphins (Gazda et al., 2005), ravens (Yosef and Yosef, 2010), and hawks (Bednarz, 1988), whereas the majority of in-depth discussion on coordination mechanisms focuses on chimpanzee behavior. Chimpanzees hunt for meat in all known populations, with the red colobus monkeys being the primary prey where both species exist (Newton-Fisher,

Departments of [1]Statistics and [2]Psychology, University of California Los Angeles, Los Angeles, California, USA.
[3]Department of Cognitive Science, University of California San Diego, La Jolla, California, USA.
[4]Department of Communication, University of California Los Angeles, Los Angeles, California, USA.

2007). Anthropological studies based on field observations suggest that chimpanzees exhibit sophisticated human-like cooperation, such as playing complementary roles during hunting, which includes drivers, blockers, chasers, and ambushers (Boesch, 2005). Consequently, understanding the motivation of such coordinated behavior can provide insights on understanding the evolutionary history of human cooperation.

Sharing rewards has been considered as a major incentive for animals' success in coordinated hunting, especially for chimpanzees. It seems to encourage participation in group hunting, which leads to higher hunting success (Samuni et al., 2018). Moreover, the literature claims that sharing rewards further contributes to chimpanzee bonding through reciprocity (Nishida et al., 1992; Mitani and Watts, 2001), reducing begging harassment (Gilby, 2006; John et al., 2019), exchanging meat for sex (Nishida et al., 1992), and securing dominance (Nishida et al., 1992).

However, since existing animal studies are mostly observational, they can only indicate a correlation, while the causal effects of sharing rewards on coordination remain unclear due to the lack of causal evidence from formal experimental manipulations. In fact, it has been argued that sharing rewards is neither necessary nor sufficient for coordinated hunting. It could be unnecessary, since chimpanzee coordinated hunting may not be based on any expectation of reward sharing, and rather be mainly driven by individual selfish interests (Tomasello et al., 2005). In such a case, during hunting, each individual chimpanzee simply takes actions to maximize its self-interest based on other agents' locations. Sharing rewards could be insufficient to support coordinated behavior as well. One critical challenge for coordination is the free-rider problem (Olson, 1989): Rational individuals benefit from the shared public goods even if they do not pay individual action costs. However, the free-rider problem has not been highlighted in the above observational studies. Theories of human cooperation suggest that other cognitive infrastructures are necessary to solve this issue, including cheater detection and punishment (Clutton-Brock and Parker, 1995), commitment (Gilbert, 1999), and fairness (Arneson, 1982). Invoking these normative and moral concepts in coordinated hunting requires a stronger definition of cooperation beyond sharing rewards (Tomasello, 2009).

With the aforementioned conflicting evidence and viewpoints from observational studies, here we study the causal effects of sharing rewards on the performance of coordinated hunting from a modeling perspective. Several studies have investigated the ultimate mechanisms explaining cooperation focusing on the consequences at the population level (Boyd et al., 2003; Nowak, 2006; Gross and De Dreu, 2019). In this study, we focus on the proximate mechanisms (Mayr, 1961; Tinbergen, 1963) and study the causal effects of sharing rewards on coordinated hunting by modeling individual agents' actions (Bonabeau, 2002) using reinforcement learning (RL). RL is a prominent model for animal learning with deep roots in psychology and neuroscience (Liu et al., 2021; Peterson et al., 2021). Through trial and error, RL approximates the optimal action policy through maximizing long-term expected rewards (Sutton and Barto, 2018). RL is also a state-of-the-art artificial intelligence model, which, combined with deep neural networks (LeCun et al., 2015), is able to generate complex intelligent behaviors, reaching human expert-level performance in games like Atari (Mnih et al., 2015) and Go (Silver et al., 2016, 2017).

Multi-agent reinforcement learning (MARL), as an extension of RL, is consistent with the empiricist theory that social skills are accumulated through trial and error (Hayek, 2011). It has been successfully applied to various challenging group coordination scenarios, such as autonomous driving coordination (Shalev-Shwartz et al., 2016) as well as teaming in Dota 2 (Berner et al., 2019) and StarCraft (Vinyals et al., 2019). MARL offers a generic solution to different applications simply through adjusting the relationship between agents' reward functions. For competition, the reward functions are zero-sum. Critically, MARL defines cooperation as having agents align their rewards through the same reward function (Busoniu et al., 2008), effectively splitting the group reward evenly among all cooperating agents. With the critical position taken by reward sharing in MARL, it is theoretically important to reveal the causal role of reward distribution in generating coordinated hunting with this model.

One particular algorithm in MARL, multi-agent deep deterministic policy gradient (MADDPG) (Lowe et al., 2017), has been successfully applied to a multi-agent-coordinated hunting game through agents sharing rewards. It shows that a group of predators can learn from scratch to coordinate the hunting of an intelligent prey. The algorithm is decentralized at the top level, with each agent learning its own model, instead of having a unified policy copied for all cooperative agents (Fig. 1). The training is cognitively realistic in two ways. First, the training involves cognitive constraints. When taking actions, each agent can only refer to its own observation, without accessing observations from other agents. Second, the model

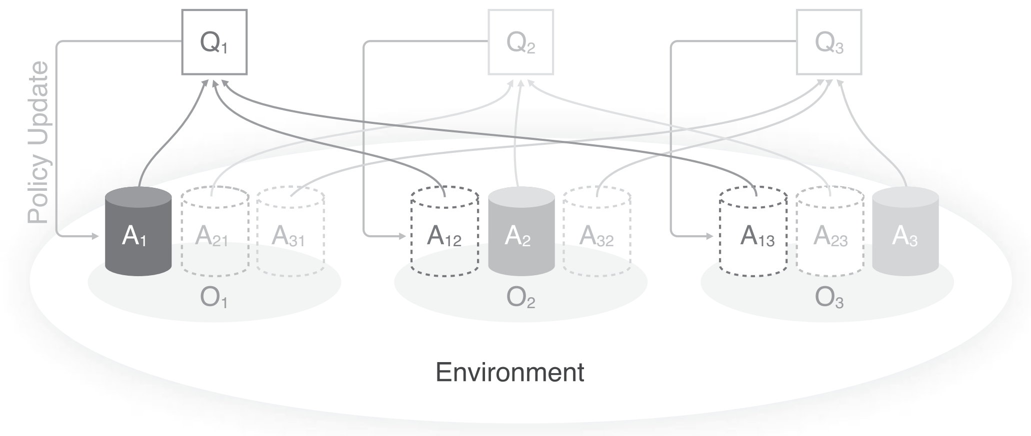

FIG. 1. Illustration of MADDPG algorithm with three agents. The MADDPG algorithm incorporates centralized evaluation and decentralized acting: Each agent i only has access to its own observations O_i to choose action A_i. When evaluating its action, each agent incorporates all other agents' observations to predict their actions (A_{ij} denotes agent i's prediction of agent j's action-to-take) and then use the observations and action predictions to form a centralized evaluation Q_i for its own action. Agents would then update their policies to output actions that generate improved values. MADDPG, multi-agent deep deterministic policy gradient.

is cognitively intelligent. Each agent treats others as actual agents, instead of random objects in the environment, and predicts what they will do next, a process that can be interpreted as a primitive version of Theory of Mind (Wellman, 1992). Moreover, an agent's evaluation of an action is based on the context of all agents' states and predicted action, echoing Tomasello's theory of coordinated hunting (Tomasello et al., 2005). Agents will then update their policies to output actions that would improve this evaluation. Still, the planning is individualized because agents are only concerned with their own actions while evaluating the situation stemmed from a holistic perspective.

However, as the focus of the MADDPG study is not in explaining realistic animal behavior, there are critical artificial components that make the conclusion not generalizable to animal hunting. First, the framework takes sharing rewards as an assumption and does not provide a comparison with cases using individual rewards, which fails to provide causal evidence for the effect of reward distribution. Second, the predators and prey have no action cost in the environment; thus, the free-rider problem is avoided altogether, since the only motivation for free riding is to avoid individual costs in cooperation. Third, to achieve better results in training, the environment rewards predator agents for ''bites,'' instead of ''kills,'' to create frequent reward signals, which helps with the model training. However, such a setting is unrealistic and even opposite in the real-world hunting scenarios, where only kills matter and can provide substantial material rewards. Bites without kills may in fact provide a negative reward to the predator, as it introduces chances of injuries and costs efforts. As such, it remains nebulous whether the computational model can indeed handle scenarios with only kill signals.

To test the causal effects of sharing rewards in modeling animal-coordinated hunting, we employ a hunting game (Fig. 2) with the following settings: (1) Predators have individual action costs that are proportional to the force they exert. (2) Predators are rewarded *only* after they successfully kill the prey, which happens in 20% of all biting instances (defined as collisions with the prey). (3) After killing the prey, the distribution of the spoils is systematically manipulated, ranging from selfishly dominant to evenly sharing. More details are explained in the following section.

2. MODEL DESIGNS

In the experiments, we systematically test MADDPG's performance in coordinated hunting with model experiment manipulations inspired by anthropological and animal studies.

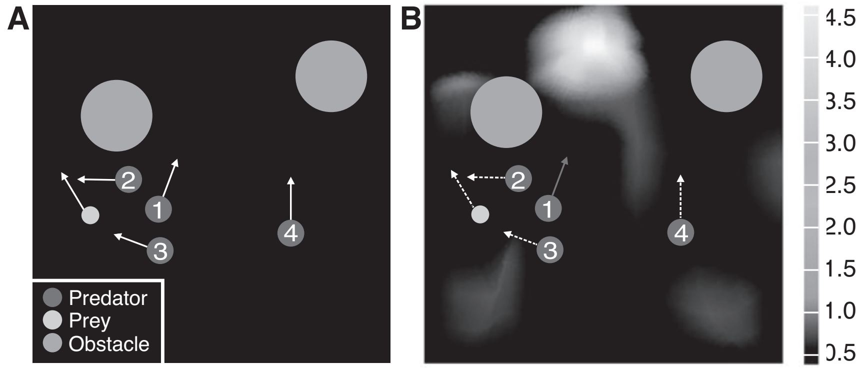

FIG. 2. An illustration of a successful coordination policy. **(A)** One frame of agents' motions during coordination. Predator 1 chooses to move upward, potentially to block the prey's movement (as a ''blocker''), instead of directly toward the prey (as a ''chaser''), indicating a sophisticated coordination strategy; **(B)** Predator 1's learned value landscape, with brighter shading indicating higher future expected rewards, conditioned on others' positions and predator 1's predictions of others' actions. The predator's policy encourages itself to move toward the position with a greater value.

2.1. Reward distribution among predators

Field observations indicate that proximity to prey at the moment it was killed plays an essential role in how chimpanzees decide to split the spoil (John et al., 2019). Here we control the reward distribution among predators as a function of the distance-to-kill. Sensitivity to the distance-to-kill is essentially a selfish index. With a high selfish index, the reward distribution concentrates on the predators close to the kill. When the predators are purely selfish (i.e., with an infinite selfish index), after one predator kills the prey, it will only reward itself since itself is the one closest to the prey. With a low selfish index, the rewards will be broadly dispersed. When the predators are purely unselfish (i.e., zero selfish index), the reward will be evenly distributed, regardless of agents' distance-to-kill. All predators follow the same mechanism in the same condition. Formally, we define the reward distribution as an exponential function of the distance-to-kill, such that

$$R_i \propto (d_i + 1 - k)^{-s} \tag{1}$$

where an agent i with d_i distance-to-kill receives R_i proportion of the reward, with selfish index s. The constant k denotes the minimum distance between two agents.

2.2. Action costs inducing the free-rider problem

Agents are motivated to free ride in coordinated hunting only to avoid the individual costs (Olson, 1965, 1989). As the action cost increases, agents would prefer to stay static to reduce the individual action costs while at the same time obtain allocated rewards. Accordingly, to test the severity of free-rider problems as a function of the reward distribution, we introduce action costs as negative rewards added to agents' reward functions. At each step, the action cost is proportional to the force exerted by agents, with the action cost for agent i, $C_i = a * F_i$, where F_i is the force exerted by agent i, which is the output from the agent's policy network and denotes the action taken at the time step; a denotes the action cost ratio in the specific condition. The action costs are applied to individual agents no matter which reward mechanism they take.

2.3. Group size

Evidence in animal studies has shown that hunting party size is positively correlated with the success of the hunters for many different species (Bednarz, 1988; Creel and Creel, 1995; Mitani and Watts, 1999; MacNulty et al., 2014; Samuni et al., 2018). In this study, we test how the reward distribution interacts with the group size by having different numbers of predators in the group.

2.4. Prey agility

Hunting risks influence the hunting behavior of many species. Wolves display a higher level of participation when hunting is riskier (MacNulty et al., 2014). Chimpanzee hunting has a low success rate and thus renders hunting an unnecessary activity for some groups, and some choose to hunt only in a situation of full nutritional abundance (Newton-Fisher, 2007). To evaluate the interaction between reward distribution and hunting risks, here we manipulate one factor of hunting risks, prey agility, by varying the speed of prey.

3. RESULTS

Our results indicate that there are significant main effects for all four variables, and selfish index is significantly intertwined with all three other variables (second-order multiple regression model, $p < 0.001$ for all terms, see Supplementary Data). In our further analysis, we investigate the two-way interaction terms by aggregating the other two conditions (Fig. 3). Our results indicate that the performance (i.e., the number of kills in one episode) of selfish agents increases linearly with the group size [linear regression, $p = 0.004$, line (B)], while unselfish agents' performance remains the same or even drops when having a larger group [line (A), $p = 0.013$]. Taking the group size of 6 as an example [line (c)], without loss of generality, the more selfish the predators are, the better their performance ($p < 0.001$). The change is monotonic, with the most selfish predators achieving the best performance (Fig. 3A).

The performance of all agents decreases as the action cost increases ($p < 0.001$), and increasing action costs hurt unselfish agents more than selfish agents—unselfish agents fail to obtain any rewards as the action cost reaches 0.01 level, whereas selfish agents still maintain a high level of performance even under the largest action cost condition (Fig. 3C). Furthermore, more selfish agents have their action force less sensitive to increase in action costs. The most unselfish agents decide almost not to move at all when there

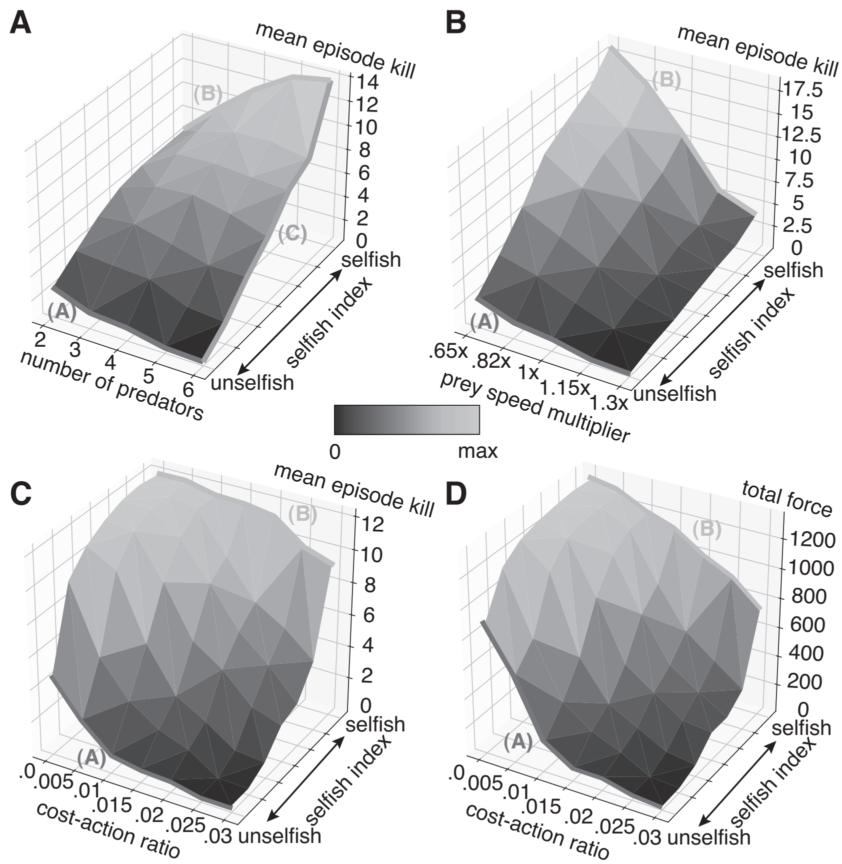

FIG. 3. Modeling results visualized as landscapes spanned by selfish index and one of the other three variables. **(A)** Performance of agents under different selfish index and group size, with highlighted lines denoting performance of unselfish agents [line (A)], selfish agents [line (B)], and performance of groups of size 6 [line (C)]. **(B)** Performance of agents under different selfish index and prey speed. **(C)** Performance of agents under different selfish index and action cost. **(D)** Agents' total force exerted under different selfish index and action cost.

is a small action cost (Fig. 3D). *Such a result strongly indicates the presence of the free-rider problem under the reward-sharing mechanism.* Lastly, as the speed of prey increases, the predators' performance decreases under all reward mechanisms, with the most selfish agents performing the best in all conditions (Fig. 3B).

3.1. Equilibrium testing

Having shown that sharing rewards cannot find a robust coordination strategy through learning, we further explore whether, given a successful coordinated policy, coordination can be maintained under sharing rewards. It will demonstrate whether successful coordination is a Nash Equilibrium through sharing rewards. In this study, we pretrained models from the condition with the highest group performance, which comes from the selfish predator group, for [30,000, 60,000, 90,000] episodes to manipulate the coordination level of the agents. The models were then further trained under both zero and infinitely high selfish indices with 120,000 training episodes for each condition in total.

Our results indicate that agents' performance remains similar before and after switching when there is no action cost, as no free-rider problem is involved (Fig. 4A). However, under conditions with action costs, the models significantly deteriorate after switching to the reward-sharing mechanism (Fig. 4B). This is evidenced by comparing models' final performance (after 120,000 episodes) with the performance at the switch ($t = -16.41,\ p < 0.001$) for switching at 30,000 episodes; $t = -50.35,\ p < 0.001$ for switching at 60,000 episodes; $t = -12.03,\ p < 0.001$ for switching at 90,000 episodes. While final performance of the models switched to training with individual rewards significantly drops, performance of models trained with selfish rewards all along increases monotonically during training (Fig. 4). More details on training are attached in the Appendix (included in the Supplementary Material). We conclude that successful coordination through sharing rewards is not an equilibrium under conditions with action costs, since all agents' policies deviate from it in further training.

4. DISCUSSION

Our results suggest that sharing rewards is neither necessary nor sufficient for modeling animal-coordinated hunting behavior with RL. It is unnecessary since models without any sharing (selfish agents) achieve good training results in the environment and even outperform agents that share rewards (unselfish agents). It is insufficient for three reasons. First, our results indicate a free-rider problem for unselfish

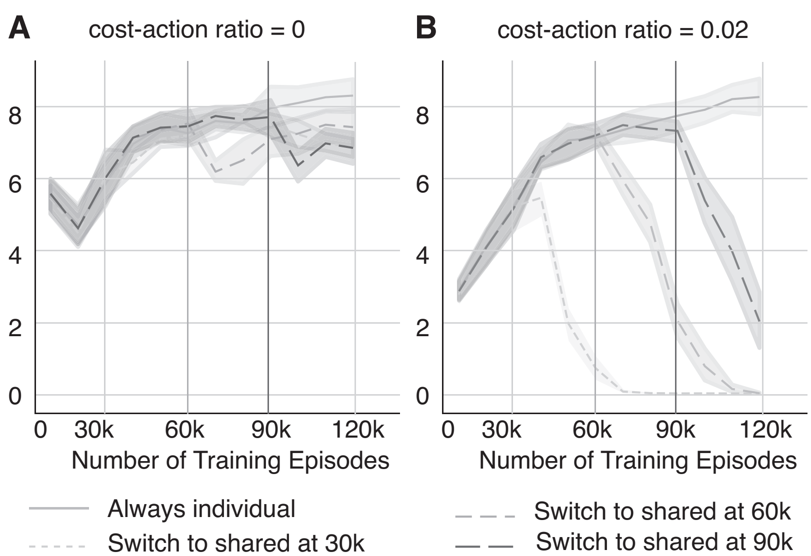

FIG. 4. Model performance with further training for equilibrium testing. Models are trained using the individual reward mechanism for [30,000, 60,000, 90,000] episodes, after which they are further trained with either individual reward or shared reward until 120,000 training episodes in total. The figures illustrate agents' performance with a 95% confidence interval under **(A)** zero action cost and **(B)** 0.02 action cost.

agents. Specifically, when agents share rewards and have their movements subject to individual action costs, they become reluctant to move, which negatively affects the group's performance. Second, unselfish agents' hunting performance does not improve when the group size increases, which contrasts with the observational evidence that hunting success should be positively correlated with the group size. Third, the reward-sharing mechanism cannot maintain a coordinated performance, and agents' actions deviated from a well-trained policy, possibly due to the free-rider problem.

Hence, our results support the theory that agents with selfish interests (e.g., chimpanzees) are capable of forming successful coordinated actions (Tomasello et al., 2005; Rossano, 2018). As such, sharing rewards might simply be a byproduct of chimpanzee hunting, instead of an intelligent strategy or the cause of a strong coordination performance, since the hunting performance deteriorates as the reward distribution gets more distributed. Note that our conclusion is based on modeling individual agents' reward-maximization behavior with internal structures and therefore applies to small-group hunting, which is the common scenario for mammals. Modeling coordination behavior in groups with hundreds or thousands of agents may require population-based algorithms.

MARL has been taken as a competitive model of cooperation through agents sharing rewards. However, our results indicate that this mechanism is not a required precondition for generating coordinated behavior and might even produce worse performance than training without such assumption. Suffering from the free-rider problem, the coordinated hunting behavior generated by multiple agents sharing rewards is indeed a special case when the action cost is zero. Failure of MARL through sharing rewards also suggests the limitation of trial and error in forming social coordination, which may explain why while all animals can learn from trial an error, none of them can form human-like social institutes. Sharing, as a crossculturally acknowledged virtue, should play a critical role in modeling coordination. To go beyond animal-like selfishness-based coordination, human-like cognitive processes need to be incorporated beyond using a generic learning algorithm.

One such framework is shared agency, which proposes that humans form a shared intention and work on tasks as a whole during cooperation (Tomasello, 2009). Recently, the idea of shared agency has been modeled through Bayesian theory of mind as inferring the relationship of utilities among agents (Kleiman-Weiner et al., 2016; Shum et al., 2019). In our work, we have extended the shared agency model into coordinated hunting using a framework called Imagined We. In this framework, agents jointly imagine that they are controlled by a super-agent *We* and would then make action plans based on the intention of the *We* agent, instead of thinking only from their own perspective. In a task where the superior agility of prey makes successful hunting depend on predators coordinating and persistently pursuing a single prey among multiple prey, it has been shown that the Imagined We model outperforms RL algorithms both under sharing and individual reward conditions (Tang et al., 2020, 2022).

AUTHORS' CONTRIBUTIONS

The authors confirm contribution to the article as follows: Conceptualization: M.Z., N.T., Y.Z., F.R., and T.G.; Methodology: M.Z., N.T., Y.Z., F.R., and T.G.; Funding acquisition: T.G.; Writing—original draft: M.Z., Y.Z., F.R., and T.G.; Writing—review and editing: M.Z., N.T., A.L.D., Y.Z., F.R., and T.G. All authors reviewed the results and approved the final version of the article.

AUTHOR DISCLOSURE STATEMENT

The authors declare they have no conflicting financial interests.

FUNDING INFORMATION

This work is supported by DARPA CREATE PA 19-03-01 (T.G.) and ONR MURI project N00014-16-1-2007 (T.G.).

SUPPLEMENTARY MATERIAL

Supplementary Data

REFERENCES

Arneson, R.J. 1982. The principle of fairness and free-rider problems. *Ethics* 92, 616–633.

Bednarz, J.C. 1988. Cooperative hunting Harris' hawks (*Parabuteo unicinctus*). *Science* 239, 1525–1527.

Berner, C., Brockman, G., Chan, B., et al. 2019. Dota 2 with large scale deep reinforcement learning. *arXiv* arXiv: 1912.06680.

Boesch, C. 2005. Joint cooperative hunting among wild chimpanzees: Taking natural observations seriously. *Behav. Brain Sci.* 28, 692–693.

Bonabeau, E. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl Acad. Sci. U. S. A.* 99, 7280–7287.

Boyd, R., Gintis, H., Bowles, S., et al. 2003. The evolution of altruistic punishment. *Proc. Natl Acad. Sci. U. S. A.* 100, 3531–3535.

Busoniu, L., Babuska, R., and De Schutter, B. 2008. A comprehensive survey of multiagent reinforcement learning. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev*. 38, 156–172.

Clutton-Brock, T.H., and Parker, G.A. 1995. Punishment in animal societies. *Nature* 373, 209–216.

Creel, S., and Creel, N.M. 1995. Communal hunting and pack size in African wild dogs, *Lycaon pictus*. *Anim. Behav*. 50, 1325–1339.

Gazda, S.K., Connor, R.C., Edgar, R.K., et al. 2005. A division of labour with role specialization in group–hunting bottlenose dolphins (*Tursiops truncatus*) off cedar key, Florida. *Proc. R. Soc. B Biol. Sci.* 272, 135–140.

Gilbert, M. 1999. Obligation and joint commitment. *Utilitas* 11, 143–163.

Gilby, I.C. 2006. Meat sharing among the Gombe chimpanzees: Harassment and reciprocal exchange. *Anim. Behav*. 71, 953–963.

Gross, J., and De Dreu, C.K. 2019. The rise and fall of cooperation through reputation and group polarization. *Nat. Commun.* 10, 1–10.

Hayek, F.A. 2011. *The Constitution of Liberty*. University of Chicago Press, Chicago, Illinois, USA.

Holekamp, K.E., Smale, L., Berg, R., et al. 1997. Hunting rates and hunting success in the spotted hyena (*Crocuta crocuta*). *J. Zool.* 242, 1–15.

John, M., Duguid, S., Tomasello, M., et al. 2019. How chimpanzees (*Pan troglodytes*) share the spoils with collaborators and bystanders. *PLoS One.* 14, e0222795.

Kleiman-Weiner, M., Ho, M.K., Austerweil, J.L., et al. 2016. Coordinate to cooperate or compete: Abstract goals and joint intentions in social interaction. In: *Proceedings of the 38th Annual Conference of the Cognitive Science Society 2016*. Cognitive Science Society, Seattle, Washington, USA.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. *Nature* 521, 436–444.

Liu, Y., Mattar, M.G., Behrens, T.E., et al. 2021. Experience replay is associated with efficient nonlocal learning. *Science* 372, eabf1357.

Lowe, R., Wu, Y., Tamar, A., et al. 2017. Multi-agent actor-critic for mixed cooperative- competitive environments. *Adv. Neural Inf. Process. Syst.* 30, 6379–6390.

MacNulty, D.R., Tallian, A., Stahler, D.R., et al. 2014. Influence of group size on the success of wolves hunting bison. *PLoS One* 9, e112884.

Mayr, E. 1961. Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science* 134, 1501–1506.

Mitani, J.C., and Watts, D.P. 1999. Demographic influences on the hunting behavior of chimpanzees. *Am. J. Phys. Anthropol*. 109, 439–454.

Mitani, J.C., and Watts, D.P. 2001. Why do chimpanzees hunt and share meat? *Anim. Behav.* 61, 915–924.

Mnih, V., Kavukcuoglu, K., Silver, D., et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 529–533.

Newton-Fisher, N.E. 2007. Chimpanzee Hunting Behavior. *In:* Henke, Winfried, and Tattersall (eds): *Handbook of Paleoanthropology.* pp. 1295–1320. Springer-Verlag, New York, New York, USA.

Nishida, T., Hasegawa, T., Hayaki, H., et al. 1992. Meat-sharing as a coalition strategy by an alpha male chimpanzee. *Top. Primatol.* 1, 159–174.

Nowak, M.A. 2006. Five rules for the evolution of cooperation. *Science* 314, 1560–1563.

Olson, M. 1965. *Logic of Collective Action: Public Goods and the Theory of Groups (Harvard Economic Studies. v. 124).* Harvard University Press.

Olson, M. 1989. Collective Action. *In:* Eatwell, J., Milgate, M., and Newman, P. (eds): *The Invisible Hand*. The New Palgrave, Palgrave Macmillan, London, UK.

Peterson, J.C., Bourgin, D.D., Agrawal, M., et al. 2021. Using large-scale experiments and machine learning to discover theories of human decision-making. *Science* 372, 1209–1214.

Rossano, F. 2018. Social manipulation, turn-taking and cooperation in apes: Implications for the evolution of language-based interaction in humans. *Interact. Stud.* 19, 151–166.

Samuni, L., Preis, A., Mielke, A., et al. 2018. Social bonds facilitate cooperative resource sharing in wild chimpanzees. *Proc. R. Soc. B Biol. Sci*. 285, 20181643.

Shalev-Shwartz, S., Shammah, S., and Shashua, A. 2016. Safe, multi-agent, reinforcement learning for autonomous driving. *arXiv* arXiv:1610.03295.

Shum, M., Kleiman-Weiner, M., Littman, M.L., et al. 2019. Theory of minds: Understanding behavior in groups through inverse planning. *Proc. AAAI Conf. Artif. Intell.* 33, 6163–6170.

Silver, D., Huang, A., Maddison, C.J., et al. 2016. Mastering the game of go with deep neural networks and tree search. *Nature* 529, 484–489.

Silver, D., Schrittwieser, J., Simonyan, K., et al. 2017. Mastering the game of go without human knowledge. *Nature* 550, 354–359.

Sutton, R.S., and Barto, A.G. 2018. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, Massachusetts, USA.

Tang, N., Stacy, S., Zhao, M., et al. 2020. Bootstrapping an imagined we for cooperation. *In: Proceedings of the 42nd Annual Conference of the Cognitive Science Society 2020*. Cognitive Science Society, Seattle, Washington, USA.

Tang, N., Gong, S., Zhao, M., et al. Exploring an imagined ‘‘we’’ in human collective hunting: Joint commitment within shared intentionality. *In: Proceedings of the 44th Annual Conference of the Cognitive Science Society, 2022.* Cognitive Science Society; Seattle, Washington, USA.

Tinbergen, N. 1963. On aims and methods of ethology. *Zeitschrift fürtierpsychologie.* 20, 410–433.

Tomasello, M. 2009. *Why We Cooperate*. MIT Press, Cambridge, Massachusetts, USA.

Tomasello, M., Carpenter, M., Call, J., et al. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691.

Vinyals, O., Babuschkin, I., Czarnecki, W.M., et al. 2019. Grandmaster level in Starcraft II using multi-agent reinforcement learning. *Nature* 575, 350–354.

Wellman, H.M. 1992. *The Child's Theory of Mind*. The MIT Press, Cambridge, Massachusetts, USA.

Yosef, R., and Yosef, N. 2010. Cooperative hunting in brown-necked raven (*Corvus rufficollis*) on Egyptian Mastigure (*Uromastyx aegyptius*). *J. Ethol*. 28, 385–388.

Address correspondence to:
Dr. Tao Gao
Department of Statistics
University of California Los Angeles
Los Angeles, CA 90095
USA

E-mail: tao.gao@stat.ucla.edu