# Exploring an Imagined "We" in Collective Hunting: Joint Commitment within Shared Intentionality

Siyi Gong*, Ning Tang†, Minglu Zhao‡, Chenya Gu†, Jifan Zhou†, Mowei Shen†, Tao Gao*‡
*Department of Communication, UCLA
†Department of Psychology, Zhejiang University
‡Department of Statistics, UCLA

***Abstract*—Philosophical theories as well as empirical evidence from developmental psychology suggest that humans, having shared intentionality as an underlying cognitive structure, may be able to form joint commitment in pursuing a collective goal out of many comparable goals without communication. By conducting experiments in a real-time cooperative hunting game, we demonstrated that humans established and maintained robust cooperation with high-quality hunting, even with a large number of potential targets. Additionally, we showed that a Bayesian imagined "We" (IW) model within a joint commitment framework, could capture humans' robustness in resisting alternative targets with relatively high quality of hunting. This poses a contrast with a Reward Sharing (RS) model that, despite performing proficiently in pursuing a single goal, mostly exhibited low-quality hunting and whose teaming fell apart as available targets increased. In a hybrid team simulation experiment, the IW model could better mimic the intentions of human hunters compared to the RS model. Together, the success of the persevered group commitment in humans suggests that shared intentionality is a pivotal element in human cooperation.**

## I. Introduction

Collective hunting is a complex group activity commonly seen in ecology, in which hunters pursue prey in a group effort. This behavior is evolutionarily significant in the animal kingdom as it extends cooperation beyond kinship to genetically unrelated group mates, including friends, nonfriends [e.g., 27], and even heterospecifics. For example, Taï Chimpanzees regularly hunt for red colobus monkeys in small groups [e.g., 4], and coyotes and badgers were observed to group-hunt ground squirrels [23]. Moreover, collective hunting and foraging has been viewed as a breakthrough in hominid evolution and provides a basis for humans' unique large-group cooperation later in phylogenesis [33, 34].

In collaborative activities such as collective hunting and foraging, humans are committed to achieving a goal together. Studies show that toddlers regulate their own [15] and other's commitment [38, 12] in a joint activity, and express guilt when they fail to maintain their commitment [36]. These behaviors, however, were not found in parallel experiments conducted in chimpanzees who are primarily motivated by individual desire [21, 38, 13]. Along with empirical evidence [e.g., 33], it is theorized that the species-specific cognitive representation at the core of this divergence between cooperation of human and other species is shared intentionality [e.g., 5, 35, 26, 11], the conception of a collective representation of the self and others, or "individuals as a group" [25].

The notion of shared intentionality has been conceptualized throughout history. One consensus regarding shared intentionality is that it is irreducible to a sheer summation, aggregation, or distributive pattern of individual intentionality, but rather is a qualitatively different structure of the mind [25]. Among many possible ways to interpret the irreducibility claim, Gilbert focused on what she believed as a definitive feature of the shared intentionality structure—joint commitment, proposing that a joint intention is only realized when two or more individuals are willing to be "jointly committed to espousing a goal as a body" [11]. The "goal" here may include a variety of intentions, beliefs, and acceptance [25], whereas "as a body" indicates an indivisible whole. For example, when individuals A and B jointly believe X as a body, they are committed to forming a single supraindividual agent C that believes X. A and B, in this case of joint commitment, constitute a plural subject that possesses the shared intentional state and cannot be broken down into two single subjects. This notion of plural subject, as we will articulate in detail later, can be understood as a distinct agent with its own actions and mind, including a full set of belief, desire, and intention. Moreover, once the joint commitment to the plural subject is established, any member cannot rescind this commitment unilaterally, and all members in the commitment are normatively responsible for each other. That is, each of them is obligated to act in accordance with their joint goal and entitled to demand others' continuation of the joint action.

Another well-accepted idea, the individual ownership claim, emphasizes that shared intentionality is fundamentally "had by individuals"—it is one's own shared intentionality [25]. This suggests that each individual voluntarily commits, owns, and understands their intention. Gilbert formulates this idea as each individual will need to engage in "whatever behavior" to demonstrate their "readiness" to willingly commit to a goal as a body, which appeals to all as common knowledge [10]. Based on a rich body of literature on the perception of intention and commitment, [e.g., 8, 29], it is highly likely that the readiness for initiating, as well as maintaining, joint commitment can be expressed simply from coordinated motions. Collectively, these pieces of evidence suggest that humans may be able to establish a sustained joint commitment in a real-time visual-grounded collective hunting task without explicit communication.

So far, current psychophysics works on the perception of

intention have been primarily conducted in individual settings and do not concern interactions between multiple agents. For example, a line of studies shows that humans are able to identify the intentions of prey and predators in an online, real-time chasing paradigm [e.g., 8, 9, 22]. These works heavily revolve around the tension between predators and prey, but rarely involve cooperation between predators. Additionally, participants mostly took the role of observers but not actual players, with a few exceptions where they actually controlled the prey [8]. Of the few studies that did use displays of cooperative chasing [40, 7], their focus was on perception alone, instead of generating cooperative actions based on perception. Moreover, in these cases, the goal of chasing was fixed to a single target and did not involve the challenge of maintaining joint commitment among many possible goals. These studies generated fruitful results that provide invaluable evidence for humans' ability to infer others' intentions in hunting tasks. However, they cannot be easily generalized to cooperative hunting scenarios in which individuals are not only observers but also participants that generate cooperative behaviors. In such cases, aside from inferring others' intentions and generating action plans accordingly, it is equally important to constantly align one's own intention with others' to converge on a collective goal "as a body" in an enduring fashion. It is thus worth exploring whether humans as engaging players can achieve good cooperation in a real-time hunting task by overcoming such challenges.

Building on top of these works, we first aim to examine whether a group of three humans can exhibit robust joint commitment while playing a virtual collective hunting game. Following this, we built a computational model of shared intention, named imagined "We," directly inspired by Gilbert's theory of commitment. Our goal is to show that this model can indeed capture important aspects of human cooperative hunting. As a baseline, we also employ a Reward Sharing model without any representation of shared intention. By revealing human performance and comparing it with model performance, we aim to better understand whether joint commitment plays an important role in human cooperation.

## II. Collective Hunting Experiment

To test the joint commitment in cooperative hunting, we developed a real-time game in a 2-D environment (fig. 1) with 3 hunters played by humans and 1, 2, or 4 stags as targets which are played by a machine. We want to explore whether human participants consistently converge on the same goal even without communication during the hunting process. Furthermore, we aim to test whether their cooperation can be resistant to an increase in the number of available targets, following the logic that once the joint commitment is achieved, participants should at least be able to secure one target, if not more. Demos of the cooperative hunting process can be found at https://ucla.box.com/s/ckclmb9kh4wu9qbi5mb796up6fkelgo9

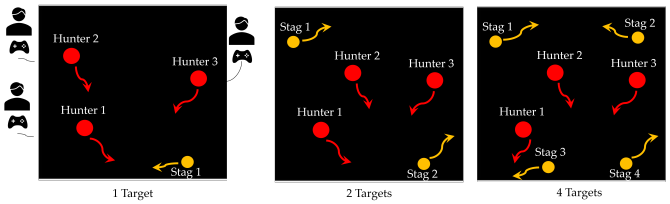


Fig. 1. Cooperative Hunting Task.

### A. Cooperative Hunting Task

Hunters aim to successfully catch the stags, while the stags aim to avoid the hunters. The stags move faster than the hunters, thus requiring hunters to collaborate by persistently chasing a single stag. In multiple-target scenarios, hunters have no predetermined target. Nevertheless, simply for the purpose of improving performance and maximizing accumulated rewards, it is best for them as a group to go after one target at a time. Agents in the environment can take actions within a range of magnitude of force from any direction at a given time step in order to achieve their respective goals. Task performance is evaluated through achieved rewards in a fixed period. The hunters receive a joint reward (+1) upon any successful touch of a stag, defined as any frame in which the hunter-target distance 2.5°. The accumulated reward at the end of each trial is used as a dependent measure of the model's performance.

### B. Computational Model

To model human performance, we built an imagined "We" model with a shared intentionality framework to test whether joint commitment is indeed the pivotal mechanism underlying human cooperation.

The previous modeling work on shared intention traces back to early logical models in artificial intelligence, including Grosz and Kraus's "SharedPlans" model [1996] and Levesque et al's definition of a "joint persistent goal" [1990]. More recently, some promising modeling works on social agency combined Bayesian inference and Theory of Mind (ToM) approaches with a focus on how agents coordinate their action plans to settle a strategic decision of whether to cooperate or compete in the social world [18, 28]. Under this framework, their formation of cooperation was an abstract planning procedure that implements shared intention as a joint policy that optimizes the team's joint reward, which was then turned into an individual policy by marginalizing the actions of others. In a recent work by [39], agents in the infer what sub-task of a cooking task other agents are working on, and plan accordingly whether they should help with the sub-task or not. During a sub-task cooperation, each agent samples a fictitious centralized planner that controls actions of all participating agents, resonating with the concept of a shared intention though not focusing on the joint commitment. This model, referred to as the Bayesian Delegation model, focuses on inferring the sub-task distribution given a fictitious centralized planner of a fixed shared goal but does not intend to solve the process of inferring a centralized planner when sampling a shared goal out of multiple equivalent alternatives.

Fig. 2. Imagined “We” Representation. The graphical model in each of the two dashed boxes represents a supraindividual agent “We”, which has its own mind containing belief, desire, and intention. Using those mental states, it can rationally control the joint action of the individual agents constituting “We”. Each dashed box represents a unique version of the unreal, imagined supraindividual agent “We” inferred by each of the two collaborating individuals here.

1) Imagined “We” Model: Directly motivated by previous Bayesian ToM modeling works [18, 28] and in parallel with the Bayesian Delegation model [39], our model employs similar idea of a ctitious centralized planner but speci cally draws inspiration from Gilbert's theory of joint commitment. In our case, human cooperation is assumed, and the focus is on how to model cooperation with a stronger constraint to cohere the team. This is consistent with the perspective that cooperation is qualitatively different from competition [31]. Preliminary modeling results showing the feasibility of this model with only two collaborators were reported in [30] without comparisons to human performance and another baseline model.

Here, we especially focus on the tension between the two well-accepted claims about shared intentionality, being that the collective attitude beyond an individual is, at least to some degree, incompatible with the idea that the intention an individual has cannot escape their own mind. Here, we aim to reconcile this discrepancy by using the imaginative capacity of a causal model [24] implemented as the Bayesian Theory of Mind [2, 17]. While “We” as a supraindividual agent is not real, each agent can nevertheless imagine the mind of “We” from a collective, “bird's-eye” perspective in their own individual mind [31]. Speci cally, “We” re ects the collective wills of all individuals, but is also a single autonomous agent with its own mind and action just like any ordinary agent, as suggested by Gilbert's theory [2006]. Its mental states can be further parsed into belief, desire, and intention which together rationally control this agent's actions. Thus, we can infer the mental states of “We” from its action using ToM, where its state space and action space are simply a concatenation of the state spaces and action spaces of individual agents.

Crucially, this supraindividual agent does not exist in reality—it is ultimately realized by an individual's own imagination about “We” through reasoning counterfactually about “how can an agent explain its own and others' actions if such actions have indeed been rationally controlled by a supraindividual agent `We'?” For this reason, we call our model the imagined “We” (IW) model ( g. 2), in part following the classic term of “imagined community,” that suggests many communities are rst constructed by the imagination [1].

Upon the readiness for joint commitment from all collaborating individuals, each of them, without communication, infers their own version of the imagined “We” by observing the joint action of themselves and their partners in the shared environment. Each agent conceptualizes their own version of “We” and acts by asking “what does `We' expect me and others to do?” Aside from taking its own action following the intention of “We”, an individual agent also expects others to take the actions demanded by “We” (eq. (1)). Newly generated actions from all agents can be observed and used for each individual agent to update their own inference about “We” for the next time step. An agent's inference is conditioned on the environment in order to capture the intuition that the mind is in uenced by the surrounding environment (eq. (2)). Eventually, joint commitment will be achieved when all individual versions of “We's” are aligned or converged.

$$\text{Joint action} \quad P(\text{Joint action} \mid \text{“We” mind}, \text{Environment}) \tag{1}$$

$$P(\text{“We” mind} \mid \text{Joint action}, \text{Environment}) \propto P(\text{Joint action} \mid \text{“We” mind}, \text{Environment}) \, P(\text{“We” mind} \mid \text{Environment}) \tag{2}$$

Essentially, this is a process of determining what “We” believes or what “We” wants by observing what “We” has done. Speci c to the context of the cooperative hunting task, the environment is fully observable without any uncertainty. The only uncertainty surrounds the intention of “We,” concerning “which prey should `We' pursue persistently?” We model the inference of “We” intention using a bootstrapping method following three steps of computation ( g. 3).

Goal Sampling: At each time step $t$, each agent $i$ maintains a distribution of intention of “We” as the probability of each target being the joint goal $GW_{i(t)}$. To decide how to act next, it will draw a sample from this distribution and use it as the estimation for what is the current intention of “We” (eq. (3)). This process of eliminating the freedom of choice and thus avoiding an open-ended future abides the exclusive property of intention explored in a recent study [6].

$$GW_{i(t)} \quad P(GW_{i(t)}) \tag{3}$$

Planning: Given a goal, each agent forms a plan of how “We” should pursue that goal rationally. The output of this planning process is a joint action, including its own action to take, as well as an expectation of the other agents' actions. Thus, each agent is simulating a centralized planner. We implemented this rational planning by using a joint policy that was learned through a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm with only one goal to pursue [20].

Fig. 3. Bootstrapping Imagined "We". In this case, there are two agents in total, with each loop representing one agent's way to bootstrap an imagined "We". Within an agent's loop, each "IW" node represents a unique distribution of mental states the agent inferred from its imagined "We" agent; the solid "a" node represents the agent's chosen action given its inferred distribution; the dashed "a" node is its expectation of the other agent's action. Actions actually taken by all agents are then observed and used by each agent to update its imagined "We" for the next time step. Here, two agents are used for illustration purposes, while the model can be generalized to multiple agents.

This joint policy (eq. (4)) de nes the probability of joint actions ($A_{joint\ (t)} = action_{1(t)}; :::; action_{i(t)}; :::; action_{I(t)}$) conditioning on the current states of "We" ($S_{\backslash we"(t)} = State_{1(t)}; :::; State_{i(t)}; :::; State_{I(t)}$) and the goal ($S_{GW_{i(t)}}$). Each agent then samples a joint action from the policy distribution and takes its own part of the joint action. Empirically, MADDPG was found to optimize group reward when only one target was present in a hunting scenario [41]. Note that this rational planning phase does not necessarily imply that human cognition employs the exact policy we utilize here. Rather, it is an engineering approximation of the assumption that humans generally act rationally to optimize their joint utility.

$$P(A_{joint}\ jGW_{i(t)}) = P(A_{joint}\ jS_{\backslash we";GW_{it}}) \quad (4)$$

Inference: After taking one's own action based on the policy determined in the planning phase, each agent observes the actions actually taken by all agents. This enables a Bayesian ToM inference process (eq. (5)): Conditioning on the observed actions, each hunter computes the posterior probability of a given target being their joint goal. After updating the posterior of the Imagined "We" mind, each agent goes back to step 1, sampling a new goal and repeating the process.

$$P(GW_{i(t+1)}\ jA_{joint\ (t)}) \ / \ P(A_{joint\ (t)}jGW_{i(t)})P(GW_{i(t)}) \quad (5)$$

2) Baseline Model: To explore the necessity of modeling a shared intention framework, we additionally examine cooperation in a Reward Sharing (RS) model as a baseline. Here we use the MADDPG, an algorithm within the MARL framework. MARL adopts the perspective that social skills are learned through trial-and-error [16]. It has been successfully applied to complex multi-agent coordination tasks [3, 37], in which it splits group rewards evenly among all agents by assigning them the same reward function. Since we used the MADDPG algorithm to train the joint policy in the one-target scenario, the RS model and the IW model are identical in the case of a single target. For the multiple target conditions, we further trained a separate RS model for each set size of targets. Note that this separate training is not required in the IW model as the inference of a goal is achieved by Bayesian inference and the same one-target policy is applied to all conditions for joint goal inference and planning.

Despite being a reinforcement model, the RS model displays several interesting components that can be considered precursors of ToM. For example, it acts based on predicting what other agents will do given their current states. Then it evaluates the utility of its own action given the current joint state of all agents plus its prediction of other agents' actions. As a type of reinforcement learning model, it has a generic framework that can be universally applied to any multi-agent scenario, including both cooperation and competition, which only differ by whether agents' rewards are aligned or opposed. At its core, it is the opposite of the IW model that assumes cooperation is qualitatively different from competition and thus requires a brand new cognitive scaffold. In short, collaborations in the RS model are encouraged by sharing rewards without any reference to commitment, whereas teaming behaviors in the IW model are enforced by shared intention.

3) Model Task & Prediction: The same task completed by human participants was used to test the IW model and the RS model and compare their performance to that of humans. We aim to explore whether human performance in cooperative hunting can be better captured by the stronger-constrained IW model or the weaker-constrained MARL model.

## C. Human Experiment

Thirty-three (3 participants in 1 group, 11 groups total) students (14 females, 19 males) participated in this experiment for payment. All were between the ages of 21 and 28 ($M_{age} = 23, SD = 2.0$).

1) 2: 0 trials were set up for each condition with different set sizes of hunting. The hunting task was presented on a 38.0° x 38.0° window displayed simultaneously on three monitors, one for each participant. Each participant was instructed to use an Xbox controller to control the simulated physical forces they could apply to drive the hunters on the screen. Three hunters were represented by circular shapes with a diameter of 1.9° and colors red, green, and blue. The stags are circular shapes with a diameter of 1.3° and could be distinguished by yellow colors of different brightness.

2) Results: Overall Performance. We rst analyzed the accumulated reward of the hunting game, which re ected the overall performance of humans and models ( g. 4). One-way ANOVA revealed a signi cant main effect of set size ($F(2, 30) = 6.58, p < .01, \eta_p^2 = 0.31$). The post-hoc comparisons showed that when the number of targets increased to 4, the performance was even higher than the 1 and 2 target conditions ($ps < .01$). Overall, when the number of targets increased, the accumulated reward of human hunters did not decrease but

Fig. 4. Results of the accumulated rewards.

Fig. 5. Results of the percentages of different quality rewards.

Fig. 6. Results of the entropy of touched target distribution.

increased instead. For the IW model, the main effect of set size was not signi cant ($F(2, 30) = 1.15, p = .331, \eta_p^2 = .07$). The results revealed that the performance of the IW model did not decrease as the number of targets increased. For the RS model, a signi cant main effect of set size was revealed ($F(2, 30) = 87.79, p < .01, \eta_p^2 = .86$). The post-hoc comparisons showed that when the number of targets increased, the performance gradually decreased ($ps < .05$ for both set size 1-2 and set size 2-4 comparisons).

Quality of Hunting. Besides the above quantitative analysis of the overall performance, we further explore the quality of hunting in humans and models ( g. 5). Here we indicate the quality of hunting by measuring the "duration of touch," de ned by the number of consecutive time steps in which at least one hunter touches the stag. In real life, a short touch duration suggests a hit to the target, but not necessarily a catch, whereas a long touch duration indicates a greater likelihood for a real catch or kill, as hunter(s) may have cornered the stag. Thus, the quality of raw rewards was categorized into classes in terms of touch duration: low (1 time step), median (2 time steps), or high (3 or more time steps). The percentages of different qualities of rewards in the total rewards were then measured.

For the set size 2 condition, one-way ANOVA revealed a signi cant main effect of player type in both low- and high-quality conditions ($F(2, 30) = 523.69, p < .001, \eta_p^2 = 0.97; F(2, 30) = 672.49, p < .001, \eta_p^2 = 0.98$). The post-hoc comparisons showed that humans received low-quality reward less often than the IW model ($p < .001$), which received it less often than the RS model ($ps < .001$). On the contrary, humans received high-quality reward more often than the IW model ($p < .001$) , which received it more often than the RS model ($p < .001$). Similar main effects of player type ($F(2, 30) = 576.61, p < .001, \eta_p^2 = 0.97; F(2, 30) = 803.57, p < .001, \eta_p^2 = 0.98$) were found in set size 4 condition with similar post-hoc comparisons ($ps < .001$). These results collectively suggest that human hunters achieved the largest proportion of high-quality hunting, followed by the IW hunters, while the RS hunters achieved the smallest proportion. Notably, the IW model had a relatively high quality of hunting, though there was still room for improvement.

Goal Consistency. Beyond task performance, we further analyzed the goal consistency among hunters in a team ( g. 6). Here we measured the entropy of the distribution of the touched target, of which a lower entropy value indicates a higher convergence or concentration on the same goal from all hunters. Both set size 2 and 4 conditions showed a signi cant main effect of agent type ($F(2, 30) = 110.11, p < .01, \eta_p^2 = 0.87; F(2, 30) = 13.38, p < .001, \eta_p^2 = 0.44$). For the set size 2 condition, the entropy of the touched target distribution of humans was higher than that of the IW model ($p < .001$), but was lower than that of the RS model ($p < .001$). For the set size 4 condition, the difference between the entropy of humans and that of the IW model was not signi cant ($p = .24$), but both of them were higher than the entropy of the RS model ($ps < .01$). These results suggested that the way humans pursued their goals could be better captured by the IW model than the RS model.

### D. Hybrid Team Simulation

1) Overview: Thus far we have only examined the performance of each type of player within their homogeneous group. Here, we take one step further to measure how well they can cooperate with each other to investigate the compatibility between their hunting strategies. We conducted a hybrid team simulation experiment based on the pre-recorded trajectories of all agents in the human experiment. To see how well they could cooperate with each other, we replaced a human hunter with an IW or RS model hunter while leaving the trajectories of all other agents untouched. As the new hunter was "invisible" to the pre-recorded stags, we expected an overall increase in the performance of the hybrid team compared to the original all-human team, but we were more interested to discover whether the models could align their goals with the rest of the human hunters. We examined the matching of goal consistency by measuring the cross entropy of the touched target distributions between the hybrid teams and the original all-human team. If the models can successfully infer the goal of human hunters and cooperate by coordinating their hunters' behaviors to commit to the same goal, then the touched target distribution of the hybrid team would be similar to that of the all-human team with a low cross entropy.

Fig. 7. Results of the cross entropy of the touched target distributions between the hybrid teams and the original all-human team.

2) Results: For both set size 2 and 4 conditions ( g. 7), the cross entropy between the IW-human team and all-human team was signi cantly lower than that between the RS-human team and all-human team ($t(20) = -17.13, p < .001, d = 7.67$; $t(20) = -37.22, p < .001, d = 16.65$). This shows that the IW model could better "replicate" the intention of the human player they replaced, than the RS model.

## III. Conclusion

The human behavioral results demonstrate a successful expansion of human social perception from individual chasing tasks to a multi-agent cooperation task, which involves the integration of perception and planning. We thus show that humans were indeed capable of, and in fact good at, achieving effective collaboration in a hunting task facing multiple temptations even without any form of explicit communication. This evidence corroborates the theory that communication may have only emerged in environments where collaboration already existed [32]. Moreover, the reasonably robust cooperation in humans, the IW model, and their teaming, in comparison to the limited performance in the RS model, support that shared intentionality is a key mechanism in enabling humans to stay robustly committed in cooperation. It is nevertheless also true that humans exceeded the IW model in cooperative hunting in both quantity and quality, potentially re ecting greater exibility in human cooperative behaviors. We believe that addressing the challenge of how to integrate exible task assignments while maintaining the constraint of shared intentionality will be the next step in advancing cooperation modeling in the future.

## References


[1] Benedict Anderson. Imagined communities: Re ections on the origin and spread of nationalism. Verso books, 1983.

[2] Chris L Baker, Rebecca Saxe, and Joshua B Tenenbaum. Action understanding as inverse planning. Cognition, 113 (3):329–349, 2009.

[3] Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemys aw Debiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Chris Hesse, et al. Dota 2 with large scale deep reinforcement learning. arXiv preprint arXiv:1912.06680, 2019.

[4] Christophe Boesch and Hedwige Boesch. Hunting behavior of wild chimpanzees in the tai national park. American journal of physical anthropology, 78(4):547–573, 1989.

[5] Michael Bratman. Rational and social agency: Re ections and replies. Rational and Social Agency, pages 294–343, 2014.

[6] Shaozhe Cheng, Ning Tang, Wei An, Yang Zhao, Jifan Zhou, Mowei Shen, and Tao Gao. Intention beyond desire: Humans spontaneously commit to future actions. In Proceedings of the Annual Meeting of the Cognitive Science Society, volume 43, 2021.

[7] Jipeng Duan, Zhangxiang Yang, Xiaoyan He, Meixuan Shao, and Jun Yin. Automatic attribution of social coordination information to chasing scenes: evidence from mu suppression. Experimental brain research, 236 (1):117–127, 2018.

[8] Tao Gao, George E Newman, and Brian J Scholl. The psychophysics of chasing: A case study in the perception of animacy. Cognitive psychology, 59(2):154–179, 2009.

[9] Tao Gao, Brian J Scholl, and Gregory McCarthy. Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. Journal of Neuroscience, 32(41):14276–14280, 2012.

[10] Margaret Gilbert. A theory of political obligation: Membership, commitment, and the bonds of society. Oxford University Press, 2006.

[11] Margaret Gilbert. Joint commitment: How we make the social world. Oxford University Press, 2013.

[12] Maria Gräfenhain, Tanya Behne, Malinda Carpenter, and Michael Tomasello. Young children's understanding of joint commitments. Developmental psychology, 45(5): 1430–1443, 2009.

[13] Julia R Greenberg, Katharina Hamann, Felix Warneken, and Michael Tomasello. Chimpanzee helping in collaborative and noncollaborative contexts. Animal Behaviour, 80(5):873–880, 2010.

[14] Barbara J Grosz and Sarit Kraus. Collaborative plans for complex group action. Arti cial Intelligence, 86(2): 269–357, 1996.

[15] Katharina Hamann, Felix Warneken, and Michael Tomasello. Children's developing commitments to joint goals. Child development, 83(1):137–145, 2012.

[16] Friedrich August Hayek. The Constitution of Liberty. University of Chicago Press, 5 edition, 2011.

[17] Julian Jara-Ettinger, Hyowon Gweon, Laura E Schulz, and Joshua B Tenenbaum. The naïve utility calculus: Computational principles underlying commonsense psychology. Trends in cognitive sciences, 20(8):589–604, 2016.

[18] Max Kleiman-Weiner, Mark K Ho, Joseph L Austerweil, Michael L Littman, and Joshua B Tenenbaum. Coordinate to cooperate or compete: abstract goals and joint intentions in social interaction. In A Papafragou, D Grodner, D Mirman, and J C Trueswell, editors, Proceedings of the 38th Annual Conference of the Cognitive Science