Prosocial Norm Emergence in Multiagent Systems

MEHDI MASHYEKHI, North Carolina State University, USA
NIRAV AJMERI, University of Bristol, UK
GEORGE F. LIST, North Carolina State University, USA
MUNINDAR P. SINGH, North Carolina State University, USA

Multiagent systems provide a basis for developing systems of autonomous entities and thus find application in a variety of domains. We consider a setting where not only the member agents are adaptive but also the multiagent system viewed as an entity in its own right is adaptive. Specifically, the social structure of a multiagent system can be reflected in the social norms among its members. It is well recognized that the norms that arise in society are not always beneficial to its members. We focus on prosocial norms, which help achieve positive outcomes for society and often provide guidance to agents to act in a manner that takes into account the welfare of others.

Specifically, we propose Cha, a framework for the emergence of prosocial norms. Unlike previous norm emergence approaches, Cha supports continual change to a system (agents may enter and leave) and dynamism (norms may change when the environment changes). Importantly, Cha agents incorporate prosocial decision making based on inequity aversion theory, reflecting an intuition of guilt arising from being antisocial. In this manner, Cha brings together two important themes in prosociality: decision making by individuals and fairness of system-level outcomes. We demonstrate via simulation that Cha can improve aggregate societal gains and fairness of outcomes.

1 INTRODUCTION

Major practical applications of information technology can be understood as multiagent systems viewed from a sociotechnical perspective in terms of a social tier (comprising social entities such as people and organizations) and a technical tier (comprising computational and other resources) [37]. The agents represent the social entities computationally and reflect the interests of the entities they represent. As Singh [67] points out, the governance of such systems is naturally characterized by the social norms that apply at the social tier. Indeed, norms help regulate interactions of autonomous agents [75]. Accordingly, it is not surprising that norms apply in resource-sharing settings broadly [68], such as social media [61] and for road sharing by autonomous vehicles [19]. In general, discussions of norms can be framed in terms of agent behavior; here, we focus on interactions between agents, to clarify that we are interested in agent behaviors that are of concern to others and are visible to others.

Authors' addresses: Mehdi Mashyekhi, North Carolina State University, Raleigh, NC, USA, USA, mmashay2@ncsu.edu; Nirav Ajmeri, University of Bristol, Bristol, UK, UK, nirav.ajmeri@bristol.ac.uk; George F. List, North Carolina State University, Raleigh, NC, USA, USA, gflist@ncsu.edu; Munindar P. Singh, North Carolina State University, Raleigh, NC, USA, USA, mpsingh@ncsu.edu.

The recent burgeoning interest in AI ethics and safety has drawn increasing attention to regulating how intelligent agents act and interact [77]. In this regard, it is helpful to distinguish macro ethics (focused on a system) from micro ethics (focused on an individual agent) [16]. Micro ethics is concerned with how agents interact in light of the norms [14, 32], and macro ethics is concerned with what norms regulate a system, including considerations of fairness [50, 78].

At its heart, a social norm defines sound or "normal" interactions among members of a social group, reflecting their mutual expectations. A norm can be violated by a member though at the cost of suffering either a moral or another sanction [51]. A norm can be descriptive, merely describing a social group, or prescriptive, describing how the social group should function. Moreover, some norms can be constitutive, meaning they provide definitions of what counts as what—in essence, the rules of the game. Or, they can be regulative, meaning they constrain the interactions of the participants [13]. A lot of the work on multiagent systems approaches them from the standpoint of explicit engineering, where stakeholder requirements are mapped to, among other elements, a set of norms [39].

This paper adopts an equally well-established alternative perspective in which the norms are not designed into a system but emerge through the interactions of the agents [35]. Specifically, this paper proposes Cha, a framework for norm emergence. Our motivations behind Cha are as follows. First, we would like to enable agents who incorporate prosocial decision making to achieve norms that avoid conflict and improve fairness. Second, we would like to tackle the challenges of flexibility in that it should be possible for the emerged norms in a multiagent system to change dynamically. Third, the process of emergence should not have to rely upon a central authority that somehow forces the agents to adopt or respect norms; instead, the norms should arise—and, when appropriate, change—in a decentralized manner solely through the interactions among the member agents.

Next, we abstract out some key desiderata from the above motivations and discuss how Cha addresses these desiderata.

Prosocial. A prosocial attitude is an attitude to benefit another. Prosocial behavior has been well studied in sociology and social psychology [21, 64]. A prosocial behavior is when an agent performs an action that benefits others even if suboptimal to itself [53, 55, 62]. Whereas existing approaches for norm emergence focus solely on decision making by agents, we relate social norms both to prosocial decision making and to societal outcomes such as fairness. We show how to incorporate prosociality into norm emergence to achieve norms that promote fair outcomes and improve social welfare for the members of a multiagent system.

Flexible and Enduring. The flexibility and endurance (in other words, the longevity) of an interaction are important aspects of openness [74]. The membership of a multiagent system can change yet there is continuity in learning norms. An example is Wikipedia, where the users are autonomous and changing but can and do build on each other's work. In contrast, existing studies of norm emergence [2, 59, 71, 79] apply social learning and assume that agents repeatedly interact in a closed system. Specifically, these approaches assume that a fixed graph is given and the agents interact with their neighbors in that graph.

Dynamic. Dynamism refers to the idea that norms may change with changes to the environment in which the system operates [36, 56, 74]. Existing works [43, 48] address dynamism inadequately. For example, once norms have emerged, they remain fixed. Existing approaches, such as Dell'Anna et al. [19, 20], support norm change at runtime but via a centralized sanction revision, which is prone to a single point of failure.

Decentralized. Decentralization refers to the absence of central authority [49]. In a decentralized system, norms arise solely through agent interactions [43, 44]. Most existing norm emergence approaches, such as Airiau et al. [2], Morales et al. [45], involve a central authority that

determines the norms; some existing approaches, such as Mashayekhi et al. [42], use a hybrid scheme. A criticism of hybrid, and especially of centralized, systems is that they assume that an all-knowing authority is present in the system [56]. Whereas a central authority may recommend norms to achieve a societal criterion, such an approach cannot guarantee the autonomy of the participating agents. In settings with unreliable or delayed communication or where components can fail, a centralized system is fragile because it has a single point of failure.

1.1 Contributions

Cha is a general dynamic and flexible framework for norm emergence that promotes prosociality while supporting decentralization. We frame our contributions as an investigation of the following themes.

Efficient resolution of conflicts captures the idea that norms emerge to avoid conflicts that arise between the members of a multiagent system. In addition, efficiency indicates that the norms that emerge avoid conflict without damaging performance. In other words, the norms should address the liveness-safety tradeoff [38] without, for example, achieving high safety at the cost of liveness.

Dynamic adaptation captures the idea that the norms in a multiagent system must be sensitive to its operating environment. That is, both the conflicts and the performance opportunities that agents face depend upon the environment in which they interact with each other. In other words, in addition to the individual member agents potentially adapting to the environment, the multiagent system as a whole would adapt through new norms.

Fairness refers to the idea that disparities in the outcomes across the various member agents are low. In general, there is a well-known tradeoff between the aggregate outcomes for the members of a society and the fairness of the individual outcomes received by its members [1]. An aspect of prosociality as we apply it here is that the agents perform actions that benefit others even if suboptimal to themselves. Therefore, in our approach the agents who are worse off are likelier to benefit from prosociality of others than the opposite, and hence fewer agents are poorly off, and the disparity in outcomes across the agents reduces.

Social welfare captures the idea of how much a society (as a whole) gains. Here, we specifically identify how a particular approach to coordination resource usage may lead to aggregate outcomes. This point can be framed as a question about the feasibility of a decentralized approach in producing social welfare relative to an approach that is fully centralized or incorporates elements of centralization through a distinguished entity to help coordinate.

1.2 Significance and Novelty

This paper synthesizes two important perspectives on prosociality. First, the individual decision-making perspective incorporates a notion of guilt based on inequity aversion [25], which posits that people may be self-interested, but their decisions are affected by how relatively poorly others fare. In a multiagent system, a prosocial agent in Cha would maintain some awareness of the outcomes received by other agents. When those outcomes are especially low for another agent, the first agent would notice the inequity and may feel guilt or exhibit aversion to that inequity. The agent would take decisions that lead to a reduction in that observed inequity.

Second, the societal perspective on prosociality is based on Rawls' [54] landmark theory of justice that focuses on designing a just society. This perspective is broadly supported by more recent works as well, for example, by Adler [1]. Specifically, Cha supports Rawls' doctrine of improving the outcome for whoever is the worst off. In other words, this doctrine advocates that we do not maximize

throughput if doing so leads to some agents starving. We adopt the Maximin formulation of Rawls' doctrine as a basis to measure fairness. By bringing in guilt, we can support norms that emerge in a bottom-up and adaptive manner without needing a central society enforcer.

1.3 Organization

Section 2 details the Cha framework. Section 3 describes how we model prosociality. Section 4 describes a simulated traffic intersection for evaluation. Section 5 discusses our results. Section 6 provides a summary of our contributions and an outlook for future work.

2 THE CHA NORMATIVE FRAMEWORK

Figure 1 depicts a schematic representation of an agent in Cha and different phases of its normative life cycle. Algorithm 1 outlines the decision loop for a Cha agent in four phases enacted via the agent's corresponding components: *Norm Generation*, *Norm Reasoning*, *Norm Updating*, and *Norm Sharing* (shown with labels in Algorithm 1). In this section, we describe norm representation in Cha and explain each phase of its normative life cycle along with the associated pseudocode that demonstrates the dynamics of Cha.
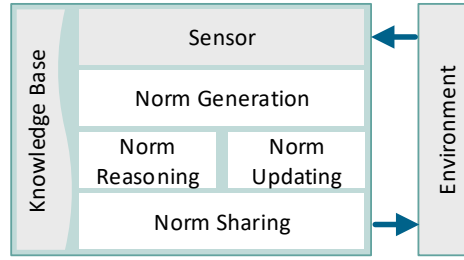

Fig. 1. Components of a Cha agent. Each Cha agent has sensors via which it perceives the environment, a knowledge base consisting of a set of norms (and norm structures), and components for generating, reasoning, updating, and sharing norms.

2.1 Norm Representation

As explained above, a norm in Cha is *regulative*: it characterizes how the agents ought to interact in specific situations. A "norm structure" is either a norm or a precursor to a norm that is based on a deontic operator proposed by García-Camino et al. [28]. We adopt a continuous notion of deontics [27], ranging from *prh* to *obl* with *may* in the middle. Initially, a norm structure is neutral. We express its neutrality using the operator *may*. The idea is that, as interactions take place, a norm structure may strengthen in one or the other direction or weaken back toward neutral. In cases of interest, a norm structure would strengthen enough in one direction or the other. Specifically, we say a norm emerges when the operator strengthens to *obl* or *prh*. Table 1 gives Cha's syntax.

Table 1. Syntax of a norm structure. Antecedent is a condition on the system state; Consequent is a deontic operator applied on an action.

Norm	::= ⟨Antecedent, Consequent⟩
Antecedent	::= Condition
Consequent	::= Operator(Action)
Operator	::= *may* | *obl* | *prh*

In Cha, given a state as it perceives, an agent applies any norm structure whose antecedent is true in that state (*Norm Generation*). It performs the action in the norm if its deontic is *obl*, does not perform it if the deontic is *prh*, and chooses either for *may* (*Norm Reasoning*). As the agent gains experience, *obl* or *prh* may begin to dominate, indicating a norm being learned (*Norm Updating*). Sections 2.2–2.5 describe these phases of normative lifecycle in detail.

2.2 Norm Generation Phase

An agent *perceives* its environment through its sensors and receives a local *view* (Line 2). A local *view* is a snapshot of the system from the agent's perspective. An agent's *view* is in general partial. The sensors are potentially specialized to each domain: a sensor is a domain-dependent function to map the state of the environment to a description of the state in the agent's representation. We express an agent's *views* using predicates over the state of the system. For example, in a traffic scenario, a possible *view* is the traffic in one part of the road that identifies positions and directions of movement of the other vehicles in reference to the agent's vehicle.

For each *view*, the agent checks whether a conflict would happen using the function f_{conflict} based on the possible next states, relying on domain information to determine which outcomes are undesirable. The agent gets the *view* v_t—the agent's view at time t, that would lead to conflict, and *conflictAgent* (Line 3). *conflictAgent* is the other agent involved in the conflicting *view* v_t.

Here *normSet* is the set of norms represented by the agent (Line 4). A norm applies to an agent if its antecedent matches the agent's current view. If no norm in normSet is applicable (Line 5), the agent generates a norm structure based on the current view (v_t) as its antecedent, initially with a *may* operator applied to the *action* that would lead to the conflict (Line 6). The generated norm structure is added to normSet (Line 7).

Algorithm 1 Decision loop for a Cha agent

1: **while** True **do**
      # Norm Generation:
2:    $view \leftarrow$ perceiveEnvironment(*Sensors*);
3:    v_t, conflictAgent $\leftarrow$ conflictDetect(*view*, f_{conflict});
4:    *normSet* $\leftarrow$ current norms;
5:    **if** $\forall \langle \text{ant}, \text{con} \rangle \in normSet : v_t \nvdash$ ant **then**
6:        $n \leftarrow \langle v_t, \text{may}(action) \rangle$;
7:        *normSet* $\leftarrow$ *normSet* $\cup \{n\}$;
      # Reasoning:
8:    $n \leftarrow$ getApplicableNorm(*normSet*);
9:    $a \leftarrow$ actionSelection(*n*, ϵ-*greedy*);
10:   *execute* (a);
11:   $a_c \leftarrow$ perceiveAction(conflictAgent, *Sensors*);
      # Updating:
12:   $r \leftarrow$ jointActionReward(*Payoffs(a,a_c)*);
13:   $U(n,t,a) \leftarrow$ *updateUtility* (n, t, a);
      # Sharing:
14:   *shareExperience()*;
15: **end while**

2.3 Reasoning Phase

The agent retrieves an applicable norm; selects an action according to its learning algorithm and an applicable norm, and performs the action (Lines 8–10). Next, the agent senses its conflicting agent's action, a_c (Line 11).

Cha agents apply reinforcement learning with the ϵ-greedy strategy for exploration and exploitation of the environment [72]. The ϵ-greedy approach offers two choices for each agent—select a random action (Exploration, with probability ϵ), or follow the applicable norm (Exploitation, with probability $1 - \epsilon$). An agent estimates ϵ via an exponential function (e^{-Em}), where E is a constant, and m is the number of times that the same situation has arisen before. Consequently, ϵ is high early (more exploration) and low later (more exploitation). Decaying exploration has been considered widely in the Reinforcement Learning literature [69]. The decay method is traditionally a linear or exponential decay. Exponential decay has been shown to lead to faster convergence [9].

2.4 Updating Phase

The agent updates its utility for the applied norm. After reasoning and selecting an action, based on the joint action of its conflicting agent, it receives a reward according to a payoff matrix (Line 12) and updates its utility via Equation 1 (Line 13).

$$U(n, t, a) = (1 - \alpha) \times U(n, t - 1, a) + \alpha \times r(n, t, a). \tag{1}$$

Here, $U(n, t, a)$ and $U(n, t - 1, a)$ stand for the utility of following norm structure n by performing action a at times t and $t - 1$, respectively; $0 \le \alpha \le 1$ is a parameter to trade off exploitation with exploration, and $r(n, t, a)$ indicates the reward based on a payoff matrix (examples in Section 4).

Each agent assigns two utility values to each norm structure for when the norm structure is, respectively, fulfilled (followed) and violated. Note that following or violating the norm structure eventually would lead to one of the operators with normative force, *obl* or *prh*, respectively, depending on whichever has the higher utility.

2.5 Sharing Phase

In Cha, each agent passes on its *experience* to incoming members of the same *type* (Line 14). Here, the experience is given by the utilities associated with different states and action outcomes and agents of the same type are those with the same goals. Thus, an incoming member obtains experience from members of the same type who have some experience to share. This approach fits in with technologies such as Vehicle-to-vehicle (V2V) communication. We assume that agents pass on their experience truthfully; false information [70] is out of our scope.

3 ACTING PROSOCIALLY IN CHA

The above approach may lead to unfair outcomes. For example, imagine that in a road traffic scenario, agent i has the right of way, and agent j (conflicting with agent i) must stop. In other words, the norm for agent i is initially *obl*(*Go*) and the norm for agent j is *prh*(*Go*). If there is heavy traffic in agent i's direction, agent j may have to wait for an arbitrarily long time. Long delays for some agents but not for others indicates unfairness.

In Cha, the agents act prosocially through inequity aversion [25]. An agent incorporates another's cumulative costs in its utility to help the latter. That is, instead of following an applicable norm that would benefit it, the agent performs an action that benefits the other party. In the traffic domain, such an action might be to wait and let another vehicle cross when you truly have the right of way. A benefit of using the cumulative cost from when an agent enters the system until it gets the opportunity to benefit from the resource is that it can help prevent starvation. We assume that agents know about each other's costs to be able to accommodate them in their decision making. This assumption is

consistent with real-world scenarios, in which people need to be aware of others' condition first, to act prosocial. In human societies, people can figure out each other's valuations enough to develop a sense of prosociality [63]. For artificial agents, this information must be either hardwired or explicitly communicated by the agents [30].

We incorporate *guilt* [41] as a *guilt disutility* to realize concessions. Please note that "guilt" here is not meant to capture psychosocial realism but is a convenient metaphor for a feeling that leads to prosocial behavior. Here, $\delta_{i,j}$, as computed in Equation 2, expresses the guilt perceived by agent i with respect to j in state s.

$$\delta_{i,j}(s) = -\beta_i \left(\max \left(f_j(s) - f_i(s) - c, 0 \right) \right). \tag{2}$$

Here, $f_x(s)$ computes the total cost paid by agent x until the present. An example of the calculation of $f_x(s)$ is discussed in Section 4 (Equation 3). Agent i's propensity toward guilt is captured in β_i: $\beta_i = 0$ means no guilt, and $\beta_i = 1$ means maximal guilt. Here, c is the threshold of inequity at which guilt kicks in.

Algorithm 2 adds prosociality to Algorithm 1 on Line 9 (actionSelection). For simplicity, we enable prosocial learning once the system has converged—otherwise, the system may never stabilize, especially with high values of β_i. Note that convergence is defined in Section 5.1. U_i^P, the utility incorporating prosociality is initialized to the converged utility (Lines 2–4). Below, $\neg n$ is the complement of norm n: it has the same antecedent but *obl* instead of *prh* or vice versa. That is, $\neg \langle p, obl(q) \rangle = \langle p, prh(q) \rangle$. Here, n is preferred by agent i and $\neg n$ by its conflicting agent j.

Agent i receives agent j's cost ($f_j(s)$, Line 5). If the difference in costs is below constant c, i follows norm n (Line 17). Otherwise, if agent i has not learned to concede ($U_i^P(n, t, a) > U_i^P(\neg n, t, a)$) (Line 6), i follows norm n (Line 11), adds its guilt disutility (a negative value) to its prosocial utility (Lines 8–9). Eventually, $U_i^P(n, t, a)$ would fall and agent i concedes to agent j (Line 14).

Algorithm 2 Prosocial learning strategy for agent i

1: **if** Converged **then**
2:     **if** U_i^P *Not Initialized* **then**
3:         $U_i^P(n, t, a) \leftarrow U_i(n, t, a)$
4:         $U_i^P(\neg n, t, a) \leftarrow U_i(\neg n, t, a)$
5:     *Receive* agent j's total cost, $f_j(s)$ via Sensors
6:     **if** $f_j(s) - f_i(s) > c$ **and** $U_i^P(n, t, a) > U_i^P(\neg n, t, a)$ **then**
7:
8:         $\delta_{i,j} = -\beta_i \left(\max \left(f_j(s) - f_i(s) - c, 0 \right) \right)$
9:         $U_i^P(n, t, a) = U_i^P(n, t - 1, a) + \delta_{i,j}$
10:         # *Follow* norm n
11:         $a \leftarrow$ actionSelection($n, \epsilon = 0$)
12:     **else if** $f_j(s) - f_i(s) > c$ **and** $U_i^P(n, t, a) < U_i^P(\neg n, t, a)$ **then**
13:         # *Concede* to agent j and *follow* norm $\neg n$
14:         $a \leftarrow$ actionSelection($\neg n, \epsilon = 0$)
15:     **else**
16:         # *Follow* norm n
17:         $a \leftarrow$ actionSelection($n, \epsilon = 0$)
18: **else**
19:     $a \leftarrow$ actionSelection(n, ϵ-greedy)

---

4 TRAFFIC SETTING AS ILLUSTRATION

We evaluate Cha in a simulated intersection understood as a multiagent system, where vehicle agents continually arrive and depart. We select this setting because it is powerful enough to illustrate Cha but not so powerful that details of the setting overwhelm our main points. Related works, for example, Mashayekhi et al. [42] and Morales et al. [45], adopt similar settings as in this paper. Please note that our focus is not to model the complexities of real-world traffic but to demonstrate the working of Cha. Cha could be potentially applied in more complex settings, for example, with a network of intersections.

As Figure 2 illustrates, we map an intersection and its vicinity to a grid. Traffic flows in four directions (north, south, west, and east). The *intersection zone (i-zone)* in the middle is composed of eight cells, highlighted in Figure 2. The light grey cells, including the center, are to be ignored. Cars travel along the grid at the speed of one cell per time tick. A vehicle may continue moving on a straight path or may randomly turn left or right in the i-zone. Agents of the same *type* are those traveling in the same direction, for example, all vehicles that are traveling east.
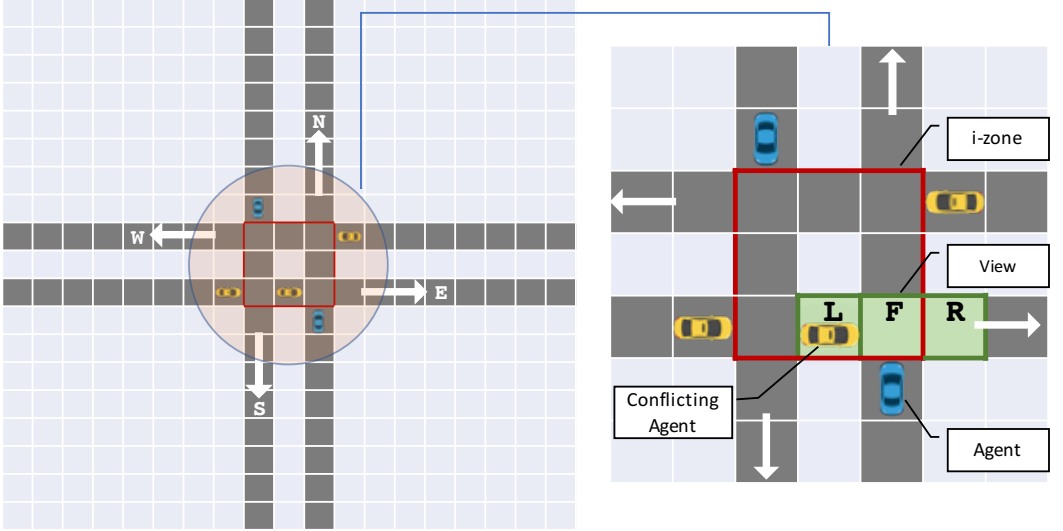

Fig. 2. The modeled intersection mapped to a grid of cells. Traffic flows in four directions (northbound, southbound, westbound, and eastbound). We consider the possibility of conflict within the i-zone. The i-zone is composed of eight cells in the middle (excluding the middle grey cell). From the perspective of the car specified as the Agent (shown in the bottom right of the inset picture), its view is composed of three cells that it must observe to detect and avoid conflicts. L refers to the cell to its left, F to the cell directly to its front, and R to the cell to its right. The car specified as Conflicting Agent is in conflict with the Agent.

A conflict arises when two *conflicting agents* are about to occupy the same i-zone cell together. Conflicting agents observe each other's actions without access to each other's internal policies (Algorithm 1 Line 11). If there is no norm in a conflicting situation, each agent creates a norm structure using the neutral operator, *may*. As in Section 2, a norm emerges as the agents gain experience and the deontic operator becomes stronger, namely, *obl* or *prh*.

The antecedent of a norm refers to the values of three cells (Left *L*, Front *F*, Right *R*) identified in Figure 2 with respect to the vehicle (agent) entering from the bottom of the intersection. *L*, *F*, and *R* constitute the view of that vehicle. Our grammar can specify cells with one of these six values:

→ (vehicle heading east), ← (vehicle heading west), ↓ (vehicle heading south), ↑ (vehicle heading north), ⊘ (empty), and ★ (any of four directions). An example norm, $L(\rightarrow) = prh(Go)$, means that an agent is prohibited from proceeding if it observes a vehicle in cell L heading east ($\rightarrow$). Figure 2 shows a similar case.

Table 2 shows our payoff matrix, which represents a *social dilemma* game, with the added twist of dynamism. Its constants are typical [2, 71]; the unselfishness term (that is, u) is novel to Cha. In a single round, the best positive payoff refers to the situation where one agent chooses a selfish action (*Go*) and the conflicting agent chooses the unselfish action (*Stop*). However, in a multi-round setting where the game is repeated, if the value of u is low, it is more rewarding to be cooperative than to be selfish, and if the value of u is high, it is more rewarding to be selfish than to be cooperative. Our flexible payoff matrix allows for the representation of various societies of agents.

Table 2. Payoff matrix (u is the unselfishnessCost).

| | | Agent j | |
		Go	*Stop*
Agent i	*Go*	−6.0, −6.0	5.0, u
	Stop	u, 5.0	u, u

The worst negative payoff relates to the situation when both agents are selfish, meaning that both decide to *Go*, thereby causing a collision. Equation 3 defines the payoff of an unselfish action for agent x in state s, that is, $f_x(s)$ in Section 3. The max ensures that the payoff of stopping is never worse than of a collision, and as a result the value of unselfish action ranges from 0 to −6.

$$unselfishnessCost_x(s) = \max(-d_x(s)^p, -6) \tag{3}$$

We use a power function to calculate the delay. Here, d_x, the delay experienced by agent x, equals the number of ticks that x must stop before entering the i-zone. The cost increases exponentially with delay. Specifically, we require the exponent of the delay term to be greater than 1, that is, $p > 1$. Intuitively, when the delay (for a specific agent) is high, the penalty that the agent assigns to stopping should be higher. Equation 3 has this property.

5 RESULTS: TESTING THE HYPOTHESES

To evaluate Cha in the traffic setting detailed in Section 4, we propose the following hypotheses based on the four themes — *efficient resolution of conflicts*, *dynamic adaptation*, *fairness*, and *social welfare* — that we motivate in Section 1. For brevity, we omit the corresponding null hypotheses indicating there are no gains under Cha.

- Hypothesis $H_{efficient}$: Cha leads to the emergence of norms that resolve conflicts while improving system-level performance outcomes.
- Hypothesis $H_{dynamic}$: Cha leads to the emergence of norms that can adapt based on changes to the environment. That is, an emerged norm can fade out once the environment changes and be replaced by a new norm.
- Hypothesis $H_{fairness}$: Cha leads to the emergence of norms under which fairness is achieved. Specifically in Cha, the agents observe the outcomes of others and act toward the benefit of those poorly off, leads to improved fairness.
- Hypothesis H_{social}: Cha leads to the emergence of norms that yield higher societal gains than both a representative central and a representative hybrid approach to sharing resources.

We now describe our simulation setup to evaluate Cha and discuss each of the hypotheses, that is, $H_{conflict}$, $H_{efficient}$, $H_{dynamic}$, $H_{fairness}$, and H_{social}, in greater detail.

5.1 Simulation Setup

We model a traffic intersection (19 cells per lane: 72 cells in all with eight cells in the i-zone) as an environment in Repast [52].

Our results are averaged over 1,000 trials. Convergence is considered to happen if the utilities associated with the norms converge to within 10^{-3} (our *convergence parameter*).

We set the initial utility values to 0 at $t = 0$, $\alpha = 0.2$ in Equation 1 [45]. We set $E = 0.05$ in the exponential function (e^{-Em}) used to set the exploration probability in the ϵ-greedy exploration approach in the reasoning phase [23].

The impact of delay, that is, p, the exponent of the delay term in Equation 3 is set to 1.1 for all agents. We choose this value somewhat arbitrarily because adjusting it would affect only the rate of convergence in our experiments, not whether convergence takes place.

For the fairness experiment to test Hypothesis $H_{fairness}$, we need to set the following parameters.

- The propensity or weight of guilt, that is, β in Equation 2. Note that $\beta = 0$ means no guilt and leads to no fairness and $\beta = 1$ means maximal guilt, which leads to repeated yielding by agents in favor of others. For simplicity of evaluation, we set the same value of β, specifically, $\beta = 0.5$, for all agents.
- The threshold of inequity tolerated by the agents, that is, c in Equation 2. For simplicity, we set the same value of c, specifically, $c = 4$, for all agents.

  Note that the fairness parameters' values (i.e., c and β) do not affect the convergence, since as mentioned in Section 3, we enable prosocial learning once the system has converged—otherwise, the system may never stabilize, especially with high values of beta. Fairness parameters' values affect the rate at which the agents become prosocial.

5.2 Testing the Efficient Resolution Hypothesis

Hypothesis $H_{efficient}$ states that emergent norms improve system-level goals—yield lower total average delays across an intersection. The corresponding null hypotheses indicate that there are no improvements in the system-level goals under Cha.

We consider a static setting in which we fix the traffic flows for the north-south and east-west directions. We first set the traffic flow distributions as the same for north-south and east-west directions. We observe that approximately half of the time (508 out of 1,000 simulation runs), north-south vehicles learn to Go, and east-west vehicles learn to Stop in conflict situations. In the remaining 492 times, the reverse norm arises. The population converges to one or the other norm depending on whether Go or Stop is more common for east-west than for north-south vehicles and thus minimizes collisions.

Our next setting is also static but with unequal traffic: the north-south orientation has a (30%) higher traffic volume than the east-west orientation. Figure 3 shows the total number of collisions per 1,000 ticks. Since there are four cells in the simulated intersection that have the potential of conflict, the maximum number of collisions is four per tick. After about 20,000 ticks, the number of collisions decreases dramatically. After 25,000 ticks (not shown here for brevity), the changes in the average utility converge to within 10^{-3}, our convergence parameter.

Based on the asymptotic reduction in collisions and the convergence in utilities, we conclude that vehicle agents have learned new norms to avoid collision. Table 3 shows the emergent norms: east-west vehicles learn to Stop in conflicting situations with north-south vehicles. Since the north-south orientation has higher traffic volume, the converged norms are efficient—average delay is lower when vehicles in the direction with the lower volume Stop and in the direction with the higher volume Go, than the other way around. Norms emerged in this experiment provide evidence to support Hypothesis $H_{efficient}$.

Table 3. Emerged norms for static setting with equal and unequal traffic for the first half and all the time, respectively.

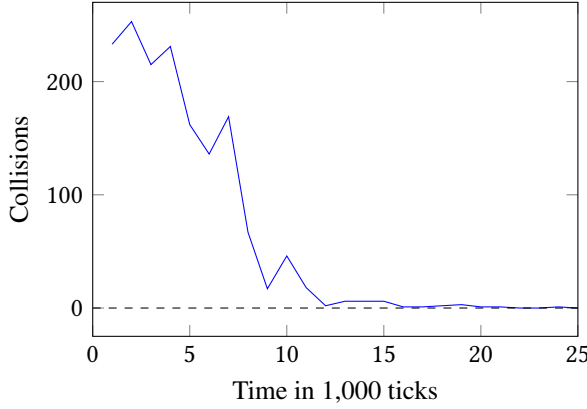|	Precondition	Modality
Eastbound and	$L(\star) \wedge F(\star) \wedge R(\leftarrow)$	$prh(\text{Go})$
Westbound	$L(\rightarrow) \wedge F(\star) \wedge R(\star)$	$prh(\text{Go})$
Southbound and	$L(\star) \wedge F(\star) \wedge R(\leftarrow)$	$obl(\text{Go})$
Northbound	$L(\rightarrow) \wedge F(\star) \wedge R(\star)$	$obl(\text{Go})$

Fig. 3. Total number of collisions per 1,000 ticks.

5.3   Testing Dynamic Adaptation Hypothesis

The Hypothesis H_{dynamic} states that Cha adapts to environmental changes—changes in traffic flow through an intersection. The corresponding null hypothesis indicates that Cha does not adapt to environmental changes. In this experiment, we start with a fixed traffic flow distribution where the north-south orientation has 30% higher traffic than the east-west orientation, let it converge, and then reverse the pattern (that is, the east-west orientation ends up with 30% more traffic). Doing so helps us determine whether norms learned in one setting persist when the traffic changes.

To test Hypothesis H_{dynamic}, we (1) measure root mean square deviation (RMSD) of average utility in a sliding window of 1,000 ticks, computed as $\text{RMSD}_t = \sqrt{\frac{\Sigma_{t=1}^n (x_i - \bar{x})^2}{n-1}}$, where t is the current tick; x_t is the utility in the current tick; $\bar{x}$ is the average utility in the current sliding window; and n is the window size, and (2) perform a two-sample Kolmogorov-Smirnov (KS) test on successive sliding windows.

Figure 4 shows the change in average utility (for a sliding window of 1,000 ticks) for westbound vehicles. By $t \approx 25,000$ ($\text{RMSD}_{25,000} = 0$ for east-west vehicles Stop; $p < 0.01$), the norm learned by east-west vehicles is to Stop in case of conflict, just as in Table 3 and Figure 3. We highlight this norm in Figure 4 with the shaded box in the middle.

After $t = 44,000$, the traffic pattern is reversed, and by $t \approx 50,000$ ($\text{RMSD}_{50,000} = 0$ for east-west vehicles Go; $p < 0.01$), the new norm is for east-west vehicles to Go in case of conflict. We highlight this norm in Figures 4 and 5 with boxes in the right parts. Eastbound and northbound vehicles have the same outcomes as westbound and southbound vehicles, respectively.
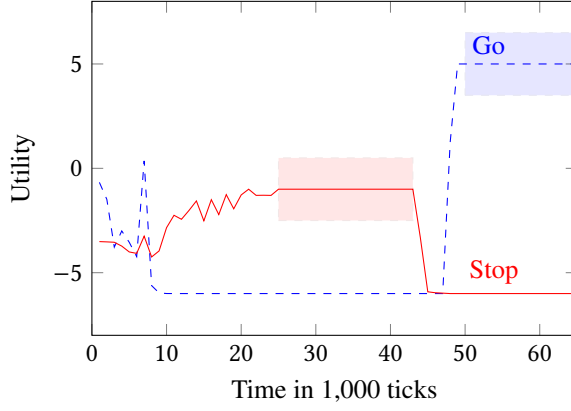
Fig. 4. Utilities of westbound vehicles for the dynamic setting. The reported figures are averaged over a window size of 1,000 ticks.
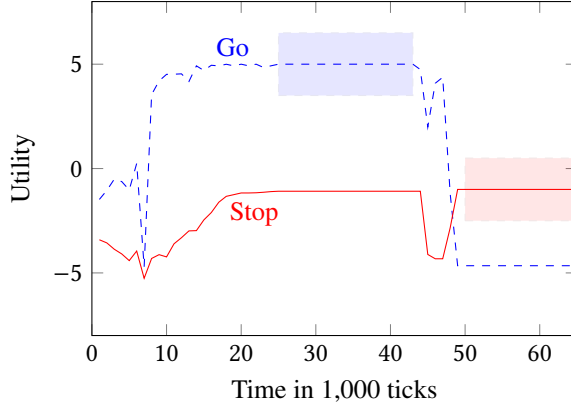

Fig. 5. Utilities of southbound vehicles for the dynamic setting. The reported figures are averaged over a window size of 1,000 ticks.

Table 4 shows how norms change when we reverse the traffic pattern, thus supporting Hypothesis $H_{dynamic}$.

Table 4. The norms as emerged in the dynamic setting when the traffic flow has been reversed.

	Precondition	Modality
Eastbound and Westbound	$L(\star) \wedge F(\star) \wedge R(\leftarrow)$	$obl(\text{Go})$
	$L(\rightarrow) \wedge F(\star) \wedge R(\star)$	$obl(\text{Go})$
Southbound and Northbound	$L(\star) \wedge F(\star) \wedge R(\leftarrow)$	$prh(\text{Go})$
	$L(\rightarrow) \wedge F(\star) \wedge R(\star)$	$prh(\text{Go})$

Note that, given that the agents have developed a notion of norms already, they tend to adapt faster to environmental changes, that is it takes $t \approx 25{,}000$ ticks to initially learn the norms, and $t \approx 6{,}000$ ticks to adapt to new environmental changes and develop new norms. If the flows of cars invert

intermittently (e.g., every x ticks), as long as the final traffic difference is significant (e.g., 30%), we still would expect Cha to be adaptive, however, it could take longer to converge.

5.4 Testing the Fairness Hypothesis

Hypothesis $H_{fairness}$ concerns disparities in resource allocation—excessive delays for some vehicles to enter the intersection while others proceed quickly. We define a fair society as one that supports the Maximin criterion [54, p. 153]. In the present setting, a fair society means one in which no agent is deprived of resources for long periods of time. We understand fairness (or the absence of fairness) as an outcome of different norms.

Consider the prosocial learning strategy given in Algorithm 2. As in Section 5.2 (unequal traffic setting), we set north-south flows to have 30% more traffic than east-west flows. We saw that east-west vehicles learned to Stop in conflicting situations with north-south vehicles. Now, we verify whether a north-south vehicle can act prosocially in a conflicting situation by yielding to an east-west vehicle if it experiences a delay above a certain threshold.

Figure 6 shows the change in the average prosocial utility (U_i^P) of southbound vehicles. Prosocial learning can be activated after convergence; we activated it at $t = 40,000$. By $t \approx 51,000$, the agents have learned to be prosocial. Northbound vehicles (figure omitted for space) show the same trend as southbound vehicles.

Below, Cha Neutral refers to the variant form without prosociality, i.e., without incorporating the guilt disutility. We evaluate Cha Neutral and Cha's performance in terms of delays. Table 5 shows the percentiles of delays over the population of vehicles. For example, 99.5% of all vehicles are delayed four or fewer ticks.

We adopt percentile values as a metric for fairness because the distribution of latency has a long tail. Specifically, agents who suffer excessively would be those concentrated in the high percentiles even though the mean delay may not vary much between fair and unfair outcomes. We compute *skewness* (γ) and *kurtosis* (κ) for the delay distribution. Skewness is defined as a measure of symmetry or lack of it [18]. A distribution is symmetric if it looks the same to the left and right of the center point. A distribution with a lower (closer to zero) skewness value is fairer. Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution [18]. A heavy-tailed distribution for the time delay indicates that a small number of agents suffer excessive time delay. A distribution with low kurtosis (light-tailed) for the time delay is fairer than a distribution with high kurtosis (heavy-tailed).

Figure 7 shows the delays of four or more ticks with and without prosociality. We observe that, by reduction, Cha which incorporates prosociality, improves the outcome for 35% of vehicles that experienced a worst-case delay of six ticks. Specifically, 225 (that is, 35%) of the 650 vehicles which experienced the highest delay (six ticks) under Cha Neutral (without prosociality), experience a delay of five ticks or fewer under Cha *with prosociality*. Each of these vehicles saw a delay reduction of at least $\frac{6-5}{6} = 16.67\%$. Cha thus promotes Rawls' [54] Maximin doctrine. Also, note that, considering the values in Figure 7, we have 425 ticks less delay for Cha *with prosociality* (i.e., $[2117 \times 4 + 973 \times 5 + 425 \times 6] - [1917 \times 4 + 948 \times 5 + 650 \times 6] = -425$). So not only Cha *with prosociality* improves the total mean delay, but also it impacts the heavy tail.

We also compute the Gini coefficient (*Gini*) to measure the degree of disparity in a distribution. A lower *Gini* indicates a lower degree of disparity. Table 6 lists *Gini* for the vehicles in 99, 99.5, 99.7, and 99.9 percentile population. We observe that *Gini* for 99, 99.5, and 99.7 percentile populations is lower for Cha compared to Cha Neutral, indicating lower disparity. However, *Gini* for 99.9 percentile population of vehicles is higher in Cha compared to Cha Neutral. Under Cha Neutral, all vehicles face the worst-case delay (six ticks), and thus the disparity is lower, whereas under Cha, the delay is either five ticks or six ticks.
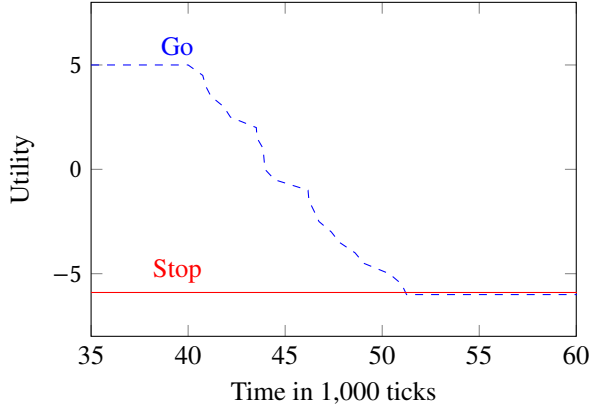
Fig. 6. Prosocial utilities of southbound vehicles.

Table 5. Delays with Cha Neutral and Cha.

	Percentiles					γ	κ
	99	99.5	99.7	99.9	100		
Cha Neutral	3	4	5	6	6	2.07	4.99
Cha	3	4	**4**	**5**	6	2.02	4.55

γ = Skewness; κ = Kurtosis

Table 6. Disparity with Cha Neutral and Cha. Lower is better.

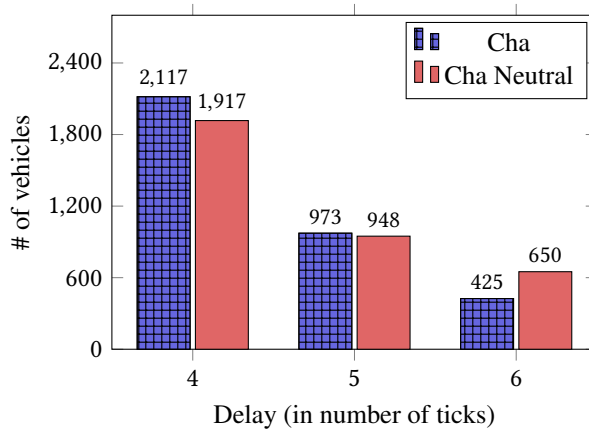| Gini Percentiles	99	99.5	99.7	99.9
Cha Neutral	0.19	0.20	0.16	5.87×10^{-17}
Cha	0.17	0.17	0.15	0.11

Fig. 7. Benefit from prosociality. Delays ($\geq$ 4) with Cha Neutral and Cha. 35% of vehicles that experience the highest delay (6 ticks) with Cha Neutral experience an improvement of $\geq$ 16.67% with Cha.

5.5 Testing the Social Welfare Hypothesis

Hypothesis. We take Hypothesis H$_{social}$ as stating that Cha wins (in aggregate societal gains) over representative central and hybrid approaches. The corresponding null hypothesis indicates no gains under Cha.

Metric. Our metric is mean travel time for the whole intersection. It is calculated by adding the best-case travel time (that is, 19 ticks, one for each cell in each lane) to the mean delay (mean number of stops for all vehicles in each tick.

Baselines. There are two main traffic control paradigms, pretimed (in which the signals change in a fixed pattern) and actuated (in which signals change in response to vehicles arriving at the intersection from different directions). The actuated paradigm is known to deliver higher performance than pretimed. Therefore, we adopt a variant of the actuated paradigm, called the fully actuated control strategy [73] as one of our baseline approaches. To compare to previous multiagent systems research, we also adopt the Silk [42] approach as another baseline because Silk is a hybrid approach and studies a similar social dilemma situation in a similar traffic setting. We highlight the difference between Cha and Silk in Section 6.1.

Fully actuated control is a traditional approach to regulating traffic. In fully actuated control, each *movement*, for example, the northbound traffic flow at an intersection, has a detector. A *phase* is a combination of nonconflicting movements. We pair northbound and southbound flows to create one phase and eastbound and westbound flows to create the other. We assume fully actuated control alternates between these phases. Each phase has a *minimum green* (set to one in our experiment); thereafter, the green can be extended indefinitely provided a vehicle keeps on being detected at the intersection.

For our comparison with Silk's framework, we used the payoffs of Table 2 in both Silk and Cha. Since Silk supports only static settings with fixed payoffs, we set the payoff of *Stop* in Silk to zero.

Elaborating on Hypothesis H$_{efficient}$ of Section 5.2, we consider a specific hypothesis that Cha yields a lower mean travel time than both Silk and fully actuated control. These experiments are amenable to statistical hypothesis testing: the null hypothesis is that there is no significant difference in mean travel times yielded by Cha, Silk, and fully actuated control. We run these experiments for the case where the north-south orientation has 30% higher traffic than the east-west orientation.

Figure 8 compares the average delay in Cha, Silk, and fully actuated control. Table 7 summarizes the travel time results. It lists the best-case travel time, mean travel time (μ_{travel}), and standard deviation (σ_{travel}) for Cha, Silk, and fully actuated control. Table 7 also lists the *p*-value from two-tailed *t*-test assuming unequal variances, effect size via Glass' Δ [29], which is measured as the difference in means divided by the standard deviation of the control group), and % improvement obtained by Cha over Silk and fully actuated control. We choose Glass's Δ to measure effect size since the standard deviations (σ) for the two groups are different.

We find that Cha reduces the delay and yields significantly less travel time than both Silk ($p < 0.01$; $\Delta = 2.06$) and fully actuated control ($p < 0.01$; $\Delta = 1.76$), reducing it by 11.79% over Silk and 18.55% over fully actuated control. Following Cohen's [17] guidelines, an effect of over 0.80 is large; the above effect sizes are substantially above Cohen's guideline.

6 DISCUSSION

Cha is a flexible and dynamic framework for norm emergence in multiagent systems that supports prosocial outcomes. In Cha, each agent reasons individually about which norms to develop. Our results support our hypotheses.

Table 7. Travel time: Cha versus Silk and Fully actuated control.

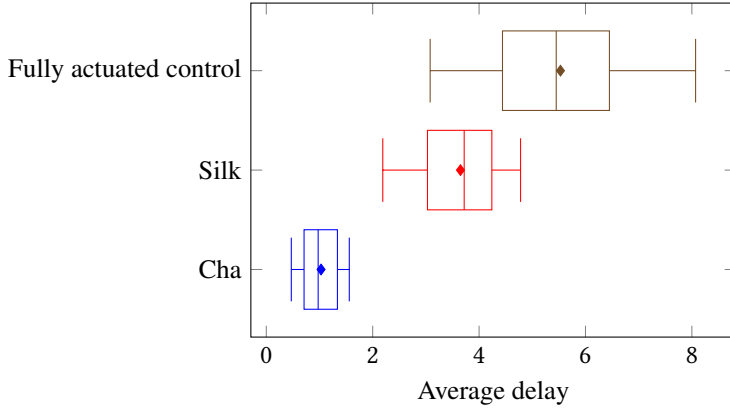|	Cha	Silk	Fully actuated control
Best case	19.00	19.00	19.00
Mean μ_{travel}	19.98	22.65	24.53
Stdev σ_{travel}	0.55	1.28	2.56
% Improvement	–	11.79	18.55
p-value	–	< 0.01	< 0.01
Glass' Δ	–	2.06	1.76

Fig. 8. Comparing average delay in Cha ($\mu = 0.98$), Silk ($\mu = 3.65$), and fully actuated control ($\mu = 5.53$). The benefits of the prosocial approach are clear.

One, norms that emerge in Cha resolve conflicts efficiently and improve system-level outcomes. Two, Cha is responsive to changes in the environment. Three, more importantly, in light of recent increasing awareness to promote and foster prosociality in artificial intelligence generally and multiagent systems, Cha supports prosocial outcomes, specifically, fairness in resource allocation [50]. Cha incorporates prosociality based on inequity aversion and naturally respects Rawls' Maximin doctrine to improve the worst-case outcome on members of the society. To our knowledge, no other approach tackles the concerns of prosociality in the study of norm emergence in multiagent systems. Four, Cha yields higher societal gains than both a representative central approach and a hybrid approach (baselines in our evaluation).

Note that Cha applies reinforcement learning to norms. Unlike work on social learning, Cha shows how incorporating guilt disutility can lead to prosocial decision making and thus to fair societal outcomes. Also, Cha does not require repeated interactions in a fixed graph. Departing agents can convey their experience to those who come after them, which facilitates norm emergence by giving endurance to the system even though individual agents have short lifetimes. Note that Cha automatically garbage collects useless norms. When the system changes, new norms emerge, and some old norms are unlearned (they lose utility and are no longer selected).

It is worth mentioning that Cha being a dynamic and decentralized framework, not only supports beginning from an initially empty set of norms but can also be used when considering an initial set of existing norms, on top of which the agents can then build new norms.

A potential critique of this work is that even though the agents are decentralized, there is a system-wide notion of conflict that leads to their behaviors being coordinated. Arguably, this setup accords

with Elsenbroich and Verhagen's [2016] idea of how functionality and complexity may shift between an agent and its context. However, the conflicting situations in our study are motivated by the domain of traffic and earn their legitimacy from that domain. In another domain, the actions and conflicts would be different. But the conflicts do not determine the outcomes and other approaches with a similar conflict structure can produce different results.

6.1 Related Work

It is worth highlighting how Cha differs from the recent literature on norm emergence. Note that we do not intend this article to be a survey of norms or normative multiagent systems in general: see, for instance, Hollander and Wu [34], Nardin et al. [51], and Andrighetto et al. [7].

Static and Closed. Many of the existing game-theoretic approaches for social norm emergence consider agents in a closed system such a graph structure, and fixed payoffs are considered for the agents' behaviors [2, 11, 60, 71, 76]. In these types of studies, agents are nodes of a graph, and the interactions between them create the edges of the graph. Agents learn their policies by playing repeated games with random agents. Through the course of these repeated games, new norms emerge in the system. In contrast, Cha considers a system that is open (agents continually leave and enter), dynamic, and addresses prosociality.

Static and Centralized. Other works consider norm emergence in a setting where a central agent with complete knowledge of the systems generates (or dictates) the norms. Morales et al. [45] proposed a mechanism called IRON along with two other variations [46, 47] for the on-line synthesis of norms. IRON, however, could result in unstable normative systems. To mitigate the instability problem, Morales et al. [48] presented an offline method called "System for Evolutionary Norm SynthEsis" (SENSE), which builds on top of IRON. These mechanisms employ a central approach that relies on global knowledge. The central agent observes the interactions of members to synthesize conflict-free norms in a top-down manner. Even though these frameworks share some commonalities with Cha in terms of norm representation and norm generation (albeit static), they are fundamentally different from it. Whereas these mechanisms use a central norm generator and learner, Cha uses a dynamic and decentralized mechanism for its four phases, as described in Section 2. Additionally, Cha agents incorporate prosocial decision making to achieve fairness. In this manner, Cha brings together two important themes in prosociality: decision making by individuals and fairness of system-level outcomes.

Static and Decentralized. Hao et al. propose two strategies based on local exploration and global exploration to facilitate norm emergence in a bottom-up manner [31]. Whereas Hao et al.'s strategies focus only on maximizing average payoff for all agents, i.e., social welfare, Cha's focus is on maximizing both social welfare and fairness. Mihaylov et al. [43] propose a decentralized approach for convention emergence in multiagent systems for a pure coordination game with fixed payoffs. Once they have learned a convention, agents will not re-learn it (i.e., have a static environment).

Static and Hybrid. Silk [42] is the closest framework to Cha in the literature. Silk is a hybrid framework for the regulation of open normative systems. Briefly, this framework is composed of a central generator, which imposes hard integrity constraints that agents must follow, and recommends norms as soft constraints. Thus, a norm can emerge based on the decision of various agents to accept or deny the recommendations. In addition, Silk's framework supports only static settings with fixed payoffs. Moreover, Silk tackles the problem of fairness through its central generator which continually monitors the performance and intervenes if necessary to improve fairness, whereas Cha agents learn to be prosocial.

Dynamic. Knobbout et al. [40] propose update semantics for norm addition to characterize the dynamics of norm addition in a formal way. Updates are parameterized over actions. However, this work does not indicate whether its proposed model can be applied to regulate agent's actions and satisfy system-level goals. Verhagen [74] puts worth the need for flexible and decentralized system. The simulation in Verhagen [74] focuses on spreading and internalization of norms in such systems. However, Verhagen assumes that a top-level entity (say, a normative advisor) is aware of the correct norm, and this group norm (g-norm) does not change. This g-norm is spread to the agents through the normative advice provided using a top-down approach by a centralized authority (i.e., there is a leader in the society). In contrast, Cha does not enforce a need for a central authority and the norms in Cha are not fixed. Savarimuthu et al. [57, 58] propose an architecture where agents can identify norms of the society in a bottom-up fashion. They address how an agent might be able to identify whether a norm is changing in society and how it might react to this situation. However, Cha is different from this work in that Cha is an on-line framework; that is, in Cha, norms are learned rather than inferred through data mining.

Prosocial. Recently there has been increasing interest in designing multiagent systems that lead to prosocial behaviors in agents [4, 53, 55, 62]. Prosocial behavior is when an agent performs an action that is unfavorable to itself but benefits others. We consider the problem of agents prosocially acting while sharing resources in a multiagent system. The connection between norms and prosociality, though conceptually important, has not received adequate attention: existing AI approaches on prosociality focus on individual decision making by agents. In contrast, we relate social norms both to prosocial decision making and to societal outcomes such as fairness. Specifically, the norms through which agents coordinate their interactions may or may not be fair.

Table 8 summarizes the difference between Cha and some leading frameworks addressing a similar setting.

Table 8. Comparison of characteristics of simulation works on norms (Yes – Considered; No – Not Considered; NA – Not Applicable; NS – Not Specified)

Model	Creation	Identification	Architecture	Openness	Dynamism	Prosocial
Cha	on-line	RL	Decentralized	Yes	Yes	Yes
Morales et al. [45–47]	on-line	Case-based Reasoning	Centralized	No	No	No
Mashayekhi et al. [42]	on-line	RL	Hybrid	Yes	No	Yes[*]
Sen and Airiau [60]	on-line	RL	Decentralized	No	No	No
Airiau et al. [2]	on-line	RL	Decentralized	No	No	No
Villatoro et al. [76]	on-line	RL	Decentralized	No	No	No
Savarimuthu et al. [57, 58]	on-line	Data Mining	Decentralized	Yes	Yes	No
Beheshti et al. [11]	on-line	Game Theory & RL	Decentralized	No	No	No
Frantz et al. [27]	on-line	RL	Decentralized	No	Yes	No
Hao et al. [31]	on-line	RL	Decentralized	No	No	No

[*]Enforced by Central Generator

6.2 Prosociality in General

Applications of our approach on practical domains would be important to validate it in broader settings.

Once following certain conditions are met, Cha can be successfully applied to other settings besides the traffic one described here:

- *Agents have asymmetric interests.* Cha assumes its population of users has asymmetric interests that may lead to conflicts.

- *Conflicts must be detectable*. Norm generation in Cha is based on detection of the conflicts (e.g., car collision), so Cha assumes conflicts are detectable.
- *Some agents' preferences can be assumed at the design time*. Cha assumes some domain information such as payoffs are provided by the system designer (e.g., the payoff of car collision).
- *Agents can observe and communicate*. Cha assumes its members are able to share their experience to agents of the same type. Additionally, agents of different types are able to receive and send accumulated costs in order to learn and act prosocially.

A natural family of application scenarios for Cha is the development of what Berners-Lee [12] calls *social machines*. We can think of social machines as sociotechnical systems [15] geared to support interactions between people or people and businesses in reference to some shared resources or coordinated decision making. Social machines today are largely not formulated to satisfy our desiderata but they could and should be. Today, there is little computational support for prosociality, flexibility, or dynamism and the underlying architecture being based on servers is decidedly centralized. Thus, these desiderata are left to the human participants to achieve informally. But we can envision a situation where decentralized agents assisted the humans and provided computational support for prosociality, flexibility, and dynamism.

Another domain of application for Cha is on-line communities, in which members continually interact by exchanging information (e.g., comments and pictures) in different forums. An on-line community is an open, dynamic system. It is open because the users can enter the system T any time. It is dynamic because users may also change their preferences and behaviors. The goal is to achieve "healthy" communities [33]. Specifically, the goal is to avoid circumstances in which a user expresses concern about the content involving that user that is shared by other users. Such content may include pictures that violate one user's privacy but may be funny in the eyes of the posting user [26]. Prosociality would be demonstrated by the posting user taking into account the preferences of the people shown in a picture. Flexibility is indicated by the users not having rigid stances for their preferences, which can shift depending on the context. Endurance is indicated by the users interacting multiple times, sometimes over the course of years. Dynamism is indicated by the fact that norms for what is acceptable (e.g., in regards to sensitive information) may change depending on the cultural or political situation. Decentralization is indicated by the users deciding individually on norms without any computational limitations imposed by any sharing platform they may be using. The users could concede to each other and positively or negatively sanction each other.

The domain of Open Source Software Development (OSSD) communities [8] is particularly compelling because it brings together considerations of ethics along with inertia (projects can last years) and group norms. An OSSD community is open and dynamic. With a defined goal, one or more developers start the development of an open source software project. Other developers who are interested in the project, contribute to the project's goal by writing software code. Prosociality in OSSD would be demonstrated by the community of developers (and users outside of the developer community) identifying issues with the ongoing project and voluntarily contributing fixes to resolve the identified issues. Similar to in on-line communities, endurance in OSSD communities is indicated by the users interacting multiple times over the course of years. Flexibility is indicated by the developer community allowing users to submit feature requests and subsequently redefining the project's goals. Dynamism is indicated by the fact that individual and group norms as well as the project's goals may evolve over time as the community grows. Decentralization is indicated by the OSSD community developers individually deciding on the norms on how they interact with other developers and users, and how they sanction them.

Another example scenario where the ideas of Cha could apply is in shared transportation. Bardaka et al. [10] describe the setting of public microtransit (i.e., last-mile ride sharing) in which a public van service is used to help members of the public who cannot use a private car (e.g., because of its expense or their poor health). This setting demonstrates the features necessary to achieve the desiderata identified for Cha. Prosociality would be demonstrated by riders adjusting their requested ride's origin, destination, or timing to accommodate the needs of other riders. Flexibility is indicated by the riders changing how they interact, e.g., in response to environmental events (is it raining) or their health (does one of them need to get to a dialysis appointment?). Endurance is indicated by the riders interacting with each other again and again (e.g., every weekday), sometimes over the course of years, because the people who need such assistance in a neighborhood are likely to be sharing rides more than once. Dynamism is indicated by the fact that norms for what is acceptable may change depending on the public health (e.g., in regards to an epidemic flaring up or dying down) or economic (e.g., during a recession, money is tight and schedules may be more rigid for workers) situation. Decentralization is indicated by the riders deciding individually on norms on how they interact. The users could help each other, thank each other for accommodating each other's requests, or negatively sanction each other, e.g., through shunning each other's company.

6.3 Outlook and Directions

An important future direction is to understand the context as a basis for overriding norms. For instance, can we allow vehicles with a sick passenger to Go when the norm suggests Stop? Such overrides may be facilitated by sharing context and explanations to justify the ethicality of deviating from a norm [3].

Another direction is to include more complex norm actions, for example, (1) communicating an obligation to Stop to let an ambulance pass; and (2) if in a rush, delegate (to the vehicle behind) a norm to help a stranded vehicle. One way to model more complex norm actions is through sanctioning, which also corresponds to an additional (state, action) pair in the payoff matrix.

Understanding societal inertia in the sense of how new norms can supersede existing norms is important; Cha is dynamic but how to evaluate and how to reduce the societal friction and inertia in arriving at new norms remains to be studied.

Cha has a flavor of group norms since agents of the same type pass on their experiences with other agents of the same type. But each agent individually reasons whether to follow an applicable norm or not. Another natural line of future research is to explore the generation of group norms as opposed to individual norms [5, 6] bring together notions of group intentions [22, 66] and group ability [65]. Group norms apply to groups of agents together. For example, in the world of autonomous cars, car agents sharing the same space can share information with each other and can decide actions and move together as a platoon of cars (i.e., creating a group norm). Group norms raise important challenges of how a group can allocate decision-making authority and accountability and whether the actions taken on behalf of a group satisfy ethical criteria both with respect to group members and with respect to outsiders.

ACKNOWLEDGMENTS

REFERENCES

[1] Matthew D. Adler. 2019. *Measuring Social Welfare: An Introduction*. Oxford University Press, New York.

[2] Stéphane Airiau, Sandip Sen, and Daniel Villatoro. 2014. Emergence of conventions through social learning. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)* 28 (2014), 779–804.

[3] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. 2018. Robust Norm Emergence by Revealing and Reasoning about Context: Socially Intelligent Agents for Enhancing Privacy. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*. IJCAI, Stockholm, 28–34. https://doi.org/10.24963/ijcai.2018/4

[4] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. 2020. Elessar: Ethics in Norm-Aware Agents. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. IFAAMAS, Auckland, 16–24. https://doi.org/10.5555/3398761.3398769

[5] Huib Aldewereld, Virginia Dignum, and Wamberto W. Vasconcelos. 2016. Group Norms for Multi-Agent Organisations. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 11, 2 (June 2016), 15:1–15:31.

[6] Natasha Alechina, Wiebe van der Hoek, and Brian Logan. 2017. Fair Decomposition of Group Obligations. *Journal of Logic and Computation* 27, 7 (Oct. 2017), 2043–2062. https://doi.org/10.1093/logcom/exx012

[7] Giulia Andrighetto, Cristiano Castelfranchi, Eunate Mayor, John McBreen, Maite López-Sánchez, and Simon Parsons. 2013. Social Norm Dynamics. In *Normative Multi-Agent Systems*. Schloss Dagstuhl, Dagstuhl, Germany, 135–170.

[8] Daniel Avery, Hoa Khanh Dam, Bastin Tony Roy Savarimuthu, and Aditya K. Ghose. 2016. Externalization of Software Behavior by the Mining of Norms. In *Proceedings of the 13th International Conference on Mining Software Repositories (MSR)*. ACM, Austin, Texas, 223–234.

[9] Avraham Bab and Ronen I. Brafman. 2008. Multi-Agent Reinforcement Learning in Common Interest and Fixed Sum Stochastic Games: An Experimental Study. *Journal of Machine Learning Research* 9, 88 (2008), 2635–2675. http://jmlr.org/papers/v9/bab08a.html

[10] Eleni Bardaka, Leila Hajibabai, and Munindar P. Singh. 2020. Reimagining Ride Sharing: Efficient, Equitable, Sustainable Public Microtransit. *IEEE Internet Computing (IC)* 24, 5 (Sept. 2020), 38–44. https://doi.org/10.1109/MIC.2020.3018038

[11] Rahmatollah Beheshti, Awrad Mohammed Ali, and Gita Reese Sukthankar. 2015. Cognitive Social Learners: An Architecture for Modeling Normative Behavior. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, Austin, 2017–2023.

[12] Tim Berners-Lee. 1999. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. Harper Business, New York.

[13] Guido Boella and Leendert W. N. van der Torre. 2004. Regulative and Constitutive Norms in Normative Multiagent Systems. In *Principles of Knowledge Representation and Reasoning: Proceedings of the 9th International Conference (KR)*. AAAI Press, Whistler, Canada, 255–266.

[14] Rémy Chaput, Jérémy Duval, Olivier Boissier, Mathieu Guillermin, and Salima Hassas. 2021. A Multi-Agent Approach to Combine Reasoning and Learning for an Ethical Behavior. In *Proceedings of AAAI/ACM Conference on AI, Ethics, and Society (AIES)*. ACM, Virtual, 13–23. https://doi.org/10.1145/3461702.3462515

[15] Amit K. Chopra and Munindar P. Singh. 2016. From Social Machines to Social Protocols: Software Engineering Foundations for Sociotechnical Systems. In *Proceedings of the 25th International World Wide Web Conference*. ACM, Montréal, 903–914. https://doi.org/10.1145/2872427.2883018

[16] Amit K. Chopra and Munindar P. Singh. 2018. Sociotechnical Systems and Ethics in the Large. In *Proceedings of the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES)*. ACM, New Orleans, 48–53. https://doi.org/10.1145/3278721.3278740

[17] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd edition ed.). Lawrence Erlbaum Associates, Hillsdale, New Jersey.

[18] Carroll Croarkin and Paul Tobias. 2012. *NIST/SEMATECH e-Handbook of Statistical Methods*. NIST/SEMATECH, Gaithersburg, Maryland. https://doi.org/10.18434/M32189 Available online: http://www. itl.nist.gov/div898/handbook.

[19] Davide Dell'Anna, Mehdi Dastani, and Fabiano Dalpiaz. 2019. Runtime Revision of Norms and Sanctions Based on Agent Preferences. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, Montreal, 1609–1617.

[20] Davide Dell'Anna, Mehdi Dastani, and Fabiano Dalpiaz. 2020. Runtime Revision of Sanctions in Normative Multi-Agent Systems. *Autonomous Agents and Multi-Agent Systems (JAAMAS)* 34, 2 (2020), 1–54. https://doi.org/10.1007/s10458-020-09465-8

[21] John F. Dovidio, Jane Allyn Piliavin, David A. Schroeder, and Louis A. Penner. 2017. *The Social Psychology of Prosocial Behavior*. Psychology Press, New York.

[22] Barbara Dunin-Kȩplicz and Rineke Verbrugge. 2010. *Teamwork in Multi-Agent Systems: A Formal Approach*. Wiley, Chichester, United Kingdom. https://doi.org/10.1002/9780470665237

[23] Samah El-Tantawy, Baher Abdulhai, and Hossam Abdelgawad. 2013. Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC): Methodology and Large-Scale Application on Downtown Toronto. *IEEE Transactions on Intelligent Transportation Systems* 14, 3 (2013), 1140–1150.

[24] Corinna Elsenbroich and Harko Verhagen. 2016. The Simplicity of Complex Agents: A Contextual Action Framework for Computational Agents. *Mind & Society* 15 (June 2016), 131–143. https://doi.org/10.1007/s11299-015-0183-y

[25] Ernst Fehr and Klaus M. Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* 114 (1999), 817–868.

[26] Ricard López Fogués, Pradeep K. Murukannaiah, Jose M. Such, and Munindar P. Singh. 2017. Sharing Policies in Multiuser Privacy Scenarios: Incorporating Context, Preferences, and Arguments in Decision Making. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 1 (March 2017), 5:1–5:29. https://doi.org/10.1145/3038920

[27] Christopher Frantz, Martin K. Purvis, Mariusz Nowostawski, and Bastin Tony Savarimuthu. 2013. Modelling Institutions Using Dynamic Deontics. In *International Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*. Springer, St. Paul, 211–233.

[28] Andrés García-Camino, Juan A. Rodríguez-Aguilar, Carles Sierra, and Wamberto Vasconcelos. 2009. Constraint rule-based programming of norms for electronic institutions. *Autonomous Agents and Multi-Agent Systems* 18 (2009), 186–217.

[29] Robert J. Grissom and John J. Kim. 2012. *Effect Sizes for Research: Univariate and Multivariate Applications*. Routledge, Abingdon-on-Thames.

[30] Jianye Hao and Ho-fung Leung. 2016. *Interactions in Multiagent Systems: Fairness, Social Optimality and Individual Rationality*. Springer, Berlin.

[31] Jianye Hao, Jun Sun, Guangyong Chen, Zan Wang, Chao Yu, and Zhong Ming. 2018. Efficient and Robust Emergence of Norms through Heuristic Collective Learning. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 12, 4 (2018), 23:1–23:20. https://doi.org/10.1145/3127498

[32] Risto Hilpinen (Ed.). 1981. *New Studies in Deontic Logic: Norms, Actions, and the Foundations of Ethics*. Synthese Library, Vol. 152. Reidel, Dordrecht, The Netherlands.

[33] David Hinds and Ronald M. Lee. 2011. Assessing the social network health of virtual communities. In *Virtual Communities: Concepts, Methodologies, Tools and Applications*, Mehdi Khosrow-Pour (Ed.). Vol. 2. IGI Global, Hershey, Pennsylvania, Chapter 49, 715–730. https://doi.org/10.4018/978-1-60960-100-3.ch304

[34] Christopher D. Hollander and Annie S. Wu. 2011. The current state of normative agent-based systems. *Journal of Artificial Societies and Social Simulation* 14, 2 (2011).

[35] Christopher D. Hollander and Annie S. Wu. 2011. Using the Process of Norm Emergence to Model Consensus Formation. In *Proceedings of the 5th IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*. IEEE, Ann Arbor, Michigan, 148–157.

[36] Xiaowei Huang, Ji Ruan, Qingliang Chen, and Kaile Su. 2016. Normative Multiagent Systems: A Dynamic Generalization. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*. IJCAI, New York, 1123–1129.

[37] Özgür Kafalı, Nirav Ajmeri, and Munindar P. Singh. 2016. Revani: Revising and Verifying Normative Specifications for Privacy. *IEEE Intelligent Systems (IS)* 31, 5 (Sept. 2016), 8–15. https://doi.org/10.1109/MIS.2016.89

[38] Özgür Kafalı, Nirav Ajmeri, and Munindar P. Singh. 2017. Kont: Computing Tradeoffs in Normative Multiagent Systems. In *Proceedings of the 31st Conference on Artificial Intelligence (AAAI)*. AAAI, San Francisco, 3006–3012.

[39] Özgür Kafalı, Nirav Ajmeri, and Munindar P. Singh. 2020. DESEN: Specification of Sociotechnical Systems via Patterns of Regulation and Control. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 29, 1 (Feb. 2020), 7:1–7:50. https://doi.org/10.1145/3365664

[40] Max Knobbout, Mehdi Dastani, and John-Jules Ch. Meyer. 2014. Reasoning about Dynamic Normative Systems. In *Proceedings of the 14th European Conference on Logics in Artificial Intelligence (JELIA)*. Springer, Funchal, Madeira, Portugal, 628–636.

[41] Emiliano Lorini and Roland Mühlenbernd. 2015. The long-term benefits of following fairness norms: A game-theoretic analysis. In *International Conference on Principles and Practice of Multi-Agent Systems (PRIMA)*. Springer, Bertinoro, 301–318.

[42] Mehdi Mashyekhi, Hongying Du, George F. List, and Munindar P. Singh. 2016. Silk: A Simulation Study of Regulating Open Normative Multiagent Systems. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*. IJCAI, New York, 373–379. https://www.ijcai.org/Abstract/16/060

[43] Mihail Mihaylov, Karl Tuyls, and Ann Nowé. 2014. A Decentralized Approach for Convention Emergence in Multi-Agent Systems. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)* 28, 5 (2014), 749–778. https://doi.org/10.1007/s10458-013-9240-2

[44] Jordi Campos Miralles, Maite López-Sánchez, Maria Salamó, Pedro Avila, and Juan A. Rodríguez-Aguilar. 2013. Robust Regulation Adaptation in Multi-Agent Systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 8, 3

(2013), 13:1–13:27. https://doi.org/10.1145/2517328

[45] Javier Morales, Maite López-Sánchez, Juan A. Rodríguez-Aguilar, Wamberto Vasconcelos, and Michael Wooldridge. 2015. Online Automated Synthesis of Compact Normative Systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 10, 1 (2015), 2:1–2:33. https://doi.org/10.1145/2720024

[46] Javier Morales, Maite López-Sánchez, Juan A. Rodríguez-Aguilar, Michael Wooldridge, and Wamberto Weber Vasconcelos. 2014. Minimality and Simplicity in the On-line Automated Synthesis of Normative Systems. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS*. IFAAMAS, Paris, 109–116.

[47] Javier Morales, Maite López-Sánchez, Juan A. Rodríguez-Aguilar, Michael Wooldridge, and Wamberto Weber Vasconcelos. 2015. Synthesising Liberal Normative Systems. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. IFAAMAS, Istanbul, 433–441.

[48] Javier Morales, Michael Wooldridge, Juan A. Rodríguez-Aguilar, and Maite López-Sánchez. 2018. Off-line Synthesis of Evolutionarily Stable Normative Systems. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)* 32, 5 (Sept. 2018), 635–671.

[49] Andreasa Morris-Martin, Marina De Vos, and Julian Padget. 2019. Norm Emergence in Multiagent Systems: A Viewpoint Paper. *Autonomous Agents and Multi-Agent Systems (JAAMAS)* 33, 6 (2019), 706–749.

[50] Pradeep K. Murukannaiah, Nirav Ajmeri, Catholijn M. Jonker, and Munindar P. Singh. 2020. New Foundations of Ethical Multiagent Systems. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. IFAAMAS, Auckland, 1706–1710. https://doi.org/10.5555/3398761.3398958 Blue Sky Ideas Track.

[51] Luis G. Nardin, Tina Balke-Visser, Nirav Ajmeri, Anup K. Kalia, Jaime S. Sichman, and Munindar P. Singh. 2016. Classifying Sanctions and Designing a Conceptual Sanctioning Process Model for Socio-Technical Systems. *The Knowledge Engineering Review (KER)* 31, 2 (March 2016), 142–166. https://doi.org/10.1017/S0269888916000023

[52] Michael J. North, Nicholson T. Collier, Jonathan Ozik, Eric R. Tatara, Charles M. Macal, Mark Bragen, and Pam Sydelko. 2013. Complex adaptive systems modeling with Repast Simphony. *Complex Adaptive Systems Modeling* 1 (2013), 1–26.

[53] Ana Paiva, Fernando P. Santos, and Francisco C. Santos. 2018. Engineering Pro-Sociality with Autonomous Agents. In *Proceedings of the 32th AAAI Conference on Artificial Intelligence (AAAI)*. AAAI, New Orleans, 7994–7999.

[54] John Rawls. 1999. *A Theory of Justice* (2nd edition ed.). Harvard University Press, Cambridge, Massachusetts.

[55] Fernando P. Santos, Jorge M. Pacheco, Ana Paiva, and Francisco C. Santos. 2019. Evolution of Collective Fairness in Hybrid Populations of Humans and Agents. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*. AAAI, Honolulu, 6146–6153.

[56] Bastin Tony Roy Savarimuthu and Stephen Cranefield. 2011. Norm Creation, Spreading and Emergence: A Survey of Simulation Models of Norms in Multiagent Systems. *Multiagent Grid Systems* 7, 1 (Jan. 2011), 21–54.

[57] Bastin Tony Roy Savarimuthu, Stephen Cranefield, and M. Purvis. 2009. Norm emergence in agent societies formed by dynamically changing networks. *Web Intelligence and Agent Systems: An International Journal* 7 (2009), 223–232.

[58] Bastin Tony Roy Savarimuthu, Stephen Cranefield, Maryam Purvis, and Martin K. Purvis. 2010. Obligation Norm Identification in Agent Societies. *Journal of Artificial Societies and Social Simulation (JASSS)* 13, 4, Article 3 (2010), 19 pages. https://doi.org/10.18564/jasss.1659

[59] Onkur Sen and Sandip Sen. 2010. Effects of Social Network Topology and Options on Norm Emergence. In *Coordination, Organizations, Institutions and Norms in Agent Systems V*. Springer Berlin Heidelberg, Berlin, Heidelberg, 211–222.

[60] Sandip Sen and Stéphane Airiau. 2007. Emergence of Norms Through Social Learning. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*. Morgan Kaufmann Publishers Inc., Hyderabad, 1507–1512.

[61] Sandip Sen, Zenefa Rahaman, Chad Crawford, and Osman Yücel. 2018. Agents for Social (Media) Change. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, Stockholm, 1198–1202.

[62] Marc Serramia, Maite López-Sánchez, Juan A. Rodríguez-Aguilar, Manel Rodriguez, Michael Wooldridge, Javier Morales, and Carlos Ansotegui. 2018. Moral Values in Norm Decision Making. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, Stockholm, 1294–1302.

[63] Alex Shaw and Kristina Olson. 2014. Fairness as partiality aversion: The development of procedural justice. *Journal of Experimental Child Psychology* 119 (2014), 40 – 53.

[64] Brent T. Simpson and Robb Willer. 2015. Beyond Altruism: Sociological Foundations of Cooperation and Prosocial Behavior. *Annual Review of Sociology* 41 (2015), 43–63.

[65] Munindar P. Singh. 1991. Group Ability and Structure. In *Decentralized Artificial Intelligence, Volume 2*, Yves Demazeau and Jean-Pierre Müller (Eds.). Elsevier/North-Holland, Amsterdam, 127–145. Revised proceedings of the *2nd European Workshop on Modeling Autonomous Agents in a Multi Agent World (MAAMAW)*, St. Quentin en Yvelines, France, August 1990.

[66] Munindar P. Singh. 1998. The Intentions of Teams: Team Structure, Endodeixis, and Exodeixis. In *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI)*. John Wiley, Brighton, United Kingdom, 303–307.

[67] Munindar P. Singh. 2013. Norms as a Basis for Governing Sociotechnical Systems. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 1 (Dec. 2013), 21:1–21:23. https://doi.org/10.1145/2542182.2542203

[68] Munindar P. Singh, Matthew Arrott, Tina Balke, Amit K. Chopra, Rob Christiaanse, Stephen Cranefield, Frank Dignum, Davide Eynard, Emilia Farcas, Nicoletta Fornara, Fabien Gandon, Guido Governatori, Hoa Khanh Dam, Joris Hulstijn, Ingolf Krüger, Ho-Pun Lam, Michael Meisinger, Pablo Noriega, Bastin Tony Roy Savarimuthu, Kartik Tadanki, Harko Verhagen, and Serena Villata. 2013. The Uses of Norms. In *Normative Multi-Agent Systems*, Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W. N. van der Torre (Eds.). Number 4 in Dagstuhl Follow-Ups. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Wadern, Germany, Chapter 7, 191–229. https://doi.org/10.4230/DFU.Vol4. 12111.191

[69] Satinder Singh, Tommi Jaakkola, Michael L. Littman, and Csaba Szepesvári. 2004. Convergence Results for Single-Step On-Policy Reinforcement-Learning Algorithms. *Machine Learning* 38 (2004), 287–308.

[70] Eugen Staab, Volker Fusenig, and Thomas Engel. 2008. Trust-Aided Acquisition Of Unverifiable Information. In *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI)*. IOS Press, Patras, 869–870.

[71] Toshiharu Sugawara. 2014. Emergence of Conventions in Conflict Situations in Complex Agent Network Environments. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, Paris, 1459–1460.

[72] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction* (2nd edition ed.). MIT Press, Cambridge, Massachusetts.

[73] Tom Urbanik, Alison Tanaka, Bailey Lozner, Eric Lindstrom, Kevin Lee, Shaun Quayle, Scott Beaird, Shing Tsoi, Paul Ryus, Doug Gettman, Srinivasa Sunkari, Kevin Balke, and Darcy Bullock. 2015. *Signal Timing Manual – Second Edition*. NCHRP Report 812. National Cooperative Highway Research Program, Transportation Research Board, Washington, DC. https://doi.org/10.17226/22097

[74] Harko Verhagen. 2001. Simulation of the Learning of Norms. *Social Science Computer Review* 19, 3 (2001), 296–306.

[75] Harko Verhagen, Martin Neumann, and Munindar P. Singh. 2018. Normative Multiagent Systems: Foundations and History. In *Handbook of Normative Multiagent Systems*, Amit Chopra, Leendert van der Torre, Harko Verhagen, and Serena Villata (Eds.). College Publications, London, Chapter 1, 3–25. http://www.collegepublications.co.uk/downloads/ handbooks00004.pdf

[76] Daniel Villatoro, Jordi Sabater-Mir, and Sandip Sen. 2011. Social Instruments for Robust Convention Emergence. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*. AAAI Press, Barcelona, Catalonia, Spain, 420–425.

[77] Michael Winikoff and Julija Sardelić. 2021. Artificial Intelligence and the Right to Explanation as a Human Right. *IEEE Internet Computing (IC)* 25, 2 (March 2021), 108–112. https://doi.org/10.1109/MIC.2021.3045821 Department: Internet Ethics.

[78] Jessica Woodgate and Nirav Ajmeri. 2022. Macro Ethics for Governing Equitable Sociotechnical Systems. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. IFAAMAS, Auckland, 1–5. Blue Sky Ideas Track.

[79] Chao Yu, Minjie Zhang, Fenghui Ren, and Xudong Luo. 2013. Emergence of Social Norms Through Collective Learning in Networked Agent Societies. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. IFAAMAS, Saint Paul, 475–482.