Prosocial Norm Emergence in Multiagent Systems

Mehdi Mashyekhi, Nirav Ajmeri, George List, and Munindar P. Singh
North Carolina State University, Raleigh, NC 27695, USA
{mmashay2, najmeri, gflist, mpsingh}@ncsu.edu
(Unpublished manuscript. Not for distribution.)

Abstract

We propose Cha, a framework for the emergence of prosocial norms in multiagent systems. 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. We demonstrate via simulation that Cha can improve aggregate societal gains as well as fairness of outcomes.

1 Introduction and Related Work

A social norm defines sound or "normal" interactions among members of a social group, reflecting their mutual expectations. Norms regulate interactions of autonomous agents and apply in resource sharing settings, such as social media [Sen *et al.*, 2018] and for road sharing by autonomous vehicles. We propose Cha, a framework for norm emergence. Cha applies in a *flexible* multiagent system (MAS) and in a *decentralized* manner, enabling norms to change *dynamically*. Crucially, Cha agents incorporate *prosocial* decision-making to achieve norms that avoid conflict and improve fairness.

Next, we discuss how Cha addresses the above aspects.

Prosocial behavior is when an agent performs an action that benefits others even if suboptimal to itself [Paiva *et al.*, 2019; Santos *et al.*, 2019; Serramia *et al.*, 2018]. Whereas existing approaches focus 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 improved social welfare.

Flexible and enduring are two aspects of openness. The membership of a MAS can change yet there is continuity in learning norms. An example is Wikipedia, where the users are autonomous and changing but can build on each other's work. In contrast, existing studies of norm emergence apply social learning [Airiau *et al.*, 2014; Sugawara, 2014] and assume that agents interact repeatedly in a closed system (e.g., with neighbors in a fixed graph).

Decentralized means there is no central party and norms arise through agent interactions [Mihaylov *et al.*, 2014]. Most approaches have a central party determine the norms [Airiau *et al.*, 2014; Morales *et al.*, 2015] and some use a hybrid [Mashayekhi *et al.*, 2016]. However, central and hybrid schemes are vulnerable to failure of the central portion.

Dynamic means the norms may change with the system (or environment) state [Savarimuthu and Cranefield, 2011; Huang *et al.*, 2016]. Current works address dynamism inadequately—e.g., norms once emerged are fixed [Mihaylov *et al.*, 2014; Morales *et al.*, 2018]. Dell'Anna *et al.* [2019] support norm change but via centralized sanction revision.

Significance and Novelty. This paper synthesizes two perspectives on prosociality. First, the individual perspective incorporates guilt based on inequity aversion [Fehr and Schmidt, 1999], which posits that people may be self-interested but their decisions are affected by how poorly others fare. Second, the societal perspective is based on Rawls' [1999] landmark theory of justice that focuses on designing a just society. Specifically, Cha supports Rawls' doctrine of improving the outcome for whoever is the worst off—e.g., don' t maximize throughput if it leads to some agents starving. We adopt Rawls' Maximin doctrine as a basis to measure fairness. By bringing in guilt, we are able to have norms that emerge in a bottom-up and adaptive manner without needing a central society enforcer.

Contributions. Cha is general decentralized framework for norm emergence that promotes prosociality while supporting decentralization and dynamism. Our findings support the following hypotheses.

- *Efficient resolution ($H_{efficient}$):* Cha norms resolve conflicts, i.e., while improving system-level outcomes.
- *Dynamic adaptation ($H_{dynamic}$):* Cha norms can adapt based on changes to the environment.
- *Fairness ($H_{fairness}$):* Cha supports prosocial outcomes, specifically, fairness in resource allocation, as formalized in Section 5.3.
- *Social welfare (H_{social}):* Cha yields higher societal gains than both a representative central and a hybrid approach.

Organization. Section 2 details the Cha framework. Section 3 describes how we model prosociality. Section 4 describes the 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

We explain these main components along with the associated pseudocode that demonstrates the dynamics of Cha. Algorithm 1 states the decision loop for a Cha agent in four phases: norm generation, reasoning, updating, and sharing (shown with labels in Algorithm 1). Below, we describe the norm representation in Cha, each phase of its normative life-cycle along with the Algorithm 1.

Norm Representation. A norm in Cha is *regulative*: it characterizes how the agents ought to behave in specific situations. A "norm structure" is either a norm or a precursor to a norm that is based on a deontic operator García-Camino *et al.* [2009]. We adopt a continuous notion of deontics [Frantz *et al.*, 2013], ranging from *prh* to *obl* with *may* in the middle. Initially, a norm structure is neutral—uses the operator *may*. 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*

Given a state as it perceives, an agent applies any norm structure whose antecedent is true in that state. It performs the action in the norm if its deontic is *obl*, doesn't perform it if the deontic is *prh*, and chooses either for *may*. As the agent gains experience, *obl* or *prh* may begin to dominate, indicating a norm being learned.

Norm Generation Phase. An agent *perceives* its environment through its sensors and receives a *view* (Line 2). In a traffic example, a view would be the positions and directions of cars. The function f_{conflict} encodes the knowledge to determine whether a next state arising from an action is a conflict. The agent takes a view v_t (at time t) that leads to a conflict along with the conflicting agent, *conflictAgent* (line 3).

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).

Reasoning Phase. The agent retrieves an applicable norm; selects an action according to its strategy 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 (RL). 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).

Algorithm 1 Decision loop for a Cha agent

```
1:  while True do
        # Norm Generation:
2:      views ← perceiveEnvironment(Sensors);
3:      v_t, conflictAgent ← conflictDetect(views, f_conflict);
4:      normSet ← current norms;
5:      if ∀⟨ant, con⟩ ∈ normSet: v_t ⊮ ant then
6:          n ← ⟨v_t, may(action)⟩;
7:          normSet ← normSet ∪ {n};
        # Reasoning:
8:      n ← getApplicableNorm(normSet);
9:      a ← actionSelection(n, ϵ-greedy);
10:     execute (a);
11:     a_c ← perceiveAction(conflictAgent, Sensors);
        # Updating:
12:     r ← jointActionReward(Payoffs(a,a_c));
13:     updateUtility (n, t, a);
        # Sharing:
14:     shareExperience();
15: end while
```

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).

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

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; $0 \leq \alpha \leq 1$ 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.

Sharing Phase. In Cha, each agent passes on its *experience*, i.e., the utilities associated with different states and action outcomes, to incoming members of the same *type* (i.e., those with the same goals) (Line 14). Thus, an incoming member obtains experience from members of the same type who have 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 [Staab *et al.*, 2008] 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) has to 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

[Fehr and Schmidt, 1999]. An agent incorporates another's costs in its utility to help the latter—i.e., the agent doesn't follow an applicable norm that would benefit it. We assume the agents know each other's costs [Hao and Leung, 2016].

We incorporate *guilt* [Lorini and Mühlenbernd, 2015] as a *guilt disutility* to realize concessions. Below $\delta_{i,j}$ expresses the guilt perceived by agent i with respect to j in state s.

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

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

Algorithm 2 adds prosociality to Algorithm 1, Line 9 (actionSelection). For simplicity, we consider prosocial learning after the system has converged—otherwise, the system may never stabilize, especially with high values of β_i. 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 6). If the difference in costs is below constant c, i follows norm n (Line 14). Otherwise, if agent i has not learned to concede ($U_i^P(n,t,a) > U_i^P(\neg n,t,a)$) (Line 7), i follows norm n (Line 8), adds its guilt disutility (a negative value) to its prosocial utility (Lines 9–10). Eventually, $U_i^P(n,t,a)$ would fall and agent i concedes to agent j (Line 12).

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) ← U_i(n,t,a)
4:          U_i^P(¬n,t,a) ← U_i(¬n,t,a)
5:      for each tick before conflict state do
6:          Receive agent j's total cost, f_j(s) via Sensors
7:          if f_j(s) − f_i(s) > c and
                U_i^P(n,t,a) > U_i^P(¬n,t,a) then
8:              Follow norm n
9:              δ_{i,j} = −β_i (max(f_j(s) − f_i(s) − c, 0))
10:             U_i^P(n,t,a) = U_i^P(n,t−1,a) + δ_{i,j}
11:         else if f_j(s) − f_i(s) > c and
                U_i^P(n,t,a) < U_i^P(¬n,t,a) then
12:             Concede to agent j
13:         else
14:             Follow norm n
15: else
16:     a ← actionSelection(n, ϵ-greedy)
```

4 Case Study

We evaluate Cha in a simulated intersection, where car agents continually arrive and depart. This setting [Morales *et al.*, 2015; Mashayekhi *et al.*, 2016] is simple yet powerful to illustrate Cha. As Figure 1 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 1. The grey cells, including the center, are to be ignored. Cars travel along the grid at the speed of one cell per time-tick. A car may continue 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, e.g., all, cars traveling east.

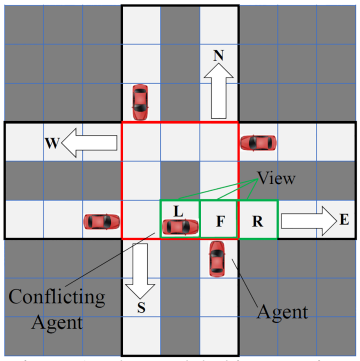

Figure 1: The modeled intersection.

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 1 with respect to the car (agent) entering from the bottom of the intersection. L, F, and R constitute the view of that car. Our grammar can specify cells with one of these six values: $\rightarrow$ (car heading east), $\leftarrow$ (car heading west), $\downarrow$ (car heading south), $\uparrow$ (car heading north), $\oslash$ (empty), and $\star$ (any of four directions). An example norm, $L(\rightarrow)$ = *prh*(*Go*), means that an agent is prohibited to proceed if it observes a car in cell L heading east ($\rightarrow$). Figure 1 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 [Airiau *et al.*, 2014; Sugawara, 2014]; the unselfishness term (i.e., u) is novel to Cha. 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*).

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

		Agent j: *Go*	Agent j: *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 (i.e., both decide to *Go*), causing a collision. Equation 3 defines the payoff of an unselfish action for agent x in state s, i.e., $f_x(s)$ in Section 3. The max ensures that the payoff of stopping is never worse than of a collision.

$$\textit{unselfishnessCost}_x(s) = \max(-d_x(s)^p, -6) \quad (3)$$

Here, d_x, the delay experienced by agent x, equals the number of ticks that x has to stop before entering the i-zone. The cost increases exponentially with delay.

5 Results: Evaluation of the Hypotheses

We now describe the simulation setup to evaluate Cha and discuss each of the hypotheses.

We model a traffic intersection (19 cells per lane: 72 cells in all with 8 cells in the i-zone) as an environment in Repast [North *et al.*, 2013]. We set the initial utility values to 0 at $t = 0$, $\alpha = 0.2$ in Equation 1 [Morales *et al.*, 2015]. 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. For the fairness experiment to evaluate $H_{fairness}$, we set $\beta = 0.5$, $c = 4$ in Equation 2, and $p = 1.1$ in Equation 3. 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*).

5.1 Evaluating the Efficient Resolution Hypothesis

$H_{efficient}$ states that emergent norms improve system-level goals—yield lower total average delays.

We consider a static setting in which the we fix the traffic flows for the north-south (N-S) and east-west (E-W) directions. We first set the traffic flow distributions are the same for N-S and E-W. We observe that approximately half of the time (508 out of 1,000 simulation runs), N-S cars learn to Go and E-W cars learn to Stop in the 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 E-W or by N-S cars, and thus minimizes collisions.

Our next setting is also static but with unequal traffic: N-S has a (30%) higher traffic volume than E-W. Figure 2 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. This trend in convergence is similar to that observed in Figures 3 and 4 of $H_{dynamic}$ (tick 0 to 44,000; before traffic pattern reverses).

Based on the asymptotic reduction in collisions and the convergence in utilities, we conclude that car agents have learned new norms to avoid collision. Table 3 shows the emergent norms: E-W cars learn to Stop in conflicting situations with N-S cars. Since N-S has higher traffic volume, the converged norms are efficient—average delay is lower when cars 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 provides evidence to support $H_{efficient}$.

5.2 Evaluating Dynamic Adaptation Hypothesis

The $H_{dynamic}$ states that Cha adapts to environmental changes—changes in traffic flow through an intersection. In this experiment, we start with a fixed traffic flow distribution where N-S has 30% higher traffic than E-W, let it converge, and then reverse the pattern (E-W has 30% more traffic). Doing so helps us determine whether norms learned in one setting persist when the traffic changes.

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 Westbound	$L(\star) \wedge F(\star) \wedge R(\leftharpoondown)$	prh(Go)
	$L(\rightharpoonup) \wedge F(\star) \wedge R(\star)$	prh(Go)
Southbound and Northbound	$L(\star) \wedge F(\star) \wedge R(\leftharpoondown)$	obl(Go)
	$L(\rightharpoonup) \wedge F(\star) \wedge R(\star)$	obl(Go)

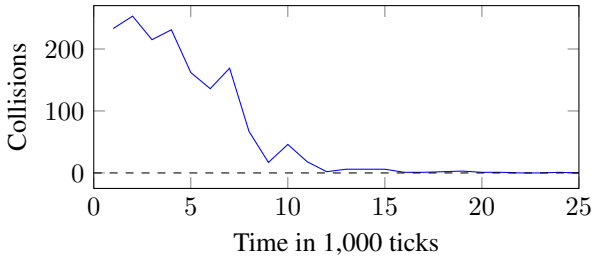

Figure 2: Total number of collisions per 1,000 ticks.

To test $H_{dynamic}$, 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 3 shows the change in average utility (for a sliding window of 1,000 ticks) for westbound cars. By $t \approx 25{,}000$ ($\text{RMSD}_{25,000} = 0$ for E-W Stop; $p < 0.01$), the norm learned by E-W cars is to Stop in case of conflict, just as in Table 3 and Figure 2. We highlight this norm in Figure 3 with the shaded box in the middle.

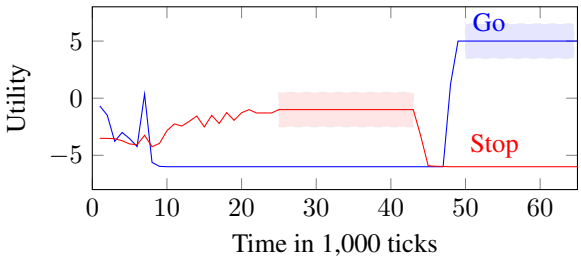

Figure 3: Utilities (averaged over a window size of 1,000 ticks) of westbound cars for dynamic setting.

After $t = 44{,}000$, the traffic pattern is reversed, and by $t \approx 50{,}000$ ($\text{RMSD}_{50,000} = 0$ for E-W Go; $p < 0.01$), the new norm is for E-W cars to Go in case of conflict. We highlight this norm in Figures 3 and 4 with boxes in the right parts. Eastbound and northbound cars have the same outcomes as westbound and southbound cars, respectively.

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

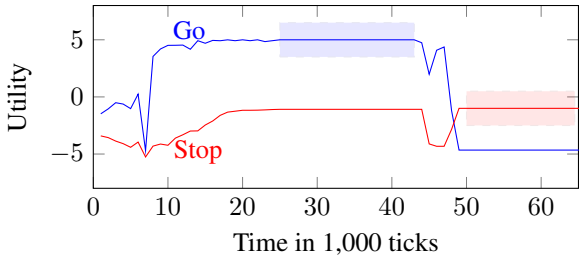

Figure 4: Utilities (averaged over a window size of 1,000 ticks) of southbound cars for dynamic setting.

Table 4: Emerged norms for dynamic setting when reversed traffic.

	Precondition	**Modality**
Eastbound and Westbound	$L(\star) \wedge F(\star) \wedge R(\leftharpoondown)$	obl(Go)
	$L(\rightharpoonup) \wedge F(\star) \wedge R(\star)$	obl(Go)
Southbound and Northbound	$L(\star) \wedge F(\star) \wedge R(\leftharpoondown)$	prh(Go)
	$L(\rightharpoonup) \wedge F(\star) \wedge R(\star)$	prh(Go)

5.3 Evaluating the Fairness Hypothesis

H_{fairness} concerns disparities in resource allocation—excessive delays for some cars to enter the intersection while others proceed quickly. We define a fair society as one that supports the Maximin criterion [Rawls, 1999, p. 153]—i.e., in which no agent is deprived of resources for long periods of time. We understand fairness (or its lack) as an outcome of different norms.

Consider the prosocial learning strategy given in Algorithm 2. As in Section 5.1 (unequal traffic setting), we set N-S flows to have 30% more traffic than E-W flows. We saw that E-W cars learned to Stop in conflicting situations with N-S cars. Now, we verify whether an N-S car can act prosocially in a conflicting situation, by yielding to an E-W car if it experiences a delay above a certain threshold.

Figure 5 shows the change in the average prosocial utility (U_i^P) of southbound cars. 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 cars (figure omitted for space) show the same trend as southbound cars.

Below, Cha refers to the original form without prosociality and ChaP to the variant with prosociality. We evaluate Cha and ChaP's performance in terms of delays. Table 5 shows the percentiles of delays over the population of cars. For example, 99.5% of all cars 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.

ChaP improves the worst-case outcome by incorporating prosociality. Figure 6 shows the delays of four or more ticks with and without prosociality.

Specifically, 225 (i.e., 35%) of the 650 cars that experienced the highest delay (six ticks) without prosociality, experience a delay of five ticks or fewer under Cha *with prosociality*. Each saw a reduction of at least $\frac{6-5}{6}$ = 16.67%. Cha thus promotes Rawls' [1999] Maximin doctrine.

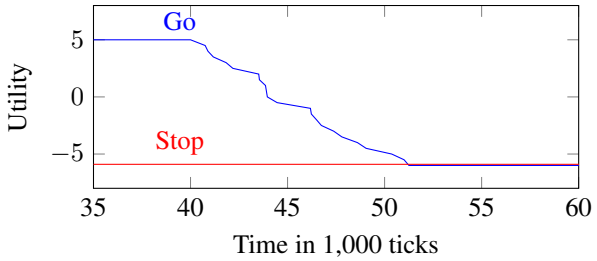

Figure 5: Prosocial utilities of southbound cars.

Table 5: Delays with Cha and ChaP.

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

γ = Skewness; κ = Kurtosis

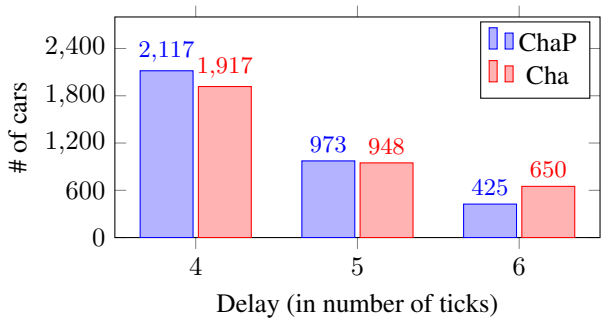

Figure 6: Benefit from prosociality. Delays ($\geq$ 4) with Cha and ChaP. 35% of cars that experience the highest delay (6 ticks) with Cha experience an improvement of $\geq$ 16.67% with ChaP.

5.4 Evaluating the Social Welfare Hypothesis

We take H_{social} as stating that ChaP wins (in aggregate societal gains) over representative central and hybrid approaches.

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

The Fully Actuated Control (FAC) strategy [FHWA, 2013] is a traditional approach to regulate traffic. In FAC, each *movement*, e.g., 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 FAC alternates between these phases. Each phase has a *minimum green* (set to one in our experiment); thereafter, the green can be extended indefinitely as long as a car is detected at the intersection.

Mashayekhi et al. [2016] proposed Silk, a hybrid framework in which agents learn norms but with hard integrity constraints called *laws* imposed on them. A law can prevent certain actions by the agent. Silk includes a central generator that recommends norms to the agents: a norm emerges if the agents accept the recommendation. Silk supports only static settings. For our comparison, we used the payoffs of Table 2 in both Silk and ChaP, but set the unselfishness cost to zero in Silk (since it supports only static settings with fixed payoffs).

Elaborating on $H_{efficient}$ of Section 5.1, we consider a specific hypothesis that ChaP yields a lower mean travel time than both Silk and FAC. These experiments are amenable to statistical hypothesis testing: the null hypothesis is that there is no significant difference in mean travel times yielded by ChaP, Silk, and FAC. We run these experiments for the case where N-S has 30% higher traffic than E-W.

Figure 7 compares the average delay in ChaP, Silk, and FAC. Table 6 summarizes the travel time results. It lists the best case travel time, mean time (μ_{travel}), and standard deviation (σ_{travel}) for ChaP, Silk, and FAC. Table 6 also lists the p-value from two-tailed t-test assuming unequal variances, effect size via Glass' Δ [Grissom and Kim, 2012] which is measured as the difference in means divided by the standard deviation of the control group), and % improvement obtained by ChaP over Silk and FAC. We choose Glass's Δ to measure effect size since the standard deviations (σ) for the two groups are different.

We find that ChaP reduces the delay and yields significantly less travel time than both Silk ($p < 0.01$; $\Delta = 2.06$) and FAC ($p < 0.01$; $\Delta = 1.76$), reducing it by 11.79% over Silk and 18.55% over FAC. Following Cohen's [1988] guidelines, an effect of over 0.80 is large; the above effect sizes are substantially above Cohen's guideline.

Table 6: Travel time: ChaP versus Silk and FAC.

	ChaP	Silk	FAC
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

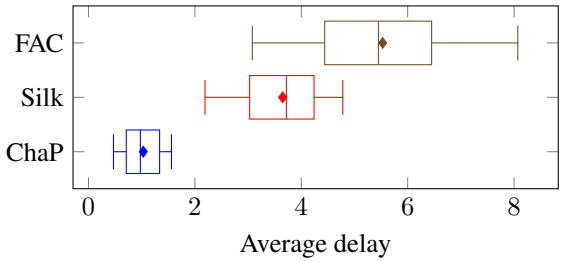

Figure 7: Comparing average delay in ChaP ($\mu = 0.98$), Silk ($\mu = 3.65$), and FAC ($\mu = 5.53$). The benefits of the prosocial approach are clear.

6 Summary and Outlook

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

One, Cha supports norm emergence to avoid conflicts in a decentralized manner. Two, norms that emerge in Cha resolve conflicts efficiently and improve system-level outcomes. Three, Cha is responsive to changes in the environment. Four, more importantly, in light of recent increasing awareness to promote and foster prosociality in MAS, Cha supports prosocial outcomes, specifically, fairness in resource allocation. 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 concerns of prosociality in the study of norm emergence. Five, Cha yields higher societal gains than both a representative central approach and a hybrid approach (baselines in our evaluation).

Note that Cha applies RL 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 doesn't require repeated interactions in a fixed graph. Departing agents can convey their experience to those 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 also worth mentioning that Cha being an *online* 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.

An important future direction is to understand context as a basis for overriding norms. For instance, can we allow cars 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 [Ajmeri *et al.*, 2018].

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 car behind) a norm to help a stranded car. One way to model more complex norm actions in a decentralized manner 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 can we evaluate and reduce the societal friction and inertia in arriving at new norms?

Another natural line of future research is to explore the generation of group norms as opposed to individual norms [Aldewereld *et al.*, 2016]. Group norms apply to groups of agents together and 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.

Cha can be applied in any MAS in which (1) its population of users have asymmetric interests that may lead to conflicts; and (2) its members can observe their neighbors and communicate. An application of our approach on practical domains would be important to validating it in the broader setting. The domain of Open Source Software Development communities [Avery *et al.*, 2016] is particularly compelling because it brings together considerations of ethics along with inertia (projects can last years) and group norms.

References

[Airiau *et al.*, 2014] Stéphane Airiau, Sandip Sen, and Daniel Villatoro. Emergence of conventions through social learning. *Autonomous Agents and Multi-Agent Systems*, 28:779–804, 2014.

[Ajmeri *et al.*, 2018] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Robust norm emergence by revealing and reasoning about context: Socially intelligent agents for enhancing privacy. In *Proc. IJCAI*, pages 28–34, 2018.

[Aldewereld *et al.*, 2016] Huib Aldewereld, Virginia Dignum, and Wamberto W. Vasconcelos. Group norms for multi-agent organisations. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 11(2):15:1–15:31, 2016.

[Avery *et al.*, 2016] Daniel Avery, Hoa Khanh Dam, Bastin Tony Roy Savarimuthu, and Aditya K. Ghose. Externalization of software behavior by the mining of norms. In *Proc. 13th International Conference on Mining Software Repositories (MSR)*, pages 223–234, 2016.

[Cohen, 1988] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum, 1988.

[Dell'Anna *et al.*, 2019] Davide Dell'Anna, Mehdi Dastani, and Fabiano Dalpiaz. Runtime revision of norms and sanctions based on agent preferences. In *Proc. AAMAS*, pages 1609–1617, 2019.

[Fehr and Schmidt, 1999] Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114:817–868, 1999.

[FHWA, 2013] FHWA. *Traffic Signal Timing Manual*. Federal Highway Administration, Washington, 2013.

[Frantz *et al.*, 2013] Christopher Frantz, Martin K. Purvis, Mariusz Nowostawski, and Bastin Tony Savarimuthu. Modelling institutions using dynamic deontics. In *Workshop on Coordination, Organizations, Institutions, and Norms in Agent Systems*, pages 211–233, 2013. Springer.

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

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

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

[Huang *et al.*, 2016] Xiaowei Huang, Ji Ruan, Quingliang Chen, and Kaile Su. Normative multiagent systems. In *Proc. IJCAI*, pages 1123–1129, 2016.

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

[Mashayekhi *et al.*, 2016] Mehdi Mashayekhi, Hongying Du, George F. List, and Munindar P. Singh. Silk: A simulation study of regulating open normative multiagent systems. In *Proc. IJCAI*, pages 373–379, 2016.

[Mihaylov *et al.*, 2014] Mihail Mihaylov, Karl Tuyls, and Ann Nowé. A decentralized approach for convention emergence in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 28:749–778, 2014.

[Morales *et al.*, 2015] Javier Morales, Maite Lopez-Sanchez, Juan A. Rodriguez-Aguilar, Wamberto Vasconcelos, and Michael Wooldridge. Online automated synthesis of compact normative systems. *ACM Transactions Autonomous and Adaptive Systems*, 10:2:1–2:33, 2015.

[Morales *et al.*, 2018] Javier Morales, Michael Wooldridge, Juan A. Rodríguez-Aguilar, and Maite Lopez-Sanchez. Off-line synthesis of evolutionarily stable normative systems. *Autonomous Agents and Multi-Agent Systems*, 32(5):635–671, 2018.

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

[Paiva *et al.*, 2019] Ana Paiva, Fernando P. Santos, and Francisco C. Santos. Engineering pro-sociality with autonomous agents. agents. In *Proc. AAAI*, pages 7994–7999, 2018.

[Rawls, 1999] John Rawls. *A Theory of Justice*. Harvard University Press, 2nd ed., 1999.

[Santos *et al.*, 2019] Fernando P. Santos, Jorge M. Pacheco, Ana Paiva, and Francisco C. Santos. Evolution of collective fairness in hybrid populations of humans and agents. In *Proc. AAAI*, pages 6146–6153, 2019.

[Savarimuthu and Cranefield, 2011] Bastin Tony Roy Savarimuthu and Stephen Cranefield. Norm creation, spreading and emergence. *Multiagent Grid Systems*, 7(1):21–54, 2011.

[Sen *et al.*, 2018] Sandip Sen, Zenefa Rahaman, Chad Crawford, and Osman Yücel. Agents for social (media) change. In *Proc. AAMAS*, pages 1198–1202, 2018.

[Serramia *et al.*, 2018] Marc Serramia, Maite Lopez-Sanchez, Juan A. Rodriguez-Aguilar, Manel Rodriguez, Michael Wooldridge, Javier Morales, and Carlos Ansotegui. Moral values in norm decision making. In *Proc. AAMAS*, pages 1294–1302, 2018.

[Staab *et al.*, 2008] Eugen Staab, Volker Fusenig, and Thomas Engel. Trust-aided acquisition of unverifiable information. In *Proc. ECAI*, pages 869–870, 2008.

[Sugawara, 2014] Toshiharu Sugawara. Emergence of conventions in conflict situations in complex agent network environments. In *Proc. AAMAS*, pages 1459–1460, 2014.