Human Perceptions of Altruism in Artificial Agents

Curry I. Guinn
Department of Computer Science
University of North Carolina Wilmington
Wilmington, NC, USA
guinnc@uncw.edu

Dan Palmer
Department of Computer Science
University of North Carolina Wilmington
Wilmington, NC, USA
dsp8301@uncw.edu

***Abstract*—Modeling realistic altruism for nonplayer characters (NPCs) is an interesting problem with substantive potential benefits to game creators and players, in the form of more believable game characters and immersive games. An experiment was conducted to investigate how humans would interpret altruistic behavior in artificial agents in a predator/prey environment. This paper describes an experiment focused on whether human observers would attribute emotional characteristics and motivations to altruistic game agents.**

Keywords—Agent-based modeling and simulation; Affective computing; Altruism; Human-computer interaction; Sentiment analysis.

I. Introduction

One motivation for this research in artificial agents is to determine human perceptions of non-player characters (NPCs) in computer games and simulations. In order to create more immersive environments, developers often attempt to imbue the artificial characters with attributes associated with affect such as facial expressions and gestures. In games where artificial agents use natural language, scripts may be developed with emotionally charged language. Game designers typically generate stories and scenarios in open-world computer games through a laborious scripting process. The process is expensive, time-consuming, error-prone, and can often result in unrealistic results that break games' immersiveness. This research presumes that perceptions of NPCs' emotions can be effectively influenced through NPC actions. This paper proposes that game agents exhibiting altruistic behavior will be interpreted by human players as being motivated by internal affective states.

One of the goals of this research project was to determine whether game players would attribute emotional motivations to artificial agents that exhibit altruistic behavior. A secondary goal was to determine how adding altruistic behaviors would change the dynamics of the game world. This paper describes experiments in simulations of predator/prey environments that demonstrate that certain types of altruistic behaviors can have positive effects on evolutionary populations. Experiments were then conducted with human subjects to see 1) whether humans would perceive certain actions as altruistic and 2) whether human observers would attribute emotional characteristics and motivations to the agents exhibiting altruistic behavior.

II. Background

A. Human Perceptions of Altruism

Altruism has been proposed to be an evolutionary trait [1]. In addition, in humans, altruistic behavior is often associated with and attributed to emotional reactions by outside observers [2]. The latter work by Sober and Wilson focuses on human perceptions of human altruism and is not in the context of artificial agents.

B. The Simulation of Predator/Prey Interactions

A common technique used to simulate predator/prey environments applies the swarming rules stipulated by Reynolds [3]. These swarming rules have also been used in game environments to integrate emergent crowd behavior into games [4, 5, 6]. However, none of these models explores human perceptions of these behaviors and whether human observers attribute emotional attributes to these agents. They do show the potential for using crowd-based rules to develop more complex and believable game agents. This paper details the implementation of swarming with additional behaviors added to the simulation.

C. Emotional modeling

Several researchers have studied the effects of emotion contagion in simulating crowd behaviors [7, 8, 9]. Tsai et al. stressed the importance of emotional contagion in producing more realistic simulations for use in what they dubbed “high-risk, high-impact” situations such as military training and virtual psychotherapy [8]. Dimas et al. tested a system whereby the emotional state of AI characters were impacted by the state and emotions of other characters and found that game players preferred this to a system without emotional modeling [7]. Xiang et al. used a heat conduction model to dynamically transmit emotional states between agents in a spectator crowd situation [9].

III. Simulation Design

The game world used in this study simulates a predator/prey environment using a two-dimensional implementation of the flocking behavior described by Reynolds [3]. Visually we represented our agents as fish with predators as “sharks” and prey as “goldfish." Three main rules govern the behavior of prey agents: 1) Agents move away from agents to prevent collisions. 2) Agents move toward the center of mass of the flock (within vision limits). 3) Agents try to match the acceleration vectors of other close agents to maintain flock cohesiveness. These three rules combine to produce flocking behavior that mimics group movement found in nature. Each prey uses these rules to calculate small adjustments to its acceleration vector on each update. For this predator/prey environment, the simulation uses some additional rules described below.

A. *Speed Limitations*

A speed limit is set for each fish's velocity vector. If a fish's velocity exceeds that limit after updating, its velocity is normalized, then scalar-multiplied by the limit. This limit preserves the fish's direction but restricts its speed. Predators move 25% faster than prey.

B. *Predator Avoidance*

Prey moves strongly to avoid predators within their limited field of view. This adjustment to their update vectors is significant enough to temporarily disrupt their flocking behavior.

C. *Energy/Hunger*

Predators have an energy/hunger level and hunt for prey to eat when its energy dropped below a threshold After eating, a predator wanders the environment with no set destination with its energy decreasing over time. A predator whose energy level reaches zero dies.

D. *Hunting/Targeting*

When a predator becomes hungry, it targets the closest prey. In order to keep it from changing its target at almost every update, it does not change its target unless a new prey is less than half the distance of its original target.

E. *Aging*

Fish behavior is tied to their age in a way that was meant to mimic the real world. Fish cannot not reproduce before or after a certain age, and fish die of old age.

F. *Reproduction*

Predators and prey reproduced asexually. We found that when using random reproductive cycles, the simulation proves to be chaotic and exhibits wild swings in populations with one or the other populations dying out after several generations. This mirrors May's description of the logistic equation that describes population growth as sensitive to chaotic behavior [10]. Interestingly, the only consistent result was that altruistic behavior (defined below) tends to make the simulation last longer. To produce longer simulations, the reproductive rates were modified so that both predators and prey give birth on regular time intervals. This regular cycle of reproduction leads to a simulation environment that is stable for a longer period of time.

G. *Altruistic Behavior*

When a predator targets any prey, the oldest available prey (within vision limits) navigates away from other prey and moves directly toward the predator. Since the predator targets the closest prey (not excepting the caveat discussed above) it may update its target selection to hunt the distracting prey moving toward it. Once the distracting prey gets close to the predator, it will change its behavior and try to flee. There is still a high probability it will be eaten. Nonetheless, this action tends to distract the predator from its original prey. Since the distracting prey is always chosen from among the oldest in the vicinity, this means that younger goldfish are more likely to survive. Only one fish will distract any given predator at a time. A screenshot from the simulation environment is given in Figure 1.

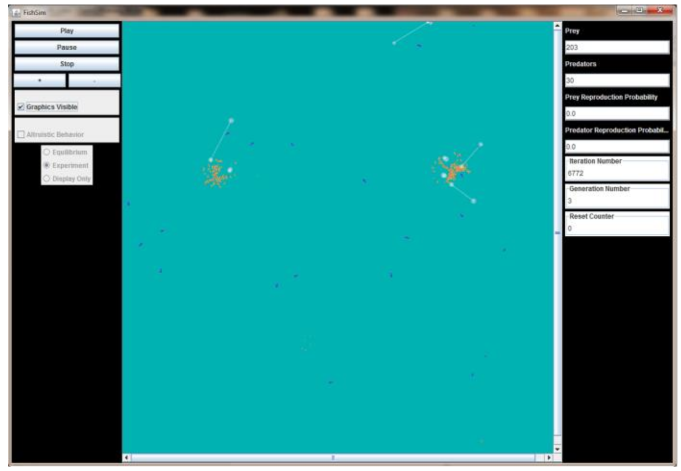

Figure 1 In this screenshot from the simulation, two large groups of prey are flocked together. Some predators are targeting prey (indicated by the lines) while other predators are not hungry and are moving randomly through the environment.

IV. The Impact of Altruistic Behavior on Predator/Prey Populations

Simulations were run without the sacrificial behaviors (Control Group) and with the sacrificial behaviors (Experimental Group). Each of these tests was repeated 100 times for each group to ensure sufficient statistical power to show a difference between the two groups.

Each iteration of the experiment for the Control Group was paired with an iteration for the Experimental Group. Initial values for position, orientation, hunger, and age were generated from random seeds created prior to runtime and read from a file. Therefore, while pseudo-randomized, these values were the same for each paired Experimental-Control iteration of the experiment. This seeding was done to ensure that differences between the control and experimental groups could be attributed to the altruistic behaviors.

A. *Outcome 1: Altruistic Behavior Detrimental to Altruistic Agent*

Experiments demonstrably show that engaging in the distracting behavior described above is, in fact, altruistic, in that it is detrimental to the prey engaging in the behavior to do so. Engaging in distracting behavior leads to the imminent death of the distractor 94.3% of the time. The only times a distracting prey is not eaten occurs when the predator is too close to its original targeted prey, not giving the distracting prey enough time to get close to it.

B. *Outcome 2: Altruistic Behavior Beneficial to Population*

However, this self-detrimental, altruistic behavior has an overall positive impact on both the prey and predator populations. The statistical results of this experiment are summarized in Table 1. Note that the p-value for all measurements is < 0.0001 for all statistics. The population whose members engaged in sacrificial behavior allows prey to live longer, reproduce more, and sustain a higher population level. Interesting, the sacrificial behavior also benefitted the predators as hunting was generally more successful and accomplished more quickly.

Table 1 Comparison of Population Statistics with and without Altruistic Behavior (n = 100)

Statistic (N = 100)	Control Group (No Altruism) Mean (s.d)	Experimental Group w/Altruism Mean (s.d)
Prey Population	502 (36.9)	579 (70.2)
Pred. Population	58 (3.6)	70 (4.3)
Prey # of Children	0.60 (0.03)	0.65 (0.04)
Pred # of Children	0.65 (0.02)	0.71 (0.02)
Prey Age	5911 (206.6)	6590 (310.1)
Pred. Age	8975 (936.9)	10691 (653.3)

V. HUMAN SUBJECT EXPERIMENTS

Human subject experiments compared the perceptions of humans viewing videos of the simulation and looked for differences in their perceptions of the agents' emotions and motivations. A research question was whether they would view the altruistic behavior as one driven by positive emotion to protect the younger fish in the flock. Two approximately one-minute videos, without the sacrificial altruistic behavior (Control Group) and with the sacrificial altruistic behavior (Experimental Group), were recorded from footage of the simulation. Each subject viewed the same video (either Control or Experimental, but not both) twice. For ease of subject reference, predators are called "Sharks", and prey are called "Goldfish." Participants were then asked to complete a pair of survey instruments designed to measure their perceptions after viewing one of the videos, selected at random.

A. *Survey Instruments*

Subjects completed both a free response questionnaire and a Likert-style questionnaire. The free response survey was administered first so as not to prime the subjects with vocabulary from the Likert-style questionnaire. In the free response survey, subjects were asked to describe and attempt to explain the behavior of the fish.

B. *Pilot Study*

A pilot study was conducted with 126 students in an introductory computer science course. The primary result from the pilot study was that it was difficult for subjects to understand what was happening in the simulation. Therefore, following the pilot study, a number of changes to the simulation were made to make it easier for a subject to observe behavior. A screenshot of the video used in the experiments is presented in Figure 2.

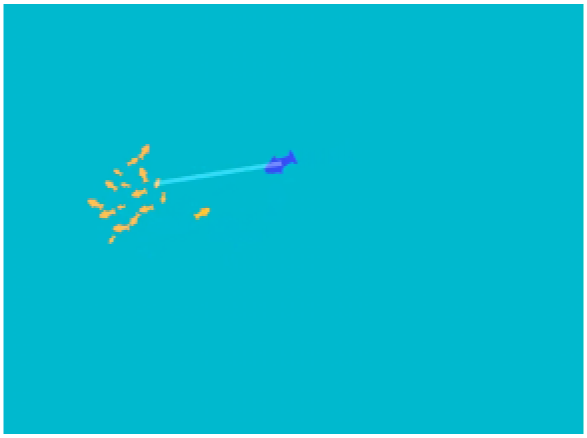

Figure 2: A screenshot of the video used in the human subjects experiment. In this image, a predator (the blue fish) has targeted a prey (a yellow fish) which is one of a school of fish. A distracting (altruistic) fish is approaching the predator but is not yet close enough for the predator to re-target.

- Having multiple predators in the simulation appeared to have been so confusing that subjects could not focus on individual behaviors of the predators. In the final study, only one predator appears in the environment at a time.
- The size difference between the predator and prey was exaggerated to make it easier for subjects to identify the predators.
- Rather than use multiple (but unconnected) scenes, the final study uses one long simulation video with four instances of predator/prey interaction.
- When a predator (shark) selects a target, a translucent line is drawn between the predator and its chosen prey. This

feature makes it easier for subjects to determine what the shark is doing.

- The final study was conducted with a smaller group, 27 subjects in the Control Group and 25 subjects in the Experimental Group.

VI. Human Perceptions of Altruistic Behavior

The Likert-style survey exhibited significant differences between the control group and the experimental group as summarized in Table 2. The scale of each Likert-style question is 1 for Strongly Disagree, 2 for Disagree, 3 for No Opinion, 4 for Agree, and 5 for Strongly Agree.

A. *Observation of Altruistic Behavior*

The most statistically significant difference between the Control and Experimental Groups was **Statement 3: "The goldfish tried to protect each other, even if it put them in danger."** With a p-value of 0.001, the Control Group mean value was 1.37 while the Experimental Group had a mean value of 2.52. First, it is worth noting that not all participants in the Experimental Group observed the altruistic behavior. Within the Experimental Group, 56% of the subjects agreed or strongly agreed with Statement 3 (opposed to the 29% of Control subjects who agreed or strongly agreed). The video was 60 seconds long, and subjects were able to view the video twice. However, with multiple agents and many predator/prey interactions, it is not surprising that not all subjects recognized behavior as being altruistic.

Subjects responded in a similar fashion to Statement 1: "The goldfish seemed to care about the other goldfish." The mean score for the Control Group was 1.70; for Experimental Group, 2.68, with a p-value of 0.003.

B. *Natural Language Descriptions of Altruistic Behavior*

Subjects were asked to describe in their own words the behaviors they observed in the video.

1) *Sample Response from Control Group*

"When the shark became hungry it would chase the closest fish (regardless of size) and would only eat that fish. The fish would stay in a group and would not flee, even when the shark was in the middle of the group. The fish being hunted would not notice until the shark was right on it and only then seemed to try to flee. The fish were constantly moving within the group but the group location would not change until the shark came through and pushed the group somewhere."

2) *Sample Response from Experimental Group*

"The shark locates a target at random. As the shark heads towards the target a fish will get in front of target and the shark will switch targets and go for the fish nearest. The fish all swim together to perhaps appear bigger in size then [sic] they really are. The fish may be sending one fish out to distract shark as the others attempt an escape. The shark seems to be randomly selecting the fish and attacking in waves."

3) *Qualitative Observation*

No subjects in the Control Group (n = 27) wrote any observations that suggested altruistic, sacrificial, or distracting behavior among the prey. Four subjects in the Experimental Group (n = 25) explicitly describe observing altruistic or sacrificial behavior among prey. However, both subject pools did use language laden with emotional words to described agent behavior.

Table 2: Results from Likert-style survey. Subjects scored each statement as follows: 1 (Strongly Disagree), 2 (Disagree), 3 (No Opinion), 4 (Agree), 5 (Strong Agree).

Statement on Likert-style Survey	Control	Experimental	p-value
1 The goldfish seemed to care about the other goldfish.	1.7	2.68	0.003
2 The goldfish tried to protect each other.	1.85	2.48	0.04
3 The goldfish tried to protect each other, even if it put them in danger.	1.37	2.52	0.001
4 The goldfish seemed to dislike the sharks.	2.37	2.64	0.185
5 The sharks seemed to care about the other sharks.	1.0	1.2	0.217
6 The sharks tried to protect each other.	0.93	1.0	0.388
7 The sharks tried to protect each other, even if it put them in danger.	0.93	1.0	0.383
8 The sharks seemed to dislike the goldfish.	0.74	1.08	0.103
9 The sharks are good hunters.	2.93	2.96	0.447
10 The goldfish are good at evading the sharks.	1.22	1.36	0.302
11 The goldfish are good at distracting the sharks.	1.52	2.24	0.012
12 The goldfish formed recognizable social groups.	2.89	3.25	0.082
13 The goldfish cared about the members of their social group.	2.07	2.44	0.125

4) Linguistic Analysis of Responses

Using the Linguistic Inquiry and Word Count (LIWC) tool, these subject responses were analyzed for words related to sentiment and emotion [11]. While statistics were gathered in all LIWC categories, we focus our analysis on the following categories as most relevant to our thesis: Family, Humans, Affect, Positive Emotion, Negative Emotion, Anxiety, Anger, Sadness, and Religion. We further broke down the Experimental Group into how those subjects scored Statement 3: "The goldfish tried to protect each other, even if it put them in danger." As can be seen from Table 3, there were some significant differences between the Control Group and the Experimental Group as well as differences based on whether the subject observed altruistic behavior based on their answers on the Likert-style survey to Statement 3.

a) Family, Anthropomorphism, Religion

An interesting finding from the LIWC analysis is that there are no words in the Control Group's free-text responses that are "Family" words. However, "Family" words appear in increasing frequency in the Experimental Group, particularly in those subjects who indicated they observed altruistic behavior based on their answers on the Likert-style survey. Similarly, while some subjects in the Control Group used some "Human" words to describe the agents, the use of "Human" words increased significantly in the Experimental Group. Words associated with "Religion" do not appear at all in the Control Group's free-text responses. "Religion" words do appear in some free-text responses in the Experimental Group.

b) Affect and Emotion Words

Interestingly, the number of affect and emotion words is not significantly different between the Control and Experimental Groups' free-text responses. There is a noticeable tendency for there to be less Negative Emotion words in the Experimental Group than in the Control Group.

Table 2: Linguistic Inquiry and World Count (LIWC) analysis of free-text responses by the Control Group and the Experimental Group (broken into three categories depending on how the subject answered Statement 3)

	Control	Exp. (Neutral, Disagree, Strongly Disagree w/ Statement 3	Exp. (Agree or Strongly Agree w/ Statement 3)	Exp. (Strongly Agree w/ Statement 3
Family	0.00	0.12	0.21	0.32
Humans	0.22	0.30	0.31	0.48
Affect	2.81	2.36	1.88	2.08
Pos Emo	1.21	1.27	1.05	0.96
Neg Emo	1.60	1.09	0.84	1.12
Anxiety	0.44	0.24	0.21	0.32
Anger	0.55	0.60	0.52	0.64
Sad	0.39	0.12	0.10	0.16
Religion	0.00	0.12	0.21	0.16

VII. Conclusions and Future Work

One goal of this project was to create an experimental platform to explore altruistic behavior in a two-dimensional game environment. Predator/prey simulations provide an environment that is rich enough for artificial agents to engage in complex behaviors, yet not be overwhelming in terms of programming complexity or overload human subjects with too much information. In our environment, we were able to develop an "altruism algorithm" that satisfied two significant constraints: first, the behavior must be truly altruistic in that it is not generally beneficial to the agent performing the action; secondly, the behavior should have a benefit for other agents. The distracting behavior of older prey almost always resulted in the death of the distractor, but the overall effect was to extend the life of younger prey resulting in more reproduction and larger population sizes. Somewhat unexpectedly, even predator populations benefitted from the altruistic behavior of prey.

Another goal of the project was to determine how human subjects would perceive the altruistic behavior. In the limited timespan that subjects viewed the videos, more than half of the experimental group did observe the altruistic behavior as indicated by scores on the Likert survey. In their natural language responses, the experimental group tended to use more words associated with Family, Humans, and Religion in describing agent behavior of the altruistic populations.

An avenue for future work includes making the simulation interactive for the human subject. In particular, giving the human subject control over a predator as it hunts for prey would provide a higher degree of immersion for the subject (as opposed to passively watching a video). We speculate that a higher degree of immersion would cause the human subject to observe and reflect on the decision-making of the prey in greater detail.

Several subjects attributed the decision-making of the prey to "group decisions" rather than the decisions of individual agents. Given the flocking behavior of prey, this attribution of group decision making is understandable. Under what conditions do human subjects attribute decisions to group processes versus individual processes? Experiments will be conducted that vary the level of swarming cohesiveness to see how much flocking behavior affects human judgment of decision-making.

The visual representation of agents in this work was very primitive with only the outline of fish. These creatures had no eyes, no mouths, no ears, and had no moving parts. The lack of such details may have had an effect on human subjects' willingness to attribute sophisticated decision-making to the agents. A prospective study would involve giving the agents more animal-like features to see whether those features affect human perceptions of these agents. As the agents become easier to anthropomorphize because of visual characteristics, are humans more likely to attribute complex inner lives to the agents?

REFERENCES

[1] G. W. Hamilton and D. William. "The evolution of altruistic behavior." *The American Naturalist*, 1963, pp. 354-356.

[2] E. Sober and D. Wilson, *Unto others: The evolution and psychology of unselfish behavior*. No. 218. Harvard University Press, 1998.

[3] C. Reynolds. "Flocks, herds and schools: A distributed behavioral model." In ACM SIGGRAPH Computer Graphics, vol. 21, no. 4, 1987, pp. 25-34.

[4] C. Chiang, C. Hoffmann, and S. Mittal. "Emergent Crowd Behavior." Computer-Aided Design and Applications 6, no. 6 (2009): 865-875.

[5] C. Reynolds. "Big fast crowds on ps3." In *Proceedings of the 2006 ACM SIGGRAPH symposium on Videogames*, pp. 113-121. ACM, 2006.

[6] S. von Mammen, and C. Jacob. "Swarming for Games: Immersion in Complex Systems." *Applications of Evolutionary Computing*, 2009, pp. 293-302.

[7] J. Dimas, G. Pereira, P. Santos, R. Prada, and A. Paiva. "I'm happy if you are happy.: a model for emotional contagion in game characters." In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, 2011, p. 2..

[8] J. Tsai, E. Bowring, S. Marsella, and M. Tambe. "Emotional contagion with virtual characters." In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*-Volume 3, International Foundation for Autonomous Agents and Multiagent Systems, 2012, pp. 1193-1194.

[9] N. Xiang, P. Zhigeng, and M. Zhang. "Using emotion transmission to simulate spectator behaviors." In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, 2012, pp. 51-54.

[10] R. May. "Simple mathematical models with very complicated dynamics." *Nature*, 1976, pp. 459-467.

[11] J. Pennebaker, M. Francis, and R. Booth. *Linguistic inquiry and word count: LIWC 2001*. Mahway: Lawrence Erlbaum Associates 71 (2001): 2001.