THE INTERPRETATION OF INTENTIONALITY FROM DYNAMIC SCENES

by

PETER C. PANTELIS

A thesis submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

Jacob Feldman, Ph.D.

and approved by

New Brunswick, New Jersey

January, 2010

ABSTRACT OF THE THESIS

The interpretation of intentionality from dynamic scenes

By PETER C. PANTELIS

Thesis Director:

Jacob Feldman, Ph.D.

This thesis explores how the mind uses the motion of animate objects to make inferences about these objects' underlying mental states, intentions, goals, or dispositions. We present dynamic scenes to subjects in which autonomously programmed triangular "agents" interact with each other and—in two of the experiments—an additional agent that is controlled by the subject. We strive for the autonomous agents to be simple in their underlying programming but to also engage in a rich array of lifelike behaviors. Subjects watch short simulations populated with these agents and then are asked questions designed to probe their perceptions of the similarities among the agents' behaviors. We use the responses to derive a multidimensional scaling (MDS) solution for the agents in our set. The aim is to relate this MDS solution to the underlying programming on one hand, and to also discover interesting structure in subjects' perception of the "agent space." Clusters in this space, for example, could provide insight into the perceptual biases subjects bring to the experiment for which types of agents are a priori more likely to be observed. The most robust result from the experiments is that subjects, in making inferences about an agent's intentions, pay special attention to how the agent reacts when another agent is at a "critical distance" of about 10-17.5 agent lengths away.

Table of Contents

List of Figures

1. Introduction

Intelligent beings can and must distinguish between animate and inanimate objects when cognitively processing the world before them. Researchers have made the claim that even infants make this distinction, and, furthermore, may harbor something akin to a naïve concept or theory of other beings mental states and intentions (e.g. Gergely, Nádasdy, Csibra, & Bíró, 1995; Johnson, 2000, for a review). For us grown-ups, our ability to conceive of others as animate, mentalistic agents with independent perceptions and motivations is evident. Indeed, adults will readily infer intentions, thoughts, and goal-directed behavior to the actions even of simple geometric figures, lines, or points moving in a display, as illustrated by the seminal work of Heider & Simmel (1944). Yet much remains to be learned about what visual information is most relevant for the mind's construal of animacy.

This research explores how people make use of one aspect of a scene—the motion of objects—to make inferences about other agents' mental architecture. For this task, motion is only one cue a person can use among many (Gelman, Durgin, & Kaufman, 1995), but one that we believe to be both evolutionarily important and quite ancient. Closely related topics have been examined extensively in fields like developmental psychology, but have rarely been studied (a) using psychophysics with adult subjects, to better understand the basic features that underlie the percept of intention, or (b) computationally, to model the inferences subjects actually make.

A handful of recent studies have ventured in this direction, demonstrating that varying the motion of simple geometric figures along certain programmable parameters (e.g. speed, trajectory) can influence peoples subjective percept of how animate they are and what their intentions might be. Dittrich & Lea (1994), for instance, highlighted the importance of an observed figures interaction with other figures, as context. Tremoulet & Feldman (2000) demonstrated that subjects can construe the intentionality of an agent from its pattern of motion alone, even in the absence of such context with which it might interact. Tremoulet & Feldman (2006) attempted to formally model related mental processes, and their experiments highlighted the role of a figures perceived passivity or reactivity to its context in the observers attribution of intention and mental architecture to that figure-agent.

Baker, Tenenbaum, & Saxe (2006, 2007) take a Bayesian approach to the question of how the human mind infers the goals of an agent from its actions, a general framework we also adopt. These authors conceive of this inference as being a sort of "inverse planning"; that is, the observer tries to find the most likely underlying goals or intentions of an agent given the observed behavior. In our study, we hope to address the *a priori* terms—$p(A)$—in Bayesian formulae of this flavor:

$$p(A|motion) \propto p(motion|A)p(A) \tag{1.1}$$

where A represents some underlying mental state, behavioral disposition, payoff matrix, or goal set—i.e. a theory of the other agent's mind that includes some reference to its intentions.

It could be that the human mind assumes any possible behavioral disposition or set of intentions to be equally likely within a relatively uniform "agent space". In this case, the subject could be said to have a neutral prior across agent types. However, it would be interesting if some underlying agent mentalities were assigned higher prior probabilities than others. It would be especially interesting if these higher priors corresponded to certain "natural kinds" that might be partially innate—the result of evolution. These natural kinds might include chasing, courting, following, guarding, fighting, or playing (Barrett, Todd, Miller, & Blythe, 2005).

We aim to shed light on the natural mental divisions that arise between categories of animate, intentional, or goal-directed motion. We ask: Given patterns of motion that are subjectively construed as reflecting underlying animacy and mentalistic processes, how do these patterns of motion naturally cluster in the mind of the observer? And, how does the human mind perform this inference? We hypothesize that when people view objects moving about a scene, they will tend to perceive certain patterns of behavior as naturally "going together." These clusters would represent modes or peaks in a subject's priors that pertain to his "agent space."

This approach to the perception of animacy resembles categorization in spirit. We indeed intend to explore how the mind partitions the world of animates into categories relevant to its behavior. Recent studies have explicitly examined categories of animate motion, presenting the above-mentioned "natural kinds" as the relevant categories a person employs when observing a dyadic interaction between agents (Blythe, Todd, & Miller, 1999; Barrett, Todd, Miller, &

Blythe, 2005; McAleer & Pollick, 2008). That is, these studies presented subjects with scenes constructed to represent these different categories of interaction, and demonstrated that subjects were reliably able to categorize these scenes even in degraded forms for which motion was the only salient cue. Additionally, these studies each reported that subjects seemed to demonstrate a bias toward categorizing an interaction as "play"—this is the type of bias in which the present study is explicitly interested.

In contrast to these studies, our aim is to show subjects a broad array of agent interactions, from a richer and more general collection of possibilities, in an attempt to allow subjects' minds to impose *their own* structure on the agent space. Another novel aspect of our methodology is that the scenes we show subjects have not been constructed beforehand to convey some category of interaction, as has usually been the case in the previous experimental literature described above. Rather, we program the agents inhabiting these scenes to behave autonomously, resulting in chaotic multi-agent interactions we cannot predict in advance.

In Experiments 1 and 2, we use multidimensional scaling (MDS) in an attempt to extract the "natural cleavages" present within this stimulus space of animate motion and behavior. In Experiment 3, we take a different, simulation-intensive approach. In all experiments, we hope to find and explain clusters in the mental stimulus space that the mind tailors for animate behaviors.

1.1 General Methodology

Each of the experiments involved subjects watching small isosceles triangles interact with each other within a simulation environment. All of these experiments were programmed using the breve Simulation Environment (Klein, 2002), an open-source software package freely available at `http://www.spiderland.org`. A snapshot from Experiment 1 is shown in Figure 1.1.

1.1.1 Programming Lifelike Automata

In designing and coding the agent behaviors, we aimed to employ a simple programming scheme that would impose minimal structure on the agents' interactions but, nonetheless, would result in a rich variety of lifelike agent behaviors. This proved to be a difficult task, and we have

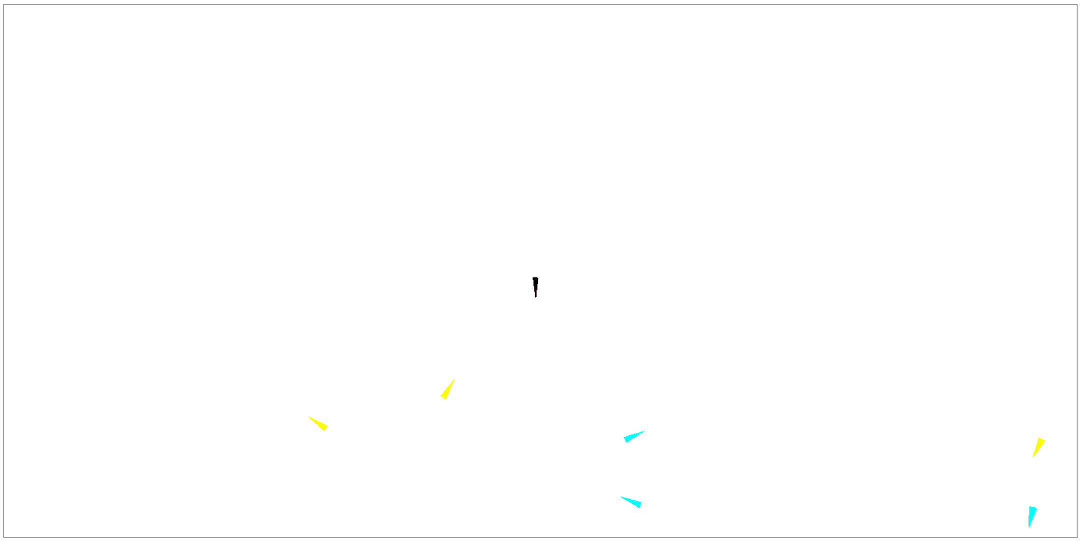

Figure 1.1: A screenshot from Experiment 1 (colors inverted).

not fully resolved this problem at this stage of the research.

It is important to note that the programming scheme we employ here is not the only possible scheme that we considered or that could be employed in other versions of these experiments. Later, I will discuss future innovations to the design of a richer and more lifelike space of agent behaviors. However, the problem of designing these automata is somewhat orthogonal to the central questions of this research. The automata are tools for addressing these questions—a pool of stimuli for the psychophysics.

We programmed the triangular agents to behave according to a number of simple rules. Automata governed by simple rules can produce vividly lifelike behaviors that may result in observers making complex inferences about their mental states and motivations (e.g., Braitenberg, 1984). In theory, these are the types of Braitenberg-inspired automata we aim to present to subjects.

We programmed each agent to orient one vertex of its triangular body (that which lay on its axis of symmetry) in the direction of its movement. We hoped this would signal to the subject that this vertex was the "head" of the agent. The agents reacted the same when they either collided with the edge of the scene or another agent—they bounced off for a split second. This sometimes resulted in jerky and unnatural-looking behaviors at collisions, so in Experiments 2 and 3 we attempted to change collision behavior (more on this later). Each agent started off in the simulation environment with a random velocity and location (within constraints).

At each iteration of a simulation, an agent would find the closest other agent within the scene and then accelerate toward or away from it according to a set of parameters. Each agent had six such parameters, each of which would control the acceleration of the agent depending on how far away the closest other agent was (0-5 units, 5-10, 10-20, 20-40, 40-70, or > 70). A schematic of these 6 radii around an agent is shown in Figure 2.1. As an example, one agent might approach an agent from afar but then veer away as it gets to a closer radius. Or, it might consistently accelerate away from another agent. Depending on how the other agent is programmed, their interaction might look like chasing/fleeing, or one pushing the other, or even one agent circling another.

We constructed a pool of 12 agents, each with 6 randomized parameters within the programming scheme.

2. Experiment 1

2.1 Method

2.1.1 Subjects

Eight undergraduate and graduate students between the ages of 18 and 24 participated in the study, which required one hour-long session. The undergraduates received course credit.

2.1.2 Stimuli

We presented scenes, 15 seconds each in duration, to subjects on a 1440 x 900 LED display, on a 15 inch MacBook Pro laptop with a 2.2 GHz dual core processor. These scenes were not pre-made animations; rather, the "agents" behaved autonomously according to a number of simple rules (described above). At the beginning of a scene, each of seven agents were placed at a random location in the display, and then the simulation was left to run.

The simulation display was 13 x 6.5 inches, and the viewing distance was approximately 18 inches. The programming library employed units that were equivalent to 22 units/inch. Velocities of agents in the simulations were programmed in units/simulation time (40 units of simulation time about equaled 1 s). The triangular agents had bases of 1 unit length and heights of 4 unit length.

2.1.3 Procedure

In each 15 second scene, the subject observed 7 agents interacting. 3 were red, 3 were blue, and 1 was white. The reds would behave according to the same parameters as the other reds, the blues according to a different set of parameters, and the lone white according to a third set of parameters. The agents were drawn from a larger 12 agent pool; thus, there were 220 possible triads of these 12 agents.[1] For each scene, one of these 220 triads was selected at random, and

[1] Strictly speaking, because the status of the white agent in each trial is special, and, as a result, during a given trial the subject cannot respond that he actually believes the blue and red agents to be most alike, 660 possible arrangements actually exist. Rather than show all 660 possibilities, we randomized the procedure so that no agent type would be more or less likely to be "white" during a trial. Nevertheless, this presents a source of noise in the data, and we altered the procedure in Experiment 2 to address this issue.

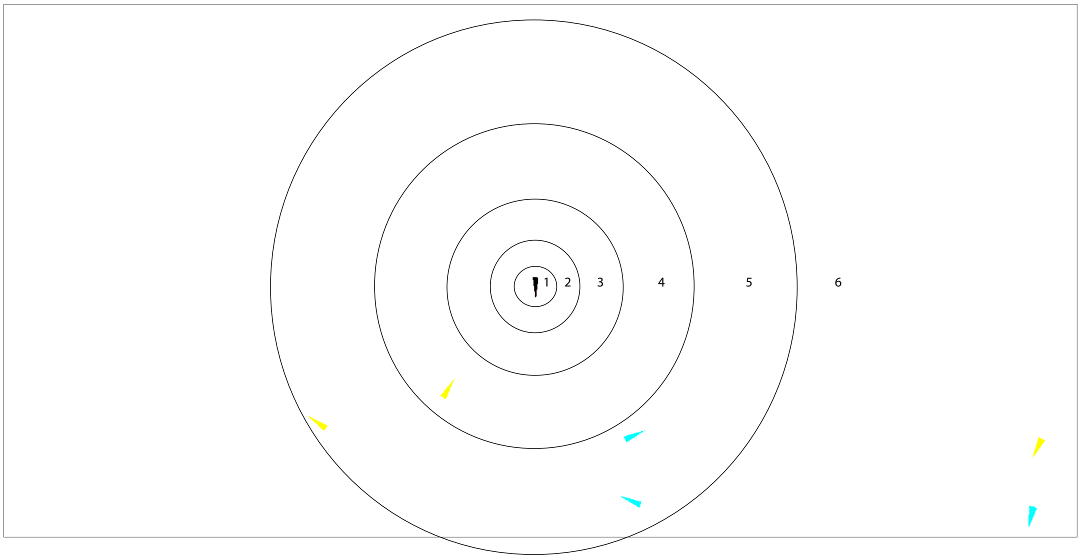

Figure 2.1: A schematic of the programming scheme for the automata. The black automaton accelerates toward or away from the closest other agent in the scene. The direction and magnitude of this acceleration depends on the distance to this closest other agent, with possible distances divided into six zones. Zone #5 seems to be the most psychologically relevant.

then each of the three programs in the selected triad was randomly assigned to either red, blue, or white. Each subject saw 220 such scenes, exhausting the possible triads.

Subjects were openly encouraged to construe the triangular agents as animate. At the end of a 15 second scene, they were asked “Is the white agent behaving more like a red, or more like a blue?” They answered by pressing a button in a dialog box.

We constructed a 12 x 12 symmetric distance matrix for each subject, to be fed into the individual differences multi-dimensional scaling (MDS) algorithm (INDSCAL/ALSCAL; Takane, Young, & de Leeuw, 1977). Within this matrix, an agent was assigned a distance of 0 from itself. As two different agents appeared in the same trial of an experimental session 10 times, the distance in this matrix between any two agents was initially set at 11.

If the subject chose “red,” then the agent whose programming was used for the red agents in this trial was made to be closer together (more similar) in this distance matrix with that of the white agent, and likewise for if the subject chose “blue.” That is, the distance between these two agents in the matrix was reduced by 1. Previous studies have used similar methodologies to gauge subject similarity ratings of visual stimuli (e.g., Kahana & Bennett, 1994; Pantelis, van Vugt, Sekuler, Wilson, & Kahana, 2008).

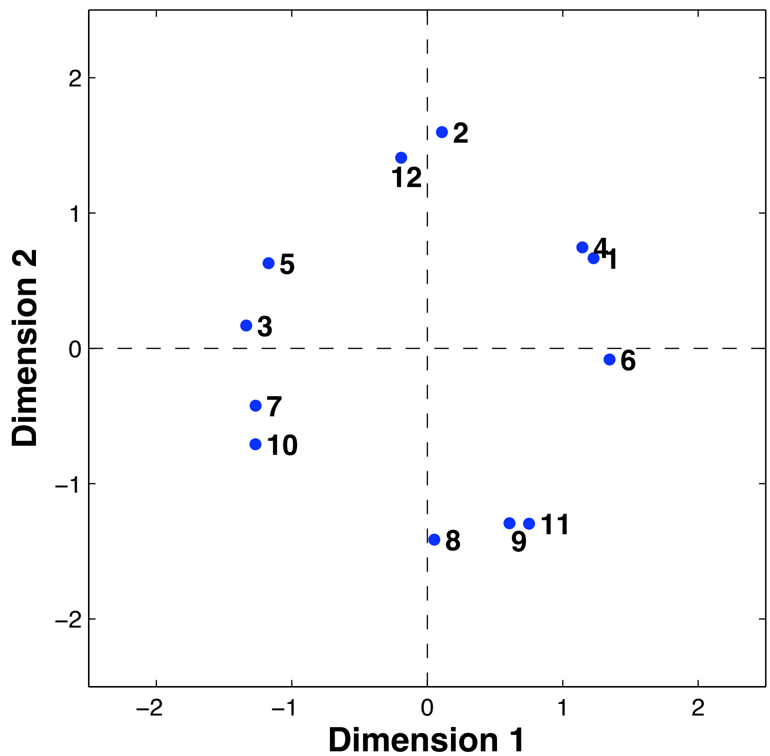

Figure 2.2: The 2-dimensional MDS solution for the 12 agents, fitting data from Experiment 1.

2.2 Results and Discussion

We derived a 2-dimensional MDS solution in order to visualize the space of agents that subjects (on average) perceived (see Figure 2.2). For this amount of points, the INDSCAL algorithm allows for fits of 2-5 dimensions. Deriving higher-dimension solutions will always result in better fits to the experimental data (see Figure 3.4). However, the higher number of dimensions would be even more difficult to interpret than the two condensed dimensions we present, and even a 5-dimensional fit would probably be a condensed version of the true amount of psychologically relevant dimensions in this agent space. The true number of dimensions could be even higher than the total number of agents in our sample.

A 2-dimensional solution allows for the easiest visualization of the inter-agent distances, an important motivation for using the MDS analysis in the first place. If interesting structure emerged only in higher-dimensional fits for these data, this might provide greater justification for using these more complex solutions. However, we actually find the clearest and most interesting structure within a 2-dimensional fit.

The most striking aspect of the space is its ring-like structure, which is a result similar to what one would get if one performed a 2-dimensional MDS on colors in the color wheel, for instance (demonstrated in Shepard, 1980). The meaning of this ring structure is not perfectly clear. On one hand, it could be a misleadingly provocative artifact of the statistical analysis we employed, or the 2-dimensional fit. On the other hand, it could be indicative of an interesting but unknown underlying dimension that—instead of having two extremes—cycles around without having any true beginning or end. On a third hand, the ring structure might suggest an interesting sort of correlation between dimensions 1 and 2. That is, for any given agent, if one is given the value of one of its dimensions, one can almost perfectly predict the *absolute* value of the other. Possible meaning behind this ring structure will be revisited in a discussion following Experiment 2.

Table 2.1: Correlations ($r[10]$) between programmed parameters (rows) and MDS dimensions (columns). Bold font represents $p < .01$

	MDS Dimension 1	MDS Dimension 2
Parameter 1	-.070	.384
Parameter 2	-.275	-.074
Parameter 3	.527	.199
Parameter 4	.411	-.375
Parameter 5	**-.801**	.093
Parameter 6	.459	.197

Subjects' perception of the agents' behaviors arises from some complex interaction of its underlying programming and the chaotic interaction with other agents that arises during each unique simulation. This contributed to there being many individual differences between subjects' results; few subjects' distance matrices showed obvious correlation. However, one of the 6 parameters with which we programmed each agent was indeed strongly correlated with one of the MDS coordinates (see Table 2.1). This parameter controlled how an agent would behave when the closest other agent in the simulation was between 40 and 70 units (about 1.8 to 3.2 inches) away from it. I discuss this finding in greater detail in the Experiment 2 discussion.

3. Experiment 2

In Experiment 2 we made some changes to the methodology in Experiment 1 that we hoped would reduce noise in the data. The most important change was to allow the subject to control one of the agents in each simulation. We hoped that the subject would use this tool to glean more information about the various agents in the simulation that would have been possible in Experiment 1. We also were able to record the movements subjects made over the course of each simulation.

3.1 Method

3.1.1 Subjects

Seven undergraduate and graduate students between the ages of 18 and 23 participated in the study, which required one hour-long session. The undergraduates received course credit.

3.1.2 Stimuli

We presented scenes similar to those in Experiment 1, 15 seconds each in duration, to subjects on an eMac with a 17 inch (16 inches viewable) monitor and a 1152 x 864 display. The monitor refresh rate was 80 Hz and the computer had a 1.25 GHz processor.

The simulation display was 10 x 6.5 inches. The simulation employed units of length that were equivalent to 22 units/inch. Velocities of agents in the simulations were expressed in units/simulation time (40 units of simulation time about equaled 1 s). The triangular agents had bases of 1 unit length and heights of 4 unit length. The subject controlled one agent with the mouse, which was a white circle with a 2 unit radius.

We programmed the agents under the same scheme as was employed in Experiment 1, with one change: When agents collided with other agents, they bounced off each other for .2 s at some random velocity vector. The automatic agents reacted to the subject-controlled agent in the same manner as any other triangular agent in the simulation.

We used the same pool of 12 agents from Experiment 1, each which had been created with 6 randomized parameters within the programming scheme.

3.1.3 Procedure

In each 15 second scene, the subject observed 6 agents and controlled 1 agent. 2 were red, 2 were green, 2 were blue, and the subject-controlled agent was white. The reds would behave according to the same parameters as the other reds, the greens according to a different set of parameters, and the blues according to a third set of parameters. The agents were drawn from a larger 12 agent pool; thus, there were 220 possible triads of these 12 agents. For each scene, one of these 220 triads was selected at random, and then each of the three programs in the selected triad was randomly assigned to red, green, or blue. Each subject saw 220 such scenes, exhausting the possible triads.

Subjects were openly encouraged to construe the triangular agents as animate, and were instructed that how agents of a certain color behaved during one trial would have nothing to do with how they behaved in subsequent trials. At the end of a 15 second scene, they were asked to determine which color of agent acted least like the other two—that is, which was most different: red, green, or blue? They answered by pressing “R”, “G”, or “B” on the keyboard, at which point the next trial began.

As in Experiment 1, we constructed a 12 x 12 symmetric distance matrix for each subject, to be fed into the individual differences multi-dimensional scaling (MDS) algorithm. For each trial, the two non-chosen agents in the odd-one-out procedure were made more similar within this distance matrix.

3.2 Results and Discussion

Once again, we derived a 2-dimensional MDS solution in order to visualize the space of agents that subjects (on average) perceived (see Figure 3.1). We again observed a ring-like structure in the space. The two MDS solutions (processed representations of the subjects’ original similarity ratings) were correlated. Dimension 1 in Experiment 1’s MDS was strongly correlated with Dimension 2 in Experiment 2’s MDS [$r(10) = .7125, p < .01$]. Dimension 2 in Experiment 1’s MDS was weakly (and negatively) correlated with Dimension 1 in Experiment 2’s MDS [$r(10) = -.5511, p = .0633$]. (The direction of these correlations is arbitrary and meaningless, but helpful in comparing the 2-dimensional MDS spaces presented in Figures 2.2 and 3.1.)

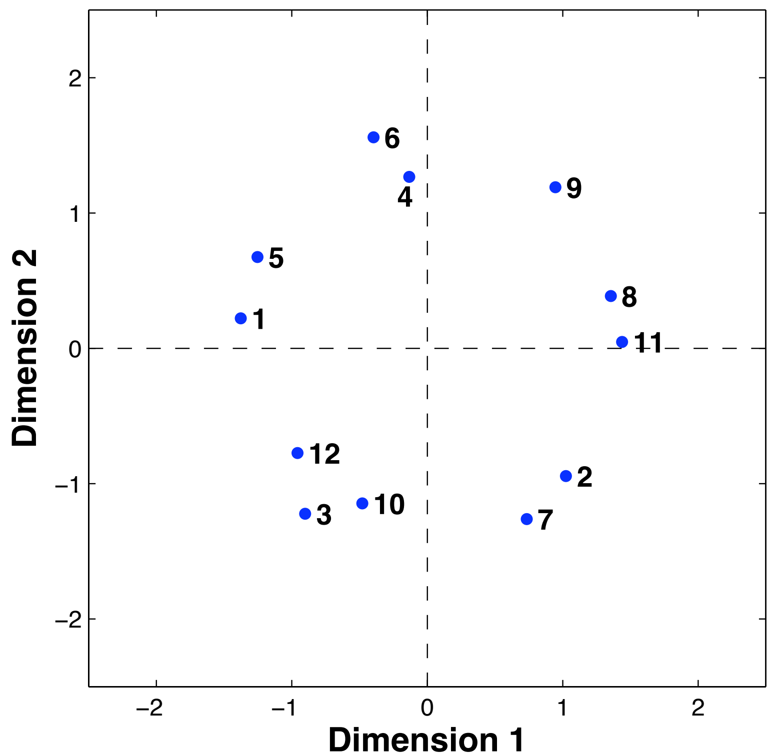

Figure 3.1: The 2-dimensional MDS solution for the 12 agents, fitting data from Experiment 2.

As shown in Table 3.1, Dimension 2 was correlated with parameter 5 in the agents' programming. This is consistent with the finding from Experiment 1, where Dimension 1 had been correlated with this same parameter. Apparently, how an automaton reacted to (i.e. accelerated toward or away from the direction of) its closest other agent in the simulation when that agent was 40-70 units away (10 to 17.5 times the length of an agent) was the most psychologically relevant programmed variable.

One might think that this result could have been the result of this inter-agent distance being the most common situation in the simulations we presented to the subjects. However, as shown in Figure 3.2, this was actually not the case. In Experiment 2, the two most common distances between an automaton and its closest other agent during a simulation were 0-5 units (0-1.25 agent lengths) and 20-40 units (5-10 agent lengths). Therefore, that 10-17.5 agent lengths would be the most psychologically critical distance is the result of some other cognitive process.

Table 3.1: Correlations ($r[10]$) between programmed parameters (rows) and MDS dimensions (columns). Bold font represents $p < .05$

	MDS Dimension 1	MDS Dimension 2
Parameter 1	-.129	-.147
Parameter 2	-.529	-.050
Parameter 3	.096	.195
Parameter 4	.548	.200
Parameter 5	-.310	**-.619**
Parameter 6	.249	.233

3.3 Further Discussion: Experiments 1 and 2

We did not discern any qualitative differences between the results of the two experiments, and therefore we here choose to pool data from all 15 subjects for the following discussion. The 2-dimensional MDS solution for these pooled subjects reveals an even cleaner ring structure (see Figure 3.3). But what does it mean as we travel around this ring?

In the combined MDS, Dimension 1 seems pretty clearly connected to how an agent behaves when the closest other agent is between 10-17.5 agent lengths away (i.e. programmed parameter #5). In this situation, the agents on the left side of the ring all tend to accelerate away from the nearest other agent, and the agents on the right side of the ring tend to accelerate toward the nearest other agent.

The meaning behind Dimension 2 is less straightforward, and furthermore this dimension is clearly not independent from Dimension 1. Because Dimension 2 does not correlate with any programmed agent parameter, in the absence of further psychophysics I can only rely on watching a number of simulations to speculate on its meaning. Ultimately, Dimension 2 alone does not seem represent anything obvious, but it may interact with Dimension 1 to produce distinct behavioral patterns. Agents around the top left section of the “ring” (Agents 6, 4, 9, 8, and 11) tend to appear generally “friendly.” Most of the other agents, occupying the majority of the “ring” (Agents 1, 5, 12, 10, 7, 2), could be characterized as generally “hostile” (Agent 3 seems to be a “friendly” exception to this characterization).

This potential “friendly” versus “hostile” dimension is neither orthogonal nor redundant with whether an agent accelerates toward or away from another agent at a certain distance—say, the distance with which programmed parameter #5 is concerned. That is, when an agent

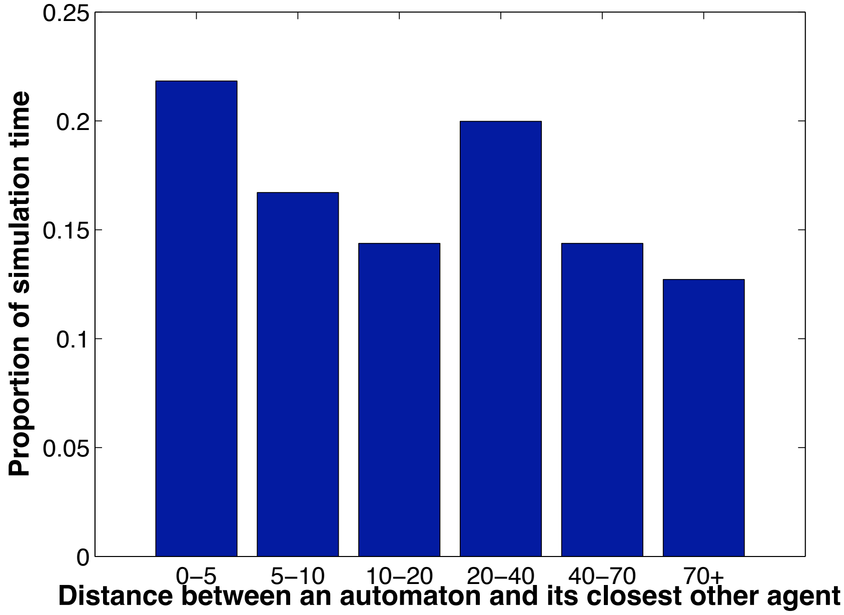

Figure 3.2: During Experiment 2, we recorded the positions of each agent in the simulation every tenth of a second. This figure shows the proportion of the time the inter-agent distance between an automaton and its closest other agent was within one of six bins, corresponding to the six underlying programmed parameters (subjects' data are pooled).

moves in the direction of another, it may appear to be aggressive or merely curious, and it is unclear what perceptual features contribute to these subjectively different characterizations. The two more basic MDS dimensions from Experiments 1 and 2 may interact to produce such characterizations.

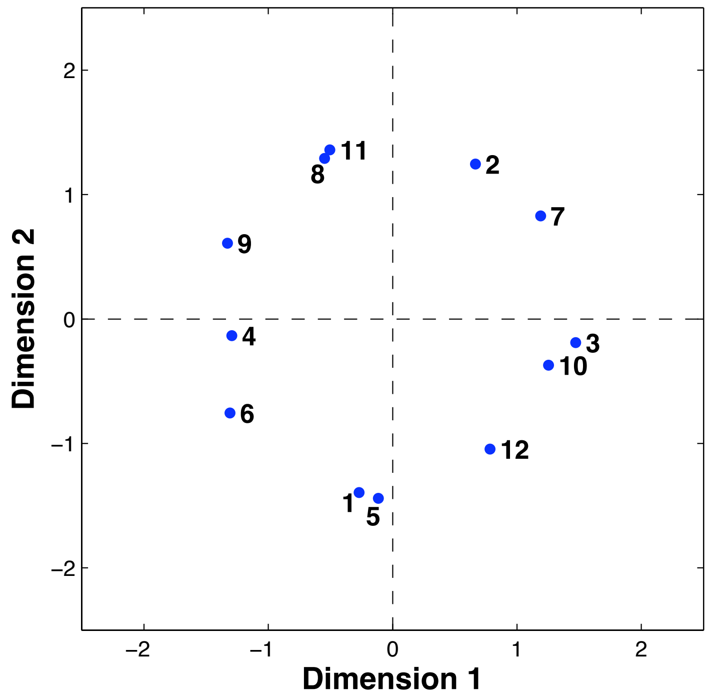

Figure 3.3: The 2-dimensional MDS solution for the 12 agents, fitting pooled data from Experiments 1 and 2.

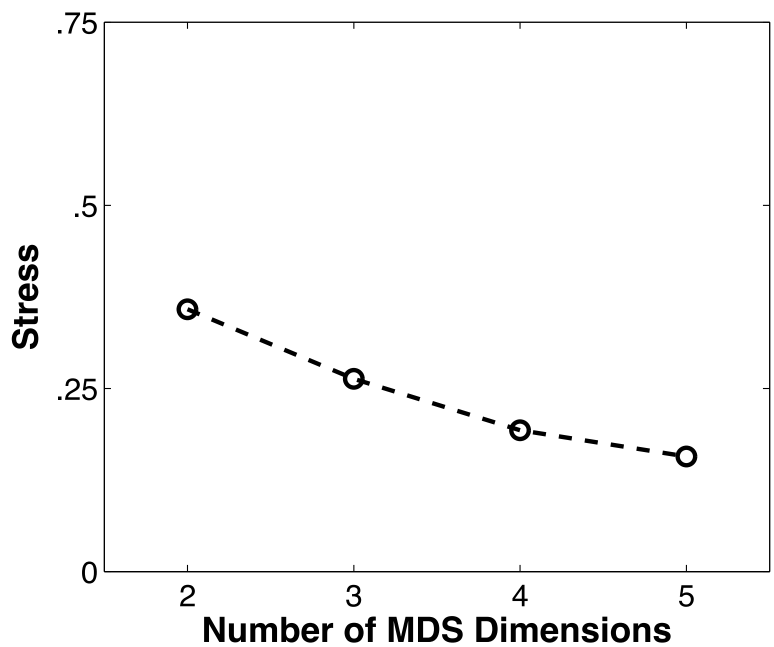

Figure 3.4: A scree plot for possible MDS solutions fitting the pooled data from Experiments 1 and 2. A greater number of dimensions will always result in a better fit, but there is no clear "elbow" in the plot favoring one particular number of dimensions over another.

4. Experiment 3

In Experiments 1 and 2, we used MDS in order to attempt to reconstruct subjects' psychological space of animate behaviors from psychophysical data. In Experiment 3, we introduce a different methodology for approaching the same problem.

As Experiments 1 and 2 illustrate, we could not reasonably have expected to detect clusters in the agent space by showing each subject only 12 different agents. In even a 2-dimensional space, 12 points provide insufficient resolution for such an endeavor. In fact, one rule of thumb for determining an appropriate number of clusters to search for with algorithms like k-means clustering is $k \approx (n/2)^{1/2}$, where k is the number of clusters and n is the number of data points (Mardia et al., 1979). So with only 12 points, one could only hope to detect 2-3 clusters at best. If more than 5 clusters actually existed in the space, one would need at least 50 points to provide a reasonable amount of resolution for their detection. If more than 7 clusters actually existed in the space, this would probably require more than 100 data points.

The primary motivation behind Experiment 3 is to address this issue.

4.1 Method

4.1.1 Subjects

Six undergraduate and graduate students between the ages of 18 and 24 participated in one half-hour experimental session.

4.1.2 Stimuli

We presented scenes to subjects on the same computer utilized in Experiment 1. The programming library employed units that were equivalent to 22 units/inch. Velocities of agents in the simulations were expressed in units/simulation time (40 units of simulation time about equaled 1 s). The triangular agents had bases of 1 unit length and heights of 4 unit length.

The agents were programmed with the same 6-parameter scheme employed in Experiment 2. We constructed two pools of agents, each with 6 randomized parameters within the programming scheme. The first pool (of 100 agents) we used as "models" to be analyzed offline. From the second pool (of 50 agents), we drew stimuli for the experimental sessions.

4.1.3 Procedure

Each subject watched one green triangle-agent interacting with six randomly-generated red triangle-agents for 15 s. They were instructed to pay special attention to the green agent, and were openly encouraged to think of the triangles as living, animate objects interacting with one another. At the end of this 15 s interval, the green agent disappeared from the scene, the six red agents were randomly relocated to new locations in the scene, and a new white triangle-agent appeared, controlled by the subject's mouse. The subject was instructed to, at this point, begin to mimic the original green agent, for 15 s. Subjects were discouraged from mimicking the green agent's exact trajectory, and instead encouraged to try and learn from what the agent had done in the previous 15 s and use this information to infer how that agent would react to the novel situation.

At the end of 15 s, the screen went gray, and the subject began the next trial by pressing the spacebar. The green triangle-agent to be emulated in each trial was drawn from the above-mentioned randomly-generated pool of 50, and each subject saw the same 50 agents over 50 trials, in random order.

Every half second, in both the watching and emulating stages of a trial, we recorded several variables describing how the green agent or subject-controlled agent (respectively) were moving about the scene. These included x-y position in the simulation, speed, distance to the closest other agent, speed in relation to the closest other agent, and the relative speed of the agent and the closest other agent (taking both agents' velocities into account). In the analyses below, the primary variable we explore is what the speed of the agent was in relation to the location of the nearest other agent in the scene.

4.2 Simulation

Offline, we ran 813 simulations of 15-second duration, each of which we populated with 7 randomly-selected agents from the above-mentioned 100-agent pool. We recorded variables relevant to their behavior in the scene every half-second, resulting in 30 time slices during each simulation at which we recorded data for each agent. We used this information (on average, about 1700 time slices per agent) to empirically determine cumulative probability distributions (CDFs) for each of these variables, for each of these 100 agents. These CDFs provide a partial summary of a given agent's typical behavior across various simulations.

The primary variable we examined was the speed of an agent in relation to the location of the nearest other agent in the scene. Thus, probabilities were of the form: In a 15 s scene in which the agent interacts with 6 other arbitrary agents, what is the probability that this agent takes on a given speed relative to the nearest other agent in the scene?

4.2.1 100 "model" agents

We wanted to get a sense of what the space of 100 "models" looked like, in a manner that was both easy to visualize and psychologically relevant. In Experiments 1 and 2 we showed that we could plot agents within a simplified 2-dimensional space that was somewhat robust between these two experiments. Because a 2D space allows for the visualization of psychological distances between the various agents, we favor it over MDS solutions of 3 or more dimensions, even though these solutions may fit the observed data better (Fig. 3.4 demonstrates exactly how much better).

Using the pooled 2-dimensional MDS solution for the 12 agents we used in Experiments 1 and 2, we employed multiple regression to analyze how these agents' 6 programmed parameters mapped onto their 2 MDS dimensions. By this method, we then formed a model for approximately placing any agent within this 6-parameter scheme into the 2-dimensional MDS space from Experiments 1 and 2.

Figure 4.2 shows the 100 models plotted in this space. While the two dimensions seem strongly correlated (and thus, it is possible that these two dimensions could be reduced to one), there are no discernible clusters among these 100 agents. Rather, the models seem evenly distributed throughout an elliptical space.

When we show subjects 50 new agents, and map their emulations onto best-fitting models in this space, we expect to observe clustering. This clustering, we argue, would be the result of subjects' observations being filtered by the prejudices of the human mind.

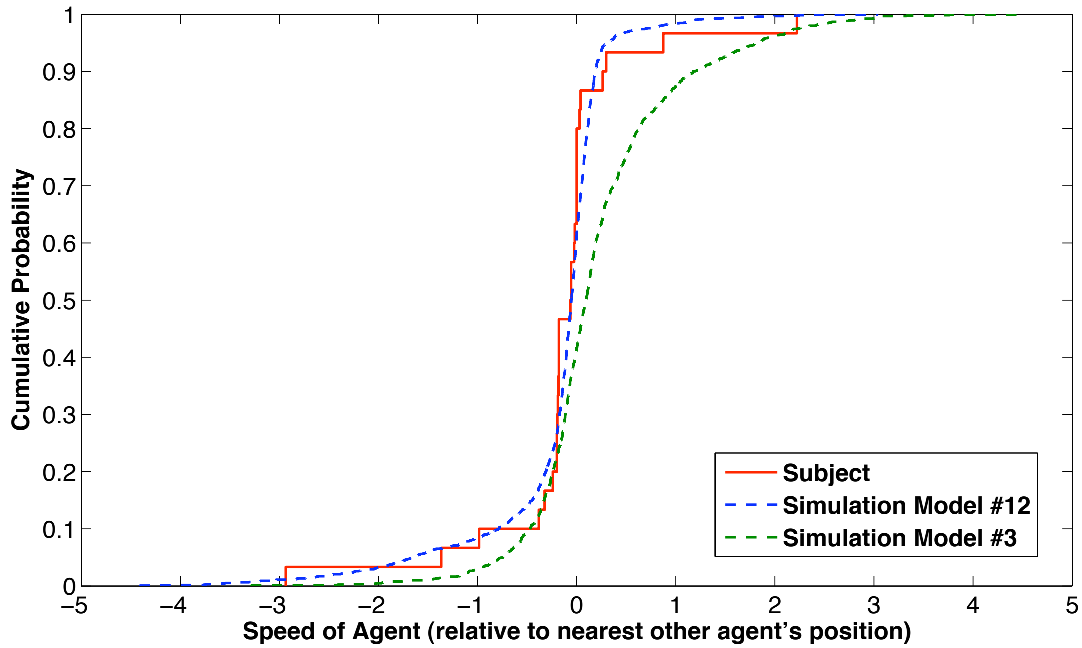

Figure 4.1: In this analysis, we examine agents' speed in relation to the position of the closest other agent in a simulation. The red curve shows the empirically determined cumulative probability distribution (CDF) for the subject-controlled agent during one trial of his experimental session. For this experimental trial, the empirically determined CDF for simulation model #12 (blue dashed line) fits the subject's data better than that of simulation model #3 (green dashed line). In fact, model #12 fits the data better than any of the other 99 models we simulated offline.

4.3 Results and Discussion

We empirically determined a cumulative probability distribution for the agent's speed over a 15 s scene, relative to the nearest other agent's location, for each of the 100 "models" we examined offline. Then, for each experimental trial, we compared the cumulative distributions of this variable for the agent during the trial with each of these 100 models, in order to find the best fit (by the Kolmogorov–Smirnov statistic). Figure 4.1 provides an illustration of this analysis, with which we asked: Given the observed behavior of an agent, if we only had this one variable to go by, which of the 100 models was it most likely to be?

If we pool the data and look at all 300 best-fitting models (50 per subject), (see Fig. 4.3), we see that some areas of the space were more heavily emulated than others (in the figure, the size of the blue dot is proportional to how many times this model was the best fit for a subject's

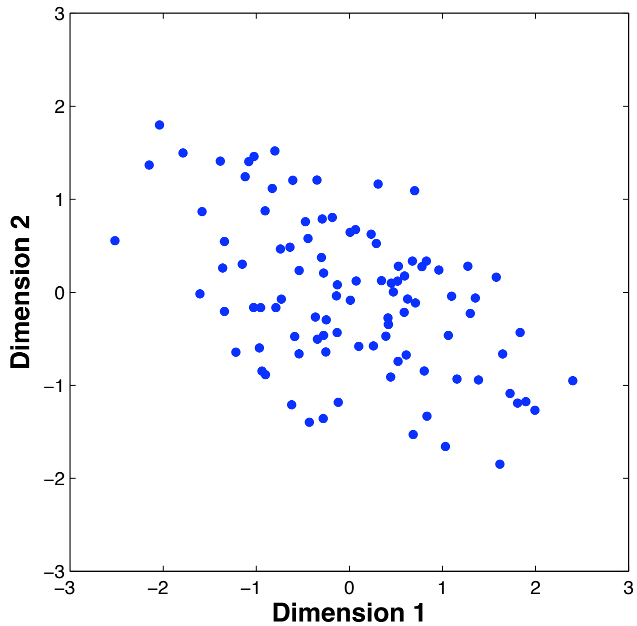

Figure 4.2: The 100 "model" agents from Experiment 3, plotted in a psychologically relevant space obtained from Experiments 1 and 2.

trial.)

Instead of using speed relative to the closest other agent, we could repeat this analysis using a number of different variables. Figure 4.4 shows the clustered space resulting from using distance to the closest other agent as the variable of interest.

It is unclear which of these variables, if either, is the appropriate one to use for this analysis. However, in both analyses there appears to be a cluster in the lower left quadrant of the space. In this area of the space, the agents are programmed along the critical parameter #5 to have a negative acceleration relative to the closest other agent. This may suggest a cluster of evasive agents in the subjects' mental space.

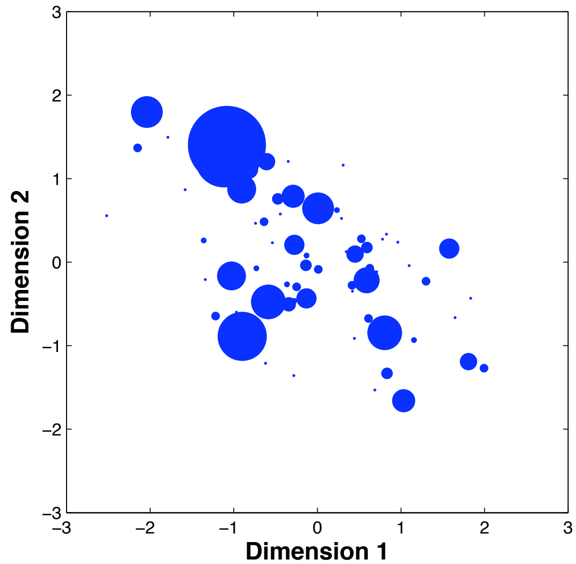

Figure 4.3: The results of 6 subjects' pooled mimicking performances in Experiment 3, plotted onto their best-fitting models in 2-dimensional MDS space, using relative speed to the closest other agent as the variable of interest. The size of a blue dot is proportional to how many times this model (one of the 100 that were analyzed offline) came up as the best-fitting one for a subject's trial.

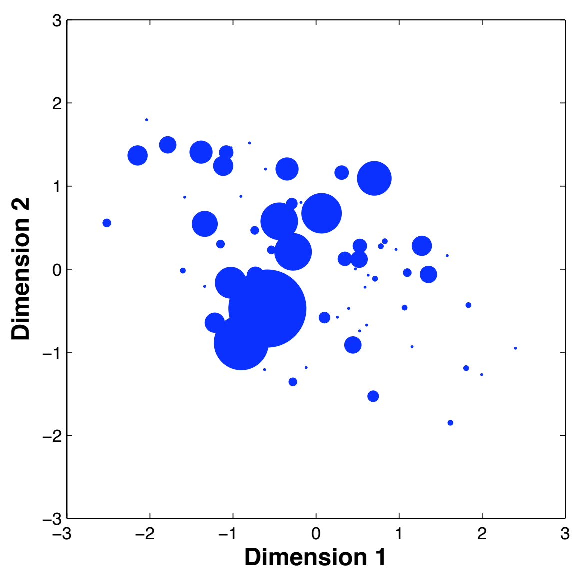

Figure 4.4: The results of 6 subjects' pooled mimicking performances in Experiment 3, plotted onto their best-fitting models in 2-dimensional MDS space, using distance from the closest other agent as the variable of interest. The size of a blue dot is proportional to how many times this model (one of the 100 that were analyzed offline) came up as the best-fitting one for a subject's trial.

5. General Discussion

Experiments 1 and 2 allowed us to get some sense of how the parameters we used to program our agents related to how subjects perceived the agent space. In Experiment 3, we attempted to use this knowledge to detect clusters in this space.

Clusters in this potentially high-dimensional agent space would represent something interesting but intuitive. When one observes the motion of an animate object, an infinite of possible intentions harbored by that agent could explain its chosen trajectory. For example, imagine you observe agent A moving slowly away from agent B. Consider 4 possible explanations for this behavior:

1. Agent A's behavior has nothing to do with agent B; agent A just happens to be moving in a direction that leads it away from agent B.

2. Agent A is avoiding a potentially dangerous agent B, and thus slowly retreats.

3. Acknowledging that agent B is also an intentional agent, agent A is trying to momentarily trick agent B into thinking he is retreating, to set up an ambush.

4. Sizing up agent B, agent A retreats for now. Tomorrow he will come back with 7 allies and successfully attack agent B.

5. Same as #4, but agent A has additionally made preparations for a catered party afterwards.

All of these can explain agent A's behavior. However, as we move down the list the explanations take on greater detail, and require a greater number of parameters. They require more inferences about agent A's mental architecture, and even about this agent's own "mind-reading" abilities. Therefore, the "higher order" explanations might be considered, a priori, less plausible in the mind of a third-party observer.

Ultimately, this may result in explanations like #2 being assigned the highest prior likelihood within the agent space: Seeing that agent A is moving slowly away from agent B, the observer concludes that agent A is retreating away from a potentially dangerous agent B.

Perhaps agent B is drawing the same conclusion about agent A; this is why a strategy like #3 can work in the first place.

5.1 A note about recursion and nested inferences

Inferences made about other mentalistic agents can take on a nested character, and can be arbitrarily long: e.g., ”I think that he thinks that I think that he thinks...”. A cognitive system that makes such inferences therefore is at peril of running into a loop of infinite recursion. At some point, the number of cognitive steps must be truncated.

Similar problems have been examined by economic game theorists. For some games, the derivation of a rationally optimal strategy or Nash equilibrium is the limit of some iterative reasoning process taken to infinite. But, empirically, experimental subjects playing these games seem to use strategies that truncate these processes to a more manageable number of “steps of iterated strategic reasoning” (Camerer & Fehr, 2006). The number of iterative steps can be termed the “order” of the strategy.

This manner of thinking can be applied to our line of research. In the above section, let us revisit the possible explanations for agent A’s behavior. Explanation #1 could perhaps be considered a 0th order explanation: It presumes very little about agent A’s mental architecture. Explanation #2 would be an example of a 1st order explanation: It acknowledges that agent A has intentions, goals, and motivations, but makes no inferences beyond a superficial matching of behaviors to intentions. Explanations #3-5 are all of a higher order.

And so, another way to ask our primary experimental questions might be: What order of explanation is most *natural* for a human observer to make. We imagine the answer is quite low. In Experiment 3, we potentially found a cluster of “evasion” in subjects’ agent space. For a simple construct like “evasion,” the order of the inferences the subject makes about the agent need not be very high at all.

5.2 Future Directions

Future experiments are necessary to determine whether the clusters we observe in Experiment 3 are a.) robust to replication and b.) match previous researchers’ intuitions about the natural categories of agent interaction. All three experiments, however, represent novel methodological approaches to the problem of how to examine subjects' inferential machinery for interpreting the interaction of animate agents.

A sensible next step for this experimental programme is to run experiments that show subjects dyadic interactions between agents. Much of the most relevant literature deals with dyadic interactions, and therefore this approach would help connect this work with previous studies. A better approach to having agents be the unit of analysis for the MDS might be to have subjects instead judge and compare the dyadic interactions.

But a more pressing task will be to develop a richer and more lifelike pool of agents with which we can populate our simulations. There is room for improvement.

5.3 A Better Pool of Agents

The programming of the next agents will take advantage of insights about cognition and action from the work of David A. Rosenbaum (e.g. Rosenbaum et al., 2001; Vaughan et al., 2001). When the motor system plans actions, it often does so in terms of achieving a goal end posture. Then, in reaching this goal state, it seems to obey certain principles, for example minimizing jerk. Most of Rosenbaum's research deals with motor actions like grasping, but the same principles might also apply to locomotion.

Consider an interaction between two agents. Under the previous programming scheme, the agent determines how far away the other agent is, and then accelerates toward or away from it based on this distance. Under the new programming scheme, the agent would have a set of subgoal points in an x-y plane (centered on the other agent—an aspect of the programming that we imagine will make the interaction seem more lifelike). The agent will then attempt to cycle through this sequence of subgoal points based on principles like minimizing jerk, and randomized parameters like its minimum/maximum speed or preferred curvature of path.

Now, these "subgoal points" would not be synonymous with the agent's internal, higher-order goals or intentions. But having an agent cycle in a lifelike way through a queue of such points has the potential to result in programmed behaviors that are both less structured and much more vivid that in our previous experiments.

The next agents may also have a capacity for first making a classification of another agent (its “kind”), and then reacting based on that classification. And, obstacles could be added into the simulations to provide additional context for the agent interactions.

Bibliography

Baker, C. L., Tenenbaum, J. B., & Saxe, R. B. (2006). Bayesian models of human action understanding. *Advances in Neural Information Processing 18*.

Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2007). Goal inference as inverse planning. *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*.

Barrett, H. C., Todd, P. M., Miller, G. F., & Blythe, P. W. (2005). Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior*, *26*, 313–331.

Blythe, P., Todd, P. M., & Miller, G. F. (1999). How motion reveals intention: categorizing social interactions. In G. Gigerenzer, P. M. Todd, G. F. Miller, & the ABC Research Group (Eds.) *Simple heuristics that make us smart*, (pp. 257–286). New York, NY: Oxford University Press.

Braitenberg, V. (1984). *Vehicles*. Cambridge, MA: MIT Press.

Camerer, C. F., & Fehr, E. (2006). When does ”economic man” dominate social behavior? *Science*, *311*, 47–52.

Dittrich, W. H., & Lea, S. E. G. (1994). Visual perception of intentional motion. *Perception*, *23*, 253–268.

Gelman, R., Durgin, F., & Kaufman, L. (1995). Distinguishing between animates and inanimates: not by motion alone. In D. Sperber, D. Premack, & A. J. Premack (Eds.) *Causal cognition: A multidisciplinary debate*, chap. 6, (pp. 151–184). New York, NY: Oxford University Press.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57*, 242–259.

Johnson, S. C. (2000). The recognition of mentalisic agents in infancy. *Trends in Cognitive Development*, *16*, 637–656.

Kahana, M. J., & Bennett, P. J. (1994). Classification and perceived similarity of compound gratings that differ in relative spatial phase. *Perception & Psychophysics*, *55*, 642–656.

Klein, J. (2002). breve: a 3d simulation environment for the simulation of decentralized systems and artificial life. *Proceedings of Artificial Life VIII, the 8th International Conference on the Simulation and Synthesis of Living Systems.*.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis*. London; New York: Academic Press.

McAleer, P., & Pollick, F. E. (2008). Understanding intention from minimal displays of human activity. *Behavior Research Methods*, *40*(3), 830–839.

Pantelis, P. C., van Vugt, M. K., Sekuler, R., Wilson, H. R., & Kahana, M. J. (2008). Why are some people's names easier to learn than others? the effects of face similarity on memory for face-name associations. *Memory & Cognition*, *36*(6), 1182–1195.

Rosenbaum, D. A., Meulenbroek, R. G., Vaughan, J., & Jansen, C. (2001). Posture-based motion planning: Applications to grasping. *Psychological Review*, *108*, 709–734.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*(4468), 390–398.

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika*, *42*(1), 7–67.

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, *29*, 943–951.

Tremoulet, P. D., & Feldman, J. (2006). The influence of spatial context and the role of intentionality in the interpretation of animacy from motion. *Perception & Psychophysics*, *68*(6), 1047–1058.

Vaughan, J., Rosenbaum, D. A., & Meulenbroek, R. G. J. (2001). Planning reaching and grasping movements: The problem of obstacle avoidance. *Motor Control*, *5*, 116–135.