The Emergence of Latent Force Representation in Human Perception of Social Interactions

Yiling Yun[1], **Yi-Chia Chen**[1], **Shuhao Fu**[1], **Hongjing Lu**[1,2]
[1] Department of Psychology, University of California, Los Angeles
[2] Department of Statistics, University of California, Los Angeles
{yiling.yun, yichiachen, fushuhao, hongjing}@g.ucla.edu

Abstract

Humans recognize social interactions effortlessly, even when presented with minimal visual information in unfamiliar displays. While force dynamics has been proposed as latent representations for perceiving social interactions, most research has approached this topic from a linguistic perspective based on conceptual knowledge, leaving open the central question of how latent force representations arise from visual inputs. The present study developed a force model that represents social interactions through two types of compositional forces: interactive forces, driven by interactions between agents; and self-propelled forces, driven by intentions of individual agents. Each force was formulated using a physics function to capture the dynamics of repulsive and attractive forces. We conducted two human experiments to measure human similarity judgments across a range of interaction animations and to evaluate recognition performance using generated animations in which the forces applied to individual agents were systematically manipulated. We found that the force model provides a parsimonious account for human judgments in both experiments. These findings suggest that mid-level representations based on compositional forces driven by different goals play an important role in social perception. We conjecture that the development of social perception may be grounded in perceptual mechanisms that support intuitive physics.

Keywords: Social Perception; Heider-Simmel; Force Representation; Human Interaction

Introduction

Rarely do we notice our ability to tell apart different kinds of social interactions, as it is so effortless. For example, we can walk into a restaurant and quickly see that some people are talking, some people are shaking hands, and some people are bidding goodbye. In fact, human minds are so sensitive to social interactions that body postures are not even necessary for differentiation: From animations of a few simple geometric shapes moving on a screen, people can see a complex story unfold involving multiple social interactions, such as seeing one shape pushing the other shape (Barrett, Todd, Miller, & Blythe, 2005; Heider & Simmel, 1944). This ability can be observed in both adults and children (Abell, Happe, & Frith, 2000; Springer, Meier, & Berry, 1996) and across different cultures (Barrett et al., 2005; Morris & Peng, 1994). Although Heider-Simmel types of stimuli with moving shapes lack body movements, facial expressions, and contextual information, they allow us to focus on the key factors that give rise to the impressions of various social interactions.

Humans are sensitive to certain social interactions, such as chasing and avoidance behavior. For instance, people naturally avert their gaze when they are caught staring as a way of social avoidance (Colombatto, Chen, & Scholl, 2020). From an evolutionary perspective, remembered dreams commonly involve scenarios of chasing and evading threats, perhaps serving as a rehearsal for real-life danger (Garfield, 2001; Revonsuo, 2000). These interactions can be described in terms of attractive and repulsive forces between the entities. While some research has explored latent representations of force, most remained primarily qualitative and focused on high-level concepts and semantics. For example, Talmy (1988) and Wolff (2012) analyzed verbs related to social interactions by explicitly separating individual intentions, the dynamics relative to the other entity, and the outcome of an action. Warglien, Gärdenfors, and Westera (2012) represented verbs using two force vectors in which one vector describes the agentive force, and the other vector describes the result of the actions. While these studies offer valuable insights into human perception of social interactions, they do not address fundamental questions: Do force representations emerge from visual input? If so, how? Here, we proposed that although forces themselves are not direct inputs to our sensory systems, latent force-like representations arise during visual processing and play a distinct role in human social perception beyond their associated low-level visual features.

There has been evidence that humans can see force in simple interactions. The classic study by Michotte (1963) demonstrated that humans perceive that one ball causes the other ball to launch. Later studies used the same paradigm to show that people perceive the strength of force based on the speed at which they moved (White, 2007). In more complicated interactions, Tang et al. (2021) showed that the human perception of chasing under the constraint of a leash can be captured by a joint inference model that includes a control force for the agent's intention and a constraint force that limits the agent's movements. These findings suggest that force representations may play a crucial role in shaping our perception of social interactions, though previous studies have only examined a limited range of interaction types, such as launching and chasing.

Another reason to study force representations in social interactions is that humans have always been exposed to forces in the physical world. Humans experience objects that move and rest according to well-documented physical laws, and show remarkable ability in physics-related perception and reasoning, as shown in many studies in intuitive physics (Kubricht, Holyoak, & Lu, 2017). We might rely on these regularities that we learned from observing the states and interactions of physical objects when interpreting social interactions.

Inspired by the idea that physical forces and intuitive physics shape humans' psychological representations of social interactions, the force formulation in physics has also been adopted to explain human behaviors. For example, Shu, Peng, Zhu, and Lu (2021) constructed a physical-social force model using Lagrangian mechanics to classify social versus physical events. Similarly, Helbing and Molnar (1995) adapted the Langevin equations from physics to model pedestrian dynamics through social forces, which are internal drives that lead to actions. Their model incorporated goal-directed acceleration, interpersonal repulsion to maintain distance, and attraction toward other pedestrians or objects, successfully describing the real-world pedestrian motion. These efforts illustrate how formal physical frameworks can contribute not only metaphorically but computationally to our understanding of social perception, bridging intuitive physics with predictive models of behavior. In the present paper, we developed a force-based model that uses the Lennard-Jones potential to capture force change between attraction and repulsion, that can account for approach and avoidance behaviors, allowing us to infer latent social dynamics from low-level visual input.

In the current study, we examine the emergence of latent representations of forces in human social perception. Specifically, we 1) evaluated how much human social representation can be attributed to force representations abstracted from visual inputs, 2) explored latent representation that can be generalized to a variety of social interactions, and 3) framed force as the causal explanation to design an intervention study of altering recognition of social interactions. We adapted a parametric function that was well tested in physics to predict the trajectories of two shapes. We conducted two human experiments using few-seconds-long Heider-Simmel type of animations that depicted various social interactions (such as hug, approach, fight, and avoid). In the first study, we collected human similarity judgments of 27 different social interactions and compared the results with model-simulated similarity matrices, that were derived from four models including low-level visual features, a deep learning model, our force model, and semantic labels. In the second study, we generated trajectories of agents using forces controlled by fitted parameters and measured the impact of imposed forces on human impression of social interactions for the generated animations.

Study 1: Human Social Impression through Similarity Judgments

We used animation stimuli in the Charade dataset developed by Roemmele, Morgens, Gordon, and Morency (2016). The Charade dataset includes a total of 1156 animations demonstrating 31 social interactions. In each animation, there were two black triangles moving on a white background. One triangle was larger than the other one. The animations were created from a charade game where human annotators were asked to demonstrate action words (such as hug and fight) by manually moving the two triangles. The trajectories and facing directions of the two triangles were recorded as the animations. The generated animations were further evaluated by a different group of participants.

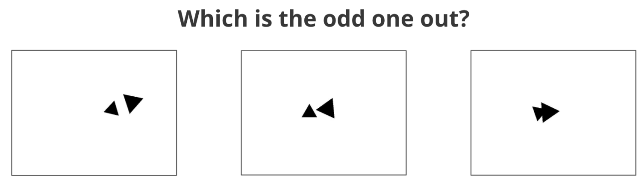

Figure 1: Illustration of a static frame of each stimulus presented in one trial. Participants were asked to select the odd-one-out among the three animations.

Methods

Participants Seventy-seven students in the Psychology department at UCLA participated in the online experiment. We excluded one participant who did not complete all the trials, five participants who self-reported not being serious throughout the experiment, and one participant who self-reported not staying in the full-screen mode throughout the experiment. We analyzed data from the remaining 70 participants (Female: 57, Male: 13; Mean age = 20.49).

Stimuli We selected 88 good-quality animations in the Charade dataset that described 27 social interactions with durations varied from 2 – 6 seconds. We further selected one animation that best described each social interaction. The selected 27 animations were associated with social interactions with labels, including hug, huddle, kiss, approach, flirt, scratch, poke, creep, tickle, hit, talk, fight, escape, lead, herd, accompany, throw, ignore, leave, avoid, bother, push, capture, follow, pull, examine, and encircle. Two researchers independently made their selections. When there was a disagreement, we asked three additional raters to break the tie. Therefore, we included a total of 27 animations to test in the experiment.

The experiment was programmed in HTML, JavaScript, CSS, and PHP. On each trial, we displayed three animations side by side in the center of the computer screen. Each animation was 300 pixels in width in its original ratio, and there was a gap of 20 pixels between each two of them (Figure 1). After watching all three animations one after another, participants clicked on a button beneath the corresponding animation to indicate their odd-one-out judgment.

Design We used the odd-one-out task to assess the similarity between each pair of the 27 animations representing different social interactions. On each trial, we displayed three animations with the prompt "Which is the odd one out?".

When a participant selected one animation as the odd-one-out, their response implied that they considered the two unselected animations to be more similar to each other than to the selected animation. We then measured the similarity between the two animations by finding the proportion of trials in which neither of the two animations was selected among all trials that contained these two animations. Afterward, we calculated the dissimilarity scores by calculating the similarity score between two animations and subtracting this similarity score from 1. The dissimilarity matrix was later tested against distance matrices derived from different models.

To test each pair of animations against all remaining 25 animations, we created the full combination of 2,925 unique trials and randomly assigned them to 65 different versions of the experiment. Each participant received one version. The order of the trials for each participant was randomized. The position of the three animations was also randomized for each trial. In total, each participant completed 45 trials.

Procedure Participants accessed the experiment from their personal laptops. They first read the instructions about the task and were shown an example animation of pushing. They then familiarized themselves with the task through one example trial. After an instruction quiz question that tested their understanding of the task, they gave consent to start the experiment. There was no time limit for their decisions. No feedback was given, so they were not guided to make judgments in a particular way. There was a progress bar at the top of the screen. Participants could watch the animations in any order and for as many times as they wanted. They could only proceed to the next trial after they had watched all three animations and selected one odd-one-out animation. After completing all the trials, we administered some survey questions to ask if they were serious throughout the experiment, had any comments about the study, or had encountered any technical issues.

Models

Low-level Visual Features To explore what visual cues in the animations contributed the most to the similarity ratings, we computed visual features for each animation: average speed, average acceleration magnitude, average velocity, average acceleration, duration, average relative distance, and average speed difference. We calculated the average speed across two entities to test the hypothesis that humans might judge two animations as more similar to each other if entities move at similar speeds. We computed acceleration magnitude to test the hypothesis that humans may perceive two animations as more similar when the entities exhibit comparable variations in speed. The low-level visual cues included velocity and acceleration to take into account directions in addition to speed and acceleration magnitude. Average relative distance and average speed difference account for the difference between the two entities.

The specific calculations of speed and acceleration magnitude were as follows: To compute the average speed, the movement distance (displacement) of each triangle between two consecutive frames was calculated and then averaged over time across two triangles (unit: px/frame); average acceleration magnitude was calculated as the change in speed of each triangle in each frame and then averaged over time across two triangles. For each animation, there was one value for average speed and one value for average acceleration magnitude. The specific calculations of velocity and acceleration were as follows: to indicate the direction in velocity, we represented the average velocity of one triangle in a vector with two elements – the average displacement in the horizontal and vertical direction over time for each triangle. We then concatenated the two vectors of the two triangles; for acceleration, we also represented it in a vector form with the average change in displacement in the horizontal and vertical direction over time for each triangle. We then concatenated the two vectors of the two triangles. Therefore, for each animation, there was one vector of four values for velocity and one vector of four values for acceleration. Duration was represented by one value. Average relative distance and average speed difference were calculated by taking the differences in location and speed between the two triangles, averaged over time. Each was represented by a single value as well. Together, we concatenated all the features and generated a vector of length 13 for each animation. We then calculated the pairwise Euclidean distances to construct a distance matrix as an estimate of the dissimilarity judgments from these basic summary features.

Deep Learning Model Since the animations involved rich visual information evolving over time, we employed a Long Short-Term Memory model (LSTM) that is designed to process sequential data such as trajectories for a supervised learning categorization task (Hochreiter & Schmidhuber, 1997). Using the LSTM architecture, we input the coordinates, orientations, and velocity vector of the two triangles for each frame as a sequence of data, with a step size of 5 frames. We trained the model through triplet loss: to decide among three animations, which one was most different from the other two, i.e., the odd-one-out task, using the animation label from the Charade dataset. Specifically, we fed in three animations as the input to the model, where two animations described the same animation label and the other one described a different animation label. The task for the model was to decide which animation depicted a different social interaction. This process mimicked the similarity judgment in the human experiment. We trained on 1129 animations from the Charade dataset using two recurrent layers with 100 hidden dimensions each. We trained the model for 1000 epochs with a learning rate of 0.00001. Note that none of the 27 animations used in the experiment was included in training. We then used the trained model to derive an embedding vector of length 64 for each of the 27 animations that we tested in the behavioral experiment. We then estimated the dissimilarity matrix by calculating the pairwise cosine distance between the embeddings of animations.

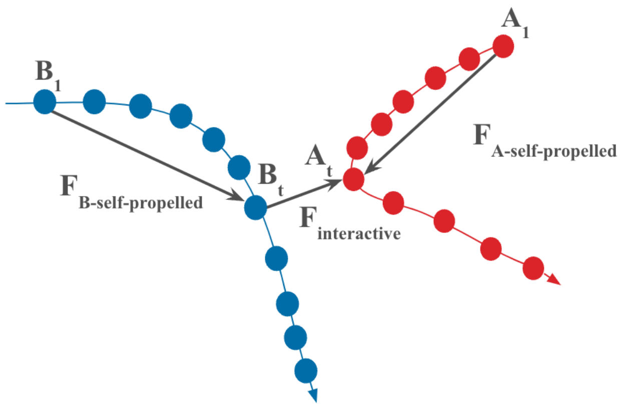

Figure 2: Force composition illustration. A_t refers to agent A's location at time t. B_t refers to agent B's location at time t.

Force Model Force has the computational advantage due to its compositional nature. Two forces can be combined together to impact the movement of an entity. We hypothesized that two types of compositional forces are crucial for capturing social interactions: interactive forces ($F_{interactive}$) driven by interactions between agents, and self-propelled forces ($F_{self-propelled}$) driven by individual intentions. For each force, the specific parametric function was inspired by particle movements in physics. Specifically, the intuition behind particle movements is that when two particles are far from each other, they have the tendency to attract each other controlled by attractive force; when two particles are too close, they have the tendency to repel each other controlled by repulsive force. The latent forces determine the movement of each entity. We used a standard function of the Lennard-Jones potential to estimate such an attractive-repulsive force as a function of distance r between two entities (Lennard-Jones, 1925):

$$F(r) = 48\varepsilon\left(\frac{\sigma^{12}}{r^{13}} - \beta\frac{\sigma^{6}}{r^{7}}\right)$$

The force function is governed by three key parameters that were latent variables σ, ε, and β. To be specific, the variable σ controls the critical distance where the attractive force would change to the repulsive force. The variable ε and β capture the changing rate of the forces as a function of the distance between two entities. In our model, given r denoting the distance between two triangles that can be calculated in the animation, we estimated force parameters σ, ε, β for every time window of 11 frames, with a step size of 5 frames. As illustrated in Figure 2, to estimate the parameters for a time window, we first predicted the locations of agent A at time point t, with t going from 3 to 11. Specifically, at each time point, we characterized $F_{interactive}$ using the Lennard-Jones function, with the distance between B_t and A_t as input. We characterized $F_{A-self-propelled}$ based on the distance and the unit vector between the inital position A_1 and observed A_t at frame t. For agent B, the same method was used to estimate parameters for the self-propelled force. The force parameters are estimated to minimize the predicted positions with the observed positions. For example, we combined $F_{interactive}$ and $F_{A-self-propelled}$ through vector addition and predicted the next location A_{t+1} using the estimated force controlled by the parameters. We used coarse grid search for the set of three parameters to find the initial values, and then followed by using MATLAB fminsearch for a fine-tuned search to find the local minimum. As a result, for each animation, every 11 frames are represented by 9 parameters. For the entire animation, we then compute the histogram of each force parameter. To derive the dissimilarity scores across animations, we calculated the Euclidean distance of the histogram of force parameters.

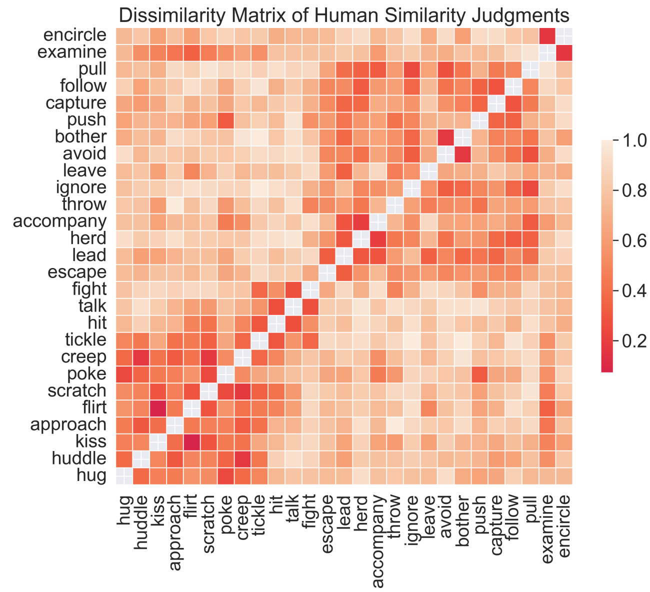

Figure 3: Dissimilarity matrix of human judgments. The darker red indicates a higher similarity (low dissimilarity) between the two corresponding animations.

Semantic Label We obtained word embedding of labels for each animation using fastText model (Bojanowski, Grave, Joulin, & Mikolov, 2016) as a high-level semantic representation of the social interaction. Each social interaction's label was represented by a vector of length 300. We used pairwise cosine distance to estimate the dissimilarity matrix.

Results

From the behavioral experiment, we calculated the dissimilarity scores based on the odd-one-out task. These scores were determined by the proportion of trials in which neither animation was chosen, considering all trials that included the given pair. Figure 3 illustrates the human dissimilarity scores, revealing distinct similarity structures across various types of social interactions.

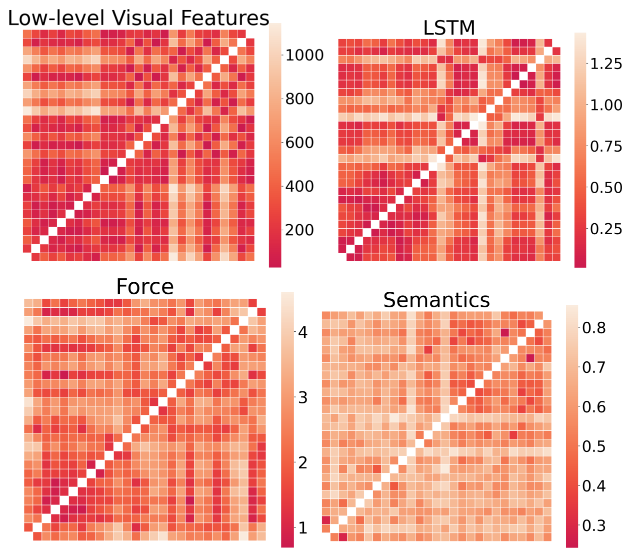

Figure 4: Dissimilarity matrices predicted by models. The display sequence of animations is the same as Figure 3. Top left: Low-level Visual Features model; Top right: LSTM; Bottom left: Force model; Bottom right: Semantic Label.

We then compared the human similarity judgments with the modeling results (Figure 4) by computing the Pearson's correlation between human dissimilarity scores and model-predicted distance scores. Figure 5 shows the result between model predictions and the human similarity judgments: the force model produced the highest correlation ($r = .499$) with human judgments, outperforming the other three models. The low-level visual features and the LSTM model generated moderate correlations ($r = .338$ and $r = .329$, respectively). Semantics has the least correlation ($r = .143$). The reliability noise ceiling calculated through split-half of human data was 0.811, which represents the highest correlation possible with the given human similarity judgments for computational models.

Using a semi-partial correlation test to control the effect of visual features and LSTM on predicting human similarity judgments, the force model provided unique contribution in account for human judgments, as revealed by a significant semi-partial correlation ($sr = .361$, $p < 0.001$). When controlling the effect of only visual features on predicting human similarity judgments, the force model also remained a significant semi-partial correlation ($sr = .371$, $p < 0.001$).

Study 2: Social Impression of Animations Created Using Force Parameters

The previous experiment demonstrated that the force model best accounts for human similarity judgments of social interactions. If human perception of social interactions depended on the intermediate level representations of forces, then manipulating the underlying force should change people's impressions of the animation. In study 2, we manipulated the

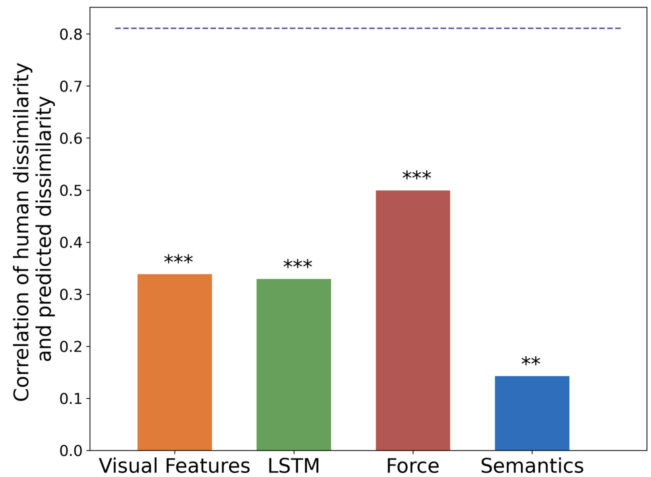

Figure 5: Correlation between human dissimilarity judgments and model-predicted dissimilarity of animations.

animations by maintaining the trajectory of one agent shape, but generating trajectories of the other agent shape based on the imposed forces from a different social interaction. We then asked subjects to label the new animations and examined if different forces would give rise to different social impressions. Note that only parameters from the force model were imposed to generate new trajectories.

Methods

Stimuli We first ran the force model to prepare for the generation of new animations. For each of the 27 animations used in study 1, we fitted the force model on its trajectories for each social interaction and obtained corresponding force parameters for interactive force, and self-propelled force of individual agents. Specifically, for every 11 frames in the trajectories, the compositional forces contained nine parameters, of which three described the interactive force, and the rest described the self-propelled force of the agents. Using the force parameters estimated from an animation, we computed the forces based on distances and then added the force to other animations to generate new trajectories of an agent. When the original animation and the forces had different durations, we trimmed whichever was the longer one. For example, when we applied the "accompany" force to the "escape" interaction, we maintained agent B's trajectory in the original "escaping" animation, and generated agent A's trajectory so that the self-propelled force and the interactive force were governed by force parameters estimated from the "accompanying" animation. Our question was whether the generated animation would give rise to an impression of "accompanying" more likely or less likely than "escaping" interaction.

We selected a subset of 10 social interactions to test human perception of them: huddle, escape, lead, herd, accompany, ignore, bother, capture, follow, and pull. Animations were chosen where agent B exhibited noticeable movement (i.e., Total distance and max speed were both greater than 10 pixels). These social interactions were selected because if agent B barely moved in two interactions, applying the same force would result in nearly identical generated trajectories for agent A as in the original animation. For agent A, its trajectory was calculated using the imposed force parameters from a different social interaction. The generated animations featured a large circle representing agent A and a small circle representing agent B.

In total, we generated 90 animations. We also included the 10 original animations with circles replacing the triangles. The animations had durations within the range of 2.28 to 5.48 seconds.

Design The study contained 100 trials in total. On each trial, participants watched one animation. After watching it, they were given the 10 social interaction labels and were asked to select only one label. The instruction was "Select the label that best describes the animation. The big circle is the agent who takes the action, and the small circle is the target of the action." There was no time limit and no feedback. Participants were allowed to play the animation as many times as they wanted before they made their judgments. The display order of the labels was the same across all trials for each participant but was randomized across participants. The order of the 100 trials was randomized across participants.

For the analysis, we calculated the proportion of trials in which participants selected the force label (the label that selects the corresponding force parameters to generate trajectories of agent A), and the proportion of trials where participants selected the original label (the label that determines the trajectory of agent B). The trials that included the original animations served as a control to measure the baseline response proportion for the corresponding labels.

Participants One-hundred-twenty-five students in the Psychology department at UCLA participated in the online experiment. We excluded nine participants who did not complete the end survey due to some technical difficulties, six participants who self-reported not being serious throughout the experiment, and two participants who self-reported not staying in the full-screen mode throughout the experiment. We analyzed data from the remaining 108 participants (Female: 90, Male: 16, Prefer non-disclosure: 2; Mean age = 20.14).

Results

For the baseline trials with original animations, participants on average selected the original label in 32.13% of the trials, notably above the random chance level of 10%. For the trials with generated animations, participants on average selected the original label in 10.14% of the trials, and they selected the force label in 15.30% of the trials. Compared to the random chance level of 10%, one-sample t-tests showed that participants selected the force label significantly more than chance ($t(107) = 13.15$, $p < 0.001$). In contrast, there was no significant difference between the selection of the original label and the random chance level ($t(107) = 0.43$, $p = 0.33$). Additionally, a paired-sample t-test showed that participants selected the force label significantly more frequently than the original label ($t(107) = -10.32$, $p < 0.001$).

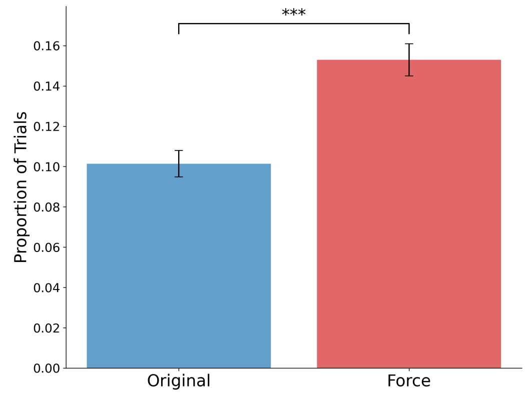

Figure 6: Proportion of responses in selecting labels to describe generated animations. The "original" label determined the trajectory of agent B; the "force" label determined the corresponding force parameters used for generating trajectories of agent A. The error bars indicated the 95% confidence interval for each selected label.

Discussion

The present study used both similarity judgments and a recognition task to reveal the emergence of latent force representations in human perception of social interactions. Thus, forces are not merely hypothetical proposal to connect perception and cognition, but serve as essential mid-level representations in human social perception.

Study 1 showed that human similarity judgments had the highest correlation with the force model, outperforming low-level visual features and LSTM. All three visual models outperformed the semantic labels. In addition, the semi-partial correlation analysis demonstrated that the force model explained unique variation in human similarity judgments. These findings suggest that people interpret social dynamics through compositional forces driven by distinct goals. Study 2 showed that imposed forces can directly alter human perception of social interactions, changing from one interaction label to the other.

In conclusion, the current study highlight a promising direction of considering force as a latent representation in service of human social perception. The force model has potential to be generalized to complex real-world social interactions such as pedestrians' movements (Farina, Fontanelli, Garulli, Giannitrapani, & Prattichizzo, 2017). The study sheds light on the development of social perception, which may build upon perceptual processes underlying intuitive physics.

Acknowledgments

This work was supported by the NSF grant BCS-2142269.

References

Abell, F., Happe, F., & Frith, U. (2000). Do triangles play tricks? attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16.

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*(4), 313–331.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). Enriching word vectors with subword information. *arXiv preprint arXiv:1607.04606*.

Colombatto, C., Chen, Y.-C., & Scholl, B. J. (2020). Gaze deflection reveals how gaze cueing is tuned to extract the mind behind the eyes. *Proceedings of the National Academy of Sciences*, *117*(33), 19825–19829.

Farina, F., Fontanelli, D., Garulli, A., Giannitrapani, A., & Prattichizzo, D. (2017). Walking ahead: The headed social force model. *PloS one*, *12*(1), e0169734.

Garfield, P. (2001). The universal dream key. *San Francisco: HarperOne*.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology*, *57*(2), 243–259.

Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical review E*, *51*(5), 4282.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *xg*, *9*(8), 1735-1780.

Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in cognitive sciences*, *21*(10), 749–759.

Lennard-Jones, J. E. (1925). On the forces between atoms and ions. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *109*(752), 584–597.

Michotte, A. (1963). The perception of causality.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and chinese attributions for social and physical events. *Journal of Personality and Social psychology*, *67*(6), 949.

Revonsuo, A. (2000). The reinterpretation of dreams: An evolutionary hypothesis of the function of dreaming. *Behavioral and brain sciences*, *23*(6), 877–901.

Roemmele, M., Morgens, S.-M., Gordon, A. S., & Morency, L.-P. (2016). Recognizing human actions in the motion trajectories of shapes. In *Proceedings of the 21st international conference on intelligent user interfaces* (pp. 271–281).

Shu, T., Peng, Y., Zhu, S.-C., & Lu, H. (2021). A unified psychological space for human perception of physical and social events. *Cognitive Psychology*, *128*, 101398.

Springer, K., Meier, J. A., & Berry, D. S. (1996). Nonverbal bases of social perception: Developmental change in sensitivity to patterns of motion that reveal interpersonal events. *Journal of Nonverbal Behavior*, *20*, 199–211.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive science*, *12*(1), 49–100.

Tang, N., Gong, S., Liao, Z., Xu, H., Zhou, J., Shen, M., & Gao, T. (2021). Jointly perceiving physics and mind: Motion, force and intention. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Warglien, M., Gärdenfors, P., & Westera, M. (2012). Event structure, conceptual spaces and the semantics of verbs. *Theoretical linguistics*, *38*(3-4), 159–193.

White, P. A. (2007). Impressions of force in visual perception of collision events: A test of the causal asymmetry hypothesis. *Psychonomic Bulletin & Review*, *14*, 647–652.

Wolff, P. (2012). Representing verbs with force vectors. *Theoretical linguistics*, *38*(3-4), 237–248.