The visual coupling between neighbors explains local interactions underlying human ‘flocking’

Gregory C. Dachner, Trenton D. Wirth, Emily Richmond, and William H. Warren*

Department of Cognitive, Linguistic, and Psychological Sciences
Brown University, Providence, RI 02912

*Corresponding Author
Email: Bill_Warren@brown.edu

ORCID IDs:

Gregory Dachner: 0000-0002-9198-330X

Trenton Wirth: 0000-0002-2510-8603

Emily Richmond: 0000-0001-8667-0969

William Warren: 0000-0003-4843-2315

Keywords:

collective behavior, crowd dynamics, pedestrian dynamics, vision-based model, agent-based model

Abstract

Patterns of collective motion in bird flocks, fish schools, and human crowds are believed to emerge from local interactions between individuals. Most 'flocking' models attribute these local interactions to hypothetical rules or metaphorical forces and assume an omniscient 3^{rd}-person view of the positions and velocities of all individuals in space. We develop a *visual model* of collective motion in human crowds based on the visual coupling that governs pedestrian interactions from a 1^{st}-person embedded viewpoint. Specifically, humans control their walking speed and direction by canceling the average angular velocity and optical expansion/contraction of their neighbors, weighted by visibility (inverse of occlusion). We test the model by simulating data from experiments with virtual crowds and real human 'swarms'. The visual model outperforms our previous omniscient model and explains basic properties of interaction: 'repulsion' forces reduce to canceling optical expansion, 'attraction' forces to canceling optical contraction, and 'alignment' to canceling the combination of expansion/contraction and angular velocity. Moreover, the neighborhood of interaction follows from Euclid's Law of perspective and the geometry of occlusion. We conclude that the local interactions underlying human flocking are a natural consequence of the laws of optics. Similar perceptual principles may apply to collective motion in other species.

Background

Human crowds exhibit patterns of collective motion in many public settings, from train stations and shopping plazas to – sometimes catastrophically – mass events [1, 2]. Similar patterns of coordinated motion are observed in bird flocks, fish schools, and animal herds, suggesting that diverse systems obey common principles of self-organization [3, 4]. It is generally believed that these global 'flocking' patterns emerge from local interactions between individuals [3-5]. The crux of the problem thus lies in understanding the nature of the local interactions.

Most models of collective motion ascribe these interactions to hypothetical rules or metaphorical forces, often inspired by physical systems, and assume an omniscient, 3rd-person view of the positions and velocities of all individuals in space [6, 7]. Such phenomenological models – including our own [8] – describe relations between individuals without offering an underlying mechanism. But humans and animals are embedded in groups and coupled to their neighbors by sensory information. Here we develop a visual model of collective motion that explains local interactions in terms of the visual coupling, based on optical variables. Not only does the visual model outperform our previous omniscient model, but basic properties of interaction follow from the laws of optics.

Understanding local interactions involves, first, identifying the *rules of engagement* that govern how an individual responds to a neighbor, and second, characterizing the *neighborhood of interaction* over which the rules operate and the influences of multiple neighbors are combined. Classical "zonal" models [9-11] posit three local rules or forces in concentric zones: (i) *repulsion* from neighbors in a near zone to avoid collisions, (ii) *alignment* with the velocity of neighbors in an intermediate zone to generate common motion, and (iii) *attraction* to neighbors in a far zone to ensure group cohesion. Influences are combined by averaging neighbors within a zone, sometimes weighted by their distance [12, 13]. An alignment rule by itself is theoretically sufficient to generate collective motion [14], as is the combination of attraction and repulsion [15]. In humans, the prominent Social Force model [16, 17] also assumes attraction and repulsion, successfully simulates key crowd scenarios [18, 19], and can produce collective motion under certain boundary conditions [20, 21]. However, it does not generate realistic individual trajectories [22] or generalize between situations without re-parameterization [17, 23].

The strength of such physics-inspired models is that they capture generic properties of collective motion, yet the same global patterns can be generated by different sets of local rules [5, 24]. To infer the actual rules, researchers have turned to behavioral experiments on local interactions [25-28]. We believe that such a 'bottom-up' approach should be grounded in the sensory coupling that actually governs these interactions. The coupling incorporates limits on the sensory range and field of view [10, 29] as well as the visibility of individual neighbors [30, 31]. Moreover, local interactions strongly depend on the visual information that controls locomotion [32, 33]. This insight has inspired recent 'vision-based' models [34-36], but the effective optical variables remains to be determined.

We take a bottom-up, experiment-driven approach called 'behavioral dynamics' [27, 37]. Our initial experiments on following in pedestrian dyads [38, 39] revealed that humans obey an alignment rule: the follower tends to match the walking direction (*heading*) and speed of the leader. To infer the neighborhood of interaction, we immersed walking participants in a virtual crowd and manipulated the motions of the avatars; we also analyzed observational data on human 'swarms' [8]. The results showed that pedestrians follow a crowd by averaging the heading directions and speeds of neighbors within a 180° field of view, with weights that decay exponentially with distance to zero at 4-5m. The findings led to an *omniscient model* of collective motion [8] based on the weighted average of neighbor headings and speeds (Figure 1a; see SM). The model successfully predicts individual trajectories in both virtual crowd experiments and real crowd data [8, 40], and the 'soft metric' neighborhood generates robust collective motion in simulation [13, 41].

Like its predecessors, however, our omniscient model relied on metaphorical forces, assumed physical variables as input, and did not account for the neighborhood of interaction. In this article we report new experiments that lead to an embedded *visual model* (Figure 1b), predicated on the optical variables that control pedestrian following [42, 43]. This new model explains the rules of engagement and the form of the neighborhood as natural consequences of the laws of optics.

Experimental Methods

Human subjects: Twelve subjects (7F, 5M) participated in Experiment 1, and ten different subjects (6F, 4M) in Experiment 2. A power analysis determined that a sample size of 8 per experiment was sufficient to achieve a power of 0.85 with $\alpha = .05$ and an effect size of 0.5 ($\eta^2 = 0.2$) [44]. All participants gave informed consent and were compensated for their time. The research protocol was approved by Brown University's Institutional Review Board in accordance with the principles expressed in the Declaration of Helsinki.

Equipment: Participants walked freely in a 12m x 14m tracking area while viewing a virtual environment in a wireless, stereoscopic head-mounted display (Oculus Rift DK1, 90°H x 65°V field of view, 640 x 800 pixels per eye, 60 Hz refresh rate). Head position and orientation were recorded with an inertial/ultrasonic tracking system (Intersense IS-900; 60 Hz sampling rate) and used to update the display with a latency of 50-67 ms.

Displays: The virtual environment (WorldViz software) consisted of a green start pole and a gray orientation pole 12.73 m apart on a granite-textured ground plane, with a blue sky. The virtual crowd consisted of animated 3D human models (WorldViz Complete Characters). These virtual humans were initially positioned on arcs with the start pole at the center, at randomly assigned eccentricities (±6˚, ±19˚, ±32˚, ±45˚) about the direction to the orientation pole, then randomly jittered.

Procedure: To elicit collective-motion responses, participants were instructed to "walk with the group of virtual humans" and "treat them as if they were real people." On each trial, the participant walked to the start pole and faced the orientation pole. The virtual crowd appeared with their backs to the participant, "Begin" was played over headphones, and the crowd began walking forward (1.0 m/s). After 5s the walking direction of some or all virtual humans was perturbed by ±10˚ (right or left); the display continued for another 7s, then "End" was played. Test trials were preceded by two practice trials to familiarize the participant with walking in a virtual environment.

Data processing: The time series of head position in the horizontal (X-Y) plane were low-pass filtered (Matlab) to reduce tracker error and oscillations due to the step cycle, then time series of heading direction and walking speed were computed. The dependent measure was *final heading*, the average heading direction during the last two seconds of each trial. Left and right perturbation trials were collapsed by multiplying the left turn heading by -1. Statistical analyses were performed in Microsoft Excel and JASP. (See SM for details.)

Experiment 1: Range of interaction

Based on crowd data, the omniscient model holds that neighbor influence decays to zero at a fixed radius of 4-5m [8]. But it seems likely that interactions with visible neighbors can occur at greater distances. To test the range of interaction, we manipulated the initial distance (1.8, 3.0, 4.0, 6.0 or 8.0 m) of a single row of virtual humans (crowd size 2, 4, or 8), with no occlusion (Figure 2a). On each trial, their headings were all perturbed in same direction (±10˚), and participants were asked to walk with the group.

Results

We observed a very gradual decay in neighbor influence over a much longer distance (Figure 2b). Final heading decreased from a maximum at 1.8m (mean M=9.55˚) to just half that value at 8m (M=5.16˚), (F (4, 44)=14.93, p<0.001, η_G^2=0.290). Simple linear extrapolation suggests an interaction range of at least 15m (y = -0.722x + 10.8, r (14) = -0.95). Consistent with averaging of neighbors, there was no effect of crowd size on final heading (F (2, 22)=0.77, p=0.476, η_G^2=0.010) and no distance x size interaction (F (8,88)=0.83 p=0.575, η_G^2=0.033).

These results clearly show that the neighborhood of interaction does not have a fixed radius of 4-5m, for pedestrians may be influenced by neighbors at three times that distance – if they are fully visible. This finding suggests that there may be two decay processes at work: a gradual decay to visible neighbors, and a more rapid decay within a partially occluded crowd.

Experiment 2: The double-decay hypothesis

The second experiment tested this 'double-decay' hypothesis that there are two decay processes that depend on distance. We manipulated a virtual crowd of 12 neighbors, randomly positioned in three rows spaced 2m apart (Figure 3a). To check the decay rate to fully visible neighbors, we varied the distance of the near row (2, 4, or 6 m). To probe the decay rate within the partially occluded crowd, we selectively perturbed the near, middle, or far row, so all neighbors in one row turned in the same direction (±10˚). Farther neighbors were thus dynamically occluded by nearer neighbors.

Results

Final heading is plotted as a function of distance to the perturbed row in Figure 3b, where each curve represents a crowd distance. Two decay rates are immediately apparent. First, the heading response decreases with the distance of the crowd (F (2,18)=26.68, p<0.001, η_G^2=0.229). In particular, the response to perturbations of the near row (diamonds) decays gradually with distance (simple effect test, F (2,18)=48.46, p<0.001), replicating Experiment 1. Linear extrapolation suggests an interaction range of at least 9m (y = -.81x + 7.33, r (2) = -.99). The decay rate (slope) is slightly steeper and responses are weaker than in Experiment 1, due to the presence of unperturbed neighbors; together they are responsible for the shorter interaction range.

Second, in each curve the heading response decreases more rapidly within the crowd (F (2,18)=86.98, p<0.001, η_G^2=0.760), steeply from the near row to the middle row (t (9)=10.82, p<0.001, Cohen's d=3.42) and the far row (t (9)=11.95, p<0.001, Cohen's d=3.77). This finding implies that dynamic occlusion by near neighbors weakened responses to the middle and far rows, almost to the floor of zero.

The evidence thus reveals that the neighborhood of interaction results from two decay processes. We propose, first, that the gradual decay to visible neighbors follows from Euclid's Law of perspective, which states that the visual angle subtended by an object (or motion) with frontal extent x diminishes with distance z as $tan^{-1}(x/z)$. Note that this predicts a larger range of interaction than simple linear extrapolation. Second, the more rapid decay within the crowd is due to the additional effect of occlusion. These findings led us to formulate a new visual model.

Visual model

To build a visual model of collective motion from the bottom up, we begin with the visual coupling between a pedestrian and a single neighbor [38, 42, 43].

Heading control

Consider a pedestrian following a neighbor who turns left (Figure 4, top row). If the neighbor is directly ahead ($\beta = 0^\circ$ eccentricity, with positive angles to the right and negative angles to the left), this generates a leftward angular velocity (negative ψ) in the pedestrian's field of view (Figure 4a). Canceling ψ would cause the pedestrian to steer left and approximately match the neighbor's heading. On the other hand, if the neighbor is on the pedestrian's right ($\beta = 90^\circ$), this generates an optical expansion (θ) in the field of view (Figure 4b). In this case, canceling θ would also cause the pedestrian to steer left and match the neighbor's heading. Critically, optical velocities (ψ, θ) decrease with neighbor distance in accordance with Euclid's Law.

These two optical variables thus trade off as a function of the neighbor's eccentricity (Figure 4c). For a left turn, angular velocity ψ (blue curve) is a cosine function of eccentricity with a minimum (leftward motion) at $\beta = 0°$; whereas expansion rate θ (red curve) is a sine function with a minimum (contraction) at $\beta = -90°$ and a maximum (expansion) at $\beta = 90°$. For a right turn, these functions flip about the horizontal axis.

The visual coupling for controlling heading (ϕ) can thus be formalized as a second-order control law,

$$\ddot{\phi}_p = -c_1(\cos\beta_i)\psi_i + c_2(\sin\beta_i)\theta_i \quad (1)$$

in which pedestrian *p* steers (angular acceleration $\ddot{\phi}$) so as to cancel the combined angular velocity (ψ) and expansion rate (θ) of neighbor *i*. Their dependence on β acts as a filter so the pedestrian is only influenced by variables that specify a turn at that eccentricity. The free parameters ($c_1 = 14.38, c_2 = 59.71$) were fit to our previous data on pedestrian following [42] and held constant.

Speed control

The control of walking speed is complementary to the control of heading (Figure 4, bottom row). If a neighbor directly ahead ($\beta = 0°$) slows down, this generates an optical expansion (θ) in the pedestrian's field of view (Figure 4d). Canceling the expansion would cause the pedestrian to slow down and approximately match the neighbor's speed. But if a neighbor to the pedestrian's right ($\beta = 90°$) slows down, this generates a rightward angular velocity (positive ψ) in the field of view (Figure 4e); canceling it would also lead the pedestrian to slow to the neighbor's speed. These two optical variables again trade off as a function of eccentricity, but with the opposite sine and cosine functions (Figure 4f). If the neighbor speeds up, the curves flip about the horizontal axis.

The visual coupling for control of radial speed (r) is thus based on the same two optical variables as in Equation 1, but the sine and cosine functions are reversed:

$$r_p = -c_3(\sin\beta_i)\psi_i - c_4(\cos\beta_i)\theta_i \quad (2)$$

Pedestrian *p* thus linearly accelerates or decelerates (r) so as to cancel the combined angular velocity (ψ) and expansion rate (θ) of neighbor *i*. But now the pedestrian is only influenced by combinations that specify a speed change at a given eccentricity. The free parameters ($c_3 = 0.18, c_4 = 0.72$) were fit to our data on pedestrian following [42] and held fixed. To normalize for variation in neighbor size, the relative rate of expansion (θ/θ) can be substituted for expansion rate (θ) [43].

Collective motion

To formulate a model of collective motion, we substitute the visual control laws for local interactions (Equations 1 and 2) into a neighborhood function that averages the influences of multiple neighbors (Equation S1):

$$\ddot{\phi}_p = \frac{1}{n}\sum_{i=1}^{n} v_i\left[-c_1(\cos\beta_i)\psi_i + c_2(\sin\beta_i)\theta_i\right] \quad (3)$$

$$r_p = \frac{1}{n}\sum_{i=1}^{n} v_i\left[-c_3(\sin\beta_i)\psi_i - c_4(\cos\beta_i)\theta_i\right] \quad (4)$$

Pedestrian *p*'s heading and speed are thus controlled by canceling the mean angular velocity (ψ_i) and rate of expansion (θ_i) of all visible neighbors ($i = 1 \ldots n$), depending on their eccentricities (β_i). The field of view is centered on the heading direction, as people tend to face in the direction they're walking [45].

Partial occlusion is incorporated by weighting each neighbor in proportion to their visibility [46], ranging from $v_i = 0$ (fully occluded) to $v_i = 1$ (fully visible). Visibility is set to 0 if its value falls below a threshold ($v_t = 0.15$), thus *n* is the number of visible neighbors above threshold. Importantly, the occluded region behind a near neighbor grows with distance, so the visibility of far neighbors tends to decrease with their separation in depth from near neighbors (Figure 1b). Consequently, the range of interaction depends on the crowd's *opacity* [47] and is limited by the complete occlusion of far neighbors.

Basic properties of physics-inspired models fall out naturally from the visual model. First, canceling optical expansion yields collision avoidance without an explicit 'repulsion' force. Second, canceling optical contraction maintains group cohesion without an explicit 'attraction' force. Third, canceling the combined angular velocity and expansion/contraction generates collective motion without an explicit 'alignment' rule. Finally, the laws of optics account for the form of the neighborhood without an explicit decay function: Euclid's Law explains the gradual decay of influence to visible neighbors, and the added effect of occlusion explains the more rapid decay within a crowd.

Model simulations

We tested the visual model (Equations 3-4) by predicting human trajectories in virtual crowd experiments and real crowd data, and compared the results to our previous omniscient model (Equations S1-S4). We find that the visual model outperforms the omniscient model (and a motion model without occlusion, see SM) and generalizes to real crowds.

To simulate each experimental trial, the models were initialized with the participant's position, heading, and speed 2s before the perturbation. For the omniscient model, the input on each time step was the position, heading, and speed of all virtual neighbors in the HMD's 90˚ field of view on that trial. For the visual model, the input was the angular velocity, expansion rate, eccentricity, and visibility of the same neighbors, calculated from their positions on each time step. The output of both models was the position, heading, and speed of the simulated agent on the next time step, represented as time series for each trial. As a measure of model performance, we computed the mean position error (ME) or root mean squared error (RMSE) between each participant's mean time series in each condition and the corresponding mean time series for the model.

Simulating Experiment 2

First, we simulated the double-decay experiment. For the omniscient model, we added a gradual exponential term to the decay function (Equation S4), estimated from the data. Because crowd speed was not manipulated in this experiment, we used the participant's recorded walking speed as input to the omniscient model. Mean final heading for the two models is plotted in Figure 3b, together with the human results. Although both models are close to the 95% confidence intervals for the human data (shaded regions), the visual model (dotted curves) lies entirely within them.

Over the whole time series, the mean heading error for the visual model ($RMSE_V$=2.47°) was significantly smaller than that for the omniscient model ($RMSE_O$=3.45°) (t (9)=14.48, p<.001, Cohen's d=1.460); a Bayes Factor indicated decisive evidence for the alternative hypothesis (BF_{10}>>100). The mean position error for the visual model (ME_V=0.241m) was also smaller than that for the omniscient model (ME_O=0.309m) (t (9)=8.46, p<.001, Cohen's d=0.294), decisive evidence (BF_{10}>>100).

In sum, the visual model predicted the neighborhood better than the omniscient model because the decay rate is not a constant function of distance, but depends on the amount of occlusion. The visual model thus accounts for the form of the neighborhood without an explicit decay function.

Re-simulating Rio, Dachner & Warren [8]

As a further test of the models, we re-simulated Rio, et al.'s [8] Experiment 2, which perturbed heading or speed and manipulated the number and distance of perturbed neighbors (Figure 5a). The virtual crowd contained 5 neighbors in the near row (1.5m) and 7 in the far row (3.5m). On each trial, a subset of neighbors, predominantly in one row, either turned ±10˚ or changed speed by ±0.3 m/s (from 1.0 m/s). Mean final heading and mean final speed appear in Figure 5b,c (solid curves). Responses were larger when near neighbors were perturbed (blue) than when far neighbors were perturbed (red), indicating a decay of influence with distance.

Simulations of the visual model (dotted curves) and the original omniscient model (dashed curves; Equation S1-S3) appear in Figure 5b,c. Both are close to the human data (solid curves), falling within the 95% confidence intervals in nearly all conditions. Over time, the mean heading error was significantly smaller for the visual model ($RMSE_V$=1.97°, $RMSE_O$=2.08°), (t (9)=6.94, p<.001, Cohen's d=0.871, BF_{10}>100), although there was no difference for the mean speed error ($RMSE_V$=0.0627 m/s, $RMSE_O$=0.0640 m/s), (t (9)=1.15, p=0.281, Cohen's d=0.208; BF_{01}=1.91, anecdotal evidence for the null hypothesis), or the mean position error (ME_V=0.193m, ME_O=0.199m), (t (9)=1.112, p=0.295, Cohen's d=0.082; BF_{01}=1.96, anecdotal). Both models thus capture the human data quite well, although the visual model performs better on heading.

The comparatively good performance of the omniscient model in this experiment stems from the fact that the decay function was originally fit to human swarms that had nearest-neighbor distances (1-3m) and densities similar to those of the virtual crowd. However, this empirical decay term did not generalize to larger distances in the double-decay experiment, whereas the visual model did so.

In sum, the visual model accounts for Rio, et al.'s [8] experiment as well or better than the omniscient model. Whereas the latter assumes physical variables as input, the former is based on optical variables available to an embedded pedestrian: far neighbors exert less influence because they have lower optical velocities and are partially occluded by near neighbors.

Human swarm simulations

To test whether our findings for virtual crowds apply to real crowds, we simulated walking trajectories in previously recorded data on 'human swarms' [8]. We attempted to predict the trajectory of an individual pedestrian from the movements of their neighbors using both models.

Three different groups of participants (n=10, 16, 20) were instructed to walk about a large tracking area (14m x 20m), veering left and right while staying together as a group, for a total of twelve 2-min trials. Head-mounted markers were recorded with 16 motion-capture cameras (Qualisys) at 60 Hz, and time series of head position, heading and speed were computed as before. We identified thirty 10s segments of data in which ≥75% of the participants were continuously tracked. For each segment, we simulated a focal participant at the back of the group and treated the tracked neighbors as input. For the visual model, we computed optical variables from neighbor positions and velocities. The omniscient model used the original decay function (Equation S3).

Two segments of simulated swarm data appear in Figure 6. The heading time series (column b) for the focal participant (red) is more closely captured by the visual model (blue) than the omniscient model (green) in both segments, whereas the speed time series (column c) is better approximated by the omniscient model in Segment 1 (top) and the visual model in Segment 10 (bottom). Over all 30 segments, the mean heading error was significantly lower for the visual model ($RMSE_V$=15.0˚) than the omniscient model ($RMSE_O$=22.9˚) (t (29)=4.48, p<0.001, Cohen's d=0.806; BF_{10}>100, decisive evidence), as was the mean position error (ME_V=0.60m, ME_O=0.80m) (t (29)=2.21, p<0.05, Cohen's d=0.338; BF_{10}=1.60 anecdotal evidence). On the other hand, the mean speed error was significantly lower for the omniscient model ($RMSE_V$=0.224 m/s, $RMSE_O$=0.146 m/s) (t (29)=6.83 p<0.001, Cohen's d=1.198; BF_{10}>>100, decisive evidence); we consider this result in the Discussion.

The visual model thus accounts for individual heading and position in real crowd data better than the omniscient model, even though the latter's decay term was fit to a sample of the same data. We attribute this advantage largely to the effect of occlusion. Whereas the omniscient model approximates the decay with distance using a fixed exponential function, the visual model incorporates dynamic occlusion and is thus sensitive to changes in visibility over time.

Discussion

Nearly all microscopic models of collective motion in humans and animals attribute local interactions to hypothetical rules or forces and assume physical variables as input. In this article we developed a visual model of human 'flocking' based on a visual coupling with optical variables as input. In contrast to previous phenomenological models, the visual model explains basic properties of interaction as natural consequences of the laws of optics.

First, *social forces* and *rules of engagement* are reduced to optical variables that control an individual's heading and speed. In place of explicit 'repulsion' and 'attraction' forces, collision avoidance results from canceling optical expansion and group cohesion is maintained by canceling optical contraction. Instead of an explicit 'alignment' rule, collective motion emerges from canceling the combined expansion/contraction and angular velocity of neighbors. The visual coupling thus behaves functionally like a force or 'optical push' [48]

Second, the *neighborhood of interaction* is explained by the laws of optics, without an explicit distance term. The gradual decay to visible neighbors follows from Euclid's Law, the diminution of optical velocity with distance. The more rapid decay within a crowd follows from the added effect of visual occlusion, which grows with the separation in depth between near and far neighbors. Consequently, the neighborhood range and number of neighbors *n* are not determined by a fixed distance but vary with crowd opacity.

The visual model thus predicts that the effective neighborhood depends on crowd density, which we have confirmed in related experiments [49]. In dense human crowds (1-2m apart), complete opacity can occur by a range of 5m. Starlings appear to adjust flock density to maintain 'marginal opacity' such that individual birds can see through the entire flock [47]. The range of interaction might also be limited by a detection threshold for optical motion. However, adding a motion threshold in our simulations did not improve the fit to the data, perhaps because it was superseded by occlusion.

Virtually all physical models assume the principle of superposition, according to which the response to a group is the linear combination of independent responses to each neighbor. But superposition is invalidated by the facts of visual occlusion: because the influence of far neighbors depends on the positions of a near neighbors, the response to the former is not independent of the latter. While this may be computationally inconvenient, visual occlusion has large effects on local interactions (see SM) and should be incorporated into future models [30, 31].

Note that Euclid's Law predicts an asymmetry in the pedestrian's response. Given a neighbor at an initial distance ahead, if they slow down, their distance decreases, whereas if they speed up, their distance increases. Consequently, the rate of expansion is greater than the rate of contraction for the same speed change. This effect explains an asymmetric speed response we previously observed in pedestrian following [38, 43].

The visual model generally outperforms the omniscient model, although they were quite similar in our re-simulation of Rio,et al.'s [8] experiment. This is attributable to the fact that the omniscient model approximates the decay with distance using an exponential function that was fit to human swarms with a similar distance and density to the virtual crowd. However, this fixed decay term did not generalize to greater crowd distances in Experiment 2, whereas the visual model did so. The visual model thus not only explains the form of the neighborhood but generalizes to new conditions without re-parameterization.

We noted a limitation of the current visual model when we were simulating the human swarm data. In five additional segments, the front of the crowd executed a 180˚ hairpin turn and walked back toward the focal participant, generating rapid expansion in the field of view. Human participants kept walking forward, but the visual model responded by slowing down and backing up to cancel the optical expansion. Similar but less extreme responses to U-turns may explain the larger speed error for the visual model reported above. Clearly, the model needs to distinguish neighbors that should be followed from obstacles that should be avoided, which may be as straightforward as discriminating the front and back of other pedestrians.

Our findings suggest that characteristic patterns of collective motion in different species might result from reliance on different sensory variables. Humans cancel optical velocities, which yields collective motion despite variation in neighbor distance, density, and size. In contrast, holding the visual angles of neighbors at a particular value would yield fish schools with a preferred spatial scale, whereas

Formatted: Font: Italic

Deleted: is

Deleted: yields robust individual responses

Deleted: .

maintaining neighbors in particular visual directions would yield bird flocks with a preferred spatial structure.

In sum, we conclude that the local interactions underlying collective motion have a lawful basis in the visual coupling between neighbors. In recent multi-agent simulations, we have also shown that the visual model generates emergent collective motion, and a report is in preparation.

Data accessibility

Data and computer code are available from the Brown Digital Repository: https://doi.org/10.26300/r4c3-dq82 [50].

Acknowledgments

Thanks to Adam Kiefer, Stephane Bonneaud, Michael Fitzgerald, and the Sayles Swarm crew for their help during crowd data collection, and to Arturo Cardenas, Eugy Han and the VENLab team for their assistance in processing and analyzing the datafiles.

Funding statement

This research was supported by the National Institutes of Health, R01EY010923 and R01EY029745 to W.W., and T32 EY018080 to Brown University; the National Science Foundation, BCS-1431406 to W.W.; and Link Foundation Fellowships to G.D. and T.W.

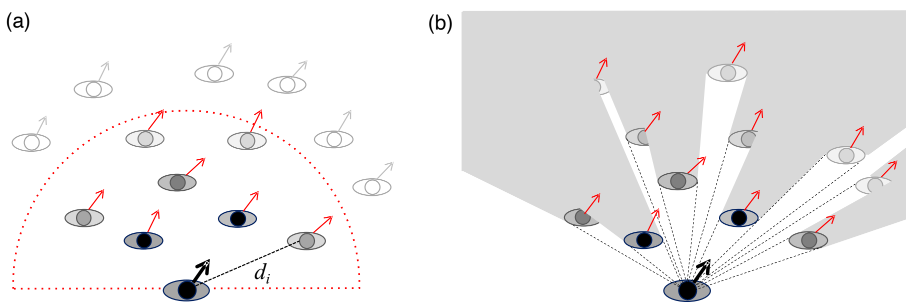

Figure 1. Omniscient and visual models of collective motion. (a) *Omniscient model:* a pedestrian (bottom) matches the average heading direction and speed of all neighbors in a 180˚ neighborhood. Neighbor weights (gray level) decay exponentially with distance d_i and go to zero at a fixed radius (dotted red curve). (b) *Visual model:* a pedestrian (bottom) cancels the average angular velocity and optical expansion of all visible neighbors. Neighbor influence decreases with distance due to Euclid's Law (gray level) and is proportional to neighbor visibility (shaded areas=occluded regions).

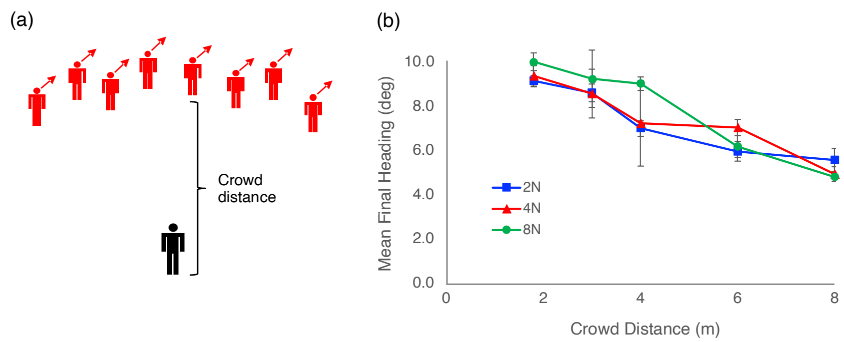

Figure 2. Experiment 1: Range of interaction, testing the decay with distance to fully visible neighbors. (a) Schematic of virtual crowd, illustrating a rightward heading perturbation (red). (b) Results: Mean final heading as a function of crowd distance, for each crowd size (curves). Error bars represent ±SEM.

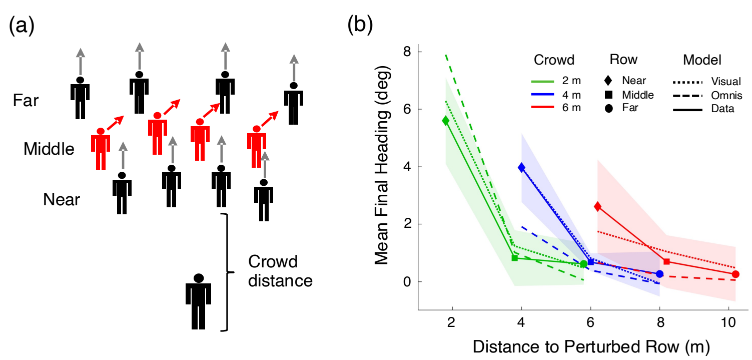

Figure 3. Experiment 2: Double-decay hypothesis. (a) Schematic of virtual crowd, illustrating a rightward heading perturbation of the middle row. (b) Results: Mean final heading as a function of distance to the perturbed row (symbols), for each crowd distance (curves). Solid curves represent human data, dotted curves the visual model, and dashed curves the omniscient model. Shaded regions represent 95% confidence intervals for the human data. (Models were not intended to reproduce gait oscillations, so their variable error is small and not represented.)

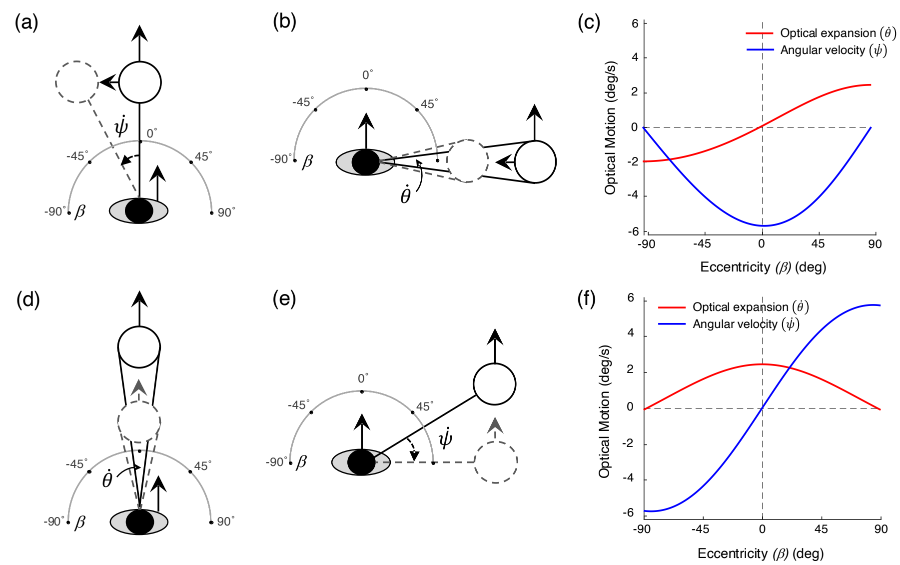

Figure 4. Visual information for control of heading (top) and speed (bottom). See text for explanation. Oval=pedestrian, open circle=neighbor, ψ=angular velocity, θ=expansion rate, β=eccentricity. Optical motions are computed for a neighbor with diameter=0.4m, distance=1m, relative speed=-1 m/s leftward (panel c) or -0.1 m/s backward (panel f).

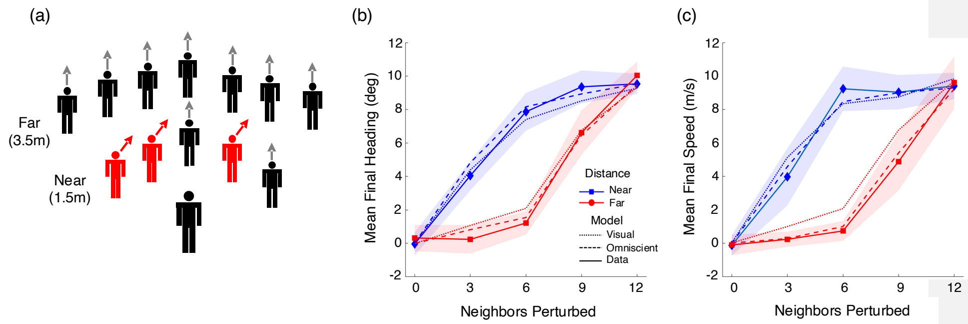

Figure 5. Rio, Dachner & Warren's [8] Experiment 2. (a) Schematic of virtual crowd (12 neighbors). A subset of neighbors (0-12) was perturbed (red), predominantly in the near or the far row. (b) Results for heading perturbation: Mean final heading as a function of the number of perturbed neighbors, for each row (curves). (c) Results for speed perturbation: Mean final speed as a function of same. Solid curves represent human data, dotted curves the visual model, dashed curves the omniscient model. Shaded regions represent 95% confidence intervals for the human data. [Modified from (8), with permission.]

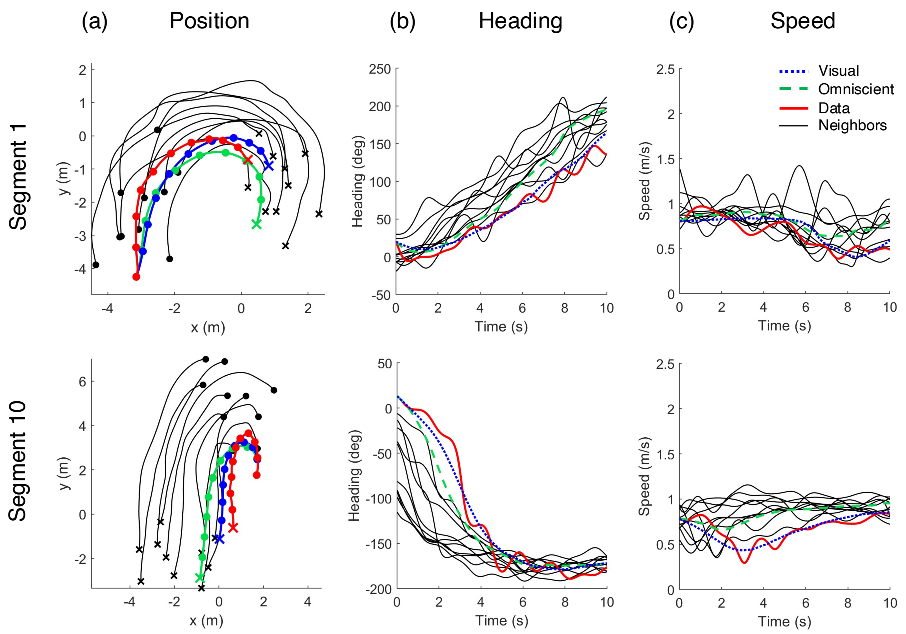

Figure 6. Sample segments (10s) from the human swarm, with focal participant (red) and simulations of visual model (blue dots) and omniscient model (green dashes). (a) Traces of position over time (Segment 1: ME_V = 0.379m, ME_O=0.818m; Segment 10: ME_V=0.275m, ME_O=1.389m). (b) Time series of heading (Segment 1: $RMSE_V$=10.67˚, $RMSE_O$=32.88˚; Segment 10: $RMSE_V$=11.81˚, $RMSE_O$=23.62˚). (c) Time series of speed (Segment 1: $RMSE_V$=0.187 m/s, $RMSE_O$=0.162 m/s; Segment 10: $RMSE_V$=0.157 m/s, $RMSE_O$ =0.178 m/s). Thin gray curves = neighbors; o = starting positions, x = final positions, dots at 1s intervals. Note that errors are higher than those in virtual crowds because they are computed on single trials rather than mean time series, and thus reflect gait oscillations and tracking errors.

References

1. Helbing D., Buzna L., Johansson A., Werner T. 2005 Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation Science* **39**(1), 1-24.
2. Ngai K.M., Burkle F.M., Hsu A., Hsu E.B. 2009 Human stampedes: a systematic review of historical and peer-reviewed sources. *Disaster medicine and public health preparedness* **3**(4), 191-195.
3. Couzin I.D., Krause J. 2003 Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior* **32**, 1-75.
4. Sumpter D.J.T. 2010 *Collective animal behavior*. Princeton, NJ, Princeton University Press.
5. Vicsek T., Zafeiris A. 2012 Collective motion. *Physics Reports* **517**, 71-140.
6. Giardina I. 2008 Collective behavior in animal groups: theoretical models and empirical studies. *HFSP Journal* **2**(4), 205-219.
7. Schellinck J., White T. 2011 A review of attraction and repulsion models of aggregation: Methods, findings and a discussion of model validation. *Ecological Modeling* **222**, 1897-1911.
8. Rio K.W., Dachner G.C., Warren W.H. 2018 Local interactions underlying collective motion in human crowds. *Proceedings of the Royal Society B* **285**(1878), 20180611, 20180611-20180619. (doi:doi:10.1098/rspb.2018.0611).
9. Couzin I.D., Krause J., James R., Ruxton G.D., Franks N.R. 2002 Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology* **218**, 1-11.
10. Huth A., Wissel C. 1992 The simulation of the movement of fish schools. *Journal of Theoretical Biology* **156**, 365-385.
11. Reynolds C.W. 1987 Flocks, herds, and schools: a distributed behavioral model. *Computer Graphics* **21**, 25-34.
12. Grégoire G., Chaté H., Tu Y. 2003 Moving and staying together without a leader. *Physica D: Nonlinear Phenomena* **181**(3-4), 157-170.
13. Cucker F., Smale S. 2007 Emergent behavior in flocks. *IEEE Transactions on automatic control* **52**(5), 852-862.
14. Vicsek T., Czirók A., Ben-Jacob E., Cohen I., Shochet O. 1995 Novel type of phase transition in a system of self-driven particles. *Physics Review Letters* **75**(6), 1226-1229.
15. Romanczuk P., Couzin I.D., Schimansky-Geier L. 2009 Collective motion due to individual escape and pursuit response. *Physical Review Letters* **102**(1), 010602.
16. Helbing D., Molnár P. 1995 Social force model of pedestrian dynamics. *Physical Review E* **51**, 4282-4286.
17. Chen X., Treiber M., Kanagaraj V., Li H. 2018 Social force models for pedestrian traffic–state of the art. *Transport Reviews* **38**(5), 625-653.
18. Chraibi M., Tordeux A., Schadschneider A., Seyfried A. 2018 Modelling of pedestrian and evacuation dynamics. In *Encyclopedia of complexity and systems science* (ed. Meyers R.A.), pp. 1-22. Berlin, Heidelberg, Springer Berlin Heidelberg.
19. Boltes M., Zhang J., Tordeux A., Schadschneider A., Seyfried A. 2018 Empirical Results of Pedestrian and Evacuation Dynamics. In *Encyclopedia of Complexity and Systems Science* (ed. Meyers R.A.), pp. 1-29. Berlin, Heidelberg, Springer Berlin Heidelberg.
20. Helbing D., Farkas I., Vicsek T. 2000 Simulating dynamical features of escape panic. *Nature* **407**, 487-490.
21. Helbing D., Molnár P., Farkas I., Bolay K. 2001 Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design* **28**, 361-383.
22. Pelechano N., Allbeck J.M., Badler N.I. 2007 Controlling individual agents in high-density crowd simulation. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Aug 3-7, San Diego, CA* (pp. 99-108. Aire-la-Ville, Switzerland, Eurographics Association.

23. Campanella M., Hoogendoorn S., Daamen W. 2009 Improving the Nomad microscopic walker model. *IFAC Proceedings Volumes* **42**(15), 12-18.
24. Weitz S., Blanco S., Fournier R., Gautrais J., Jost C., Theraulaz G. 2012 Modeling collective animal behavior with a cognitive perspective: A methodological framework. *PLoS ONE* **7**(6), e38588.
25. Gautrais J., Ginelli F., Fournier R., Blanco S., Soria M., Chaté H., Theraulaz G. 2012 Deciphering interactions in moving animal groups. *PLoS Comput Biology* **8**(9), e1002678.
26. Sumpter D.J.T., Mann R.P., Perna A. 2012 The modelling cycle for collective animal behaviour. *Interface Focus* **2**(6), 764-773.
27. Warren W.H., Fajen B.R. 2008 Behavioral dynamics of visually-guided locomotion. In *Coordination: Neural, behavioral, and social dynamics* (eds. Fuchs A., Jirsa V.), pp. 45-75. Heidelberg, Springer.
28. Moussaid M., Helbing D., Garnier S., Johansson A., Combe M., Theraulaz G. 2009 Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society B* **276**(1668), 2755-2762.
29. Pita D., Collignon B., Halloy J., Fernández-Juricic E. 2016 Collective behaviour in vertebrates: a sensory perspective. *Royal Society open science* **3**(11), 160377.
30. Strandburg-Peshkin A., Twomey C.R., Bode N.W.F., Kao A.B., Katz Y., Ioannou C.C., Rosenthal S.B., Torney C.J., Wu H.S., Levin S.A., et al. 2013 Visual sensory networks and effective information transfer in animal groups. *Current Biology* **23**(17), R709-R711.
31. Poel W., Winklmayr C., Romanczuk P. 2021 Spatial structure and information transfer in visual networks. *Frontiers in Physics: Social Physics* **9**, 716576, 716571-716514.
32. Gibson J.J. 1979 *The ecological approach to visual perception.* Boston, Houghton Mifflin.
33. Pepping G.J., Grealy M.L. 2007 Closing the gap: The scientific writings of David N. Lee. (Mahwah, NJ, Erlbaum.
34. Ondrej J., Pettré J., Olivier A.-H., Donikian S. 2010 A synthetic-vision based steering approach for crowd simulation. *ACM Transactions on Graphics* **29**(4), 123: 121-129.
35. Moussaïd M., Helbing D., Theraulaz G. 2011 How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences* **108**(17), 6884-6888.
36. Bastien R., Romanczuk P. 2020 A model of collective behavior based purely on vision. *Science Advances* **6**, eaay0792, 0791-0799.
37. Warren W.H. 2006 The dynamics of perception and action. *Psychological Review* **113**, 358-389.
38. Rio K.W., Rhea C., Warren W.H. 2014 Follow the leader: Visual control of speed in pedestrian following. *Journal of Vision* **14**(2), 4:1-16.
39. Dachner G., Warren W.H. 2014 Behavioral dynamics of heading alignment in pedestrian following. *Transportation Research Procedia* **2**, 69-76.
40. Wirth T.D., Warren W.H. 2021 Robust weighted averaging accounts for recruitment into collective motion in human crowds. *Frontiers in Applied Mathematics and Statistics: Dynamical Systems* **7**(73), 761445. (doi:10.3389/fams.2021.761445).
41. Warren W.H., Dachner G.C. 2018 Comparing simple-radius and doughnut models of collective crowd motion. *Journal of Vision* **18**(10), 1038.
42. Dachner G., Warren W.H. 2017 A vision-based model for the joint control of speed and heading in pedestrian following. *Journal of Vision* **17**(10), 716.
43. Bai J., Warren W.H. 2019 The relative rate of optical expansion controls speed in 1D pedestrian following. *Journal of Vision* **19**(10), 52.
44. Faul F., Erdfelder E., Lang A.-G., Buchner A. 2007 G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* **39**(2), 175-191.
45. Grasso R., Prevost P., Ivanenko Y.P., Berthoz A. 1998 Eye-head coordination for the steering of locomotion in humans: An anticipatory synergy. *Neuroscience letters* **253**(2), 115-118.

46. Dachner G., Warren W.H. 2019 Dynamic occlusion reduces the influence of neighbors in human crowds. In *International Conference on Perception and Action* (Groningen, The Netherlands.

47. Pearce D.J.G., Miller A.M., Rowlands G., Turner M.S. 2014 Role of projection in the control of bird flocks. *Proceedings of the National Academy of Sciences* **111**(29), 10422-10426.

48. Shaw R., Kinsella-Shaw J. 2007 Could optical 'pushes' be inertial forces? A geometro-dynamical hypothesis. *Ecological Psychology* **19**(3), 305-320.

49. Wirth T.D., Warren W.H. 2018 Metric vs. topological models of collective motion in human crowds. *Journal of Vision* **18**(10), 1036.

50. Dachner G.C., Wirth T.C., Richmond E., Warren W.H. 2022 Data and code from: The visual coupling between neighbors explains local interactions underlying human 'flocking'. Brown Digital Repository. doi.org/10.26300/r4c3-dq82 (