Supplemental Information for "Measurement-Induced Phase Transitions in Informational Active Matter"

Bryan VanSaders,[1,*] Michel Fruchart,[2] and Vincenzo Vitelli[1,3,4,†]

[1] *James Franck Institute, The University of Chicago, Chicago, Illinois 60637, USA*
[2] *Gulliver, ESPCI Paris, Université PSL, CNRS, Paris, France*
[3] *Department of Physics, The University of Chicago, Chicago, Illinois 60637, USA*
[4] *Kadanoff Center for Theoretical Physics, The University of Chicago, Chicago, Illinois 60637, USA*

Contents

S1. Molecular dynamics simulation parameters and methods 3
S1.1. Nematic flocking simulations (Figs. 1d-h&2l-m) 3
S1.2. Balloon simulations (Fig. 1i) 4
S1.3. Free energy sampling simulations (Fig. 2b-e) 4
S1.4. Information swimmer simulations (supplementary Fig. S4) 5
S1.5. Dropped grains simulations (Fig. 3a) 5
S1.6. Vibrated demon simulations (Fig. 3c) 6
S1.7. Self-propelled seeker simulations (Fig. 3b) 6
S1.8. Mixed passive-active systems (Fig. 3d-g) 6
S1.9. Pure demon gases with diameter functions of position (Fig. 4) 6
S1.10. Learning simulations (Fig. 5) 7
S1.11. Small vibrated demon simulations (Fig. 7) 8

S2. Kinetic energy conservation of the demon gas implementation 8

S3. Informational flocking and system size 8

S4. Markov model estimation for a single demon particle immersed in a passive gas 8

S5. The demon gas velocity distribution function 11
S5.1. Kinetic theory of hard disks 11
S5.2. Hard disk collisions with non-constant diameter 12
S5.3. The demon gas collision invariant 13
S5.4. Demonstration of f^d as the steady-state solution to the variable-diameter collision operator 16
S5.5. Invertibility of the matrix $\mathcal{A}$ 17
S5.6. Explicit form of $\mathcal{A}$ for select $D(\boldsymbol{v})$ 17
S5.6.1. Step function 17
S5.6.2. Anisotropic Gaussian diameter function 19

S6. Kinetic theory of the nematic flocking model 20
S6.1. Diameter function defined in the fluid flow frame 20
S6.1.1. Growth of pressure asymmetry 21
S6.1.2. Coupling of ψ 23
S6.2. Analytical constants for nematic flocking diameter function 25
S6.3. Inversion of $\mathcal{A}$ 26
S6.4. Conserved constants 26
S6.5. Pressure anisotropy 27

S7. Hydrodynamic equations for the nematic flocking demon gas model 27
S7.1. Integral balance equation 27
S7.2. Hydrodynamic equations 27

*Current affiliation: Department of Physics, Drexel University, Philadelphia, PA 19104, USA
†Electronic address: vitelli@uchicago.edu

S8. Discussion of experimental realization in a macroscopic robotic system 28

S9. Description of Supplemental Videos 29
S9.1. Supplemental video V1 29
S9.2. Supplemental video V2 29

References 29

In §S1 we describe the various simulation protocols used throughout the study, including key parameters. In §S2 we demonstrate that the demon model conserves kinetic energy via a numerical simulation. In §S3 we provide preliminary data on the scaling of orientational alignment with system size in the informational flocking model. In §S4 we detail the kinetic arguments underpinning our estimate of the swimming efficiency of a single demon particle immersed in a passive isothermal gas. In §S5 we derive the solution to Boltzmann's equation for disks with diameters that are functions of particle velocity. Here, emphasis is placed on diameter functions of laboratory-frame velocity. In §S6 the kinetic theory of a demon gas with diameter function of velocity relative to bulk flow is derived. In §S7, hydrodynamic relations for a demon gas with a flow-frame diameter function are derived. Experimental realization of a variable-diameter gas is discussed in section §S8. Section §S9 describes the accompanying supplemental videos.

S1. MOLECULAR DYNAMICS SIMULATION PARAMETERS AND METHODS

Molecular Dynamics (MD) simulations of the demon gas were performed with the open-source software package `HOOMD-Blue` (v2.9.3) [1]. All particles interact through a shifted Weeks-Chandler-Andersen potential (sWCA) [2], where the origin is shifted so that the radius of the particle is the potential's zero isoenergy surface. This potential (implemented in the function `hoomd.md.pair.slj`) takes the form:

$$V_{\mathrm{sWCA}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r-\Delta}\right)^{12} - \left(\frac{\sigma}{r-\Delta}\right)^{6}\right] + \varepsilon$$

for $r < (r_{\mathrm{cut}} + \Delta)$ and $V_{\mathrm{sWCA}} = 0$ otherwise, where $\Delta = (D_i + D_j)/2$ and D_i, D_j are the diameters of the interacting pair. The radial shifting was chosen with $\sigma = 0.5$ so that the minimum (located at $\sigma 2^{1/6}$ in the unshifted case) is at a distance of $D_0 = 2^{1/6}$. This sets the value of r_{cut} in the interaction potential to $\sigma 2^{1/6}$. The depth of the attractive well was set to $\varepsilon = 1$. Particle diameters were modified by a custom updater in accordance with a chosen diameter function, which required the following actions be taken on each demon particle every fixed interval of timesteps (i.e. the measurement interval t_m):

1. Find neighbors (of any type, including passive particles in the environment) within a distance of 1.5 times the maximum interaction cutoff distance. Find the minimum distance between the particle center and the boundary of a neighbor particle ($r^n_{\mathrm{min}} = r_{ij} - D_j/2$).
2. Evaluate the diameter function $D(\boldsymbol{x}, \boldsymbol{v})$ which determines the desired new diameter of the particle (D_{new})
3. Change particle diameter. Three cases are possible:
   (a) $D_{\mathrm{new}} > D_{\mathrm{old}}$ and $D_{\mathrm{new}}/2 > r^n_{\mathrm{min}}$. Expansion would cause particle overlap. The diameter change is discarded.
   (b) $D_{\mathrm{new}} < D_{\mathrm{old}}$ and $D_{\mathrm{old}}/2 > r^n_{\mathrm{min}}$. This particle is currently interacting with a neighbor. Reducing its diameter would cause a potential energy change. The diameter change is discarded.
   (c) $D_{\mathrm{old}}/2 < r^n_{\mathrm{min}}$ and $D_{\mathrm{new}}/2 < r^n_{\mathrm{min}}$. No overlaps will be added or removed by the diameter change, and it proceeds.

This sequence of steps is done in randomized serial order for all demon particles so as to avoid a race condition which would change the value of r^n_{min}.

Description of methods and parameters particular to various studies follow. All values listed here are given in simulation units. Individual simulation timesteps were $dt = 0.005$. Unless noted explicitly, system thermal energy was held at $k_bT = 1$, and particle mass was fixed at $m = 1$. Diameter resizing was performed at an interval of $t_m = 0.5$ in simulation, which for two-dimensional simulations shown in the main text corresponds to a value of $t_m/\tau_{\mathrm{col}} = 0.06$. Nosé-Hoover simulations utilize a Verlet integration scheme to integrate dynamics governed by the extended Hamiltonian $\mathbb{H} = \sum_i \frac{\mathbf{p_i^2}}{2ms^2} + \frac{1}{2}\sum_{i,j,i\neq j} U(\boldsymbol{r}_i - \boldsymbol{r}_j) + \frac{p_s^2}{2\tau} + gk_bT\ln(s)$. Here U includes all (time-varying) interactions, g is the number of momentum degrees of freedom, (s, p_s) are extended coordinates associated with the thermostat, and τ is a coupling constant (in this study $\tau = 0.1$). All data shown with the exception of Fig. 6 a-c, 5 d,f-h, and 7 (which employ Langevin dynamics) are integrated this way.

Langevin integration proceeds according to the equation of motion $m\dot{\boldsymbol{v}} = \boldsymbol{F}_{\mathrm{C}} - \gamma\boldsymbol{v} + \boldsymbol{F}_{\mathrm{R}}$, where $\boldsymbol{F}_{\mathrm{C}}$ represents all external and (time-varying) interaction forces, $\boldsymbol{F}_{\mathrm{R}}$ is a random force with average magnitude $2dk_bT\gamma/\delta t$ (for dimensions d and timestep δt) and zero average. To approximate the effect of vibration by a shaker table, in Fig. 6 a-c, 5 d, and 7 the vertical components of the random force $\boldsymbol{F}_{\mathrm{R}}$ are drawn from a uniform distribution ten times wider than the horizontal components.

S1.1. Nematic flocking simulations (Figs. 1d-h&2l-m)

To generate spatially resolved maps of the orientation ψ, 8,000 particles were simulated at a density of $\rho = 0.01$. The difference in diameter was $\Delta D = 3$, and the resizing updater operated at an interval of $t_m = 0.5$. Particles were thermalized (at $kT = 1$ unless otherwise noted) for $t_{\mathrm{eq}} = 1000$ without diameter resizing to homogenize the gas at the chosen density. After diameter resizing was enabled, individual particle orientations were spatially averaged. Every $\Delta t = 3$ interval a sample of orientations was collected (for this system, collision time is ~ 15), and 5 such samples were averaged to produce a single map in Fig. 1g. For information transfer between particles, a maximum

communication radius of ~ 112.4 was used, which at this density ensured approximately 500 neighbors were involved in the local computation of the virial pressure tensor. For measurements of $\langle Q \rangle$ (Fig. 2l-m), the same parameters were used, except that system size was reduced to $4,000$ particles. System pressure and velocity distribution functions were averaged by collecting 1000 samples spaced $\Delta t = 5$ apart.

S1.2. Balloon simulations (Fig. 1i)

Simulations of demon gases in flexible containers were performed using constant-energy Verlet integration of the gas and Langevin integration of the flexible container ($k_b T = 1$, damping parameter $\gamma = 2$). The container consisted of two close-packed layers of beads with diameter D_L, bonded to nearest neighbors by harmonic springs of stiffness $k = 100$. Maximum diameter change was $\Delta D = 2$. 100 samples spaced $\Delta t = 100$ were collected.

S1.3. Free energy sampling simulations (Fig. 2b-e)

The entropy of demon gases was conducted by sampling their velocity distribution functions at fixed time intervals relative to the diameter resizing event. For data in Fig. 2b, 300 distributions were sampled spread over a time interval of $3t_m$. These distributions were collected by averaging samples collected over $2,000 \times 3t_m$ cycles of diameter resizing (meaning each distribution represents $2,000$ independent samples). These systems were simulated at a density of $\rho = 0.02$, and therefore had a communication cutoff of ~ 79.5. Data in Fig.2c-e was collected identically, except that only 30 samples distributed over $3t_m$ were collected.

After measurement the entropy was fit to a simple functional form to extract estimates of the amount of free energy injected per measurement event. We subtract off the MB entropy as a baseline, and model the effect of measurement as a unidirectional coupling between the information-carrying degree of freedom (the diameters), and the velocity degrees of freedom. We assume that the free energy of information carrying degree of freedom ($\mathcal{F}_{\text{ic}}$) receives impulses of strength $\mathcal{F}_m$ due to measurement which add to any elevated free energy remaining from the previous measurement ($\Delta\mathcal{F}^*$), and relaxes exponentially with a single timescale τ,

$$\frac{d}{dt}\mathcal{F}_{\text{ic}} = -\frac{1}{\tau}\mathcal{F}_{\text{ic}} \tag{S1}$$

$$\mathcal{F}_{\text{ic}} = (\mathcal{F}_m + \Delta\mathcal{F}^*)\, e^{-\frac{t}{\tau}} \tag{S2}$$

where we take $t = 0$ as the time of measurement. Relaxation of $\mathcal{F}_{\text{ic}}$ is assumed to occur by coupling to the velocity degrees of freedom ($\mathcal{F}$). The velocity degrees of freedom are also assumed to exponentially relax to equilibrium (with the same timescale τ),

$$\frac{d}{dt}\mathcal{F} = -\frac{d}{dt}\mathcal{F}_{\text{ic}} - \frac{1}{\tau}\mathcal{F} \tag{S3}$$

$$\mathcal{F} = \left(\Delta\mathcal{F}^* + (\mathcal{F}_m + \Delta\mathcal{F}^*)\,\frac{t}{\tau}\right) e^{-\frac{t}{\tau}} \tag{S4}$$

To fit these parameters, several simulated measurement intervals t_m were observed and the resulting free energy curves are collapsed onto the range $[0, t_m]$. The parameters $\mathcal{F}_m$, $\Delta\mathcal{F}^*$, and τ can be seen in extended Fig. S1a-b, and raw free energy curves along with their fit lines can be seen in panels d-l. Bootstrap fitting error is shown in Fig. S1a-c.

To estimate the information stored in the diameter state of the particles, $1,000$ snapshots of particle diameters were saved (interval equal to the measurement interval t_m). At each pair of snapshots transitions from $D_L \to D_L$, $D_L \to D_S$, $D_S \to D_L$, and $D_S \to D_S$ were counted and normalized by the likelihood of each diameter state to obtain the mean transition matrix ($\boldsymbol{T}$). The entropy per step of this Markov system was then computed as reported in the main text,

$$I = -\sum_{i,j} \mu_i T_{ij} \ln T_{ij}, \tag{S5}$$

where $\boldsymbol{\mu}$ is the leading eigenvector of $\boldsymbol{T}$.

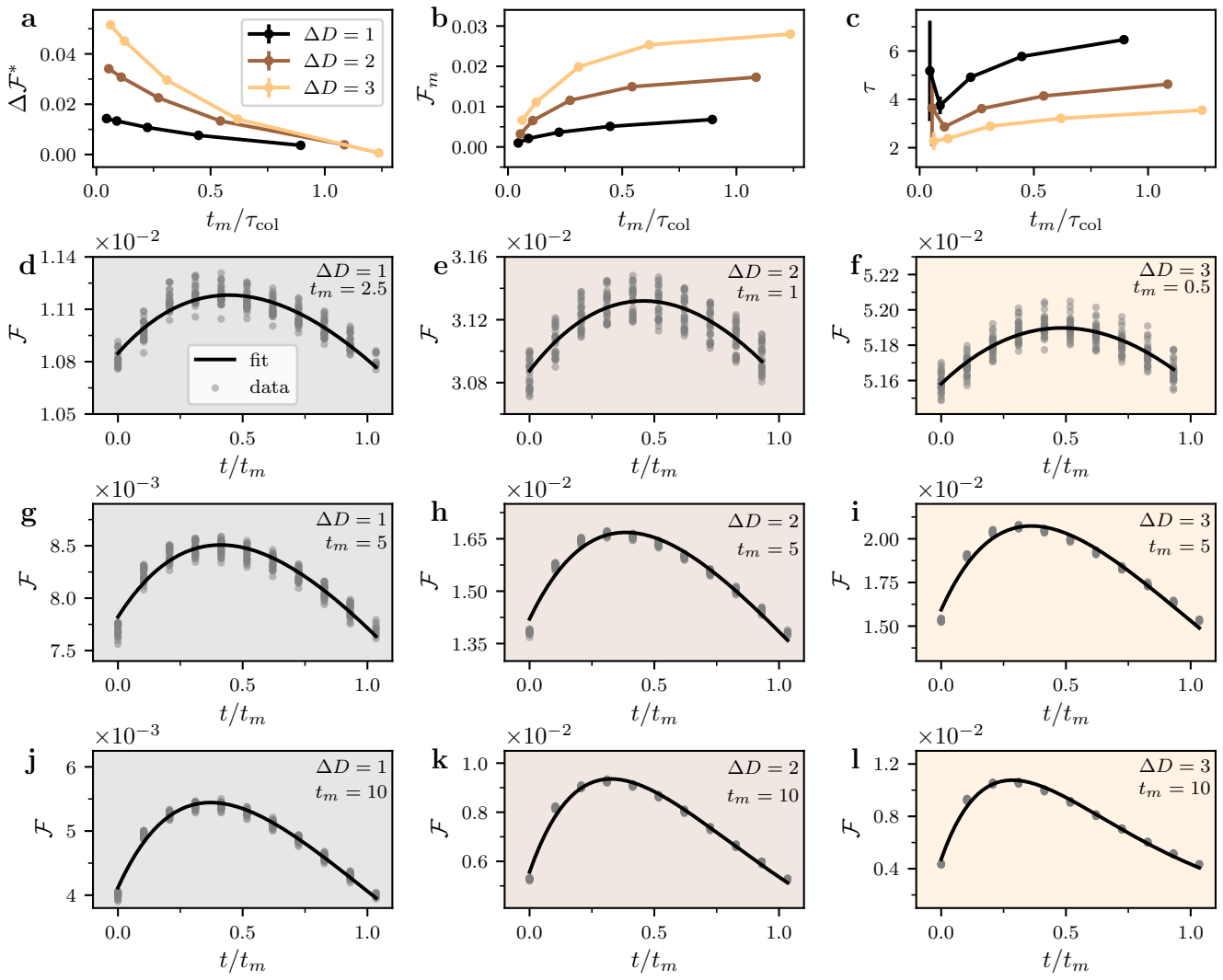


Fig. S1: **Parameter fitting of the free energy of the demon gas**. **a-c** Estimated parameters of the free energy cycles of demon gases with various measurement intervals (t_m) and diameter differences (ΔD). These parameters summarize the data presented in **d-l**. **a**. Value of free energy at the start and end of a measurement cycle. **b**. Free energy impulse delivered to the demon gas from measurement and diameter change. **c**. Relaxation parameter for free energy dynamics. **d-l**. Free energy data, fit by the functional form presented in methods to extract the parameters in **a-c**. Rows increase measurement time top to bottom, columns increase ΔD left to right. Note that the smallest measurement time that can be reliably fit is a function of ΔD.

S1.4. Information swimmer simulations (supplementary Fig. S4)

Simulations of demon particles propelling through a background gas of passive, fixed-diameter particles were performed in periodic domains with $3,600$ passive particles (diameter D_L). A small fraction of demon gas particles (400) were interspersed, and had a diameter difference of $\Delta D = 1$. Systems were equilibrated for $t_{\mathrm{eq}} = 5000$ time units and $2,000$ samples spaced 25 time units were collected.

S1.5. Dropped grains simulations (Fig. 3a)

Simulations of demon particles dropped into a container were simulated with a Langevin thermostat ($kT = 0$, linear damping term $\gamma = 0.01$). A total of $30,000$ particles were simulated, split between two species (25% red, 75% blue). Both species updated their diameters, with opposite behavior between A and B particles. The diameter difference was set to $\Delta D = 3$, and domain size was set to $L = 390 D_S$. Particles were initialized randomly above a constriction,

then fell freely under a vertical force until striking the bottom of the domain (distance $300D_S$). The constriction, bottom, and side wall surfaces of the container were implemented as described in eq. S7. The vertical force had a magnitude of $f_z = 10\gamma$. At its narrowest point, the constriction had an opening size of $10D_L$. The simulation was run for $1,000$ time units.

Particles changed their diameter according to

$$D(\boldsymbol{x}, \boldsymbol{v}) = D_S + \Delta D\mathcal{H}(-\boldsymbol{v} \cdot \boldsymbol{\xi}(\boldsymbol{x})), \tag{S6}$$

where $\boldsymbol{\xi}(\boldsymbol{x})$ was determined by computing the local gradient of 30 Gaussian functions ($\sigma = L_{\text{box}}/10\sqrt{2}$), which were distributed into the desired shape (i.e. the curve of the letter "S"). Gaussian locations were determined from an image of the desired character. The image was binarized, skeletonized, and converted to a set of points corresponding to bright pixels in the processed image. B particles behaved the same as red, except utilizing $-\boldsymbol{\xi}(\boldsymbol{x})$.

S1.6. Vibrated demon simulations (Fig. 3c)

Simulations of continuously vibrated demon particles were performed with a modified Langevin integrator (unity mean kinetic energy, vertical forces emphasized by a factor 10). Linear damping was set to $\gamma = 0.01$, and a floor and sidewalls were simulated as frictionless planes that interacted via a potential with particles:

$$V_{\text{wall}}(r_n) = 4\varepsilon\left[\left(\frac{\sigma}{r_n}\right)^{12} - \left(\frac{\sigma}{r_n}\right)^{6}\right] + \varepsilon \tag{S7}$$

for $r_n < r_{\text{cut}}$ and $V_{\text{wall}} = 0$ otherwise, where $\varepsilon = 1$, $\sigma = 1$, $r_{\text{cut}} = 2^{1/6}$, and r_n is the point-plane distance $r_n = \hat{n}\cdot\boldsymbol{x}+\boldsymbol{x_0}$ (plane normals $\hat{n}$ are chosen so that r_n is always positive). A vertical force of magnitude $f_z = 20\gamma$ kept particles confined to the vicinity of the floor. A total of $30,000$ particles (25% A and 75% B) were simulated for $t_{\text{frame}} = 600$ time units per image displayed, with 25% of that time devoted to erasing the previously shown image by driving type A particles to the domain edge and type B towards the domain center. Director fields $\boldsymbol{\xi}(\boldsymbol{x})$ for each image were found by computing the gradients of Gaussian functions distributed over the text to be displayed.

S1.7. Self-propelled seeker simulations (Fig. 3b)

Self-propelled particle simulations were performed in every way identically to vibrated demon particle simulations, except that particles have fixed diameters (the mean of the two diameter states allowed to demon particles) and self-propel with a fixed-magnitude force ($f_w = 1.5$). The direction of self-propulsion exactly follows $\boldsymbol{\xi}(\boldsymbol{x})$ (A type) or $-\boldsymbol{\xi}(\boldsymbol{x})$ (B type) as described above. To generate similar quality patterns as demon grains, a simulation time of $4,000$ was required.

S1.8. Mixed passive-active systems (Fig. 3d-g)

Simulations of mixed demon-passive systems were conducted in periodic domains with $3,600$ fixed-diameter particles (diameter D_S) and 400 demon particles. Isothermal integration was used (see above). For simulations which varied temperature, (total) density was held constant at $\rho = 0.01$. For simulations which varied density, temperature was held constant at $kT = 1$. For demon gases which moved in one direction, a constant $\boldsymbol{\xi} = [1, 0, 0]$ was used. For demon gases which moved to the center of the domain, $\boldsymbol{\xi}(\boldsymbol{x}) = -\boldsymbol{x}$ was used.

S1.9. Pure demon gases with diameter functions of position (Fig. 4)

Simulations of pure demon gases with $\boldsymbol{\xi}(\boldsymbol{x}) = -\boldsymbol{x}$ were simulated isothermally at $kT = 1$. A totall of $4,000$ particles were simulated. A total of 100 position histograms were collected (with 100 samples spaced 10 time units apart in each histogram). Diameter difference was set to $\Delta D = 1$.

S1.10. Learning simulations (Fig. 5)

Standard policy optimization algorithms [3, 4] were implemented to allow demon particle learning during MD simulations. Particles were rewarded (r) by their location within a simulated domain, effectively specifying a minimum-time locomotion task. Artificial feed-forward neural networks were used to encode two functions, the actor and critic. These functions are maps between a 'state' vector (in this case $\mathcal{S} = [x, y, v_x, v_y]$) and other quantities of interest. The critic maps states to expected future rewards (r):

$$\mathcal{C}(\mathcal{S}_t) \to \mathbb{E}\left[\int_\tau^\infty \gamma_r^{t-\tau} r_t dt\right] \tag{S8}$$

where $\gamma_r < 1$ is the discount factor that bounds predictions to a finite time horizon. This expectation is over all possible future trajectories of an agent beginning at state $\mathcal{S}_t$. The actor simply maps states to actions ($P(\mathcal{S}_t) \to a_t$). The critic network is trained using the temporal difference (TD) target:

$$C(\mathcal{S}_t) \leftarrow C(\mathcal{S}_t) + \alpha(r_t + \gamma_r C(\mathcal{S}_{(t+\Delta t)}) - C(\mathcal{S}_t)) \tag{S9}$$

where α is the learning rate and Δt is the time interval between updates. Note that we must wait until time $t + \Delta t$ to apply learning updates to functions evaluated at time t. The TD target is saved for use as an advantage estimator when training the actor network [5]. The actor function is updated in batches using the target:

$$P_\theta(\mathcal{S}_t) \leftarrow P_\theta(\mathcal{S}_t) + \alpha \nabla_\theta \ln P_\theta(\mathcal{S}_t) A_t \tag{S10}$$

where θ are the parameters of the network and A_t is the TD advantage estimator. To improve stability, the actor's evolution during a single update is clipped to a neighborhood in action probability space [4]. In these simulations, the demon gas interacts with a bath of passive, fixed-diameter particles. At the shortest timescales ($t_{\text{act}} = 1$), the demon particles choose new diameters as a function of their experienced state $\mathcal{S}$ and actor network. At a longer timescale ($t_{\text{learn}} = 5$), the parameters of the actor and critic networks are updated. State transition tuples are accumulated in a memory buffer, and the critic and actor networks are trained by stochastic descent on these samples. All particles contribute experience to the same networks, and since the demon particles comprise only 10% of simulations, the effect is equivalent to parallel simulations of single demon gas particles in passive baths with pooled network parameters (centralized learning, decentralized actions). In all learning cases a simple feed-forward neural network architecture was used with 3 hidden layers of size 10. Neurons utilized the swish activation function [6].

A variety of simulation types were used as environments for learning. In all cases, $3,600$ fixed-diameter particles and 400 demon particles were simulated. A total of 30 episodes of $1,000$ model updates were simulated. Particles were rewarded by a monotonic function of distance from the center of the domain, $r = e^{-(|\boldsymbol{x}|/3L)^2} - 1$. The critic network was updated with a learning parameter of $\alpha_{\text{C}} = 0.01$, and the actor updated with $\alpha_{\text{A}} = 0.05$. Learning rates were scaled from their initial value to 10% over the course of the simulation to facilitate settling into local minima. L2 regularization was applied to the values of all networks, for weights exceeding 2.

In the isothermal setting, a velocity Verlet integrator with Nosé-Hoover thermostat was used (temperature $k_bT = 1$) in a two-dimensional periodic simulation domain. In the vibrated granular bed model, a modified Langevin integrator was used in an xy-periodic simulation domain. Stochastic forces consistent with the fluctuation-dissipation relation ($\gamma = 0.1$, unity mean kinetic energy) were applied to particles, albeit with the z-component of these forces emphasized over in-plane components by a factor of 10. A floor was simulated as a frictionless plane according to eq. S7.

In the magnetic gas environment, isothermal conditions were applied to a 2D gas in a periodic domain. A Lorentz force with magnitude $B = 0.5$ was applied to induce curved particle trajectories. In the sheared gas environment a body force $F_{y+} = (0.1, 0)$ was applied to particles with y-position greater than zero and $F_{y-} = (-0.1, 0)$ otherwise. Sheared particles evolved under standard (isotropic) Langevin integration ($\gamma = 0.1$, $k_bT = 1$) in a periodic 2D domain. In the active setting, fixed-diameter particles were rendered 'active' by a constant-magnitude force $F_a = 5\gamma$ pointing along their director, which diffused with a rotational constant equal to the Langevin damping constant γ.

Simulations of odd gases utilized a transverse reciprocal force between particles of the form

$$F_{\text{Lub}}(r) = \varepsilon \log\left(\frac{r_{\text{cut}}}{r - \Delta}\right) \tag{S11}$$

for $r < (r_{\text{cut}} + \Delta)$ and $F_{\text{Lub}} = 0$ otherwise, where $\Delta = (D_i + D_j)/2$ and D_i, D_j are the diameters of the interacting pair. Particles were integrated under Langevin dynamics, with damping parameter $\gamma = 0.1$. The strength of the transverse force was chosen to produce a gas with $k_bT_{\text{eff}} = 1$ in the steady state.

S1.11. Small vibrated demon simulations (Fig. 7)

A $N = 125$ particle system was simulated using the same methodology as described above for continuously agitated systems. Simulation size was set to $L = 40$, $D_S = 1$, and $\Delta D = 2$. A vertical force $f_z = (0, 0, -1)$ held particles down and the mixture was an even split of particles with $\boldsymbol{\xi} = (-1, 0, 0)$ and $\boldsymbol{\xi} = (1, 0, 0)$. Histograms of particle position (21 evenly spaced bins) were collected at an interval of $t = 1.8$ for 450 total time units (250 samples).

S2. KINETIC ENERGY CONSERVATION OF THE DEMON GAS IMPLEMENTATION

By construction, the demon gas model microscopically conserves linear momentum and energy with the consequence that an initially fixed-diameter hard gas will preserve its total kinetic energy if diameter resizing is suddenly (or gradually) 'turned on'. To demonstrate this, as well as the correct implementation of the numerical scheme, we thermalize a fixed-diameter gas of 4,000 particles by Nosé-Hoover coupling to a reservoir, then uncouple the gas and integrate the dynamics forward in time (still with fixed diameters) to demonstrate the baseline energy conservation performance of the molecular dynamics implementation (Fig. S2a). We then 'turn on' diameter resizing and continue simulating (Fig. S2b). Comparison of the total energy drift in both cases demonstrates that diameter resizing has no measurable effect on the total kinetic energy of the gas.

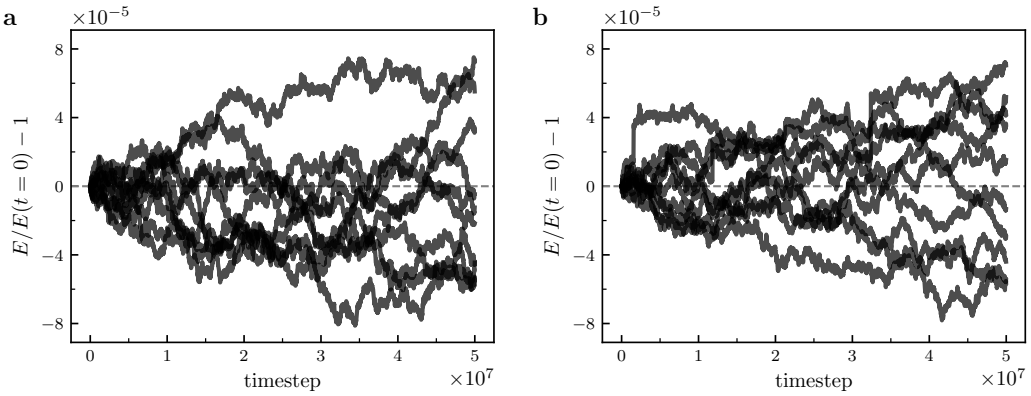


Fig. S2: **Demon gases conserve energy**. Simulations of gases prepared by thermalization and then integrated under a constant energy ensemble (no coupling to a thermal reservoir). Ten replicate simulations of 4,000 particles with timestep $t = 1 \times 10^{-3}$ are shown. **a**. Drift in total energy relative to initial energy for a fixed diameter gas. **b**. Drift in total energy for the same gases with variable diameters. Energy is shown relative to its value just before the start of variable diameter dynamics. Despite driving the system in to a non-equilibrium state, diameter changes do not effect the total energy state of the simulation.

S3. INFORMATIONAL FLOCKING AND SYSTEM SIZE

Proof of the emergence of true global orientational order in 2D flocking models [7–9] requires a fluctuation theory and accompanying large-scale numerical scaling analysis. While we defer this substantial undertaking to a future study, we have conducted a preliminary scaling analysis of the degree of orientational order present in flocking simulations of different scales (but otherwise identical parameters). Figure S3 shows how the nematic order parameter $\mathcal{Q}$ decreases as the size of the flock is increased. The ultimate fate of $\mathcal{Q}$ as system size is further increased remains an open question.

S4. MARKOV MODEL ESTIMATION FOR A SINGLE DEMON PARTICLE IMMERSED IN A PASSIVE GAS

To compliment our analysis of the rate of information acquisition and departure from equilibrium in a homogeneous system of demon particles (Fig. 2b-e), we also studied the simplified case of a single demon particle immersed in a isothermal gas of fixed-diameter particles. In this situation, scattering between the demon particle and the passive gas

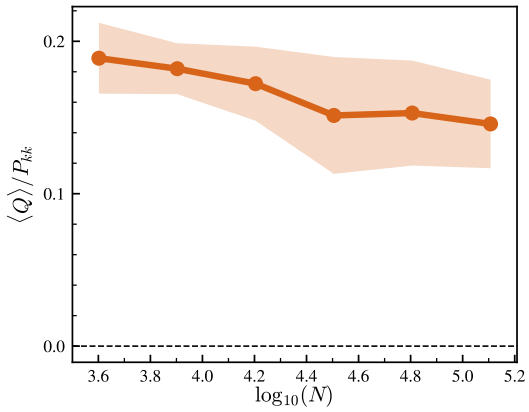


Fig. S3: **The scaling of $\mathcal{Q}$ with system size**. Mean steady state magnitude of the norm of the traceless part of the pressure tensor ($\mathcal{Q}$) plotted against system size. All simulations use a constant $\mathcal{Q}_0/P_{kk} = 0.191$, $k_bT = 1$, and density $\rho = 0.01$. Orientational fluctuations diminish the mean nematic order of the flock for larger systems.

allows for momentum transfer and spontaneous drift of the demon particle (an effect not observable in homogeneous demon systems with diameter functions of velocity only, due to linear momentum conservation). The mean drift velocity achieved by the particle is a result of a balance between collisional drag against the passive thermal gas and biased scattering due to size changes.

The information gained by a demon particle immersed in a passive, isothermal gas can be estimated from a Markov chain model of the diameter changing process. The model is constructed by first considering the diameter transitions that depend upon gas collisions and crowding, shown schematically in extended Fig. S4a. The demon particle undergoes random collision with the passive surrounding gas, with probability P_{col}, which then randomizes its velocity into the forward or reverse velocity half spaces with probabilities P_+ and P_-, respectively. If the demon particle's velocity is such that a diameter expansion is attempted, no nearby particles must obstruct the expansion. The probability that this requirement is met is denoted by $P_<$, i.e. if an overlap is created the work exerted by the demon particle must be less than the bound, which we set to zero here (as it is throughout the main text). The transition diagram can be condensed into a binary model of the form shown in extended Fig. S4b.

We utilized several elementary kinetic approximations to define the terms P_+, P_-, P_{col}, and $P_<$. Velocities are assumed to be randomly re-distributed by collisions on average, with $P_+ = P_- = 1/2$. We assume a collision timescale for the gas:

$$\tau^i_{\text{col}} = \frac{l^i_{\text{mfp}}}{v_{\text{th}}} = \frac{1}{\rho_0 \sigma^i_{\text{col}} \sqrt{2k_bT/m}} \tag{S12}$$

where superscripts indicate different values for the large and small diameter states, l^i_{mfp} is the mean free path for a gas of density ρ_0, σ^i_{col} is the collision cross section, and $v_{\text{th}} = \sqrt{2k_bT/m}$ is the thermal speed of this isothermal gas. We take the likelihood of a collision having occurred in the time between measurements t_m to be

$$P^i_{\text{col}} = 1 - e^{-t_m/\tau^i_{\text{col}}}. \tag{S13}$$

Assuming that particles are distributed over space uniformly, we can treat their occurrence within an area A as obeying a Poisson distribution, $P(n) = \frac{(\rho A)^n}{n!} e^{-\rho A}$ where n is the number of particles observed and A is the area of the annulus defined by the expanding diameter, $A = ((D_L + D_S)^2/4 - D_S^2)\pi$. This yields:

$$P_< = P(n = 0) = e^{-\rho_0 A}. \tag{S14}$$

The states shown in extended Fig. a can be condensed down to two states which mirror the condition of the memory of the demon particle, $m = 0$ or $m = 1$ (see extended Fig. S4b). The transition matrix describing the sequence of these states (and hence measurements) is

$$\mathbf{T} = \begin{bmatrix} 1 - \frac{1}{2}P^S_{\text{col}}P_< & \frac{1}{2}P^S_{\text{col}}P_< \\ \frac{1}{2}P^L_{\text{col}} & 1 - \frac{1}{2}P^L_{\text{col}} \end{bmatrix}, . \tag{S15}$$

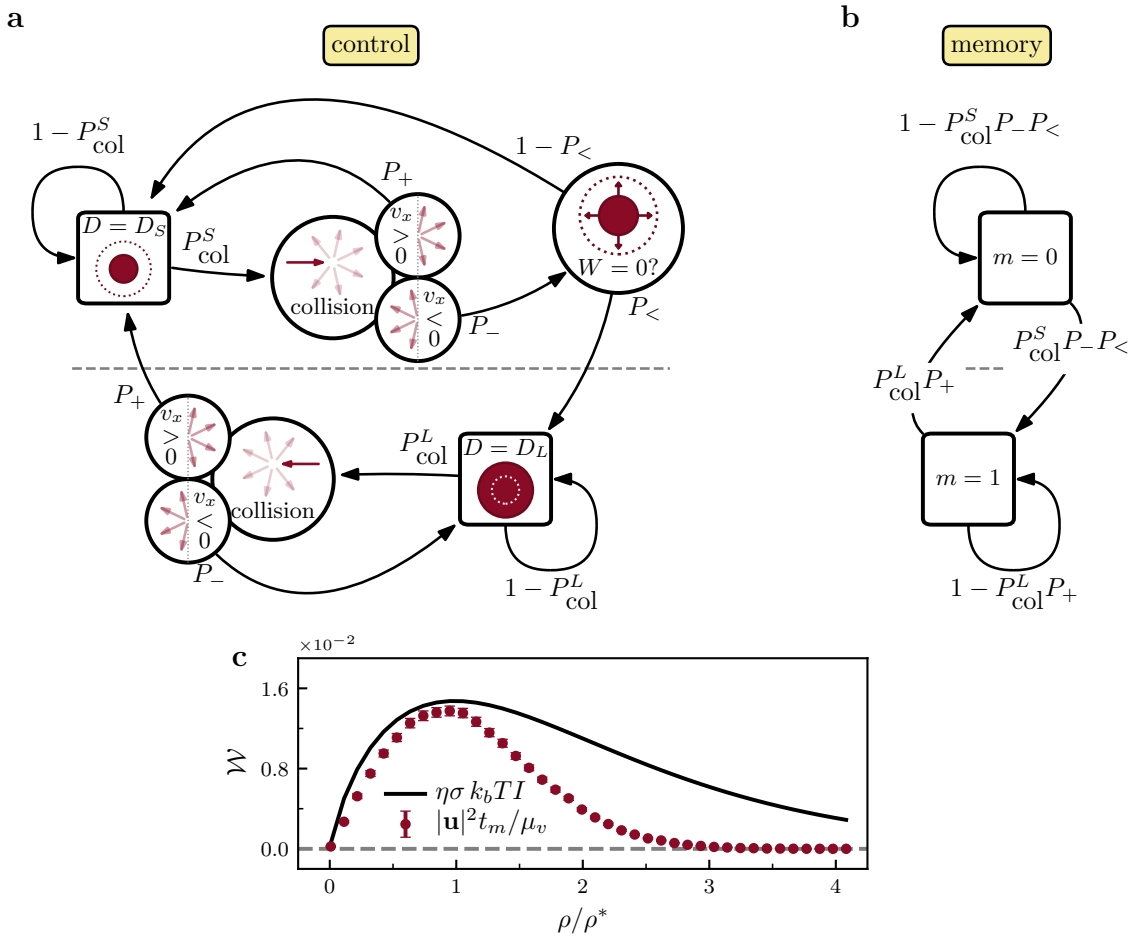


Fig. S4: **Markov chain model and simulation of a demon particle immersed in an isothermal gas**. **a**. Transition diagram for the states of a demon particle following a binary diameter rule: D_L when $v_x < 0$, D_S when $v_x > 0$. Collisions occur in the time between measurements (t_m) with probability $P^L_{\rm col}$ (for large particles or $P^S_{\rm col}$ for small) and randomize velocities (scattering into negative or positive velocities with probabilities P_- or P_+ respectively). For hard disks, shrinking diameter never requires the demon particle to exert work on the surrounding gas, but expansions are only possible if no other obstacles (i.e. particles) are nearby (probability $P_<$). **b**. Condensed two-state Markov chain model transition diagram for the one-bit measurement that demon particles collect to determine which diameter state to adopt. In the steady state, the measurement sequence $M = [m_t, m_{t+t_m}, \dots]$ has a mean entropy per measurement, $H(M)$. **c**. Work done to propel a single demon particle through a passive gas per measurement cycle (duration t_m). Power dissipated in demon particle motion is estimated from simulation by finding the density-dependent drift velocity $\boldsymbol{u}$ and mobility coefficient μ_v. The black curve is the average free energy dissipated in deleting the demon particle's memory of prior measurements, obtained from a Markov chain model of the measurement process (see SI section for additional details), where σ is the intrinsic efficiency of rectification by collisional biasing and η is a fitted constant of order one. Numerical and theoretical estimations agree on the location of the work maxima at a density of $\rho = \rho^* = 1/A$, where A is the excluded volume change as the demon particle changes between its two diameter states.

The corresponding steady state population is

$$\boldsymbol{\mu} = \left[\frac{P^L_{\rm col}}{P^L_{\rm col} + P^S_{\rm col} P_<} \quad \frac{P^S_{\rm col} P_<}{P^L_{\rm col} + P^S_{\rm col} P_<} \right], \tag{S16}$$

which is the left eigenvector with unity eigenvalue of the transition matrix $\mathbf{T}$. For error-free measurements, the information contained in an ensemble of measurements is equal to the Shannon entropy H. For a Markov chain in the steady state, the mean entropy (and therefore information) per step is given by [10]:

$$I(\mathbf{T}) = H(\mathbf{T}) = -\sum_{ij} \mu_i T_{ij} \ln T_{ij}. \tag{S17}$$

The Markov estimate of propulsive work $\mathcal{W}$ includes an efficiency term:

$$\sigma = \frac{\tau_{\rm col}^S - \tau_{\rm col}^L}{\tau_{\rm col}^S + \tau_{\rm col}^L} = \frac{\Delta D}{\Delta D + 4D_S}, \tag{S18}$$

which is the ratio of time moving forward over total time, assuming $P_+ = P_- = 1/2$. For the data shown in extended Fig. S4c, ΔD is fixed and $\sigma = 1/5$, and the fitting parameter $\eta = 5/13$.

Propulsive work was estimated from numerical data by finding the mobility constant from the self-diffusivity of the gas ($\mu_v = D_{\rm self}/k_bT$) for the the given density and temperature conditions. An effective force could then be defined from the drift velocity of the demon gas particle, $\boldsymbol{F}_{\rm eff} = \boldsymbol{u}/\mu_v$, and the work per action as $\boldsymbol{u} \cdot \boldsymbol{F}_{\rm eff} t_m$.

S5. THE DEMON GAS VELOCITY DISTRIBUTION FUNCTION

In order to derive an approximate description of the demon gas velocity distribution function, we will outline key points from the derivation of the Boltzmann equation which describes the evolution of the same distribution for a hard sphere gas. More information on the derivation of the standard Boltzmann equation and its mathematical properties can be found in texts by Cercignani et al [11] and Kardar [12]. From there, the treatment is extended to the case of hard disks with non-constant diameters (which are functions of particle velocity only). We show the form of the collision operator which describes the demon gas, and demonstrate that this operator has a modified collision invariant expression which includes the diameter function $D(\boldsymbol{v})$. We use this property to define a linear system of integral equations which permits solving for the deviation of the demon gas velocity distribution function from the equilibrium Maxwell-Boltzmann (MB) distribution.

S5.1. Kinetic theory of hard disks

The evolution of an N-particle gas in phase space is given by the Liouville equation,

$$\frac{\partial P}{\partial t} + \sum_{i=1}^{N} \boldsymbol{v}_i \cdot \frac{\partial P}{\partial \boldsymbol{x}_i} = 0 \tag{S19}$$

where P is a probability density and the influence of external forces has been neglected. Note that interaction forces do not yet explicitly appear, as they are 'internal' to the N-particle distribution. Interaction terms must appear for $s < N$ particle descriptions. The s-particle distributions are obtained by marginalization:

$$P^{(s)}(\boldsymbol{x}_1, \boldsymbol{v}_1, ..., \boldsymbol{x}_s, \boldsymbol{v}_s, t) = \int_{\mathcal{R}^{2N-2s}} P(\boldsymbol{x}_1, \boldsymbol{v}_1, \boldsymbol{x}_2, \boldsymbol{v}_2, ..., \boldsymbol{x}_N, \boldsymbol{v}_N, t) d\boldsymbol{x}_{s+1} d\boldsymbol{v}_{s+1}, ... d\boldsymbol{x}_N d\boldsymbol{v}_N \tag{S20}$$

and the corresponding evolution equation is obtained by integration of eq. S19 (see [11] for a derivation):

$$\frac{\partial P^{(s)}}{\partial t} + \sum_{i=1}^{s} \boldsymbol{v}_i \cdot \frac{\partial P^{(s)}}{\partial \boldsymbol{x}_i} = (N-s) \sum_{i=1}^{s} \int P^{(s+1)}(\boldsymbol{v}_i - \boldsymbol{v}_*) \cdot \hat{n}_i d\hat{n}_i d\boldsymbol{v}_* \tag{S21}$$

where the surface normal $\hat{n}$ has appeared since contact configurations between higher order distributions are responsible for disturbances to the s-particle streaming flow. This set of equations relating each s-particle distribution to the $s+1$ distributions is known as the BBGKY hierarchy. If a suitable closure of the hierarchy can be chosen (by approximating some distribution and truncating the sequence) then a useful evolution equation can be obtained. The Boltzmann equation is obtained in this way by focusing on the single-particle distribution $P^{(1)}(\boldsymbol{x}_1, \boldsymbol{v}_1, t)$, which is the probability of finding a particle within the interval $[\boldsymbol{x}_1 + d\boldsymbol{x}, \boldsymbol{v}_1 + d\boldsymbol{v}_1, t + dt]$. The evolution of $P^{(1)}$ in the absence of collisions proceeds only by the streaming of particles along straight trajectories ($\partial P^{(1)}/\partial t = -\boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} P^{(1)}$). Collisions between particles introduce a new operator related by the hierarchy to the 2-particle distributions:

$$\frac{\partial P^{(1)}}{\partial t} + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} P^{(1)} = \mathcal{C} \tag{S22}$$

$\mathcal{C}$ describes the rate of change of particle number (or density) within the interval $[\boldsymbol{x} + d\boldsymbol{x}, \boldsymbol{v} + d\boldsymbol{v}, t + dt]$ caused by collisions. For hard spheres, collisions occur only when two particles are found in contact. Naturally, $\mathcal{C}$ must involve the two particle distribution, $P^{(2)}$ (the probability of finding two particles in a given configuration).

If we denote pre-collision velocities for a pair of particles as $\boldsymbol{v}$ and $\boldsymbol{v}_*$, and their post-collision velocities as $\boldsymbol{v}'$ and $\boldsymbol{v}'_*$, we can write the collision operator as the difference of two rates, the gain and loss of particles within the interval of interest ($\mathcal{C} = \mathcal{C}^+ - \mathcal{C}^-$). Particles are removed from the interval by collisions (velocities are transformed from $\boldsymbol{v}, \boldsymbol{v}_*$ to $\boldsymbol{v}', \boldsymbol{v}'_*$). The likelihood of a collision is the same as the likelihood of finding a two-particle configuration in contact. For particles of fixed diameter (D_0):

$$\mathcal{C}^- = \frac{1}{2} \int_{\mathcal{R}^2} \int_{\mathcal{S}} P^{(2)}(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{x} + D_0 \hat{n}, \boldsymbol{v}_*, t) d\boldsymbol{v}_* d\hat{n} \tag{S23}$$

here the variable $\hat{n}$ is a unit vector that lies along the particle contact vector and integration over $d\hat{n}$ occurs on the surface $\mathcal{S}$ of all possible contacts. For a gas in a periodic domain (homogeneous spatial density) the spatial aspect of $P^{(2)}$ can be approximated geometrically. A rectangular area with length $|(\boldsymbol{v}_* - \boldsymbol{v}) \cdot \hat{n}|dt$ and width $|D_0 d\hat{n}| = D_0$ contains the particles with velocity $\boldsymbol{v}_*$ which are capable of striking the first particle. The probability density of finding two particles with velocities $\boldsymbol{v}, \boldsymbol{v}_*$ in contact can therefore be written as:

$$\mathcal{C}^- = \frac{1}{2} \int_{\mathcal{R}^2} \int_{\mathcal{S}^-} P^{(2)}(\boldsymbol{v}, \boldsymbol{v}_*, t) |(\boldsymbol{v}_* - \boldsymbol{v}) \cdot \hat{n}| D_0 d\boldsymbol{v}_* d\hat{n} \tag{S24}$$

where now location variables in $P^{(2)}$ have been dropped due to the assumption of spatial homogeneity. The inner integration is carried out over half the unit disk ($\mathcal{S}^-$, the half for which velocities $\boldsymbol{v}$ and $\boldsymbol{v}_*$ are approaching each other). This loss term is balanced (at steady state) by a similar term describing the gain of probability density by scattering from other regions of velocity space. The only differences being that the domain $\mathcal{S}^-$ is replaced by $\mathcal{S}^+$, the complementary hemisphere, and the velocities in consideration are $\boldsymbol{v}', \boldsymbol{v}'_*$, velocities which will undergo a scattering into $\boldsymbol{v}, \boldsymbol{v}_*$. As in the standard treatment, we will use the 'molecular chaos' assumption to replace two particle distribution terms with the product of single particle terms $P^{(2)}(\boldsymbol{v}, \boldsymbol{v}_*, t) \approx P^{(1)}(\boldsymbol{v}, t) P^{(1)}(\boldsymbol{v}_*, t)$. In the standard Boltzmann equation, the gain and loss terms are collected together under common integration limits by exploiting the time reversal symmetry of the collision process. Since we will now expand our treatment to include gases with diameters which are not constant in velocity space, we must examine the reversibility of hard disk collisions more closely.

S5.2. Hard disk collisions with non-constant diameter

We can find the velocities after collision between a pair of hard disks by conservation of momentum and energy. Assuming uniform particle masses, momentum and kinetic energy are conserved through the collision process:

$$\boldsymbol{v} + \boldsymbol{v}_* = \boldsymbol{v}' + \boldsymbol{v}'_* \tag{S25}$$

$$|\boldsymbol{v}|^2 + |\boldsymbol{v}_*|^2 = |\boldsymbol{v}'|^2 + |\boldsymbol{v}'_*|^2 \tag{S26}$$

This collision process can be viewed as the exchange of velocities projected onto the vector which points from one disk's center to the other. If $\hat{n}$ is a unit vector that points between disk centers at collision contact, then the post-collision velocities can be written as (for disks of equal mass):

$$\boldsymbol{v}' = \boldsymbol{v} - ((\boldsymbol{v} - \boldsymbol{v}_*) \cdot \hat{n}) \hat{n} \tag{S27}$$

$$\boldsymbol{v}'_* = \boldsymbol{v}_* + ((\boldsymbol{v} - \boldsymbol{v}_*) \cdot \hat{n}) \hat{n} \tag{S28}$$

For disks of fixed diameter, this scattering process is microreversible; $(\boldsymbol{v}, \boldsymbol{v}_*)$ map to $(\boldsymbol{v}', \boldsymbol{v}'_*)$ and $(-\boldsymbol{v}', -\boldsymbol{v}'_*)$ map to $(-\boldsymbol{v}, -\boldsymbol{v}_*)$, resulting in a Jacobian for the change of variables equal to -1 [11]. This one-to-one mapping allows all four velocity terms to be collected under the same unit disk integration in the standard Boltzmann collision operator.

When particle diameters are instead permitted to be functions of velocity, the microreversibility of the collision process is broken. We can see this diagrammatically in Fig. S5. In this representation, we view the two particle collision process as the scattering of a point object with velocity $\boldsymbol{v}_* - \boldsymbol{v}$ off of a stationary disk with diameter $(D(\boldsymbol{v}) + D(\boldsymbol{v}_*))/2$. By changing the location of the stationary disk, the mapping of velocities can be preserved (Fig. S5b). In a coordinate frame centered on the point of contact, the forward and reverse scattering process are symmetric provided that the

coordinates of the stationary particle are scaled by a factor $(D(\boldsymbol{v}')+D(\boldsymbol{v}'_*))/(D(\boldsymbol{v})+D(\boldsymbol{v}_*))$. The collision process can also be represented with the so-called 'impact parameter' (σ) as the integration variable (as opposed to the vector $\hat{n}$). Figure S5c shows that a rescaling of the impact parameter has an equivalent effect of preserving the velocity mapping. Naturally, the choice of rescaling one particle's position or the other is arbitrary and should produce equivalent results.

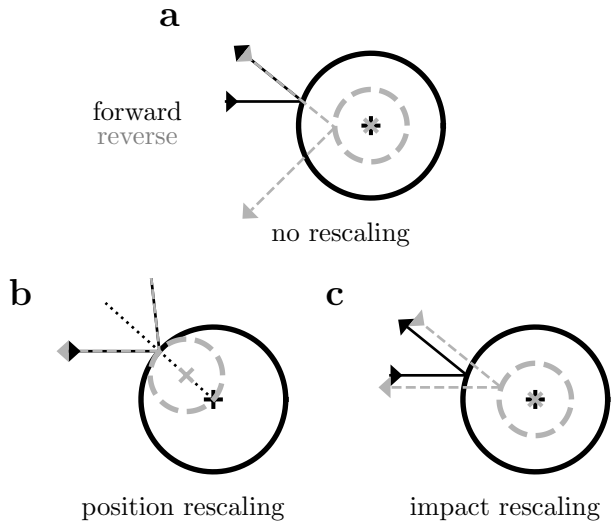


Fig. S5: **The microreversibility of hard disk collisions with velocity-dependent diameters.** Two particles of diameter $D(\boldsymbol{v}) = D$ and $D(\boldsymbol{v}_*) = D_*$ with velocities $\boldsymbol{v}$ and $\boldsymbol{v}_*$ are shown as the collision of a point object (with velocity $\boldsymbol{v}_* - \boldsymbol{v}$) with a stationary disk of diameter $D + D_*$. **a**. In the coordinate frame centered on the stationary disk, the forward (black) and reverse (grey) collisions are not symmetric. **b**. In the coordinate frame centered on the point of contact forward and reverse collision processes are symmetric, provided that the stationary disk's position is rescaled by a factor $(D' + D'_*)/(D + D_*)$. **c**. The symmetry of the collision can also be preserved in the frame centered on the stationary disk by rescaling the impact parameter of the incoming particle.

This mapping defines the procedure for exchanging unprimed and primed velocity variables in the collision operator of the demon gas. From now on, we will refer to the single particle velocity distribution function as $f(\boldsymbol{v})$, suppressing the location and time arguments as we are concerned with the homogeneous steady-state solution. For brevity, functions of velocity will have their argument suppressed but retain distinguishing marks (such as $f(\boldsymbol{v}'_*) = f'_*$ and $D(\boldsymbol{v}'_*) = D'_*$). In this notation, we write the loss term of the demon gas collision operator as:

$$\mathcal{C}^- = \frac{1}{2}\int_{\mathcal{R}^2}\int_{\mathcal{S}^-}(D + D_*)ff_*|(\boldsymbol{v}_* - \boldsymbol{v})\cdot\hat{n}|d\boldsymbol{v}_*d\hat{n} \tag{S29}$$

The gain term can be expressed by the exchange of unprimed velocities for primed ones. The term $\boldsymbol{v}_* - \boldsymbol{v} = \boldsymbol{v}'_* - \boldsymbol{v}'$ is conserved by the collision process, and so need not be transformed.

$$\mathcal{C}^+ = \frac{1}{2}\int_{\mathcal{R}^2}\int_{\mathcal{S}^+}(D' + D'_*)f'f'_*|(\boldsymbol{v}_* - \boldsymbol{v})\cdot\hat{n}|d\boldsymbol{v}'_*d\hat{n} \tag{S30}$$

These terms can be collected under a common integration, as the diameter function scaling acts a Jacobian term. The form of $\mathcal{C}$ shown in the text is achieved by redistribution of diameter terms:

$$\mathcal{C} = \mathcal{C}^+ - \mathcal{C}^- = \frac{1}{2}\int_{\mathcal{R}^2}\int_{\mathcal{S}}(D + D_*)|(\boldsymbol{v}_* - \boldsymbol{v})\cdot\hat{n}|\left[\frac{D' + D'_*}{D + D_*}f'f'_* - ff_*\right]d\boldsymbol{v}_*d\hat{n} \tag{S31}$$

Notice that the resulting collision operator is the sum of two terms that differ only by D and D_*. We will next consider the invariants of the demon gas collision operator.

S5.3. The demon gas collision invariant

We will first discuss collision invariants for the standard Boltzmann equation, and then explore the consequences of our modified demon gas collision operator. If we consider a moment of a distribution function $f(\boldsymbol{v}, t)$ which is a

solution to the Boltzmann equation with a non-time-varying test function ϕ:

$$\langle\phi\rangle_f = \int_{\mathcal{R}^2} f(\boldsymbol{v},t)\phi(\boldsymbol{v})d\boldsymbol{v} \tag{S32}$$

the evolution of this quantity is:

$$\frac{d}{dt}\langle\phi\rangle_f = \int_{\mathcal{R}^2} \phi(\boldsymbol{v})\mathcal{C}(f)(\boldsymbol{v},t)d\boldsymbol{v} \tag{S33}$$

The function ϕ can be brought into the integrand of the collision operator since it is not a function of time. As we are now considering an expression integrating over all $\boldsymbol{v}$ and $\boldsymbol{v}_*$ (and by extension $\boldsymbol{v}'$ and $\boldsymbol{v}'_*$), the choice of primed and starred variables is entirely arbitrary. A standard procedure is to average over the exchange of velocity variables [11]. We first express the right side of eq. S33 (with $\mathcal{C}$ appropriate for constant disks, and dropping velocity arguments where obvious) as:

$$\frac{1}{2}\int_{\mathcal{R}^2}\int_{\mathcal{R}^2}\int_{\mathcal{S}} \phi\left(f'f'_* - ff_*\right)|(\boldsymbol{v}_* - \boldsymbol{v})\cdot\hat{n}|D_0 d\boldsymbol{v}_* d\boldsymbol{v} d\hat{n} \tag{S34}$$

The exchange of starred and unstarred variables results in no change to the expression, allowing the average over this exchange to be written as:

$$\frac{1}{4}\int_{\mathcal{R}^2}\int_{\mathcal{R}^2}\int_{\mathcal{S}} (\phi+\phi_*)\left(f'f'_* - ff_*\right)|(\boldsymbol{v}_* - \boldsymbol{v})\cdot\hat{n}|D_0 d\boldsymbol{v}_* d\boldsymbol{v} d\hat{n} \tag{S35}$$

Exchanging primed variables (which can be done because of the microreversibility of the constant-diameter collision process), results in a change of sign. Therefore the result of averaging over both exchanges is:

$$\frac{1}{8}\int_{\mathcal{R}^2}\int_{\mathcal{R}^2}\int_{\mathcal{S}} (\phi+\phi_*-\phi'-\phi'_*)\left(f'f'_* - ff_*\right)|(\boldsymbol{v}_* - \boldsymbol{v})\cdot\hat{n}|D_0 d\boldsymbol{v}_* d\boldsymbol{v} d\hat{n} \tag{S36}$$

If ϕ is a quantity which is conserved through the collision, then $\frac{d}{dt}\langle\phi\rangle_f = 0$. This defines a summational collisional invariant for the system.

For the Boltzmann equation, summational invariants can only be linear combinations of the microscopically conserved quantities, $\mathcal{M} = 1$, $\boldsymbol{v}$, $|\boldsymbol{v}|^2$, which are unchanged during collision by construction [13]. Therefore there exist constants $(a, \boldsymbol{b}, c)$ such that:

$$h(\boldsymbol{v}) = a + \boldsymbol{b}\cdot\boldsymbol{v} + c|\boldsymbol{v}|^2 \tag{S37}$$

The preceding discussion has been for the standard Boltzmann collision operator. We can follow the same logic for the demon gas collision operator described in eq. S31. If we adopt the shorthand notation of the gain term from the main text ($g_+ = (f'f'_*)B(\hat{n},\Delta\boldsymbol{v})\alpha/2$, with $\alpha = (D'+D'_*)/(D+D_*)$ and $B(\hat{n},\Delta\boldsymbol{v}) = |(\boldsymbol{v}_*-\boldsymbol{v})\cdot\hat{n}|$ and the loss term as $g_- = (ff_*)B(\hat{n},\Delta\boldsymbol{v})/2$, the result of averaging over primed and unprimed variables for the demon gas can be written as:

$$\frac{1}{4}\int_{\mathcal{R}^2}\int_{\mathcal{R}^2}\int_{\mathcal{S}} (\phi(D+D_*)+\phi_*(D+D_*)-\phi'(D'+D'_*)-\phi'_*(D'+D'_*))\,(g_+-g_-)\,d\boldsymbol{v}_* d\boldsymbol{v} d\hat{n} \tag{S38}$$

where the leading term involving the function ϕ can be written as:

$$\phi(D+D_*)+\phi_*(D+D_*)-\phi'(D'+D'_*)-\phi'_*(D'+D'_*) = (\phi D+\phi_* D_*-\phi' D'-\phi'_* D'_*)+(\phi D_*+\phi_* D-\phi' D'_*-\phi'_* D'). \tag{S39}$$

If the entirety of eq. S39 is zero at all points in the velocity plane, then the quantity ϕ is stationary on the distribution f. The second term on the r.h.s. of eq. S39 contains only cross terms between starred and unstarred variables, and so all terms within it will integrate to the same value and the total term reduces to zero. An expression for the time

evolution of a test function ϕ for the demon gas can therefore be written with only the first term in the r.h.s. of eq. S39 as:

$$\frac{d}{dt}\langle\phi\rangle_f = -\frac{1}{2}\int_{\mathcal{R}^2}\int_{\mathcal{R}^2}\int_{\mathcal{S}} D\phi\left(f'f'_*\frac{D'+D'_*}{D+D_*} - ff_*\right)|(\boldsymbol{v}_*-\boldsymbol{v})\cdot\hat{n}|d\boldsymbol{v}_*d\boldsymbol{v}d\hat{n} \tag{S40}$$

From this representation, we can define a new test function, $h = D(\boldsymbol{v})\phi$. In order for the function ϕ to be stationary under the action of the demon collision operator, the quantity h must be a collisional invariant (unlike the case for fixed-diameter disks, where ϕ itself must be invariant). For a collisionally invariant h there must exist constants which express it as a linear combination of microscopically conserved quantities: $h = a+\boldsymbol{b}\cdot\boldsymbol{v}+c|\boldsymbol{v}|^2$. Note that an invariant test function for the MB distribution does not generically produce an invariant h for the demon collision operator, due to the diameter rescaling. We will use this relation to define the velocity distribution function of the demon gas. Beginning with the MB velocity distribution function in 2D:

$$f^{\mathrm{MB}}(\boldsymbol{v}) = \frac{m\rho}{2\pi kT}e^{-\frac{m|\boldsymbol{p}|^2}{2kT}} \tag{S41}$$

where $\boldsymbol{p} = \boldsymbol{v}-\langle\boldsymbol{v}\rangle = \boldsymbol{v}-\boldsymbol{u}$ is the velocity of the gas in the flow frame, we express the demon gas distribution function as $f^d = (1+\mathcal{D})f^{\mathrm{MB}}$, where $\mathcal{D}$ is a function of velocity only. Since we seek the steady-state form of $\mathcal{D}$, we can treat $(1+\mathcal{D})$ as a non-time-varying test function operating on the MB distribution. We are searching for the form of this test function that has stationary expectation values on the MB velocity distribution under the action of the demon gas collision operator. By substituting $\langle(1+\mathcal{D})\rangle_{f^{\mathrm{MB}},\mathcal{C}^d}$ for $\langle\phi\rangle_f$ in eq. S40, we see that steady state will be reached when $D(1+\mathcal{D})$ is a summational collision invariant. Finding the form of $\mathcal{D}$ amounts to specifying the four constants a, b_x, b_y, c which define the linear mixing of microscopically conserved quantities. These constants can be found if we require that the first four moments of the demon gas velocity distribution are equal to the first four moments of the equilibrium MB distribution.

This requirement reflects the fact that demon gas collisions microscopically conserve number, momentum, and kinetic energy, just as elastic collisions between fixed diameter disks. This requires that any re-sizing of the disks occurs between collisions. From a theoretical point of view, the gas is sparse enough to ignore the cases in which a diameter change post-collision must be delayed until the disks have sufficiently separated. This limitation on the ability of the disks to change their size is likely a significant contributor to deviations from the theory at high densities or large maximum diameters.

The requirement that the first four moments of the demon gas distribution are equal to those of the MB and can be written as:

$$\begin{aligned}\int_{\mathcal{R}^2}\mathcal{M}f^{\mathrm{MB}}d\boldsymbol{v} &= \int_{\mathcal{R}^2}\mathcal{M}(1+\mathcal{D})f^{\mathrm{MB}}d\boldsymbol{v}\\ 0 &= \int_{\mathcal{R}^2}\mathcal{M}\mathcal{D}f^{\mathrm{MB}}d\boldsymbol{v}\end{aligned} \tag{S42}$$

which highlights that the expectation value of the deviation of the demon distribution ($\mathcal{D}$) must be zero for the first four moments. Note that this restriction does not constrain the third-order moments which define the heat flux tensor, or even the individual entries of the pressure tensor (derived from second-order moments). Only the trace of the pressure tensor is fixed relative to the MB distribution.

Using the expression for the demon gas collision invariant, $h = D(1+\mathcal{D})$, we can trivially reorder to obtain $(1+\mathcal{D}) = hD^{-1}$. As a scalar function which must be greater than zero, division by D presents no problems. In practice, D need not even be non-zero everywhere, but instead only where the probability density of f^{MB} is not infinitesimal. Using this expression for the demon perturbation, and the notation for expectation value $\langle\phi\rangle^{\mathrm{MB}} = \int\phi f^{\mathrm{MB}}d\boldsymbol{v}$, we can write:

$$\begin{aligned}\langle\mathcal{M}\rangle^{\mathrm{MB}} &= \langle\mathcal{M}(1+\mathcal{D})\rangle^{\mathrm{MB}}\\ &= \langle\mathcal{M}hD^{-1}\rangle^{\mathrm{MB}}\\ &= a\langle\mathcal{M}D^{-1}\rangle^{\mathrm{MB}} + b_x\langle\mathcal{M}v_xD^{-1}\rangle^{\mathrm{MB}} + b_y\langle\mathcal{M}v_yD^{-1}\rangle^{\mathrm{MB}} + c\langle\mathcal{M}|\boldsymbol{v}|^2D^{-1}\rangle^{\mathrm{MB}}\end{aligned} \tag{S43}$$

For each of the conserved moments ($\mathcal{M} = 1,\ v_x,\ v_y,\ |\boldsymbol{v}|^2$) a new expression is found, allowing for the definition of a full rank system of linear equations:

$$\mathcal{A} \cdot G = \mathcal{B} \tag{S44}$$

$$\mathcal{A} = \begin{bmatrix} \langle \frac{1}{D} \rangle_m & \langle \frac{v_x}{D} \rangle_m & \langle \frac{v_y}{D} \rangle_m & \langle \frac{|\boldsymbol{v}|^2}{D} \rangle_m \\ \cdot & \langle \frac{v_x^2}{D} \rangle_m & \langle \frac{v_x v_y}{D} \rangle_m & \langle \frac{v_x |\boldsymbol{v}|^2}{D} \rangle_m \\ \cdot & \cdot & \langle \frac{v_y^2}{D} \rangle_m & \langle \frac{v_y |\boldsymbol{v}|^2}{D} \rangle_m \\ \cdot & \cdot & \cdot & \langle \frac{|\boldsymbol{v}|^4}{D} \rangle_m \end{bmatrix} \tag{S45}$$

$$G = \begin{bmatrix} a & b_x & b_y & c \end{bmatrix}^\top \tag{S46}$$

$$\mathcal{B} = \begin{bmatrix} \rho & \rho u_x & \rho u_y & \rho |\boldsymbol{u}|^2/2 + \rho k T \end{bmatrix}^\top \tag{S47}$$

where $\mathcal{A}$ is symmetric and redundant entries are not shown. Using the solution of this system of equations the demon gas velocity distribution can be written as:

$$f^d = f^{\mathrm{MB}}(1 + \mathcal{D}) = f^{\mathrm{MB}} D^{-1} h = f^{\mathrm{MB}} D^{-1} \left(\boldsymbol{M} \cdot (\mathcal{A}^{-1} \mathcal{B}) \right) \tag{S48}$$

where $\boldsymbol{M} = [1, v_x, v_y, |\boldsymbol{v}|^2]$.

S5.4. Demonstration of f^d as the steady-state solution to the variable-diameter collision operator

We can show directly that the above procedure for finding an demon gas distribution function solves $\mathcal{C}[f^d] = 0$ by examining the collision operator written as:

$$\mathcal{C} = \frac{1}{2} \int_{\mathcal{R}^2} \int_{\mathcal{S}} B(\hat{n}, \Delta \boldsymbol{v}) \left[(D' + D'_*) f' f'_* - (D + D_*) f f_* \right] d\boldsymbol{v}_* d\hat{n} \tag{S49}$$

in which the cross-section terms have been redistributed and the kernel term has been collected into a function B for brevity. Focusing on the term in brackets (which we will refer to as C), we can replace f with $f^d = D(1 + \mathcal{D}) f^{\mathrm{MB}}$:

$$C \equiv (D' + D'_*)(1 + \mathcal{D}')(f^{\mathrm{MB}})'(1 + \mathcal{D}'_*)(f^{\mathrm{MB}})'_* - (D + D_*)(1 + \mathcal{D}) f^{\mathrm{MB}} (1 + \mathcal{D}_*) f^{\mathrm{MB}}_* \tag{S50}$$

First, note that the Maxwellian distribution is the exponential of a collision invariant, or:

$$\ln\ f^{\mathrm{MB}} + \ln\ f^{\mathrm{MB}}_* - \ln\ (f^{\mathrm{MB}})' - \ln\ (f^{\mathrm{MB}})'_* = 0 \tag{S51}$$

$$\frac{f^{\mathrm{MB}} f^{\mathrm{MB}}_*}{(f^{\mathrm{MB}})'(f^{\mathrm{MB}})'_*} = 1 \tag{S52}$$

Equation S52 allows for eq. S50 to be written as:

$$\begin{aligned} C = \ & f^{\mathrm{MB}} f^{\mathrm{MB}}_* \left[D'(1 + \mathcal{D}') + \mathcal{D}'_* D'(1 + \mathcal{D}') + D'_*(1 + \mathcal{D}'_*) + \mathcal{D}' D'_*(1 + \mathcal{D}'_*) \right] \\ & + f^{\mathrm{MB}} f^{\mathrm{MB}}_* \left[D(1 + \mathcal{D}) - \mathcal{D}_* D(1 + \mathcal{D}) - \mathcal{D}_*(1 + \mathcal{D}_*) - \mathcal{D} D_*(1 + \mathcal{D}_*) \right] \end{aligned} \tag{S53}$$

Recall that the quantity $h = D(1 + \mathcal{D})$ is also a collision invariant. We simplify eq. S53 to:

$$C = f^{\mathrm{MB}} f^{\mathrm{MB}}_* \left[h' + h'_* - h - h_* \right] + f^{\mathrm{MB}} f^{\mathrm{MB}}_* \left[h' \mathcal{D}'_* + h'_* \mathcal{D}' - h \mathcal{D}_* - h_* \mathcal{D} \right] \tag{S54}$$

The first term in eq. S54 is zero, by the definition of collision invariants. The second term is also equal to zero, because each of the four terms individually integrate to zero. The logic for each term is identical, so we will focus on $f^{\mathrm{MB}} f^{\mathrm{MB}}_* h \mathcal{D}_*$:

$$\begin{aligned} f^{\mathrm{MB}} f^{\mathrm{MB}}_* h \mathcal{D}_* &\to -\frac{1}{2} \int_{\mathcal{R}^2} \int_{\mathcal{S}} B(\hat{n}, \Delta \boldsymbol{v}) f^{\mathrm{MB}} f^{\mathrm{MB}}_* h \mathcal{D}_* d\boldsymbol{v}_* d\hat{n} \\ &= -\frac{1}{2} h f^{\mathrm{MB}} \int_{\mathcal{R}^2} \int_{\mathcal{S}} B(\hat{n}, \Delta \boldsymbol{v}) \mathcal{D}_* f^{\mathrm{MB}}_* d\boldsymbol{v}_* d\hat{n} \end{aligned} \tag{S55}$$

Recall that the collision kernel B for hard disk collisions is a linear function of relative velocity. Therefore, eq. S55 is a first-order moment of $\mathcal{D} f^{\mathrm{MB}}$, which must equal zero by construction (see eq. S42). The other terms in eq. S54 can all be likewise shown to equal zero by exchange of velocity variables and use of eq. S52.

S5.5. Invertibility of the matrix $\mathcal{A}$

$\mathcal{A}$ can be written as the element-wise expectation over $f^{\mathrm{MB}}D^{-1}$ of the symmetric matrix:

$$\mathbb{E}_{f^{\mathrm{MB}}D^{-1}}\begin{bmatrix}1 & v_x & v_y & |\boldsymbol{v}|^2\\ . & v_x^2 & v_xv_y & v_x|\boldsymbol{v}|^2\\ . & . & v_y^2 & v_y|\boldsymbol{v}|^2\\ . & . & . & |\boldsymbol{v}|^4\end{bmatrix} \tag{S56}$$

This matrix can be written as the outer product of two vectors (therefore it has rank one):

$$\begin{bmatrix}1 & v_x & v_y & |\boldsymbol{v}|^2\end{bmatrix}^\top\begin{bmatrix}1 & v_x & v_y & |\boldsymbol{v}|^2\end{bmatrix} \tag{S57}$$

Despite this, $\mathcal{A}$ is generically invertible, as the element-wise expectation promotes the matrix to full rank. If we temporarily adopt the notation of $\langle\phi\rangle = \int \phi D^{-1} f^{\mathrm{MB}} d\boldsymbol{v}$, $\mathcal{A}$ can be written as:

$$\begin{bmatrix}\langle 1\rangle & \langle v_x\rangle & \langle v_y\rangle & \langle|\boldsymbol{v}|^2\rangle\\ \langle v_x\rangle & \langle v_x^2\rangle & \langle v_xv_y\rangle & \langle v_x|\boldsymbol{v}|^2\rangle\\ \langle v_y\rangle & \langle v_xv_y\rangle & \langle v_y^2\rangle & \langle v_y|\boldsymbol{v}|^2\rangle\\ \langle|\boldsymbol{v}|^2\rangle & \langle v_x|\boldsymbol{v}|^2\rangle & \langle v_y|\boldsymbol{v}|^2\rangle & \langle|\boldsymbol{v}|^4\rangle\end{bmatrix} \tag{S58}$$

The determinant of this matrix can be found by cofactor expansion along the first column. If we examine the determinant of the first of the four 3×3 matrices which make up $\mathcal{A}$, we find that a term composed of the difference of permutations of expectation values is produced:

$$\langle 1\rangle\begin{vmatrix}\langle v_x^2\rangle & \langle v_xv_y\rangle & \langle v_x|\boldsymbol{v}|^2\rangle\\ \langle v_xv_y\rangle & \langle v_y^2\rangle & \langle v_y|\boldsymbol{v}|^2\rangle\\ \langle v_x|\boldsymbol{v}|^2\rangle & \langle v_y|\boldsymbol{v}|^2\rangle & \langle|\boldsymbol{v}|^4\rangle\end{vmatrix} \tag{S59}$$

$$= \langle 1\rangle\left(\langle v_x^2\rangle\langle v_y^2\rangle - \langle v_xv_y\rangle^2\right)\langle|\boldsymbol{v}|^4\rangle$$

$$+\langle 1\rangle\left(\langle v_xv_y\rangle\langle v_x\rangle\langle v_y\rangle + \langle v_x\rangle^2\langle v_y\rangle^2 - \langle v_x^2\rangle\langle v_y\rangle^2 - \langle v_x\rangle^2\langle v_y^2\rangle\right)\langle|\boldsymbol{v}|^2\rangle^2 \tag{S60}$$

which will be nonzero due to the non-independence of the terms being exchanged through the expectation operator. All four sub-matrices follow this same pattern, with no requirements on the form of D^{-1}. Provided that $D > 0$ across the velocity plane, $\mathcal{A}$ can be numerically inverted without difficulty.

S5.6. Explicit form of $\mathcal{A}$ for select $D(\boldsymbol{v})$

The matrix $\mathcal{A}$ which defines the demon gas velocity distribution function involves finding the expectation value (over the MB distribution) of several functions of velocity multiplied with the inverse diameter function (D). For arbitrary D these Gaussian integrals may become too cumbersome to find analytically, and numerical integration is expedient. In this appendix we will present explicit equations for the $\mathcal{A}$ matrices of two special D functions - a diameter step function, and an anisotropic Gaussian diameter function. These matrices may in principle be inverted analytically through a process such as finding the adjugate matrix. The equations for such a process rapidly become unwieldy however, so we will instead use a semi-analytical approach of finding $\mathcal{A}$ exactly, then performing a numerical inverse and estimating small-value derivatives of the relevant constants a, b_x, b_y, and c numerically as required. This numerical procedure can also be extended to estimate the values of the $\mathcal{A}$ matrix for diameter functions which do not admit feasible analytic integration.

S5.6.1. Step function

A convenient diameter function can be written using the Heaviside step function:

$$D(\boldsymbol{v}) = D_S + (D_L - D_S)\mathcal{H}\left(-\boldsymbol{v}\cdot\boldsymbol{\xi}\right) \tag{S61}$$

As we will require D^{-1} in many formulas, it is useful to express this also with the Heaviside function:

$$D^{-1}(\boldsymbol{v}) = \frac{1}{D_L} + \left(\frac{1}{D_S} - \frac{1}{D_L}\right)\mathcal{H}\left(-\boldsymbol{v}\cdot\boldsymbol{\xi}\right) \tag{S62}$$

where D_S and D_L are the small and large diameters respectively, and $\boldsymbol{\xi}$ is the normal vector of the large/small decision boundary. We will only consider fixed $\boldsymbol{\xi} = \begin{bmatrix}1 & 0\end{bmatrix}$ here. The value of the Heaviside step function at zero will here be taken to be zero (as opposed to 1/2). With this choice of $\boldsymbol{\xi}$, integrals over the velocity plane can be split into an integral over the entire plane in addition to an integral over the half plane ($x > 0$). Using well-known expressions for Gaussian integrals of the form $\int x^n \mathcal{N}(x)$ [14], the following components of the $\mathcal{A}$ matrix can be found for a Maxwell-Boltzmann distribution in 2D with bulk velocity $\boldsymbol{u}$, thermal energy k_bT and gas density ρ. Where it appears, erf() refers to the standard error function.

$$\mathcal{A}_{11} = \int D^{-1} f^{\mathrm{MB}} = \frac{1}{D_l} + \frac{1}{2}\left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(1 - \mathrm{erf}\left(-u_x\sqrt{\frac{m}{2k_bT}}\right)\right) \tag{S63}$$

$$\mathcal{A}_{12} = \mathcal{A}_{21} = \int v_x D^{-1} f^{\mathrm{MB}} = \frac{1}{D_l}u_x + \left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(\frac{1}{2}u_x\left(\mathrm{erf}\left(-u_x\sqrt{\frac{m}{2k_bT}}\right) - 1\right) - \sqrt{\frac{k_bT}{2\pi m}}e^{-\frac{m}{2k_bT}u_x^2}\right) \tag{S64}$$

$$\mathcal{A}_{13} = \mathcal{A}_{31} = \int v_y D^{-1} f^{\mathrm{MB}} = \frac{1}{D_l}u_y + \left(\frac{1}{D_s} - \frac{1}{D_l}\right)\frac{1}{2}u_y\left(1 - \mathrm{erf}\left(-u_x\sqrt{\frac{m}{2k_bT}}\right)\right) \tag{S65}$$

$$\mathcal{A}_{14} = \mathcal{A}_{41} = \int |\boldsymbol{v}|^2 D^{-1} f^{\mathrm{MB}} = \int v_x^2 D^{-1} f^{\mathrm{MB}} + \int v_y^2 D^{-1} f^{\mathrm{MB}} = \mathcal{A}_{22} + \mathcal{A}_{33} \tag{S66}$$

$$\begin{aligned}\mathcal{A}_{22} = \int v_x^2 D^{-1} f^{\mathrm{MB}} = \quad & \frac{1}{D_l}\left(u_x^2 + \frac{k_bT}{m}\right) + \\ & \frac{1}{2}\left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(\left(u_x^2 + \frac{k_bT}{m}\right)\left(1 - \mathrm{erf}\left(-u_x\sqrt{\frac{m}{2k_bT}}\right)\right) + u_x\sqrt{\frac{k_bT}{2m\pi}}e^{-\frac{m}{2k_bT}u_x^2}\right)\end{aligned} \tag{S67}$$

$$\mathcal{A}_{33} = \int v_y^2 D^{-1} f^{\mathrm{MB}} = \frac{1}{D_l}\left(u_y^2 + \frac{k_bT}{m}\right) + \frac{1}{2}\left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(u_y^2 + \frac{k_bT}{m}\right)\left(1 - \mathrm{erf}\left(-u_x\sqrt{\frac{m}{2k_bT}}\right)\right) \tag{S68}$$

$$\begin{aligned}\mathcal{A}_{23} = \mathcal{A}_{32} = \int v_x v_y D^{-1} f^{\mathrm{MB}} = \quad & \frac{1}{D_l}u_x u_y + \\ & \left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(\frac{u_x u_y}{2}\left(1 - \mathrm{erf}\left(-u_x\sqrt{\frac{m}{2k_bT}}\right)\right) + u_y\sqrt{\frac{k_bT}{2\pi m}}e^{-\frac{m}{2k_bT}u_x^2}\right)\end{aligned} \tag{S69}$$

$$\mathcal{A}_{24} = \mathcal{A}_{42} = \int v_x|\boldsymbol{v}|^2 D^{-1} f^{\mathrm{MB}} = \int v_x^3 D^{-1} f^{\mathrm{MB}} + \int v_x v_y^2 D^{-1} f^{\mathrm{MB}} \tag{S70}$$

$$\begin{aligned}\int v_x^3 D^{-1} f^{\mathrm{MB}} = & \frac{1}{D_l}u_x\left(u_x^2 + \frac{3k_bT}{m}\right) + \\ & \left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(\frac{u_x}{2}\left(\left(u_x^2 + \frac{3k_bT}{m}\right)\left(1 - \mathrm{erf}\left(-u_x\sqrt{\frac{m}{2k_bT}}\right)\right) + \left(\frac{u_x^2}{2} + \frac{k_bT}{m}\right)\sqrt{\frac{2k_bT}{\pi m}}e^{-u_x^2\frac{m}{2k_bT}}\right)\right)\end{aligned}$$

$$\begin{aligned}\int v_x v_y^2 D^{-1} f^{\mathrm{MB}} = & \frac{1}{D_l}u_x\left(u_y^2 + \frac{k_bT}{m}\right) + \\ & \left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(u_y^2 + \frac{kT}{m}\right)\left(\frac{1}{2}u_x\left(1 - \mathrm{erf}\left(-u_x\sqrt{\frac{m}{2k_bT}}\right)\right) + \sqrt{\frac{k_bT}{2\pi m}}e^{-\frac{m}{2k_bT}u_x^2}\right)\end{aligned}$$

$$\mathcal{A}_{34} = \mathcal{A}_{43} = \int v_y |\boldsymbol{v}|^2 D^{-1} f^{\mathrm{MB}} = \int v_y v_x^2 D^{-1} f^{\mathrm{MB}} + \int v_y^3 D^{-1} f^{\mathrm{MB}} \tag{S71}$$

$$\int v_y v_x^2 D^{-1} f^{\mathrm{MB}} = \frac{1}{D_l} u_y \left(u_x^2 + \frac{k_b T}{m}\right) + \left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left[\frac{1}{2} u_y \sqrt{\frac{m}{2k_b T}} \left(u_x^2 + \frac{k_b T}{m}\right)\left(1 - \mathrm{erf}\left(-u_x \sqrt{\frac{m}{2k_b T}}\right)\right) + \frac{u_x u_y}{2\sqrt{\pi}} e^{-u_x^2 \frac{m}{2k_b T}}\right]$$

$$\int v_y^3 D^{-1} f^{\mathrm{MB}} = u_y \left(u_y^2 + \frac{3k_b T}{m}\right)\left[\frac{1}{D_l} + \frac{1}{2}\left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(1 - \mathrm{erf}\left(-u_x \sqrt{\frac{m}{2k_b T}}\right)\right)\right]$$

$$\mathcal{A}_{44} = \int |\boldsymbol{v}|^4 D^{-1} f^{\mathrm{MB}} = \int v_x^4 D^{-1} f^{\mathrm{MB}} + 2 v_x^2 v_y^2 D^{-1} f^{\mathrm{MB}} + v_y^4 D^{-1} f^{\mathrm{MB}} \tag{S72}$$

$$\int v_x^4 D^{-1} f^{\mathrm{MB}} = \left(u_x^4 + \frac{6k_b T}{m} u_x^2 + \frac{3k_b T^2}{m^2}\right) \times \left[\frac{1}{D_l} + \frac{1}{2}\left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(1 - \mathrm{erf}\left(-u_x \sqrt{\frac{m}{2k_b T}}\right)\right)\right] + \left(\frac{1}{D_s} - \frac{1}{D_l}\right)\sqrt{\frac{k_b T}{2\pi m}} u_x e^{-u_x \sqrt{\frac{m}{2k_b T}}} \left(u_x^2 + 5\frac{k_b T}{m}\right)$$

$$\int 2 v_x^2 v_y^2 D^{-1} f^{\mathrm{MB}} = \frac{2}{D_l}\left(u_x^2 + \frac{k_b T}{m}\right)\left(u_y^2 + \frac{k_b T}{m}\right) + \left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(u_y^2 + \frac{k_b T}{m}\right)\left[\left(u_x^2 + \frac{k_b T}{m}\right)\left(1 - \mathrm{erf}\left(-u_x \sqrt{\frac{m}{2k_b T}}\right)\right) + \sqrt{\frac{2k_b T}{\pi m}} u_x e^{-u_x^2 \frac{m}{2k_b T}}\right]$$

$$\int v_y^4 D^{-1} f^{\mathrm{MB}} = \left(u_y^4 + \frac{6k_b T}{m} u_y^2 + \frac{3k_b T^2}{m^2}\right)\left[\frac{1}{D_l} + \left(\frac{1}{D_s} - \frac{1}{D_l}\right)\left(1 - \mathrm{erf}\left(-u_x \sqrt{\frac{m}{2k_b T}}\right)\right)\right]$$

S5.6.2. Anisotropic Gaussian diameter function

Another form for D with comparatively simple integrals to compute is an anisotropic Gaussian:

$$D = e^{-\gamma_x \frac{m}{2k_b T}(v_x - u_x)^2} e^{-\gamma_y \frac{m}{2k_b T}(v_y - u_y)^2} \tag{S73}$$

where $\boldsymbol{u}$ is again the average velocity of the gas. The relative values of the constants γ_x and γ_x control the degree of anisotropy. Note that this diameter function tends to zero as velocity increases. As such, a well-behaved distribution function for a gas with this diameter function only exists for $\gamma_x < 1$, $\gamma_y < 1$. Otherwise, regions of the velocity plane with finite density of particles will have zero diameter, acting as particle sinks. Nevertheless, this diameter function is particularly convenient because it can be merged with the MB distribution and integrated using the same functional forms:

$$\mathcal{A}_{11} = \int D^{-1} f^{\mathrm{MB}} = \frac{1}{\sqrt{1-\gamma_x}\sqrt{1-\gamma_y}} \tag{S74}$$

$$\mathcal{A}_{12} = \mathcal{A}_{21} = \int v_x D^{-1} f^{\mathrm{MB}} = \frac{u_x}{\sqrt{1-\gamma_y}\sqrt{1-\gamma_x}} \tag{S75}$$

$$\mathcal{A}_{13} = \mathcal{A}_{31} = \int v_y D^{-1} f^{\mathrm{MB}} = \frac{u_y}{\sqrt{1-\gamma_y}\sqrt{1-\gamma_x}} \tag{S76}$$

$$\mathcal{A}_{14} = \mathcal{A}_{41} = \mathcal{A}_{22} + \mathcal{A}_{33} \tag{S77}$$

$$\mathcal{A}_{22} = \int v_x^2 D^{-1} f^{\mathrm{MB}} = \frac{1}{\sqrt{1-\gamma_y}\sqrt{1-\gamma_x}}\left(u_x^2 + \frac{k_b T}{m}\frac{1}{(1-\gamma_x)}\right) \tag{S78}$$

$$\mathcal{A}_{33} = \int v_y^2 D^{-1} f^{\mathrm{MB}} = \frac{1}{\sqrt{1-\gamma_y}\sqrt{1-\gamma_x}} \left(u_y^2 + \frac{k_b T}{m} \frac{1}{(1-\gamma_y)} \right) \tag{S79}$$

$$\mathcal{A}_{23} = \mathcal{A}_{32} = \int v_x v_y D^{-1} f^{\mathrm{MB}} = \frac{u_x u_y}{\sqrt{1-\gamma_x}\sqrt{1-\gamma_y}} \tag{S80}$$

$$\mathcal{A}_{24} = \mathcal{A}_{42} = \int v_x |\boldsymbol{v}|^2 D^{-1} f^{\mathrm{MB}} = \int v_x^3 D^{-1} f^{\mathrm{MB}} + \int v_x v_y^2 D^{-1} f^{\mathrm{MB}} \tag{S81}$$

$$= \frac{u_x}{\sqrt{1-\gamma_x}\sqrt{1-\gamma_y}} \left(u_x^2 + u_y^2 + \frac{k_b T}{m} \left(\frac{3}{1-\gamma_x} + \frac{1}{1-\gamma_y} \right) \right) \tag{S82}$$

$$\mathcal{A}_{34} = \mathcal{A}_{43} = \int v_y |\boldsymbol{v}|^2 D^{-1} f^{\mathrm{MB}} = \int v_y v_x^2 D^{-1} f^{\mathrm{MB}} + \int v_y^3 D^{-1} f^{\mathrm{MB}} \tag{S83}$$

$$= \frac{u_y}{\sqrt{1-\gamma_x}\sqrt{1-\gamma_y}} \left(u_x^2 + u_y^2 + \frac{k_b T}{m} \left(\frac{3}{1-\gamma_y} + \frac{1}{1-\gamma_x} \right) \right) \tag{S84}$$

$$\mathcal{A}_{44} = \int v_x^4 D^{-1} f^{\mathrm{MB}} + 2 \int v_x^2 v_y^2 D^{-1} f^{\mathrm{MB}} + \int v_y^4 D^{-1} f^{\mathrm{MB}} \tag{S85}$$

$$\begin{aligned} &= \frac{1}{\sqrt{1-\gamma_x}\sqrt{1-\gamma_y}} \left(u_x^4 + u_y^4 + 2u_x^2 u_y^2 \right) \\ &\quad + \frac{2}{\sqrt{1-\gamma_x}\sqrt{1-\gamma_y}} \frac{k_b T}{m} \left(u_x^2 \left(\frac{3}{1-\gamma_x} + \frac{2}{1-\gamma_y} \right) + u_y^2 \left(\frac{3}{1-\gamma_y} + \frac{2}{1-\gamma_x} \right) \right) \\ &\quad + \frac{3}{\sqrt{1-\gamma_x}\sqrt{1-\gamma_y}} \left(\frac{k_b T}{m} \right)^2 \left(\frac{1}{(1-\gamma_x)^2} + \frac{1}{(1-\gamma_y)^2} \right) \end{aligned} \tag{S86}$$

S6. KINETIC THEORY OF THE NEMATIC FLOCKING MODEL

In this section, we develop a version of the demon gas which selects diameter changes based only on information local to each particle, i.e. not involving any global or laboratory-frame coordinate information. Prior examples assumed that the velocity or position of particles relative to the lab frame was known and used in determining the diameter of particles. Instead, here a local coordinate system is used to determine the diameter of particles. This allows us to study the spontaneous emergence of long-range order as a common coordinate orientation is adopted system-wide.

S6.1. Diameter function defined in the fluid flow frame

In this section we consider a diameter function defined on a coordinate frame that moves with bulk fluid flow, i.e. a function of the peculiar velocity $\boldsymbol{p} = \boldsymbol{v} - \boldsymbol{u}$. Our diameter function will depend only on the in-plane direction (θ) of velocity relative to a local orientation vector,

$$D(\theta) = \Delta D \mathcal{H}(\cos^2\theta - \frac{1}{2}) + D_S \tag{S87}$$

where $\mathcal{H}$ is the Heaviside step function D_S is the minimum possible diameter, and the largest diameter $D_L = D_S + \Delta D$. The in-plane direction θ will be determined relative to a direction ψ as $\theta = \mathrm{atan2}(p_y, p_x) - \psi$, where $p_{x,y}$ are peculiar velocity components expressed in lab frame basis. The angle ψ is determined locally as the in-plane rotation that diagonalizes the local virial pressure tensor (P),

$$P_{ij} = -\frac{1}{\Omega} \sum_{k \in \Omega} m^{(k)} (v_i^{(k)} - u_i)(v_j^{(k)} - u_j), \tag{S88}$$

where Ω is the region surrounding the particle we are considering. The pressure tensor P is computed in the laboratory frame, and its basis can be changed (and therefore diagonalized) by rotation,

$$R = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \tag{S89}$$

$$R^T P R = \begin{pmatrix} P^\psi_{xx} & 0 \\ 0 & P^\psi_{yy} \end{pmatrix} = P^\psi. \tag{S90}$$

Pressure tensors without superscripts are expressed in the lab frame, and the ψ superscript indicates the diagonal frame. Note that every particle independently computes a separate local virial pressure and diagonalization angle. The angle ψ can be defined from the lab-frame pressure tensor using the two-argument arctangent:

$$\psi = \frac{1}{2}\text{atan2}(2P_{xy},\, P_{xx} - P_{yy}). \tag{S91}$$

Note that the pressure terms in eq. S91 are laboratory-frame pressures. As it will return frequently, we will use the symbol $\Delta P \equiv P_{xx} - P_{yy}$ to denote pressure asymmetry. For $P_{xy} = 0$ and $P_{xx} > P_{yy}$, $\psi = 0$. Conversely for $P_{xy} = 0$ and $P_{xx} < P_{yy}$, $\psi = \pi/2$. In the case where $P_{xy} = 0$ and $P_{xx} = P_{yy}$, $R = \mathbf{I}$ and we naturally take $\psi = 0$. Equation S87 indicates that the diameter function takes on two discrete values, one in the arcs $\theta - \psi = [-\pi/4,\ \pi/4]$ and $[3\pi/4,\ 5\pi/4]$, and a different value in the arcs $\theta - \psi = [\pi/4,\ 3\pi/4]$ and $[5\pi/4,\ 7\pi/4]$. For ease, we will refer to those two diameters as D_x and D_y respectively. The diameter function in eq S87 is designed to induce asymmetry between P^ψ_{xx} and P^ψ_{yy}. Here we are using the superscript to denote the the diagonal basis. In an extended system $\psi(\boldsymbol{x})$, two questions are immediately raised:

1. Does the asymmetry in pressure that defines $\psi(\boldsymbol{x})$ grow locally? We can denote this asymmetry as $\Delta P^\psi = P^\psi_{xx} - P^\psi_{yy}$. An equivalent quantity can be obtained from the Frobenius norm of the traceless pressure tensor, $\sqrt{2}|P - \text{tr}P| = \Delta P^\psi$, which is of course the same for P in any basis
2. Assuming ΔP^ψ grows to a finite value, does $\psi(\boldsymbol{x})$ at all points tend to a common value in the long time limit, i.e. is there a spontaneous rotational symmetry breaking whereby the entire system chooses an orientation?

S6.1.1. Growth of pressure asymmetry

The growth of the average value associated with a test function (of velocity) can be written as:

$$\frac{d}{dt}\langle\phi\rangle_f = \int_{\mathcal{R}^2}\int_{\mathcal{S}} \phi(\boldsymbol{v})\mathcal{C}[f](\boldsymbol{v},t)d\boldsymbol{v}d\hat{n} \tag{S92}$$

where $\mathcal{C}$ is the collision operator. This integral is substantially difficult to evaluate in large part because of the presence of the collision kernel (which describes the geometry of collisions) inside the operator $\mathcal{C}$. Also required is an analytical form for the probability distribution function, f. From prior sections, we know that f^d for a given diameter function D is ultimately given by:

$$f^d = \left(a + \boldsymbol{b}\cdot\boldsymbol{p} + c|\boldsymbol{p}|^2\right)\frac{1}{D}f^{\text{MB}}. \tag{S93}$$

Note that $\boldsymbol{p}$ is being used here, a slight modification from prior expressions that changes the values of the required constants $[a, b_x, b_y, c]$ (which can be determined by a linear system of moment equations). In section S6.2 these constants have been worked out for a controller gas with diameter function as written in eq. S87. Notably, $b_x = b_y = c = 0$ for this choice of diameter function, and the remaining constant takes the simple form of

$$a = \frac{2D_xD_y}{D_x + D_y} \tag{S94}$$

The test function of interest here is the difference in pressure components (in the diagonal basis),

$$\phi = p_xp_x - p_yp_y, \tag{S95}$$

or

$$\phi = p^2\cos^2\theta - p^2\sin^2\theta = p^2(1-2\sin^2\theta) \tag{S96}$$

in polar coordinates.

Since the steady-state form of the distribution function is known, we might avoid the difficulty of handling eq. S92 and compute the pressure asymmetry in the long-time limit directly as:

$$\Delta P^{(N)} = \int_0^\infty dp \int_0^{2\pi} d\theta p^3 (1-2\sin^2\theta)(a + \boldsymbol{b}\cdot\boldsymbol{p} + c|\boldsymbol{p}|^2)\frac{f^{\mathrm{MB}}}{D(\theta)} \tag{S97}$$

As the constants $[a, b_x, b_y, c]$ are not functions of $\boldsymbol{p}$, this integral closely resembles those that were used to compute the constants. The simple form of the constants for the specific diameter function of interest here permits a very simple equation for the pressure asymmetry (see section S6.5):

$$\Delta P^{\psi} = \rho\left(\frac{2kT}{m}\right)\frac{2}{\pi}\frac{D_y - D_x}{D_y + D_x} \tag{S98}$$

where D_x and D_y are the constant values of the diameter within the arcs of the diameter function. How does equation S98 grow with time, when one of the constant diameter values is changed? By setting $D_x = D_0$ and $D_y = D_0 + \delta$, we see that

$$\Delta P^{\psi} = \rho\left(\frac{2kT}{m}\right)\frac{2}{\pi}\frac{\delta}{2D_0 + \delta}. \tag{S99}$$

Assuming that $\delta(t)$, and that $d\delta/dt$ is much slower than gas collision dynamics, the time evolution of the pressure asymmetry is

$$\frac{d\Delta P^{\psi}}{dt} = \rho\left(\frac{2kT}{m}\right)\frac{2}{\pi}\left(1 - \frac{\delta}{2D_0+\delta}\right)\frac{1}{2D_0+\delta}\frac{d\delta}{dt} \tag{S100}$$

Which has a limit of $\mathrm{sign}(d\delta/dt)0$ for $\delta \to \infty$. The value of pressure asymmetry has the limit

$$\lim_{\delta\to\infty}\Delta P^{\psi} = \frac{2}{\pi}\left(\frac{\rho 2kT}{m}\right) \tag{S101}$$

This simple, monotonic form of the pressure differential paints a clear picture - even infinitesimal differences between D_x and D_y cause the leading eigenvalue of the pressure tensor (i.e. P^{ψ}_{xx}) to grow at the expense of the other eigenvalue (i.e. P^{ψ}_{yy}). This picture is compatible with the single relaxation time approximation of the collision operator. In our case, this approximation reads

$$\mathcal{C}_{\mathrm{lin}} = \frac{1}{\tau}\left(\frac{af^{\mathrm{MB}}}{D(\theta-\psi)} - f\right). \tag{S102}$$

As for all moments (ϕ), this produces $d\phi/dt = (\tilde{\phi} - \phi)/\tau$, where $\tilde{\phi}$ is the value of the moment over the reference distribution (i.e., eq. S93). This yields an equation for each term of the pressure tensor in the lab frame:

$$\partial_t P_{ij} = \frac{1}{\tau}\left(\tilde{P}ij - P_{ij}\right) \tag{S103}$$

$$\tilde{P} = R(-\psi)^T P^{\psi} R(-\psi) \tag{S104}$$

$$P^{\psi} = \begin{pmatrix} P_0 + \Delta P^{\psi}/2 & 0 \\ 0 & P_0 - \Delta P^{\psi}/2 \end{pmatrix} \tag{S105}$$

where P_0 is the pressure of the gas in the absence of activity. In order to write out an explicit expression for $\tilde{P}_{ij}$, we need to rotate the diagonalized pressure tensor P^{ψ} (upon which the activity is acting to cause $P^{\psi}_{xx} \neq P^{\psi}_{yy}$) by $-\psi$.

This is because ψ is defined to be the angle that diagonalizes the lab-frame pressure tensor. Note that ΔP^ψ is given by eq. S98. We begin by explicitly writing the terms of $\tilde{P}$:

$$\tilde{P}_{xx} = P_0 + \frac{1}{2}\Delta P^\psi \cos 2\psi \tag{S106}$$

$$\tilde{P}_{xy} = \Delta P^\psi \cos\psi \sin\psi \tag{S107}$$

$$\tilde{P}_{yy} = P_0 - \frac{1}{2}\Delta P^\psi \cos 2\psi \tag{S108}$$

Into which we now substitute the expression for the diagonalization angle, $\psi = \text{atan2}(2P_{xy}, P_{xx} - P_{yy})$. Note that the pressure terms in the definition of ψ are in the laboratory frame. We use the identities

$$\cos(\arctan(x)) = \left(x^2+1\right)^{-1/2} \tag{S109}$$

$$\cos(\arctan(x)/2)\sin(\arctan(x)/2) = \frac{x}{2}\left(x^2+1\right)^{-1/2} \tag{S110}$$

Substituting these yield:

$$\tilde{P}_{xx} = P_0 + \frac{1}{2}\Delta P^\psi \left(\frac{4P_{xy}^2}{(P_{xx}-P_{yy})^2}+1\right)^{-1/2} \tag{S111}$$

$$\tilde{P}_{xy} = \Delta P^\psi \frac{P_{xy}}{P_{xx}-P_{yy}} \left(\frac{4P_{xy}^2}{(P_{xx}-P_{yy})^2}+1\right)^{-1/2} \tag{S112}$$

$$\tilde{P}_{yy} = P_0 - \frac{1}{2}\Delta P^\psi \left(\frac{4P_{xy}^2}{(P_{xx}-P_{yy})^2}+1\right)^{-1/2} \tag{S113}$$

Finally, the explicit equations for the time evolution of the pressure tensor in the laboratory frame are:

$$\partial_t P_{xx} = \frac{1}{\tau}\left(P_0 + \frac{\Delta P^\psi}{2\sqrt{\frac{4P_{xy}^2}{(P_{xx}-P_{yy})^2}+1}} - P_{xx}\right) \tag{S114}$$

$$\partial_t P_{xy} = \frac{1}{\tau}\left(\left(\frac{P_{xy}}{P_{xx}-P_{yy}}\right)\frac{\Delta P^\psi}{\sqrt{\frac{4P_{xy}^2}{(P_{xx}-P_{yy})^2}+1}} - P_{xy}\right) \tag{S115}$$

$$\partial_t P_{yy} = \frac{1}{\tau}\left(P_0 - \frac{\Delta P^\psi}{2\sqrt{\frac{4P_{xy}^2}{(P_{xx}-P_{yy})^2}+1}} - P_{yy}\right) \tag{S116}$$

The term ΔP^ψ is given by eq. S98 and therefore a parameter we are free to choose. In the main text, we express the relations given by eqs. S114, S115, S116 compactly using the traceless part of the pressure tensor $Q_{ij} = P_{ij} - (P_{kk}/2)\delta_{ij}$ (in these equations, ΔP^ψ appears as Q_0). See section S7 for a development of that representation.

When only considering one point in the gas, the orientation of the pressure asymmetry does not evolve with time. As soon as $D_x \neq D_y$, local fluctuations set the orientation of $\psi(\boldsymbol{x})$, and the leading eigenvalues in that basis begin to grow. However, in an extended system, the diagonal basis for the pressure tensor is a function of location, $\psi(\boldsymbol{x})$. How do nearby regions undergoing this eigenvalue growth impact each other? If coupling between nearby locations is strong enough, $\psi(\boldsymbol{x})$ will approach a single value at all locations, analogous to a magnetic or nematic transition.

S6.1.2. Coupling of ψ

We will consider an extended system where $\psi = 0$ (i.e. the lab frame and diameter function frames are aligned) everywhere except for a small region. In this region, a local fluctuation has resulted in $\psi \neq 0$. Using ψ to denote the orientation of the diagonal pressure tensor in the small region, we are interested in the equation $d\psi/dt$ - will the misalignment grow or decay?

The total derivative of the two-argument arctangent is

$$\frac{d}{dt}\text{atan2}(y,x) = \frac{x(dy/dt) - y(dx/dt)}{x^2+y^2}, \tag{S117}$$

so we will naturally focus on the time evolution of the quantities P^i_{xy} and ΔP^i. The superscript indicates that these are pressure components of the isolated region (with misaligned ψ) in the laboratory frame. To begin, we can express the components of P^i using the angle ψ through the change of basis transformation $P^i = R^T(-\psi)P^\psi R(-\psi)$,

$$\Delta P^i = \Delta P^\psi \cos 2\psi - 2P^\psi_{xy}\sin 2\psi \tag{S118}$$

$$P^i_{xy} = P^\psi_{xy}\cos 2\psi + \Delta P^\psi \cos\psi \sin\psi. \tag{S119}$$

Note that the terms $P^\psi_{xy} = 0$ by construction. The sum of squared arguments to the arctangent function is therefore

$$(\Delta P^i)^2 + 4(P^i_{xy})^2 = (\Delta P^\psi)^2 \tag{S120}$$

The full derivative is then

$$2\frac{d\psi}{dt} = \frac{\partial\psi}{\partial 2P^i_{xy}}\frac{\partial 2P^i_{xy}}{\partial t} + \frac{\partial\psi}{\partial\Delta P^i}\frac{\partial\Delta P^i}{\partial t} = \frac{2\cos 2\psi}{\Delta P^\psi}\frac{\partial P^i_{xy}}{\partial t} - \frac{2\sin\psi\cos\psi}{\Delta P^\psi}\frac{\partial\Delta P^i}{\partial t} \tag{S121}$$

Now we consider the time derivatives of the ΔP^i and P^i_{xy} terms. Each contains a contribution from the local pumping, as well as coupling to other nearby regions by gradient relaxation. Here, we are considering an isolated region that is surrounded by a uniform bulk. We will indicate the pressure components of this uniform bulk by P^B. Since we are considering a bulk with pressure asymmetry axis aligned with the lab coordinate frame, $P^B_{xy} = 0$.

We will take a linear approximation of the coupling between nearby pressure tensors. The time derivatives then take the form

$$\partial_t P^i_{xy} = \partial_t\Delta P^\psi \cos\psi\sin\psi + \frac{1}{\omega}\left(P^B_{xy} - P^i_{xy}\right) = \left(\partial_t\Delta P^\psi - \frac{\Delta P^\psi}{\omega}\right)\cos\psi\sin\psi \tag{S122}$$

$$\partial_t\Delta P^i = \partial_t\Delta P^\psi\cos 2\psi + \frac{1}{\omega}\left(\Delta P^B - \Delta P^i\right) = \left(\partial_t\Delta P^\psi - \frac{\Delta P^\psi}{\omega}\right)\cos 2\psi + \frac{\Delta P^B}{\omega} \tag{S123}$$

where ω is a linear coupling constant. Substituting equations S122 and S123 into equation S121 results in cancellation of most terms, leaving

$$\frac{d\psi}{dt} = -\frac{1}{\omega}\frac{\Delta P^B}{\Delta P^\psi}\cos\psi\sin\psi, \tag{S124}$$

which can be expanded for small ψ as

$$\frac{d\psi}{dt} \approx -\frac{1}{\omega}\frac{\Delta P^B}{\Delta P^\psi}\psi. \tag{S125}$$

Using appropriate boundary conditions ($t = 0$,$\psi = \psi_0$ and $t = \infty$,$\psi = 0$), eq. S125 is solved by

$$\psi(t) = \psi_0 \exp\left[-\frac{1}{\omega}\frac{\Delta P^B}{\Delta P^\psi}t\right]. \tag{S126}$$

By construction, ΔP^ψ and ΔP^B are positive, so eq. S126 indicates that small deviations from alignment with nearby regions decay, for any finite value of coupling strength $1/\omega$. The relative strength of pressure asymmetry in between the two regions and their coupling strength determine the timescale of alignment.

S6.2. Analytical constants for nematic flocking diameter function

We will use a slight modification of our method of moments to work in the co-moving frame, with a benefit being significant simplification of the required Gaussian integrals. In general, we will compute integrals of the form

$$\int_0^\infty dp \int_{\theta_1}^{\theta_2} d\theta g(\theta) p^n C e^{-bp^2} \tag{S127}$$

where C and b are obtained from the normal form of the Maxwell-Boltzmann distribution, p is the magnitude of $\boldsymbol{p}$, and $g(\theta)$ is one of several trigonometric functions, depending on which components of $\boldsymbol{p}$ are involved in the moment. The integration over θ is sector-wise, where the value of the diameter function is constant in each sector.

As detailed earlier, the solution begins with constructing the matrix of moments, $\mathcal{A}$. The components of this matrix for equation S87 are listed here. For convenience, we define the following symbols which only depend on D_x and D_y:

$$\Delta = \frac{1}{D_x} + \frac{1}{D_y} \tag{S128}$$

$$\phi_x = \frac{1}{D_x}\left(\frac{1}{2} + \frac{1}{\pi}\right) + \frac{1}{D_y}\left(\frac{1}{2} - \frac{1}{\pi}\right) \tag{S129}$$

$$\phi_y = \frac{1}{D_x}\left(\frac{1}{2} - \frac{1}{\pi}\right) + \frac{1}{D_y}\left(\frac{1}{2} + \frac{1}{\pi}\right) \tag{S130}$$

$$\mathcal{A}_{11} = \int\int \frac{p}{D(\theta)} f^{\mathrm{MB}} dp d\theta = \frac{\rho}{2}\Delta \tag{S131}$$

$$\mathcal{A}_{12} = 0 \tag{S132}$$

$$\mathcal{A}_{13} = 0 \tag{S133}$$

$$\mathcal{A}_{14} = \int\int \frac{p^3}{D(\theta)} f^{\mathrm{MB}} dp d\theta = \rho\frac{kT}{m}\Delta \tag{S134}$$

$$\mathcal{A}_{22} = \int\int \frac{p^3 \cos^2\theta}{D(\theta)} f^{\mathrm{MB}} dp d\theta = \rho\frac{kT}{m}\phi_x \tag{S135}$$

$$\mathcal{A}_{23} = \int\int \frac{p^3 \cos\theta \sin\theta}{D(\theta)} f^{\mathrm{MB}} dp d\theta = 0 \tag{S136}$$

$$\mathcal{A}_{24} = 0 \tag{S137}$$

$$\mathcal{A}_{33} = \int\int \frac{p^3 \sin^2\theta}{D(\theta)} f^{\mathrm{MB}} dp d\theta = \rho\frac{kT}{m}\phi_y \tag{S138}$$

$$\mathcal{A}_{34} = 0 \tag{S139}$$

$$\mathcal{A}_{44} = \int\int \frac{p^5}{D(\theta)} f^{\mathrm{MB}} dp d\theta = 4\rho\left(\frac{kT}{m}\right)^2 \Delta \tag{S140}$$

S6.3. Inversion of $\mathcal{A}$

Obtaining an expression for the velocity distribution function requires that $\mathcal{A}$ be inverted, which we will accomplished using the adjugate matrix.

First, the determinant of $\mathcal{A}$ is given by,

$$\det\mathcal{A} = \left(\frac{\rho kT}{m}\right)^4 \Delta^2 \phi_x \phi_y. \tag{S141}$$

Below are the terms of the adjugate of $\mathcal{A}$, $\tilde{\mathcal{A}}$. Note that terms $\tilde{\mathcal{A}}_{22}$, $\tilde{\mathcal{A}}_{23}$, and $\tilde{\mathcal{A}}_{33}$ will not contribute to the final expressions (because $\langle (v_x - u_x) f^{\mathrm{MB}} \rangle = \langle (v_y - u_y) f^{\mathrm{MB}} \rangle = 0$), and so are not listed.

$$\tilde{\mathcal{A}}_{11} = \rho^3 \left(\frac{kT}{m}\right)^4 4\Delta \phi_x \phi_y \tag{S142}$$

$$\tilde{\mathcal{A}}_{12} = 0 \tag{S143}$$

$$\tilde{\mathcal{A}}_{13} = 0 \tag{S144}$$

$$\tilde{\mathcal{A}}_{14} = -\rho^3 \left(\frac{kT}{m}\right)^3 \Delta \phi_x \phi_y \tag{S145}$$

$$\tilde{\mathcal{A}}_{24} = 0 \tag{S146}$$

$$\tilde{\mathcal{A}}_{34} = 0 \tag{S147}$$

$$\tilde{\mathcal{A}}_{44} = \rho^3 \left(\frac{kT}{m}\right)^2 \frac{\Delta \phi_x \phi_y}{2} \tag{S148}$$

S6.4. Conserved constants

The conserved constants will define the distribution, as

$$f^d = \frac{a + \boldsymbol{b} \cdot \boldsymbol{p} + c|\boldsymbol{p}|^2}{D} f^{\mathrm{MB}} \tag{S149}$$

These constants are:

$$a = \frac{1}{\det\mathcal{A}} \left(\rho \tilde{\mathcal{A}}_{11} + \rho \frac{2kT}{m} \tilde{\mathcal{A}}_{14}\right) = \frac{2}{\Delta} = \frac{2 D_x D_y}{D_x + D_y} \tag{S150}$$

$$b_x = \frac{1}{\det\mathcal{A}} \left(\rho \tilde{\mathcal{A}}_{21} + \rho \frac{2kT}{m} \tilde{\mathcal{A}}_{24}\right) = 0 \tag{S151}$$

$$b_y = \frac{1}{\det\mathcal{A}} \left(\rho \tilde{\mathcal{A}}_{31} + \rho \frac{2kT}{m} \tilde{\mathcal{A}}_{34}\right) = 0 \tag{S152}$$

$$c = \frac{1}{\det\mathcal{A}} \left(\rho \tilde{\mathcal{A}}_{41} + \rho \frac{2kT}{m} \tilde{\mathcal{A}}_{44}\right) = 0 \tag{S153}$$

S6.5. Pressure anisotropy

The pressure difference along the principle and secondary axes for this gas is given by

$$\Delta P^{\psi} = \int (p_x^2 - p_y^2) \frac{a f^{\mathrm{MB}}}{D(\theta)} = a(\mathcal{A}_{22} - \mathcal{A}_{33}) = a \left(\frac{\rho k T}{m} \right) (\phi_x - \phi_y) = \left(\rho \frac{2kT}{m} \right) \frac{2}{\pi} \frac{D_y - D_x}{D_x D_y} \tag{S154}$$

S7. HYDRODYNAMIC EQUATIONS FOR THE NEMATIC FLOCKING DEMON GAS MODEL

We start with the Boltzmann equation $\mathscr{D}(f) = \mathscr{C}(f)$ in which $\mathscr{C}(f)$ is the collision integral and

$$\mathscr{D}(f) = \partial_t f + v_\mu \partial_{x_\mu} f + F_\mu \partial_{v_\mu} f \tag{S155}$$

in which v_μ is the microscopic velocity, and F_μ are external forces.

S7.1. Integral balance equation

Consider some quantity $A(t, \boldsymbol{x}, \boldsymbol{v})$. Integrating A times the Boltzmann equation yields the integral balance equation

$$\int A(t, \boldsymbol{x}, \boldsymbol{v}) \mathscr{D}(f) d\boldsymbol{v} = \int A \mathscr{C}(f) d\boldsymbol{v} \tag{S156}$$

The local density is $n = \int f d\boldsymbol{v}$ and we define averages as

$$n \langle X \rangle = \int X f d\boldsymbol{v}. \tag{S157}$$

It is convenient to express the LHS of the integral balance equation as follows. First, we express everything in terms of the peculiar velocity $\boldsymbol{p} = \boldsymbol{v} - \boldsymbol{u}$ where $\boldsymbol{u} = \langle \boldsymbol{v} \rangle$. Note that this does modify the meaning of $\mathscr{D}(f)$ (see [15]). Observables are then of the form $A(t, \boldsymbol{x}, \boldsymbol{p})$ and the integral balance equation (IBE) reads

$$\int A(t, \boldsymbol{x}, \boldsymbol{p}) \mathscr{D}(f) d\boldsymbol{p} = \int A \mathscr{C}(f) d\boldsymbol{p} \tag{S158}$$

In addition (see [15] §3.13)

$$\begin{aligned} \int A \mathscr{D}(f) d\boldsymbol{p} &= \frac{D n \langle A \rangle}{Dt} + n \langle A \rangle \operatorname{div}(\boldsymbol{u}) + \operatorname{div}(n \langle A \boldsymbol{p} \rangle) \\ -n \left\{ \left\langle \frac{DA}{Dt} \right\rangle + \left\langle \boldsymbol{p} \cdot \frac{\partial A}{\partial \boldsymbol{x}} \right\rangle + \left(\boldsymbol{F} - \frac{Dv}{Dt} \right) \cdot \left\langle \frac{\partial A}{\partial \boldsymbol{p}} \right\rangle - \left\langle \frac{\partial A}{\partial p_\mu} p_\nu \right\rangle \frac{\partial u_\mu}{\partial x_\nu} \right\} \end{aligned} \tag{S159}$$

in which

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla_{\boldsymbol{x}} \tag{S160}$$

S7.2. Hydrodynamic equations

Applying the IBE to $A = 1$, we obtain

$$\frac{Dn}{Dt} + n \operatorname{div}(\boldsymbol{u}) = \int \mathscr{C}(f) d\boldsymbol{p} = 0 \tag{S161}$$

We now apply the IBE to $A = p_\mu$. In this case, $\langle A \rangle = 0$, $\partial_{\boldsymbol{x}} A = 0$, $\partial_t A = 0$. We end up with

$$\rho \frac{D u_\mu}{Dt} = \rho F_\mu - \partial_\nu P_{\mu\nu} + \int p_\mu \mathscr{C}(f) d\boldsymbol{p} \tag{S162}$$

where $\rho = mn$ and

$$P_{\mu\nu} = \rho \, \langle p_\mu p_\nu \rangle \, . \tag{S163}$$

Assuming that collisions are momentum-conserving, the integral vanishes.

Finally, we apply the IBE to $A = p_\mu p_\nu$. Multiplying by m, we obtain

$$m \int p_\mu p_\nu \mathscr{D}(f) d\boldsymbol{p} = \frac{DP_{\mu\nu}}{Dt} + P_{\mu\nu}\mathrm{div}(\boldsymbol{u}) + 2\partial_\rho R_{\mu\nu\rho} + (\delta_{\nu\alpha}P_{\mu\beta} + \delta_{\mu\alpha}P_{\nu\beta})\frac{\partial u_\alpha}{\partial x_\beta} \tag{S164}$$

resulting in

$$\frac{DP_{\mu\nu}}{Dt} + 2\partial_\rho R_{\mu\nu\rho} + P_{\mu\nu}\partial_\beta u_\beta + P_{\mu\beta}\partial_\beta u_\nu + P_{\nu\beta}\partial_\beta u_\mu = m \int p_\mu p_\nu \mathscr{C}(f) d\boldsymbol{p} \tag{S165}$$

in which we have defined the generalized heat flux tensor $R_{\mu\nu\rho} = \rho \, \langle p_\mu p_\nu p_\rho \rangle \, /2$. The usual heat flux tensor is obtained by contracting appropriate indices, see below. The right-hand side is obtained using eqs. S114, S115, S116 as

$$(\partial_t P_{\mu\nu})_{\mathrm{coll}} = \frac{1}{\tau}\left[P_0\delta_{\mu\nu} - P_{\mu\nu} + (\Delta P^\psi/2)N_{\mu\nu}\right], \tag{S166}$$

in which we have defined $N = L/\|L\|_T$ where $L_{\mu\nu} = P_{\mu\nu} - \|P\|_T\delta_{\mu\nu}$ and $\|A\|_T \equiv A_{\mu\nu}A_{\nu\mu}/2$.

The collision integral does not vanish in general. Assuming that energy is conserved during collisions, one can however show that the trace ($\mu = \nu$) vanishes. This motivates us to define the traceless tensor

$$Q_{\mu\nu} = P_{\mu\nu} - \frac{1}{d}P_{\rho\rho} \tag{S167}$$

which is symmetric by construction. The trace $P_{\rho\rho}$ is related to the energy/temperature in the usual way. Namely, at zeroth order in a Chapman-Enskog expansion we have $P_{\rho\rho} = d\, nk_BT$. The tensor $Q_{\mu\nu}$ reassembles a nematic tensor: indeed, it describes the nematic alignment of velocities in the fluid. Removing the trace from the full equation, we also find

$$\frac{DQ_{\mu\nu}}{Dt} = 2\partial_\rho[R_{\mu\nu\rho} - \frac{1}{d}R_{\mu\mu\rho}] + Q_{\mu\nu}\partial_\rho u_\rho + Q_{\mu\rho}\partial_\rho u_\nu + Q_{\nu\rho}\partial_\rho u_\mu = m \int p_\mu p_\nu \mathscr{C}(f) d\boldsymbol{p} \tag{S168}$$

The RHS is the same as above (which is already traceless so no change is required).

As for the trace, it corresponds to energy conservation so taking the trace of the $P_{\mu\nu}$ equation, we find

$$\frac{ndk_B}{2}\frac{DT}{Dt} + \partial_\rho Q_\rho^{\mathrm{heat}} + P_{\mu\nu}\partial_\nu v_\mu = 0 \tag{S169}$$

where T is temperature and $Q_\rho^{\mathrm{heat}} = R_{\mu\mu\rho}$.

S8. DISCUSSION OF EXPERIMENTAL REALIZATION IN A MACROSCOPIC ROBOTIC SYSTEM

The demon gas model is intended as a toy model to highlight the distinct effects of information processing and feedback control in active systems. However, direct realizations of variable-diameter particles with feedback control are possible, particularly at the macroscopic scale (cm). A particularly illustrative example of robotic elements which change their size is presented in [16], in which 'particle robots' shrink and swell to collectively locomote and navigate obstacles. To directly mimic our simulations, some modifications of the devices presented in [16] would be needed;

1. These particles would need to exist on a low-friction surface (such as an air table) and be agitated (perhaps by oscillating bumpers at the side of the table).
2. Particles must be capable of sensing the proximity of other particles (to avoid exerting work on them with a diameter change). This could possibly be achieved with ultrasonic emitters/microphones on each unit. A demonstration of principle with a central controller could be realized with an overhead camera tracking the particles.
3. Particles must be able to actuate quickly enough to change size between collisions. However this interval is controlled by particle density and mean speed, so even relatively slow actuation could be made to fit between collisions if the particles are slow and sparse. Note also that for our model size changes need not exert work on other particles, so light, rapidly responding mechanisms would be feasible.

These real robots would be exerting work to overcome friction and change their size, however those size changes would result in a 'redirection' of a fraction of the particle's kinetic energy into biased motion. Therefore, in the limit where the mean kinetic energy of the particle gas is much larger than the energy required for actuation, we expect to recover the qualitative behaviors presented here.

S9. DESCRIPTION OF SUPPLEMENTAL VIDEOS

S9.1. Supplemental video V1

In this video, two species of demon particles are dropped into a hard-sided box. They fall under a vertical force, while a drag force (linear in velocity) causes them to slowly lose energy and settle into a layer. Feedback control of particle diameters allows the two species to separate and form a pre-programmed pattern.

S9.2. Supplemental video V2

In this video, two species of demon particles are continuously agitated while contained in a hard-sided box. Feedback control of particle diameters allow the two species to form patterns, which are changed over time. The agents can therefore write sequences of symbols, as shown here with an equation.

---

[1] J. A. Anderson, J. Glaser, and S. C. Glotzer, Computational Materials Science **173**, 109363 (2020).
[2] J. D. Weeks, D. Chandler, and H. C. Andersen, The Journal of Chemical Physics **54**, 5237 (1971), ISSN 0021-9606, URL `http://aip.scitation.org/doi/abs/10.1063/1.1674820`.
[3] R. S. Sutton and A. G. Barto, *Reinforcement Learing: An Introduction* (MIT Press, Cambridge, MA, 2018), 2nd ed.
[4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, CoRR **abs/1707.06347** (2017), 1707.06347, URL `http://arxiv.org/abs/1707.06347`.
[5] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, arXiv preprint arXiv:1506.02438 (2015).
[6] P. Ramachandran, B. Zoph, and Q. V. Le, CoRR **abs/1710.05941** (2017), 1710.05941, URL `http://arxiv.org/abs/1710.05941`.
[7] J. Toner and Y. Tu, Phys. Rev. E **58**, 4828 (1998).
[8] H. Chaté and A. Solon, Physical Review Letters **132**, 268302 (2024).
[9] B. Mahault, F. Ginelli, and H. Chaté, Physical review letters **123**, 218001 (2019).
[10] T. M. Cover and J. A. Thomas, *Elements of information theory* (Wiley-Interscience, 2006).
[11] C. Cercignani, R. Illner, and M. Pulvirenti, *The Mathematical Theory of Dilute Gases*, Applied Mathematical Sciences (Springer New York, 1994), ISBN 9780387942940.
[12] M. Kardar, *Statistical physics of particles* (Cambridge University Press, 2007).
[13] C. Cercignani, Journal of statistical physics **58**, 817 (1990).
[14] N. E. Korotkov and A. N. Korotkov, *Integrals Related to the Error Function* (CRC Press, 2020).
[15] S. Chapman and T. G. Cowling, *The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases* (Cambridge university press, 1990).
[16] S. Li, R. Batra, D. Brown, H.-D. Chang, N. Ranganathan, C. Hoberman, D. Rus, and H. Lipson, Nature **567**, 361 (2019).