

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 15-06-2017	2. REPORT TYPE Final Report	3. DATES COVERED (From - To) 16-Aug-2016 - 15-May-2017
---	--------------------------------	---

4. TITLE AND SUBTITLE Final Report: Computational Mathematics STIR: Density Functional Theory as a New Mathematical Framework for Predicting Population Flows	5a. CONTRACT NUMBER W911NF-16-1-0433
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102

6. AUTHORS Tomas A Arias, Itai Cohen	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Cornell University Office of Sponsored Programs 373 Pine Tree Road Ithaca, NY 14850 -2820	8. PERFORMING ORGANIZATION REPORT NUMBER
--	--

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 69767-MA-II.1

12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited
--

13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT World events are at a critical juncture where predictive, quantitative models of the flow of human beings from crowds at concerts to international human migration would greatly improve responses to developing crises. Current agent-based models of population dynamics can reproduce many behaviors, ranging from random milling to flocking and schooling, but often must postulate difficult to validate rules for agent interactions with each other and their environment. <i>Over the period of the contract, in contrast to such bottom-up approaches, we have developed a top-down approach</i>

15. SUBJECT TERMS Population flow, density functional theory, agent based model, human crowds
--

16. SECURITY CLASSIFICATION OF:	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT UU	UU		Tomas Arias
b. ABSTRACT UU			19b. TELEPHONE NUMBER 607-255-0450
c. THIS PAGE UU			

Report Title

Final Report: Computational Mathematics STIR: Density Functional Theory as a New Mathematical Framework for Predicting Population Flows

ABSTRACT

World events are at a critical juncture where predictive, quantitative models of the flow of human beings from crowds at concerts to international human migration would greatly improve responses to developing crises. Current agent-based models of population dynamics can reproduce many behaviors, ranging from random milling to flocking and schooling, but often must postulate difficult to validate rules for agent interactions with each other and their environment.

Over the period of the contract, in contrast to such bottom-up approaches, we have developed a top-down approach to infer the rules for mass behaviors directly from observations of local crowd density and to quantitatively predict mass behavior under new circumstances. Underlying this approach is a statistical framework employing concepts from classical density-functional theory. We demonstrated the power of this theory-based, data-driven approach in a model crowd consisting of walking fruit flies by extracting functions describing preferences for location and crowding. We then used these functions to make quantitatively accurate predictions for fly distributions in new environments, which we then confirmed experimentally to within statistical sampling errors. Finally, we have preliminary results showing how this density-functional approach can be generalized to quantify interactions from observations of groups of fish and pedestrians. Our method opens the way to a new approach for real-time measurement and prediction of mass behavior that may find uses in fields ranging from spatial ecology and active matter to demography, economics, and crowd and population management and security. Finally, the work was presented at the annual March Meeting of the American Physical Society, and a manuscript for submission to Nature Physics is currently under preparation.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

<u>Received</u>	<u>Paper</u>
-----------------	--------------

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

<u>Received</u>	<u>Paper</u>
-----------------	--------------

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 1.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Book

TOTAL:

Received

Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Under Graduate students supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:..... 0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME

PERCENT SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See attachment

Technology Transfer

Density Functional Theory as a New Mathematical Framework for Predicting Population Flows

Abstract:

World events are at a critical juncture where predictive, quantitative models of the flow of human beings from crowds at concerts to international human migration would greatly improve responses to developing crises. Current agent-based models of population dynamics can reproduce many behaviors, ranging from random milling to flocking and schooling, but often must postulate difficult to validate rules for agent interactions with each other and their environment.

Over the period of the contract, in contrast to such bottom-up approaches, we have developed a top-down approach to infer the rules for mass behaviors directly from observations of local crowd density and to quantitatively predict mass behavior under new circumstances. Underlying this approach is a statistical framework employing concepts from classical density-functional theory. We demonstrated the power of this theory-based, data-driven approach in a model crowd consisting of walking fruit flies by extracting functions describing preferences for location and crowding. We then used these functions to make quantitatively accurate predictions for fly distributions in new environments, which we then confirmed experimentally to within statistical sampling errors. Finally, we have preliminary results showing how this density-functional approach can be generalized to quantify interactions from observations of groups of fish and pedestrians.

Our method opens the way to a new approach for real-time measurement and prediction of mass behavior that may find uses in fields ranging from spatial ecology and active matter to demography, economics, and crowd and population management and security. Finally, the work was presented at the annual March Meeting of the American Physical Society, and a manuscript for submission to *Nature Physics* is currently under preparation.

Introduction:

Invariably, any behavior involving humans such as migrations and crowd gatherings involve social factors that complicate their study and modeling, including the inability to ethically conduct controlled experiments on human populations and the difficulty of collecting data on highly stressed populations. Nevertheless, simplified models in which persons are treated as active agents governed by a simple set of equations or in which the crowd is treated as an effective fluid can give important insights relating to global behavior. However, progress using such approaches has been hampered by the difficulty of positing, quantifying and validating agent-based rules or appropriate continuum equations and the challenge of developing more rapid and nimble quantitative frameworks for encapsulating the mechanisms underlying human crowd behavior.

We report here significant progress on these issues from a two-pronged approach. First, we have developed a novel theoretical description which combines the best aspects of agent-based and continuum flow approaches by capturing the behaviors of individual agents in a mathematical description which uses the population density at different locations as the key variable whose behavior is to be studied. This not

only radically simplifies data gathering and computational cost, but also focusses the extraction of behavioral rules to precisely those rules relevant to behaviors of the population as a whole. Second, we have developed a model, living “social” system on which to test and validate our approach. Our system consists of the confinement of large collections of fruit flies to two-dimensional chambers with different environments and stimuli. We then not only are able to track the movement of individuals and the evolution of the overall population in unprecedented detail but also can apply a wide range of stimuli and study populations under extreme stress without the ethical concerns which experiments on human populations would entail. Finally, we present data indicating that our approach applies not only to our model fruit-fly system, but to other living systems, including zebrafish and humans.

Theoretical Framework — To uncover the best-suited coarse-grained description for prediction of crowd behavior, we do not impose a particular continuum description from the outset. Rather, we adopt a bottom-up approach. We begin with a realistic, albeit simple, class of agent-based descriptions that we can *solve exactly in closed form with local density as the only variable*. We then find that the density-only solutions for crowd behavior within this broad class of models lead naturally to a density-functional based description which proves ultimately to give highly accurate and detailed quantitative descriptions and predictions for crowds in a wide variety of living systems, *without the need for time-consuming detailed agent-based simulations*.

Simple model and key concepts — We base our simple class of agent-based models on utility optimization, positing that each agent individually seeks out an optimal location as judged by a combination of external influences and crowding effects. To quantify the effects of external influences such as heat, food, presence of light, feelings of safety, perceived likelihood of escape, *et cetera*, we assign to each location x a local utility value $V(x)$. To include the effects of interactions among agents, we posit that the relative favorability of each location x is affected by additional factors, such as mutual attraction and avoidance of overcrowding *et cetera*, which can be quantified according to the density $n(x)$ at that location. Accordingly, we introduce a second, density-dependent utility term $f'(n(x))$ at each location x , so that the quantity which each agent actually seeks to optimize is $V(x)+f'(n(x))$. (Our motivation for writing the derivative of $f(n)$ instead of the function itself will be apparent in our exact solution below.)

Note that, here and throughout, we follow the Physics (rather than Economics) convention, in which agents seek to *minimize* (rather than maximize) utilities. Correspondingly, we refer to the “vexation” $V(x)$ and “frustration” $f(n)$, as quantities the agents seek to minimize, with $V(x)$ playing a role analogous to an external potential-energy function.

Exact solution to deterministic model: Under conditions in which agents strictly seek to optimize their utility, agents will adjust their locations until no location x has a lower net utility $V(x)+f'(n(x))$ than any other, at which point all motion will cease and a static equilibrium is established. At this point, the population is distributed so that the utility function for all locations has taken on a constant value,

$$\mu = V(x) + f'(n(x)) . \quad (1)$$

To interpret Eq. (1), we note that this condition corresponds precisely to minimization of the following global crowd-utility function H under the constraint of fixed number of agents $N \equiv \int n(x) dA$,

$$H = F[n(x)] + \int V(x) n(x) dA, \quad (2)$$

where μ is the Lagrange-multiplier maintaining the fixed number of agents, and $F[n(x)] \equiv \int f(n(x)) dA$ is a functional (a map from an entire function $n(x)$ to a single real number) of the population density alone,

regardless of the external influences as characterized by $V(x)$.¹ From the role of μ in Eqs. (1,2), it is clear that μ is directly analogous to the chemical potential from statistical physics, which takes on a constant value across a system in which particles have freedom of movement.

The great practical utility of Eqs. (1,2) is that, once the frustration function $f(n)$ (or a more general functional $F[n(x)]$) is known, Eq. (1) (or its generalization from minimizing Eq. (2)) can be used to infer, up to an arbitrary additive constant, the vexation $V(x)$ for any new environment directly from observation of an equilibrium density $n(x)$ in that environment, *without the need for running multiple, lengthy agent-based simulations to perform parameter fitting*. Instead, one may simply evaluate $V(x)=\mu-f'(n(x))$ (or $V(x)=\mu-\delta F[n(x)]/\delta n(x)$, when using Eq. (2)). Moreover, once the vexation $V(x)$ has been inferred for a given environment from, for example, a small population *or even a single agent* in that environment, by simply changing μ and inverting Eq. (1) (or the corresponding generalization from Eq. (2)), the distribution $n(x)$ for much larger populations in that same environment can be predicted, *again without recourse to agent-based simulation*. Below we demonstrate just such capability within our population experiments.

Stochastic generalization — In general, we do not expect living agents to immediately and strictly optimize their utility, and so we expect, and indeed observe, density fluctuations in equilibrium populations. The deterministic model above posits that for each given time interval, an agent at location x will move to any available location x' with probability $P=1$ if that motion improves the net utility ($\Delta H < 0$) and probability $P=0$ if it worsens that utility ($\Delta H > 0$). To account for lack of perfection in utility optimization among agents and the fact that time may be required for an agent to make a determination to move, we include a stochastic aspect to the agent behavior, generalizing to a smooth transition probability function $P(\Delta H)$ which

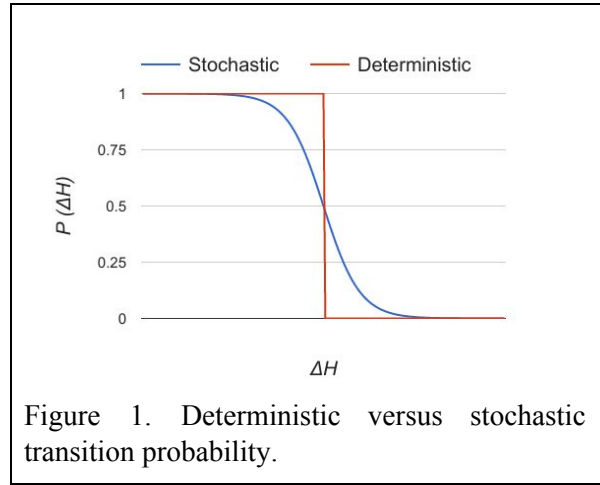


Figure 1. Deterministic versus stochastic transition probability.

varies from unity to zero as ΔH varies from negative to positive values. One very natural, symmetric choice for this function, $P(\Delta H)=1/(e^{\Delta H}+1)$ (Figure 1), is particularly convenient because it then follows mathematically² that the probability for any given arrangement of agents is then proportional to the classic Boltzmann factor $\exp(-H)$, where H is defined as in Eq. (2). Because Eq. (2) depends only on the density, we thus have a direct, detailed prediction of the density fluctuations, *without the need for agent-based simulations*, and which we can test directly against our population experiments.

Connection to experiment and predictions: To make quantitative contact with experiments, we anticipate processing video frames taken at discrete times t_i , dividing the field of view into bins $b=1\dots B$ of equal area A , counting the number of agents $N_b(t_i)$ in each bin b for each time frame t_i (Figure 2), and then

¹ We write Eq. (2) using a generic functional form $F[n(x)]$ to allow also for the possibility of more general frustration effects at each point x that may depend on more than simply the local density $n(x)$. The particular form corresponding to Eq. (1) is known as a “local density approximation” (LDA) in the density functional theory literature.

² This follows because the agents then execute exactly a form of the Metropolis algorithm for sampling the distribution $\exp(-H)$. Note that this holds regardless of the form of H , even forms which depend on the detailed arrangement of agents and not solely the density $n(x)$ as in Eq. (2).

associating with the center x_b of each bin b an estimated density $n(x_b, t_i) \equiv N_b(t_i)/A$. The fundamental observables in such experiments are thus the bin counts N_b for each video frame t_i .

To predict the probability of observing a given set of bin counts N_b , we assume the variations in $V(x)$ and $n(x)$ to be sufficiently smooth to allow discretization of the integrals for the two contributions to H . Specifically, we take $\int V(x) n(x) dA \approx \sum_{b=1}^B v_b N_b$ and $F[n(x)] \equiv \int f(n(x)) dA \approx \sum_{b=1}^B f_{Nb}$, where v_b is the average vexation $V(x)$ over bin b , $f_{Nb} \equiv f(N_b/A) A$, and we have taken the simple local-density approximation (LDA) form for $F[n(x)]$ in which agent frustration with local x depends on the density only at that location x . Finally, we note that, whereas the Boltzmann factor $\exp(-H)$ gives the probability for any given arrangement of agents, for a given set of bin counts $\{N_b\}_{b=1}^B$, there are actually $N!/(N_1! \dots N_B!)$ corresponding arrangements of the (in principle) distinguishable agents. We therefore predict the probability for any set of bin counts to be

$$P(\{N_b\}) = Z^{-1} N! \prod_{b=1}^B \exp(-v_b N_b - f_{Nb}) / N_b!, \quad (3)$$

where Z is a normalization factor.

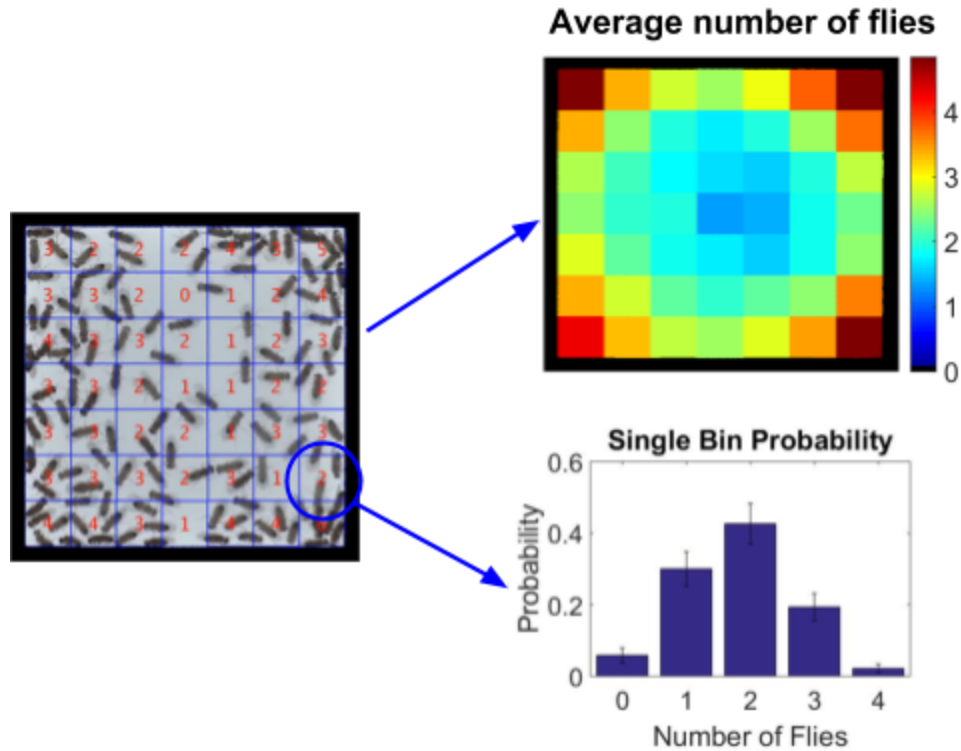


Figure 2. Analysis of experimental data, from raw video pixel images for different times t_i to bin counts at discrete times $N_b(t_i)$, and then to bin number distributions $P_b(N_b)$ and average densities $\langle N_b \rangle$.

From (3), we expect counts in different bins to be independently distributed and thus uncorrelated (apart from small corrections related to the constraint of fixed total number of agents). Moreover, we expect the distribution $P_b(N)$ of counts N in each bin b (illustrated in Figure 2) to follow the distribution

$$P_b(N) = z_b^{-1} [\exp(-v_b)^N / N!] \times \exp(-f_N), \quad (4)$$

where z_b is a normalization factor. This distribution is Poisson in nature (as expected for independently distributed agents) but modified by a multiplicative “frustration” factor $\exp(-f_N)$ that captures all impacts

of agent-agent interactions and *depends solely on the number of agents in the bin and is otherwise completely bin-independent*.

As described below, our experimental observations confirm these predictions to high accuracy and in exquisite detail. Before proceeding to the experimental results, it is worth noting the assumptions leading up to Eqs. (3,4) and how these predictions could fail in principle. First, the Boltzmann distribution $\exp(-H)$, which is the basis of Eq. (3), presumes that the measured sets of bin counts $N_b(t_i)$ for $b=1\dots B$ for each time frame t_i are drawn as independent samples from an equilibrium distribution. This requires the system to be sufficiently ergodic to allow the drawing of independent samples over the time-scale of the experiment.

Next, the discretization of the integrals leading to the separable product form of Eq. (3) and the consequent independence of counts in separate bins requires that bin sizes be chosen smaller than the typical scale of spatial variations in $V(x)$ and $n(x)$. However, if the requisite bins become so small that at most only one or two agents appear in each bin, knowledge of the count in a single bin will be insufficient to determine the local density and a non-local form will be required for $F[n(x)]$, thereby leading to correlations in the counts among bins.

Finally, it could be simply that the utility of a given point x is not separable as a sum of external influences $V(x)$ and a density-only dependent part $f'(n(x))$, or that some sort of underlying utility model does not describe the crowd behavior at all. In the former case, we would not expect to find a single bin-independent multiplicative factor $\exp(-f'_N)$ accounting for the impact of agent interactions to appear in the bin distribution functions as in Eq. (4), but rather a separate factor $\exp(-f^{(b)}_N)$ for each particular bin. In the case where no utility model applies, it is hard to imagine how a single parameter v_b for each bin and a single, bin-independent parameter f'_N for each possible bin count could accurately describe the distribution for each bin.

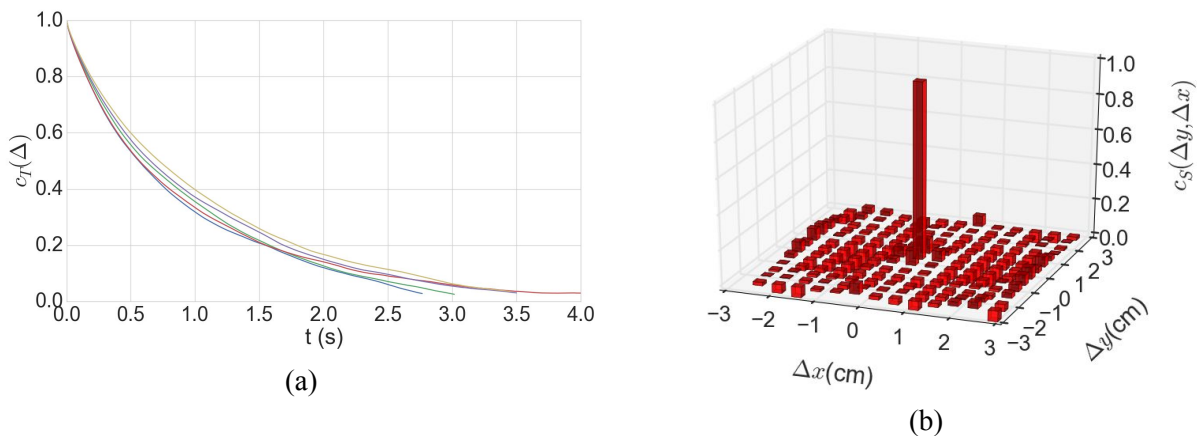


Figure 3. Global correlation functions in population of 127 flies in two-dimensional square chamber of dimensions $3.06 \text{ cm} \times 3.06 \text{ cm}$: (a) time-dependent autocorrelation of bin counts illustrating timescale of approach to equilibrium, (b) spatially-dependent correlation of bin counts confirming *predicted* statistical independence among bins.

Experimental Framework —

As a living model of a milling human crowd, we confine Wild-type male *Drosophila melanogaster* in 1.5 mm tall chambers enclosed between two acrylic plates on top of a light table. The chambers are tall

enough for flies to walk freely, but short enough to prevent flight or flies from climbing over one another. We then record overhead videos of the flies' motion for 10 minutes after being introduced into the chamber

To analyze the group behavior of the system, we divide the accessible space into bins of equal area A and count the number of flies $N_b(t_i)$ in each bin b for each frame t_i , as in Figure 2. To count the flies in each bin, we employ one of two methods. In the first, we track the individual fly positions employing custom algorithms in Matlab. The second approach that we employ underscores the advantages and simplicity of using density as the fundamental variable. For this approach, we threshold images so that only silhouettes of fly bodies are seen and then simply approximate the number of flies in each bin by dividing the number of pixels corresponding to the fly-body silhouettes by the average area of a fly and then rounding to the nearest whole number of flies. Finally, from the observed bin counts $N_b(t_i)$, we can study correlations among the bins and also construct separate count distribution functions $P_b(N)$ for each bin b .

Validation of assumptions underlying theoretical analysis —

As described above, our raw data consists in collections of bin counts at a series of discrete sample times, $N_b(t_n)$. To verify attainment of equilibrium, we consider the normalized autocorrelation function $c_r(\Delta t) \equiv \langle \sum N_b(t) N_b(t+\Delta t) \rangle_t / \langle \sum N_b(t) N_b(t) \rangle_t$, which shows the expected rapid exponential decay (Figure 3(a)) and whose integral gives the correlation time $\tau = 0.92$ s. (These data and all following, unless otherwise specified, are from a square chamber of dimension $3.06 \text{ cm} \times 3.06 \text{ cm}$ containing 127 flies.) Indeed, we find this correlation time to be quite short, typically on the order of a few seconds, for all of our experimental runs. This decay time is two orders of magnitude smaller than the typical runtime and does not vary significantly when computed in different time sub-windows, strongly suggesting rapid mixing and stationarity of the random process, thereby allowing the interchange of time and ensemble averages, and establishing the existence of equilibrium in the timescales under study. Our video frames thus represent hundreds of independent samples drawn from the equilibrium ensemble underlying our analysis

Confirmation of theoretical prediction of density distribution —

We next consider whether the bins are truly independently distributed as predicted in Eq. (3). Accordingly, we consider the normalized time-averaged spatial-correlation function $c_s(\Delta) \equiv \langle N_b(t) N_{b+\Delta}(t) \rangle_{b,t} / \langle N_b(t) N_b(t) \rangle_{b,t}$, where Δ is the two-dimensional vector displacement between bins (Figure 3(b)). The data show essentially no correlation between bins, thereby verifying the product form of the global bin distribution function in Eq. (3). This confirms not only that we have chosen appropriately sized bins but also, more fundamentally, establishes that there are little or no fly-fly interaction effects between bins, so that the local density approximation (LDA) form for the frustration, $F[n(x)] = \int f(n(x)) dA$, indeed gives a good representation of the behavior of the fly populations at the scale of our chosen 7×7 binning.

With the individual bin distributions $P_b(N)$ established as independent and ergodically sampled, it remains to verify that the individual distributions follow the predicted modified Poisson-like form of Eq. (4). Rearranging the result into an effective free-energy form $P_b(N) \equiv \exp(-A_b(N)) / N!$, we find that the *bin free energy* should follow $A_b(N) \equiv -\ln (N! P_b(N)) = \ln(z_b) + \nu_b N + f_N$, corresponding to separate, *bin-dependent* linear behavior ($\ln(z_b) + \nu_b N_b$) for each bin, along with a single, bin-independent correction (f_N) which depends only upon the number of agents N , in the particular bin at the given time.

To test the above prediction, we have performed least chi-square fits of the bin free energy $A_b(N)$ to the above form. For these fits, we determine the uncertainties in the $P_b(N)$ from the experimental counting statistics, and adjust two fit parameters for each bin b (z_b and ν_b) and a *single global, bin-independent parameter* f_N for each observed bin count $N=0 \dots N_{\max}$, where N_{\max} is the maximum number of flies observed in any bin throughout the experiment. The variety among the curves in Figure 3(a) underscore how the flies exhibit different preferences among bins. (As per Figure 2, the flies show strong preference

for the corners of the square chamber in this experiment.) Moreover, the clear non-linearities in $A_b(N)$ evident in Figure 4(a) establish that the interaction terms as described by the frustration are significant. Indeed, performing the χ^2 fit without the frustration term f_N leads to a least chi square value *per degree of freedom* value of 9.726 for 110 degrees of freedom, strongly indicating that significant effects of fly-fly interactions are evident in the data. When the fit is performed including a bin-independent frustration value f_N for each possible fly count (just 8 more adjustable parameters), the chi-square *per degree of freedom* reduces to 0.43, well within what we expect from pure statistical error³ and establishing that frustration effects are indeed bin-independent and dependent solely on the local density of flies. *This completes the full confirmation of our prediction for the detailed density-fluctuation distribution in Eq. (3) to within experimental statistical sampling uncertainty.*

Finally, to allow for a graphical illustration of the success of Eq. (4), we consider the *bin frustration functions* $f_N^{(b)} \equiv A_b(N) - (\ln(z_b) + v_b N)$, which are the bin free-energy functions but with the fitted linear bin-dependent parts $(\ln(z_b) + v_b N_b)$ removed. Our analysis above indicates that these functions should take on the bin-independent value $f_N^{(b)} = f_N$. Figure 4(b) displays the result, showing the predicted curve collapse among the distributions for all 49 bins in the data set, along with the statistical sampling error bars. The remarkably complete collapse among all 49 bins establishes the reliability of Eqs. (3,4) and reveals a smooth underlying functional form for the local frustration function $f(n=N_b/A) \equiv f_{Nb}/A$, which is “universal” in the sense that it does not depend upon the local external environment but only on the flies’ internal tendencies regarding packing

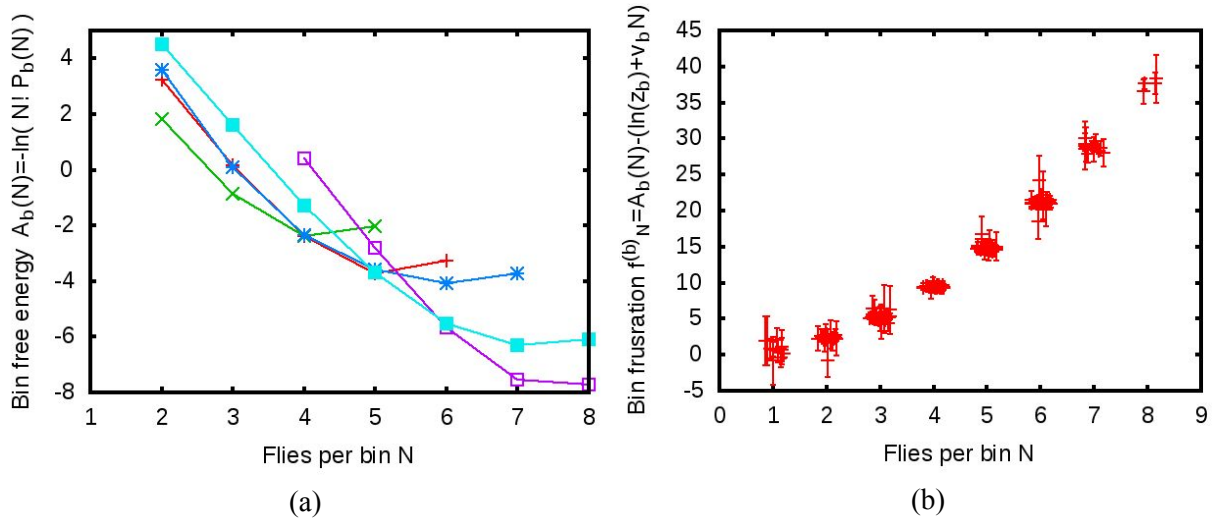


Figure 4. Bin-distribution characteristics for 127 flies in two-dimensional square chamber of dimensions $3.06 \text{ cm} \times 3.06 \text{ cm}$ divided into 7×7 square bins: (a) bin free energies $A_b(N) \equiv -\ln(N! P_b(N))$ for six representative bins, showing the wide variety of the distributions and clear nonlinearity indicative of the non-Poisson (e.g., interacting) nature of the distributions, (b) bin frustration functions $f_N^{(b)} \equiv A_b(N) - (\ln(z_b) + v_b N)$ for all 7×7 bins, whose collapse confirms Eq. (4) and the existence of a bin-independent frustration function f_N corresponding to a smooth local, *universal* frustration function $f(n)$.

³ We recognize that this χ -square value is somewhat less than what should be expected. We ascribe this to some degree of overfitting as seen by the fact that our fit parameters v_b and f_N result in smooth curves, indicating that, in principle, fits with even fewer parameters should be possible.

Density-functional theory and prediction of crowd behavior —

With our prediction for the crowd density-fluctuation confirmed in detail, we now return to the possibility of using Eq. (1) (or its generalization from minimization of Eq. (2)) to predict how large crowds distribute themselves in new environments. As per the previous discussion of Eqs. (1,2), such prediction requires *a priori* quantitative knowledge of the frustration $f(n)$, from which the vexation $V(x)$ for the new environment may be inferred from observation of a small crowd, or even of a single individual, in the new environment. Then, with both $f(n)$ and $V(x)$ known, the density $n(x)$ for a large crowd in the new environment can be predicted from inverting Eq. (1) (or its generalization from Eq. (2)). Above, we demonstrated a remarkable capability of extracting the unknown frustration function $f(n)$ to high accuracy directly from observations of crowd density fluctuations (Figure 3(b)), and so such predictions of crowd behavior should now be possible.

A significant challenge arises, however, because Eq. (1) and its extension from minimizing Eq. (2) were derived to give a final static equilibrium crowd distribution $n(x)$ in the context of a deterministic model which entirely ignores crowd fluctuations of precisely the kind we observe, so that this predictive capability does not apply directly. Related to this issue is the fact that although we have confirmed our prediction that $P[n(x)] \propto \exp(-H[n(x)])$, each of the crowd densities $n(x)$ in this relation is an *instantaneous* density which can take on any value in principle. To understand how a crowd distributes itself in a new environment, we would rather predict a different quantity, the average equilibrium crowd density over time $n_{\text{ave}}(x)$.

Prediction of the average equilibrium density $n_{\text{ave}}(x)$ of a fluctuating many-body system which follows the Boltzmann distribution $P \propto \exp(-H)$ for some Hamiltonian H is the subject of Density-Functional Theory (DFT) from the field of Condensed Matter Physics. Specifically, Ebner's classical physics extension [Ebner] of the Hohenberg-Kohn-Mermin [HK,KM] theorems of density-functional theory applies to any such system in which the the coupling to the environment appears in the Hamiltonian H as an additive linear term of the form $\int V(x) n(x) dA$, precisely as in Eq. (2). In particular, it follows from Ebner that there exists a density functional $G[n_{\text{ave}}(x)]$ dependent solely on the agent-agent interactions, such that the *average equilibrium density* $n_{\text{ave}}(x)$ can be determined by minimization of $G[n_{\text{ave}}(x)] + \int V(x) n_{\text{ave}}(x) dA$.⁴

For the special case of interactions described according to the local-density form $F[n(x)] = \int f(n(x)) dA$, which we also validated above for our crowd experiments, the resulting functional G is also local, $G[n_{\text{ave}}(x)] = \int g(n_{\text{ave}}(x)) dA$, with $g(n_{\text{ave}})$ computable directly from our experimentally extracted frustrations $f(n)$.⁵ Given that a local frustration function describes our population model well, we thus obtain the average equilibrium density condition, $\mu = V(x) + g'(n_{\text{ave}}(x))$, so that measurements of the average density $n_{\text{ave}}(x)$ from a single measurement on a small population or individual can be combined with our knowledge of $g(x)$ determine the vexation $V(x)$ for a new environment. By increasing the value of μ and inverting the same relation, we can then predict the average distribution $n_{\text{ave}}(x)$ of a large crowd in the new environment.

To test our prediction capability, we observe the motion of *a single fly* in a chamber to determine the corresponding average density for a single fly in a new environment. From this measurement, we have

⁴ As a technical point, Ebner's generalization and the conclusions of this paragraph to this point hold true regardless of the form of the interactions among objects, which need not even be of the purely density-dependent form of $F[n(x)]$ in Eq. (2), but may require for their expression explicit two-, three- and higher-order many- body terms.

⁵ Specifically, for the local-density case, $g(n_{\text{ave}}) A$ is the Legendre transformation with respect to μ of $\Sigma_N \exp(-\mu N - f_N)$. (Here, A is the area of each bin.)

two pathways to predict the average density of a large crowd in the new environments. First, we can ignore fly-fly interactions, treat the flies in the crowd as uncorrelated entities, and simply scale the observed single or small-number fly density by the number of flies in the large crowd. Second, we can account accurately for fly-fly interactions using our density-functional theory procedure. Specifically, we can employ our knowledge of the frustration $f(n)$ (extracted from experimental observations as in Figure 4(b)) to infer the vexation $V(x)$ for the new environment from small-crowd density. Then, we can use this experimentally extracted frustration and vexation information to predict the average density $n_{ave}(x)$ for a large crowd of flies in the chamber.

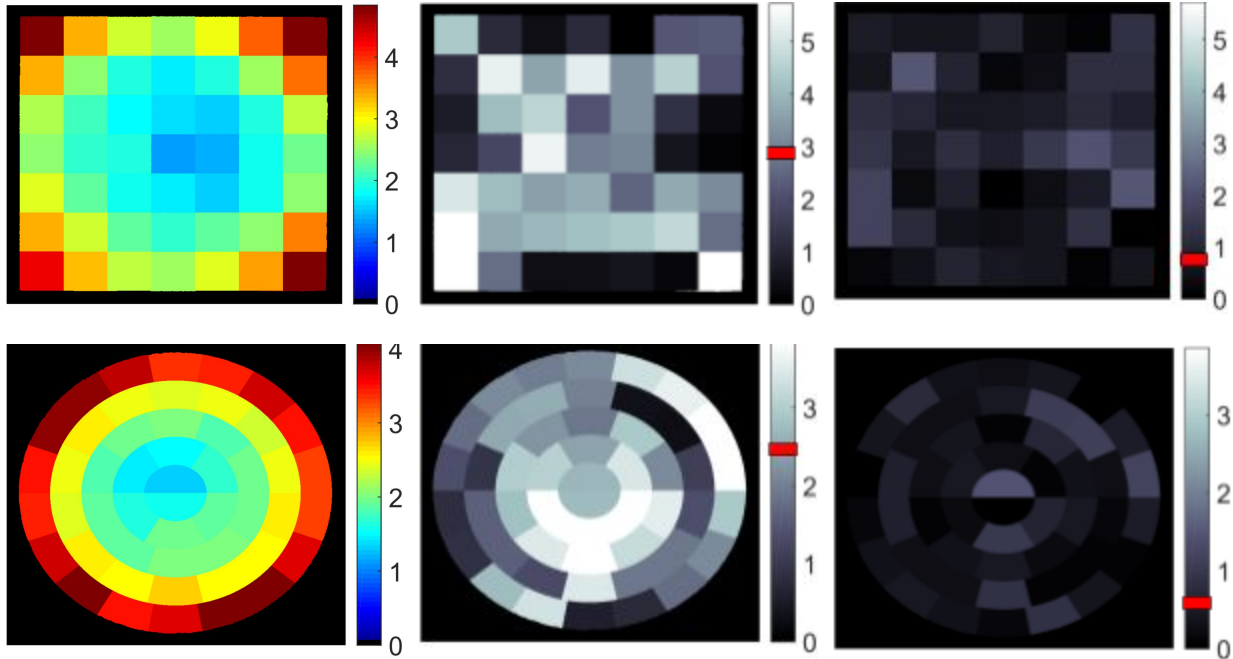


Figure 5. Prediction of distribution of large crowds in new environments: 127 flies in a square $3.06 \text{ cm} \times 3.06 \text{ cm}$ chamber (top row) and 130 flies in a circle chamber of radius 3.5 cm (bottom row), experimentally observed density (in units of flies per bin) of large crowd in environment (first column), absolute error in prediction from ignoring interactions and scaling small-crowd density (1 fly for square, 21 flies for circle) by crowd size (second column), absolute error in prediction from density-functional theory (third column), average error per bin across bins for each experiment (horizontal red bars on grey scale bars). Errors measured in units of the uncertainty σ for each bin that result from experimental counting statistics. Bin areas equal in both experiments. (Circular chamber appears elliptical due to linear distortion in plot rendering.)

Figure 5 compares both of these prediction methods with the actual observed density for two different environments, a square chamber of size 3.06 cm (top row) and a circular chamber of radius 3.5 cm (bottom row). The figure shows the experimentally observed average densities for large crowds (127 and 130 flies, respectively) in each of these environments (first column), and the comparison from the two prediction methods (second and third columns, respectively, are simple scaling and our density-functional theory procedure) for extrapolating from the behavior of a single fly (square chamber) and a small group of 21 flies (circular chamber). Figure 6 presents a more detailed comparison of these approaches for the case of the square chamber experiment after applying symmetry to improve the sampling statistics. The comparisons from Figures 5 and 6 show that the simple scaling approach leads to

errors far outside the experimental uncertainty from counting statistics, whereas our density-functional approach fully accounts for interactions and gives highly accurate predictions to within the experimental uncertainty. Although we do note that the density-functional theory extrapolation from the small crowd (21 flies) is somewhat more accurate than extrapolating from the behavior of an isolated individual, our ability to predict the behavior of large crowds *without lengthy agent-based simulations* by the computationally trivial inversion of the relation $\mu = V(x) + g'(n(x))$ is truly remarkable, especially considering that we are extrapolating by over two orders of magnitude, from the behavior from a single individual to a crowds of 127 individuals in the case of the square chamber.

Frustration as a quantifier of crowd mentality —

Whereas vexation $V(x)$ varies from one environment to the next, the frustration functional $F[n(x)]$ is *universal* in that it depends only upon the nature of the interactions between individuals and is otherwise independent of the local environment. Frustration therefore provides a quantification of the overall state or “mood” of a crowd. For example, we expect that as a crowd transitions from a relaxed state to a panicked state, that the agents may change the minimum distances that they prefer to tolerate from each other. Such a change in state may not be apparent from an overall view of the crowd because the average crowd density across the environment $\langle n_{ave}(x) \rangle = N/A_{tot}$ is set by the total number of individuals N and the total area A_{tot} of the environment. However, a shift in frustration, as we have seen, will manifest itself in the form of a change in the nature of the fluctuations in the number of individuals in each bin. Our above demonstrated ability to extract and frustration directly from experimental data affords the opportunity to quantitatively anticipate changes in the quantify the state or “mood” of a crowd in real time, simply from observations of fluctuations in the local crowd density.

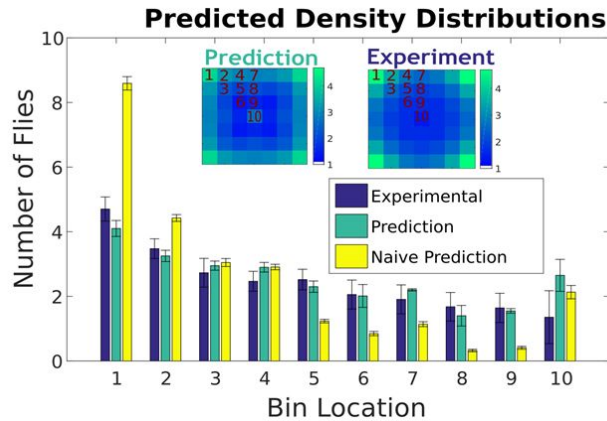


Figure 6. Comparison in square chamber of large-crowd measurements (blue bars) with prediction by ignoring interactions and scaling average density from single-fly experiment (yellow bars) and by employing density-functional theory (green bars). Parameters and analysis as in Figure 5. Here, to improve experimental sampling statistics and to simplify the display, single-individual and large-crowd densities are first symmetrized with the full square symmetry group, leaving 10 distinct bins.

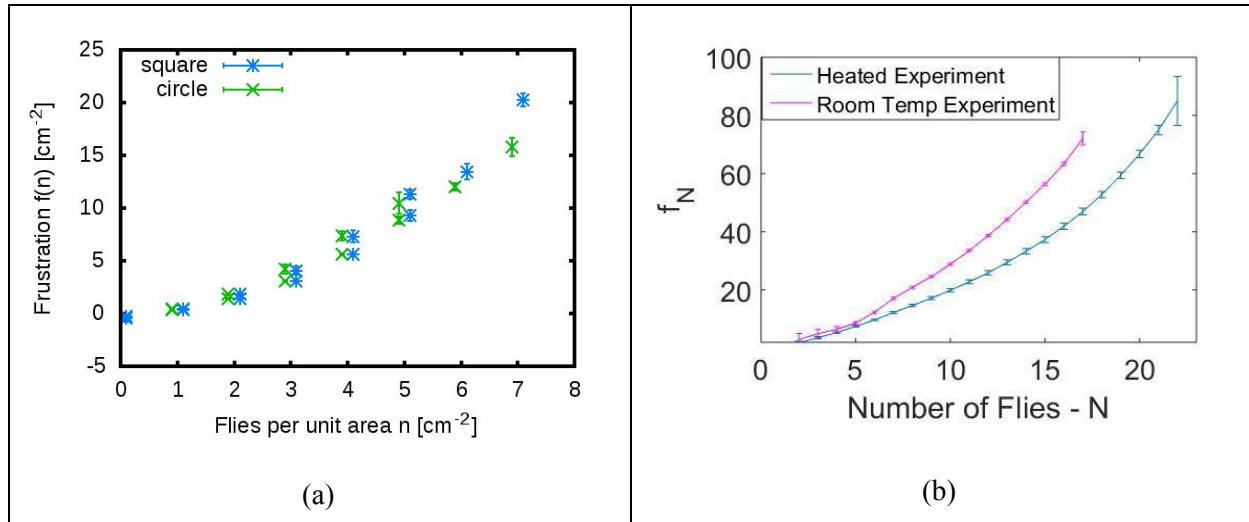


Figure 7. Exploration of “mood” dependence of frustration: (a) frustration extracted from experiments on flies in square (blue) and circular (green) chambers of varying population (53 and 125, and 52 and 130, respectively, for the square and circular chambers) but otherwise identical preparation, and (b) from an experiment in a staircase shaped chamber in which the flies are driven to panic by exposure to high heat (blue curve) and then returned to room temperature and allowed to relax . The agreement among the normal preparation square and circle experiments shows the reproducibility and stability of the extracted frustration as a measure of crowd “mood”, whereas the staircase chamber experiments show that high-heat induces a change in psychological state that can be reflected in changes in frustration.

Figure 7 explores the dependence of frustration $f(n)$ on “mood” by comparing the frustrations extracted from six different scenarios. The first four scenarios (Figure 7(a)) are of flies given the same preparation (placed into the chamber while unconscious and then observed for ten minutes after waking up) in both square and circular chambers, each at a high and intermediate density of flies. These four frustrations show remarkable agreement with each other, demonstrating that frustration is indeed stable and reproducible for crowds of different densities and in different environments but in the same “mood”. The final two scenarios (Figure 7(b)) reflect a different state of “mood” of the crowd. Here, the flies are in a staircase shaped chamber, where they are first heated to a high temperature, which creates extreme panic as the flies search desperately for an escape route from the chamber, and then allowed to relax as the chamber is returned to room temperature. While panicked, the flies show little care for overcrowding and so the frustration allows for higher local densities by exhibiting notably less upward curvature.

Finally, we have succeeded in applying our approach to other species than fruit flies, including zebrafish and walking groupings of humans. Figure 8(a) compares the zebrafish frustration $f(n)$ taken from two experiments in a rectangular tank, one with uniform lighting and one with two regions of shade, which the fish find attractive and thus correspond to regions of lower vexation $V(x)$. Despite the change in the environment, the frustration remains stable. We note also that zebrafish, unlike fruit flies, exhibit attractive “schooling” behavior favoring larger densities, behavior which the frustration is able to capture by exhibiting downward curvature. (Eventually, we expect the frustration to increase as we reach the physical packing limit for the fish, but the density of fish is too low to observe this in the experiment because the fish here are able to move in three dimensions.) Finally, Figure 8(b) shows our first results for humans, considering people walking freely along a pathway. The persons show a natural tendency to clump into small groups, which again manifests as a downward curvature in frustration.

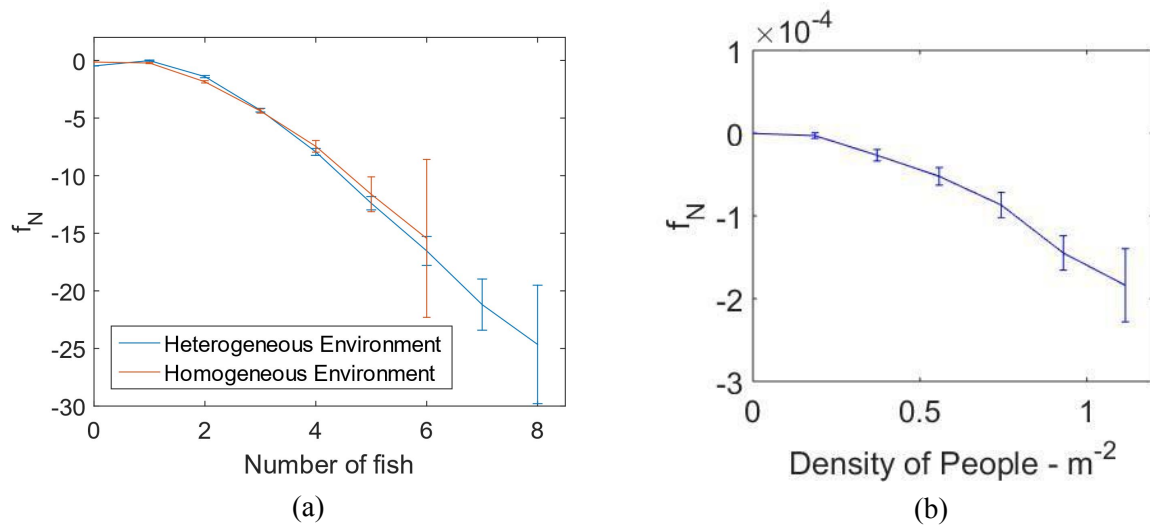


Figure 8. Frustrations extracted for other species: (a) zebrafish in a tank under uniform illumination (red curve) and with two regions of shade (blue curve), (b) humans walking in groups along a path. The zebrafish data again demonstrate the stability and environment-independence of extracted frustrations. Both zebrafish and human datasets demonstrate that attractive interactions manifest as downward curvature in the frustration.