

# Measuring the coupling strength of living organisms

Zengru Di<sup>1</sup>, Bertrand M. Roehner<sup>2</sup>

**Abstract** For any system, whether physical or non-physical, knowledge of the form and strength of inter-individual interactions is a key-information. In an approach based on statistical physics one needs to know the interaction in order to write the Hamiltonian of the system:  $H = H_{\text{free}} + H_{\text{interaction}}$ . For non-physical systems, based on qualitative arguments similar to those used in physical chemistry, interaction strength gives useful clues about the macroscopic properties of the system, e.g. for an institution the dropout rate is expected to be smaller when inter-individual attraction is stronger.

Even though our ultimate objective is the understanding of social phenomena, we found that systems composed of insects (or other living organisms) are of great convenience for investigating *group effects*. In this paper we show how to design experiments that enable us to estimate the strength of interaction in groups of insects. By repeating the same experiments with increasing numbers of insects, ranging from less than 10 to several hundreds, one is able to explore key-properties of the interaction. The data turn out to be consistent with a global interaction that is independent of distance (at least within a range of a few centimeters). Estimates of the average cross-correlation will be given for ants, beetles and fruit flies. The experimental results clearly exclude an Ising-like interaction, that is to say one that would be restricted to nearest neighbors. In the case of fruit flies the average cross-correlation appears to be negative which means that instead of an inter-individual attraction there is a (weak) repulsive effect.

In our conclusion we insist on the fact that such “physics-like experiments” on insect populations provide a valuable complement or alternative to computer simulations. When testable group effects are predicted by a model, the required experiments can be set up *within a short time*, thus permitting to confirm or disprove the model. This marks a significant progress with respect to modeling of social systems where, all too often, the requested statistical data just do not exist, thus obstructing any fruitful dialogue between theory and observation.

Key-words: Group effects, collective behavior, physical chemistry, statistical physics, living organisms, insects, ants, drosophila, coupling strength, inter-correlation, correlated random variables, exchangeable random variables.

PACS classification: Interdisciplinary (89.20.-a) + Correlation, collective effects (71.45.Gm)

1: Department of Systems Science, Beijing Normal University, Beijing, China.

Email: zdi@bnu.edu.cn

2: Institute for Theoretical and High Energy Physics (LPTHE), University Pierre and Marie Curie, Paris, France.

Email: roehner@lpthe.jussieu.fr

In a first version of the paper the title made reference to “statistical physics” rather than to “physical chemistry”. Although chemistry itself plays no role in our investigation, we think it is important to emphasize that at this point it relies rather on the approach of “physical chemistry” than on the more formal approach of statistical mechanics. The reasons for that will be explained later on<sup>1</sup>.

## Rationale and motivations

By using the theoretical framework of statistical mechanics one can derive the macroscopic properties of a system from the characteristics of its microscopic elements. This is a major achievement and so it is hardly surprising that researchers from other disciplines (e.g. biology, demography, sociology or economics) have been tempted to adapt such a powerful tool to their own field. In light of the successful record of statistical mechanics in physics there is little doubt that such extensions appear highly desirable. Yet, to our best knowledge, in spite of many attempts in this direction such attempts have not been highly successful so far<sup>2</sup>.

### Obstacles

As a matter of fact this is hardly surprising for there are indeed many obstacles.

- Statistical physics is fundamentally a theory of systems in equilibrium. For systems which are (strongly) out of equilibrium the very concept of temperature becomes meaningless.
- Statistical physics relies on the connection between ensemble averages (which are predicted theoretically) and time-averages (which are measured in experiments). This so-called ergodic hypothesis may be valid for physical systems which move from one state to another every picosecond so that there are trillions of transitions during an observation time of a few seconds. Yet, it is not obvious that such an assumption can still be accepted for socio-economic systems for which the transition rates are much slower<sup>3</sup>.
- Last but not least, one should not forget that in order to use the theoretical framework of statistical mechanics one needs to know the Hamiltonian  $H$  of the system which indicates how energy is distributed. Generally  $H$  includes three parts:

$$H = H_0 + H_{\text{inter}} + H_{\text{exo}}$$

---

<sup>1</sup>A preliminary but extended version (some 120 pages) of the present paper is available on the following website: <http://www.lptthe.jussieu.fr/roehner/effusion.pdf>. Those readers who, on the contrary, wish to get a quick account can read the vivid and clear review posted on the following website:

<http://www.technologyreview.com/view/509486/first-physics-like-experiments-for-measuring-group-behaviour-of-living-creatures/>

<sup>2</sup>Recently, some promising breakthroughs were made in this direction by a group of Japanese economists; see Aoki and Yoshikawa (2007), Iyetomi et al. (2011), Iyetomi (2012).

<sup>3</sup>The highest transition rates are probably those in currency exchange markets with hundreds of orders (worldwide) every second. Recently so-called high speed trading, that is to say transaction orders passed by computers, has reduced transition times to a few micro-seconds at least for a number of actively traded securities.

where  $H_0$  stands for the free particles,  $H_{\text{inter}}$  for the interaction energy between them and  $H_{\text{EXO}}$  for the energy of the particles when an exogenous field is involved. For instance  $H_0 \sim \sum p_i^2/m$  for a system containing the molecules of a gas,  $H_{\text{inter}} \sim \sum \frac{1}{(r_i-r_j)^6}$  when one wants to take into account the van der Waals forces between the molecules, and  $H_{\text{EXO}} \sim \sum S_i H(i)$  for the energy of a set of spins in an external magnetic field.

Whereas the third term can possibly be omitted when the experimental device can be shielded from external fields, the interaction term must *always* be taken into account<sup>4</sup>. Needless to say, there are almost no biological or social systems for which one has a clear knowledge of their interactions. It is precisely the main purpose of the present paper to explain how such interactions can be measured.

### Reasons for optimism

The previous list of obstacles could appear discouraging especially if one realizes that there are many other problems in non-physical systems just for defining key-variables such as velocity or energy. However, there are also good reasons for optimism as we will see now.

First, it can be observed that the theory of phase transitions has been used to describe the transition between ordinary hadronic matter and quark-gluon plasma. As such states are characterized by temperature of the order of  $10^{12}\text{K}$  and life-times of the order of  $10^{-20}\text{s}$ , it means that this theory is applied well beyond the limits of the phenomena<sup>5</sup> for which it was originally developed. Does the ergodic assumption hold for such extremely short time intervals? Nobody knows and probably nobody cares. The strategy of physicists is to use this framework without giving too much concern to underlying assumptions. If sensible results emerge, this will provide so to say *ex post* justification.

Secondly, it can be observed that the title of this paper does not refer to statistical physics but to physical chemistry. Why?

- Although the objective of physical chemistry is also to explain the properties of macroscopic systems in terms of molecular interactions, there are two main differences with the approach of statistical mechanics. First, physical chemistry considers a broad range of molecules rather than just the simplest ones as is done in physics. Thus, because many cases are being considered, it becomes indispensable to adopt a comparative perspective. Why is the melting point of argon lower than the melting point of water? Why is the equilibrium vapor pressure higher for ethanol than for water? And so on and so forth.

---

<sup>4</sup>Even in order to use a mean field approximation one must know the form of  $H_{\text{inter}}$ .

<sup>5</sup>E.g. second order phase transitions such as the paramagnetic-ferromagnetic transition in iron.

- Because it would be an almost impossible task to propose (and solve) full-fledged models for all these cases, physical chemistry will rather resort to qualitative arguments. For instance, a standard argument is to observe that the stronger are molecular interactions in a liquid, the fewer molecules will be able to escape which in turn will lead to a low equilibrium vapor pressure above the liquid. Whereas this argument relies on a specific mechanism describing how molecules leave the liquid, it does not require any of the assumptions that we listed previously. Even equilibrium is not strictly required. Indeed, if the container is left open, no equilibrium will take place and no equilibrium vapor pressure can be defined, but the same argument can nevertheless be used for explaining differences in the evaporation rate.

Such kind of argument can be used with success to explain many physical properties. For instance, the boiling temperature of alkanes ( $C_nH_{2n+2}$ ) is expected to increase with  $n$  because the so-called London attraction forces (due to induced polarization which create short-lived dipoles that attract one another) exist between *all* atoms and therefore, in the absence of any other force, attraction will be stronger for big molecules than for small ones. Through a similar argument one would also expect the heat of vaporization to increase with  $n$ . These predictions are indeed confirmed by experimental data; two graphs displaying such data can be found in Roehner (2004, p. 663).

In short, once one knows the strength of interaction in a system, one should be able to derive several of its macroscopic properties. Thus, we are again confronted to the same key-question: *how can we measure interaction strengths?*

The simplifications that we have already made consisted firstly in saying that we do not need to care too much about the underlying hypotheses of statistical physics, secondly that (at least in a first stage) there is no need to use the mathematical framework of statistical mechanics. Now, for the purpose of measuring coupling strength, we introduce a third simplification which concerns the kind of systems on which the experiments will be performed.

It is often said that for socio-economic systems one cannot make experiments<sup>6</sup>. However, this is only partially true. In fact, social sciences researchers are in the same position as astrophysicists. While they cannot perform *any* observation that they would like to do, nevertheless they can make such observations as allowed by the statistical data that exist<sup>7</sup>. Yet, one must recognize that in many investigations the

---

<sup>6</sup>In the discussion which follows we leave apart so-called class-room experiments that are performed with small groups of students. Such experiments can be useful to study how people will react in specific circumstances such as in response to auction rules. However, one does not see how collective behavior can truly be studied in such a way because the experiment will only reflect genuine behavior if the people are *not* told that they are involved in an experiment. In the 1970s and 1980s the psycho-sociologist Stanley Milgram performed experiments of this kind. However, such an approach raises major ethical problems and should rather be avoided.

<sup>7</sup>Researchers who have appropriate funding can even organize surveys in order to collect data that would not be

very data that one would need turn out to be unavailable. This is a serious obstacle. The task of designing appropriate measurement methods is difficult enough in itself; it would become altogether impossible if at each step progress is hindered by a lack of data.

There is a simple solution. Instead of studying people we can study populations of living organisms such as bacteria, insects or small fishes. For all these populations there exists a broad range of species. Different species will have different inter-individual interactions. Thus, one is very much in the same position as in physical chemistry. In what follows we will limit ourselves to populations of insects.

Our goal is to study groups of insects not at all as an entomologist would do but from the perspective of physical chemistry. In this respect living organisms have another important advantage over social or economic systems. Energy is a key-notion in physics. While it is not obvious how to define the “energy” of a set of stocks or a sample of companies, it is easy to define the velocity and kinetic energy of a group of ants. In other words, systems of living organisms are much closer to physical systems than are socio-economic systems.

In the next section we explain how we designed and implemented our experiments. In the following sections we propose some consistency tests of our results and we discuss their significance.

## Design of the experiments

The experiment will be described for ants but their design is fairly similar for other insects such as fruit flies or beetles.

A number  $n$  of ants are contained in a rectangular box (15cm long and 5cm wide, 4mm high) (see Fig. 1). In this box one defines two parts: an area  $A$  and the part  $B$  of the box which does not belong to  $A$ . For the sake of simplicity we can think of  $A$  as being the left-hand side of the box. However, one should keep in mind that  $A$  can also be much smaller than one half of the box. This allows to explore the behavior of the ants at smaller scales.

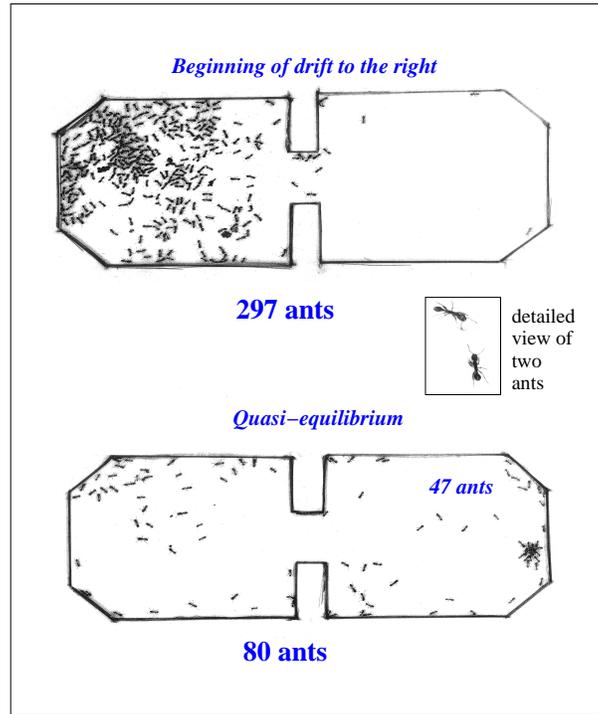
The ants can choose the compartment in which they wish to go or to stay. We record the number  $n_A(t)$  of ants which are in compartment  $A$  at time  $t$ .

The idea of the experiment is the following.

- Suppose for a moment that the movements of the ants are completely correlated. This means that if one ant goes from  $A$  to  $B$  (or from  $B$  to  $A$ ) all the others will follow. Thus at each time step  $n_A(t)$  will experience huge jumps, either from  $n$  to 0

or vice versa.

- Suppose now that there is a zero correlation between the movements of the ants. This means that if one ant goes from  $A$  to  $B$  it will not be imitated by others. Of course, other ants may make the same move but they will do so independently from one another. As a result their moves will follow a binomial process. A move of all the ants together is not completely excluded but it will occur with a probability of  $(1/2)^n$  that decreases exponentially when  $n$  increases.



**Fig. 1 Experiments with ants in a two-compartment device.** The picture shows two different initial conditions. (i) The situation where all ants are on the same side is the starting point of “evaporation” experiments. (ii) The situation with similar numbers on each side is the starting point for the measurement of the average cross-correlation.

This argument suggests that there is a connection between the standard deviation of  $n_A(t)$  and the average correlation between the movements of the ants<sup>8</sup>. Needless to say, we wish to know the mathematical form of this relationship. Then, by recording the fluctuations of  $n_A(t)$  we will be able to compute its variance and to derive the average correlation between ants. This average correlation can be considered as a measure of their interaction strength.

## Formalization

<sup>8</sup> $n_A(t)$  records only the *balance* of the moves from 1 to 2 and 2 to 1. Would it not be better to record these moves separately? After all if they happen to be identical  $n_A(t)$  will remain constant and fail to reflect those separate fluctuations. However, if the moves are independent random variables, say  $X_{12}(t)$  and  $X_{21}(t)$ , identical changes on each side are unlikely. As a matter of fact, under this assumption,  $\sigma^2(n_A(t)) = \sigma^2(X_{12}(t)) + \sigma^2(X_{21}(t)) = 2\sigma^2(X_{12}(t))$ . In other words,  $n_A(t)$  will reflect individual moves adequately.

To each ant  $i$  we associate a random variable  $X_i$  which takes the value 1 when  $i$  is in compartment  $A$  and 0 otherwise. Thus, at any moment  $t$ , the number of ants in compartment  $A$  will be given by:  $S_n = \sum_1^n X_i$ . If  $n_A(t)$  is a *stationary* random function, it is reasonable<sup>9</sup> to assume that the variance computed from the time series  $n_A(t)$  coincides with the probabilistic (i.e. ensemble) variance of the random variable  $S_n$ .

Various assumptions can be made regarding inter-individual interaction. Each assumption leads to different correlations between the  $X_i$ . We will examine two specific cases: global correlation which means that  $r_{ij} = r(X_i, X_j)$  is basically the same for all pairs  $(i, j)$  and correlations which decrease exponentially when the difference  $i - j$  increases:  $r_{ij} \sim \eta^{|i-j|}$

### Global correlation

In this case,  $\sigma^2(S_n) = \sigma^2(n_A)$  is given by the following proposition.

**Variance of a sum of correlated variables.** We consider a sum  $S_n$  of  $n$  identically distributed random variables  $X_i$  of variance  $\sigma^2$ . We assume that between  $X_i, X_j, i \neq j$  there are cross-correlations  $r_{ij}$ , the average of which is denoted by  $\bar{r} = \frac{1}{[n(n-1)/2]} \sum_{i < j}^n r_{ij}$ . Then, the variance of  $S_n = X_1 + \dots + X_n$  is given by:

$$\frac{\sigma^2(S_n)}{n\sigma^2} = (n-1)\bar{r} + 1 \quad (1)$$

The proof is fairly straightforward and is given in Appendix A.

Four observations are of interest in relation with formula (1).

(1) The factor  $n\sigma^2$  represents the variance of  $S_n$  when the variables are uncorrelated. Therefore the ratio on the left-hand side represents the variance of  $S_n$  divided by what it would be if the correlations are switched off. Subsequently, this ratio will be denoted by  $g^2$ .

(2) In the special case where  $\bar{r} = 1$ , formula (1) gives:  $\sigma^2(S_n) = n^2\sigma^2$ . This result can be confirmed by observing that  $\bar{r} = 1$  means that all variables  $X_i$  are identical that is to say take the same values (with probability 1). Thus,  $S_n = nX_1 \Rightarrow \sigma^2(S_n) = \sigma^2(nX_1) = n^2\sigma^2$

(3) A negative average correlation reduces the variance instead of increasing it. This would correspond to a repulsive force between the individuals. It is of interest to observe that  $\bar{r}$  cannot become smaller than  $-1/(n-1)$ . In this case the variance is reduced to zero. Intuitively, this corresponds to a situation where the moves of any individual are countered by the moves of the others in a way which leaves  $S_n$

---

<sup>9</sup>While of course necessary, the stationarity condition is not sufficient to guaranty ergodicity of the standard deviation. The specific mathematical condition that  $n_A(t)$  must satisfy is given in Papoulis (1965, p. 330).

unchanged.

(4) Formula (1) applies to any random variables  $X_i$ . For the problem in which we are interested, the  $X_i$  have a special meaning from which results that:

$$\sigma^2(X_i) = E(X_i^2) - E^2(X_i) = P\{X_i = 1\}1 - (P\{X_i = 1\}1)^2 = p(1 - p)$$

where  $p$  is the fraction of  $A$  with respect to the total area.

### Ising-like correlations

When the interaction is restricted to nearest neighbors as in the one-dimensional Ising model for spins, the correlation between the  $X_i$  decreases exponentially when the “distance” between the spins increases (Glauber 1963 p. 300). In this case the variance of  $S_n$  is given by the following proposition.

**Variance of a sum of Ising-like correlated variables.** We consider a sum  $S_n$  of  $n$  identically distributed random variables  $X_i$  of variance  $\sigma^2$ . We assume that between  $X_i, X_j, i \neq j$  there is a cross-correlation  $r_{ij} = \eta^{|j-i|}$ ,  $0 < \eta < 1$ . Then, the variance of  $S_n = X_1 + \dots + X_n$  is given by:

$$\frac{\sigma^2(S_n)}{n\sigma^2} = \frac{1 + \eta}{1 - \eta} - \frac{2\eta}{n(1 - \eta)^2} \quad (2)$$

The proof is fairly similar to the proof of the first proposition and it is outlined in Appendix A. According to this result, the ratio  $g^2(n) = \sigma^2(S_n)/(n\sigma^2)$  is slightly increasing when  $n$  increases (see Appendix A). However, when  $n$  becomes large the term involving  $n$  becomes negligible with respect to the first term. Thus, it is legitimate to say that for large  $n$ ,  $g^2(n)$  is almost constant.

Can one explain the difference between case 1 and 2 intuitively? We have already observed that if  $\bar{r}$  is close to 1, almost all insects will cross from one side to the other at the same time which will result in big fluctuations of  $n_A(t)$  between 0 and  $n$ . In the second model the parallel of such a high correlation would be  $\eta$  close to 1, e.g.  $\eta = 0.9$ . Yet, even with such a value of  $\eta$  the correlation between  $i$  and its neighbors will fall off rapidly when the distance increases. This means that when  $i$  will change side, only a small number ( $f$ ) of neighbors will follow. As  $f$  depends only upon  $\eta$  (and not upon  $n$ ) one sees that  $g^2(n)$  does not increase with  $n$ .

In short, for the models that we considered the ratio  $g^2(n)$  can behave in three different ways as a function of  $n$ .

- (1) It decreases linearly when  $\bar{r} < 0$
- (2) It is almost constant when  $r_{ij}$  decreases exponentially with respect to  $|i - j|$ .
- (3) It increases linearly when  $\bar{r} > 0$ .

We will see that only cases 1 and 3 occur in our observations.

## Experimental results

### Procedure

The experimental procedure involves the following steps.

- First one must spread  $n$  ants fairly uniformly in the whole container. Then pictures will be taken every  $\delta$  seconds <sup>10</sup>.
- These pictures will allow us to record the numbers  $n_A(t)$ . Once the variance of this time series has been computed one gets the ratio  $g^2(n)$ .
- By repeating this procedure for different number of ants one gets results which can be represented as a set of points  $(n - 1, g^2(n))$  (see Fig. 2).
- A linear regression performed on this set of points gives an estimate of the slope  $\bar{r}$ .

### Results

See the graphs in Fig. 2a,b,c.

An important observation is in order regarding the magnitude of the estimated average correlation. First, it must be emphasized that  $\bar{r}$  is very different from the correlation estimated from a scatter-plot. In the latter case a correlation as low as 0.01 would be non-significant (in the sense that the confidence interval would contain 0) except if the scatter-plot contains several thousand data points. Here, however, the correlation was obtained as the slope of a regression line and its estimate is quite significant as can be seen from the size of the error bars.

In order to get an intuitive understanding of  $\bar{r}$ , one should compare the actual trajectories of the insects to those shown in the simulation of Fig. 5 a,b. Broadly speaking, the comparison will reveal that at individual level the actual trajectories of the insects are even more random than those in Fig. 5b. In spite of this high degree of randomness there is an observable effect at the macro level. The situation is somewhat the same as for a gas or a liquid. In spite of the randomness of the movements of individual molecules there are nevertheless well defined macroscopic properties.

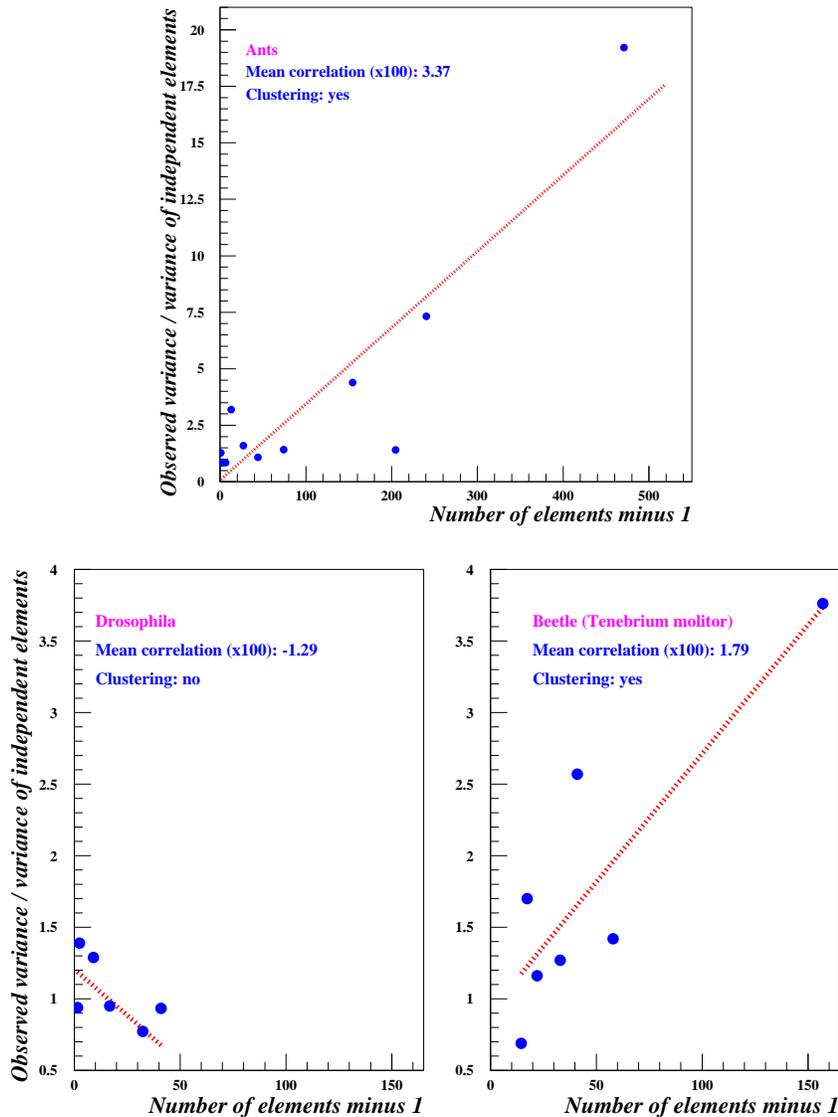
### Experimental problems

Although the procedure may appear fairly straightforward, there are a number of hurdles; while some are purely technical others are of more fundamental importance. Let us begin with the latter.

Ideally, in order to remain in a stationary equilibrium situation one would like  $n_A(t)$

---

<sup>10</sup>An “appropriate” time interval is important for the accuracy of the measurement. Of course, it is useless to take pictures when nothing happens that is to say when  $n_A(t)$  does not change. On the other hand, simulations show that one can greatly improve the accuracy of the measurement by increasing the number of pictures. In our experiments, depending on the activity and number of insects  $\delta$  was between 10s and 120s.



**Fig. 2a,b,c Relation between the variance of the number of individuals in a compartment  $A$  and the size of the group.** We suppose that the whole domain which contains the ants has been divided into two parts and we observe the fluctuations  $n_A(t)$  of the number of individuals in part  $A$ . The slope of the regression line gives an estimate of the mean correlation between the moves of individual elements. The negative correlation observed for drosophila can be interpreted as the result of repulsive forces between individuals. In the beetle case, each experiment (for a specific group size) has been repeated 10 times.

The ability to form clusters can be seen as revealing the existence of attractive inter-individual forces. Thus, this characteristics comes as a confirmation of the sign of the correlation.

The confidence intervals (at a probability level of 0.90) are as follows: ants:  $100\bar{r} = 3.37 \pm 0.9$ , drosophila:  $100\bar{r} = -1.29 \pm 1.16$ , beetles:  $100\bar{r} = 1.79 \pm 0.75$ .

*These experiments were performed between June and November 2012 in three different places, first in Paris (ants, drosophila), then in Beijing (drosophila, beetles) and finally in Kunming, Yunnan Province, China (beetles).*

to fluctuate around  $1/2$ . Actually, for ants as well as for beetles,  $n_A(t)$  can become very different from  $1/2$ . This is due to the fact that in such cases the individuals tend

to form a cluster in one part of the container. Thus, if the cluster is in  $A$ , the ratio  $n_A(t)/n$  will become close to 1, whereas it will decrease toward 0 if the cluster is in part  $B$ .

In a sense, this clustering behavior is good news because it is a direct proof of the existence of an inter-individual attraction. On the other hand, however, it introduces a bias in the measurement of  $\bar{r}$ . A correction procedure was introduced to take this effect into account.

There is a problem which arises especially for drosophila, namely the fact that once introduced in the observation device only a few of the insects will move. In the case of drosophila this may take the following form: in a group of some 50 only about 5 to 10 will move at one moment and they will do so with great speed going from one end of the container to the other without seemingly caring about the 45 others<sup>11</sup>. Another circumstance which will prevent the insects from moving is when they form a cluster. Although a correction can be introduced in the analysis to take into account such “frozen” elements, it is clear that the analysis eventually becomes meaningless when the proportion of frozen elements is too high.

The formation of clusters also leads to a more practical difficulty namely the fact that once ants are part of a cluster their spatial density becomes so high that it is difficult to count them. As they form several layers, counting becomes nearly impossible even on high resolution pictures. Recently, we have tried an alternative method which consists in weighing rather than counting. This method is well suited for small beetles whose unit weight is of the order of 15mg (see Fig 3). It is more difficult for ants whose typical weight (e.g. workers of “*Formica japonica*”) is about 3mg. It is altogether impossible for drosophila whose typical weight is around 0.2mg.



**Fig. 3: Container with weighing device on one side.** The compartments  $A$  and  $B$  are slightly (0.5mm) disjoined along the blue and red lines respectively so that the weight measured by the scale corresponds only to the beetles contained in the left-hand side part but that the beetles can nevertheless cross from  $A$  to  $B$  and vice-versa. Here most of the beetles have formed a cluster in a corner. The weight is 357 mg which, when divided by 15 mg, gives a total of 24 beetles.

## Consistency tests

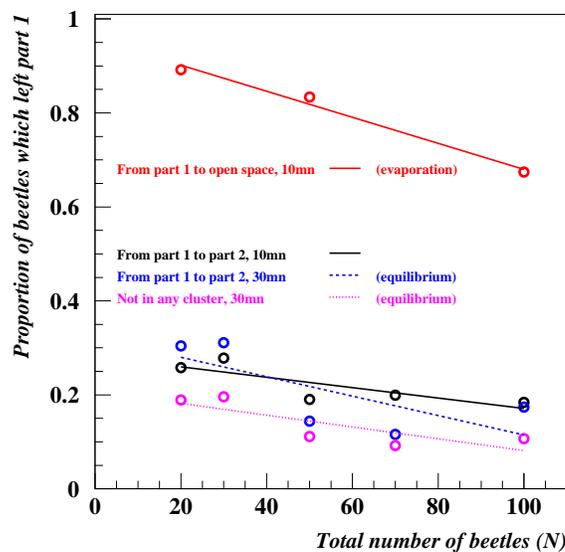
For a liquid, inter-molecular attraction can be estimated through various means and variables: evaporation rate, equilibrium pressure of vapor, boiling temperature, heat of vaporization. It is the fact that such estimates are (at least most often) consistent with one another which gives us confidence in them. One would like to do the same here.

<sup>11</sup>Whereas ants will tend to slow down or stop every time they come close to another ant.

A simple qualitative consistency test is provided by the following “evaporation” experiment. One takes a test tube containing some 50 drosophila and one makes them all move to the bottom of the tube by hitting the tube on a table. Then, very quickly<sup>12</sup> one puts the tube on the table in horizontal position. Let us assume that the bottom of the tube is on the left. After a few seconds, some 5 flies will have reached the right-hand side, and may be 10 others will be in the middle of the tube. If one waits 5mn, the flies will be distributed fairly uniformly throughout the tube.

If one repeats the same experiment with “*Tenebrio molitor*” beetles it will be seen that after 5mn almost all insects are still together on the left-hand side of the tube.

This experiment can be repeated in a more precise way by using the following procedure.

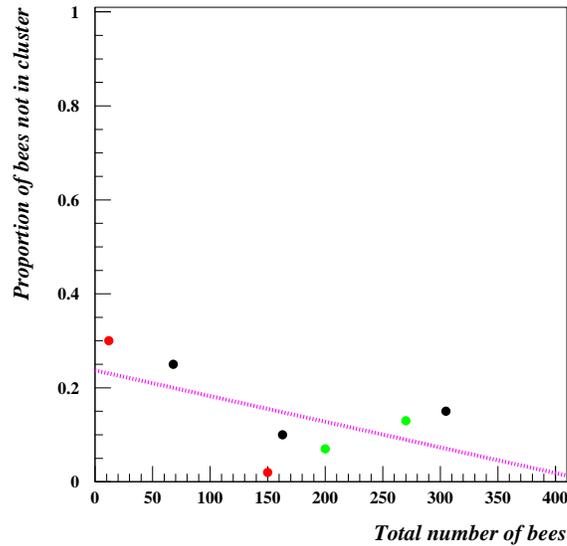


**Fig. 4a: “Evaporation” experiment with beetles.** In the evaporation version of the experiment (top) the beetles move from the container into open space (i.e. the laboratory table) whereas in the equilibrium version they move from part 1 of the container into part 2 of same size. In the first case almost no beetles come back into the container just like the molecules in the evaporation of a liquid. The graph shows that the dropout rate decreases when the size of the population increases pointing to greater attraction power of larger groups. In physics similar effects can be observed. For instance the vapor pressure around droplets of liquid decreases when the droplets increase in size (Kelvin equation) and the melting point of gold particles increases with the diameter of the particles (Buffat and Borel 1976, p. 2294).

For each value of  $n$  the experiment was repeated 10 times, which means that 80 experiments were performed altogether. For the 10 repetitions the coefficient of variation  $\sigma/m$  was around 50%. The slopes of the regression lines (with the numbers of beetles expressed in thousands) are as follows (the error bars correspond to a probability level of 0.90):

evaporation:  $-2.8 \pm 0.5$ ; 1 to 2, 10mn:  $-1.10 \pm 0.7$ ; 1 to 2, 30mn:  $-2.2 \pm 1.8$ ; not in cluster:  $-1.3 \pm 0.8$ . The average slope is  $a = -2.0$ . *The experiments were done in November 2012 by Ms. Mengying Feng and Shuying Lai from Beijing Normal University, Department of Systems Science.*

<sup>12</sup>This movement must be fast because drosophila have a natural tendency to go upward.



**Fig. 4b: “Not in cluster” experiment with bees.** After formation of a bee cluster, the number of those outside of the cluster were counted. The duration of each experiment was comprised between one hour and one hour and a half. The three different colors correspond to slightly different experimental conditions. For instance, for the black data points there was a single cluster whereas for the red points two clusters formed. In the latter case we divided all numbers by 2. The slope of the regression line (expressed per 1,000 bees) for the 7 experiments, namely  $a = -0.55 \pm 0.68$ , is 2.3 times smaller than the “not in cluster” slope in the beetles experiment. *The experiments were done in June and July 2012 by Mrs. J. Darley and B. Roehner in Val Fleury (western suburb of Paris). The bees were *Apis Mellifera mellifera*.*

The experiment starts after a number  $n$  of beetles has been introduced into a container that we will call part 1. In the “evaporation rate” version of the experiment, the beetles can just walk out into open space. In the equilibrium version of the experiment the opening of part 1 leads to a container of same size. In this case, most often, the beetles formed a cluster both in part 1 and in part 2. However not all the beetles were in the clusters. This leads to the definition of two different variables:  $n_2(t)$ , the number of beetles in part 2 at time  $t$ , and  $n_3(t)$ , the number of beetles which are not in a cluster. It is this latter variable which is the analog of the molecules in the vapor phase over a liquid. The observations summarized in the figure show that whether in the non-equilibrium case of evaporation or in the quasi-equilibrium case, the escape rate decreases when the number of beetles increases. A natural interpretation is that the combined attraction of  $n$  beetles on one of them increases with  $n$ <sup>13</sup>.

## Significance of the experimental results

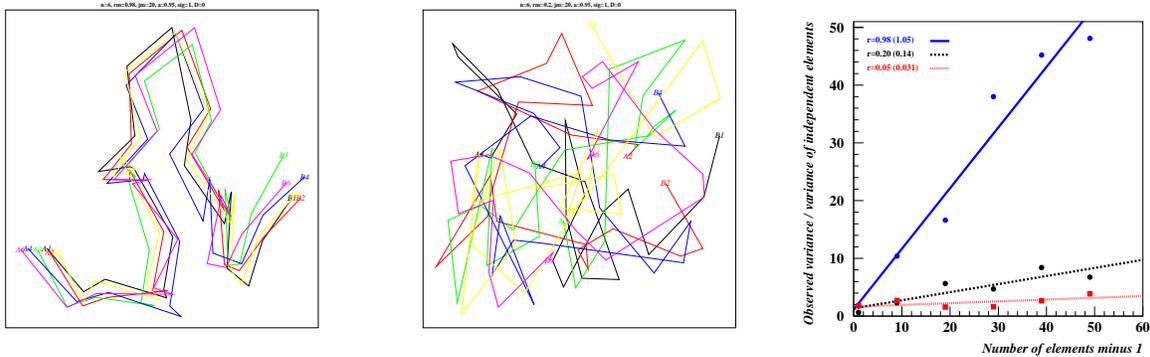
In this section we examine the significance of the results presented above from three

<sup>13</sup>More precisely, one can say that the experiment displays two competing forces: (i) attraction and (ii) increased volatility. The increased volatility likely comes with the beetles’ new environment. Indeed, when they are left alone for a long time they cluster together instead of occupying the whole available area, a typical liquid-like behavior.

perspectives: (i) How reliable are they? (ii) What is their physical significance? (iii) What is their biological significance?

## Reliability of the measurement method

In some experiments involving a limited number of big animals it is possible to record the trajectories of all individual elements<sup>14</sup>.



**Fig. 5 a,b,c Testing the cross-correlation measurement method on simulated data.** Simulated data were generated with different correlation levels, from a high correlation of 0.98 to a medium correlation of 0.20 and to a low correlation of 0.05; the latter has same order of magnitude as the correlations observed in ant experiments. The two graphs on the left-hand side show the simulated trajectories. The number of elements was limited to 6 for the clarity of the picture. The  $A_i$  letters indicate the starting point while the  $B_i$  letters are the end points. For  $\bar{r} = 0.2$  the trajectories are already fairly chaotic. Yet, the graph  $(n - 1, g^2)$  on the right-hand side shows that the measurement method based on  $\sigma^2(n)$  gives reliable results even for a correlation as low as 0.05. The length of the time series (namely 100) used in the simulation was similar to that which is used in the experiments.

For our experiments this would require marking hundreds of ants with different colors so as to be able to follow their movements on a video, a fairly impossible task. By merely restricting ourselves to counting the number of ants on one side of the container, the recording task becomes easier but at the same time one loses much information. The question is whether or not this limited amount of information is sufficient to make reliable measurements.

The best way to answer this question is to generate simulated data and to see if the results produced by the measurement procedure are consistent with the input correlations.

The simulated data were generated as follows. First,  $n$  auto-regressive processes  $Z_i(t)$  were generated through the equation  $Z_i(t) = aZ_i(t - 1) + B(t)$ , where  $B(t)$  is a white noise term (Gaussian white noise with mean zero and standard deviation 1) and where the parameter  $a$  regulates the time-continuity of the  $Z_i(t)$ . In our application we took  $a = 0.98$  which implies a fairly high continuity similar to what can be observed for individual ants. So far, the  $Z_i$  are uncorrelated. An average

<sup>14</sup>For instance in the paper by Gautrais et al. (2012) the movements of up to 30 fishes each about 25cm in size were recorded every 1/12 of a second during 2mn-long observations.

cross-correlation  $\bar{r}$  is introduced by using the Cholesky procedure. In this procedure a triangular matrix is created which defines new random variables  $X_i$  as linear combinations of the  $Z_i$ . Once the  $X_i$  have been produced it can of course be checked that they have indeed the right cross-correlations. This procedure must be repeated twice in order to produce the 2 Cartesian coordinates of each ant.

The simulated data are then used in the same way as observed data. Fig. 5c shows that the estimated correlations (namely the numbers within parentheses) are indeed consistent with the theoretical correlations. This graph also shows that smaller average cross-correlations require larger values of  $n$ . This is of course not surprising; when  $\bar{r}$  is small the term  $\bar{r}(n - 1)$  will remain lost in the background noise unless  $n$  becomes large “enough”. How large  $n$  must become of course depends on the level of the background noise.

The purpose of this simulation was to check the feasibility of the methodology. In addition, it gave useful indications about how to improve the accuracy of the measurement.

So far, it was assumed that  $\bar{r}$  is constant and in particular that it does not depend on  $n$ . What would be changed if  $\bar{r}$  depends on  $n$ ? If  $\bar{r}$  increases with  $n$  the relationship between  $g^2$  and  $n$  would no longer be linear but would have a parabolic (or even faster) growth. Similarly, if  $\bar{r}$  is supposed to decrease with  $n$  the function  $g^2(n)$  will be replaced by a function which grows slower than  $n$  but which remains above 1. This question will become important only once the accuracy of the measurement has been improved. At the present stage, we can accept the simple assumption that  $\bar{r}$  is constant. A more important point is whether  $g^2(n)$  remains above 1 or falls below 1, for the second case indicates a negative  $\bar{r}$  that is to say a repulsion. These conclusions are summarized in the following table.

### Conclusions about the interaction derived from $g^2(n)$

$$\begin{array}{l}
 g^2(n) > 1 \text{ and increasing } \rightarrow r > 0 : \text{ attraction} \\
 g^2(n) < 1 \text{ and decreasing } \rightarrow r < 0 : \text{ repulsion}
 \end{array}
 \left\{ \begin{array}{ll}
 \begin{array}{l}
 g^2(n) \text{ increases faster than } n \rightarrow r \text{ increases with } n \\
 g^2(n) \text{ increases linearly} \rightarrow r \text{ is constant} \\
 g^2(n) \text{ increases slower than } n \rightarrow r \text{ decreases with } n
 \end{array} \\
 \begin{array}{l}
 g^2(n) \text{ decreases slower than } n \rightarrow |r| \text{ decreases with } n \\
 g^2(n) \text{ decreases linearly} \rightarrow r \text{ is constant} \\
 g^2(n) \text{ decreases faster than } n \rightarrow |r| \text{ increases with } n
 \end{array}
 \end{array} \right.$$

We now examine the physical and biological significance of the previous measurements.

### Physical significance

Roughly speaking there are two main classes of interactions, namely local pair-wise interactions (thereafter called class  $L$ ) and global interactions (thereafter class  $G$ ) As

an example of the first kind, one can mention the interaction of an atom in a solid with its nearest neighbors; as one knows the Ising model provides a basic theoretical description of this situation.

Interaction through pheromone emission/reception or through visual (or auditive) contact (as in schooling fishes<sup>15</sup>) are examples of more global interactions.

In a random variable description there is a fundamental difference between the two classes. As soon as the population is of a substantial size (say  $n > 30$ ) most of the elements of a class- $L$  correlation matrix will be equal to zero (this point is explained in more detail in Appendix B). Thus, the average correlation decreases to zero when  $n$  increases which in turn implies that the variance will decrease toward the variance of the uncorrelated variables, i.e.  $g^2 \rightarrow 1$ . Clearly, this is not what was observed in Fig. 2.

- Thus, for the case of ants one can conclude that within a range of a few centimeters the interaction is of global type.
- For drosophila it is true that  $g^2$  decreases but as it tends toward zero rather than toward 1 this excludes an attractive Ising-like interaction. The fact that drosophila do not display any clustering behavior is another argument for excluding any attractive interaction.

### Biological significance

Although at this stage it is still too early to draw important biological implications<sup>16</sup> one can nevertheless emphasize two points.

- Clearly, well organized colonies such as those of ants or bees necessitate a sophisticated interaction network for control and regulation purposes. In other words, one can be certain that social insects do have many meaningful interactions. However, as the same argument cannot be used for non-social insects, does this mean that they have little interactions? Our beetles experiments as well as common observation about clustering contradicts such a conclusion. When clustering occurs spontaneously and not in response to an external stimulus (such as light, lower temperature or a supply of food) it reveals the existence of an inter-individual attraction. Observation of the free behavior of the beetles used in the experiment of Fig. 2c shows that they form a single cluster within a time span comprised between half an hour and several hours<sup>17</sup>. This comes as a confirmation of the positive cross-correlation measured in Fig. 2c. In fact, the formation of large clusters of hundreds or even

---

<sup>15</sup>Let us recall the distinction between schooling which means *swimming* together in the same direction and shoaling which means *staying* together in the same area.

<sup>16</sup>A similar situation prevailed in physical chemistry one century ago. It is only once a sufficient mass of observations had been recorded that a clear picture began to emerge. For an illustration of this process in the case of the alkanes see Roehner (2007, p. 10-11, Fig. 1.2).

<sup>17</sup>The process takes longer in a container without corners nor edges such as the lower half of a torus.

thousands of individuals can be observed for many insects, either on a permanent basis or just at specific times. Is there a relationship between the size of the clusters and the strength of the attraction<sup>18</sup>? This question will be addressed in forthcoming experiments.

- So far, entomologists have paid only scant attention to the issue of interaction strength. One of the objectives of this paper is to convince them that the interaction strength is a meaningful parameter.

## Conclusions

In this concluding section, we first summarize the main results, then we explain what makes this study fairly new. Finally we list some of the questions left for further investigation.

### Reproducible experiments

Like other animals, insects are influenced by many exogenous factors such as light, humidity, temperature, vibrations, and so on. Because of this background noise it was not obvious that measuring an interaction strength would be feasible. The good news is that it is. In other words the most important result is probably that the experiments described in this paper are *reproducible* with a signal-to-noise ratio which allows reasonable accuracy. It can be added that our experiments were done in distant places, in summer as well as in fall and with experimental devices which were not always the same. In spite of such changing conditions they lead to consistent results. This robustness was certainly favored in an important way by the large numbers of insects involved.

The results were not only consistent with one another, they were also consistent with what was expected from qualitative field observation. For the interaction strength they led to the following ranking:

Ants (3.4), Beetles (1.8), *Drosophila* (−1.3)

### Two different approaches

It must be realized that two fairly distinct approaches are possible. To explain this important point it is best to illustrate it through an example from the study of molecular interactions.

(1) One possible objective is to study the *detailed* mechanism of hydrogen bonding in water molecules. This is a demanding task which requires measurements at atomic level and a solid competence in quantum mechanics. In addition, hydrogen bonds are not the only interaction between water molecules; the so-called London

---

<sup>18</sup>In fact, the significant factor should be the balance between the attraction and the level of background noise.

interaction is another, albeit much weaker. In short, studying hydrogen bonds is a difficult and in a sense, endless) task.

(2) A second possible approach relies on comparative analysis. It is made possible by the fact that one does not need to know the precise mechanisms of the interactions in order to measure their *global* effect. This can be done by measuring the amount of energy that is required to break *all links* between water molecules. Such measurements lead to coupling estimates expressed as interaction energies. Taken alone, any single measurement of that kind is not of great significance but once such interactions have been measured for *various* compounds, then their connection with macroscopic properties will become apparent, which in turn will provide valuable predictive power. In short, this approach fundamentally relies on comparative analysis.

The two approaches are certainly useful and in a sense they complement one another. Whereas the vast majority of studies (especially those done by biologists) follow the first approach, in this paper we tried the second one.

### **Connection with the approach pioneered by Prof. Deborah Gordon**

Although Prof. Gordon's investigations mostly concerned a specific species, namely harvester ants in Arizona, the way she conducted them actually has much to do with the comparative approach outlined above.

Why?

The main reason is because she asked broad questions that make sense for any species of ants and indeed for any colony of insects. As illustrations one can for instance mention the following issues.

- What is the long-term growth pattern of ant colonies?
- How in a colony is food supply adjusted to needs?

In such investigations the emphasis is more on stable patterns than on minute details that would be species specific. This orientation is also quite apparent from the titles and subtitles of Gordon's books (1999, 2010). The organizational issues addressed by Deborah Gordon are not easy ones. We believe that interaction strengths are basic parameters whose knowledge may shed some light on organizational problems. Can we hope to explain superconductivity without a quantitative knowledge of the interaction between electrons?

### **Experiments done by physicists**

In past decades a number of interesting ant experiments were done by physicists; they were a source of inspiration for us. For instance one can mention the experiments conducted by the teams of Ernesto Alstshuler (2005), Jean-Louis Deneubourg (e.g. Theraulaz et al. (2003)) or Guy Theraulaz (Theraulaz et al. (2002), Gautrais et al.

(2012).

However, almost all these studies are of the “one species - one model” kind. By this expression we mean that after detailed observations have been recorded for *one* species, a model is built (which by construction will of course be in agreement with the experimental data) but little is attempted to determine by further experiments what is the field of validity of this model. Although these studies were pioneering in several respects and pointed out many interesting phenomena, one can regret that, to our best knowledge, they did not cover a broad range of species in order to identify *core mechanisms*. This would be crucial in order to be able to offer *testable predictions*<sup>19</sup>.

### Testable predictions

It is well known that the ability to offer testable predictions is the hallmark of scientific investigations. What testable prediction can be derived from the results given in the present paper?

(1) It can be predicted that if a species has a negative cross-correlation (as in the case of the drosophila), then it cannot at same time exhibit a clustering behavior

(2) It can be predicted that if the correlation and evaporation experiments are repeated for other species, then for any species characterized by a positive average cross-correlation one should observe a negative slope in the the “evaporation” experiment. Moreover, a higher cross-correlation should give a more negative slope.

At this point such predictions should rather be seen as conjectures. The important point is that they can be tested. If they are disproved by observation one will be lead to a re-examination of the present model. It is through such a continual dialogue between theory and observation that further progress can and will be achieved,

### The route ahead

Collective phenomena are particularly difficult to explain. In physics the theory of superconductivity came centuries after Galileo and decades after it was first discovered. In other words, one should be prepared for a long-term investigation. This is a field where there is much to explore.

Let us just mention one question not mentioned so far but that is certainly important, namely the role of temperature. Some preliminary observations convinced us that in group effects temperature plays a role which is fairly similar to what can be seen in chemistry and statistical physics. Which temperature do we mean, the external

---

<sup>19</sup>A few centuries ago when physicists studied the phenomenon of “free” fall they did not confine themselves to falling apples. Indeed, comparative observation was the only way to demonstrate that, at least in air, the law is fairly independent of the shape and density of the falling object. This was a milestone in the development of classical mechanics. In water, on the contrary, both shape and density play a role. Thus, in this case, the limits of validity of the law of free fall are well defined.

temperature or the endogenous temperature of the insects as defined in the sense of statistical physics? Observation suggests that the two are closely connected (data can be found in the long write-up version mentioned at the beginning). This provides a means for controlling the endogenous temperature and for studying its effect on many phenomena (e.g. clustering, self-diffusion, evaporation).

In physics real progress occurs when there is a fruitful dialogue between theory and observation. This is currently one of the problems faced by string theory. There is a similar problem with computer simulations of social phenomena because of the fact that they rarely lead to testable predictions and when they do, most often, the requested statistical data turn out to be unavailable. For experiments on groups of insects the situation is much more favorable.

**Acknowledgments** For their helpful comments we would like to thank the researchers who attended a seminar given by one of us at the Department of Economics of the University of Tokyo on 28 November 2012.

One of the experiments reported above was done in Kunming at the Eastern Bee Institute of Yunnan Agricultural University, China; we wish to express our appreciation to Prof. He and Tan for their hospitality and to Dr. Chen and Wang for their help.

We also want to express our gratitude to Ms. Mengying Feng, Shuying Lai and Mr. Geng Li from the Department of Systems Science of Beijing Normal University in relation with the results mentioned in our discussion of the “evaporation” rate of beetles.

We are very grateful to Prof. Guy Theraulaz for accepting to share with us the data described in Gautrais et al. (2012). By giving an example of a cross-correlation which decreases when the number of elements increases, these data attracted our attention to this possibility.

## Appendix A: Variance of a sum of correlated variables

We proceed in several steps.

For the sake of simplicity we first consider the average of a sum of three correlated random variables  $X_1, X_2, X_3$  of mean  $m$  and identical standard deviation  $\sigma$ . Our objective is to compute the variance of  $S_3 = X_1 + X_2 + X_3$ .

By definition of the variance  $\sigma^2(S_3) = E[(S_3 - E(S_3))^2]$ . One knows that the expectation of a sum of random variables is always equal to the sum of the expectations, whether the variables are correlated or not. Thus:  $E(S_3) = E(X_1) + E(X_2) + E(X_3)$ .

Consequently:

$$\sigma^2(S_3) = E \left[ \left( \sum_{i=1}^3 \hat{X}_i \right)^2 \right], \quad \text{where: } \hat{X}_i = X_i - E(X_i)$$

Thus,

$$\sigma^2(S_3) = \sum_{i=1}^3 E(\hat{X}_i^2) + 2 \left[ E(\hat{X}_2\hat{X}_3) + E(\hat{X}_3\hat{X}_1) + E(\hat{X}_1\hat{X}_2) \right]$$

We express the expectations of the products by introducing the coefficient of correlation of the  $X_i$ :  $r_{ij} = E(\hat{X}_i\hat{X}_j)/\sigma^2$ . Thus:  $\sigma^2(S_3) = 3\sigma^2 + 2\sigma^2(r_{23} + r_{31} + r_{12})$

From that point on, we will consider two cases.

## Global correlations

Introducing the mean of the  $r_{ij}$ ,  $\bar{r} = (r_{23} + r_{31} + r_{12})/3$ , we obtain:

$$\sigma^2(S_3) = 3\sigma^2[1 + 2\bar{r}]$$

This formula has an obvious generalization to an arbitrary number  $n$  of random variables:

$$\sigma^2(S_n) = n\sigma^2g^2, \quad g^2 = (n-1)\bar{r} + 1$$

where:

$$\bar{r} = \frac{1}{[n(n-1)/2]} \sum_{i \neq j}^n r_{ij}$$

## Ising-like correlations

For a one dimensional Ising spin system the correlation between spins  $i$  and  $j$  is:  $r_{ij} = \eta^{|i-j|}$  where  $\eta$  can be expressed (if one wish) as a function of the parameters which define the interaction between the spins (see Glauber 1963 p. 299, formulas (56) and (57)).

Introducing this expression of  $r_{ij}$  gives:  $\sigma^2(S_3) = \sigma^2(3 + 4\eta + 2\eta^2)$

In extending this formula to any  $n$ , one needs to express the finite sum  $f(\eta) = \sum_{i=0}^{n-2} \eta^i$  (as well as its derivative  $f'(\eta)$ ). Instead of using the exact expression  $f(\eta) = (1 - \eta^{n-1})/(1 - \eta)$  we will consider that the term  $\eta^{n-1}$  is negligible with respect to 1, which means that we approximate the finite sum by the corresponding infinite series. This approximation is acceptable for our experiments because most of the time  $n > 20$ . Of course the approximation is no longer valid when  $\eta \rightarrow 1$  but  $\eta = 1$  is a case of uniform correlation already considered above.

Under this assumption one obtains finally:

$$\sigma^2(S_n) = n\sigma^2 \left[ 1 + \frac{2\eta}{1-\eta} \left( 1 - \frac{1}{n(1-\eta)} \right) \right]$$

or:

$$g^2(n) = \frac{\sigma^2(S_n)}{n\sigma^2} = \frac{1+\eta}{1-\eta} - \frac{2\eta}{n(1-\eta)^2}$$

Due to the approximation made in the derivation, this formula is not valid when  $n$  is close to 1. We have seen above that for  $n = 2, 3$  one gets:

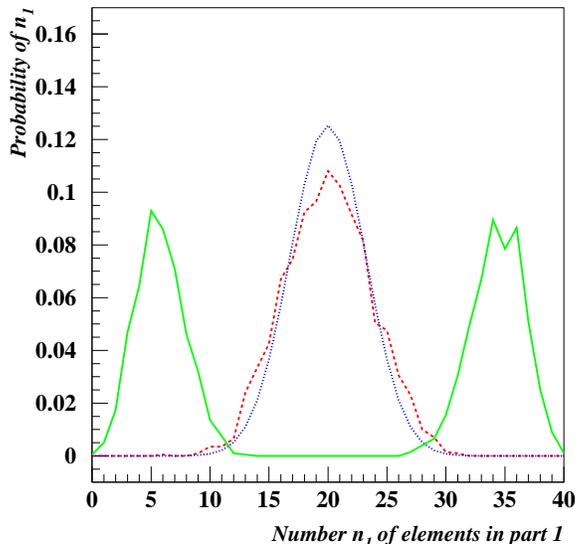
$$g^2(2) = 1 + \eta, \quad g^2(3) = 1 + (4/3)\eta + (2/3)\eta^2$$

which shows that the function  $g^2(n)$  increases toward its asymptotic limit  $(1 + \eta)/(1 - \eta)$ .

**Remark** Can the Ising case be seen as a special instance of the previous case? Formally, it may seem so. However, the real picture emerges when we consider large values of  $n$ . In the Ising case, due to the exponential decrease, all elements in the correlation matrix are almost equal to zero except

for a zone around the first diagonal whose width depends only upon  $\eta$ . Consequently, for such a matrix the average correlation goes to zero when  $n$  becomes larger.

This observation shows three things. (i) It would be irrelevant to treat the Ising case as a special instance of the first case. (ii) The fact that in the Ising case  $\bar{r} \simeq 0$  helps to explain that the ratio  $g^2(n)$  remains basically constant instead of increasing. (iii) It explains why we used the expression “global correlations” to designate the first case. The correlations are global in the sense that when  $n \rightarrow \infty$  the number of elements of the correlation matrix that are “substantially” different from zero must remain of the same order of magnitude as  $n$ . For a distance-dependent correlation, this means that the decrease with distance must be slow enough.



**Fig. A1 Distribution of a sum  $S$  of  $n$  correlated Bernoulli variables  $X_i$ .** In our experiments  $S = \sum_{i=1}^n X_i$  represents the number of elements in part 1 of the container. The (blue) dotted curve is the binomial distribution corresponding to a zero correlation; the (red) broken line corresponds to a correlation  $r = 0.01$ , and the (green) solid line to  $r = 0.5$ . A bifurcation process occurs at some point which in our experiments translates into a clustering process. The  $n$  variables were generated through the Lunn and Davies algorithm which gives correlated exchangeable variables (“exchangeable” means that they have a symmetric joint distribution function). We took  $n = 40$  and for each curve we generated 1,000 realisations of  $S$ .

## Simulations

So far we did not need to make the assumption that the  $X_i$  are Bernoulli variables, that is to say variables taking only the values 0 and 1<sup>20</sup>. However, if one wishes to carry out a simulation there is a convenient algorithm which works for Bernoulli variables (Lunn and Davies 1998). The relevant formulas can be summarized as follows:

**Simulation of global correlation between  $n$  Bernoulli variables**  $Z$  and  $Y_i$  are  $\text{Ber}(p)$  random variables while the  $U_i$  are  $\text{Ber}(\sqrt{r})$  random variables. Then, the variables  $X_i$  defined as:

$$X_i = (1 - U_i)Y_i + U_iZ, \quad i = 1, \dots, n$$

are correlated Bernoulli variables with the following properties:

$$E(X_i) = p, \quad E(X_i^2) = p, \quad \text{Cor}(X_i, X_j) = r, \quad i \neq j$$

<sup>20</sup>When  $P\{X = 1\} = p$  such a variable will be noted as  $\text{Ber}(p)$ .

It can be noted that this algorithm works only for positive correlations between the variables.

**Simulation of correlated Ising-like Bernoulli variables**  $Y_i$  are  $\text{Ber}(p)$  random variables while the  $U_i$  are  $\text{Ber}(\eta)$  random variables. Then, the variables  $X_i$  defined as:

$$X_1 = Y_1, \quad X_i = (1 - U_i)Y_i + U_i X_{i-1}, \quad 2 \leq i \leq n$$

are correlated Bernoulli variables with the following properties:

$$E(X_i) = p, \quad E(X_i^2) = p, \quad \text{Cor}(X_i, X_j) = \eta^{|j-i|}, \quad i \neq j$$

## Appendix B: Experimental “toolkit”

Just in order to convince readers that experiments with insects can be done fairly easily we give some practical hints.

Basically, the needs can be summarized as follows.

First, one needs to get the *living organisms*.

- Ants can be easily collected (at least in spring and summer) by putting appropriate food as a bait on a Bristol board just a few centimeters away from the entrance of a colony. Within one hour and depending on the species a few hundred ants may gather on the Bristol board.
- *Drosophila* can be obtained from biology laboratories.
- Flies and beetles can be bought in the form of larvae (worms) destined to fishermen or for feeding big aquarium-fishes. The waiting time between the larvae stage and the emergence of the adults ranges from less than one week to a few months depending on species, temperature and time of year.

Secondly, in many cases, one needs a small bottle of carbon dioxide to make the insect sleep in order to be able to handle them easily. Carbon dioxide has an almost instantaneous anesthetic effect on all these insects. According to a paper published in the “Journal of Experimental Biology” (Ribbands 1950) anesthesia through carbon dioxide does not infer a memory loss and changes only slightly the behavior of bees. It is probably safe to assume that the effect on the other insects mentioned above is similar.

Next one needs an appropriate container. A simple solution is to cut it into a piece of flexible plastic (such as PVC) of adequate thickness (3mm to 5mm is usually enough). This is illustrated in Fig. 1.

Finally, one needs a counting device. Taking pictures and counting by hand is a simple solution but not always satisfactory especially for counting the elements in a cluster. For this reason we have developed a weighing method (illustrated in Fig. 3).

Clustering phenomena also occur among bacteria and micro-organisms that are present in so-called biofilms which form at the surface of liquids. Because of the small size and high numbers of such elements one is in a situation fairly similar to physical systems. For instance, it can be mentioned that inter-molecular forces such as van der Waals forces play a significant role in the movements of such micro-organisms. Studying the *collective behavior* of such populations from the perspective of physics seems a promising field. However, in contrast to the study of insects, it requires special laboratory devices and equipment.

## References

- Altshuler (E.), Ramos(O.), Núñez (Y.), Fernández (J.), Batista-Leyva (A.J.), Noda (C.) 2005: Symmetry breaking in escaping ants. *American Naturalist* 166,6,643-649.
- Aoki (M.), Yoshikawa (H.) 2007: *Reconstructing macroeconomics: a perspective from statistical physics and combinatorial stochastic processes*. Cambridge University Press, New York.
- Buffat (P.), Borel (J.-P.) 1976: Size effect on the melting temperature of gold particles. *Physical Review A*, 13,2287-2298.
- Gautrais (J.), Ginelli (F.), Fournier (R.), Blanco (S.), Soria (M.) Chaté (H.), Theraulaz (G.) 2012: Deciphering interactions in moving animal groups. *PLoS Comput Biol* 8(9): e1002678. doi:10.1371/journal.pcbi.1002678.
- Glauber (R.J.) 1963: Time-dependent statistics of the Ising model. *Journal of Mathematical Physics* 4,294-307.
- Gordon (D.M.) 1999: *Ants at work. How an insect society is organized*. Free Press, New York.
- Gordon (D.M.) 2010: *Ant encounters: interaction networks and colony behavior. Primers in complex systems*. Princeton University Press, Princeton.
- Iyetomi (H.) 2012: Labor productivity distribution with negative temperature. *Progress of Theoretical Physics Supplement* 194, 135-143.
- Iyetomi (H.), Nakayama (Y.), Aoyama (H.), Fujiwara (Y.), Ikeda (Y.), Souma (W.) 2011: Fluctuation-dissipation theory of input-output interindustrial relations. *Physical Review E* 83, 016103-1 to 016103-12.
- Lunn (A.D.), Davies (S.J.) 1998: A note on generating correlated binary variables. *Biometrika* 85,2,487-490.
- Papoulis (A.) 1965: *Probability, random variables and stochastic processes*. McGraw-Hill Kogakusha, Tokyo.
- Ribbands (C.R.) 1950: Changes in the behavior of honeybees following their recovery from anaesthesia. *Journal of Experimental Biology* 27,3,302-309.
- Roehner (B.M.) 2004: A bridge between liquids and socio-economic systems: the key-role of interaction strengths. *Physica A* 348, 659-682.
- Roehner (B.M.) 2007: *Driving forces in physical, biological and socio-economic phenomena. An investigation of social bonds and interactions*. Cambridge University Press, Cambridge.
- Roehner (B.M.) 2012: *Estimating interaction strength in populations of living organisms*. Working report, LPTHE, University Pierre and Marie Curie, Paris. Available on the following website:  
<http://www.lpthe.jussieu.fr/roehner/effusion.pdf>

- Theraulaz (G.), Bonabeau (E.), Nicolis (S.C.), Solé (R.V.), Fourcassié (V.), Blanco (S.), Fournier (R.), Joly (J.-L.), Fernández (P.), Grimal (A.), Dalle (P.) 2002: Spatial patterns in ant colonies. *Proceedings of the National Academy of Science* 99,15,9645-9649.
- Theraulaz (G.), Gautrais (J.), Camazine (S.), Deneubourg (J.-L.) 2003: The formation of spatial patterns in social insects: from simple behaviours to complex structures. *Philosophical Transactions of the Royal Society in London A*, 361,1263-1282.