Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

Modelling social behavior of Drosophila Melanogaster under the effect of drugs

Relatore:
Prof. Armando Bazzani

Correlatore:
Dott. Nico Curti

Presentata da:
Davide Ravaglia

Anno Accademico 2018/2019

Contents

Abstract

This thesis faces the problem of the study of the cognitive behavior in the animals carried out through the methods of statistical analysis and stochastic dynamical models. In the experiment analyzed were utilized specimens of Drosophila Melanogaster which were confined inside an arena and to which were administered different drugs in different experiments; the influence of these drugs are a core feature of the study.
The intent is to examine cognitive behavior in these animals thanks to the statistical analysis of the encounters between specimens, in particular we want to distinguish whether these encounters happen by chance or are a decision of the specimens involved. To highlight the presence of this social interaction it has been produced a toy model of the experiment: a set of parameters extracted from the real data allows for the generation of random walks. Assuming its capacity of representation, significant differences from the toy model will indicate the presence of cognitive behavior.

Sommario

Questa tesi affronta il problema di studiare il comportamento cognitivo negli animali attraverso metodi di analisi statistica e modelli dinamici stocastici. Negli esperimenti analizzati sono stati utilizzati degli esemplari di Drosophila Melanogaster confinati in un'arena a quali sono state somministrate droghe diverse in diversi esperimenti; parte integrante dello studio riguarderà l'influenza di queste droghe.
Lo scopo è esaminare il comportamento cognitivo degli animali attraverso l'analisi statistica degli incontri tra esemplari che possono avvenire sia per una situazione casuale che per una decisione dell'animale stesso. Per evidenziare la presenza di questa interazione sociale è stato prodotto un modello nullo dell'esperimento: diversi parametri estratti dai dati in possesso permettono la generazione di percorsi casuali. Assumendo la sua capacità di rappresentazione, differenze significative dal modello nullo indicheranno la presenza di comportamenti cognitivi.

Chapter 1

Behavior Analysis

Recent studies have pointed out how the understanding of the behavior is fundamental in the neuroscience studies of all branches such as molecular, cellular, cognitive, behavioral and computational neuroscience, translational research, neural circuits and many more branches[1]. These studies have shown as well the relationships between neuroscientific findings and conceptual research, fostering the an open debate on the subject which can include psychology and philosophy[2].
Neuroscience is deeply connected to behavioral neuroscience and decision theory which both address the topic of the interaction between the brain and its environment. This relationship is expressed through behavior.
Behavior is fundamental to all living organism, spacing from bacteria to humans, every animal has multiple responses to stimuli from the internal and external environment[3]. Along the pathways that leads from genes to behavior there are several steps which include the habitat, the past experiences, the social structure and the immediate surroundings in the moment of the manifestation of the behavior. Because of these different causes, behavior research presents a deep bond with various different subject of study and standardized methods have been developed by behaviorists to study behavior objectively.
Since most of the behavior are expressed towards the environment or the other animals, both of the same species and not, its study offers insights on the relationship that the animal expresses with its surrounding. Behavior allows us to better understand how to tackle problem regarding the protection of endangered species and how much the human impact is felt. Moreover, by studying animals that have departed from the branch of the evolution that humans have followed, we can produce estimation regarding the evolution of behavior and the physiological mechanisms behind it. Reversely the study of the behavior of the primates has often offered valuable perspectives into the evolution of individual, social, and reproductive human actions. At last it has to be considered that since complex behavior has evolved in million of years, it often represent the most efficient answer to a problem or a task. The transposition of the problem and the corre-

sponding behavior/response into our modern world is often utilized, such as in the smart design of cities.
We will start by analyzing the possible motives of behavior, the fundamentals of the decision-making process and how to measure the preferences expressed by the animals. At last we will briefly examine 5 experiments in which it has been observed the shift in behavior caused by different drug administration to animals and we will discuss what it has been learned from these experiments.

1.1 The causes of behavior

As stated, behavior is a response to a stimulus, if the stimulus is external the information is carried by the senses, a different point has to be made about internal stimuli. Internal stimuli, or appetites, are generally caused by the tendency for an organism to maintain internal equilibrium, a concept called homeostasis[4]. Possible drives of behavior are: hunger, thirst, the need for sleep, the need to regulate body temperature. Homeostasis should not be thought as a source of homogenization of behavior since more often it causes its differentiation. This becomes clear when thinking that different behavior are expressed by the animals in a changing environment, such as different season, in order to keep their balance.
When an animal is placed in an environment where it cannot express the behavior needed to maintain its internal equilibrium it often performs what are known as displacement behavior. To offer an easy to relate example let us think at humans who adjust their hair or continuously change their arms placement when in situation where they do not feel in comfort. The repetition of such behavior can cause great mental and physical discomfort, such as birds constantly pulling out their feathers or mammals grooming regularly. The study of these distress behavior and of the identification of the not expressed behavior which causes them is of great importance when considering animals is captivity and the treatment for anxiety and depression.
Another possible cause of stress is the presence of conflicting behavior. Focusing on one activity at a time is usually more profitable than attempting to simultaneously achieve multiple ones, as a consequence for most of the species it is present a latter of behavioral priorities. The study of the decision making process between different behavior is an area of research that only in the later years have found the first answers.

1.2 Measuring animal preferences

In order to understand whether animals have preferences in the choice they make, so whether they make conscious choice or not, several controlled experiments have been performed during the years. In these studies it has been measured both positive and neg-

ative motivations; positive when it was shown a wish to complete an activity or engage in a social environment and negative when it was shown a wish to avoid such activity. In both cases it can be associated a strength of the motivation. Both the motivation and its strength can be influenced by many factors such as age, experience, time of the day, weather, sex and so on[5].
When having to choose a behavior over many different profitable ones, the strength of the motivation creates a ladder of preferences. The preferences may be specific for the individual or common to the whole species. In any case the preference implies a decision which develops into a behavior. The ability to modify the preferences is a key factor for the adaptability of the species in different or changing environments and situations.
In order to examine animal choice behavior with less possible factors that may influence it, these experiments are often performed in standardized test situations in the laboratory. This feature also allows for better repeatability. To improve the level of control over the experiment even more and not have animals confused by the surrounding during the experience, these tests are often performed with captive-born animals. At last, often the behavior of the animals which are subject to the test are compared to a control group which undergoes a similar experience which differ only in one of the possible causes of the expressed behavior. This way its influence can be better observed.

1.3 Past experiments regarding the response to drugs

In the past researchers have considered the problem of how the administration of different drugs could alter behavior and what can be learned from these experiences. In all of these experiments it was present a control group to which it was not given the drug.

Giving octopuses MDMA: serotonin and sociality

Serotonin is present all over the animal kingdom but it is not simple to understand how widespread is its role in the sociality related behavior. Since in humans the usage of methylendioxymethamphetamine (MDMA) highly enhances social behavior due to the increase in the serotonin level in the brain that it generates, this substance has been administered to a different animal to see how the response may vary. The choice fell on the octopuses for two different reasons: it is mainly a solitary and asocial animal and its evolutionary path diverged from ours over 500 million years ago. In this experiment it was given to the octopus the choice either to play alone with some toys or to spend time with another octopus[6]. The octopuses which received MDMA where highly more likely to choose to spend time with one of their peer relative to the octopuses of the control group which did not receive the drug.
The influence caused by serotonin is very similar to the one shown in humans, so it was highlighted the ancient role of serotonin in sociality. Moreover, since this role is common

both in humans and in octopuses it becomes very plausible that is shared by most of the animal kingdom.

Giving bees cocaine: symptoms on insects and mammals

Cocaine shows very different effects in humans and in insects, for the latter in fact this substance is often deadly and consequently is considered an insecticide. This is often explained by assuming different assimilation processes in the two Phylum (divisions of animal kingdom) of insects and mammals. However studying the behavior of honey bees when given small doses of cocaine gave opposite responses, as the bees acted similar to mammals[7]. In particular it was observed the changes in the symbolic dances performed to indicate the location of flowers: when given cocaine the rate of these dances increased, causing the other bees to overestimate the value of pollen collected. Furthermore, reduction and cessation of the administration of cocaine caused responses similar to the withdrawal behavior shown in humans.
These findings suggested that insects and mammals do not assimilate the drug in a completely different way but that the symptoms in insects, often deadly, were probably due to intoxication caused by a too high drug intake.

Giving flies ethanol: its bond with the reward system

When behavior required for species survival, such as sex, foraging and consuming food, social interaction, are performed, the brain reinforces the neural pathways connected with these behavior, a process known as the reward system. There are drugs which parallel the response of the brain and can create addiction if associated with a need for survival. This study was conducted to understand if this parallelism is present in behavior shown by humans in situations that could be considered merely influenced by the society perspective[8]. Scientist gave the option of ethanol assumption to male Drosophila Melanogaster after these were either rejected by a partner or successfully reproduced with her. The study revealed that the rejected flies were highly more likely to drink several times the amount drank by the sexually appeased ones, showing in this way a link between sexual experience, the brain reward system and ethanol consumption: the reward system encouraged the rejected flies to drink alcohol to strengthen the same neural pathways that have been strengthen in the other flies by the successful sexual experience.
This study is particularly interesting to look under a drug addiction perspective as it shows how addiction can be generated by the impossibility of the completion of a behavior linked with survival.

Giving rats opioids: a vaccine against drugs

Another family of drugs that link to the neural pathways of the reward system causing addiction are opioids. Opioids in fact numb pain stimuli and release large amount of dopamine inside the brain. Given the public health crisis related to these drugs, scientist investigated the possibility of developing a drug vaccination and tried to do so on male Wistar rats[9]. In a first stage they combined the opioid chosen, in this case oxydocone, to a carrier protein that would induce the immune system to create antibodies against the compound and treated half of the rats with this combined molecule. Then they trained the rats to self-administer the pure drug. They found that, while all of the rats who were not previously treated with the compound developed an addiction, only half of the treated rats did. Moreover, when the scientist increased the difficulty of obtaining a drug dose, the rats which were previously treated dropped the habit significantly faster. Once again these findings are of great importance for addiction treatment, where during the recovery a vaccine against the drug of abuse might help to reintroduce a subject into society. A possible relapse of usage in fact would not kick in the same reward system mechanism.

Giving mice psilocybin: neurogenesis and fear response

Other then social behavior, serotonin plays an important role in the creation of new brain cells, phenomenon connected to the formation of new memories. In particular the aim of this experiment was to observe how long would a past induced fear response last and if its persistence could be influenced. This was studied on mice, part of which were given a small dose of psilocybin (found in hallucinating mushrooms is a potent drug that releases a high dose of serotonin in the brain) before the experiment[10]. The scientist played a note and concurrently shocked with electricity the mice and repeated this procedure until all of the mice freeze in fear hearing the note even if not receiving the shock. The scientist then observed how many times it was necessary to repeat the note without the shock before the mice stopped freezing. It was found that mice injected with psylocybin extinguished the fear response significantly more rapidly than those which did not receive the drug. Moreover the dose given was so small that the mice did not seem to present any side effect, such as hallucinations.
The importance of these findings is of particular interest for the treatment of post-traumatic stress disorder. The doctor could give small doses of psylocybin or related drugs to the patient during exposure therapy, where the patient usually faces his fears.

Chapter 2

Introduction

2.1 Goals of the study

The objective of this study is the identification and classification of a cognitive behavior by a physical-mathematical approach and how it is influenced by the administration of different drugs. The main target of this behavior that we aim to look into concerns the meeting between Drosophila specimens, its frequency as well as the recurrence of the different possible meetings. Since the investigation is regarding meeting between Drosophila it will be necessary to define the features that characterize a meeting event. To recognize a cognitive behavior we will have to compare the data relative to it with another set of analogous data. In fact, if the proprieties of the data can be considered part of a known distribution, it becomes possible the computation of the p-value. This expresses the likelihood of an event, which in this case would be the hypothetical cognitive behavior.
In order to obtain this set of analogous data we will create a toy model as faithful as possible to the real experiments but based on random walks. This random walks will have to be similar to the real trails, so we will have to maintain the most important kinematic properties, such as the probabilities about the duration of the movement, the speed and the velocity direction variation.

2.2 The experiment

In total were performed three different experiments. The difference relies on the substances given to each Drosophila before the experiment, these drugs:

- **Caffeine**
- **Ethanol**

- **Sugar**

These drugs were administered from 8 hours before the start of the data collection. The Drosophila have then been placed in a circular arena with a diameter of $100mm$ and a low ceiling. The height of the ceiling forced the Drosophila to walk inside the arena, making the identification process easier and more reliable. Moreover this forced them to move in a 2 dimensional surface, a feature that the experimenters hope will induce more contact between the specimens.
All of the specimens are male with the same age, from 8 to 10 days, of the same strain. At the beginning of experiment the flies are left in the dark inside the arena for about 5 minutes. After this the arena is lit with an homogeneous white light and no visual target. At this point the data collection begins and lasts for about 24 minutes ($\approx 1466.7sec$).

2.3 DataSet

The data is collected thanks to a video camera placed above the arena, perpendicular to it and directed to its center. From the video which is recorded, every 1/15 of a second the software collects several features relative to that instant for a total of 22000 frames. Those features span from kinematic information, such as position, speed, direction and much more, to information regarding the closest Drosophila[11]. All of these features are obtained by the program by assigning to each Drosophila an ellipse, where the main axis is the head-tail axis, with an orientation in the direction of the head. The main body of the animal in fact has a well defined elliptical shape as shown in Fig.2.1[12] where is shown a representation of a Drosophila and its scale. For the features which regard a measure of length, the program converts from pixels to millimeters, this ratio for this experiment is set to $8.251\frac{pixel}{mm}$. This means that a single Drosophila is represented by dozens of pixels, which is to say that is well recognizable and it can be excluded any error regarding the identification process.

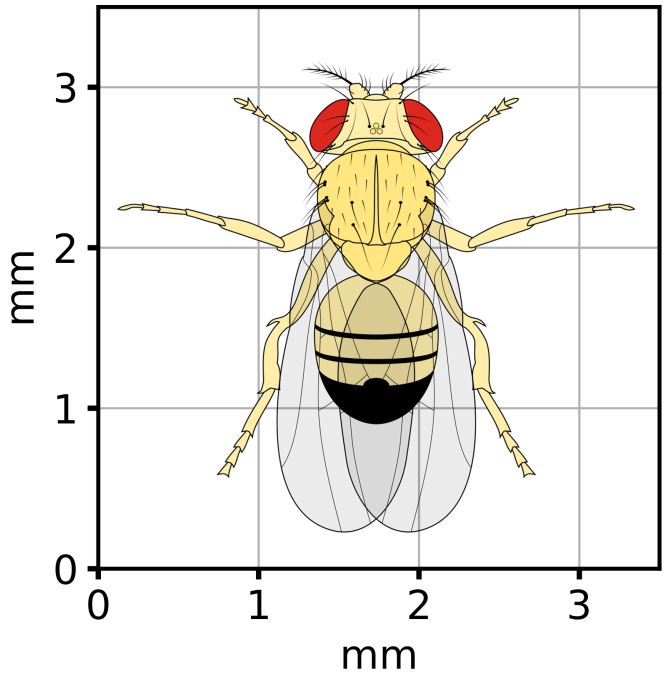

Figure 2.1: Representation of a Drosophila and measure of its width and height. It is possible to see how the body of the animal has an elliptical shape.

Beside the use of the identification code, the main features obtained by the program which have been utilized for the study are:

- **Time**: starting from $0sec$ at the first frame, it increases of $\frac{1}{15}sec$ each subsequent frame.
- **Position**: x and y coordinates, using the center of the arena as the origin (mm).
- **Speed**: computed as the distance travelled, relative to the Drosophila's center, between the current frame and the previous one divided by the time interval of $\frac{1}{15}sec$ ($\frac{mm}{sec}$).
- **Velocity direction**: computed from the position of the Drosophila in the current frame and the previous one (rad).
- **Orientation**: computed by assigning to each Drosophila an ellipse with the major axis in the tail-head direction (rad).
- **Distance to the closest wall**: the distance, based on the Drosophila's center, is to the closest point to the border of the arena (mm).

- **Distance to the closest Drosophila**: the distance, based on the Drosophila's center, is to the closest center of another Drosophila (mm).

In Fig.2.2 is shown a sample of these raw data. It has to be considered that for all of these data the accuracy of the measure is not on the least significant digit as this would imply an unrealistic precision in the data collection.

Index	id	theta	timestamps	x_mm	y_mm	velmag_ctr	phi	dist2wall	dcenter
0	fly#_001_1	-2.3994	0	11.8963	-44.8828	0	-2.5529	2.06	10.587
1	fly#_001_1	-2.4114	0.066668	10.6763	-45.6988	22.031	-2.6728	1.5641	11.956
2	fly#_001_1	-2.7551	0.13334	9.39432	-46.1518	20.436	-2.828	1.3953	13.269
3	fly#_001_1	-2.7965	0.2	8.08932	-46.5378	20.4	-2.8838	1.2576	14.612
4	fly#_001_1	-2.8168	0.26667	6.73032	-46.8548	20.934	-2.8673	1.1583	15.934
5	fly#_001_1	-2.8295	0.33334	5.38332	-47.2998	21.28	-2.8437	0.88872	17.36
6	fly#_001_1	-2.8136	0.40001	4.11232	-47.6578	19.806	-2.8989	0.65832	18.663
7	fly#_001_1	-2.818	0.46668	3.37532	-47.7968	11.25	3.132	0.57798	19.375
8	fly#_001_1	-2.9697	0.53334	3.31032	-47.6508	2.4007	1.3814	0.72832	19.403
9	fly#_001_1	-3.0161	0.60001	3.40332	-47.6498	1.4007	-0.098615	0.72235	19.3
10	fly#_001_1	-3.0248	0.66668	3.45132	-47.6648	0.74433	-0.17227	0.70446	19.229
11	fly#_001_1	-3.0355	0.73335	3.47132	-47.6618	0.31249	-0.22634	0.70563	19.213
12	fly#_001_1	-3.0155	0.80002	3.46532	-47.6678	0.13127	-1.9818	0.70016	19.218
13	fly#_001_1	-2.999	0.86668	3.45932	-47.6898	0.34171	-1.6139	0.6786	19.221

Figure 2.2: Part of the raw data utilized relative to one of the Drosophila in the experiment regarding caffeine. The column in order represent: identification code, orientation, time, x coordinate and y coordinate of position, speed, velocity direction, distance to the closest wall, distance to the closest Drosophila.

These features enable us to represent the dynamical proprieties of the trails such as displayed in Fig.2.3. Here it has been reconstructed the trail of one of the Drosophila of the experiment regarding caffeine obtained from 5000 frames ($\approx 333.3sec$).

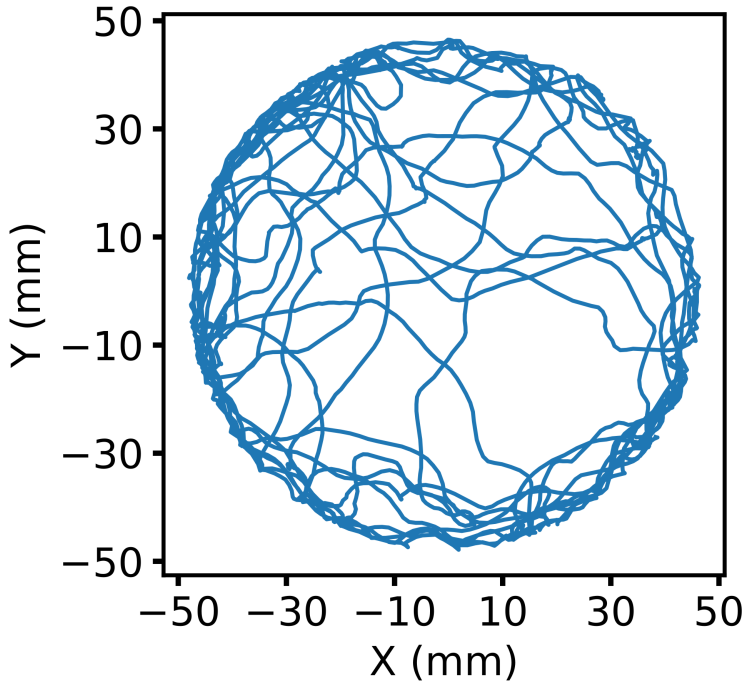

Figure 2.3: Part of the trail of one Drosophila of the experiment regarding caffeine obtained from 5000 frames ($\approx 333.3sec$).

Chapter 3

Data Analysis and Models

In the following section it will be discussed the preliminary analysis needed to obtain a reliable toy model which will be used as a comparison of the real data, with the aim of highlighting possible cognitive behavior. This analysis concerns the features before listed (speed, orientation and velocity direction) and the alternating kinematic state of the Drosophila between movement and stop.
Since the data have a timestep of $\frac{1}{15}sec$ the toy model will utilize the same time interval between consecutive frames. Also, since one of the main objective is to observe if there are differences in these cognitive behavior that could be caused by the drug used in the experiment, most of the following data extraction will be differentiated by the drug type. Doing this it will be possible to produce random walks fitter to the three distinct experiments. The differentiation of the random walks will be useful as well to verify if the random walk creation was successful for some experiment but not for others, this would give a clue of some kinematic information lost in the process of modeling the real trails.

3.1 Implementation of the running average

Instead of utilizing the raw data of speed and velocity direction of the Drosophila, it has shown necessary the usage of the running average of this quantities. In fact in both of these features it occurs a frequent oscillations around a fixed value or around an increasing or decreasing trend as is shown in Fig.3.1, which shows the raw data of the speed compared to its different running average for part of one of the trails of the experiment regarding caffeine. These oscillations can be attributed to the data acquisition which is made frame by frame and may cause these discontinuities due to the instant nature of the measure. However these oscillations do not correlate to an erratic movement as shown in Fig.3.2, where is shown the trail travelled by the same Drosophila as the one in Fig.3.1 in the same time interval.

Moreover in the speed values are present some very sharp peak (in Fig.3.1 next to the $44sec$ mark) which are either caused by an error in the data collection process or by a short fly of a Drosophila, immediately interrupted by the low ceiling. Not knowing the exact origin of these values we have thought to maintain the continuity of the mobility paths by smoothening these peaks rather then just excluding them. At last, later in this chapter it will be shown how each trail is divided into two categories based on a speed threshold: it has been verified that by introducing a running average, even if based on a small time interval, the occurrences in which the threshold was crossed for only a fraction of a second (e.g.: for only one frame the Drosophila speed drops to $0\frac{mm}{sec}$ while the values around it are much higher) was greatly reduced.
Summing up, doing so we hope that by filtering the fast oscillations we will be able to isolate the macro behavior.

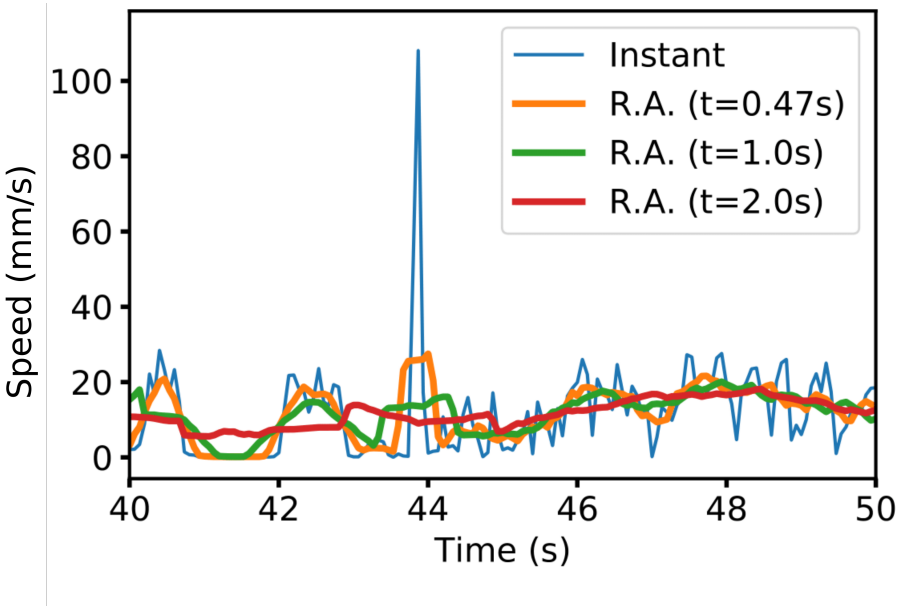

Figure 3.1: Instant speed of the raw data compared to different simple running average of the same speed. The graph shows the data relative to a time interval of one of the Drosophila in the experiment regarding caffeine. The running average chosen in the orange one since it allows the filtering of the fast oscillations but does not alter the macro behavior.

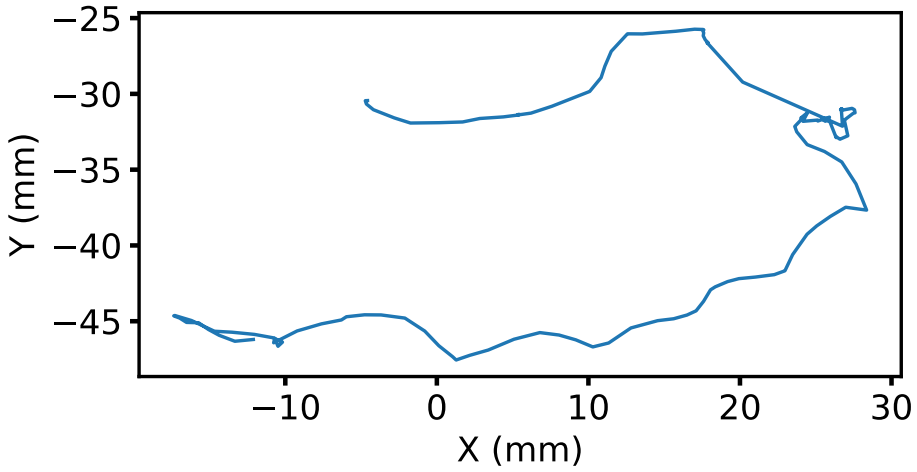

Figure 3.2: Trail travelled by the same Drosophila as the one in Fig.3.1 in the same time interval. It can be seen that frequent oscillations in speed do not correlate to a completely erratic movement.

The running average used is the simple unweighted mean of the N data around a certain value (N preferably being an odd integer, in order to have the exact center of the considered time interval in the value examined). So the running average of the i-eth value of the velocity direction (φ) using N steps will be given by:

$$\varphi_{R.A.} = \frac{\varphi_{i-M} + \ldots + \varphi_{i+M}}{N} \qquad with \qquad M = \frac{N-1}{2}.$$

The graph in Fig.3.1 helps as well to choose which running average is the most suitable to point out the relevant infromation for our purpose. In fact, knowing that it will be necessary to separate the mobility path from the stopping points of each Drosophila, we can see that the running average with an interval of approximately $0.47sec$ is the fastest to reach the null speed where the Drosophila stops (see Fig.3.1 around the $42sec$ mark along the x-axis) but at the same time it smooths out almost completely the oscillation around a fixed value (observable in Fig.3.1 around the $48sec$ along the x-axis).
This is not the only reason why it was chosen to use this exact value of running average which consists of 7 frames ($7frames \cdot \frac{1sec}{15frames} \approx 0.47sec$). We have in fact considered the distribution of the differences between the raw data and the possible running averages. These distributions are shown in Fig.3.3. It can be observed that it is a Lorentzian distribution when the running average is based on the $0.47sec$ time interval but it loses

this feature for time intervals above $1sec$, where, instead of a smooth and steep descent, two humps appear close the center. This is an indication that the running average is not being a faithful representation of the real data since a considerable portion of the values of the running average differ of constant values (the center of the humps) from it.

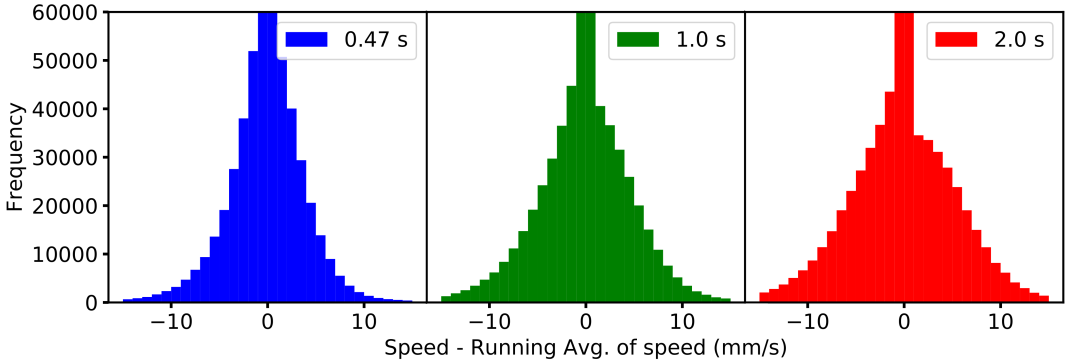

Figure 3.3: Frequencies of the differences between the speed and the running average of the speed, for different running averages. It can be noticed how for the higher value of the time interval used for the running average the distribution loses its lorentzian shape.

At last the same graph shown in Fig.3.1 can be replicated using the velocity direction instead of the speed and obtaining similar results regarding both the smoothening of the oscillations and the proximity to the curve as shown in Fig.3.4.
From now on when we will refer to speed, orientation and velocity direction it will always be implied the running average just described of these quantities, not their raw values.

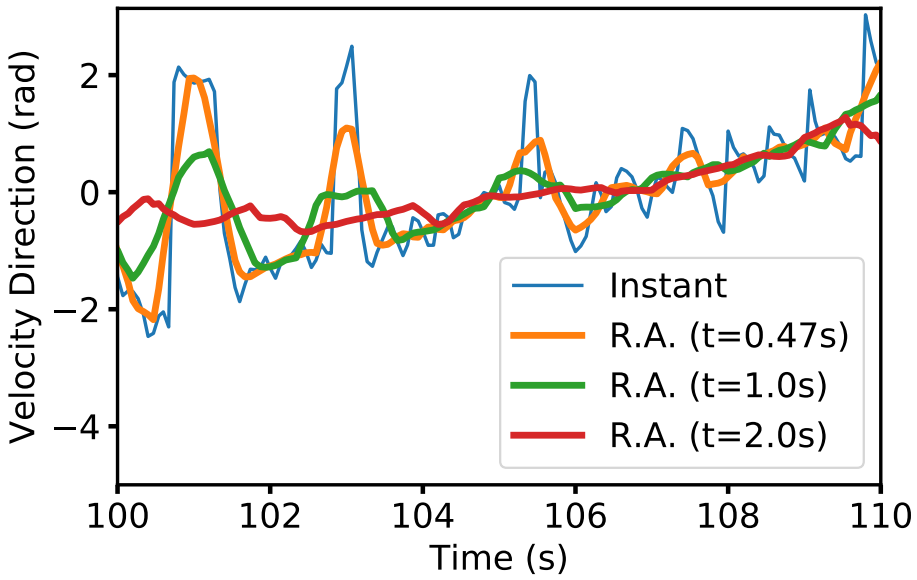

Figure 3.4: Instant velocity direction of the raw data compared to different simple running average of the same velocity direction for a time interval of one of the trails in the experiment regarding caffeine. The running average chosen in the orange one since it allows the filtering of the fast oscillations but does not alter the macro behavior.

3.2 Alternation between movement and stop

The kinematic behavior of the Drosophila can be divided into two possibilities: movement or stop. Basically the Drosophila moves for a certain time and then stops for another interval, in this latter intervals they do not completely stop but drastically reduce their speed. In the toy model we will have to represent this alternating behavior so we need to study its dynamical property. To do this we divide the kinematics of the trail into the two categories based on the value of the speed, as the threshold it has been chosen $2.0\frac{mm}{sec}$. The motivation of this choice relies on two main observations.
The first one is obtained from the graph in Fig.3.5 where is shown, for all of the Drosophila in the three experiments, the probability of a given value of speed both in normal and in logarithmic scale. It can be seen how the likelihood of a given speed decreases with the increase of the speed value even more rapidly then a exponential decrease and around the value of $2\frac{mm}{sec}$ it reaches a stable point in the decrease.

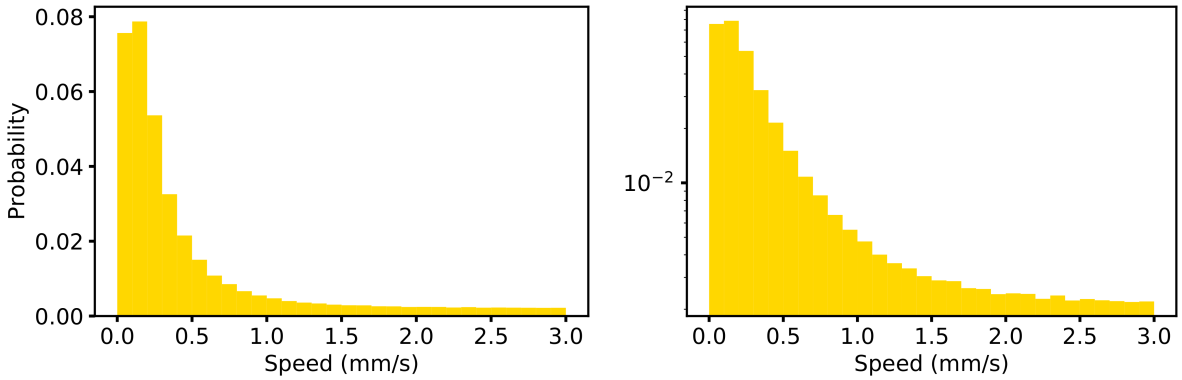

Figure 3.5: Probability of instantaneous speed for all of the Drosophila in the three experiments both in normal and in logarithmic scale. It can be observed how the decrease is faster then an exponential decrease and how around the $2\frac{mm}{sec}$ mark it stops decreasing.

The second observation comes from the variation of orientation and in the relationship between orientation and velocity direction which are shown in the graphs in Fig.3.6 and Fig.3.7. The first two graphs show the probability of the differences in orientation between consecutive frames for speeds above (on the left) and below (on the right) the threshold chosen. The second two graphs show the differences between orientation and velocity direction in the same timestep, for speeds above (on the left) and below (on the right) the threshold chosen.
The threshold of $2.0\frac{mm}{sec}$ appears appropriate to distinguish the different behavior which the Drosophila has in movement and when it stops. In fact it can be seen how the variation of orientation is minimum while in stop and far greater when in movement. Conversely the difference between orientation and velocity direction in minimum while in movement while has a completely different distribution for the stop part. This indicates that when in stop the Drosophila moves around keeping its orientation while in movement it moves approximately towards the direction is oriented to and is free to change this direction.
The peaks around π and $-\pi$ shown in Fig.3.7 for the values of the speed under the threshold of $2.0\frac{mm}{sec}$ indicate how the Drosophila tend to move backwards as well while in stop.

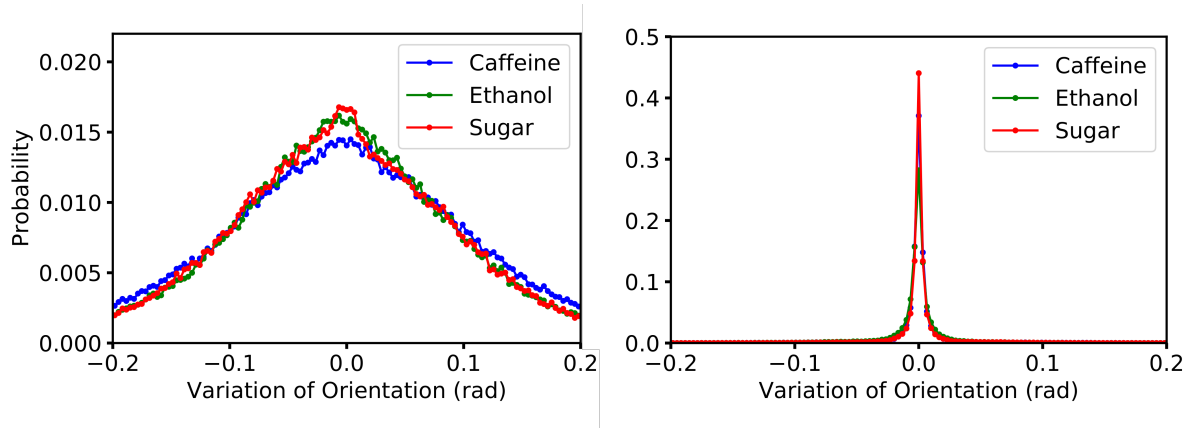

Figure 3.6: Probability of certain change in orientation between subsequent frames. For $v > 2.0\frac{mm}{sec}$ on the left and for $v < 2.0\frac{mm}{sec}$ on the right. It can be observed how much the graph on the right is centered around 0 radiant meaning that the Drosophila do not change their orientation while in stop.

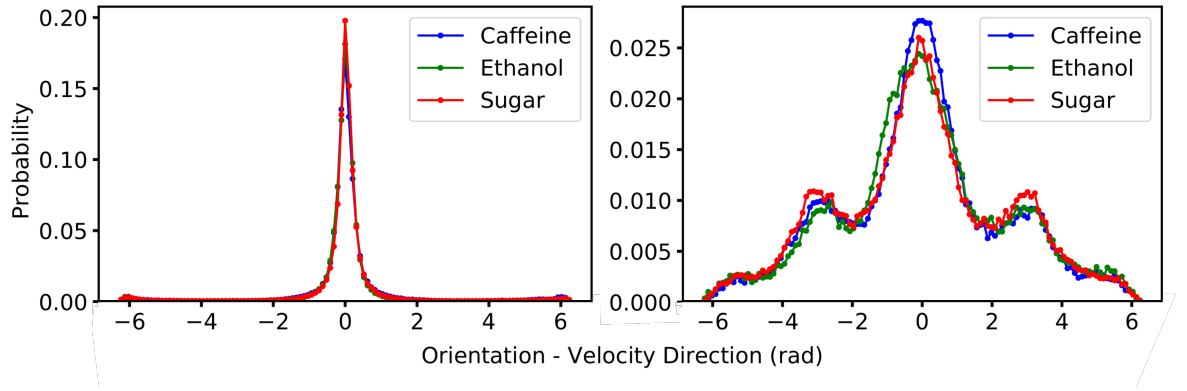

Figure 3.7: Probability of certain difference between orientation and velocity direction in the same instant. For $v > 2.0\frac{mm}{sec}$ on the left and for $v < 2.0\frac{mm}{sec}$ on the right. It can be observed how much the graph on the left is centered around 0 radiant meaning that the Drosophila almost always move towards the direction they are facing when in movement.

Dividing the kinematics based on this threshold we can obtain a probability distribution of the duration of movement and stop which are shown in Fig.3.8 and Fig.3.9, differentiated by the drug type, both of the graphs have the representation in logarithmic scale as well. It can be observed that, while the decrease of duration of movement is exponential for all of the drugs, the decrease of the duration of stop is not: it presents different trends for the different drugs and different peaks of probability as well, such as

the one immediately before the $2sec$ mark. These probability function distributions will later be used for the toy model.

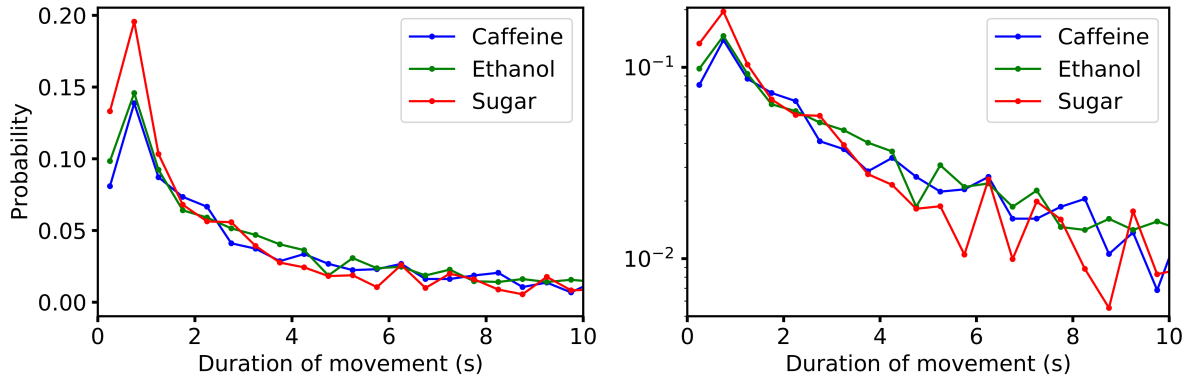

Figure 3.8: Probability of movement duration, differentiated by the drug type, both in normal and in logarithmic scale. It can be observed that the decrease is approximately exponential.

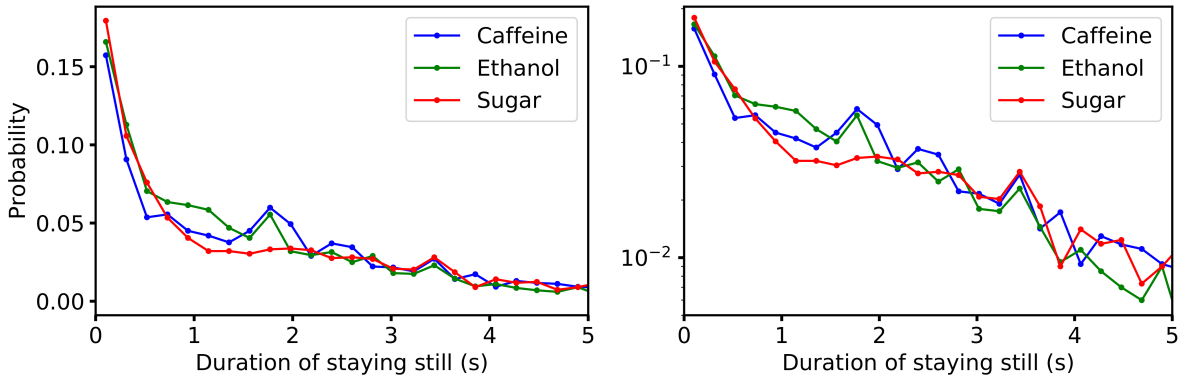

Figure 3.9: Probability of stop duration, differentiated by the drug type, both in normal and in logarithmic scale. It can be observed that there is not only an exponential decrease but several trends which differentiate for the drug type.

3.3 Correlation between speed and velocity direction

As stated in Par.3.2 the toy model will carry the information of alternation between movement and stop and it will have to represent the different features in speed and velocity direction of these two dynamical states. A faithful representation is needed in order not to assign randomly extracted velocity direction changes to randomly extracted values of speed.
To do this we have to consider the relationship between these two physical properties, always differentiating between movement and stop. The relationship between speed and the change of velocity direction can be seen in Fig.3.10 and Fig.3.11, where the values of the speed have been partitioned depending on the value of the change in the velocity direction measured in the same instant. This means that the interval $[-2\pi, 2\pi]$ has been divided into a number of smaller intervals (301 intervals) and for each of these intervals are collected all of the values of speed correlated to such a change in velocity direction. Then, assuming that the speed distribution for each interval is a Gaussian distribution, from all of these values of speed it has been computed the average and the standard error on the mean which are represented by the dot and the light bar around it. The formulas which have been utilized are:

$$v(\Delta\theta)_{avg} = \frac{\sum_{i=1}^{N} v(\Delta\theta)_i}{N} \qquad \sigma_{avg} = \sqrt{\frac{\sum_{i=1}^{N}[v(\Delta\theta)_i - v(\Delta\theta)_{avg}]^2}{(N-1)N}}$$

where N is the number of speed entries inside the interval and $v(\Delta\theta)_i$ is one of the entries of the speed value in the bin relative to a certain $\Delta\theta$. From both of these graphs it can be observed a correlation between higher values of speed and small changes in velocity direction, this correlation appears to be stronger for the higher values of the speed. Moreover, in this graphs we can see a clear difference between the types of drugs as in the first graph the values of average for the experiment relative to sugar are significantly lower compared to the other two and in the second one the same values for the experiment relative to ethanol are significantly higher compared to the other two. For both of these graphs, the range on the x-axis that has been shown is the only part with a considerable amount of entries, as it will later be proven.
It has been computed as well the probability of a given interval simply dividing the entries in the interval by the total number of entries. The graphs of the probability are shown in Fig.3.12 and Fig.3.13 where is shown the same part of the x-axis to the previous graphs, which is, as is now clear, the only part with a relevant probability associated to it. From these graphs it can be observed that, especially for high values of speed, most of the time there is only a small change in velocity direction, so the Drosophila tend to keep their direction of movement.

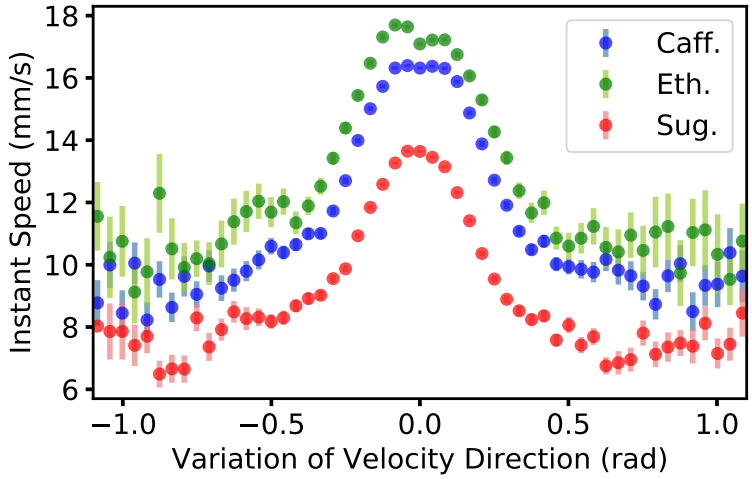

Figure 3.10: Average and standard error on the average obtained from all the values of the speed (v>2.0$\frac{mm}{sec}$) which are related to a certain change in velocity direction. It can be seen a strong correlation between high values of speed and small changes of velocity direction.

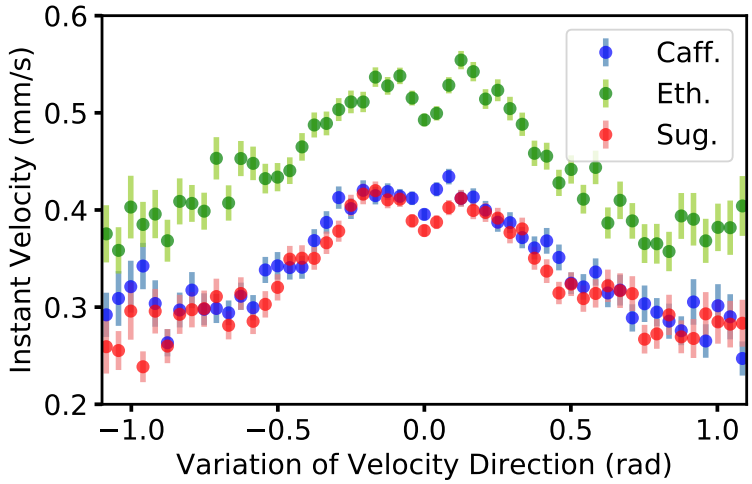

Figure 3.11: Average and standard error on the average obtained from all the values of the speed (v<2.0$\frac{mm}{sec}$) which are related to a certain change in velocity direction. It can be seen a correlation between higher values of speed and small changes of velocity direction.

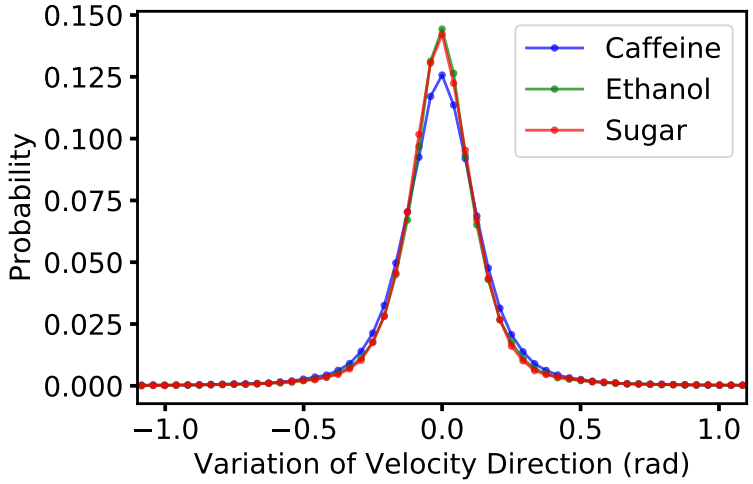

Figure 3.12: Probability distribution function of the changes in velocity direction (v>2.0$\frac{mm}{sec}$). It can be seen how the small changes are the most likely ones.

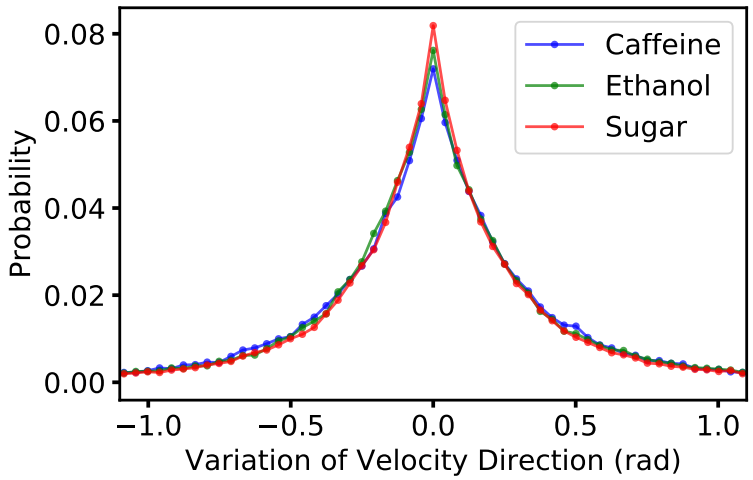

Figure 3.13: Probability distribution function of the changes in velocity direction (v<2.0$\frac{mm}{sec}$). It can be seen how the small changes are the most likely ones.

3.4 Modelling Drosophila mobility as a random walk

All of the information gathered above will now be utilized for the definition of a random walk model. Before illustrating the process we will tackle the problem of the

border restraint.

3.4.1 The boundary conditions

Since the random walks will have to be a realistic dynamical representation of the real trails we want them to be confined inside an arena with the same diameter as the one in the experiment. In order to have this confinement it has been chosen the following method to apply (see Fig.3.14 for more clarity):

- For every new point generated it is computed the distance from the center. If the distance is greater than $50mm$ the new point will have to be replaced. We will call B the point outside of the circle and A the last point of the trail inside the circle.
- Since we want to replace B but also to keep the same distance originally travelled (otherwise the random walk will suffer a loss in trail lenght) we find the two intersections (C and D) between the border and the circle having A as the center and B as a point of the circumference.
- Between C and D, we identify the closest intersection to point B and substitute it with B. This way we will be able to keep as best as possible the original direction of movement. Supposing that the closest intersection is the point C, we have effectively replaced in the trail sequence the point B with the point C.
- Since the generation of new points of the trail is partly based on the variation of velocity direction, the velocity direction has to be computed again and it will be given by $arctan(y_C - y_A, x_C - x_A)$. In $arctan(y, x)$ the quadrant is chosen so that $arctan(y, x)$ is the signed angle in radians between the ray ending at the origin and passing through the point (1,0), and the ray ending at the origin and passing through the point (y, x).

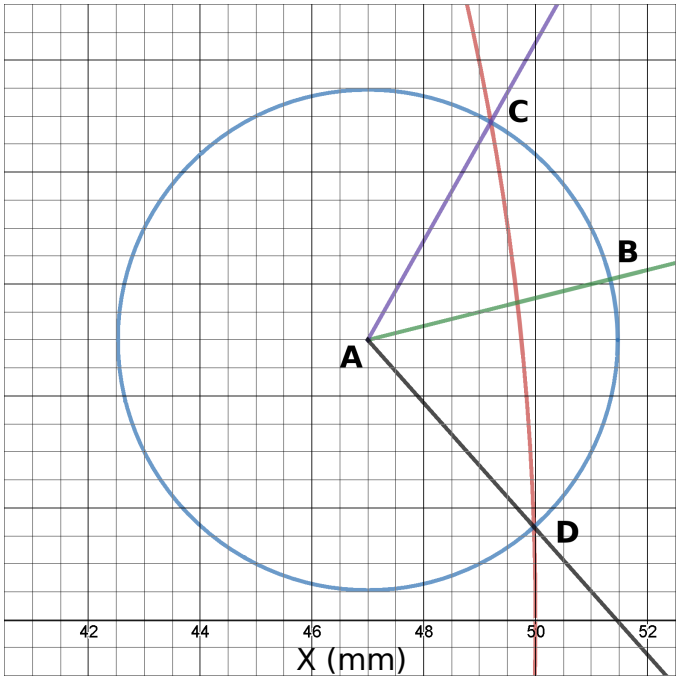

Figure 3.14: Illustration of the procedure followed to restrain the random walk trails inside the arena. The trail is at point A when point B is computed as the next one to be in the trail. Being out of the boundary (the red curve), point B will be replaced by point C which is the closest intersection between the boundary and the circle having A as center and B as point on the circumference.

3.4.2 The generation process

Now we have all of the information needed for the process used to generate the random walks, which goes as follows:

- A point inside the arena is randomly chosen with a random velocity direction in the interval $[\pi, -\pi]$.
- Still randomly is chosen if the trail will start with a stop or movement state, each of the state having 50% likelihood of extraction. From here on this states will alternate.
- Based on what state has been chosen, the duration of the state is randomly extracted from the probability distributions shown in Fig.3.8 and Fig.3.9. The extraction will be based on different distributions depending on the type of drug. The closest integer to the result of the multiplication between the duration extracted and 15 will represent how many steps to do in this state.

- To every timestep is associated a change in velocity direction randomly extracted from the probability distributions shown in Fig.3.12 and Fig.3.13. The extraction will be based on different distributions depending on the type of drug.
- To every timestep is associated a speed randomly extracted from the Gaussian distribution, shown in Fig.3.10 and Fig.3.11, relative to the interval of the change in velocity direction. The extraction will be based on different distributions depending on the type of drug.
- The coordinates of the new point of the trail are computed based on the speed and the velocity direction.
- Once the steps of the current state have been completed, the state will change from stop to movement and vice versa, so another duration of the state will be extracted.
- Whenever a random walk exits the arena the trail is edited as explained in Sect.3.4.1 in order to keep the trail inside.
- The process ends when 22000 steps have been generated.

Every completed set of random walks has 10 trails of 22000 steps for each drug, meaning that for the generation of the first 10 trails are used different probability function distributions then the second 10 trails and so on.
Let us clarify what we mean when we refer to a certain value randomly extracted from a given distribution. Given an integrable probability density function (PDF) $f(x)$, such as the ones shown above in Fig.3.8, 3.9, 3.10 and 3.11, and a random uniform distribution $g(x')$ in the interval [0, 1]:

$$f(x)dx = g(x')dx' = dx'$$

since $g(x') = 1$ being it a uniform random distribution in the interval [0, 1]. Integrating both sides we will obtain $F(x) = x'$ from which it follows $x = F^{-1}(x')$. Here $F(x)$ is the cumulative distribution function (CDF).
The same logic can be applied as well with our PDFs even if they are not continuous by creating a discrete CDF. Now, once we have extracted a random number x in the interval [0, 1], we will choose the value of the quantity studied where the value of the CDF in a certain interval is higher then x but is lower in the previous interval. In order to have a more continuous extraction of values of the quantity, the value chosen will not actually be the exact center of the identified interval but a random uniform extraction in the range $[X_{center} - \xi, X_{center} + \xi]$ where X_{center} is the center of the interval and ξ its half width.

3.5 Definition of encounter

It is now necessary the definition of an encounter between two Drosophila. The only requirement made to the Drosophila which interact is to simultaneously be still for a certain amount of time, which is yet to be defined, in the same region of the arena, which size is yet to be defined. No requirements are made about the relative orientation since the Drosophila have a range of vision which almost spans 360° so they always see one another and studying the relative orientation when in proximity to other Drosophila it does not emerge any kind of preference.
Let us define the threshold for minimum time spend simultaneously still and maximum distance of interaction. Regarding the first parameter, the main objective is not to consider, for the real data, the times where two Drosophila meet next to the border of the arena just because they are following it in inverse direction and then continue on their paths. In fact, as it is visible in Fig.2.3, the Drosophila spend a lot of time around the border so most of the encounters will happen there but will be short and not of interest for our study. Therefore, in order not to consider these, and possible other, category of very brief encounters, the time threshold chosen is $0.5sec$. Regarding the choice of the relative distance not to exceed during an encounter, it has been graphed the probabilities of the possible distances to the closest Drosophila, shown in Fig.3.15. The graph on the left shows the complete spectrum of distances while the graph on the right is the close up of the part near the peak (the y-axis is the same for the two graphs). Here it is possible to see a peak of frequency for distances between $2.5mm$ and $5mm$.

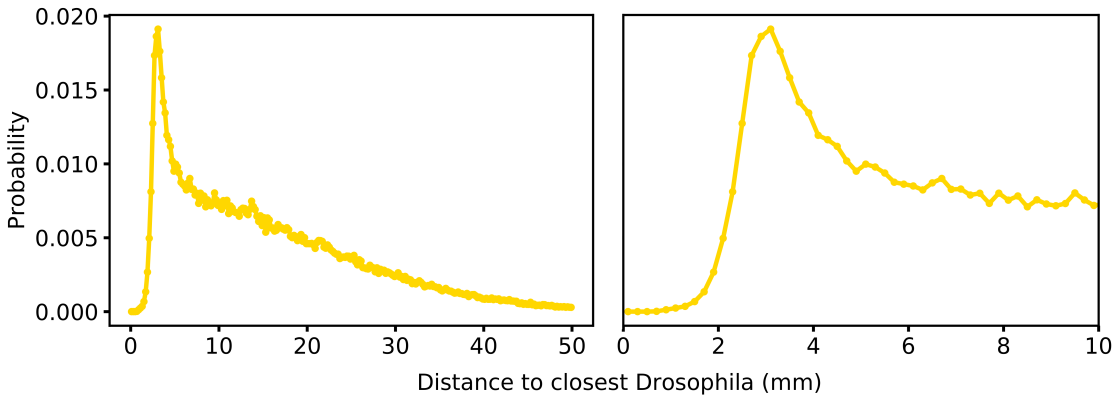

Figure 3.15: Probability distribution function of the distance between a Drosophila's center and the closest other Drosophila's center, obtained from all of the data. Closer distances then $2mm$ are very unlikely due to the animal body size (see Fig.2.1).

Closer distances then $2mm$ are very unlikely, the distance in fact is computed based on the position of their centers so the physical size of the animal (see Fig.2.1) becomes

the minimum distance obtainable. By looking at these graph it would appear that the distance to consider for the encounter is about the same value of the peak shown, 2.5-5mm. It has although to be considered that when two Drosophila meet at the border in the process of avoiding each other they keep a relative distance which is approximately in the range of the peak shown in the graph. This hints at the fact that the peak could not be representative of all of the possible meetings, in fact during the experiments appear several events which could be considered encounters where the distance kept is greater then 5mm. One of the examples is shown, also to better display how little the Drosophila move during the part of stop, in Fig.3.16. The graph on the left is relative to the full duration of the encounter between the two Drosophila in the middle, highlighted by the red circle around them (their trails are almost single points because of they are in stop), and it shows the evolution of the 10 trails during 5.1sec. The graph on the right is part of the trails evolution shown on the left and it displays the simultaneous stop of three Drosophila for 2.3sec, always shown by the single dots in the center of the arena and highlighted by the red circle. For these 3 Drosophila, the distance to the closest one is about 8-9mm.

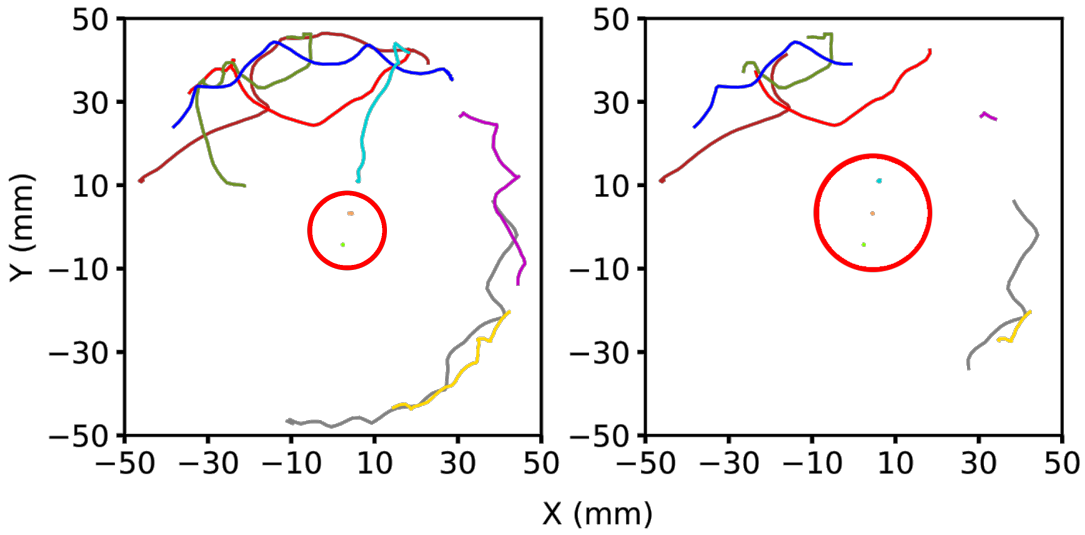

Figure 3.16: Representation of the simultaneous trail of the 10 Drosophila in the experiment involving caffeine. Highlighted by the red circle are the Drosophila which are in stop and in relative proximity, so what could be considered an encounter. The graph on the right (evolution during 2.3sec) shows part of the trail evolution shown in the graph on the left (evolution during 5.1sec).

From these observation we conclude that we must consider as well possible encounters at further distances then the range previously identified in Fig.3.15. For this reason the

subdivision into regions of the arena has been performed as follows: the smallest possible square which encases the arena with parallel sides to the x and y axis of the coordinates has been divided in 36 square regions of equal area by dividing his sides into 6 portions (so is divided in squares with sides $\approx 16.7mm$).
So the definition of encounter is the simultaneous stop for at least $0.5sec$ in the same region into which the arena has been divided. In order to consider possible meetings between Drosophila in two neighboring regions at a close distance one to the other, each region has a border with a width of a fifth of the side of the square ($\approx 3.3mm$) overlapping the other regions which is considered part of both regions. For the sake of clarity, in the example given in Fig.3.17 the two Drosophila, shown by the yellow ellipses, are considered in the same region since one is inside the region, the dark orange section, and the other is inside the border, the light orange section.

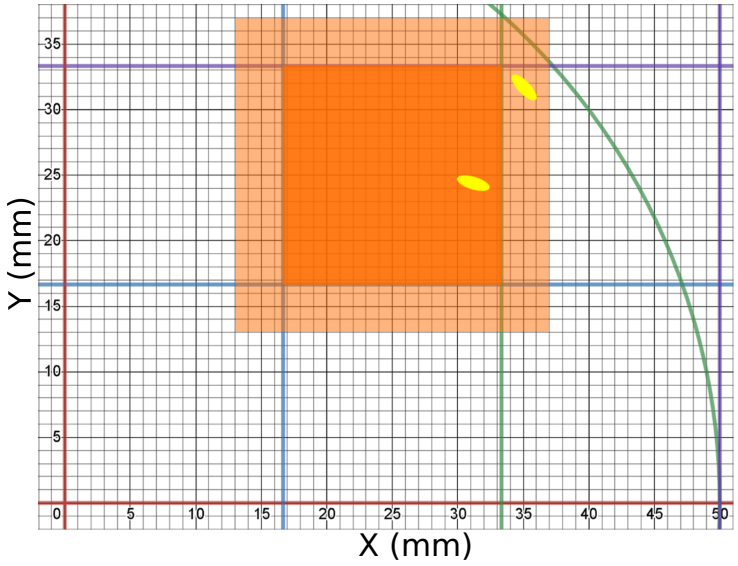

Figure 3.17: Illustration of two Drosophila, shown by the yellow ellipses, inside the arena. The arena is divided into square regions with side $\approx 16.7mm$ and with a border of width $\approx 3.3mm$, one of these region with its border is shown in orange. The event represented could be an encounter between the Drosophila since the one on the left is inside the region while the one on the right is inside the border of the region so they are considered in proximity.

Chapter 4

Data Processing

In this chapter it will be examined some features that can indicate weather or not the toy model created is a faithful dynamical representation of the real data. Then the statistics regarding the meeting will be analyzed in order to look for what can be considered a cognitive behavior.

4.1 Comparison between trails

Let us start with a visual comparison between the real trails and the random walks. In Fig.4.1 is shown on the left the trail previously shown in Fig.2.3 while on the right a random walk. The trail is from the experiment regarding caffeine and the random walk has been generated based on the distribution relative to caffeine. For both of the graphs are shown only 5000 out of 22000 points, this way there is not too much overlapping and it is possible to see clearly both the micro trends, such as the sharp turns and the straight paths, and the macro ones, such as the tendency of sticking to the walls.
The following observation in Sect.4.1.1 and 4.1.2 regarding the toy model will be done exclusively on one of the sets of random walks generated which contain 10 complete trails for each of the three experiments.

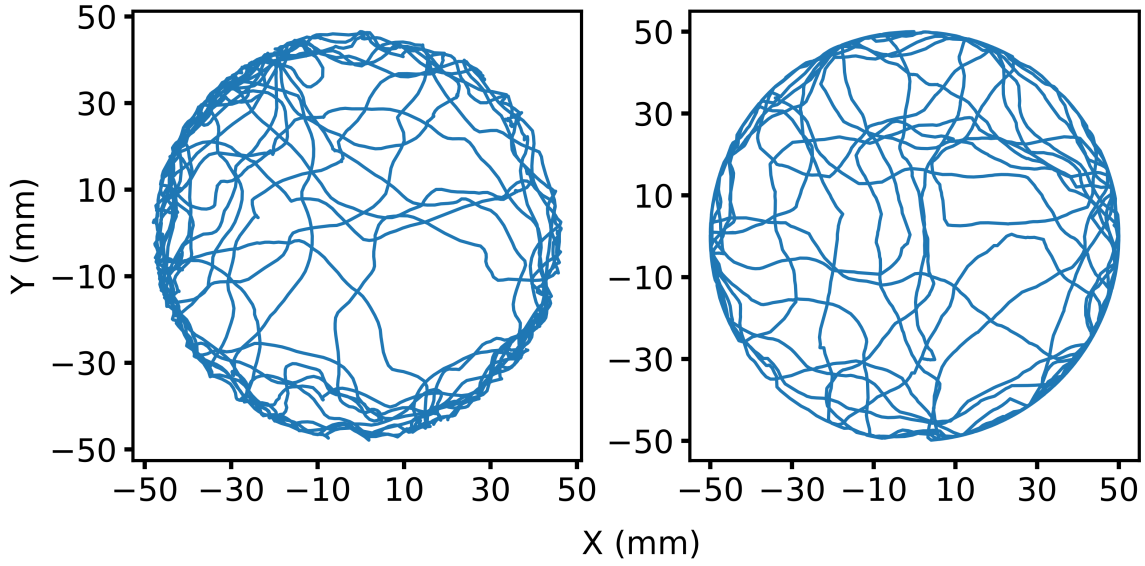

Figure 4.1: Evolution of the trail obtained from the real data (on the left) and from the random walk (on the right), both relative to the experiment regarding caffeine. Both trails are constituted by 5000 points, so $\approx 333.3sec$ of trail evolution. Except for the part next to the border the two trail appear very similar.

4.1.1 Distance to the wall

It is worth noting that even if in Fig.4.1 it may look like way less time was spent next to the border in the random walk compared to the real trail, this is not necessarily the case. In fact it could be caused by the fact that, when the real Drosophila run along the border, their distance to it varies approximately from $1mm$ to $4mm$, generating in the graph of the trail evolution a thick stripe next to it. Meanwhile when the random walk follows the border is often because it tends to exit it but is forced back on its circumference as explained in Sect.3.4.1, so all of the paths next to border tend to precisely overlap, removing the thickness.
It could be possible to consider the effect of a local repulsive force from the boundary to compensate this phenomenon. This effect is visible as well in the distribution observed in Fig.4.2. Here is shown the PDFs of different distances to the wall for both the real trails and the set of random walks, it has been graphed as well the CDFs obtained from the two.

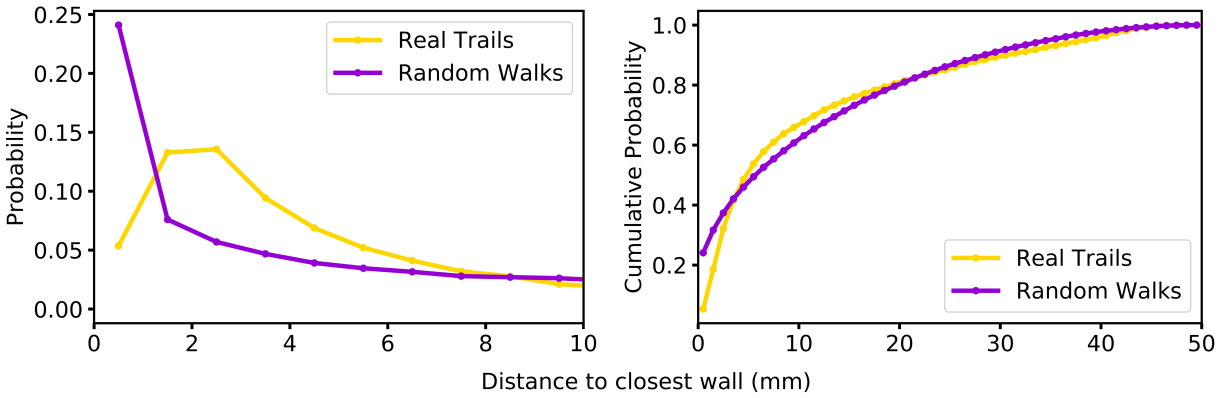

Figure 4.2: Probabilities and cumulative probability of the distance between a Drosophila and the closest wall. The yellow dots refer to all of the real data and the purple dots to one of the sets of random walks produced. It can be observed that while the PDFs differ notably, the CDFs share many similarities.

4.1.2 Rectifiability of the trails

For a better understanding of the meaning of the study on rectifiability, in Fig.4.3 is shown an example of the procedure which has been utilized.

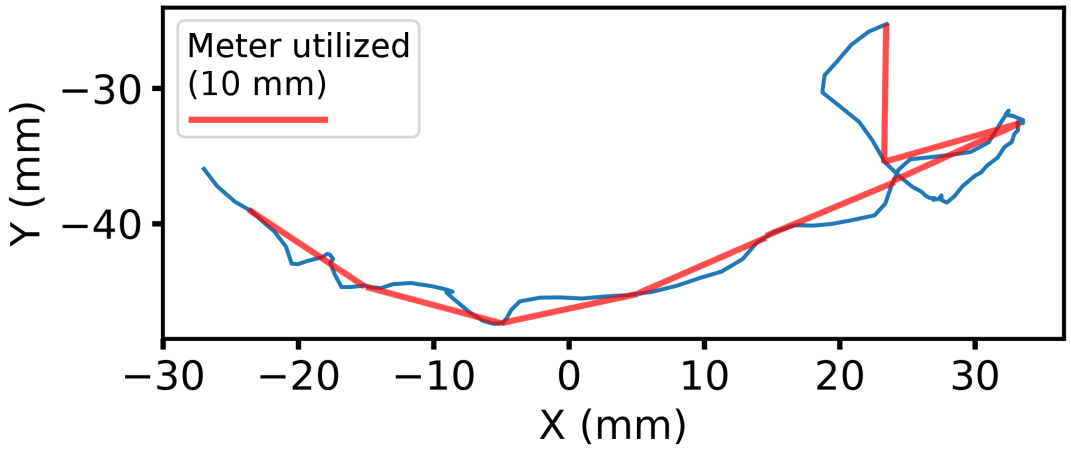

Figure 4.3: Illustration of the procedure followed to measure the extent of the trail with a given minimal unit of length. In this case this unit is $10mm$ and since 8 of these units are counted, the extend computed is $80mm$.

Here is visible a part of a real trail (in blue) which has for starting point the end of

the trail present at the top of the graph. The extent of the trail is measured using as minimal unit of length of $10mm$, which is shown by the red segments. Beginning at the starting point of the trail, which will be the first point of reference, it will be added to the extent of the trail $10mm$ every time that is found a following point of the trail at such a distance from the current point of reference, then the point found will be the next point of reference. Following this procedure, the extent of the trail in Fig.4.3 turns out to be $80mm$, since 8 minimal units of length are used. If the length were to change from $10mm$ to another value then the measure of the extend of the trail may vary as well.
It is now graphed the rectifiability of the trails, in Fig.4.4 is shown the average (obtained from the 10 trails) of the extent of the trails as a function of the minimal unit of length used to measure the trails. This average is then divided for the maximum value of the extent obtained, which for all of the graphs is around the $7mm$ mark on the y-axis. This means that for all of all of the three experiments and both for the real data and the toy model the maximum extent of the trail is measured when using a minimum unit of length of about $7mm$.
The discrepancies between the trends obtained from the toy model and from the real data which can be observer are that for caffeine and sugar when the minimal unit of length $< 30mm$, where the random walks have an average length of the trail higher in comparison with the real trails, and for ethanol when the minimal unit of length $> 30mm$, where instead the average lengths of the trails are higher.

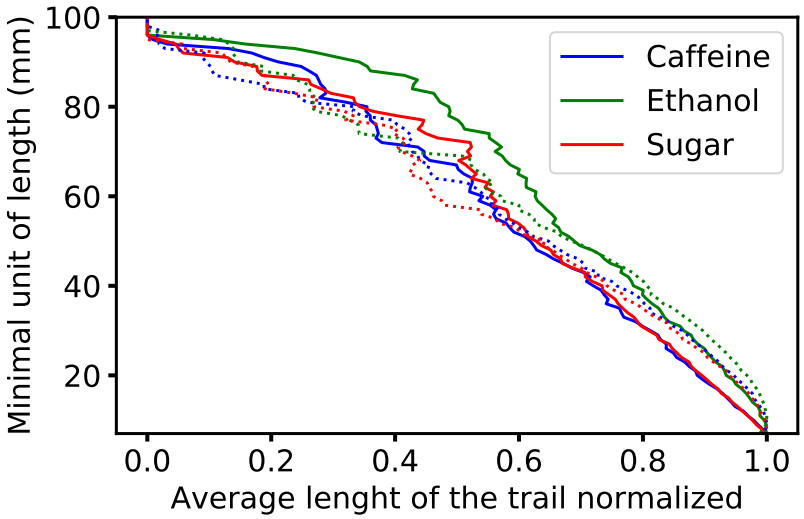

Figure 4.4: Average length of the trails, computed from the 10 trails, for different minimal units of length used for the measurement. The value is then divided by the maximum average obtained. It can be observed how the results of the toy model overlap with the ones from the real data.

4.2 Comparison between encounters

Here it will be compared the duration of the encounters obtained with the definition given in Sect.3.5. Then it will be shown how it was possible to associate a probability to these encounters.
In this section when referring to certain trait of the random walks, such as the number of encounters during the whole experience, it will be implied the average of the trait measured in the totality of the 40 sets of random walks generated. To the average it will be associated the standard deviation. The value of the traits regarding the real data will be shown with full dots and a continuous line. Meanwhile the average of the value of these traits regarding the toy model will be shown by the dashed line and its relative error will be highlighted with a light bar which surrounds it.

4.2.1 Duration and amount of the encounters

In Fig.4.5 are shown the frequencies of the duration of the encounters. It can be observed that all of the frequencies related to the real data exceed their toy model counterparts. In order to better compare the trends of duration of the encounters, in Fig.4.6 are shown four graphs, all with the probability of the different duration of the encounters. On the upper left is shown the one relative to caffeine, on the upper right the one relative to ethanol, on the bottom left the one relative to sugar and on the bottom right the three previous ones combined. It is noticed how the PDFs relative to the real data are often not inside the error bar of the toy model. This happens especially for the experiment regarding sugar, where it can be seen a rise in the probability of the duration of the meeting around the values of $3 - 5sec$.
In Tab.4.1 are presented, divided by drug type, the number of encounters performed both in the toy model, for which it has been computed the average and standard deviation, and in the real data. Is reported the ratio between the two as well.

	Caffeine	Ethanol	Sugar
Toy model	150 ± 15	139 ± 15	220 ± 16
Real data	235	298	420
Ratio	0.638 ± 0.064	0.466 ± 0.050	0.524 ± 0.038

Table 4.1: Number of encounters, divided by experiment, during the whole experience both for the toy model and for the real experiment. It is computed the ratio between the two.

The ratios between the number of the encounters relative to different drug types, both for the real data and for the toy model, have been computed. For the toy model the error has been propagated with the formula:

$$\Delta Ratio = \frac{N}{M}\sqrt{\frac{(\Delta N)^2}{N^2} + \frac{(\Delta M)^2}{M^2}}.$$

The results are reported in Tab.4.2.

	Ratios of number of encounters		
	Caff./Eth.	Eth./Sug.	Sug./Caff.
Toy model	1.079 ± 0.159	0.632 ± 0.082	1.467 ± 0.181
Real data	0.789	0.710	1.787

Table 4.2: Ratios between the different experiments of the number of encounters during the whole experience, both for the real data and from the toy model.

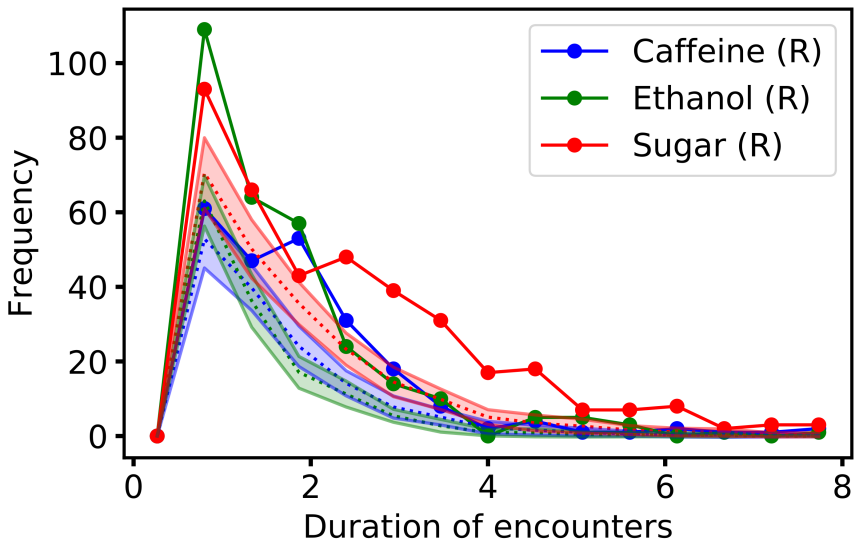

Figure 4.5: Frequencies of duration of the encounters for the real data and for the toy model. To the data of the toy model, which is the dashed line, is associated an error shown by the light bar. It can be observed how the frequencies of the real data exceed the one of the toy model.

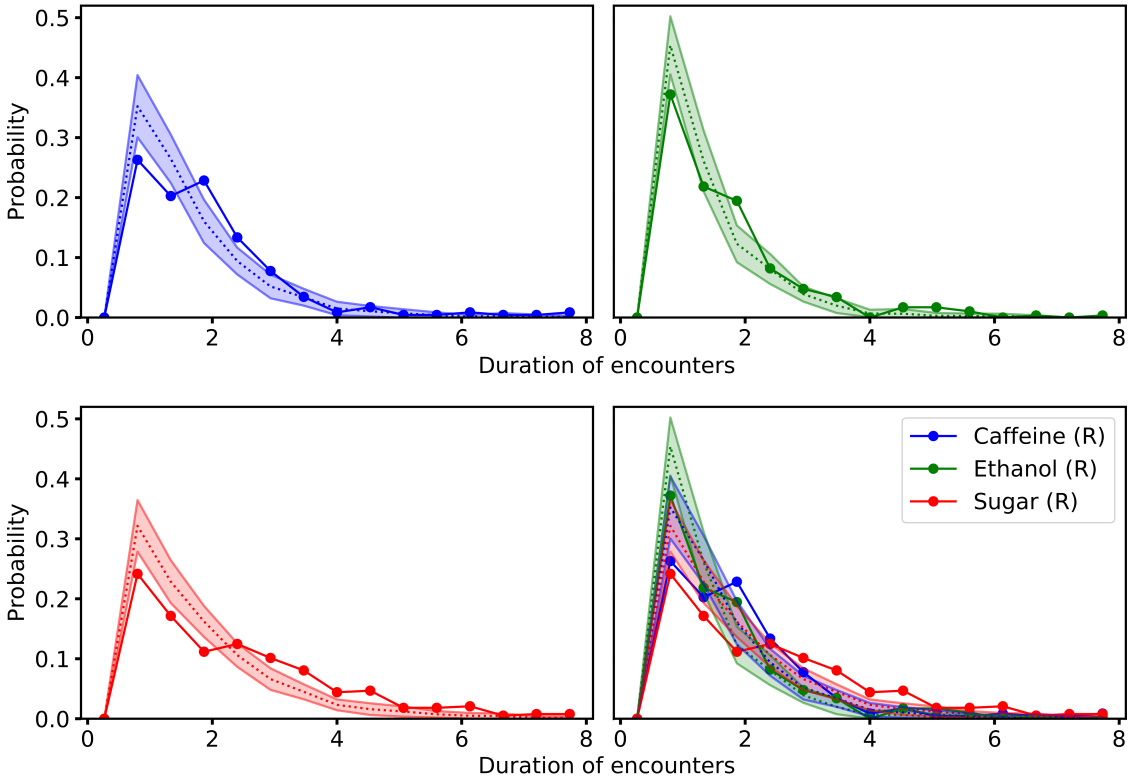

Figure 4.6: Probabilities of the duration of the encounters for the real data and for the toy model shown in order for: caffeine, ethanol, sugar and all combined. To the data of the toy model, which is the dashed line, is associated an error shown by the light bar. It can be observed how, in particular for the data relative to the experience with sugar, the distributions do no match.

4.2.2 Probability of the meeting combinations

In order to assign a given probability to the encounters made by a Drosophila it has to be considered the totality of the encounters made during the experiment. In particular, since we want to understand whether or not the Drosophila have any preferences regarding which other Drosophila to meet, we will have to assign a probability relative to the number of encounters accomplished with the other specimens. This is made possible using the binomial distribution. Indeed, if we consider that in theory each Drosophila has the same probability of meeting another specimen, the probability of meeting a precise specimen is $\frac{1}{9}$. Knowing this, the probability that a Drosophila who made a total of N encounters will do exactly m encounters with another given Drosophila is given by:

$$P(m; N, \frac{1}{9}) = \binom{N}{m} \frac{1}{9}^m (1 - \frac{1}{9})^{N-m} \qquad where \qquad \binom{N}{m} = \frac{N!}{m!(N-m)!}.$$

However this is the probability of the exact event, in order to have an estimate of the likelihood such an event it will be used its corresponding p-value. In a binomial distribution the p-value is given by the sum of all the probabilities in the bins as far or further from the center then the bin of the event which is examined.
Since we are not considering anymore the single encounter but the m encounters between a Drosophila and another specimen, to each Drosophila we will be able so associate 9 different p-values for a total of 90 p-values for experiment.
Now it will be studied the distributions of the values of the p-values for the three different experiments. In Fig.4.7 are shown three graphs, all showing the PDF of the p-values. The first is the one relative to caffeine, then there is the one relative to ethanol and at last the one relative to sugar. The main observation is the peak present both for sugar and ethanol. This peak is the manifestation of a large number of meeting combinations with a very low p-value, which means very unlikely. In Fig.4.8 are shown the CDFs of the previous PDFs. Here is possible to see the influence of the peak which causes the CDFs relative to caffeine and sugar to rise immediately. It is also observed how the CDFs relative to the toy model are indistinguishable and resembling a straight line, which implies that there is no particular outcome of p-value more likely then the others.

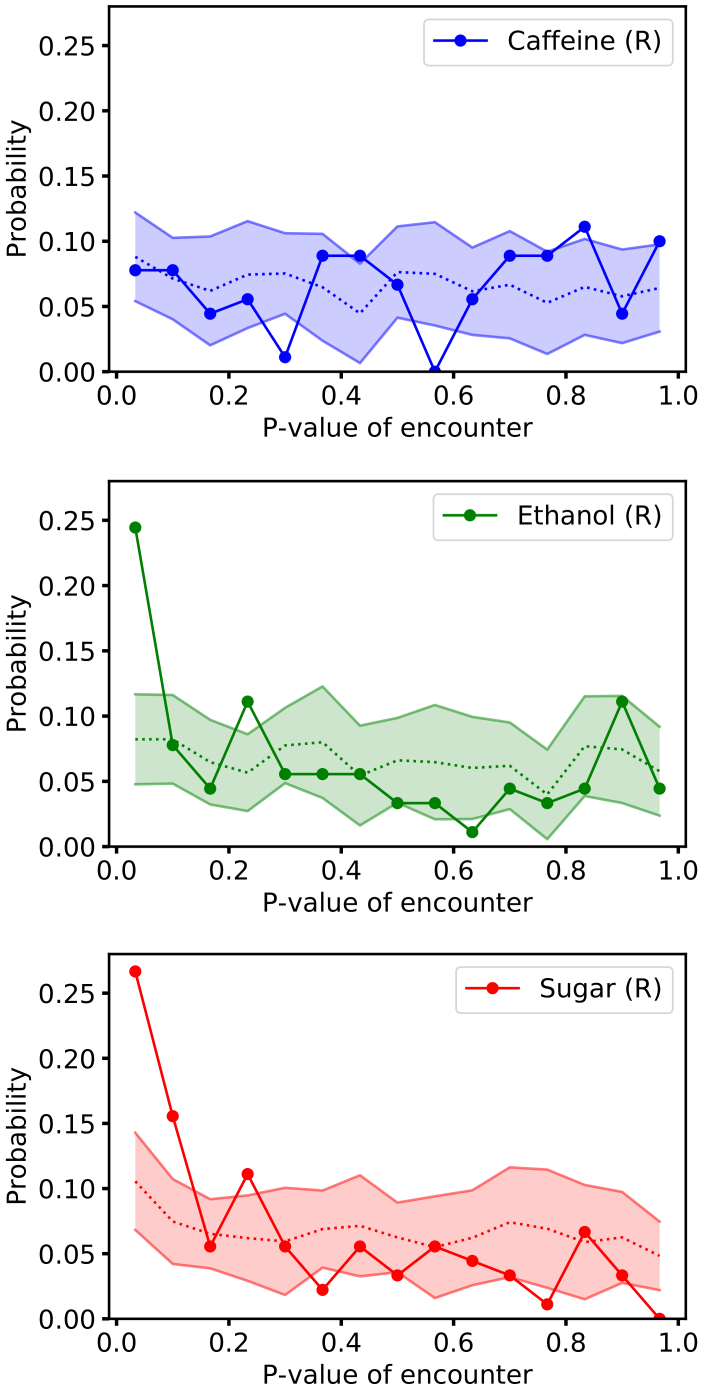

Figure 4.7: Probability distribution functions of the p-values of the meeting combinations for the real data and for the toy model, split in three graphs by drug type. The spike present in the graph relative to ethanol and sugar indicate the presence of several meeting combinations with a low p-value.

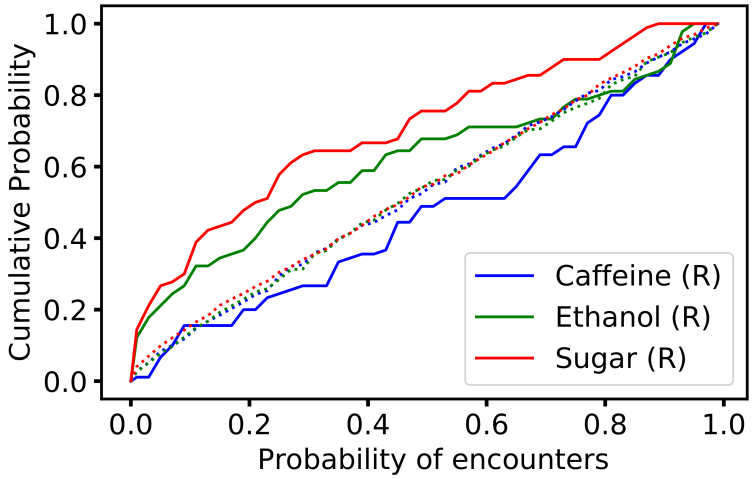

Figure 4.8: Cumulative distribution functions of p-values of the meeting combinations for the real data and for the toy model, split by drug type. It can be seen how the CDFs relative to ethanol and sugar have a spike for the low values and how the CDFs of the toy model resemble a straight line, meaning that there is no p-value more likely then the others.

In Tab.4.3 are listed, for all the three different drug and for the both the real experiment and the toy model, the number of meeting combinations (out of the possible 90) which have a p-value < 0.10 and < 0.05. The data regarding the toy model are the average, with the standard deviation, of the 40 random walks.

	p-value < 0.10 frequency		p-value < 0.05 frequency	
	Real Data	Toy Model	Real Data	Toy Model
Caffeine	14	10.65 ± 3.08	4	6.18 ± 2.61
Ethanol	24	11.10 ± 4.21	16	5.98 ± 2.64
Sugar	27	12.85 ± 3.56	20	7.75 ± 3.13

Table 4.3: P-values of meeting combinations frequencies under 0.10 and 0.05 for the real data and toy model, split by drug type.

In Tab.4.4 is shown, only for the real data and for the meeting combinations having a p-value (p-v) < 0.10 and < 0.05, how many of these meeting combinations have a ratio $\frac{occurrences}{avg.occurrences} > 1$ or < 1, where $avg.occurrences = \frac{tot.occurrences}{9}$. The *occurrences* are the number of times the two Drosophila involved in the meeting combination encountered and the *tot.occurences* are the total number of encounters made by a Drosophila during the experiment. A higher frequencies of meeting combination with the ratio $\frac{occurrences}{avg.occurrences} > 1$

indicate that the Drosophila involved in "rare" meeting combination encountered an unusually high amount of times, a ratio < 1 indicates the opposite.

	p-value < 0.10 frequency		p-value < 0.05 frequency	
	$\frac{occ.}{avg.occ.} > 1$	$\frac{occ.}{avg.occ.} < 1$	$\frac{occ.}{avg.occ.} > 1$	$\frac{occ.}{avg.occ.} < 1$
Caffeine	9	5	3	1
Ethanol	16	8	10	6
Sugar	15	12	13	7

Table 4.4: P-value of meeting combinations frequencies under 0.10 and 0.05, divided by the value of the ratio $\frac{occurrences}{avg.occurrences}$, so divided based on the peculiarity of the meeting combination: the two Drosophila involved in the meeting combination met an unusually high number of times if the ratio is > 1 and an unusually low number of times if the ratio is < 1.

At last, from the result in Tab.4.3 it have been computed how many standard deviations (σ) separate the number of meeting combinations (with a p-value < 0.10 or < 0.05) of the real data from their average of the toy model (parameter called $\sigma_{distance}$). It has been reported as well the probability of such a distance from the center of the distribution (μ), obtained by integrating the standard normal distribution in the intervals: $[-\infty, \mu - \sigma \cdot \sigma_{distance}]$ and $[\mu + \sigma \cdot \sigma_{distance}, \infty]$. The results are listed in the Tab.4.5.

	p-value < 0.10		p-value < 0.05	
	$\sigma_{distance}$	Probability	$\sigma_{distance}$	Probability
Caffeine	1.09	0.27572	0.84	0.40090
Ethanol	3.06	0.00222	3.80	0.00014
Sugar	3.97	0.00008	3.91	0.00010

Table 4.5: Distance, measured in units of standard deviation, between the number of meeting combinations with a p-value under 0.10 and 0.05 for the real data to the corresponding average obtained from the toy model. Is reported the probability of such a distance as well.

Chapter 5

Conclusions

The toy model

From the graphs shown in Fig.4.2 we can see how, even if the two PDFs of the distances to the wall appear very different, the relative CDFs show many similarities. In fact the two CDFs cross twice at $4mm$ and $20mm$ and they differ the most around 0-$1mm$. So, for this feature, the main discrepancy of the trails is when moving close to the borders, such as previously supposed in Sect.4.1.1. This difference will not influence the rate of the meeting since this difference disappears in a few millimeters, thus these distances are way below the size of the sides of the region ($\approx 16.7mm$). The improvement that it could have been applied is the influence of a stress which would force the points of the random walks closer to a $10mm$ distance from the walls since in this way the CDFs would likely overlap more significantly.
From the graph in Fig.4.4 we can observe the similarities of the rectifiability features of the toy model with the features of the real data counterparts for the different drugs, the trends are in fact very similar, with some exceptions. In particular the drugs for which the two trends differ the most are caffeine and sugar for the minimal unit of length $< 30mm$, where the random walks have an average length of the trail higher in comparison with the real trails, and ethanol for the minimal unit of length $> 30mm$, where instead the average lengths of the trails are higher. The first discrepancy is likely due to a higher likelihood of uncorrelated consecutive changes of velocity direction in the random walks; in fact, while the value of the changes are equally likely with the ones in the real trails, in the toy model there is no memory storage of the past variation of velocity direction so the random walks do not follow a main direction or direction variation as much as their real counterparts. This causes more oscillation of direction in the paths, which increases the length of the trails significantly if measured with a shorter meter. The second discrepancy could be caused by the fact that the Drosophila which received ethanol tend to stick to the walls more then the others. Doing a round trip of the arena (that has a diameter of $100mm$), the extent of the trails will be measure even if using a

minimal unit of measurement of $70mm$, which explains the difference observed.
Having proven how the toy model can be considered a satisfactory approximation of the spatial and kinematic features of the real trails, we move on to the study performed in Sect.4.2.

The proprieties of the meetings

The results shown in Sect.4.2.1 starts to exibit significant differences in the behavior between different sets of data, these dissimilarities appear both based on the type of drug administered and between the real data and the random walks. Observing the results in Tab.4.1 we can see that for the real data, for all of the drugs, there are unquestionably more encounters compared to the ones of the toy model. However only the ratios of these quantities regarding ethanol and sugar are comparable within the error, which means that in proportion either happened more encounter regarding caffeine in the toy model or less of these encounters in the real experiment, we will later try to pin down which of the two possibility is more likely. This inequality causes the fact that in Tab.4.2 only the ratio Ethanol/Sugar of the toy model comes out comparable within the error with the same result of the real experiment. What is instead remarkably similar are the graphs obtained of the probability of different duration of such encounters, which are shown in Fig.4.6. In these graphs the drug for which the PDF relative to the real experiment differs the most from the one relative to the toy model is sugar, where most of the probability shown are outside, even if not considerably so, the error bar obtained from the toy model. This indicates the will of these Drosophila to interact for longer times then what would be expected from the toy model.
The most interesting findings are reported in the following section, Sect.4.2.2. In Fig.4.7 it can be observed the similarities between the graphs relative to ethanol and sugar and how the common trend is completely lost for the one relative to caffeine. The first two in fact show a sharp peak of probability for the real data around extremely low p-values which is not present in the latter graph. This information is shown as well in the CDFs obtained which are shown in Fig.4.8, where the red and green continuous line present a high rise for low p-values; in this graph it can be seen how the averages obtained from the toy model resemble straight lines with slope = 1, which indicated that there is no particular probability of meeting combinations preferred. The peak mentioned above indicated the presence of several unlikely meeting combination, which means the presence of couples of Drosophila which met an unusual number of times. The quantity of these couples is expressed in Tab.4.3 and in Tab.4.4. In this second table the couples are divided based on the value of the ratio $\frac{occurrences}{avg.occurrences}$. This ratio indicates if the frequence of these couples is due to a reciprocal interest of meeting (ratio > 1) or to a reciprocal interest of avoidance (ratio < 1). As it is shown the number of meeting combinations drop drastically for caffeine between the constraint of p-value < 0.10 and of p-value $<$

0.05 while lose only about $\frac{1}{3}$ of the frequencies in the case of ethanol and sugar. This illustrates how a large amount of these combinations are extremely unlikely to happen for the experience with ethanol ($\frac{16}{90}$) and for the one with sugar ($\frac{20}{90}$) . Moreover, for all of the drugs, these meeting combinations are about twice as likely to have a ratio $\frac{occurrences}{avg.occurrences} > 1$, so these rare combinations refer to Drosophila with an intention to meet, not to avoid, each other. At last some more accurate observation can be made with the data listed in Tab.4.5. The two frequencies of meeting combination which have a p-value < 0.10 and < 0.05 relative to the experience regarding caffeine are comparable within the error. Meanwhile the frequencies of the meeting combinations for th experiences regarding sugar and ethanol are constantly higher for the real data and the significance of the difference between the frequencies is shown by the probability related to $\sigma_{distance}$. In fact for both drugs and both constraints $\sigma_{distance} > 3$, which causes the probability related to be almost insignificant. In particular for the case of the sugar this probability is ≤ 0.0001.

Overall results

It was possible to create a toy model based on the kinematic of the real trails which is an acceptable approximation of the real experiments and which has been used to study the abundance, duration and the probability of the encounters between Drosophila. From the first featured it emerged a significant higher tendency of meeting between real Drosophila and an imbalance in the number of the meetings performed relative to caffeine, although it was not possible to know what was the cause of this last finding. The higher tendency of meeting alone can be considered a primitive form of cognitive behavior so its proprieties have been examined further. From the study regarding the duration of the encounters it was possible to show that the Drosophila which were administered sugar were more likely to be part of longer meetings then what it was predicted by the toy model. Ultimately, by studying the probabilities of the meeting combinations it clearly emerged how the Drosophila in the experiments regarding ethanol and sugar performed several extremely unlikely meeting combinations, what could be considered social relationship between different Drosophila specimens. Another evidence of the existence of this behavior is the fact that it can be altered by the administration of drugs. In fact, while in the experience regarding ethanol the number of meetings performed drops but the behavior continues to show, such behavior is completely lost in the experiment regarding caffeine. This justifies as well the imbalance cited above which can now be safely assumed to be caused by the lack of development of social behavior between these caffeine-drugged Drosophila. This latter experience is for the most part comparable with the toy model, so with a set of data based on randomness. Summing up: it was possible to prove the existence of a cognitive behavior and the distinctive influences caused by the administration of different drugs.

Bibliography

[1] Robert G. SHULMAN. *Brain Imaging: What it Can (and Cannot) Tell Us About Consciousnes*. Oxford Scholarship Online, 2013, p. 59.

[2] Antonello PUZELLA Massimo ARAGONA Georgios D. KOTZALIDIS. "The many faces of empathy, between phenomenology and neuroscience". In: *Archives of Psychiatry and Psychotherapy* 4 (2013), pp. 5–12.

[3] Stephen B. HAGER. "The Diversity of Behavior". In: *Nature Education Knowledge* 4(2).66 (2010).

[4] Leticia SANCHEZ Micheal D. BREED. "What Functions of Living Systems Underlie Behavior?" In: *Nature Education Knowledge* 3(10).67 (2010).

[5] Leticia SANCHEZ Gro V. AMDAM. "Measuring Animal Preferences and Choice Behavior". In: *Nature Education Knowledge* 3(10).74 (2011).

[6] Eric EDSINGER Gul DOLEN. "A Conserved Role for Serotonergic Neurotransmission in Mediating Social Behavior in Octopus". In: *Current Biology* 28.19 (2018), pp. 3136–3142.

[7] Paul G. HELLIWELL Andrew B. BARRON Ryszard MALESZKA. "Effects of cocaine on honey bee dance behaviour". In: *Journal of Experimantal Biology* 212 (2009), pp. 163–168.

[8] Reza AZANCHI Galit SHOHAT-OPHIR Karla R. KAUN. "Sexual Deprivation Increases Ethanol Intake in Drosophila". In: *Science* 335 (2012), pp. 1351–1355.

[9] Jacques D. NGUYEN. "Prophylactic vaccination protects against the development of oxycodone self-administration". In: *Neuropharmacology* 138 (2018), pp. 192–303.

[10] Jiri HORACEK Filip TYLS Tomas PALENICEK. "Neurobiology of the Effects of Psilocybin in Relation to Its Potential Therapeutic Targets". In: *Neuropathology of Drug Addictions and Substance Misuse* 2 (2016), pp. 782–793.

[11] John BENDER Kristin BRANSON Alice A. ROBIE. "High-throughput ethomics in large groups of Drosophila". In: *Nature Methods* 6(6) (2009), pp. 451–457.

[12] Wikipedia. *Drosophila melanogaster*. URL: `https://en.wikipedia.org/wiki/Drosophila_melanogaster`. (accessed: 06.09.2019).