# Dirichlet Process Mixtures of Multinomials for Data Mining in Mice Behaviour Analysis

Matteo Zanotto, Diego Sona, Vittorio Murino
Pattern Analysis and Computer Vision
Istituto Italiano di Tecnologia,
via Morego, 30, 16163 Genova (Italy)
name.surname@iit.it

Francesco Papaleo
Neuroscience and Brain Technologies
Istituto Italiano di Tecnologia
via Morego, 30, 16163 Genova (Italy)

Dipartimento di Scienze del Farmaco
Università degli Studi di Padova
Largo Meneghetti, 2, 35131 Padova (Italy)

## Abstract

*Automatic analysis of rodents behaviour has received growing attention in recent years as rodents are the reference species for large scale pharmacological and genetic screenings. In this paper we propose a new method to identify prototypical high-level behavioural patterns which go beyond simple atomic actions. The method is embedded in a data mining pipeline thought to support behavioural scientists in exploratory data analysis and hypothesis formulation. A case study is presented where the method is capable of learning high-level behavioural prototypes which help discriminating between two strains of mouse having known differences in their behaviour.*

## 1. Introduction and Related Work

As already observed in many other disciplines, an increasingly wide quantity of data is being collected in many sectors of biological sciences. In particular, this has been a challenge in reference to the recent generation and study of an enormous amount of genetically modified mice around the world. While this opens huge opportunities for understanding the specific mechanistic role played by gene and gene mutations, scientists have to face new challenges to analyse them, especially when relevant information cannot be captured by simple statistics. This is often the case when dealing with complex phenomena such as relations, sequences, co-occurring events or repetitive patterns, especially if they are evolving over time. This observation calls out for new advanced methods for data mining which can support scientists in the stage of formulating new hypotheses. We hereby propose a method to support behavioural scientists in the complex task of understanding important traits of mice behaviour. This is achieved by summarising with an intelligible representation the statistical structure automatically extracted from the data.

While large datasets are required to evaluate behaviours in a statistically significant way, extensive video footage poses serious issues to be solved. First of all, the enormous amount of time required for watching wide video collections makes it impractical in most cases. For this reason the experiments are commonly confined to short time intervals. Secondly, it is almost impossible to evaluate complex phenomena like the occurrence of specific behavioural patterns by simply watching a video, especially if they extend over long time periods. Except for simple cases, mathematical formalism is needed to capture this type of information which might otherwise be lost.

The interest of both the computer vision community and the biological scientists in automatic analysis of mouse behaviour is witnessed by the increasing number of publications appearing in the field. Jhuang *et al*. [6] presented a computer vision system trained to recognise eight types of actions from video recordings of single mice living in their home cage. Burgos-Artizzu *et al*. [2] proposed a system to automatically recognise thirteen different actions relevant to mouse social interactions. In this case the dataset consists of a large collection of 10-minute videos showing two interacting mice. Weissbrod *et al*. [13] went beyond that and combined computer vision and a radio-frequency identification system to obtain actions of both individual mice and groups while interacting. Giancardo *et al*. [5] presented a machine learning system for mouse tracking capable of finely quantifying social and non social behaviours of multiple mice interacting for long periods of time.

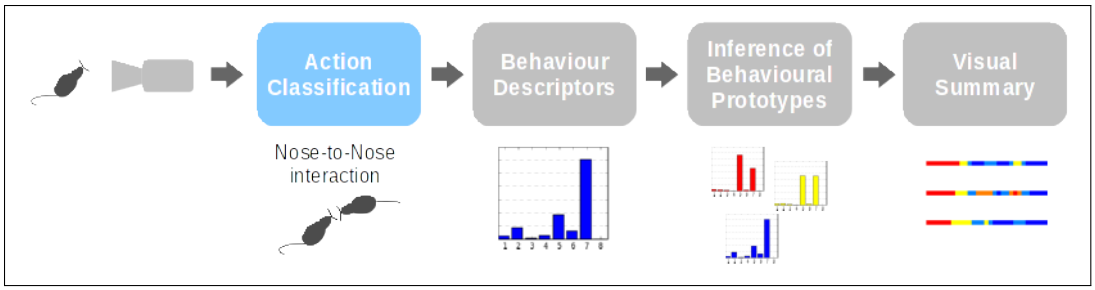



Figure 1. Pipeline of the proposed method. In order to address behavioural studies of different kind, action classification is considered as an independent module. Different action classification systems could be used for different studies. Detailed descriptions of each module are provided in Section 2.

While the common trend is focusing on single actions, the importance of behavioural patterns composed by a set of actions should not be overlooked. Many examples can be found in nature and one is the honeybees waggle dance. While the single actions of turning left and right carry no meaning if considered by themselves, the dance as a whole indicates direction and distance of new food sources. Behavioural patterns have also been investigated in ants by Reznikova *et al*. [12] who considered how simple actions (or behavioural units) are combined in hunting routines. Looking for patterns of actions is therefore important and can give a richer interpretation of the observed behaviours. This is especially true when studying phenomena linked to social interaction in animals. For this reason, we expect that fragmenting mice behaviour in single actions might result in a loss of information. Moreover, sequences of actions [8] and the repetition of behaviours [9] are important to highlight important features of animal behaviour which cannot be captured by analysing single actions independently. The work by de Chaumont *et al*. [3] goes in this direction as some of the events detected by the proposed system are the concatenation of two different elementary actions. In addition, the authors report an analysis of the transitions between successive actions as performed by two mouse strains and show the emergence of different patterns. While interesting, this work still focuses on single actions (or sets of two actions) which are codified in the system. An attempt to gain flexibility is the work proposed by Kabra *et al*. [7] where actions are not built into the model but learnt at runtime on the basis of what the experimenter annotates as interesting. This allows to perform a considerably wider variety of analyses but still requires knowledge about what behaviour the scientist is interested in.

In this paper we propose a method aimed at moving beyond these works by inferring behavioural prototypes spanning long time periods without any human intervention. This is especially useful both when interesting behavioural patterns can only be appreciated over long time intervals, and when such behaviours are not known *a priori*. Specifically, we present a pipeline for exploratory data analysis in mice behavioural experiments. Such behaviour-oriented data mining could be very useful to get a first understanding of the data and can guide the formulation of hypothesis and the subsequent design of further experiments.

The core of the method is represented by a Bayesian Nonparametric model (a Dirichlet Process Mixture of Multinomials) which is able to find an appropriate collection of behavioural prototypes in a completely unsupervised way.

These prototypes are then used to characterise mice behaviour and present the user with a high level aggregation of the data capturing salient behavioural features.

The paper is structured as follows: methodological details are provided in Section 2. A case study applying our pipeline to look for known behavioural traits of autism in mouse models is presented in Section 3, followed by concluding remarks in Section 4.

## 2. Behaviour Analysis with Dirichlet Process Mixtures

We assume that there exist a collection of prototypical behaviours that mice follow with some variation. The hypothesis is that mice follow these behaviours in different proportions and that the way of mixing them is useful to characterise important traits of the mouse itself.

Such behaviours must necessarily be at a higher level of abstraction than atomic actions in order to capture richer information. For this reason, behavioural patterns are commonly represented by neuroscientists as histograms of atomic actions over the entire experiment. In order to achieve a finer level of granularity, we compute histograms of actions over shorter time windows whose length is significant for the observed phenomena (see Sec. 3 for details). Underlying this choice is the assumption that the sequence of the behavioural prototypes is important, while the sequence of atomic actions is not stable enough to carry high-level significant information.

While we assume that a collection of behavioural prototypes exists, we do not impose any constraint on their shape or number. In particular, we allow such collection to include

potentially a countably infinite number of behaviours. To do so, we build a Bayesian Nonparametric model capable of inferring both the appropriate number of prototypes and their shape. Given the specific formulation of the problem, the model used is a Dirichlet Process Mixture of Multinomials. Even though alternative models like K-means exist, they would require to fix the number of the behavioural prototypes beforehand. This is a strong limitation as the choice would necessarily be arbitrary and could cause wrong results. Figure 1 presents the pipeline of the proposed method. Each stage is described in detail in the following sections.

### 2.1. Action Classification

Behaviours are considered to be high-level entities obtained by aggregating lower-level atomic actions. Any action classification system can be used in our pipeline which is independent of its specific formulation. For the experimental part of this work we relied upon the method by Giancardo *et al*. [5]. What happens briefly is that in each experiment a group of four mice is placed in an open arena and their activity is recorded for about one hour with a thermal camera. Individual mice are tracked and frame-by-frame action labelling is performed by a classifier (random forest) working on a pool of features. These include relative distance between mice, mice shape and movement measurements. For further details please refer to [5].

A set of eight actions of interest are recognised by the system. Six of them focus on the following social interactions: nose-to-body, nose-to-nose, nose-to-back, mouse above another mouse, mice standing together, mouse following another mouse. The remaining two actions pertain mice not interacting with others: walking alone, standing alone.

### 2.2. Building Behaviour Descriptors

Once action classification has been performed on a frame-by-frame basis, behaviour descriptors are obtained for each mouse. This is done by computing histograms of atomic actions over overlapping time windows spanning the whole video recording. The length of these time windows is fixed according to the phenomena scientists want to analyse. The amount of overlap is chosen to avoid introducing artefacts due to the specific time slicing.

### 2.3. Inferring the Behavioural Prototypes

As previously mentioned, a Bayesian model is built to describe the generative process of the observed behaviours. Since behaviours are described as histograms of actions, it is natural to model them as Multinomial distributions.

From a generative point of view, we assume that there exist a possibly infinite mixture of Multinomial distributions (each one representing a behavioural prototype) from which observations (*i.e.* actions) are generated. The histograms obtained as behaviour descriptors can hence be seen as empirical estimates of the parameters of the Multinomial representing the latent behaviour. The generative process is outlined in Figure 2. Thanks to the Dirichlet Process [4], the number of mixture components is unbounded [1] and is found during inference along with their parameters.

Inference is performed by Gibbs sampling [11] and the details of the specific algorithm are presented in Algorithm 1. Basically all the behavioural descriptors are considered as observations and the Gibbs sampler finds the mixture of Multinomial distributions which is needed to explain the observed behaviours. This is done by iteratively sampling the assignment of observation to mixture components on the basis of the conditional distributions $c_i|c_{-i}, y_i$, where $c_i$ is the assignment of observation $i$, $c_{-i}$ is the assignment of all the observations except for $i$, and $y_i$ is the value of observation $i$. Specifically, the probability of assigning observation $i$ to component $c$ given $c_{-i}, y_i$ is

$$P(c_i = c|c_{-i}, y_i) = b\frac{n_{-i,c}}{N-1+\alpha}\int F(y_i,\phi)dH_{-i,c}(\phi) \tag{1}$$

while the probability of generating a new component $c^*$ associated to observation $i$ is

$$P(c_i = c^*|G_0(\phi), y_i) = b\frac{\alpha}{N-1+\alpha}\int F(y_i,\phi)dG_0(\phi) \tag{2}$$

where $b$ is a normalising constant which can be ignored in the computation, $n_{-i,c}$ is the number of observations assigned to component $c$ except for observation $i$, $N$ is the total number of observations, $\alpha$ is the concentration parameter of the Dirichlet Process, $F(y_i, \phi)$ is the likelihood of $y_i$ given the parameters of the associated mixture component, $G_0(\phi)$ is the prior over the mixture components' parameters and $H_{-i,c}$ is the posterior over parameters $\phi$ given the prior $G_0$ and all the observations associated to component $c$ except for $i$.

In our specific parametrisation we have that $F(y_i, \phi) \sim Multinomial$ and $G_0(\phi) \sim Dirichlet$. Thanks to conjugacy of these two distributions we obtain that $H_{-i,c} \sim Dirichlet$ and that the integrals in equations 1 and 2 admit a closed form solution.

It is important to underline that the model is completely unsupervised and is hence suitable to perform data mining in situations like the one described, obtaining relevant information from data without imposing any constraint *a priori*.

### 2.4. Visualisation

In order to make the obtained behavioural prototypes useful, a proper visualisation needs to be presented to the scientists interested in the analysis. The visualisation

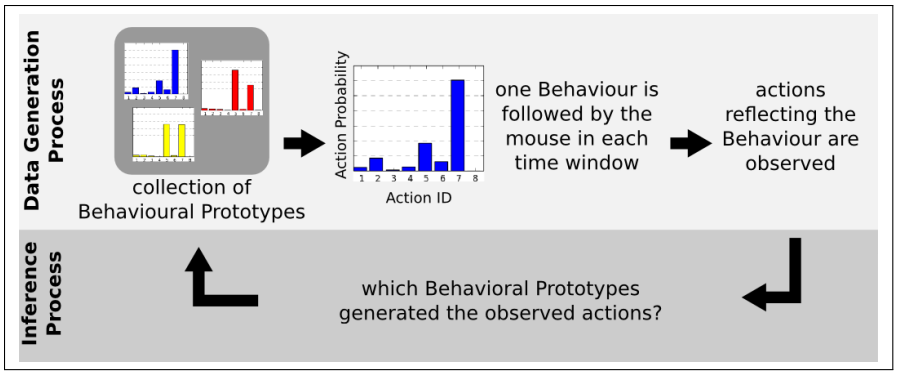



Figure 2. Outline of the data generation and inference processes. The data generation process describes how actions are generated from the collection of behavioural prototypes. The proposed method deals with the inference process estimating the latent behavioural patterns from the observed actions.

**Algorithm 1:** Gibbs Sampler

| | |
|---|---|
| $i$ | observation index |
| $t$ | iteration index |
| $y_i$ | value of observation $i$ |
| $c$ | component index |
| $c^*$ | index of new component |
| $c_i$ | component assignment of obs. $i$ |
| $c_{-i}$ | component assignment of all the obs. excluding $i$ |
| $n_{-i,c}$ | number of obs. assigned to comp. $c$ except for $i$ |
| $N$ | number of observations |
| $T$ | number of iterations |
| $K_t$ | number of mixture component at iteration $t$ |
| $c_i \vert c_{-i}, y_i$ | distribution of $c_i$ given $c_{-i}$ and $y_i$ |
| $\circ$ | is the conditioning set $c_{-i}, y_i$ |
| $\bullet$ | is the conditioning set $G_0(\phi), y_i$ |

**for** $t = 1 \dots T$ **do**
  /* Sample observation-to-component assignment */
  **for** $i = 1 \dots N$ **do**
    **if** *Observation $i$ is the only one associated to component $c$* **then**
      Remove component $c$ from the current mixture model
    **for** $c = 1 \dots K_t$ **do**
      $P(c_i = c \vert c_{-i}, y_i) \leftarrow$ eqn. (1)
    $P(c_i = c^* \vert G_0(\phi), y_i) \leftarrow$ eqn. (2)
    Draw a new value for $c_i$ from the conditional distribution $c_i \vert c_{-i}, y_i \sim$ Discrete with parameters $(P(c_i = 1 \vert \circ), \dots, P(c_i = K_t \vert \circ), P(c_i = c^* \vert \bullet))$
    **if** $c_i = c^*$ *is picked* **then**
      Add a new component to the mixture model

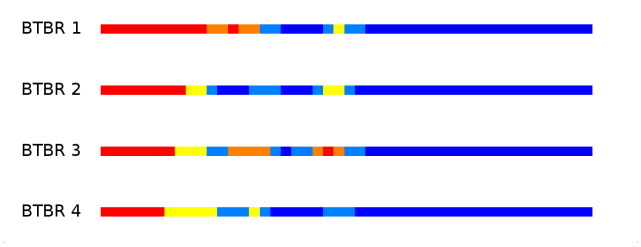
should be customary to the specific task, but in general colour coding of the behaviours can help understanding whether specific patterns tend to appear.

## 3. Experiments and Results

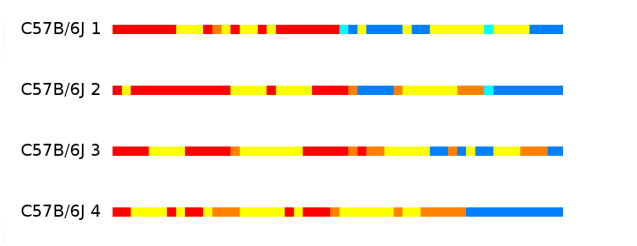
In order to evaluate whether the method is able to detect different types of behaviour, we analysed recordings coming from two different strains of mice: C57B/6J (commonly used as the reference background strain in genetically modified mice) and BTBR T+tf/J (a mouse model of autism-like behaviours). The two types of mice are known to have different behaviours [10] and serve well to our purpose. In particular, when compared to C57B/6J, BTBR T+tf/J mice are known to show repetitive and stereotyped patterns of behaviour and high inactivity levels [10]. A decreasing motility over time is also reported [6].

In order to test our system, a dataset with recordings from BTBR T+tf/J and C57B/6J has been used. Basically, four mice of the same strain were placed in an open-field arena and recorded for one hour with a thermal camera at 30 fps. Two cages of each strain were involved in the study, each having four mice. Every mouse was tracked and frame-by-frame action labelling was performed by the automated classifier. The obtained action labelling is taken as an input to find the behavioural prototypes. While Giancardo *et al*. [5] focused on atomic actions, in this work we go beyond that and focus on behavioural patterns. These patterns capture richer information than single actions through their distribution, which can express important aspects concerning how the animal is modifying its behaviour.

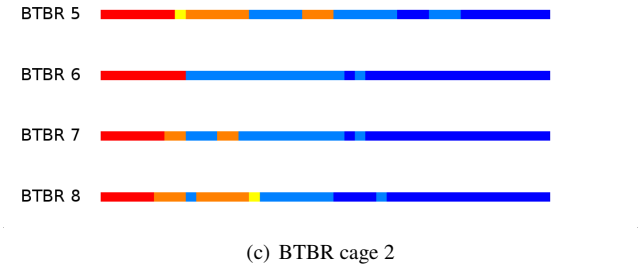
As a first stage, action histograms (i.e. our behaviour descriptors) are computed over 5-minute time windows sliding over the entire sequence with a 75% overlap between consecutive ones. This period of time has been chosen in order to capture interesting aspects of mice behaviour while avoiding the huge variability that would have been obtained considering shorter intervals. The high overlap, on the other



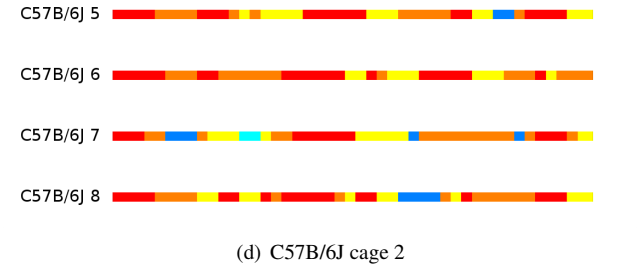

(a) BTBR cage 1



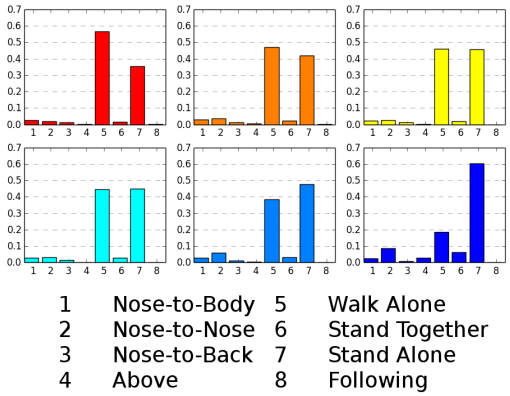

(b) C57B/6J cage 1




(c) BTBR cage 2




(d) C57B/6J cage 2




(e) Behavioural Prototypes

Figure 3. Results obtained by analysing the behaviours of BTBR and C57B/6J mice. 3(a)-3(d) show the classification of the behaviour of 16 different mice. Each line represents the recording of a single mouse and is divided in coloured squares representing the 5-minute time windows used for computing the behavioural descriptors. Each square is assigned the colour of the behavioural prototype associated to the time window. 3(e) shows the inferred behavioural prototypes.

hand, reduces the importance of the specific windowing and the likelihood of introducing artefacts due to the particular splitting.

Once the behaviour descriptors are computed for each mouse of each experiment, they are pooled together as the observations coming from the Mixture of Multinomials. The Gibbs sampler is then run on these observations in order to find the behavioural prototypes. Since the percentage of time spent by mice in performing the eight different actions is considerably different, the Dirichlet prior over the Multinomial parameters is built to reflect this imbalance. The results presented have been obtained setting $\alpha$ (see eqn.1 and 2) to 1 to favour the emergence of a compact set of prototypical behaviours.

Figure 3 reports the obtained results. Figures 3(a)-3(d) represent four different experiments which were considered in the analysis. Each line within the plots represents the complete recording of a single mouse and is composed by a sequence of coloured squares representing the 5-minute time windows. Such squares are colour coded on the basis of the behavioural prototypes they are associated with (see Figure 3(e)). The prototypes are ordered so that the ones with a higher proportion of motion-related activities (walk alone and following) are on the red end of the colour-map, while the blue end is associated to more static behaviours.

As previously discussed, the aim of the work is finding a

set of prototypical behaviours which can help neuroscientists finding interesting and recurrent patterns in animal behaviour. As a case study we analysed the recordings from BTBR T+tf/J and C57B/6J mice as they present demonstrated and clear differences in behavioural patterns (*e.g.* repetitive stereotypical behaviours and high inactivity levels in the BTBR T+tf/J strain).

By analysing the behaviour classification in Figure 3 we can observe the following:

1. The behaviour of BTBR mice tends to be more homogeneous (cross experiment) w.r.t. that of C57B/6J.
2. BTBR mice show a short initial period of high motility (characterised by the red colours) and then are more static and inactive (blue prototype characterised by high percentage of *Stand Alone* action i.e. inactivity). C57B/6J mice, on the other hand, have considerably higher motility throughout the whole recording.
3. BTBR mice tend to be more repetitive as shown by the fact that they keep performing the same behaviour for longer periods of time (i.e. the lines show longer segments with constant colour).

These three observations are basically revealing the characterising traits of the two mouse strains previously described. This suggest that the presented method is capable of finding behavioural prototypes capturing important and discriminating aspects of mice behaviour.

## 4. Discussion and Future Work

In this paper a pipeline to support behavioural scientists in data mining and exploratory data analysis has been presented. By modelling behaviours at a higher level of abstraction with respect to single actions, the proposed methodology aims at capturing richer information than other previous systems performing automatic classification of single independent actions. Being based on a Dirichlet Process Mixture, the model can infer behavioural prototypes in a completely unsupervised way, learning both their number and their shape from data.

A case study has been considered to evaluate whether the method could highlight differences in the behaviour of two strains of mouse. As reported in the experimental section, the inferred behavioural prototypes allow to visually assess the presence of traits characteristic of the autism-relevant BTBR mouse model. This result suggests that the method is capable of extracting interesting structure from the data and aggregating it in a way which is useful to scientists.

Given the promising results obtained, future work will be aimed at evaluating the method in other data mining tasks on behavioural experiments. Moreover, given its flexibility, the model could be tested in other domains characterised by phenomena evolving over time.

In conclusion, this new system might provide a new and highly effective tool in the assessment of complex behavioural patterns, applicable in large pharmacological and genetic screenings in mouse studies.

## References


[1] C. E. Antoniak. Mixtures of Dirichlet Processes with Applicatioins to Bayesian Nonparametric Problems. *The Annals of Statistics*, 2(6):1152–1174, 1974. 3

[2] X. P. Burgos-Artizzu, P. Dollár, D. Lin, D. J. Anderson, and P. Perona. Social Behaviour Recognition in Continuous Video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1

[3] F. de Chaumont, R. D.-S. Coura, P. Serreau, A. Cressant, J. Chabout, S. Granon, and J.-C. Olivo-Marin. Computerized Video Analysis of Social Interactions in Mice. *Nature Methods*, 9(4), April 2012. 2

[4] T. S. Ferguson. A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, 1(2):209–230, March 1973. 3

[5] L. Giancardo, D. Sona, H. Huang, S. Sannino, F. Managò, D. Scheggia, F. Papaleo, and V. Murino. Automatic Visual Tracking and Social Behaviour Analysis with Multiple Mice. *PLoS ONE*, 8(9), September 2013. 1, 3, 4

[6] H. Jhuang, E. Garrote, X. Yu, V. Khilnani, T. Poggio, A. D. Steele, and T. Serre. Automated Home-Cage Behavioural Phenotyping of Mice. *Nature Communications*, 1, September 2010. 1, 4

[7] M. Kabra, A. A. Robie, M. Rivera-Alba, S. Branson, and K. Branson. JAABA: Interactive Machine Learning for Automatic Annotation of Animal Behaviour. *Nature Methods*, 10(1), January 2013. 2

[8] J. Kain, C. Stokes, Q. Gaudry, X. Song, J. Foley, R. Wilson, and B. de Bivort. Leg-tracking and Automated Behavioural Classification in Drosophila. *Nature Communications*, 4, May 2013. 2

[9] M. Langen, M. J. Kas, W. G. Staal, H. van Engeland, and S. Durston. The Neurobiology of Repetitive Behavior: Of Mice.... *Neuroscience and Behavioural Reviews*, 35:345–355, 2011. 2

[10] H. G. McFarlane, G. K. Kusek, M. Yang, J. L. Phoenix, V. J. Bolivar, and J. N. Crawley. Autism-like Behavioural Phenotypes in BTBR T+tf/J mice. *Genes, Brain and Behaviour*, pages 152–163, 2008. 4

[11] R. M. Neal. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, June 2000. 3

[12] Z. Reznikova, S. Panteleeva, and Z. Danzanov. A New Method for Evaluating the Complexity of Animal Behavioural Patterns Based on the Notion of Kolmogorov Complexity, with Ants' Hunting Behavior as an Example. *Neurocomputing*, 84:58–64, 2012. 2

[13] A. Weissbrod, A. Shapiro, G. Vasserman, L. Edry, M. Dyan, A. Yitzhaky, L. Hertzberg, O. Feinerman, and T. Kimchi. Automated Long-Term Tracking and Social Behavioural Phenotyping of Animal Colonies within a Semi-Natural Environment. *Nature Communications*, 4, June 2013. 1