

Big behavioral data analysis:

Computational methods for the study of continuous recordings of behavior

Jose Antonio Espinosa Carrasco

---

TESI DOCTORAL UPF / 2016

DIRECTORS DE LA TESI

Dr. Cédric Notredame, Bioinformatics and Genomics Program

Dra. Mara Dierssen, Systems Biology Program

CENTER FOR GENOMIC REGULATION (CRG)



Universitat  
Pompeu Fabra  
Barcelona





A mis dos soletes, mis padres y hermano



*“Live as if you were to die tomorrow.  
Learn as if you were to live forever.”*

*Mahatma Gandhi*



## Acknowledgment

Foremost, I would like to express my gratitude to my two supervisors Cedric Notredame and Mara Dierssen, who gave me the opportunity to carry out this challenging project. I am particularly grateful for all the time they dedicated to guide me in the process of becoming a critical scientist and for providing me with the ability to work independently. From Mara, I learnt that with enthusiasm everything is possible, something that we should always keep in mind in our daily life. Furthermore, she makes me realize that you can always push your own limits. From Cedric, I learnt that science was not an engineering project. In the long way to understand a biological problem many times we get stuck or we arrive to dead ends; it is at this precise moments when it is more important to rethought the problem and find new angles that may be much more interesting than the initial ones. That is why he always finds some time to have an interesting scientific discussion with his students about their projects.

In addition to my supervisors, I would also like to thank the members of my thesis committee, Matthieu Louis, Fyodor Kondrashov and Robert Castelo, who help to bring interesting views and new thoughts to the project.

I further thank all my co-authors, since without them this thesis would have not been possible. Entre mis colaboradores me gustaría destacar a algunos de ellos. La primera mención por supuesto es para Ionas Erb. Para mi es algo más que un colaborador, ha sido mi postdoc de cabecera y un amigo que siempre ha estado ahí en los momentos difíciles del doctorado. Ionas el más “latino” de entre los alemanes, me ha ayudado a entender muchas de aquellas ecuaciones que en un principio, como buen biólogo de formación, me aterrorizaban. Yo a cambio he conseguido convertirlo a la religión de R aunque nunca conseguí hacerle abandonar sus scripts en editores de texto plano. Además le agradezco muchísimo toda la ayuda que he recibido por su parte en la revisión de este manuscrito y todas las discusiones científicas que he mantenido con él a lo largo de estos años. También me gustaría mencionar a Marta Fructuoso con la que sobretodo en los albores de la tesis compartimos muchos días de trabajo en las mazmorras del PRBB, el estabulario, y que cuando la situación parecía no poder ser más complicada siempre mostraba su mejor sonrisa. Con Silvina Catuara hemos compartido dos papers y aunque a veces los nervios y la presión del momento nos hicieron tener pequeños encontronazos, le tengo que agradecer toda la tenacidad que puso para que los proyectos salieran adelante.

Por supuesto me gustaría agradecer a todos los miembros que han pasado por mis dos grupos toda la discusión científica que hemos compartido y también aquellos momentos

más distendidos que hemos compartido, sobretodo en los retreats Salamanca y Granada, donde descubrí que además de compañeros de grupo a muchos de vosotros os puedo considerar amigos.

Merecen también una mención especial todos aquellos con los que comencé este camino hace ya 5 años Ramón, Javi, Irene, Marta, Jessica, Gabriel, Ignasi y Guillem. Con ellos he compartido muchos momentos tanto de discusión científica como personal y alguna que otra subrealista mientras tomábamos un café o alguna cerveza. Esta tesis ya ha valido la pena solo por tener nuevos amigos como ellos.

En un plano más personal me gustaría pedir perdón a mis amigos de siempre, Ester, Miguel, Moe, Néstor, Núria y Toni por todo el tiempo que no les he podido dedicar durante estos años. También al grupo de las cotorras y compañía. El objetivo lo merecía y espero que ahora os pueda compensar.

Me gustaría también agradecer a mi familia todo el respaldo que me han dado. A ellos también les debo una disculpa por la poca atención que les he podido prestar durante algunos momentos de esta tesis, muy a mi pesar. Mis padres, Gabriel y Maribel, han sido siempre un puntal muy importante en mi vida y si a alguien le debo poder haber escrito esta tesis es a ellos. Gracias a ellos me he podido dedicar a lo que me gusta, ya que siempre alimentaron desde pequeño la sed de conocimiento que cualquier niño posee. En algunos momentos no os lo demuestro pero podéis estar seguros de que os quiero y que os lo agradezco en lo más profundo de mi corazón. Mi hermano Gabriel también es una persona muy importante para mí. En estos últimos años me ha servido de ejemplo para darme cuenta de lo importante que es trabajar duro para conseguir lo que se quiere. Aunque parezca un poco raro el hermano pequeño ha dado ejemplo al mayor. Espero que en tu nueva andadura por las Américas seas feliz junto a Aurora y ambos podáis conseguir todo aquello que se propongan, tanto en un plano científico como sobretodo personal.

Por último la persona que se merece todos los agradecimientos del mundo es Pili. Eres mi mejor amiga, mi compañera y pareja. Creo que juntos hacemos un buen equipo y gracias a ti he podido ver las cosas de manera diferente en los momentos más duros de la tesis y valorar las cosas realmente importantes de la vida. Ahora compartimos también una hija, Sol, y todavía nos queda la parte más importante de nuestro camino juntos por delante. Sé que durante estos primeros meses no he podido estar a la altura pero ahora que la tesis llega a su fin espero poder compensarlo. Quién sabe, quizá algún día será Sol quien pensará en nosotros al escribir su tesis.



This research was supported by a PhD fellowship “Ajuts per a la contractació de personal investigador novell” FI FI-DGR 2012 from the Catalan Agency for Management of University and Research Grants (AGAUR) granted for Jose Espinosa-Carrasco.

The work was also funded by Fondation Jérôme Lejeune (Paris, France), EU (Era Net Neuron PCIN-2013-060), MINECO (SAF2013-49129-C2-1-R and BFU2011-28575), EU (Era Net Neuron PCIN-2013-060), AGAUR (“Grups consolidats” SGR 2014/1140) and Center for Genomic Regulation Severo Ochoa excellence grant SEV-2012-0208.



## Abstract

Behavioral measurements have long been considered a potential means to uncover underlying neural and molecular processes. However, limitations in the classical methods to study behavior have impeded progress in this field and have put the reproducibility and translational power of the results in question. A main problem is that stand-alone tests, due their reliance on pre-selected variables, reduce the complexity of the behavioral responses and do not capture their temporal dimension. Recently, technological improvements have made it possible to develop novel automated methods for the high-throughput screening of behavior. These methods make it possible to record fine-grained longitudinal behaviors in a systematic manner. Although this wealth of data presents a great opportunity to uncover the connections between behavior and the nervous system both in health and disease, the huge amount of data generated confronts biologists with a ‘Big Data’ problem that requires a new generation of data mining tools to be developed and deployed. Such tools involve integration of very diverse data sources, complex correlation analyses and multidimensional visualizations. In this thesis, I have developed novel tools for the exploration, integration and comprehensive analysis of high-throughput behavioral data. These tools, combined in a new software suite, “Pergola”, take advantage of existing genomic data standards developed to store and manipulate continuous data and their associated meta-data. I also explored multidimensional analysis techniques to help reducing the huge spatio-temporal dimensionality derived from behavioral recordings, and the high variability associated to all behavioral paradigms. Classical single-variate measures are insufficient in detecting the variance that explains the behavioral domain we wish to understand. This problem is addressed adapting Principal Component Analysis (PCA) for statistical inference on complex behaviors such as the recognition of learning strategies.



## Resum

La mesura del comportament s'ha considerat una eina per a descobrir els processos neuronal i moleculars subjacents. Malgrat això, les limitacions dels mètodes clàssics per a estudiar el comportament han impedit el progrés en aquest camp, i han posat en dubte la reproductibilitat dels resultats. Els principals problemes són que els test simples redueixen la complexitat de les respostes de comportament i no capten la dimensió temporal de la conducta perquè es basen en variables preseleccionades. Recentment, les millores tecnològiques han fet que esdevingui possible desenvolupar nous mètodes per a l'enregistrament d'alt rendiment i automatitzat del comportament. Aquests mètodes fan possible l'enregistrament de comportaments longitudinals detallats d'una manera sistemàtica. Tot i que la aquesta riquesa de dades presenta una gran oportunitat per a descobrir les connexions entre el comportament i el sistema nerviós tant en individus sans com malalts, l'enorme quantitat de dades que generen aquests sistemes confronta als biòlegs amb el problema del 'Big Data', que requereix el desenvolupament d'una nova generació d'eines de mineria de dades. Aquestes eines han d'involucrar la integració de dades de divers origen, l'anàlisi de correlacions complexes i la visualització multidimensional. En aquesta tesi, he desenvolupat noves eines per a l'exploració, la integració i l'anàlisi exhaustiva de les dades de comportament procedents de sistemes d'alt rendiment. Aquestes eines, combinades en un nou conjunt de programari, "Pergola", s'aprofiten de formats estàndards per a dades genòmiques pre-existents i que van ser desenvolupats per emmagatzemar i manipular dades contínues acompanyades de les seves meta-dades. També he explorat tècniques d'anàlisi multidimensionals per ajudar a reduir l'enorme dimensió espai-temporal derivada dels enregistraments de comportament, i l'alta variabilitat associada a tots els paradigmes de comportament. Les mesures clàssiques d'una sola variable no són suficients per a detectar la variància que pot explicar un domini conductual d'interès. He adreçat aquest problema mitjançant l'Anàlisi de Components Principals (PCA) per la inferència estadística de comportaments complexos com per exemple, el reconeixement de les estratègies d'aprenentatge.



## Preface

The study of behavior has been a longstanding question in neuroscience, since to understand brain function, it is not enough to uncover the molecular and physiological mechanisms taking place in the brain but we also need to contextualize and elucidate their relationship with the behavioral responses they encode. Moreover, from the clinical point of view, behavioral change is the only relevant outcome for diagnosing, treating and monitoring mental or neurological disorders.

The study of behavior, however, is not trivial. If we analyze its structure, it immediately becomes clear that behavior is a very complex, multidimensional process unfolding in multiple spatiotemporal scales, which cannot be accounted for using traditional stand-alone tests. Over the last ten years, a new generation of devices has made it feasible to acquire continuous recordings of behavior, following several animals during long time series. This revolution in technology allows for the high-throughput screening of behavior, which will lead to many opportunities for brain research, but will also pose challenges for the effective computational modeling of these data.

It is this precise problem what I have addressed in this thesis, by developing new methods for the visualization, integration and analysis of longitudinal behavioral recordings. The work presented here explores some new methods to visualize, handle and analyze critical aspects of behavior. This is well aligned with the interests of the Cellular and Systems Neurobiology Group at the Center for Genomic Regulation led by Mara Dierssen, who have been long interested in unraveling the dynamics of the development of disorders related to altered brain function. A main line of research of the group has been to understand disease as a complex phenotype with multiple behavioral manifestations. In this respect, classical behavioral studies are limited in terms of their temporal resolution and their oversimplified approach to tackling a multidimensional biological question.

The Comparative Bioinformatics Group, also at the Center for Genomic Regulation and headed by Cedric Notredame, has developed widely used bioinformatics tools mainly in the framework of multiple sequence analysis. However, we have also been interested in other types of sequences: temporal sequences analysis of different processes. In this regard, Cedric has collaborated with groups from the field of sociology to apply methods commonly used for the analysis of biological sequences in the modeling of life trajectories.

My work has taken an original approach: Capitalizing on the strong similarities between biological genomic sequences and longitudinal recordings, in which a given system can evolve across a discrete number of states over time, we decided to take advantage of

already existing genomic tools. The hypothesis was that longitudinal behavioral recordings can be expressed as sequences and can be dealt with accordingly. In the work associated with this thesis, we went one step further and decided to tap into the wealth of genome analysis tools developed for large-scale sequencing. Taking advantage of the strong conceptual analogy between high throughput sequences and automatically recorded behaviors, we have shown that existing genomic tools provide a natural standard for the storage, the visualization and the analysis of behavioral data. Moreover, in a multidisciplinary framework, my thesis aimed to provide the tools for studying neural disorders in an integrative manner taking into account all the possible factors influencing its development along time. Readouts resulting from the study of such complex processes as behavior are largely accompanied by noisy signals. Therefore, we have tried also to solve this question by using multidimensional approaches and in this way prioritize which parts of a complex, multidimensional signal could explain the behavioral phenotypes observed. The tools developed in this thesis should also help to provide new ideas to the field of behavioral neuroscience by offering a framework to adopt good practices in terms of computational reproducibility and replicability and improving data comparison and sharing between studies.



# Table of contents

<b>1. INTRODUCTION</b> .....	<b>1</b>
<b>1.1 Behavior to understand the nervous system and brain function</b> .....	<b>1</b>
1.1.1. Why do we study behavior? .....	1
1.1.2. Classical approaches to the study of behavior .....	2
<b>1.2. The challenges of studying behavior</b> .....	<b>4</b>
1.2.1. Behavior is a multidimensional, complex process .....	4
1.2.2. Temporal dimension of behavior .....	5
1.2.3. Behavior is highly variable .....	7
<b>1.3. Automated recordings of behavior</b> .....	<b>8</b>
1.3.1. Towards high-throughput behavioral analysis.....	8
1.3.2. Main benefits of automated behavioral recordings .....	9
<b>1.4. Analysis of big behavioral data</b> .....	<b>13</b>
1.4.1. Visualization of big behavioral data .....	13
1.4.2. Analysis and visualization of longitudinal big behavioral data.....	15
1.4.3. Analysis and visualization of multidimensional big behavioral data .....	17
<b>1.5. Case studies</b> .....	<b>19</b>
1.5.1. Obesity as a case study of automated longitudinal behavioral data. 19	
1.5.2. The MWM as a model for the study of a multidimensional behavioral cognitive process.....	21
<b>2. OBJECTIVES</b> .....	<b>23</b>
<b>3. RESULTS</b> .....	<b>25</b>
<b>3.1. Pergola: a new paradigm for longitudinal data visualization and analysis</b> .....	<b>25</b>
<b>3.2. Time-course and dynamics of the development of obesity-related behavioral changes induced by two types of energy-dense foods in mice</b> ....	<b>43</b>
<b>3.3. Principal Component Analysis of the Effects of Environmental Enrichment and (-)-Epigallocatechin-3-Gallate on Age-Associated Learning Deficits in a Mouse Model of Down Syndrome</b> .....	<b>75</b>
<b>3.4. Combined treatment with environmental enrichment and (-)-epigallocatechin-3-gallate ameliorates learning deficits and hippocampal alterations in a mouse model of Down syndrome</b> .....	<b>91</b>
<b>4. DISCUSSION</b> .....	<b>129</b>
<b>5. CONCLUSIONS</b> .....	<b>133</b>
<b>6. LIST OF COMMUNICATIONS</b> .....	<b>134</b>
<b>7. APPENDIX A</b> .....	<b>135</b>
<b>8. BIBLIOGRAPHY</b> .....	<b>145</b>



## **Abbreviations**

BBD: Big Behavioral Data

HMM: Hidden Markov Models

IR: infrared

LDA: Linear Discriminant Analysis

MWM: Morris water maze

PCA: Principal Component Analysis

RDoC: Research Domain Criteria

RFID: Radio Frequency Identification

WT: Wild Type



# 1. INTRODUCTION

## 1.1 Behavior as a way to understand the nervous system and brain function

### 1.1.1. Why do we study behavior?

Behavior has been defined as the “observable activity of an organism; anything an organism does that involves action and/or response to stimulation” (Wallace, Sanders, and Ferl 1991). Behavioral responses involve a coordinated pattern of sensory-motor and associative neural activity (Beck et al. 1981), functions that are under the direct control of the nervous system and especially the brain (Anderson and Perona 2014). Behavior refers to all the actions that organisms perform innately, but also those shaped to adapt to a changing environment in the broadest sense of the word (Mery and Burns 2009). Some of them can be genetically hardwired and thus constitute evolutionary species-dependent traits that are already observed in the neonate (Roubertoux et al. 1996; Suh et al. 2004). However, organisms have also the ability to vary their behavioral patterns in response to environmental changes, a property called behavioral plasticity (Sugi, Nishida, and Mori 2011). Adaptation, learning, or memory leading to changes in adult behavior, are among experience-dependent behavioral plasticity changes. As an example of behavioral modifications induced by environmental factors through integration of sensory input we can take the case of *Drosophila* larvae chemotaxis (Davies, Louis, and Webb 2015). This behavioral plasticity involves a large spectrum of behavioral mechanisms and properties that are dependent on many physiological patterns, and is usually associated with neuronal plasticity, although it may also include other changes.

We can discriminate two types of behavioral responses: innate behavioral responses and learned behaviors (Mazur 2015). The former consists of all the behavioral responses that have been selected through evolution to respond to environmental challenges and are selected at a population level over many generations. Hence, a behavior is innate, genetically determined when an animal can display it without learning due to the “hard-wiring” of the nervous system (Breed and Sanchez 2010). Genetic determinants constitute an important basis for innate behavior, but ontogenetic and, sometimes lifelong shaping by learning is required in addition, to fine tune this innate behavior. Learning would thus relate to the capacity of individuals to change behavioral responses using the previous experience (Mery and Burns 2009). However, we cannot consider innate behavioral responses and learning as isolated behavioral phenomena. These behavioral responses are intrinsically related, since the innate capacity of

individuals to perform a behavior and the individual experience received during their lifetime interact to shape the final behavioral response. However, this differentiation provides us with a useful conceptualization for the study of behavior.

Given that behavior is under the direct control of the brain, its study gives us an excellent way of understanding complex brain functions (Anderson and Perona 2014; Gomez-Marin et al. 2014). By studying behavioral responses strongly related to these brain functions, and how they emerge and evolve in order to overcome environmental challenges (Breed and Sanchez 2010), we can obtain valuable information on how these neuronal functions are shaped (Chiel and Beer 1997). This is because neural systems adapt to detect, process and respond to environmental cues by modifying behavior. Behavior can also change as a consequence of altered brain function, as occurs in genetic or environmental brain disorders (Mega et al. 1996; Silverman et al. 2010; Dierssen 2012). These disorders disrupt the ability to express environmentally appropriate behavior. Thus, the study of behavior could also provide important hints on brain dysfunction and could also establish pre-symptomatic behavioral deviations as robust biomarkers of brain disorders. As an example, diagnostic tools for mental disorders benefit from the increasing knowledge of fundamental programs of behavior, boosting initiatives such as the research domain criteria (RDoC) project, a new framework for classifying mental disorders based on observable behaviours and neurological measures (S. E. Morris and Cuthbert 2012).

However, the study of behavior is an onerous task given its complexity, high dimensionality and dynamicity. Besides, its nature is multiscale and displays multiple dependencies. Under this conceptual framework, one part of the work on my thesis has been devoted to the development of new tools for the study of innate behavioral responses such as behavioral eating patterns. Another important part of the work of this thesis was dedicated to the creation of a multidimensional framework for the study of learning-related behavioral responses.

### 1.1.2. Classical approaches to the study of behavior

Historically, two main approaches have been followed to study animal behavior. On one hand, ethologists have tried to understand innate behavior of free animals in their natural environment (K. Lorenz 1937; Tinbergen 1951; Von Frisch 1955). The ultimate aim of ethology is to understand how innate behaviors are organized and those evolve over time, under the premise that behavior is under natural selection as any other adaptive trait (K. Z. Lorenz 1958). On the other hand, psychology-influenced schools designed behavioral assays in controlled environmental settings to understand parts of complex behavioral phenotypes with a more reductionist approach. The concept behind

this is that those behaviors expressed in a controlled environment are representative of physiologically relevant processes or of their disruption. This field focused on learned behaviors, which could be predicted, controlled, trained and measured in a systematic manner. Examples of such paradigms include the classical conditioning (Pavlov 1927) and the operant conditioning (Skinner 1937)<sup>1</sup>. These paradigms propose that the environment can be described by its effect on behavior. In this manner, neutral stimuli would have no predictable effects whether presented or withdrawn, reinforcing stimuli are associated with an increase in behavior when presented following that behavior, aversive stimuli are associated with decrease in behavior when presented following this given behavior, and finally discriminative stimuli are a special type of antecedent stimuli that set the occasion for behavior to occur. This field established the basis to explain learning and motivation-developing principles.

These early schools, psychology and ethology, laid the foundations for the development of disciplines such as neuroethology and behavioral neuroscience. The former tries to understand the neural mechanisms behind natural behaviors following Krogh's principle. Krogh proposed "for a large number of problems there will be some animal of choice, or a few such animals, on which it can be most conveniently studied". Under these conceptual frameworks, neuroethologist studied innate behaviors of specialist animals in natural environments. Therefore, they did not focus only on common model organisms and regular experimental procedures, but instead, their approach consisted in characterizing behaviors as a complex sequence of smaller components that are organized and evolve over time in natural environments and in trying to find the neural basis of the behavioral patterns they observed (Hoyle 1984).

Instead, psychology-based approaches address the same question of how the nervous system orchestrates specific behaviors in a controlled but usually stand-alone manner, and with special emphasis on processes such as learning, memory, motivation, perception or disease (Donaldson 2010). This discipline uses a hypothesis-driven approach, in many instances with a more translational-oriented focus. To do so, in most cases model organisms are generated, often genetically modified with a gene or genes of interest, to create suitable paradigms for the study of specific disease-related behaviors,

---

<sup>1</sup> Classical conditioning experiments rely on associative learning processes by which a natural response to a stimulus in the environment (unlearned) is paired to a neutral stimulus, becoming then the neutral stimulus also associated (learned) to the natural response the former produced. On the other hand, operant conditioning experiments study associative learning by reinforcement processes. This means that a given behavior can be strengthened or weakened in function of its consequences. In this manner, if a given behavior produces a reward, the individual will repeat it, conversely, if it produces a punishment, the individual will avoid its repetition.

behavioral expression of internal states, or behavioral problems. Species, such as rats (*Rattus norvegicus*), mice (*Mus musculus*), zebra fish (*Danio rerio*), the fruit fly (*Drosophila melanogaster*) and the round worm (*Caenorhabditis elegans*) have been the common choice for this type of studies.

Obviously, both approaches present clear limitations that often lead to irreproducible or irrelevant results (J. C. Crabbe, Wahlsten, and Dudek 1999; Berry M. Spruijt et al. 2014). First, the use of stand-alone behavioral strategies does not allow the reproduction of real situations. Second, until recently, it was only possible to track a reduced number of dimensions that do not capture the full complexity of behavioral response. This rather reductionist approach has led to an isolated view based on a single or limited number of parameters. In the last years, reproducibility issues and the failure of many preclinical studies to be demonstrated in clinical trials have made clear that the study of a complex behavior cannot rely on a limited number of isolated parameters usually based on the experimenter's opinion of appropriate measures since relevant pieces of behavioral information could be undetected or get discarded.

## 1.2. The challenges of studying behavior

### 1.2.1. Behavior is a multidimensional, complex process

Behavior is multidimensional both in terms of the different biological aspects that are involved in a behavioral response (behavioral domains) and in terms of the quantifiable parameters that can be used to measure behavior (Hersen 2005).

From a biological perspective, psychologists designed experimental setups that constrained the possible behavioral outcomes, based on the hypothesis to be validated. The idea is that a specific task will be informative of a given behavioral dimension (Stepanichev et al. 2014). In this way, this approach tries to avoid tackling at the same time all the spectrum of possible behavioral domains of a behavioral response that unfolds in a natural environment. However, a prerequisite for measuring behavior in such way is to be able to operationally define the target behavior. This is not trivial, because behavior is a high-dimensional problem with distinct but related dimensions that introduce incremental degrees of freedom. To reduce the possible outcomes, neuroscientists have created well-controlled situations, in which animals are forced to choose among a small discrete set of behaviors. Widely used paradigms of this type are the open-field test for the study of locomotor activity and anxiety (Hall 1934) or the Morris water maze (MWM) (R. Morris 1984; Vorhees and Williams 2006) and the Barnes maze (Barnes 1979; Fiona E. Harrison et al. 2006) for the study of spatial learning and memory. The behavioral output of these tests is still evaluated today by



single-variate measures. However, behavioral domains establish complicated dependencies that are not always easy to disentangle (Gallagher and Burwell 1989). In addition, this oversimplified approach assumes that a single behavioral domain can be grasped by a single test (John C. Crabbe and Morris 2004) or that to avoid false positive results it is sufficient to perform a battery of tests evaluating the same behavioral domain (Crawley 2008).

From a more quantitative perspective, a very simple behavior, such as moving from a point A to a point B can be measured by multiple variables, including the distance travelled by the individual, the time spent to travel the distance, the speed, the tortuosity of the path, etc. In this way, a behavior that seems rather simple can be described in multiple ways, and could include factors influencing the behavioral response, such as environmental changes or cues. Each of these measures could inform about distinct behavioral domains to a different degree. For example, the time spent to travel the distance, or the speed, can help us understand the motor component of the behavior. However, a shorter execution time can also be informative of the cognitive capabilities of the animal, and more precisely of spatial learning, if the path has been previously learnt. Hence, these measures display clear interdependencies that can hinder the comparison between different individuals.

As we have just discussed, these classical single-variate measures are an oversimplified, subjective approach to the problem of understanding different behavioral domains (Graziano, Petrosini, and Bartoletti 2003). As such, although a simplistic approach can be sometimes helpful in terms of the tractability of the problem, univariate measures do not address the problem of understanding behavioral multidimensionality in its whole complexity.

### 1.2.2. Temporal dimension of behavior

The brain orchestrates behavior as a set of actions to appropriately respond to a fast changing environment (Hong et al. 2015) while evaluating the individual state, which also evolves along time. In this sense, behavior is an adaptative process that acts on different temporal scales depending on the environmental cues to which it responds. For this reason, the nature of behavior is intrinsically dynamic (Marr 1992; Benjamini et al. 2011; Gomez-Marin et al. 2014). Behavioral processes often involve transitions between different states, single events and cyclic episodes whose duration varies over time (Goulding et al. 2008). These rhythmic episodes affect most behavioral categories including eating, drinking, cage exploration, self-grooming, or social interaction. But the time scale of this variation can be different across lifetime. For instance, rodents maintain a stable rank order of time engaged in different behaviors through

development, and the time animals are engaged in these behaviors changes with age, displaying a clear ontogenetic pattern (Castro and Andrade 2005). To identify all these discrete behavioral episodes along a time course, behaviors are segmented or parsed. Segmentation of behavior poses a conceptual challenge since the temporal scale at which different behaviors unfold differs, exhibiting also a hierarchical organization (Gomez-Marin et al. 2014). For instance we can segment the behavioral trajectory of a rodent into circadian active and inactive periods. However, active periods at the same time can be further annotated into feeding, drinking and resting events that are organized on a shorter time scale (Goulding et al. 2008). Figure 1 illustrates this point by representing several behaviors that unfold on different time scales and with hierarchical relationships (Egnor and Branson 2016). Taking as example motion, we can see how steps are short-duration, less variable behaviors (bottom left) with a low position in the hierarchy (behavioral primitives or sub-behavior). These sub-behaviors can be shared by several high-level behaviors. For instance, running or foraging share steps as a ‘sub-behavioral’ annotation. High-level behaviors tend to be more variable and unfold over longer periods of time (top right). Therefore, behavior is multi-scale both in terms of time and variability.

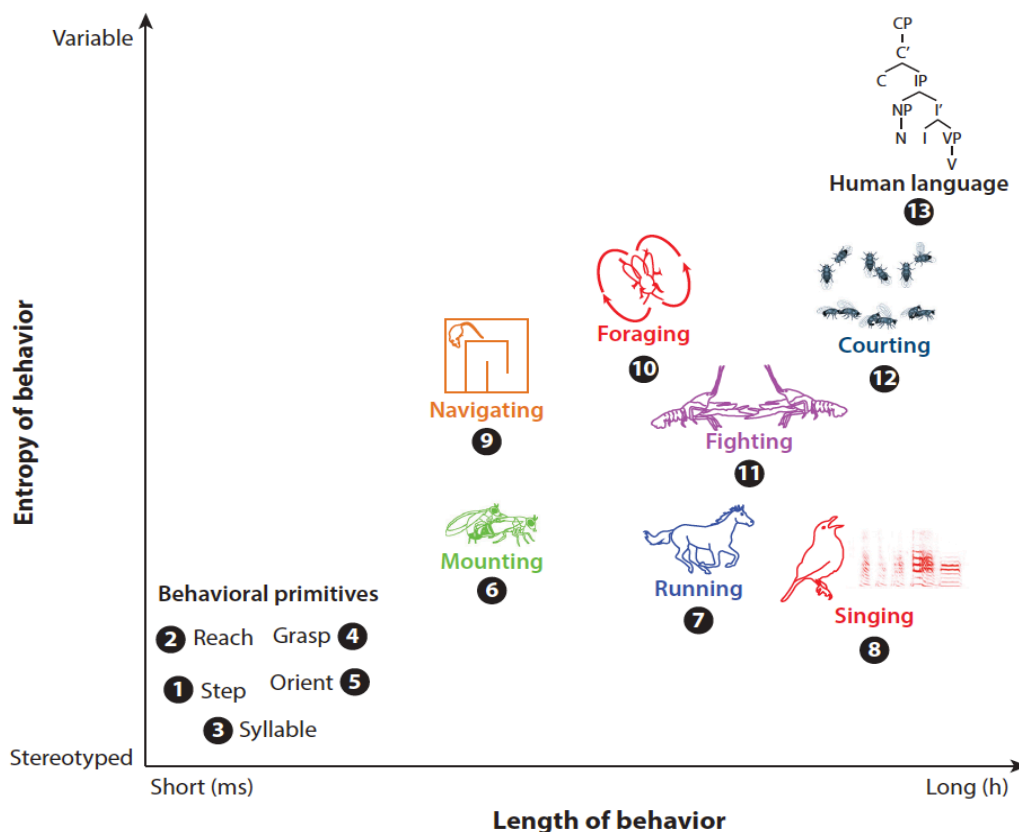


Figure 1 | Behaviors have to be measured on different time and variability scales. Behavioral primitives unfold in a short time scale and display low variability. Conversely, high-level behaviors unfold on longer time periods and display higher variability. There is a hierarchical relationship between high-level behaviors and behavioral primitives that allow the segmentation of the former into repeated elements of the latter (adapted from Egnor and Branson 2016).

Once a temporal sequence of behaviors is characterized, it is possible to infer the causal associations between environmental changes, behaviors and the physiological mechanisms behind the behavioral response. It can also help to understand whether specific properties, such as for example irregularity in neonate behavior, are intrinsic to the dynamics of motor activation or the result of random perturbations. Furthermore, the comparative analysis of how altered behaviors diverge from normality along time in a neural disorder can provide very relevant hints on the neural basis of behavior both in health and disease (Tecott and Nestler 2004). All these aspects reinforce the concept that to grasp the whole significance of behavioral repertoires, we need thorough time series analyses that would reveal more details about the behavioral trajectories. Each behavioral feature also has a rich dynamical profile with differential properties and patterns of oscillations. Thus, the temporal dimension emerges as indispensable to correctly grasp the complexity of the behavioral response.

### 1.2.3. Behavior is highly variable

Individual differences in behavior and their underlying mechanisms are an increasingly relevant research field. Behavioral readouts present a high degree of intraindividual and interindividual variability (Baker 2011). Individual behavioral tendencies can be temporally consistent, meaning that the individual's traits stay similar across time. However, they can also present variability even on short temporal scales due to specific processes such as habituation, or sensitization. Behavioral tendencies can be heritable to some extent, may have significant fitness consequences, and may be organized hierarchically, so that multiple traits correlate to form a higher organizational level. This variation in traits is added to the variation in behavioral responses measured during an experiment, i.e. "task performance". In fact, these variations are highly adaptive, since a strong consistency in behavior would limit individual capability to adjust behavior to deal with a new situation in an optimal way (Niemelä et al. 2013).

The sources of this intra- and inter-individual variability are diverse. Interindividual differences and similarities in patterns of intraindividual changes can be due to variation in individual lives, i.e. variability in developmental trajectories. Besides, they can be molded by the diverse biological and social conditions the individual is subjected to. All these differences that occur over the lifetime give rise to unique patterns of behavioral responses. Environment is an important factor accounting for experimental variability in behavioral studies that can interact not only with intrinsic genetic factors, but also with external perturbations, whether imposed or not, and hence affect experimental results (Richter, Garner, and Würbel 2009; Reardon 2016). For these reasons, a huge effort has

been devoted to standardizing laboratory practices by homogenizing environmental conditions. Remarkably, those efforts to reduce variability resulting from the environment have been unfruitful (J. C. Crabbe, Wahlsten, and Dudek 1999) and therefore, some authors argued that the solution is just the opposite: not trying to standardize environmental conditions among laboratories, under the assumption that if a result is robust enough it would be replicated even in heterogeneous experimental conditions regardless of the variability caused by the environment (Crawley 2008). This approach will avoid that results from a given laboratory are not just explanatory for this rather artificial and isolated environment (Richter, Garner, and Würbel 2009).

In this context, we could consider experimental variability just as the noise of a multidimensional, dynamic readout such as behavior. However, when we perform an experiment we are indeed interested in grasping the part of the variability that can explain the differences among animals grouped by categories of interest such as genotype, treatment, time, or combinations of these, or other factors. Therefore, the problem that we face is how to distinguish the part of the variability that explains the behavioral process from the noise inherent to a complex readout such as behavior. To address this problem, some authors proposed mathematical approaches for the modeling of complex behaviors (Wolfer and Lipp 2000; Graziano, Petrosini, and Bartoletti 2003; Stephens et al. 2008; Kumar et al. 2011). These approaches rely on multidimensional techniques to reduce the noise of the multiple variables automatically recorded when studying complex behaviors. To this aim, they take advantage of automated behavioral recordings systems that allow the extraction of a richer variable space. One of the challenges is to extract the relevant information from these multidimensional processes. Traditional behavioral evaluation methods have focused on observer-defined categories to reduce behavioral dimensionality. More complete records of an animal's behavior, such as those obtained from automated imaging and data analysis tools require new approaches to extract the relevant information, and one possibility is to extract lower-dimensional representations

## 1.3. Automated recordings of behavior

### 1.3.1. Towards high-throughput behavioral analysis

Sometimes technological advances are the force unleashing revolutions in a scientific field. As examples, the development of microbiology would have been impossible without the invention of the microscope, X-ray crystallography allowed the discovery of DNA structure (Watson and Crick 1953) opening the era of molecular biology and, more recently, next-generation sequencing produced a paradigm shift in genomics by allowing us to pose questions at a genome-wide scale and across large population

samples (Koboldt et al. 2013). New behavioral phenotyping technologies represent a breakthrough in the field of neuroscience and blur traditional borders between disciplines such as psychology, ethology and physiology (Schaefer and Claridge-Chang 2012). In parallel, the emerging field of computational behavioral analysis has been defined as “the use of modern methods from computer science and engineering to quantitatively measure animal behavior” (Egnor and Branson 2016).

Indeed, these two breakthroughs of biological technology (next-generation sequencing and high-throughput behavioral phenotyping) are closely related. The expansion of genetics and genomics, thanks to next-generation sequencing, highlighted the lack of high-quality phenotypic data that could help to uncover the genotype-phenotype relationship (Robert Gerlai 2002).

Automation defines a major shift in the way behavior can be studied, bringing substantial improvements to behavioral science. In short, some of the main benefits that automation offers is the possibility to observe behaviors that are otherwise undetectable for the naked eye, in different temporal resolutions, and during short or long periods of time. Besides, it allows to record single or multiple variables, on one or several individuals, and to reduce confounding factors derived from the handling of the animals or from subjective researcher criteria. Another benefit is that it enables to use more natural experimental setups and to use less time and human power, hence increasing experimental throughput. In this section, we will introduce how these systems can help to tackle the two different types of behavioral problems we have studied in this thesis. For an extended review of the current state-of-the-art of automated recording systems see Appendix A.

### 1.3.2. Main benefits of automated behavioral recordings

Systems for the automated recording of behavior provide neuroscience with unprecedented detailed data for the investigation of behavior. These new automated systems thus provide some obvious advantages. For instance, they allow for long-term continuous observations. Besides, automated systems limit the handling and transportation of the animals, reducing stress levels and hence avoiding biased results. They also allow for experiments consisting of multiple protocols optimizing the use of animals. Another important advantage is the possibility of recording for extended periods of time, which is especially relevant for non-invasively measuring disease progression. On the whole, the main feature that summarizes all the benefits of these automated systems is that they enable the observation of the temporal dynamics of behavior, i.e. how behavior evolves along time. This overcomes the limitations of the

common strategy of performing behavioral assays that can only grasp the behavioral state at a given moment. The resulting longitudinal, fine-grained recordings can be annotated generating organized temporal sequences of behaviors that resemble exhaustive classical ethological descriptions in the form of ethograms (Branson et al. 2009; Schaefer and Claridge-Chang 2012). Figure 2 displays an example of these annotated behavioral trajectories together with two of the continuous measures tracked by the automatic computer vision system (speed and angle), CTRAX (Branson et al. 2009). Interestingly, we can observe how speed and angle correlate or anti-correlate with the annotated behavioral bouts.

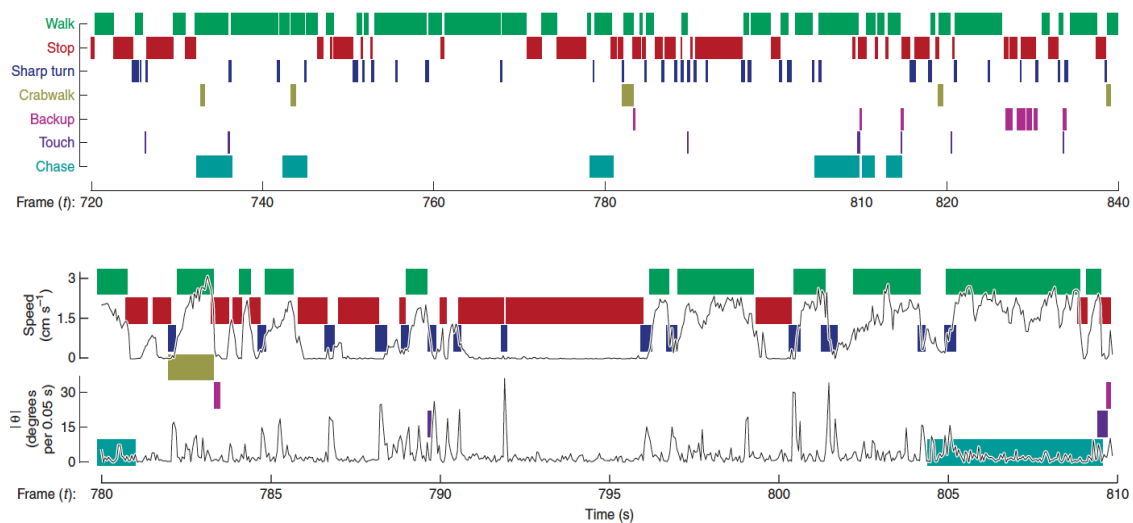


Figure 2 | On top a typical ethogram resulting from computer vision algorithmic tools (CTRAX in this case) to classify behaviors based on the measures obtained by the video tracking of multiple animals. Colored bars represent the duration of several classified behaviors. On the bottom classified behaviors with two of the motion variables obtained from the video tracking. Remarkably, the variations displayed by the variables can be explanatory of the annotated behaviors (adapted from Branson et al. 2009).

In contrast with experiments that are performed only at a given time point or during a reduced period of time, longer behavioral recordings can provide useful insights into changes of behavior that otherwise might remain undetected. To illustrate this point, we can use the work of Fonio et al. (Fonio, Golani, and Benjamini 2012; Fonio, Benjamini, and Golani 2012). One of the measures used to quantify anxiety-like behavior is the percentage of time the animal spent in the center of the arena within a certain period of time<sup>2</sup>. Interestingly, Fonio et al. found opposite results when comparing BALB/c mice,

<sup>2</sup> Rodents have a natural aversion to open spaces to avoid the possible hazards, however, they also show a strong tendency towards the exploration of novel environments. The open-field and related paradigms take advantage of these ethological behaviors to assess anxiety-like behaviors. Decreased anxiety will thus be characterized by an increased exploratory behavior, while on the contrary, increased anxiety will result in a preference to stay close to the walls of the field (thigmotactic behavior) (Ramos 2008).

an inbred mouse strain which displays elevated levels of anxiety (Belzung and Berton 1997) versus its wild progenitor, *Mus musculus domesticus*, during short and long time periods (Figure 3). During the first 30 minutes of the experiment, BALB/c mice displayed higher levels of anxiety-like behavior as measured by reduced percentage of time spent in the center of the arena in comparison with wild mice. Nonetheless, during the rest of the following 45 hours period the situation is reversed. Besides advocating for the importance of habituation phases to minimize the possible confounding effect of the so-called habituation to novelty process (Fonio, Benjamini, and Golani 2012), this work highlights how taking into account the temporal dimension of behavior could lead to distinguish transient from stable behavioral states.

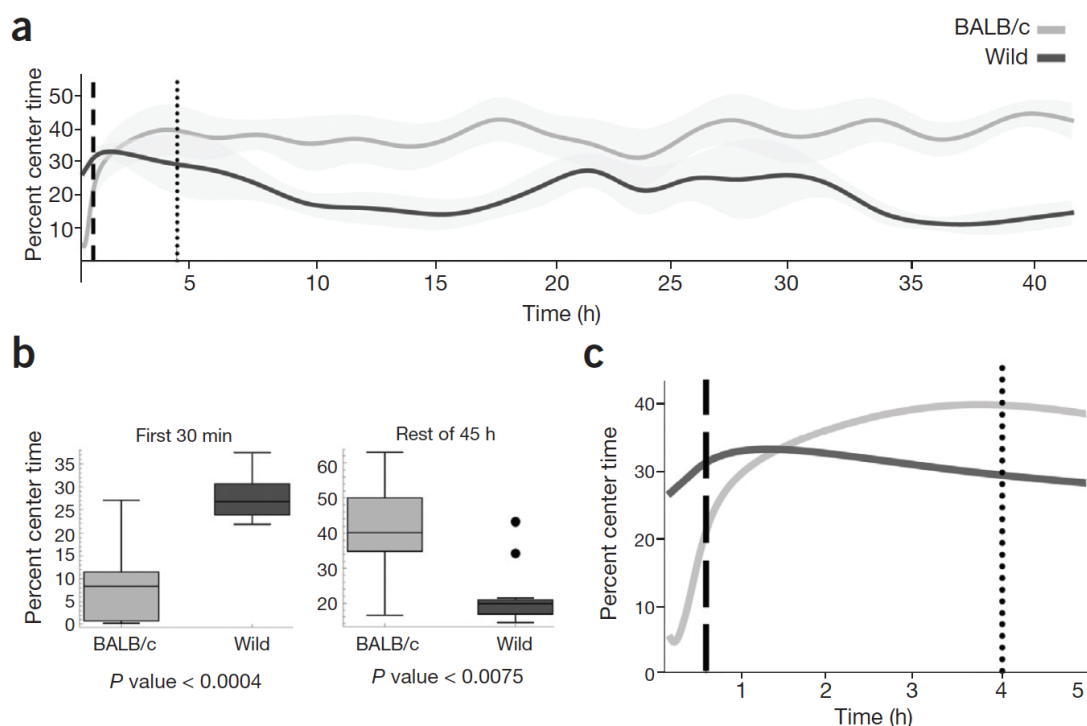


Figure 3 | Comparison of levels of anxiety-like behavior between BALB/c and wild mice. (a) Percentage of time spent in the center of an open field recorded during 45 hours as mean of the group (curve) +/- s.e.m. (gray shadow). The dotted line represents the quantified habituation period that should be used in order to avoid the detection of transient anxiety-like behavior. (b) Boxplots representing the comparison of anxiety-like behavior during the first 30 minutes in the open field, common duration of the habituation phase and the rest of the 45 hours of recording. Box edges represent lower and upper quartile and the black bar horizontal line depicts the median of the groups. Minimum and maximum values within a maximum of 1.5 times the interquartile distance from the box are represented by the whiskers. Extreme values are represented as single dots. (c) Percentage of time spent in the center of an open field during the first 5 hours of the recording comparing BALB/c vs. wild mice. Lines represent the same time periods that in a. BALB/c n=12; wild mice=9. (Fonio, Golani, and Benjamini 2012)

Automated systems for the recording of behavior also provide an unprecedented level of resolution (Steele et al. 2007). As shown in Figure 2, video-based system can track behaviors with time resolutions beyond seconds. Even though sensor-based system provide a slightly coarser signal (Schaefer and Claridge-Chang 2012), usually in the range of seconds, both types of systems provide a fine-grained signal that can be used to detect subtle changes of behavior on very short time scales (Goulding et al. 2008; Heyne et al. 2009; Wiltschko et al. 2015). In a nice example, Heyne et al. used these high-dense behavioral readouts to model the fine structure of feeding events in rodents (Heyne et al. 2009).

Another advantage of automated trackings is that they allow the recording of spontaneous behaviors closer to the natural behavior exhibited by animals since they are measured in “semi-natural” setups (Tecott and Nestler 2004; Schaefer and Claridge-Chang 2012). In fact, classical stand-alone behavioral paradigms exploit behaviors that animals exhibit in ecologically relevant conditions to obtain reliable results (R. Gerlai and Clayton 1999). In the open-field test, which exploits the natural tendency of rodents to avoid open spaces, anxiety-like behaviors are reported by assessing the balance between two opposed natural behaviors: the motivation to explore new environments and the necessity to avoid predators. Conversely, automation enables a more ethological perspective of behavioral studies (Tecott and Nestler 2004; Schaefer and Claridge-Chang 2012). Although it can be argued that experiments are still performed in controlled laboratory setups and with widely used model organisms (John C. Crabbe and Morris 2004), it is nevertheless also true that progressively, neuroscientists have developed more natural environments (Galsworthy et al. 2005; Lewejohann et al. 2009; Rudenko et al. 2009).

Automation of behavioral recordings also increased the number of possible readouts from behavioral experiments. In fact, the indexes derived from behavioral paradigms in the past were very coarse and prone to subjective scoring (Wahlsten et al. 2003; Patel et al. 2014). However, video-based monitoring systems together with algorithmic tools for the analysis of the data provide richer readouts (see Appendix A) that provide a more global and objective assessment of an animal’s performance on a behavioral test.

In addition, the combination of both semi-natural and experimentally constrained automated recordings can be very enlightening with respect to different aspects of the behavior by providing complementary views of a single question. Thus, combined approaches offer nice frameworks to work out different aspects of the behavioral response (see Appendix A for detailed examples). Two additional benefits of these systems are that first, they allow the social housing of animals and second, they minimize the handling of animals, hence reducing two of the most common



confounding factors found in many behavioral studies: social deprivation (Würbel 2001) and handling stress (Sorge et al. 2014).

Behavioral experiments have been historically a very time-consuming and expensive task, due to their manual nature both in terms of conducting experiments and annotating the readouts (Schaefer and Claridge-Chang 2012). Researchers were forced to reduce the sample size of the experiments at the expenses of statistical power. Therefore, the fact that automatic systems make the recording of multiple animals easier can help to increase the number of animals recorded at the same time. Besides, automatic recording systems allow for the tracking of specific behaviors in smaller animals that could not be detected before, allowing for the simultaneous recording of multiple individuals (Ramot et al. 2008; Swierczek et al. 2011a).

All in all, automated systems for the tracking of behavior provide the neuroscience community with very powerful systems that can record behavior during long periods of time, under semi-naturalistic conditions and from multiple animals. In addition, they allow obtaining multidimensional readouts, reducing experimental variability and increasing throughput. Nonetheless, the huge amount of data produced by high-throughput behavioral systems, the so-called big behavioral data (BBD) (Gomez-Marin et al. 2014), poses a challenge to how to achieve the highest amount of useful information. To this end, behavioral neuroscience needs to develop tools that enable the visualization and the analysis of those BBD. Furthermore, neuroscientists should apply approaches to modeling BBD that regardless of facing the huge complexity of the data still enable to obtain relevant biological results.

## 1.4. Analysis of big behavioral data

### 1.4.1. Visualization of big behavioral data

A good data analysis starts with a good visualization of the data. Quoting John Tukey “The simple graph has brought more information to the data analyst’s mind than any other device.” By failing to adequately visualize the data, fundamental aspects of the data structure might be missed. Furthermore, graphical representation is as important as numerical calculations for understanding the data, as Francis Ascombe nicely discussed using his well-known quartet (Anscombe 1973). Ascombe presented four datasets (Figure 4) that have identical numerical properties (mean, variance, correlation and regression line). Therefore, we can conclude that these datasets are pretty similar in terms of their basic statistical properties. However, their visualization revealed that, indeed, the four datasets were completely different, and that the fact that they share

numerical properties does not tell us anything about specific properties such as the presence of noise, outliers, or a non-linear relationship between the variables involved. This illustrates that visualization is a key aspect when analyzing data and not simply an additional step that we can ignore.

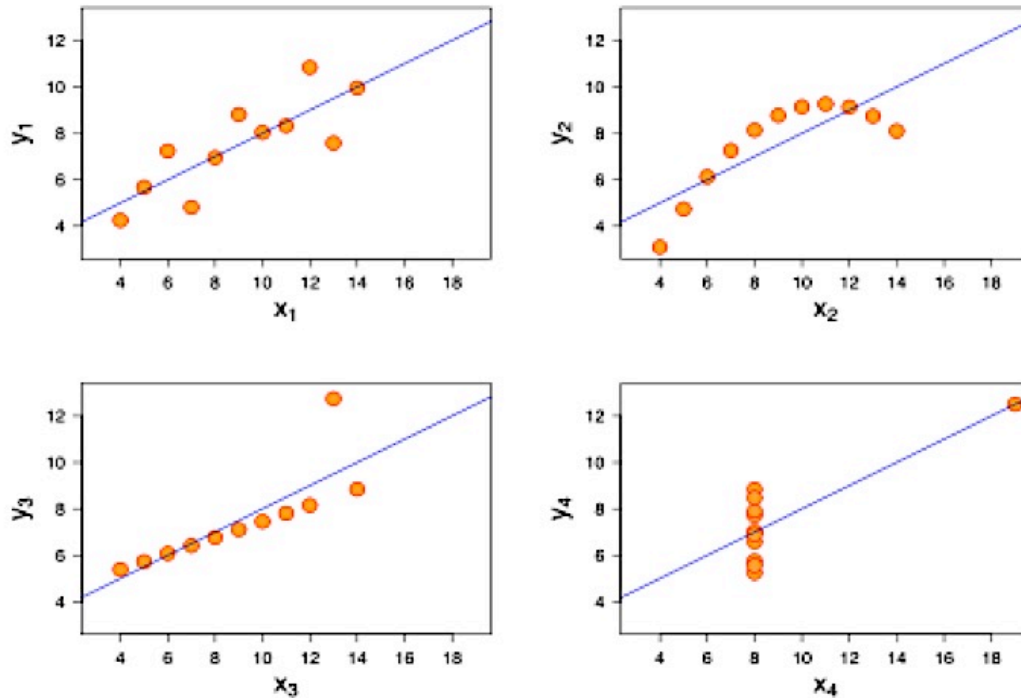


Figure 4 | Ascombe's quartet. These four datasets have exactly the same statistical properties (mean, variance, correlation and regression line), however their visualization clearly shows that these datasets in fact are very different. Ascombe wanted to illustrate that graphs constitute a prime tool for the analysis of the data contradicting the common assumption that “numerical calculations are exact but graphs are rough” (Creative Commons).

This is especially relevant when dealing with multidimensional behavioral data since simple measures of tendency or dispersion do not capture important aspects of these data (Pastizzo, Erbacher, and Feldman 2002). Hence, comprehensive data visualizations become indispensable for their exploration since otherwise it is impossible to understand their complexity. In many cases, we are especially interested in how variables change along the temporal dimension. Temporal data are structured as a unidirectional ordered sequence of events and usually manifest repetitive patterns (Streit and Gehlenborg 2015). As we have seen, longitudinal behavioral data perfectly fit this definition, showing ordered sequences that can be represented in ethograms (Branson et al. 2009) as we already discussed in section 1.3.2 (Figure 2). However, longitudinal behavioral data also exhibit repetitive patterns. Possibly the best known example of a repetitive pattern in biology are circadian rhythms. Circadian rhythms consist in biological processes that exhibit a stereotyped oscillation of about 24 hours, among which we find behavior (Vitaterna, King, and Chang 1994). We can find other repetitive behavioral patterns unfolding in shorter or longer (ultradian) periodicities (from minutes

to ultradian), some as obvious as locomotion patterns (Delcomyn 1980). Uncovering such repetitive structures from the data highly relies on powerful tools for the visualization of the data. One such representation is the raster-plot or actogram, which vertically stacks several periods of the same duration and thus provides a powerful representation to observe circadian or ultradian periodicities of the data (Strogatz 2013), as we will discuss further on section 1.3.2. Sometimes, however, we are not so interested in the temporal aspect of the data and therefore, we may want to identify globally representative data patterns. A collection of techniques that allow to combining the original variables to obtain linear combinations of them that account for the greatest variance in the data has become very popular. Most of these techniques rely on singular value decomposition approaches in order to factorize a data matrix (with observations encoded in the rows and variables in the columns) and obtain relationships of the original variables (in form of composite measures) that are more explicative of the data (Greenacre 2010). We can then represent the direction of the original variables in this new reduced space of composite measures and see for example if some of them are correlated. In a similar way, by plotting the observations in this new space we can obtain clusters informative of observations that share commonalities in terms of the new composite variables. Other possibilities that have been proposed is to organize the dimensions according to their similarity by rearranging data dimensions that display similar behaviors side by side, namely similarity clustering. These approaches use heuristic algorithms such as k-means to perform the clustering and give very interesting hints about group patterns in the data. They have been used, for example, to characterize population-wise patterns of behavior that constitute lifestyle-risk factors (Hofstetter et al. 2014).

#### 1.4.2. Analysis and visualization of longitudinal big behavioral data

A key aspect when dealing with longitudinal behavioral data is how to aggregate data in meaningful discrete behavioral events along time. These behavioral events, as seen in section 1.1.4, expand on multiple time scales. For instance, to understand the structure of feeding and drinking we need to annotate feeding bouts (meals) and drinking bouts (Zorrilla et al. 2005). However, these feeding bouts at the same time can be composed of several smaller discrete events or feeding acts (Goulding et al. 2008; Heyne et al. 2009). This way, indeed a meal is formed by several of these smaller subunits or feeding acts that are separated by small pauses without feeding. Hence, failing to understand this microstructure of the feeding bouts can lead to miss-annotate a feeding act belonging to a single meal as a complete meal. Eating and drinking bouts are

discrete behaviors that unfold in relatively short periods of time. Similar to these bouts it is also possible to annotate activity bouts, grooming bouts or social engagement episodes, among others. Once these short duration behaviors are annotated, it is possible to analyze them at a lower temporal resolution, to detect whether short behavioral events form coordinated sequences during given periods. For instance, combined analysis of feeding, drinking and locomotor activity has yielded very informative clues about how these behaviors distribute along the circadian cycle in different mouse models of obesity or overeating (Goulding et al. 2008; Edelsbrunner, Herzog, and Holzer 2009). Remarkably, Goulding et al. demonstrate how detailed data analysis allowed to identify behavioral organization on different time scales (for instance global active and inactive behavioral states unfold in longer times than feeding and drinking or locomotor behaviors). Besides, the aggregation of behavioral events on different time scales can yield very interesting hints about how this behavioral organization is modified in genetic models. In another interesting work, Steele et al. show how the annotation of behavioral bouts in longitudinal behavioral readouts can help to detect the progression of several altered behavioral phenotypes in two models of brain degenerative disorders such as Huntington's and prion diseases.

Behavioral annotations have then to be contextualized and integrated with behavior-related variables (environmental, experimental stages, etc.) and treated with analytical measures such as the evolution of the median values across time, or other data-derived measures on specific time intervals of interest. As discussed above, visualizations play a main role in this process. To illustrate this point, we can take an adaptation of the classical raster-plotted actograms used by Jud et al. (Jud et al. 2005) to visually inspect possible disruptions of the circadian locomotor activity. Figure 5 represents daily activity of a group of lynx living in a given region of the world as horizontally stacked lines (Heurich et al. 2014). Each pixel corresponds to an interval of 5 minutes with averaged activity of the group ranging from high (red) to low (blue). At the same time, the graph depicts the 24 hours of a day on the x-axis and the days of the year on the y-axis. On top, we find lines representing daily season-dependent periods of light. By integrating the behavioral and the environmental information in the same graph, this visualization allowed Huerich et al. to detect how lynx circadian rhythms are season-dependent when they compare two regions of Europe that differ in the number of hours with light during the year.

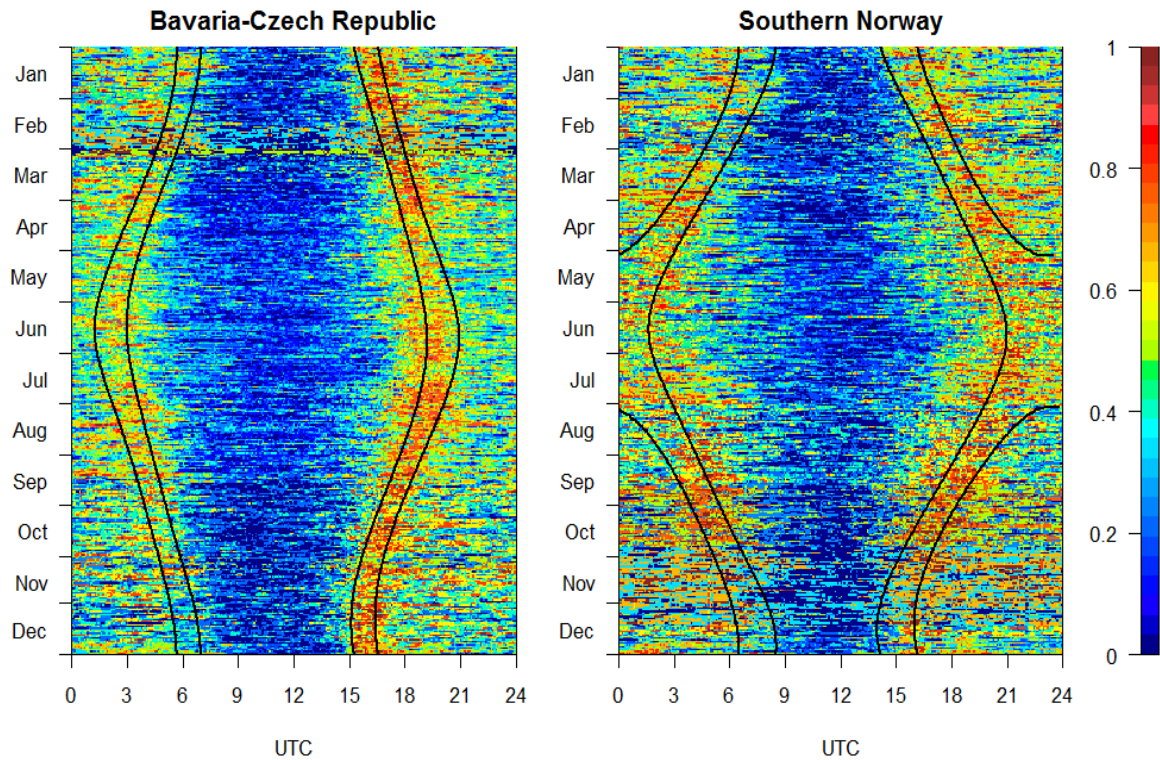


Figure 5 | Data visualizations of behavioral data can give very interesting insights into time-dependent or environment-dependent patterns. Activity was measured as the average activity of a group of lynx along time windows of five minutes. Each pixel corresponds to one of these windows and is colored blue when there was a lack of activity and red when activity was high. The x-axes represent the hours of the day while the y-axis corresponds to months of the year. Black lines on top of the plot represent the beginning of nautical twilight, sunrise, sunset, and end of nautical twilight. The left actogram shows the activity of a Central European lynx population and the one on the right of a Scandinavian population. This updated version of the classical actogram is a powerful way of visualizing the data to observe how behavioral patterns of activity are intrinsically related with environmental cues. We observe how the activity pattern of these two lynx populations differ between them depending on the hours of light in their respective regions (adapted from Heurich et al. 2014).

### 1.4.3. Analysis and visualization of multidimensional big behavioral data

Most of the biologically relevant behaviors cannot be reduced to a single behavioral variable. This is due to the huge spatio-temporal and behavioral dimensionality of the recordings derived from behavioral paradigms and their high variability. However, classical single-variate measures select specific variables in the hope that those may suffice to explain the behavioral domain we want to understand. Many authors have already detected this pitfall and have tried to apply suitable methods to avoid it. Multidimensional analysis techniques have been a good approach to address these concerns since they enable to reduce the number of dimensions by creating composite measures, principal components, of the original variables that capture the biggest amount of the experimental variance. The assumption behind this approach is that variables that vary greatly are important for the problem under consideration while those remaining almost unchanged can be considered noise. Usually, the main direction

of the variance will not be captured by a single of the original variables, but by a combination of them.

On the other hand, one or a few linear combinations of the original variables are often enough to explain a significant portion of the variance and thus, the new composite variables could be accurate descriptors of the system we are studying. In this manner, we can obtain a reduced set of quantitative descriptors powered enough to explain a complex behavior (Kumar et al. 2011), which is known as dimensionality reduction. Among these multidimensional techniques, Principal Component Analysis (PCA) is the most widely used (Hotelling 1933; see also Abdi and Williams 2010). PCA and derived approaches have been already used for analyzing several complex behaviors. For instance, PCA has been applied to the forward genetic screenings of thousands of mice by analyzing the data from an open field paradigm and a pharmacological intervention (Kumar et al. 2011). In this study Kumar et al. nicely showed how they can identify three independent principal components that account for different behavioral domains and how combinations of these components characterize the families of mutants that show subtle behavioral differences when compared to controls. In another study, authors applied PCA to the analysis of around 3000 mice trajectories in the MWM (Wolfer and Lipp 2000). These authors conclude that a large part of the variance from the MWM was explained by behavioral domains unrelated to spatial learning, and thus warned the community against univariate approaches to interpret the behavioral results of such paradigms. Another technique that relies on linear combinations of variables to analyze data is the Linear Discriminant Analysis (LDA) (Fisher 1936), see also (Bishop 2007). The main difference with PCA is that LDA is a supervised technique. Its approach consists essentially in finding those linear combinations of the original variables that best explain the separation of a number of predefined groups. LDA has been used for the recognition of swimming strategies on the MWM (Graziano, Petrosini, and Bartoletti 2003). LDA uses a specific distance measure that takes within-group variances into account, which further improves group separation. However, a discriminant analysis can be performed by simply determining the principal axes of group-wise aggregated variables (Greenacre 2010). This is the approach we have taken in the two last papers presented in the Results section (3.3 and 3.4).

Learning in the MWM is the consequence of an accumulated experience by being challenged with the same problem at consecutive times. Therefore we should emphasise this temporal aspect for its analysis and visualization. In univariate analysis, progression of learning is assessed by learning curves, which are obtained when plotting the variable of interest along time (see Figure 8d). One of the main benefits of the multidimensional techniques is that they allow to obtain informative representations of complex datasets that otherwise are difficult to visualize (Greenacre 2010). However, these

methodologies treat time as just another variable present in the dataset and hence it is difficult to separate its contribution from the rest of variables. To maintain a representation of the temporal aspect within this kind of display, our approach was to connect observations obtained at different time points in form of a trajectory (see results 3.3 and 3.4).

## 1.5. Case studies

To validate our visualization and analysis framework and our multidimensional approaches, we selected two common problems in biology: the development of a behavioral change upon introducing a perturbation, as occurs in wild type animals when introducing a diet change, and the learning process, studied in a well-established test, namely the MWM.

### 1.5.1. Obesity as a case study of automated longitudinal behavioral data

One of the fundamental problems in Clinical Neuroscience and in Psychopharmacology is that alterations of the nervous system that lead to a myriad of complex disorders lead to progressive behavioral changes that in the symptomatic phase develop into characteristic behavioral manifestations. However, the unfolding of these behavioral biomarkers is not abrupt, and many disorders display subtle changes in otherwise normal behavioral patterns precluding their onset. The progression of all the preclinical and clinical symptoms of such disorders, including patterns of behavior, is what is commonly known as the natural history of a disease. Examples of these disorders include neuropsychological disorders such as schizophrenia (Davidson et al. 1999a), neurodevelopmental disorders such as autism (Zwaigenbaum et al. 2005), inherited brain disorders such as Huntington's disease (Nguyen et al. 2006), some types of eating disorders (Polivy and Herman 2002), addiction to substance of abuse (Koob, Sanna, and Bloom 1998) and neurodegenerative diseases such as Alzheimer's disease (Dubois et al. 2016), among others. Two well-known examples are Alzheimer's disease (Sperling, Aisen, Beckett, Bennett, Suzanne, et al. 2011) and schizophrenia (Davidson et al. 1999a), both showing a long preclinical phase within which behavioral biomarkers that precede the actual onset of the disease are already detectable (as depicted in Figure 6). It is noteworthy that an early detection of prodromal behavioral patterns offers an opportunity for an early intervention that can improve disease prognosis (Dawson 2008; Petersen 2010; Sommer et al. 2016).

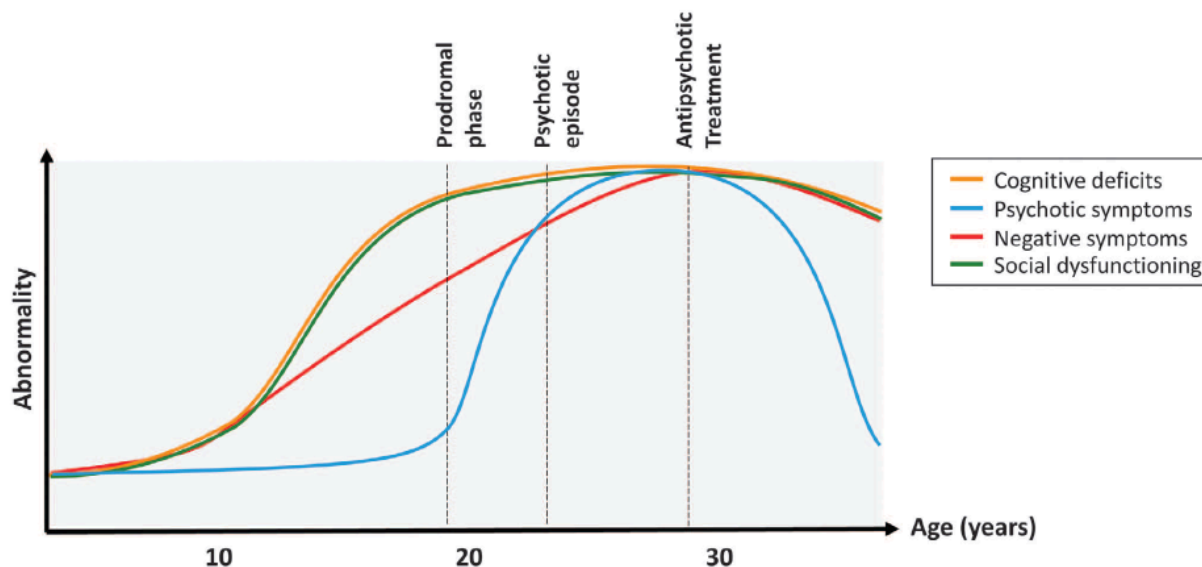


Figure 6 | Natural history of schizophrenia, a neural disorder with a long prodromal phase within which several abnormal behavioral patterns can be detected in a subtle manner. These behavioral patterns worsen with time until they become diagnosable with current medical procedures (adapted from Sommer et al. 2016).

Even being aware of its limitations, the common strategy followed by behavioral neuroscience to study chronic disorders with a long progression along time has been to study constrained time windows that are especially relevant for the biological question under study. Under this framework, the easiest approach consists in carrying out the experiments that detect differential patterns of behavior once the physiological mechanisms of the disorder are established. Another way to tackle the question has been to perform repeated assays along time. However, both strategies suffer from an obvious pitfall: these readouts are just snapshots of a longer behavioral trajectory. Therefore, it can be the case that we are not capturing the precise time when relevant behaviors are taking place. In addition, more subtle behaviors may unfold at time points not considered by the experimental design, and thus results can be dramatically affected by studying the wrong time window (Fonio, Benjamini, and Golani 2012). New systems to track spontaneous behaviors during long periods of time (see Appendix A) provide a more natural way of studying disorders that unfold during these long time periods.

In this thesis we have used the development of diet-induced obesity to validate our longitudinal behavioral analysis framework, Pergola. Obesity is an important health problem with a strong environmental component. Our current environments with high availability of caloric dense foods promote overeating potentially leading to obesity (Hill 1998). In agreement with these findings, it has been shown that the availability and palatability of food potentially leads to eating disorders (Heyne et al. 2009), which are characterized by a lack of control that is common with other addictive-like behaviors



like drug addiction (Kenny 2011; Latagliata et al. 2010). Abnormal eating behaviors such as binge eating (Curtis and Davis 2014; Colles, Dixon, and O'Brien 2007) or snacking are among the commonly observed behavioral alterations in obese individuals. Thus, this is an ideal model for the characterization of the time course of the appearance of abnormal eating patterns<sup>3</sup>.

### 1.5.2. The MWM as a model for the study of a multidimensional behavioral cognitive process

A behavioral process is usually composed of a range of variables spanning many behavioral domains. Therefore, the approach we take to select the relevant variables for the behavior under study is of great importance. The MWM is an example of a widely used assay trying to understand a brain function as complex as spatial learning or memory. In brief, as shown in Figure 7, the water maze consists of a circular pool with a hidden platform that a rodent has to find by swimming, using spatial cues that are around the pool in order to escape from the water (again this assay exploits a natural behavior). The MWM is run in several consecutive sessions. In the first one, the platform is visible and therefore the animal can learn to escape from water using the cues (procedural learning). In successive sessions (visuo-spatial learning sessions), the platform remains hidden, and in this way we can assess the learning process by measuring how efficiently the mice swim to the platform using the available extra-maze cues. In additional sessions it is possible to assess reference memory by removing the platform from the pool or to reveal how flexible are animals in terms of learning by placing the platform in a new position (Vorhees and Williams 2006). Using this paradigm, we can measure multiple variables ranging from variables that are identical or similar to the recordings themselves, to indexes that have been created as a more conceptual way to characterize the performance of the animal in terms of the learning strategy. Both types of variables have been used classically to assess the phenotypes or the effects of a treatment on animal models of disorders with affected cognitive capabilities. For instance, among the variables closer to recorded measures, we find the latency to escape from the water. When an individual learns to orient itself inside the pool, it will find the target more quickly and thus, it will reduce the time to achieve the platform along the sessions by presenting a better swim strategy (Figure 7). An example of a more conceptual variable is the Wishaw index, which measures the percentage of time that an animal spent in an ideal corridor connecting the platform and the place of release into the pool (Whishaw and Jarrard 1996). In this case good learners will reduce the value of the Wishaw index along sessions, since their swim strategy will be close to

---

<sup>3</sup> Details of the system we used for the recording of animal behavior are provided in Appendix A, see PheCOMP page.

optimal and they will tend to be in this ideal corridor. However, these two variables are just proxies to quantify relevant learning-related behaviors. In this thesis we have tried to grasp the behavioral domain of cognition in a more global manner using all the commonly measured variables at the same time. To achieve this goal we used a multidimensional approach since it enables to reduce the dimensionality of the MWM readout combining the information of several of these variables.

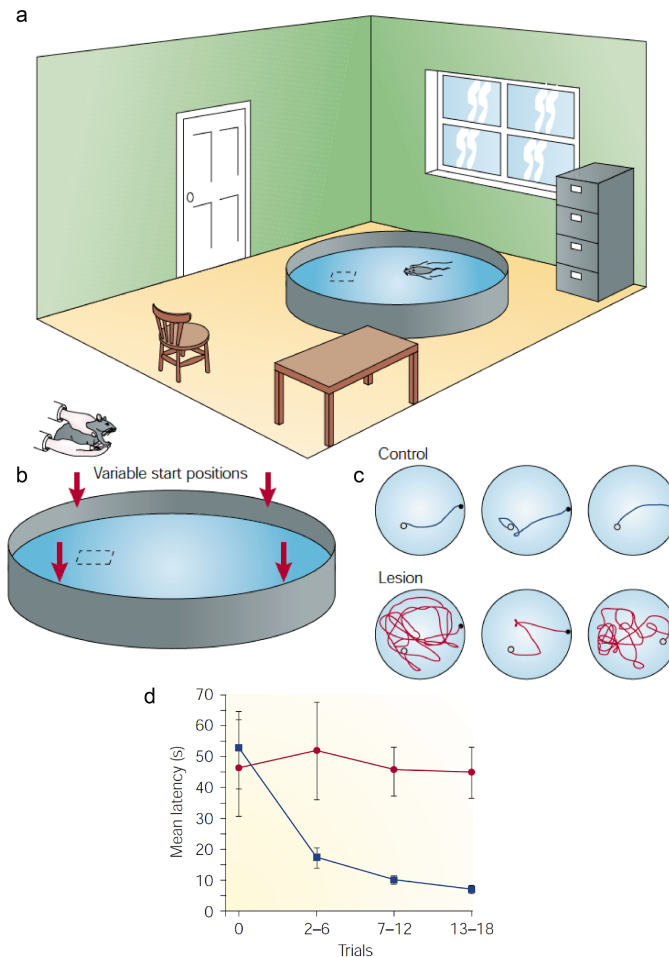


Figure 7 | Some aspects of the widely-used behavioral paradigm of the MWM. **(a)** Typical configuration of the assay consisting of a pool with a hidden platform, not visible by the animal, surrounded by visual cues. **(b)** The rodent is introduced in consecutive trials from four starting positions. **(c)** Animals exhibit different strategies to reach the platform depicted by their swim path (trajectory). Blue trajectories on the top correspond to a control animal and red trajectories on the bottom to individuals with a lesion (hippocampal damage) that impairs its spatial learning capability. **(d)** Example of learning curves. The common way to assess the performance of the animals in the MWM is to measure the latency to reach the platform of the animal. As shown in the plot of the groups of animals, in this case WT (blue) has spatial learning intact, so they will keep improving along the sessions and reducing the latency to find the platform. However, impaired animals (red) will not decrease their latency to reach the platform (adapted from Eichenbaum 2000)

## 2. OBJECTIVES

The big data problem in life sciences is the consequence of a paradigm shift. High-throughput technologies, which classically included genomics, transcriptomics, proteomics or metabolomics, now also involve phenomics, enabling the complete mapping of behavioral components in any organism. This involves enormous quantities of data whose visualization and analysis can only be addressed creating new frameworks. My thesis has aimed to provide a useful tool for the visualization and analysis of big behavioral data, to overcome the limitations of handling big behavioral data, which require processing of different data sets and formats. I also explored multidimensional analysis techniques to help reducing the huge spatio-temporal dimensionality derived from behavioral recordings. The specific objectives have been:

1- The design of a computational combining a new generation of tools allowing the effective manipulation and representation of long high-throughput behavioral data (BBD). This framework should standardize file formats and provide access to bioinformatics tools making it possible to programmatically filter, analyze, display and compare the readouts from any of the apparatus currently available for automated behavioral capture.

2- Proof of concept of our computational BBD tool for the study of the dynamics of appearance of disease behavioral signatures using longitudinal behavioral data.

3- Implementation of a PCA-based methodological framework for the study of multidimensional behaviors.

4- Proof of concept of the PCA-based methodological framework for the study of learning.



## **3. RESULTS**

### **3.1. Pergola: a new paradigm for longitudinal data visualization and analysis**

#### **Preface**

One milestone in this thesis was to develop a framework for the exploration and analysis of automated longitudinal recordings of behavior and associated metadata. In this publication we present Pergola (Python bEHavioRal GenOmetools LibrAry), a python library that enables the visualization, manipulation, modeling and integration of longitudinal behavioral data. Pergola offers a set of own tools to analyze and explore the data and capitalizes on a myriad of efficient, widely used genomic tools. In this manner we provide the community with a bioinformatics framework that unifies and boosts the analysis of big behavioral data contextualized with all its rich metadata including environmental information, statistical modeling and physiological information. The integrative analysis of high-throughput behavioral recordings along with the relevant information that shape behavioral responses could enable the understanding of how brain function orchestrates complex behavior. Therefore, our free tool represents a relevant resource for the neuroscience community.

Espinosa-Carrasco J., Erb I., Dierssen M., Notredame C. Pergola: a new paradigm for longitudinal data visualization and analysis. In preparation.



# **Pergola: a new paradigm for longitudinal data visualization and analysis**

Jose Espinosa-Carrasco<sup>a,b,c</sup>, Ionas Erb<sup>a,c</sup>, Mara Dierssen<sup>b,c</sup>, Cedric Notredame<sup>a,c</sup>

<sup>a</sup>Comparative Bioinformatics, Bioinformatics and Genomics Program, Barcelona Institute of Science and Technology, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain

<sup>b</sup>Cellular & Systems Neurobiology, Systems Biology Program, The Barcelona Institute of Science and Technology, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain

<sup>c</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain

Correspondence should be addressed to: C Notredame: [cedric.notredame@crg.eu](mailto:cedric.notredame@crg.eu)

## **Abstract**

Exploratory data analysis is key to understanding longitudinal behavioral data. Pergola (<http://cbcr.github.io/ pergola>) processes longitudinal recordings by representing them in data structures originally developed for genome analysis. The equivalence of data formats allows for the visualization, analysis, and advanced modeling of longitudinal data using state-of-the-art genome analysis tools. We show how Pergola simplifies and accelerates explorative analysis of behavioral data analysing recordings of two major model organisms (*C. Elegans* and *Drosophila*) and can potentially be applied to any analysis involving longitudinal recordings.

Fine-grained longitudinal data is one of the fastest growing corpus of biological data (Rose 2016). Behaviors are a prime target for longitudinal studies, as they constitute one of the most complex available phenotypes, linking genetics, development, neurobiology, evolution and environmental interaction such as drugs and stress responses. Furthermore, the dynamic nature of behavior makes its study particularly reliant on temporal structure exploration. Behavioral studies have long been hampered by data acquisition limitations, an issue recently alleviated by the emergence of high-throughput behavior monitoring methods (Hong et al. 2015; Ohyama et al. 2015). Of course, as in other fields, the data burst induced by such a technological shift requires a new generation of tools for handling, analysing and visualizing data. New data also poses new challenges, the most pressing one being reproducibility. In that respect, a major difficulty stems from the lack of uniform standards that would allow the seamless comparison of data gathered on different platforms. Current systems look very much

like the so-called data silos, a commonly used concept in biological databases referring to fragmented non-interoperable data. Data silos are very common in rapidly evolving fields. The problem is usually addressed through a lengthy community effort on normalization and standardization. We demonstrate here that solutions are available to dramatically accelerate this process in the field of behavioral analysis. Indeed, we show the many solutions and standards developed for over a decade in genomic analysis perfectly fit the current needs of behavioral analysis. We have illustrated this demonstration by reproducing with standard genomic tools, three distinct high-impact behavioral analyses, originally carried out on three different platforms.

High-throughput screenings of live behaviors are very diverse. They may be broadly separated in two classes: systems that physically monitor physiological intakes such as feeding, general activity or oxygen consumption in mice (PHECOMP, PHENOMASTER, CLAMS) and systems that, using videos of behaving animals and computer-vision techniques, annotate complex behavioral trajectories. These are available for most animal models, including *C. elegans* (Ramot et al. 2008; Yemini et al. 2013; Swierczek et al. 2011), mice (de Chaumont et al. 2012; Hong et al. 2015), *D. melanogaster* adults (Branson et al. 2009), larvae (Ohyama et al. 2015), zebrafish (Rihel et al. 2010) or multipurpose (Gomez-Marin et al. 2012). These techniques all deliver longitudinal data, along with more or less elaborate annotation. While most of these platforms provide good ad-hoc analysis capacities and visualization tools, they all stop short of allowing the seamless comparison of data gathered across different platforms. The range of data handling tools effectively implemented by the authors also limits the range of data handling tools. We argue here that barriers between these platforms must be broken and we have addressed this issue by developing Pergola (Python bEHavioRal GenOmetools LibrAry).

Pergola is open-source software designed for the visualization, manipulation, modeling and integration of (high-throughput) longitudinal behavioral data including its associated metadata. It does so using a repurposed combination of state of the art genome analysis software (Quinlan and Hall 2010; Dale et al. 2011; Robinson and Thorvaldsdóttir 2011; Pohl and Beato 2014)(Quinlan and Hall 2010; Dale, Pedersen, and Quinlan 2011; Robinson and Thorvaldsdóttir 2011). Longitudinal data recordings share a strong conceptual similarity with genomic sequence recordings. While they differ in their observation units - time vs. nucleotide positions - both data types are nonetheless sequential by definition, and are usually associated with highly heterogeneous metadata - behaviors vs gene annotation. These data types also require analogous processing such as quality based filtering, binning, averaging discretization or differential annotation with either quantitative or categorical variables. In both systems, final modelling often involves combining direct observations with higher level



measurements or contextual information. The natural correspondence between longitudinal and genomic data would be merely anecdotal if genomics had not developed powerful standard formats for data representations (Kent et al. 2002), sophisticated visualization tools (Robinson and Thorvaldsdóttir 2011; W. James Kent et al. 2002) and efficient data analysis tools (Quinlan and Hall 2010), to cite only a few. The genomic framework has been in production for over a decade and is regularly updated to scale up with the incoming data. As such, its repurposing towards behavioral data constitutes a major opportunity.

Pergola makes data interoperable by transforming any temporal records into standard genomics file formats. This process relies on a customizable mapping between the behavioral and the genomic ontologies. This reference mapping, invisible to the user, associates terms referring to analogous features of the records, such as *StartTime* in the behavioral file with *start* in the genome-like representation (see Online methods and Supplementary Table 1, 2 and 3). Besides supporting comma-separated files (CSV) as input, Pergola is also ISA-tab compliant (Rocca-Serra et al. 2010; Sansone et al. 2012) thus insuring compatibility of data acquired under this novel data-sharing standard to annotate experimental meta-information.

Genomic formats provide a perfect scaffold for storing longitudinal behavioral data (Online Methods). Formats such as BED (the Browser Extensible Data format) (Kent et al. 2002) or GFF (the General Feature Format)(<http://gmod.org/wiki/GFF3>) make it possible to record sequence features along with any kind of quantified annotation. Pergola adapts any longitudinal data consisting of time intervals, such as behavioral events, to these files. Similarly, continuous-valued data is perfectly suited for the bedGraph format or BigWig (W. J. Kent et al. 2010), two formats originally designed for storing quantitative information associated with sequence features. These data formats can therefore store any continuous behavioral measure derived from animal locomotion or statistical scores of a model applied to data.

Once formatted, data is also ready to be processed by powerful analysis algorithms, like BEDtools (Quinlan and Hall 2010), Pybedtools (Dale, Pedersen, and Quinlan 2011) or bwtool (Pohl and Beato 2014). These tools allow a wide operational diversity including joining, subtracting, intersecting across several channels containing longitudinal information or the aggregation, smoothing or summarizing within the same channel. It must be stressed that the repurposing of these tools does not require any modification: it is a mere consequence of the reformatting. Likewise, reformatted data makes it possible to use any genomic browser (Robinson and Thorvaldsdóttir 2011; W. James Kent et al. 2002) and repurpose it into a behavioral data visualizer thus taking advantage of sophisticated features allowing the simultaneous display of data and metadata as well as

multi-scale navigation. This combination between BED, GFF, BEDtools and genome browsers defines a mature and versatile framework for the uptake of any data for which sequences constitute the underlying structure. This makes it an ideal fit for so-called big behavioral data (Gomez-Marin et al. 2014).

As a proof of principle we have used Pergola to reprocess data gathered on three distinct platforms using worms, drosophila and mice models. The first source of data (Yemini et al. 2013) is a set of worm trajectories recorded using Worm Tracker 2.0, a video tracking system allowing trajectory typing. We extracted locomotion-related behaviors from the source database, mapped the associated control vocabulary on pergola ontology and visualized the resulting data with an adapted version of the Integrative Genomics Viewer (IGV) (Robinson and Thorvaldsdóttir 2011). For a single control individual, converted tracks display the direction of the worm at each point of the trajectory (Fig. 1a) aligned with several locomotion phenotypes measured frame by frame (Fig. 1b). This display clearly shows the correlation between the speed of the different worm body parts, as reported in the original study. In that same original analysis, the authors carried out mid-body speed comparison. We replicated these results by using bwtool (Pohl and Beato 2014)) to smooth data across a specific number of frames. Pybedtools (Dale, Pedersen, and Quinlan 2011) (Fig. 1d) was then used to intersect speed tracks with the direction of the worm trajectory (Fig 1e) thus recapitulating the original report on the distribution of mid body speed by the motion state for a mutant strain (*unc-16*) and its difference from the controls (Fig. 1f). The second source of data (Kabra et al. 2013) is a set of drosophila trajectories gathered using the JAABA software suite, another image tracking software. It is focused on the analysis of chasing behaviors in flies. Following a similar procedure as shown for worm, the original data descriptors were mapped onto pergola ontology, thus allowing a smooth translation into the BED format that was used to display the original data with the IGV browser. The result is a rendering of the original data of chase social behaviors along with their confidence scores in which one can simultaneously browse several individual trajectories. The third application involved mice data obtained using PHECOMP cages. These cages provide fine-grained longitudinal monitoring of mice intake behaviors (solid and liquid) and can follow a single individual for several weeks. Using the pergola ontology remapping, we were able to re-map the original data into BED format, and could use this data to reproduce the main findings of the original analysis (see Supplementary Section).

These three examples illustrate how very heterogeneous datasets could be rapidly re-processed so as to become suitable for high quality standard tools. Doing so has several major advantages. The most obvious is reproducibility and stability. By using popular tools, one is much less likely to be affected by the effect of poor software development.

The analyses become easier to replicate and compare across labs and groups. Furthermore, the high usage level of these tools ensures higher quality of documentation, stricter quality monitoring by the community, not to mention the obvious advantage of using the same programmatic tools to process diverse types of data in the lab. Yet, from a broader perspective, Pergola does much more than merely providing access to new tools. In the high throughput era science is going through, definition of novel standards is a common process. It usually involves complex negotiations between stakeholders. These are followed by lengthy periods of tool developments whose reliability often depends on usage levels. We show here that behavioral analysis may skip all this and immediately move into production mode thanks to the strong analogy between behavioral and genomics data. Pergola dramatically broadens the scope and effectiveness of longitudinal analysis. It does so with very limited overhead. Genomics is an extremely active field of investigation and the capacity to tap into its rapidly expanding spectrum of statistical analysis may rapidly generate key synergies. Genome analysis software are very modular, and the bridge we are opening will doubtlessly result in new modules for data visualization and analysis to be made available at a minimum cost for the community. The stakes are high, especially at a time when novel types of longitudinal data are being recorded at a massive scale in hospitals (Jensen et al. 2014), or even societal contexts (de Montjoye et al. 2013; de Montjoye et al. 2015).

## Figures

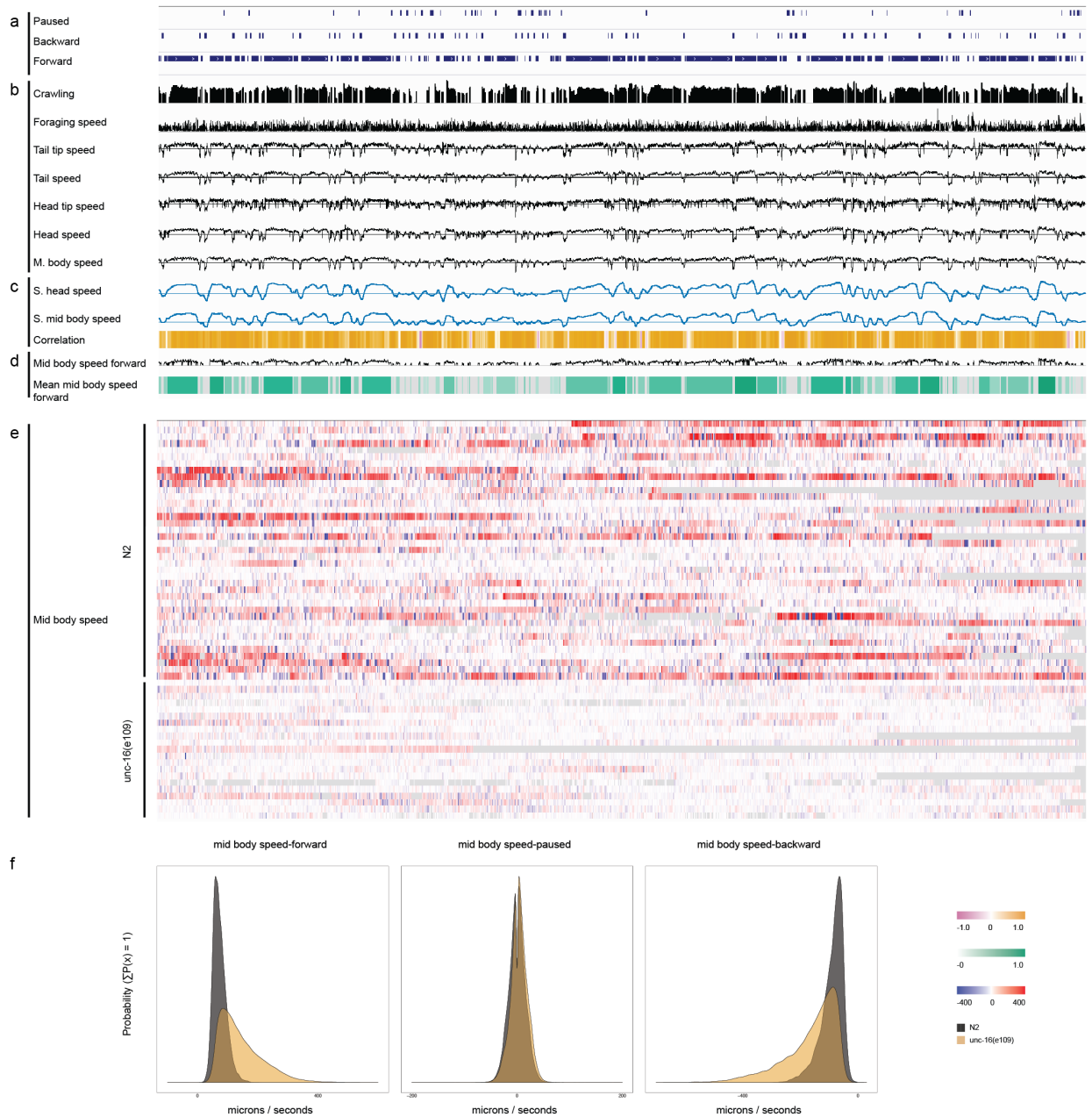
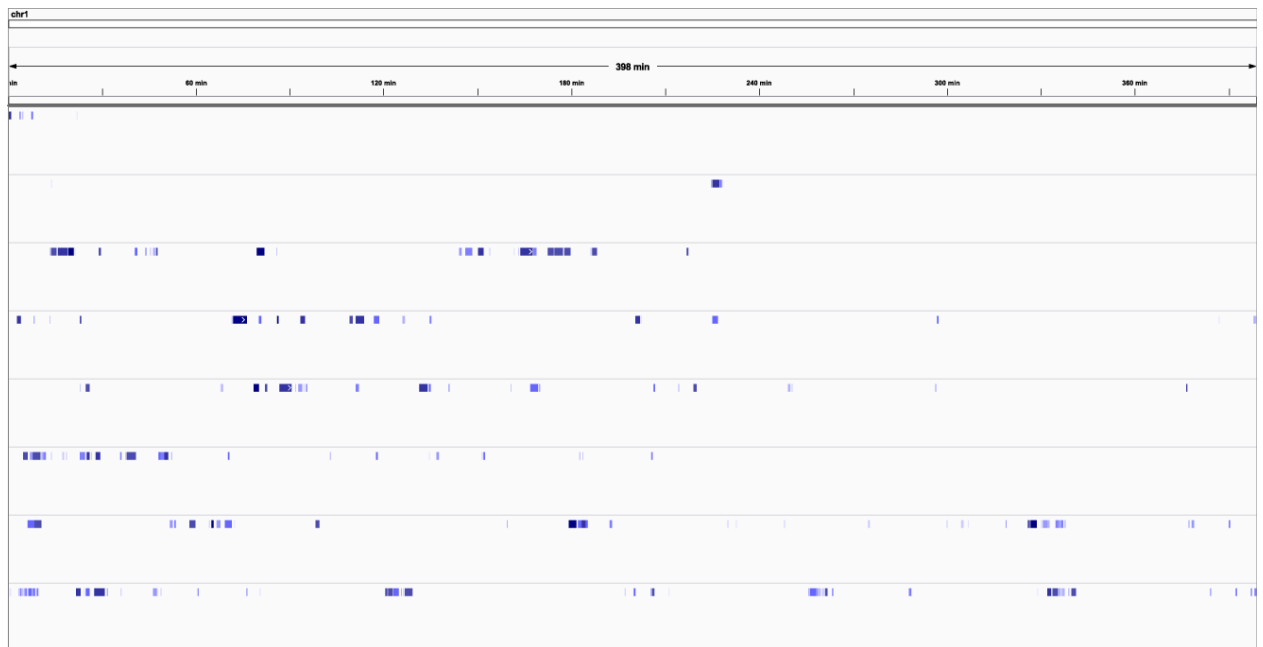


Figure 1. *C. elegans* motor behavior visualisation and analysis using Pergola and genomic software. (a) Each blue track displays the intervals corresponding to each motion state (paused, backward or forward) of a single N2 individual. (b) Several locomotion measures of the same individual namely: crawling amplitude (0, 50), foraging speed (0, 730), tail tip, tail, head, and mid body speed (-700, 700) and finally head tip speed (-1050, 810), numbers in parenthesis indicate the displayed variable range. All speed measures are expressed in microns/second except crawling amplitude

expressed in degrees/second and foraging speed in degrees. (c) Head and midbody speed signal (blue) smoothed (S.) using a sliding window (Online Methods). Correlation track represent the intensity of the correlation between smoothed head and mid body speed within each of the windows, color depicts high correlation (orange) or high anticorrelation (reddish purple). (d) Mid body speed measures intersected with forward motion periods. Below, each track segment corresponds to a forward motion period and the intensity of the color to the mean value of mid body speed during the period. (e) Individual tracks represent mid body speed of an individual during the tracking period of approximately 15 minutes. Red values indicate positive speeds and blue values negative speeds. Tracks of N2 control group are shown on the top (n=40) and unc-16 individuals on the bottom (n=20). (f) Distribution of speed intersected with forward, backward and paused motion state for N2 control strain (black) and unc-16 strain (yellow).



**Figure 2. *Drosophila* chasing behavior displayed with the adapted IGV Genome Browser.** Each track in the figure corresponds to the behavioral trajectory of a single fly. Each bar in the trajectory depicts a chasing bout and the color intensity reflects the score of the annotation, i.e. the confidence JAABA classifier provides for its annotation. We use the sample data of fly trajectories provided by JAABA to train a chase classifier and using it we annotate the bouts of chasing behavior.

## References

Branson, Kristin, Alice A. Robie, John Bender, Pietro Perona, and Michael H. Dickinson. 2009. "High-Throughput Ethomics in Large Groups of *Drosophila*." *Nature Methods* 6 (6). Nature Publishing Group: 451–57.

- Dale, Ryan K., Brent S. Pedersen, and Aaron R. Quinlan. 2011. "Pybedtools: A Flexible Python Library for Manipulating Genomic Datasets and Annotations." *Bioinformatics* 27 (24): 3423–24.
- de Chaumont, Fabrice, Renata Dos-Santos Coura, Pierre Serreau, Arnaud Cressant, Jonathan Chabout, Sylvie Granon, and Jean-Christophe Olivo-Marin. 2012. "Computerized Video Analysis of Social Interactions in Mice." *Nature Methods* 9 (4). Nature Publishing Group: 410–17.
- de Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. "Unique in the Crowd: The Privacy Bounds of Human Mobility." *Scientific Reports* 3: 1376.
- de Montjoye, Yves-Alexandre, Laura Radaelli, Vivek Kumar Singh, and Alex Sandy Pentland. 2015. "Identity and Privacy. Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata." *Science* 347 (6221): 536–39.
- "GFF3 - GMOD." 2015. Accessed November 20. <http://gmod.org/wiki/GFF3>.
- Gomez-Marin, Alex, Nicolas Partoune, Greg J. Stephens, Matthieu Louis, and Björn Brembs. 2012. "Automated Tracking of Animal Posture and Movement during Exploration and Sensory Orientation Behaviors." *PloS One* 7 (8): e41642.
- Gomez-Marin, Alex, Joseph J. Paton, Adam R. Kampff, Rui M. Costa, and Zachary F. Mainen. 2014. "Big Behavioral Data: Psychology, Ethology and the Foundations of Neuroscience." *Nature Neuroscience* 17 (11): 1455–62.
- Hong, Weizhe, Ann Kennedy, Xavier P. Burgos-Artizzu, Moriel Zelikowsky, Santiago G. Navonne, Pietro Perona, and David J. Anderson. 2015. "Automated Measurement of Mouse Social Behaviors Using Depth Sensing, Video Tracking, and Machine Learning." *Proceedings of the National Academy of Sciences of the United States of America* 112 (38): E5351–60.
- Jensen, Anders Boeck, Pope L. Moseley, Tudor I. Oprea, Sabrina Gade Ellesøe, Robert Eriksson, Henriette Schmock, Peter Bjødstrup Jensen, Lars Juhl Jensen, and Søren Brunak. 2014. "Temporal Disease Trajectories Condensed from Population-Wide Registry Data Covering 6.2 Million Patients." *Nature Communications* 5 (May): 1–10.
- Kabra, Mayank, Alice A. Robie, Marta Rivera-Alba, Steven Branson, and Kristin Branson. 2013. "JAABA: Interactive Machine Learning for Automatic Annotation of Animal Behavior." *Nature Methods* 10 (1). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 64–67.
- Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. 2002. "The Human Genome Browser at UCSC," 996–1006.
- Kent, W. J., A. S. Zweig, G. Barber, A. S. Hinrichs, and D. Karolchik. 2010. "BigWig and BigBed: Enabling Browsing of Large Distributed Datasets." *Bioinformatics* 26 (17): 2204–7.
- Ohyama, Tomoko, Casey M. Schneider-Mizell, Richard D. Fetter, Javier Valdes Aleman, Romain Franconville, Marta Rivera-Alba, Brett D. Mensh, et al. 2015. "A Multilevel Multimodal Circuit Enhances Action Selection in *Drosophila*." *Nature* 520 (7549): 633–39.
- Pohl, Andy, and Miguel Beato. 2014. "Bwtool: A Tool for bigWig Files." *Bioinformatics* 30 (11): 1618–19.
- Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42.
- Ramot, Daniel, Brandon E. Johnson, Tommie L. Berry Jr, Lucinda Carnell, and Miriam B. Goodman. 2008. "The Parallel Worm Tracker: A Platform for Measuring

- Average Speed and Drug-Induced Paralysis in Nematodes.” *PloS One* 3 (5). Public Library of Science: e2208.
- Rihel, Jason, David A. Prober, Anthony Arvanites, Kelvin Lam, Steven Zimmerman, Sumin Jang, Stephen J. Haggarty, et al. 2010. “Zebrafish Behavioral Profiling Links Drugs to Biological Targets and Rest/Wake Regulation.” *Science* 327 (5963): 348–51.
- Robinson, J. T., and H. Thorvaldsdóttir. 2011. “Integrative Genomics Viewer.” *Nature* 29 (1): 24–26.
- Rocca-Serra, Philippe, Marco Brandizi, Eamonn Maguire, Nataliya Sklyar, Chris Taylor, Kimberly Begley, Dawn Field, et al. 2010. “ISA Software Suite: Supporting Standards-Compliant Experimental Annotation and Enabling Curation at the Community Level.” *Bioinformatics* 26 (18): 2354–56.
- Rose, Suzanne. 2016. “Huge Data-Sharing Project Launched.” *Cancer Discovery* 6 (1): 4–5.
- Sansone, Susanna-Assunta, Philippe Rocca-Serra, Dawn Field, Eamonn Maguire, Chris Taylor, Oliver Hofmann, Hong Fang, et al. 2012. “Toward Interoperable Bioscience Data.” *Nature Genetics* 44 (2). Nature Publishing Group: 121–26.
- Swierczek, Nicholas A., Andrew C. Giles, Catharine H. Rankin, and Rex A. Kerr. 2011. “High-Throughput Behavioral Analysis in *C. Elegans*.” *Nature Methods* 8 (7). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 592–98.
- Yemini, Eviatar, Tadas Jucikas, Laura J. Grundy, André E. X. Brown, and William R. Schafer. 2013. “A Database of *Caenorhabditis Elegans* Behavioral Phenotypes.” *Nature Methods* 10 (9): 877–79.

## Supplementary Section - Methods

**Implementation and availability.** The package is implemented in Python and released under GNU general public license version 3. Latest version, source code, documentation and issue tracking are available at <https://github.com/cbcrg/pergola>. All the scripts and configuration files used in the analysis are available from Github: (<https://github.com/cbcrg/pergola/blob/master/examples/>).

Pergola releases are provided as Docker containers in a public dockerhub account (<https://hub.docker.com/>). The version used for the analysis on this paper is accessible under `cbcrg/pergola:celegans`

**Pergola functioning overview.** Pergola uses genomic files formats widely used in genomics to encode sequential data. To achieve this task, our software needs: a file encoding the sequential data to be converted, a file to set the equivalence between the sequential and the genomic space (mapping file), and user selected options.

**Input data.** Pergola currently accepts three input data formats: (i) character-separated values (CSV) format, a common export format for tabular data available from most commercial and noncommercial tools. The minimum input is a two column file containing the timestamps and the variable measured (ii) ISA-Tab format, a wrapper

format for data generated in high-throughput studies that provides a general framework to annotate experimental metadata. (iii) JAABA output format when saved in scores.mat MATLAB format (MathWork).

**Mapping file.** The mapping file is a text file used to declare the equivalences between the Pergola internal controlled terms and those used in the input data (i.e. the column names). In practice this file contains two terms per line, on the left the user-defined term and on the right the Pergola equivalent, as selected among Pergola ontology (Supp Table 1). The mapping file follows the Gene Ontology Consortium community standards as defined in (“External Mapping File Format | Gene Ontology Consortium” 2015) from the Gene Ontology Consortium (Ashburner et al. 2000). An example of a generic mapping file is shown on Supplementary Table.2. The example below shows how timestamps in input data are to be mapped onto start interval position in the pergola ontology.

*behavioral\_file:event\_start > pergola:start*

Where behavioral\_file corresponds to the tag designating the input file, event\_start is the column name from the original file\*, pergola is the tag used for pergola ontology and start is the item in the pergola ontology.

*\*It is possible to use files with unnamed columns using pergola flag -nh and giving the list of field names as a Pergola parameter after -s flag.*

**BED and GFF output format for discrete-valued data.** One of the most common types of behavioral data consists in a sequence of a few behavioral states along a temporal trajectory. A simple example is the alternation between animal forward, backward and paused motions states (Fig. 1). Formats used to encode genomic annotations or describe sequence features provide a perfect scaffold to define such behavioral trajectories. Pergola supports two of these genomic formats: the Browser Extensible Data (BED)(Kent et al. 2002) or the General Feature Format (<http://gmod.org/wiki/GGF3>).

**BedGraph output format for continuous-valued data.** A sequence annotation may be continuous rather than discrete, in which case every time point will receive a value from some interval of the real numbers. Such representations are typical of modelling efforts in which given time intervals get associated with a probability rather than a discrete state. Likewise, these formats can be easily adapted to describe statistical parameters or scores resulting from a behavioral trajectory. To this end, Pergola supports the BedGraph format, a genomic standard for the storage of continuous-valued data.



**Reference sequence** The repurposing of genome analysis tools requires the presence of a sequence used as a reference for the mapping of the rest of the data. In practice, Pergola generates a dummy string stored in FASTA format whose length is set to the number of time units in the recorded experiment. Reference sequences come along with the description of the chromosomal cytogenic bands declared with the cytoband file format. Pergola adapts these higher level annotations to complement the records with extra signals such as the alternation between days and nights (Fig. X/supp figure Y) or any environmental or metadata mapped onto the records.

**C. elegans data set.** We obtained worm motor behaviors from the “C.elegans behavioural database” (Yemini et al. 2013) available in (<http://wormbehavior.mrc-lmb.cam.ac.uk/>). Specifically we used data from unc-16 (n=16) a mutation affecting locomotion and N2 control strain (n=40). Each individual recording was available in a MAT, HDF5-formatted file (Hierarchical Data Format Version 5). Files consisted in a time series of raw phenotypic features frame by frame besides all the experimental metadata (see Yemini et al for a detailed description). As a result of parsing MAT raw files by means of python scripts, we produced CSV Pergola compatible files containing the information of a single recording. All scripts are available in our public repository.

**C.elegans motor behavioral data analysis.** After extracting individual phenotypic features into CSV files, we followed a standard Pergola analysis. A mapping file w2p.txt (Supp Table 3) was defined and used to process the worm behavioral time series celegans\_behavioral\_features.csv into BedGraph files using the following command:

```
pergola_rules.py -i celegans_behavioral_features.csv -m w2p.txt -f bedGraph -w 1
```

Here, -f bed sets the output format to bedGraph and -w sets a time window within which data is aggregated (among other possible operations). In this case a window of 1 maintains the original input time series. In a similar way, worm motion states can be converted into BED files. The resulting tracks can then be directly loaded into IGV for visualization (Fig. 1a, b and e). Moreover, BEDtools enables the calculation of statistics and/or summary operations on the overlapping intervals of two bed files. The following command determines the mean value of a given behavioral feature for each forward motion period (Fig. 1 d):

```
mapBed -a motion_state.bed -b celegans_behavioral_feature.bed -c 5 -o mean -null 0 | awk 'BEGIN{OFS="\t";} {print $1,$2,$3,$10}' > mean_behavioral_feature_motion_state.bed
```

(Note that the actual BEDtool command is piped into the awk command-line tool to enable the extraction of the required columns.) Here -a motion\_state.bed specifies motion state intervals, -b celegans\_behavioral\_feature.bed the file containing the values to be overlapped (map) with motion state intervals, -c sets the file column of “b” that contains values, -o mean indicates the operation to be performed on -c, -null 0 outputs

zero when no overlapping intervals are found for a motion state interval, and finally awk is used to format the output as a bedGraph file since native mapBed output format is bed. To obtain the phenotypic features specifically occurring during forward, backward and paused states, locomotion features bedGraph files were intersected with BED files containing motion using the following BEDtools command:

```
intersectBed -i celegans_behavioral_feature.bedGraph -b motion_state.bed >
behavioral_feature_i_motion_state.bed
```

The direct rendering of tracks can be very helpful to visually detect disparity for a certain measurable behavior between two groups of animals. For instance in Fig. 1e midbody speed tracks of N2 strain worms clearly differ from those of the unc-16 strain. These observations can be confirmed and further analyzed using the above-mentioned intersected data, which can then be easily fed to R (R Core Team 2015), or a similar statistical software, to compare for example the distribution of mid body speed at a certain motion state (forward, backward and paused) as shown on Fig 1f.

From the visualization it became evident that speed of different body parts was correlated. We transformed bedGraph files obtained using Pergola into bigWig format using the following command:

```
bedGraphToBigWig celegans_behavioral_feature.bedGraph chrom.sizes
celegans_behavioral_feature.bw
```

BigWig is a compressed, binary format and allows the analysis of the data using tools available from the UCSC browser (Kent et al. 2010) (bigWigSummary, bigWigCorrelate, among others) or the bwtool command-line utility. These utilities can be used to calculate the overall correlation between two bigWig tracks, for instance speeds of different body parts of the same individual:

```
bigWigCorrelate celegans_behavioral_feature_1.bw
celegans_behavioral_feature_2.bw
```

Users can access other functionalities available on bwtool such as smoothing the data by applying a sliding window:

```
bwtool window 127 celegans_behavioral_feature.bw -fill=0 -center | awk -
f window_avg.awk > celegans_behavioral_feature.windows.bedGraph
```

Here, window 127 sets the length of the window, -fill tag sets values that are not available to 0 and -centers avoids the creation of overlapping intervals since they are not valid for bedGraph format. The result is piped into a simple awk script that averages the values of each window.

We used the bwtool window command to obtain the correlation between each window of two body parts of a worm:

```
bwtool window 127 celegans_behavioral_feature_body_p1.bw -fill=0 -center  
> celegans_behavioral_feature_body_p1.windows  
bwtool window 127 celegans_behavioral_feature_body_p2.bw -fill=0 -center >  
celegans_behavioral_feature_body_p2.windows  
win_cor.R w1=celegans_behavioral_feature_body_p1.windows  
w2=celegans_behavioral_feature_body_p1.windows > cor_head_midbody.bedGraph
```

In this case, two files corresponding to two different body parts of the same individual are processed using bwtool and the resulting files are given as an input to a small R script that just takes all the values of a windows of each file and calculates the correlation between them. In Figure 1c we represented the smoothed speeds of head and midbody of an N2 control worm and the results of the correlation between the smoothed windows.

**JAABA data visualization.** GAL4 flies lines data used in (Kabra et al. 2013) was obtained from [http://sourceforge.net/projects/jaaba/files/Sample%20Data/sampled\\_data\\_v0.5.zip/download](http://sourceforge.net/projects/jaaba/files/Sample%20Data/sampled_data_v0.5.zip/download). These tracking data relate to a line of *Drosophila Melanogaster* showing an increased propensity towards chasing. Chasing bouts were identified using JAABA, a machine learning-based software that annotates behaviors assigning them a confidence score. Resulting files were fed to Pergola.

**IGV adapted version.** We adapted IGV (Thorvaldsdóttir, Robinson, and Mesirov 2013), a popular desktop genome browser, to display temporal measurements instead of genomic features. We forked the IGV repository from github <https://github.com/igvteam/igv> and the modified version is available for download at <https://github.com/JoseEspinosa/IBB.git>. While this modified version has been used to produce the rendering shown on Fig 1a to e, the native application could generate the same visualization as our modifications involved only cosmetic improvements. Descriptions of how to visualize data are available at the Pergola documentation site (<http://cbcr.github.io/pergola/>).

**Analysis reproducibility.** Reproducibility is one of the main concerns of science nowadays (Goodman 2016), particularly in the neuroscience community (Button et al. 2013; Otchy et al. 2015). Contributing to the reproducibility crisis is an inability to recompute results (replicability). In the effort to ensure reproducibility, beside the adoption of standard formats and tools, we provide the users with all the resources to reproduce the Pergola paper results on our Github repository. By using Nextflow ([www.nextflow.io](http://www.nextflow.io)), a workflow framework that supports Docker, we guarantee the stability of the code, environment and third-party software through the version control and the containerization provided by the integration of both technologies. In short, Nextflow provides a way to code the analysis as independent tasks of a more complex

pipeline. Each of these tasks can be run in a single or several containers available on public repositories that sandbox all the software and environment needed for performing a reproducible analysis. The version of a whole analysis pipeline likewise can be controlled using the native integration of nextflow with Github, provided that a tag or revision has been previously specified. In this way to reproduce the analysis performed in Yemini et al., we just have to run this command:

```
nextflow run cbcrg/pergola/examples/N2_vs_case_strain_distros.nf
-r v0.1 --path_files="path2case_mat_files/" --
ctrl_path_files="path2ctrl_mat_files/" -tag_results="tag_output_folder"
```

### Supplementary Data - Table S1. Pergola Controlled Vocabulary

Term name	Mandatory	Definition
chrom_start	yes	Refers to start time points of each interval of the original data. If “chrom_end” is not set all “chrom_start” should be equidistant and intervals will be set to the delta between time points.
data_values	yes	Refers to associated values consider for the representation of data.
chrom_end	no	Refers to the end of each time interval.
track	no	Refers to each of the experimental entities present in the file.
data_types	no	Refers to each of the different features annotated in the file.
chrom	no	Refers to different phases of the experiment.
dummy	no	All additional fields in the original input data not used by pergola

### Supplementary Data - Table 2. Generic conversion file

<i>! Mapping of behavioral fields into pergola ontology terms</i>		
<i>! comments</i>		
<i>behavioral_file:phase_exp</i>	<i>&gt;</i>	<i>pergola:chrom</i>
<i>behavioral_file:individual_id</i>	<i>&gt;</i>	<i>pergola:track</i>
<i>behavioral_file:event_start</i>	<i>&gt;</i>	<i>pergola:start</i>
<i>behavioral_file:event_end</i>	<i>&gt;</i>	<i>pergola:end</i>
<i>behavioral_file:type_of_event</i>	<i>&gt;</i>	<i>pergola:data_types</i>
<i>behavioral_file:value</i>	<i>&gt;</i>	<i>pergola:data_value</i>

### Supplementary Data - Table S3. C.elegans conversion file

```
! Mapping of c.elegans behavioral fields into pergola ontology terms
!
behavioral_file:frame_start      >   pergola:start
behavioral_file:frame_end       >   pergola:end
behavioral_file:direction       >   pergola:data_types
behavioral_file:value           >   pergola:data_value
```

### Supplementary Data - Table S4. JAABA conversion File

```
! Mapping of behavioral JAABA fields into pergola ontology terms
!
behavioral_file:animal          >   pergola:track
behavioral_file:startTime      >   pergola:start
behavioral_file:endTime        >   pergola:end
behavioural_file:dataType      >   pergola:data_types
behavioral_file:value           >   pergola:data_value
```

## References

- Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25 (1). Nature America Inc.: 25–29.
- Button, Katherine S., John P. a. Ioannidis, Claire Mokrysz, Brian a. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews. Neuroscience* 14 (5). Nature Publishing Group: 365–76.
- "External Mapping File Format | Gene Ontology Consortium." 2015. Accessed November 19. <http://geneontology.org/page/external-mapping-file-format>.
- "GFF3 - GMOD." 2015. Accessed November 20. <http://gmod.org/wiki/GFF3>.
- Goodman, Steven N. 2016. "STATISTICS. Aligning Statistical and Scientific Reasoning." *Science* 352 (6290): 1180–81.
- Kabra, Mayank, Alice A. Robie, Marta Rivera-Alba, Steven Branson, and Kristin Branson. 2013. "JAABA: Interactive Machine Learning for Automatic Annotation of Animal Behavior." *Nature Methods* 10 (1). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 64–67.
- Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. 2002. "The Human Genome Browser at UCSC," 996–1006.

- Kent, W. J., A. S. Zweig, G. Barber, A. S. Hinrichs, and D. Karolchik. 2010. "BigWig and BigBed: Enabling Browsing of Large Distributed Datasets." *Bioinformatics* 26 (17): 2204–7.
- Otchy, Timothy M., Steffen B. E. Wolff, Juliana Y. Rhee, Cengiz Pehlevan, Risa Kawai, Alexandre Kempf, Sharon M. H. Gobes, and Bence P. Ölveczky. 2015. "Acute off-Target Effects of Neural Circuit Manipulations." *Nature* 528 (7582): 358–63.
- R Core Team. 2015. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Thorvaldsdóttir, Helga, James T. Robinson, and Jill P. Mesirov. 2013. "Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration." *Briefings in Bioinformatics* 14 (2): 178–92.
- Yemini, Eviatar, Tadas Jucikas, Laura J. Grundy, André E. X. Brown, and William R. Schafer. 2013. "A Database of *Caenorhabditis Elegans* Behavioral Phenotypes." *Nature Methods* 10 (9): 877–79.

## 3.2. Time-course and dynamics of the development of obesity-related behavioral changes induced by two types of energy-dense foods in mice

### **Preface**

The analysis of long behavioral trajectories can provide very interesting hints on how disorders with a long natural history and a strong behavioral component, such as obesity, develop and progress. In this paper, we identified how altered behavioral patterns appear upon the exposure to two types of widely-craved foods, chocolate and high-fat content food. To this end, we used Pergola, our computational framework for the analysis and visualization of longitudinal behavioral data. We show how this analysis framework is a very powerful way of characterizing the disruption of subtle behavioral patterns and how it allows obtaining interesting hints on different temporal scales. We also perform classical end point characterization of obesity and show how our approach provides a powerful tool to describe, analyze and interpret the emergence of behavioral patterns along time. Remarkably, we demonstrate how the analysis of the time course of complex disorder can lead to the identification of early disruptions of behavior that can be a very interesting biomarker for the early therapeutical intervention.

Espinosa-Carrasco J, Burokas A, Fructuoso M, Erb I, Martín-García E, Gutiérrez-Martos M, et al. [Time-course and dynamics of obesity-related behavioral changes induced by energy-dense foods in mice](#). *Addict Biol.* 2018 Mar;23(2):531–43. DOI: 10.1111/adb.12595

### 3.3. Principal Component Analysis of the Effects of Environmental Enrichment and (-)-Epigallocatechin-3-Gallate on Age-Associated Learning Deficits in a Mouse Model of Down Syndrome

#### **Preface**

The MWM test has become one of the most widely used behavioral paradigms for the assessment of cognitive capabilities. The common practice to infer the cognitive performance of individuals on the MWM is the use of univariate indexes. Nonetheless, this is usually an underpowered and rather subjective approach to grasp the complexity of a behavioral dimension such as cognition. Therefore, my contribution to this paper was the development of a multidimensional analysis based on principal component analysis (PCA). We applied this methodology to an experiment aiming to assess the effects of a therapeutical intervention on cognition in the Ts65Dn mouse model of down syndrome at the age of onset of cognitive decline. We demonstrate that a supervised PCA is able to capture the main contribution of several behavioral variables to a composite measure of learning. On top of this, we developed a nonparametric (permutation) test to assess the differences of this composite measure of learning among the different combinations of genotype and treatment. Our results provide a proof of concept of how this approach leads to a better, global assessment of behavioral multidimensionality and how we can use this methodology to unravel differences between treatment groups that otherwise are difficult to detect with classical univariate analysis.

Catuara-Solarz S, Espinosa-Carrasco J, Erb I, Langohr K, Notredame C, Gonzalez JR, et al. [Principal Component Analysis of the Effects of Environmental Enrichment and \(-\)-epigallocatechin-3-gallate on Age-Associated Learning Deficits in a Mouse Model of Down Syndrome](#). *Front Behav Neurosci*. 2015 Dec 11;9:330. DOI: 10.3389/fnbeh.2015.00330



### 3.4. Combined treatment with environmental enrichment and (-)-epigallocatechin-3-gallate ameliorates learning deficits and hippocampal alterations in a mouse model of Down syndrome

#### **Preface**

In this work we applied the same multidimensional approach to a second experiment assessing the effects in cognition of a therapeutical intervention on the same mouse model of Down Syndrome (Ts65Dn), but this time on an early age developmental stage. In this paper, we demonstrate that our approach is robust enough to be systematically used for the assessment of different behavioral domains on behavioral paradigms aimed to understand precise behavioral domains.

Catuara-Solarz S, Espinosa-Carrasco J, Erb I, Langohr K, Gonzalez JR, Notredame C, et al. [Combined Treatment With Environmental Enrichment and \(-\)-Epigallocatechin-3-Gallate Ameliorates Learning Deficits and Hippocampal Alterations in a Mouse Model of Down Syndrome](#). *eNeuro*. 2016 Nov 10;3(5). DOI: 10.1523/ENEURO.0103-16.2016

## 4. DISCUSSION

Behavioral neuroscience has undergone a paradigm shift with the advent of technologies for the high-throughput recording of behavioral data (Gomez-Marin et al. 2014). Hence, the era of big data has arrived to behavioral neuroscience. These big behavioral data offer the opportunity of studying brain function with an unprecedented amount of quantitative behavioral recordings. However, the exploration, analysis and modelling of such data represents a great challenge (Hidalgo-Mazzei et al. 2016).

A comprehensive analysis of long-term longitudinal behavioral recordings by computational methods has to deal with heterogeneous data types. These data include the long-term behavioral recordings of multiple individuals and the associated-behavioral data. To date, there is a lack of tools that are flexible enough to integrate all these large volumes of heterogeneous behavioral data on multiple individuals for an explorative analysis. Beside the commercial systems (Phecomp, Phenomaster, Ethovision), some open-source tools allow to visualize variables derived from video trackings of posture and movements together with locomotion or social behaviors and to browse the data along their timeline (Kabra et al. 2013; de Chaumont et al. 2012). None of these tools provide, however, a framework for the unified analysis of longitudinal recordings of behavior. Moreover, in most instances, we want to combine those data with physiological or molecular readouts and results of the annotation or the modelling of the data. Each of these data types are generated by different systems and comes along with its own tools for its analysis and visualization, if available, making the analysis even more difficult. Therefore a key aspect when analyzing behavior is how to cope with this diversity. To overcome all these limitations of handling big behavioral data, my thesis has aimed to provide a bioinformatics tool for the visualization and analysis of longitudinal big behavioral data. This led to the development of Pergola (Results 4.1), an open-source software that unifies the visualization of the raw variables resulting from behavioral tracking systems, the behavioral events produced by annotation software, derived of the experiment and any other relevant temporal information for the contextualization of behavior. Therefore, Pergola enables to capture the subtle behavioral variations in response to specific environmental signals.

The originality of Pergola is that it converts these third-party temporal datasets into widely used standard genomic file formats (Sims et al. 2014). The advantage of the adoption of genomic file formats is that they provides an entry point to genomic bioinformatics tools. This is a great opportunity, because the genomics community has developed a myriad of efficient computational tools for the visualization and explorative analysis of genomic data (Hinrichs et al. 2006; Quinlan and Hall 2010; J. T. Robinson and Thorvaldsdóttir 2011; Dale, Pedersen, and Quinlan 2011; Pohl and Beato 2014).

We reasoned that since both genomic and behavioral data share a sequential structure, these tools could be used for the analysis of behavioral data. Genome browsers (Kent et al. 2002; J. T. Robinson and Thorvaldsdóttir 2011), for instance, allow the visualization of behavioral longitudinal data, while processing tools such as BEDTools (Quinlan and Hall 2010) or bwtool (Pohl and Beato 2014) enable the unified analysis of behavioral data and its annotations. Furthermore, I also developed in Pergola a set of own utilities for the manipulation of the data and the annotation of periodic environmental cues.

The next step was to validate the Pergola framework in a real experimental situation. To this aim we decided to use it on a quite challenging question, with a strong translational relevance: the study of a disorder with a long natural history. This aim was inspired by the observation that some relevant brain disorders progress during a long time without clinical symptoms. Schizophrenia (Davidson et al. 1999b), Alzheimer Disease (Sperling, Aisen, Beckett, Bennett, Craft, et al. 2011) or some types of eating disorders (Polivy and Herman 2002) are among the most widely known examples. We chose diet-induced obesity as a model, and hypothesized that altered behavioral patterns could emerge during the long presymptomatic phase, i.e., before significant weight increases are established. This is especially relevant, since some authors have proposed that the intervention during early disease stages is the strategy to follow to ameliorate the prognosis of patients affected by these long-term disorders (Dawson 2008; Sommer et al. 2016). However, the temporal progression of the eating behaviors altered during obesity development has not been previously reported.

In our study we used Pergola for the characterization of the time course of behavioral disruptions provoked by exposure to two of the most widely craved foods. The aim was to identify behavioral patterns that can be used as biomarkers of diet-induced obesity during its preclinical phase, since this could pave the way to early intervention in diet-induced obesity. An intervention driven by the early detection of behavioral biomarkers could represent an opportunity to restrict the long-term irreversible consequence of obesity (Wabitsch 2000; Kopelman 2000). We demonstrated that behavioral disruption occurs very early after the exposure to high-caloric foods (Results 4.2), much earlier than the common phenotypic trait used for its diagnostic, namely weight, becomes visible. Moreover, we established common and differential behavioral changes that discriminate the different types of diets. These experiments showed that Pergola was extremely useful for data visualization, but also to establish short- and long-phase oscillations in eating behavior. Moreover, we were able to combine this description with the different analytical tools, extracting relevant information about the fine-grained characteristics of feeding behavior. Therefore, we have demonstrated that high-resolution temporal studies of spontaneous behavior are a powerful tool to unravel the dynamics of disease progression in disorders with a long preclinical history, and that

Pergola is very efficient and overcomes some limitations of big behavioral data visualization and analysis.

Pergola is easy-to-use and freely available to the scientific community on GitHub (<https://github.com/cbcrg/pergola/>), and licensed under the General Public License version 3.0 (GPLv3). Overall, Pergola offers a unified computational framework for a comprehensive computational analysis of longitudinal recordings of behavior in its rich context of associated data sets. In addition, the Pergola website (<http://cbcrg.github.io/pergola/>) hosts several hands-on tutorials where users can learn how to use Pergola for the analysis of longitudinal behavioral recordings of feeding (Results 4.2). Future efforts will be devoted to release a new version of Pergola that incorporates the option of interactively visualizing the data using Shiny (<http://shiny.rstudio.com/>), utilizing a similar approach as Sleuth, a tool for the differential analysis of RNA-Seq data (Pimentel et al. 2016). This feature will allow users to render and browse the temporal visualizations together with insightful charts in a web application. Finally, Pergola can be a useful tool for other fields of research that are rapidly accumulating big corpora of sequential data, examples include societal and health system contexts (de Montjoye et al. 2013; Jensen et al. 2014; de Montjoye et al. 2015).

The final work of my thesis was aimed at the study of complex, multidimensional behaviors. Classically, the spatial learning capacity of an animal on a multidimensional behavioral task such as the MWM (D'Hooge and De Deyn 2001) is assessed by univariate indexes. However, these approaches rely on the rather simplistic, subjective assumption that a complex behavioral response such as cognition can be grasped by a unique index. Therefore the cognitive concepts derived from these indexes should be interpreted with caution (Wolfer and Lipp 2000; Graziano, Petrosini, and Bartoletti 2003). This pitfall led some authors to propose the use of multidimensional techniques to tackle the study of processes made up of multiple behavioral domains (Wolfer and Lipp 2000; Graziano, Petrosini, and Bartoletti 2003; Kumar et al. 2011). We reported in this thesis a PCA-based supervised multidimensional approach for the study of a complex behavioral process (Results 4.3).

These results provide a way of understanding different learning-related cognitive domains considering the different dimensions affecting behavioral response as opposed to classical single-variate measures. Therefore, this method constitutes a better approximation to obtain reliable behavioral measures. Another remarkable feature of our approach is that it allows to reduce a confounding factor that commonly masks results derived from univariate measures, namely the experimental variance that is unrelated to the behavioral domain of interest. Finally, we show how our approach enables the statistical comparison among different genotypes and experimental

conditions (treatments) using a nonparametric permutation test. Our results, thus, provide the first example, to the best of our knowledge, of a PCA-based methodology that also assesses the statistical significance of behavioral differences found by the multidimensional approach. Interestingly enough, we applied the same methodology to a second experiment reporting similar comparable results (results 4.4) and thus, we can conclude that our methodology is robust enough to be used on the complex behavioral readouts of similar behavioral paradigms.

All in all, in this thesis we have advanced in the understanding of behavior by the analysis of automated behavioral recordings. As we have discussed here, classical behavioral studies suffer from several shortcomings. Among the most important ones we can mention lack of good quantitative measures, limited temporal resolution, subjective scoring and poor reproducibility. Big behavioral data offers the opportunity to ameliorate some of these problems. Nonetheless, in order to achieve this rather ambitious goal, behavioral neuroscience needs to develop tools capable of dealing with an overwhelming amount of behavioral data. We strongly believe that a new generation of computational methods that incorporate the wisdom acquired by the different schools studying behavior, would enable to shed light on one of the most intriguing biological questions: how the brain works.

## 5. CONCLUSIONS

1. The characterization of longitudinal high-throughput behavioral recordings and its complete spectrum of behavioral-relevant data requires a comprehensive study of its time course. Therefore, we have developed a visualization and analytical framework for the explorative analysis of the time-dependent changes of longitudinal behavioral data.

2. Previous tools do not allow the unified study of longitudinal behavioral data on a flexible interface. We have wrapped up our methodological framework in one open-source tool, Pergola, which we have made available to the behavioral neuroscience community providing access to a myriad of ready-to-use genomic analysis tools.

3. Using Pergola we were able to identify the dynamics of development of altered behavioral patterns upon the exposure to two types of widely craved foods, chocolate and high-fat content food.

4. The common approach to tackle complex multidimensional behaviors is based on univariate measures. In this thesis, we have adapted PCA to grasp the different components of a complex behavioral domain in a composite variable. To assess significance among different experimental conditions we implemented a nonparametric permutation test.

5. We have validated the relevance of our multidimensional approach on two independent experiments using a common behavioral test for the assessment of different aspects of the cognitive domain. Thus, we have demonstrated the robustness of our approach.

6. Our results in a preclinical therapeutic trial on a Down syndrome mouse model provide a proof of concept of how this approach leads to a better, global assessment of behavioral multidimensionality and how we can use this methodology to unravel difference between treatment groups that otherwise are difficult to detect with classical univariate analysis.

7. We have also demonstrated that our approach is robust enough to be systematically used for the assessment of different behavioral domains on behavioral paradigms aimed to understand precise behavioral domains.

## 6. LIST OF COMMUNICATIONS

1. Espinosa-Carrasco J., Erb I., Dierssen M., Notredame C. Pergola: a new paradigm for longitudinal data visualization and analysis. In preparation.
2. Espinosa-Carrasco J., Aurelijus B., Fructuoso M., Erb I., Martín-García E., Gutiérrez-Martos M., Notredame C., Maldonado R., Dierssen M. Time-course and dynamics of the development of obesity-related behavioral changes induced by two types of energy-dense foods in mice. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*. Manuscript under revision, (2016).
3. Espinosa-Carrasco, J., Erb I., Catuara-Solarz S., Langohr K., Notredame C., Gonzalez J. R., and Dierssen M. Principal Component Analysis of the Effects of Environmental Enrichment and (-)-Epigallocatechin-3-Gallate on Age-Associated Learning Deficits in a Mouse Model of Down Syndrome. *Frontiers in Behavioral Neuroscience*. 2015; 9 (December): 330.
4. Catuara-Solarz S., Espinosa-Carrasco J., Erb I., Langohr K., Notredame C., Gonzalez J. R., Dierssen M. Combined treatment with environmental enrichment and (-)-epigallocatechin-3-gallate ameliorates learning deficits and hippocampal alterations in a mouse model of Down syndrome. *E-Neuro*, In press (2016).

## 7. APPENDIX

### Methods for the automated recording of behavior.

#### Automation of behavioral assays

Behavioral assays have been extensively used to systematically understand the functions of genes expressed in the nervous system by using transgenic, knockout or other newer genetically engineered mice (Crawley 2008). The goal of these assays is to evaluate capacities such as sensory discrimination (Cybulska-Klosowicz and Kossut 2001), cognitive performance (D’Hooge and De Deyn 2001; F. E. Harrison, Hosseini, and McDonald 2009; Deacon and Rawlins 2006; Leger et al. 2013), motor function (Brooks and Dunnett 2009), feeding (Heyne et al. 2009), or assess responses such as fear (Davis 2006), pain (Cavanaugh et al. 2009), anxiety (Holmes 2001) and drug addiction (Maldonado et al. 2011). The characterization of these behavioral phenotypes has been successfully applied to define or validate gene functions, characterize the role of receptor isoforms, to study disease phenotypes using translational animal models (Chen et al. 2000; Reeves et al. 1995; Yamamoto, Lucas, and Hen 2000) and evaluate possible treatments (Pons-Espinal, Martinez de Lagran, and Dierssen 2013; Yamamoto, Lucas, and Hen 2000; D’Amico et al. 2016; Brunner, Nestler, and Leahy 2002). Behavioral assays aim to characterize behavioral domains in a systematic and reproducible manner, and to achieve this goal, these tests minimize the space of possible behavioral outputs by reducing the uncertainties in the outcomes by constraining the experimental setup at expense of a greater experimental rigidity. For example, operant conditioning paradigms evaluate responses as simple as lever presses to obtain a reward and their associated measures (latencies, escalation, extinction), or T-maze tests measure the tendency of animals to either explore a new environment or to obtain a reward by choosing between the left and right arms of a maze shaped as the letter T. Strategies such as combining several tests are commonly employed as a means to discard possible confounding factors (an animal is not performing correctly a test because it is blind or motor impaired) or as a way to evaluate several behavioral processes occurring in human disorders (Bućan and Abel 2002; Seong, Seasholtz, and Burmeister 2002).

As even a reader without behavioral testing expertise may suspect, these paradigms can be very tedious, time-consuming, expensive and also error prone when manually conducted. Hence, since the initial attempts to apply them, researchers have tried to automatize these behavioral paradigms as much as possible (Skinner 1932; Winter and Schaefer 2011). However, it was not until relatively recently that automation and monitoring software solutions became a game-changer of the field. Data acquisition was one of the first important improvements information technologies provided. Scores



annotated by direct observations can lead to unreliable behavioral indexes even when dealing with measures as simple as the time spent eating during a deprivation test in spite of using the same criteria (Wahlsten et al. 2003). Because of this, many of the behavioral readouts such as the number of lever presses in operant conditioning boxes, food or liquid consumption, activity or time spent in a given zone of a maze are automatically annotated in almost all current devices using equipments such as photoelectric sensors, lickometers or weight transducers. A further advantage of automated systems over hand-scored versions of the paradigms is that they enable to characterize behaviors that otherwise remain undetectable or unquantifiable, for instance, reaction time on the scale of milliseconds (Abraham et al. 2004; Uchida and Mainen 2003). However, although automated versions of behavioral tests reduce experimental time, the animals have to still be transferred in and out of the home cage with the notorious risk of increasing experimental variability due to the animal handling (Andrews and File 1993; Sorge et al. 2014; Tecott and Nestler 2004). This pitfall has led to the proliferation of systems that integrate mouse housing and experiments to avoid the physical manipulation of the animals. The operating principle of these devices relies on the use of Radio Frequency IDentification (RFID) transponders that uniquely identify animals. Animals can be then housed together in a single cage, and the systems control the access of each animal to the operant compartment or to conditioning boxes where animals perform the programmed behavioral tasks. Since some model organism such as mice or rats are among social animals (Ferhat et al. 2015; Latham and Mason 2004/6; Vanderschuren, Niesink, and Van Pee 1997), these devices have the added benefit of not socially depriving the animals under study, a factor that can affect the results of behavioral tests (Olsson and Westlund 2007; Wilkinson et al. 1994). Equally important, these systems enable the continuous observation of the animals, an issue which we will come back to below.

### **Sensor-based systems for the continuous monitoring of single animal behaviors**

In recent years many different commercial and non-commercial systems have been developed for the tracking of single animals using different types of sensors. Sensors used for the detection of behavior include photobeams for recording of general and vertical activity (Sakkou et al. 2007; Goulding et al. 2008; Bura et al. 2010) or actions such as nose-poking (Abraham et al. 2004) or feeding (Goulding et al. 2008), infrared (IR) sensors (Dell’Omo et al. 2002; Pendergast et al. 2013) mechanical vibrations sensors (Van de Weerd et al. 2001; Megens et al. 1987) for the monitoring of general activity, weighing transducers (Bura et al. 2010) for the monitoring of food and liquid consumption, or gas sensors for the monitoring of oxygen requirements. One of the most extensively used animal models in behavior is mouse because of the ease of its genetic manipulation, the abundance of techniques to obtain relevant physiological

readouts and the translational relevance of its study (Bućan and Abel 2002; Nestler and Hyman 2010). Consequently, the number of devices for the recording of rodent behavior has grown significantly in recent years. Many of these systems consist in a regular rodent home cage equipped with different types of the above-mentioned sensors assembled by the researchers in a modular manner (Goulding et al. 2008) or arranged on a ready-to-use commercial setup (Van de Weerd et al. 2001; Tang, Orchard, and Sanford 2002; Bura et al. 2010). Some popular sensor-based commercial systems include the PheCOMP System (<http://www.panlab.com/en/products/phecomp-system-panlab>) that enables the monitoring of eating, drinking, vertical and horizontal activity (shown in Figure 8 as it has been the system that we used for recording mice feeding behavior), the Phenomater systems (<http://www.tse-systems.com/products/metabolism/home-cage/phenomaster/index.htm>) a modular platform that allows for the combination of several components for the recording of readouts such as eating, drinking, vertical/horizontal activity and calorimetry among other possibilities and LABORAS (<http://www.metris.nl/laboras/laboras.htm>), which monitors the vibrations produced by the animals to distinguish between behaviors such as drinking, eating, grooming and different types of activities. Moreover, several companies offer IR frames that can be situated around the cage to track spontaneous activity of rodents inside their regular cages. IR actimeter (<http://www.panlab.com/es/productos/actimetro-ir-actimeter-panlab>) or Actimot2 (<http://www.tse-systems.com/products/metabolism/home-cage/phenomaster/activity/actimot2.htm>) are just two examples. Sensor-based approaches record a coarse-grained signal when compared with video-based approaches (Hueihan Jhuang et al. 2010) (see below). However, some of the sensors described, as for instance weighing transducers or mechanical vibrations sensors, offer the advantage of not relying on the quality of the images and the presence of light and do not suffer from an interruption of the tracking of the animals by objects masking the signal (Bailoo, Bohlen, and Wahlsten 2010), hence they allow for instance the placement of objects in the home cage for environmental enrichment. All in all, the use of sensor-based home cages has proved to be an effective strategy to identify phenotypic activity differences on disease mouse models (Dell’Omo et al. 2002; Tang, Orchard, and Sanford 2002; L. Robinson et al. 2013) or to identify differential feeding patterns in animal models of obesity (Goulding et al. 2008; Heyne et al. 2009; Edelsbrunner, Herzog, and Holzer 2009).

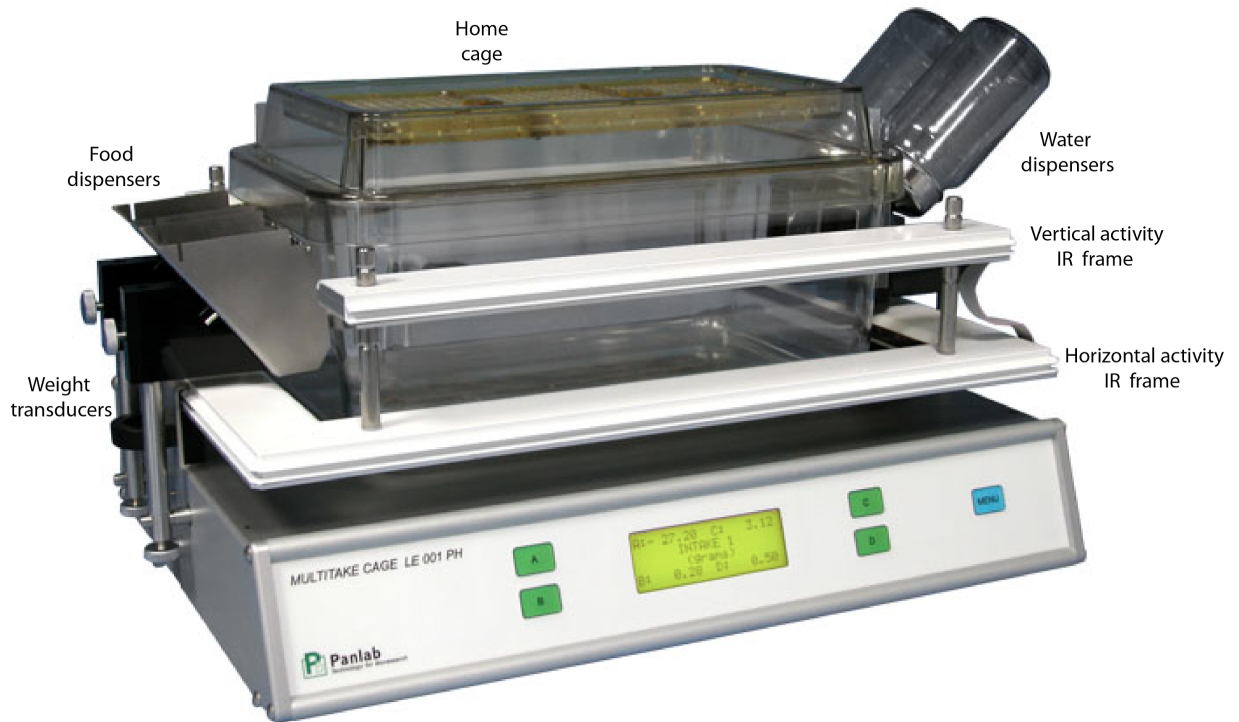


Figure 8 | The PheCOMP multi-take metabolism and activity cage (Panlab-Harvard Instruments, Barcelona, Spain). The PheCOMP cage is a high-resolution monitoring system that allow for the longitudinal monitoring of eating and drinking acts. Cages are equipped with two feeders and two drinking dispensers. The system records food/water intake by means of weight transducer. Besides, the PheCOMP cage incorporates an IR frame to continuously record horizontal (locomotor) and vertical (rearing) activity during the whole recording time.

### Video-based monitoring systems

Sensor-based monitoring methodologies can yield relevant insights into the temporal organization of behavior and the possible implications for the underlying biological processes that might lead to different behavioral phenotypes, both in health and disease. However, the range of commercially available detectors restricts the possible behavioral readouts captured by these systems and only can provide a relatively coarse signal (Hueihan Jhuang et al. 2010; Schaefer and Claridge-Chang 2012). Recent advances in computer vision systems (Dollar et al. 2005; H. Jhuang et al. 2007; Moeslund, Hilton, and Krüger 2006) have fostered the development of a myriad of methodologies for the continuous video-monitoring of animal behavior. Video-based systems offer advantages such as the fine-grained nature of the signal, the potential capability of recording any macroscopic behavior and the low cost of its implementation, among other aspects that we will discuss. Obviously, sensor and video-based approaches can be combined to obtain very powerful methodologies and reveal all the spatio-temporal dynamics of behavior.

The study of behavior has for a long time aimed at the video monitoring of relevant behavioral readouts. First attempts to achieve this goal can be considered nowadays rather elemental and manual implementations (Buelthoff, Poggio, and Wehrhahn 1980; Moran, Fentress, and Golani 1981; Crawley et al. 1982; B. M. Spruijt and Gispen 1983). Nonetheless, these pioneers foresaw the potential of the usage of automatic direct observations of behavior. Since then, video recording systems, thanks in part to the ubiquitous presence of consumer electronics in our society, have become relatively cheap and widely accessible, making the acquisition, storage and processing of enormous amounts of fine-grained video recordings affordable for researchers (Gomez-Marín et al. 2014). This technological development has led to a parallel progress of computational solutions for the problems behavioral science has to face.

### **Video-based monitoring systems for the monitoring of individual animals**

First video tracking systems enabled only the extraction of features of single animals due to technical limitations. Indeed, animal-background segmentation is not a trivial algorithmic problem and the segmentation of several individuals increases the difficulty of the task (Egnor and Branson 2016). Nonetheless, even a coarse monitoring of animal behavioral features capable of yielding space-locomotion measures such as speed or distance travelled can provide very informative quantitative data for some behavioral paradigms such as the MWM, fear conditioning or the open-field test. As in the case of sensor-based monitoring systems, we can also find commercial or open-source systems. Commercial software provides ready-to-use solutions for a specific paradigm at the expense of a poor adaptability to particular requirements. Some interesting solutions are FreezeFrame of Actimetrics (<http://actimetrics.com/products/freeze-frame/>) for fear conditioning, SMART (<http://www.panlab.com/en/products/smart-video-tracking-software-panlab>), VideoMot (<http://www.tse-systems.com/products/behavior/videotracking/videomot2.htm>) or EthoVision (Noldus, Spink, and Tegelenbosch 2001) (<http://www.noldus.com/animal-behavior-research/products/ethovision-xt>) three software tracking solutions that can be adapted for several behavioral tasks relying on mazes and several model organisms and finally, Phenotyper (<http://www.noldus.com/animal-behavior-research/products/phenotyper>) or HomeCageScan ([http://cleversysinc.com/csi\\_products/homecagescan/](http://cleversysinc.com/csi_products/homecagescan/)) two video-based system to monitor spontaneous activity of mice in their home cage. Many of these solutions are subject to constant evolution to incorporate additional features and although first they were only able to track one individual or monitor a very precise behavior they have been extended to other model organisms (Martin 2004). However, many researchers preferred their own, tailor-made solutions which they openly distributed to the community. In this manner, users with some programming skills enjoy

the additional benefit of being able to adapt the code to particular needs (Aguiar, Mendonça, and Galhardo 2007; Gomez-Marin et al. 2012). OpenControl or Autotyping for instance, are open source softwares designed for addressing the tracking of rodents in several behavioral mazes (Aguiar, Mendonça, and Galhardo 2007; Patel et al. 2014). The Worm Tracker allows the extraction of motor and morphological features of *C. elegans* (Yemini et al. 2013). Similar solutions are available for the tracking of body posture and motion of *Drosophila* adults (Martin 2004; Valente, Golani, and Mitra 2007) and larvae (Aleman-Meza, Jung, and Zhong 2015). Interestingly, SOS software is able to track several model organisms (Gomez-Marin et al. 2012).

### **Video-based monitoring systems for the monitoring of multiple animals**

As already mentioned, social deprivation can affect behavior and furthermore some behaviors (courtship, mating, aggression, etc.) are only observable in a social context. However, the identification of individual animals in video recordings is a hard algorithmic problem. Consequently, development of systems capable of simultaneously tracking multiple animals has been in the focus of intense research efforts. To simplify the segmentation problem, some systems make use of color dyes (Noldus, Spink, and Tegelenbosch 2001) or bleach patterns on the back of the animal (Ohayon et al. 2013) for the unambiguous tracking of animals. Since this approach is unfeasible when dealing with small animals like flies, some authors created tools to automatically detect complex social behaviors such as courtship and aggression tracking only two flies (Nilsen et al. 2004; Hoyer et al. 2008; Dankert et al. 2009). Shortly after, MiceProfiler allowed the study of social interaction of mice in a dual setup (de Chaumont et al. 2012). However, these systems were not able to track large populations of untagged animals. To overcome this limitation, a new generation of methods was developed enabling to keep the identity of numerous animals (Gilestro and Cirelli 2009; Branson et al. 2009; Kohlhoff et al. 2011; Swierczek et al. 2011b; Straw et al. 2011; Pérez-Escudero et al. 2014). Noteworthy, Pysolo allows for the high-throughput screenings of *Drosophila* locomotion and sleep in a low-cost setup based on webcams (Gilestro and Cirelli 2009). In the case of *C. elegans*, the Multi-Worm tracker provides the community with a very nice real-time method capable of extracting multiple round worm behaviors with a screening system. Another interesting tool, CTRAX, enables to track the position and orientation of walking flies and based on these parameters, it uses machine-learning techniques to automatically annotate behaviors (Branson et al. 2009) as shown in Figure 2. Finally, idTracker uses animal “fingerprints” to robustly follow the trajectories of several animals (Pérez-Escudero et al. 2014). Interestingly, it exploits the “fingerprints” of the animals to avoid the propagation of possible errors and in this manner it can keep track of especially difficult cases such as mice. Although some of these computational methods have been originally designed for a given model organism,

it is often the case that they work well for other model organisms. This is the case of CTRAX and idTracker, to mention just two examples.

### **Combined approaches**

Direct observations of behavior can be very enlightening with respect to underlying disorders or phenotypic differences. On the other hand, behavioral paradigms often help us to understand precise behavioral domains. Thus, sometimes approaches that combine both can be very efficient for grasping different aspects of the behavioral response. As an example, the social housing of animals whose access to the operant modules is controlled by RFID transponders allows to combine operant experiments, classically used in psychology, with longitudinal observations of spontaneous behaviors, classically studied in ethology (The Intellicage from Newbehavior ([www.newbehavior.com](http://www.newbehavior.com)) is an outstanding example of this type of systems). Spontaneous behaviors comprise raw signals such as the number and duration of entries in operant compartments, licking events, nosepokes or behavioral bouts derived from video tracking. These longitudinal readouts can be interpreted in terms of relevant behavioral signals such as total activity, exploration or place preference. The combination of both strategies during prolonged periods of time can be fruitful to reveal insights about the progression of neurodegenerative diseases (Rudenko et al. 2009; Codita et al. 2010; Oakeshott et al. 2011; Lewejohann et al. 2009) or the impact of treatments on mental disorders (Branchi et al. 2013). To obtain more ecologically relevant results, Lewejohann et al. built a semi-naturalistic experimental setup with RFID readers placed in strategic positions (Figure 9). The experiment allowed to understand long-term patterns of dominance, movement and place preference of a mouse model of Alzheimer's disease. Furthermore, just the data recorded during periods with free access to all cage compartments can already be turned into informative behavioral profiles that discriminate between different mouse models (Vannoni et al. 2014). Video-based monitoring systems have also been successfully combined with experimentation. Remarkably, a recent study reported how video-based monitoring of a social home cage where animals were challenged with different tests can effectively uncover behavioral signatures for a Huntington's disease genotype (Alexandrov et al. 2016).

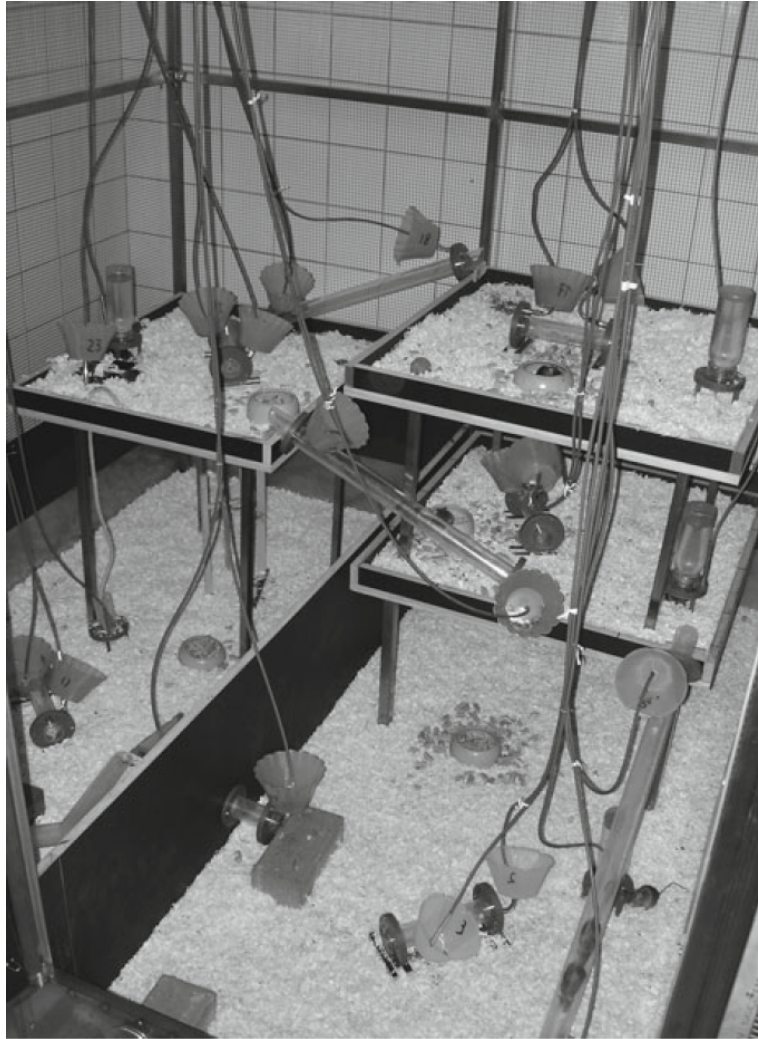


Figure 9 | Picture from a self-made, semi-naturalistic social housing that used RFID transducers to obtain the movement patterns, place preference and social interactions of a mouse model of Alzheimer's disease for behavioral phenotyping (adapted from Lewejohann et al. 2009).

Beside behavior and environment, another main ingredient for understanding brain functioning is neurophysiology. To gain insights into the role of a given protein (or any product of a gene) in normal brain function, researchers develop approaches relying on the knocking-out of genes and the subsequent screening for the phenotypes that may be modified when compared with genetically unaltered animals, which is known as reverse genetics (Bućan and Abel 2002). Behavior is one of the possible phenotypes that can be characterized in order to hypothesize which possible molecular mechanism leads to the behavioral alteration (John C. Crabbe and Morris 2004; Bućan and Abel 2002). The current state-of-the-art in the field of electrophysiology allows for the obtention of longitudinal electrophysiological recordings (Buzsáki 2004; Warby et al. 2014). As an example, it has been possible to monitor the flying behavior of fixed-head *Drosophila* and contextualize it with the underlying neural activity (Maimon, Straw, and Dickinson 2010). Furthermore, it is now possible to use optogenetics to modulate neural

populations and trigger certain behaviors (Deisseroth 2015). Therefore, the integration of dense behavioral recordings, long high-quality electrophysiological monitoring and interventions in the system become an important approach for understanding brain function in its broad context (Maimon, Straw, and Dickinson 2010; Anikeeva et al. 2012; Chiappe et al. 2010).

### **Tools for the automatic annotation of high-throughput behavioral recordings**

All these high-throughput methodologies for the screening of behavior are resulting in collections of big chunks of longitudinal data or BBD (Gomez-Marin et al. 2014). The analysis of these longitudinal data may reveal the connection between complex phenotypic traits like behavior and the neural systems behind. However, the first step to accomplish this goal relies on a correct annotation of the sequence of behavioral bouts. The high throughput of these systems makes it unfeasible, in practical terms, to manually annotate these data. Hence, the explosion of methodologies for the high-throughput screening of behavior has given rise to a subsequent explosion of tools for the automatic annotation of behavior (Egnor and Branson 2016). Indeed, many of the presented software applications for the monitoring of longitudinal behaviors already incorporate more or less sophisticated algorithmic solutions for the computational behavioral annotation of the gathered recordings (Branson et al. 2009; de Chaumont et al. 2012). Remarkably, JAABA, an interactive tool for the automatic annotation of complex sequences (Kabra et al. 2013), allows users to use as input the longitudinal posture and motion readouts extracted from video-based recordings of different trackers (Branson et al. 2009; Dankert et al. 2009; Swierczek et al. 2011b). Most of these computational solutions use either supervised or unsupervised machine-learning methods or rule-based classifiers to automatically annotate behavioral bouts from quantitative measurement of animal behavior (Egnor and Branson 2016). Given the strong lateral dependency of behavioral series, Hidden Markov Models (HMM) approaches are a popular method to model the transition probabilities between contiguous behavioral annotations (Hueihan Jhuang et al. 2010; Wiltchko et al. 2015). Regardless of how behavioral events are characterized, a paramount aspect for the successful annotation of behavior is to contextualize as much as possible the readouts provided by the model. This will help to detect possible data artifacts and hence will avoid data misinterpretations. Indeed, this concerns other types of readouts, and explorative data analysis has a paramount importance in the correct modelling of the data.





## 8. BIBLIOGRAPHY

- Abdi, H., and L. J. Williams. 2010. "Principal Component Analysis." *Wiley Interdisciplinary Reviews: Wiley Online Library*. <http://onlinelibrary.wiley.com/doi/10.1002/wics.101/full>.
- Abraham, Nixon M., Hartwig Spors, Alan Carleton, Troy W. Margrie, Thomas Kuner, and Andreas T. Schaefer. 2004. "Maintaining Accuracy at the Expense of Speed: Stimulus Similarity Defines Odor Discrimination Time in Mice." *Neuron* 44 (5): 865–76.
- Aguiar, Paulo, Luís Mendonça, and Vasco Galhardo. 2007. "OpenControl: A Free Opensource Software for Video Tracking and Automated Control of Behavioral Mazes." *Journal of Neuroscience Methods* 166 (1): 66–72.
- Aleman-Meza, Boanerges, Sang-Kyu Jung, and Weiwei Zhong. 2015. "An Automated System for Quantitative Analysis of Drosophila Larval Locomotion." *BMC Developmental Biology* 15 (February): 11.
- Alexandrov, Vadim, Dani Brunner, Liliana B. Menalled, Andrea Kudwa, Judy Watson-Johnson, Matthew Mazzella, Ian Russell, et al. 2016. "Large-Scale Phenome Analysis Defines a Behavioral Signature for Huntington's Disease Genotype in Mice." *Nature Biotechnology* 34 (8): 838–44.
- Anderson, David J., and Pietro Perona. 2014. "Toward a Science of Computational Ethology." *Neuron* 84 (1): 18–31.
- Andrews, N., and S. E. File. 1993. "Handling History of Rats Modifies Behavioural Effects of Drugs in the Elevated plus-Maze Test of Anxiety." *European Journal of Pharmacology* 235 (1): 109–12.
- Anikeeva, Polina, Aaron S. Andalman, Ilana Witten, Melissa Warden, Inbal Goshen, Logan Groseknick, Lisa A. Gunaydin, Loren M. Frank, and Karl Deisseroth. 2012. "Optetrode: A Multichannel Readout for Optogenetic Control in Freely Moving Mice." *Nature Neuroscience* 15 (1): 163–70.
- Anscombe, F. J. 1973. "Graphs in Statistical Analysis." *The American Statistician* 27 (1). [amstat.tandfonline.com](http://amstat.tandfonline.com): 17–21.
- Bailoo, Jeremy D., Martin O. Bohlen, and Douglas Wahlsten. 2010. "The Precision of Video and Photocell Tracking Systems and the Elimination of Tracking Errors with Infrared Backlighting." *Journal of Neuroscience Methods* 188 (1): 45–52.
- Baker, Monya. 2011. "Animal Models: Inside the Minds of Mice and Men." *Nature* 475 (7354): 123–28.
- Barnes, C. A. 1979. "Memory Deficits Associated with Senescence: A Neurophysiological and Behavioral Study in the Rat." *Journal of Comparative and Physiological Psychology* 93 (1): 74–104.
- Beck WS, Liem KF, Simpson GG. *Life, An Introduction To Biology*. 3rd edn Harper Collins; New York: 1991
- Belzung, C., and F. Berton. 1997. "Further Pharmacological Validation of the BALB/c Neophobia in the Free Exploratory Paradigm as an Animal Model of Trait Anxiety." *Behavioural Pharmacology* 8 (6-7): 541–48.
- Benjamini, Yoav, Ehud Fonio, Tal Galili, Gregor Z. Havkin, and Ilan Golani. 2011. "Quantifying the Buildup in Extent and Complexity of Free Exploration in Mice." *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 3 (September): 15580–87.
- Bishop, C. 2007. "Pattern Recognition and Machine Learning (Information Science and Statistics), 1st Edn. 2006. Corr. 2nd Printing Edn." Springer, New York.

- Branchi, Igor, Sara Santarelli, Sara Capoccia, Silvia Poggini, Ivana D'Andrea, Francesca Cirulli, and Enrico Alleva. 2013. "Antidepressant Treatment Outcome Depends on the Quality of the Living Environment: A Pre-Clinical Investigation in Mice." *PloS One* 8 (4): e62226.
- Branson, Kristin, Alice A. Robie, John Bender, Pietro Perona, and Michael H. Dickinson. 2009. "High-Throughput Ethomics in Large Groups of *Drosophila*." *Nature Methods* 6 (6). Nature Publishing Group: 451–57.
- Breed, M., and L. Sanchez. 2010. "Both Environment and Genetic Makeup Influence Behavior." *Nature Educ Knowl.* nature.com. <http://www.nature.com/scitable/knowledge/library/both-environmentandgenetic-makeup-influence-behavior-13907840>.
- Brooks, Simon P., and Stephen B. Dunnett. 2009. "Tests to Assess Motor Phenotype in Mice: A User's Guide." *Nature Reviews. Neuroscience* 10 (7). Nature Publishing Group: 519–29.
- Brunner, Dani, Eric Nestler, and Emer Leahy. 2002. "In Need of High-Throughput Behavioral Systems." *Drug Discovery Today* 7 (18 Suppl): S107–12.
- Bućan, Maja, and Ted Abel. 2002. "The Mouse: Genetics Meets Behaviour." *Nature Reviews. Genetics* 3 (2): 114–23.
- Buelthoff, H., T. Poggio, and C. Wehrhahn. 1980. "3-D Analysis of the Flight Trajectories of Flies (*Drosophila Melanogaster*)." *Zeitschrift Für Naturforschung C.* degruyter.com. <http://www.degruyter.com/view/j/znc.1980.35.issue-9-10/znc-1980-9-1024/znc-1980-9-1024.xml>.
- Bura, S. Andreea, Aurelijus Burokas, Elena Martín-García, and Rafael Maldonado. 2010. "Effects of Chronic Nicotine on Food Intake and Anxiety-like Behaviour in CB(1) Knockout Mice." *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology* 20 (6): 369–78.
- Buzsáki, György. 2004. "Large-Scale Recording of Neuronal Ensembles." *Nature Neuroscience* 7 (5): 446–51.
- Castro, Eric Campos Vieira, and Miriam Mendonça Morato Andrade. 2005. "Longitudinal Study of the Spectral Composition of Behavioral Rhythms in the Rat." *Biological Rhythm Research* 36 (1-2): 131–40.
- Cavanaugh, Daniel J., Hyosang Lee, Liching Lo, Shannon D. Shields, Mark J. Zylka, Allan I. Basbaum, and David J. Anderson. 2009. "Distinct Subsets of Unmyelinated Primary Sensory Fibers Mediate Behavioral Responses to Noxious Thermal and Mechanical Stimuli." *Proceedings of the National Academy of Sciences of the United States of America* 106 (22): 9075–80.
- Chen, G., K. S. Chen, J. Knox, J. Inglis, A. Bernard, S. J. Martin, A. Justice, et al. 2000. "A Learning Deficit Related to Age and Beta-Amyloid Plaques in a Mouse Model of Alzheimer's Disease." *Nature* 408 (6815): 975–79.
- Chiappe, M. Eugenia, Johannes D. Seelig, Michael B. Reiser, and Vivek Jayaraman. 2010. "Walking Modulates Speed Sensitivity in *Drosophila* Motion Vision." *Current Biology: CB* 20 (16): 1470–75.
- Chiel, H. J., and R. D. Beer. 1997. "The Brain Has a Body: Adaptive Behavior Emerges from Interactions of Nervous System, Body and Environment." *Trends in Neurosciences* 20 (12): 553–57.
- Codita, Alina, Astrid Gumucio, Lars Lannfelt, Pär Gellerfors, Bengt Winblad, Abdul H. Mohammed, and Lars N. G. Nilsson. 2010. "Impaired Behavior of Female Tg-ArcSwe APP Mice in the IntelliCage: A Longitudinal Study." *Behavioural Brain Research* 215 (1): 83–94.

- Colles, S. L., J. B. Dixon, and P. E. O'Brien. 2007. "Night Eating Syndrome and Nocturnal Snacking: Association with Obesity, Binge Eating and Psychological Distress." *International Journal of Obesity* 31 (11): 1722–30.
- Crabbe, J. C., D. Wahlsten, and B. C. Dudek. 1999. "Genetics of Mouse Behavior: Interactions with Laboratory Environment." *Science* 284 (5420): 1670–72.
- Crabbe, John C., and Richard G. M. Morris. 2004. "Festina Lente: Late-Night Thoughts on High-Throughput Screening of Mouse Behavior." *Nature Neuroscience* 7 (11): 1175–79.
- Crawley, Jacqueline N. 2008. "Behavioral Phenotyping Strategies for Mutant Mice." *Neuron* 57 (6): 809–18.
- Crawley, J. N., S. Szara, G. T. Pryor, C. R. Creveling, and B. K. Bernard. 1982. "Development and Evaluation of a Computer-Automated Color TV Tracking System for Automatic Recording of the Social and Exploratory Behavior of Small Animals." *Journal of Neuroscience Methods* 5 (3): 235–47.
- Curtis, Claire, and Caroline Davis. 2014. "A Qualitative Study of Binge Eating and Obesity from an Addiction Perspective." *Eating Disorders* 22 (1): 19–32.
- Cybulska-Klosowicz, A., and M. Kossut. 2001. "Mice Can Learn Roughness Discrimination with Vibrissae in a Jump Stand Apparatus." *Acta Neurobiologiae Experimentalis* 61 (1): 73–76.
- Dale, Ryan K., Brent S. Pedersen, and Aaron R. Quinlan. 2011. "Pybedtools: A Flexible Python Library for Manipulating Genomic Datasets and Annotations." *Bioinformatics* 27 (24): 3423–24.
- D'Amico, Davide, Thomas Gener, Maria Martínez de Lagrán, Maria V. Sanchez-Vives, Mónica Santos, and Mara Dierssen. 2016. "Infralimbic Neurotrophin-3 Infusion Rescues Fear Extinction Impairment in a Mouse Model of Pathological Fear." *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, August. doi:10.1038/npp.2016.154.
- Dankert, Heiko, Liming Wang, Eric D. Hoopfer, David J. Anderson, and Pietro Perona. 2009. "Automated Monitoring and Analysis of Social Behavior in Drosophila." *Nature Methods* 6 (4): 297–303.
- Davidson, M., A. Reichenberg, J. Rabinowitz, M. Weiser, Z. Kaplan, and M. Mark. 1999a. "Behavioral and Intellectual Markers for Schizophrenia in Apparently Healthy Male Adolescents." *The American Journal of Psychiatry* 156 (9): 1328–35.
- 1999b. "Behavioral and Intellectual Markers for Schizophrenia in Apparently Healthy Male Adolescents." *The American Journal of Psychiatry* 156 (9): 1328–35.
- Davies, Alex, Matthieu Louis, and Barbara Webb. 2015. "A Model of Drosophila Larva Chemotaxis." *PLoS Computational Biology* 11 (11): e1004606.
- Davis, Michael. 2006. "Neural Systems Involved in Fear and Anxiety Measured with Fear-Potentiated Startle." *The American Psychologist* 61 (8): 741–56.
- Dawson, Geraldine. 2008. "Early Behavioral Intervention, Brain Plasticity, and the Prevention of Autism Spectrum Disorder." *Development and Psychopathology* 20 (3): 775–803.
- Deacon, Robert M. J., and J. Nicholas P. Rawlins. 2006. "T-Maze Alternation in the Rodent." *Nature Protocols* 1 (1): 7–12.
- de Chaumont, Fabrice, Renata Dos-Santos Coura, Pierre Serreau, Arnaud Cressant, Jonathan Chabout, Sylvie Granon, and Jean-Christophe Olivo-Marin. 2012. "Computerized Video Analysis of Social Interactions in Mice." *Nature Methods* 9 (4). Nature Publishing Group: 410–17.
- Deisseroth, Karl. 2015. "Optogenetics: 10 Years of Microbial Opsins in Neuroscience." *Nature Neuroscience* 18 (9): 1213–25.

- Delcomyn, F. 1980. "Neural Basis of Rhythmic Behavior in Animals." *Science* 210 (4469): 492–98.
- Dell’Omo, Giacomo, Elisabetta Vannoni, Alexei L. Vyssotski, Michele Angelo Di Bari, Romolo Nonno, Umberto Agrimi, and Hans-Peter Lipp. 2002. "Early Behavioural Changes in Mice Infected with BSE and Scrapie: Automated Home Cage Monitoring Reveals Prion Strain Differences." *The European Journal of Neuroscience* 16 (4): 735–42.
- de Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. "Unique in the Crowd: The Privacy Bounds of Human Mobility." *Scientific Reports* 3: 1376.
- de Montjoye, Yves-Alexandre, Laura Radaelli, Vivek Kumar Singh, and Alex Sandy Pentland. 2015. "Identity and Privacy. Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata." *Science* 347 (6221): 536–39.
- D’Hooge, R., and P. P. De Deyn. 2001. "Applications of the Morris Water Maze in the Study of Learning and Memory." *Brain Research. Brain Research Reviews* 36 (1): 60–90.
- Dierssen, Mara. 2012. "Down Syndrome: The Brain in Trisomic Mode." *Nature Reviews. Neuroscience* 13 (12): 844–58.
- Dollar, P., V. Rabaud, G. Cottrell, and S. Belongie. 2005. "Behavior Recognition via Sparse Spatio-Temporal Features." In *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 65–72.
- Donaldson, Zoe R. 2010. "We’re the Same... but Different: Addressing Academic Divides in the Study of Brain and Behavior." *Frontiers in Behavioral Neuroscience* 4 (July). doi:10.3389/fnbeh.2010.00041.
- Dubois, Bruno, Harald Hampel, Howard H. Feldman, Philip Scheltens, Paul Aisen, Sandrine Andrieu, Hovagim Bakardjian, et al. 2016. "Preclinical Alzheimer’s Disease: Definition, Natural History, and Diagnostic Criteria." *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association* 12 (3): 292–323.
- Edelsbrunner, Martin E., Herbert Herzog, and Peter Holzer. 2009. "Evidence from Knockout Mice That Peptide YY and Neuropeptide Y Enforce Murine Locomotion, Exploration and Ingestive Behaviour in a Circadian Cycle- and Gender-Dependent Manner." *Behavioural Brain Research* 203 (1): 97–107.
- Egnor, S. E. Roian, and Kristin Branson. 2016. "Computational Analysis of Behavior." *Annual Review of Neuroscience* 39 (1): 217–36.
- Eichenbaum, Howard. 2000. "A Cortical–hippocampal System for Declarative Memory." *Nature Reviews. Neuroscience* 1 (1). Nature Publishing Group: 41–50.
- Ferhat, Allain-Thibeault, Anne-Marie Le Sourd, Fabrice de Chaumont, Jean-Christophe Olivo-Marin, Thomas Bourgeron, and Elodie Ey. 2015. "Social Communication in Mice--Are There Optimal Cage Conditions?" *PloS One* 10 (3): e0121802.
- Fisher, R. A. 1936. "THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS." *Annals of Eugenics* 7 (2). Blackwell Publishing Ltd: 179–88.
- Fonio, Ehud, Yoav Benjamini, and Ilan Golani. 2012. "Short and Long Term Measures of Anxiety Exhibit Opposite Results." *PloS One* 7 (10): e48414.
- Fonio, Ehud, Ilan Golani, and Yoav Benjamini. 2012. "Measuring Behavior of Animal Models: Faults and Remedies." *Nature Methods* 9 (12): 1167–70.
- Gallagher, M., and R. D. Burwell. 1989. "Relationship of Age-Related Decline across Several Behavioral Domains." *Neurobiology of Aging* 10 (6): 691–708.
- Galsworthy, Michael J., Irmgard Amrein, Pavel A. Kuptsov, Inga I. Poletaeva, Pascal Zinn, Anton Rau, Alexei Vyssotski, and Hans-Peter Lipp. 2005. "A Comparison of

- Wild-Caught Wood Mice and Bank Voles in the Intellicage: Assessing Exploration, Daily Activity Patterns and Place Learning Paradigms.” *Behavioural Brain Research* 157 (2): 211–17.
- Gerlai, R., and N. S. Clayton. 1999. “Analysing Hippocampal Function in Transgenic Mice: An Ethological Perspective.” *Trends in Neurosciences* 22 (2): 47–51.
- Gerlai, Robert. 2002. “Phenomics: Fiction or the Future?” *Trends in Neurosciences* 25 (10): 506–9.
- Gilestro, Giorgio F., and Chiara Cirelli. 2009. “pySolo: A Complete Suite for Sleep Analysis in Drosophila.” *Bioinformatics* 25 (11): 1466–67.
- Gomez-Marin, Alex, Nicolas Partoune, Greg J. Stephens, Matthieu Louis, and Björn Brembs. 2012. “Automated Tracking of Animal Posture and Movement during Exploration and Sensory Orientation Behaviors.” *PloS One* 7 (8): e41642.
- Gomez-Marin, Alex, Joseph J. Paton, Adam R. Kampff, Rui M. Costa, and Zachary F. Mainen. 2014. “Big Behavioral Data: Psychology, Ethology and the Foundations of Neuroscience.” *Nature Neuroscience* 17 (11): 1455–62.
- Goulding, Evan H., A. Katrin Schenk, Punita Juneja, Adrienne W. MacKay, Jennifer M. Wade, and Laurence H. Tecott. 2008. “A Robust Automated System Elucidates Mouse Home Cage Behavioral Structure.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (52): 20575–82.
- Graziano, Alessandro, Laura Petrosini, and Alessandro Bartoletti. 2003. “Automatic Recognition of Explorative Strategies in the Morris Water Maze.” *Journal of Neuroscience Methods* 130 (1): 33–44.
- Greenacre, M. J. 2010. “Biplots in Practice: Fundación BBVA.”
- Hall, C. S. 1934. “Emotional Behavior in the Rat. I. Defecation and Urination as Measures of Individual Differences in Emotionality.” *Journal of Comparative Psychology* 18 (3). Williams & Wilkins Company: 385.
- Harrison, F. E., A. H. Hosseini, and M. P. McDonald. 2009. “Endogenous Anxiety and Stress Responses in Water Maze and Barnes Maze Spatial Memory Tasks.” *Behavioural Brain Research* 198 (1): 247–51.
- Harrison, Fiona E., Randall S. Reiserer, Andrew J. Tomarken, and Michael P. McDonald. 2006. “Spatial and Nonspatial Escape Strategies in the Barnes Maze.” *Learning & Memory* 13 (6): 809–19.
- Hersen, M. 2005. *Encyclopedia of Behavior Modification and Cognitive Behavior Therapy: Volume I: Adult Clinical Applications Volume II: Child Clinical Applications Volume III: Educational Applications*. SAGE Publications.
- Heurich, Marco, Anton Hilger, Helmut Küchenhoff, Henrik Andréén, Luděk Bufka, Miha Krofel, Jenny Mattisson, et al. 2014. “Activity Patterns of Eurasian Lynx Are Modulated by Light Regime and Individual Traits over a Wide Latitudinal Range.” *PloS One* 9 (12): e114143.
- Heyne, Christoph Kiesselbach, Ignasi Sahún, Jerome McDonald, Monica Gaiffi, Mara Dierssen, and Jochen Wolffgramm. 2009. “An Animal Model of Compulsive Food-Taking Behaviour.” *Addiction Biology* 14 (4): 373–83.
- Hidalgo-Mazzei, Diego, Andrea Murru, María Reinares, Eduard Vieta, and Francesc Colom. 2016. “Big Data in Mental Health: A Challenging Fragmented Future.” *World Psychiatry: Official Journal of the World Psychiatric Association* 15 (2): 186–87.
- Hill, J. O. 1998. “Environmental Contributions to the Obesity Epidemic.” *Science* 280 (5368): 1371–74.

- Hinrichs, A. S., D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, et al. 2006. "The UCSC Genome Browser Database: Update 2006." *Nucleic Acids Research* 34 (Database issue): D590–98.
- Hofstetter, Hedwig, Elise Dusseldorp, Pepijn van Empelen, and Theo W. G. M. Paulussen. 2014. "A Primer on the Use of Cluster Analysis or Factor Analysis to Assess Co-Occurrence of Risk Behaviors." *Preventive Medicine* 67 (October): 141–46.
- Holmes, A. 2001. "Targeted Gene Mutation Approaches to the Study of Anxiety-like Behavior in Mice." *Neuroscience and Biobehavioral Reviews* 25 (3): 261–73.
- Hong, Weizhe, Ann Kennedy, Xavier P. Burgos-Artizzu, Moriel Zelikowsky, Santiago G. Navonne, Pietro Perona, and David J. Anderson. 2015. "Automated Measurement of Mouse Social Behaviors Using Depth Sensing, Video Tracking, and Machine Learning." *Proceedings of the National Academy of Sciences of the United States of America* 112 (38): E5351–60.
- Hotelling, H. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24 (6). Warwick & York: 417.
- Hoyer, Susanne C., Andreas Eckart, Anthony Herrel, Troy Zars, Susanne A. Fischer, Shannon L. Hardie, and Martin Heisenberg. 2008. "Octopamine in Male Aggression of *Drosophila*." *Current Biology: CB* 18 (3): 159–67.
- Hoyle, Graham. 1984. "The Scope of Neuroethology." *The Behavioral and Brain Sciences* 7 (3): 367.
- Jensen, Anders Boeck, Pope L. Moseley, Tudor I. Oprea, Sabrina Gade Ellesøe, Robert Eriksson, Henriette Schmock, Peter Bjødstrup Jensen, Lars Juhl Jensen, and Søren Brunak. 2014. "Temporal Disease Trajectories Condensed from Population-Wide Registry Data Covering 6.2 Million Patients." *Nature Communications* 5 (May): 1–10.
- Jhuang, H., T. Serre, L. Wolf, and T. Poggio. 2007. "A Biologically Inspired System for Action Recognition." In *2007 IEEE 11th International Conference on Computer Vision*, 1–8.
- Jhuang, Hueihan, Estibaliz Garrote, Xinlin Yu, Vinita Khilnani, Tomaso Poggio, Andrew D. Steele, and Thomas Serre. 2010. "Automated Home-Cage Behavioural Phenotyping of Mice." *Nature Communications* 1 (September). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 68.
- Jud, Corinne, Isabelle Schmutz, Gabriele Hampf, Henrik Oster, and Urs Albrecht. 2005. "A Guideline for Analyzing Circadian Wheel-Running Behavior in Rodents under Different Lighting Conditions." *Biological Procedures Online* 7 (July): 101–16.
- Kabra, Mayank, Alice A. Robie, Marta Rivera-Alba, Steven Branson, and Kristin Branson. 2013. "JAABA: Interactive Machine Learning for Automatic Annotation of Animal Behavior." *Nature Methods* 10 (1). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 64–67.
- Kenny, Paul J. 2011. "Common Cellular and Molecular Mechanisms in Obesity and Drug Addiction." *Nature Reviews. Neuroscience* 12 (11). Nature Publishing Group: 638–51.
- Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. 2002. "The Human Genome Browser at UCSC." *Genome Research* 12 (6): 996–1006.
- Koboldt, Daniel C., Karyn Meltz Steinberg, David E. Larson, Richard K. Wilson, and Elaine R. Mardis. 2013. "The next-Generation Sequencing Revolution and Its Impact on Genomics." *Cell* 155 (1): 27–38.

- Kohlhoff, Kai J., Thomas R. Jahn, David A. Lomas, Christopher M. Dobson, Damian C. Crowther, and Michele Vendruscolo. 2011. "The iFly Tracking System for an Automated Locomotor and Behavioural Analysis of *Drosophila Melanogaster*." *Integrative Biology: Quantitative Biosciences from Nano to Macro* 3 (7): 755–60.
- Koob, G. F., P. P. Sanna, and F. E. Bloom. 1998. "Neuroscience of Addiction." *Neuron* 21 (3): 467–76.
- Kopelman, P. G. 2000. "Obesity as a Medical Problem." *Nature* 404 (6778): 635–43.
- Kumar, V., K. Kim, C. Joseph, L. C. Thomas, H. Hong, and J. S. Takahashi. 2011. "Second-Generation High-Throughput Forward Genetic Screen in Mice to Isolate Subtle Behavioral Mutants." *Proceedings of the National Academy of Sciences* 108 (Supplement\_3): 15557–64.
- Latagliata, Emanuele Claudio, Enrico Patrono, Stefano Puglisi-Allegra, and Rossella Ventura. 2010. "Food Seeking in Spite of Harmful Consequences Is under Prefrontal Cortical Noradrenergic Control." *BMC Neuroscience* 11 (January): 15.
- Latham, Naomi, and Georgia Mason. 2004/6. "From House Mouse to Mouse House: The Behavioural Biology of Free-Living *Mus Musculus* and Its Implications in the Laboratory." *Applied Animal Behaviour Science* 86 (3–4): 261–89.
- Leger, Marianne, Anne Quiedeville, Valentine Bouet, Benoît Haelewyn, Michel Boulouard, Pascale Schumann-Bard, and Thomas Freret. 2013. "Object Recognition Test in Mice." *Nature Protocols* 8 (12). Nature Research: 2531–37.
- Lewejohann, Lars, Anne Marie Hoppmann, Philipp Kegel, Mareike Kritzler, Antonio Krüger, and Norbert Sachser. 2009. "Behavioral Phenotyping of a Murine Model of Alzheimer's Disease in a Seminaturalistic Environment Using RFID Tracking." *Behavior Research Methods* 41 (3): 850–56.
- Lorenz, Karl. 1937. "On the Formation of the Concept of Instinct." *Natural Sciences* 25 (19): 289–300.
- Lorenz, K. Z. 1958. "The Evolution of Behavior." *Scientific American* 199 (6): 67–74 passim.
- Maimon, Gaby, Andrew D. Straw, and Michael H. Dickinson. 2010. "Active Flight Increases the Gain of Visual Motion Processing in *Drosophila*." *Nature Neuroscience* 13 (3): 393–99.
- Maldonado, R., F. Berrendero, A. Ozaita, and P. Robledo. 2011. "Neurochemical Basis of Cannabis Addiction." *Neuroscience* 181 (May): 1–17.
- Marr, M. J. 1992. "Behavior Dynamics: One Perspective." *Journal of the Experimental Analysis of Behavior* 57 (3): 249–66.
- Martin, Jean-René. 2004. "A Portrait of Locomotor Behaviour in *Drosophila* Determined by a Video-Tracking Paradigm." *Behavioural Processes* 67 (2): 207–19.
- Mazur, J. E. 2015. *Learning and Behavior: Seventh Edition*. Taylor & Francis.
- Mega, M. S., J. L. Cummings, T. Fiorello, and J. Gornbein. 1996. "The Spectrum of Behavioral Changes in Alzheimer's Disease." *Neurology* 46 (1): 130–35.
- Megens, A. A., J. Voeten, J. Rombouts, T. F. Meert, and C. J. Niemegeers. 1987. "Behavioral Activity of Rats Measured by a New Method Based on the Piezo-Electric Principle." *Psychopharmacology* 93 (3): 382–88.
- Mery, Frederic, and James G. Burns. 2009. "Behavioural Plasticity: An Interaction between Evolution and Experience." *Evolutionary Ecology* 24 (3): 571–83.
- Moeslund, Thomas B., Adrian Hilton, and Volker Krüger. 2006. "A Survey of Advances in Vision-Based Human Motion Capture and Analysis." *Computer Vision and Image Understanding: CVIU* 104 (2–3): 90–126.



- Moran, Greg, John C. Fentress, and Ilan Golani. 1981. "A Description of Relational Patterns of Movement during 'ritualized Fighting' in Wolves." *Animal Behaviour* 29 (4): 1146–65.
- Morris, R. 1984. "Developments of a Water-Maze Procedure for Studying Spatial Learning in the Rat." *Journal of Neuroscience Methods* 11 (1): 47–60.
- Morris, Sarah E., and Bruce N. Cuthbert. 2012. "Research Domain Criteria: Cognitive Systems, Neural Circuits, and Dimensions of Behavior." *Dialogues in Clinical Neuroscience* 14 (1): 29–37.
- Nestler, Eric J., and Steven E. Hyman. 2010. "Animal Models of Neuropsychiatric Disorders." *Nature Neuroscience* 13 (10): 1161–69.
- Nguyen, Huu Phuc, Philipp Kobbe, Henning Rahne, Till Wörpel, Burkard Jäger, Michael Stephan, Reinhard Pabst, et al. 2006. "Behavioral Abnormalities Precede Neuropathological Markers in Rats Transgenic for Huntington's Disease." *Human Molecular Genetics* 15 (21): 3177–94.
- Niemelä, Petri T., Anssi Vainikka, Jukka T. Forsman, Olli J. Loukola, and Raine Kortet. 2013. "How Does Variation in the Environment and Individual Cognition Explain the Existence of Consistent Behavioral Differences?" *Ecology and Evolution* 3 (2): 457–64.
- Nilsen, Steven P., Yick-Bun Chan, Robert Huber, and Edward A. Kravitz. 2004. "Gender-Selective Patterns of Aggressive Behavior in *Drosophila melanogaster*." *Proceedings of the National Academy of Sciences of the United States of America* 101 (33): 12342–47.
- Noldus, L. P., A. J. Spink, and R. A. Tegelenbosch. 2001. "EthoVision: A Versatile Video Tracking System for Automation of Behavioral Experiments." *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc* 33 (3): 398–414.
- Oakeshott, Stephen, Fuat Balci, Igor Filippov, Carol Murphy, Russell Port, David Connor, Ahmad Paintdakhi, et al. 2011. "Circadian Abnormalities in Motor Activity in a BAC Transgenic Mouse Model of Huntington's Disease." *PLoS Currents* 3 (April): RRN1225.
- Ohayon, Shay, Ofer Avni, Adam L. Taylor, Pietro Perona, and S. E. Roian Egnor. 2013. "Automated Multi-Day Tracking of Marked Mice for the Analysis of Social Behaviour." *Journal of Neuroscience Methods* 219 (1): 10–19.
- Olsson, I. Anna S., and Karolina Westlund. 2007. "More than Numbers Matter: The Effect of Social Factors on Behaviour and Welfare of Laboratory Rodents and Non-Human Primates." *Applied Animal Behaviour Science* 103 (3–4): 229–54.
- Pastizzo, Matthew J., Robert F. Erbacher, and Laurie B. Feldman. 2002. "Multidimensional Data Visualization." *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc* 34 (2): 158–62.
- Patel, Tapan P., David M. Gullotti, Pepe Hernandez, W. Timothy O'Brien, Bruce P. Capehart, Barclay Morrison, Cameron Bass, James E. Eberwine, Ted Abel, and David F. Meaney. 2014. "An Open-Source Toolbox for Automated Phenotyping of Mice in Behavioral Tasks." *Frontiers in Behavioral Neuroscience* 8 (October): 349.
- Pavlov, I. P. 1927. *Conditional Reflex, An Investigation of The Psychological Activity of the Cerebral Cortex*. New York, Oxford University press.
- Pendergast, Julie S., Katrina L. Branecky, William Yang, Kate L. J. Ellacott, Kevin D. Niswender, and Shin Yamazaki. 2013. "High-Fat Diet Acutely Affects Circadian Organisation and Eating Behavior." *The European Journal of Neuroscience* 37 (8): 1350–56.

- Pérez-Escudero, Alfonso, Julián Vicente-Page, Robert C. Hinz, Sara Arganda, and Gonzalo G. de Polavieja. 2014. "idTracker: Tracking Individuals in a Group by Automatic Identification of Unmarked Animals." *Nature Methods* 11 (7): 743–48.
- Petersen, Ronald C. 2010. "Alzheimer's Disease: Progress in Prediction." *Lancet Neurology* 9 (1): 4–5.
- Pimentel, Harold J., Nicolas Bray, Suzette Puente, Páll Melsted, and Lior Pachter. 2016. "Differential Analysis of RNA-Seq Incorporating Quantification Uncertainty." *bioRxiv*. doi:10.1101/058164.
- Pohl, Andy, and Miguel Beato. 2014. "Bwtool: A Tool for bigWig Files." *Bioinformatics* 30 (11): 1618–19.
- Polivy, Janet, and C. Peter Herman. 2002. "Causes of Eating Disorders." *Annual Review of Psychology* 53: 187–213.
- Pons-Espinal, Meritxell, Maria Martinez de Lagran, and Mara Dierssen. 2013. "Environmental Enrichment Rescues DYRK1A Activity and Hippocampal Adult Neurogenesis in TgDyrk1A." *Neurobiology of Disease* 60 (December). Elsevier Inc.: 18–31.
- Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42.
- Ramos, André. 2008. "Animal Models of Anxiety: Do I Need Multiple Tests?" *Trends in Pharmacological Sciences* 29 (10): 493–98.
- Ramot, Daniel, Brandon E. Johnson, Tommie L. Berry Jr, Lucinda Carnell, and Miriam B. Goodman. 2008. "The Parallel Worm Tracker: A Platform for Measuring Average Speed and Drug-Induced Paralysis in Nematodes." *PloS One* 3 (5): e2208.
- Reardon, Sara. 2016. "A Mouse's House May Ruin Experiments." *Nature News* 530 (7590): 264.
- Reeves, R. H., N. G. Irving, T. H. Moran, A. Wohn, C. Kitt, S. S. Sisodia, C. Schmidt, R. T. Bronson, and M. T. Davisson. 1995. "A Mouse Model for Down Syndrome Exhibits Learning and Behaviour Deficits." *Nature Genetics* 11 (2): 177–84.
- Richter, S. Helene, Joseph P. Garner, and Hanno Würbel. 2009. "Environmental Standardization: Cure or Cause of Poor Reproducibility in Animal Experiments?" *Nature Methods* 6 (4): 257–61.
- Robinson, J. T., and H. Thorvaldsdóttir. 2011. "Integrative Genomics Viewer." *Nature* 29 (1): 24–26.
- Robinson, Lianne, Andrea Plano, Stuart Cobb, and Gernot Riedel. 2013. "Long-Term Home Cage Activity Scans Reveal Lowered Exploratory Behaviour in Symptomatic Female Rett Mice." *Behavioural Brain Research* 250 (August): 148–56.
- Roubertoux, P. L., B. Martin, I. Le Roy, J. Beau, C. Marchaland, F. Perez-Diaz, C. Cohen-Salmon, and M. Carlier. 1996. "Vocalizations in Newborn Mice: Genetic Analysis." *Behavior Genetics* 26 (4): 427–37.
- Rudenko, Olga, Vadim Tkach, Vladimir Berezin, and Elisabeth Bock. 2009. "Detection of Early Behavioral Markers of Huntington's Disease in R6/2 Mice Employing an Automated Social Home Cage." *Behavioural Brain Research* 203 (2): 188–99.
- Sakkou, Maria, Petra Wiedmer, Katrin Anlag, Anne Hamm, Eve Seuntjens, Laurence Ettwiller, Matthias H. Tschöp, and Mathias Treier. 2007. "A Role for Brain-Specific Homeobox Factor Bsx in the Control of Hyperphagia and Locomotory Behavior." *Cell Metabolism* 5 (6): 450–63.
- Schaefer, Andreas T., and Adam Claridge-Chang. 2012. "The Surveillance State of Behavioral Automation." *Current Opinion in Neurobiology* 22 (1). Elsevier Ltd: 170–76.

- Seong, Eunju, Audrey F. Seasholtz, and Margit Burmeister. 2002. "Mouse Models for Psychiatric Disorders." *Trends in Genetics: TIG* 18 (12): 643–50.
- Silverman, Jill L., Mu Yang, Catherine Lord, and Jacqueline N. Crawley. 2010. "Behavioural Phenotyping Assays for Mouse Models of Autism." *Nature Reviews. Neuroscience* 11 (7): 490–502.
- Sims, David, Nicholas E. Illott, Stephen N. Sansom, Ian M. Sudbery, Jethro S. Johnson, Katherine A. Fawcett, Antonio J. Berlanga-Taylor, Sebastian Luna-Valero, Chris P. Ponting, and Andreas Heger. 2014. "CGAT: Computational Genomics Analysis Toolkit." *Bioinformatics* 30 (9): 1290–91.
- Skinner, B. F. 1932. "Drive and Reflex Strength." *The Journal of General Psychology* 6 (1): 22–37. 1937. "Two Types of Conditioned Reflex: A Reply to Konorski and Miller." *The Journal of General Psychology* 16 (1): 272–79.
- Sommer, Iris E., Carrie E. Bearden, Edwin van Dellen, Elemi J. Breetvelt, Sasja N. Duijff, Kim Maijer, Therese van Amelsvoort, et al. 2016. "Early Interventions in Risk Groups for Schizophrenia: What Are We Waiting For?" *NPJ Schizophrenia* 2 (March): 16003.
- Sorge, Robert E., Loren J. Martin, Kelsey A. Isbester, Susana G. Sotocinal, Sarah Rosen, Alexander H. Tuttle, Jeffrey S. Wieskopf, et al. 2014. "Olfactory Exposure to Males, Including Men, Causes Stress and Related Analgesia in Rodents." *Nature Methods* 11 (6): 629–32.
- Sperling, Reisa A., Paul S. Aisen, Laurel A. Beckett, David A. Bennett, Suzanne Craft, Anne M. Fagan, Takeshi Iwatsubo, et al. 2011. "Toward Defining the Preclinical Stages of Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease." *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 7 (3). Elsevier: 280–92.
- Sperling, Reisa A., Paul S. Aisen, Laurel A. Beckett, David A. Bennett, Craft Suzanne, Anne M. Fagan, Iwatsubo Takeshi, et al. 2011. "Toward Defining the Preclinical Stages of Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease." *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 7 (3): 280–92.
- Spruijt, Berry M., Suzanne M. Peters, Raymond C. de Heer, Helen H. J. Pothuizen, and Johanneke E. van der Harst. 2014. "Reproducibility and Relevance of Future Behavioral Sciences Should Benefit from a Cross Fertilization of Past Recommendations and Today's Technology: 'Back to the Future.'" *Journal of Neuroscience Methods*, March. doi:10.1016/j.jneumeth.2014.03.001.
- Spruijt, B. M., and W. H. Gispen. 1983. "Prolonged Animal Observation by Use of Digitized Videodisplays." *Pharmacology, Biochemistry, and Behavior* 19 (5): 765–69.
- Steele, Andrew D., Walker S. Jackson, Oliver D. King, and Susan Lindquist. 2007. "The Power of Automated High-Resolution Behavior Analysis Revealed by Its Application to Mouse Models of Huntington's and Prion Diseases." *Proceedings of the National Academy of Sciences of the United States of America* 104 (6): 1983–88.
- Stepanichev, Mikhail, Nikolay N. Dygalo, Grigory Grigoryan, Galina T. Shishkina, and Natalia Gulyaeva. 2014. "Rodent Models of Depression: Neurotrophic and Neuroinflammatory Biomarkers." *BioMed Research International* 2014 (June): 932757.

- Stephens, Greg J., Bethany Johnson-Kerner, William Bialek, and William S. Ryu. 2008. "Dimensionality and Dynamics in the Behavior of *C. Elegans*." *PLoS Computational Biology* 4 (4): e1000028.
- Straw, Andrew D., Kristin Branson, Titus R. Neumann, and Michael H. Dickinson. 2011. "Multi-Camera Real-Time Three-Dimensional Tracking of Multiple Flying Animals." *Journal of the Royal Society, Interface / the Royal Society* 8 (56): 395–409.
- Streit, Marc, and Nils Gehlenborg. 2015. "Points of View: Temporal Data." *Nature Methods* 12 (2). Nature Research: 97–97.
- Strogatz, S. H. 2013. *The Mathematical Structure of the Human Sleep-Wake Cycle*. Lecture Notes in Biomathematics. Springer Berlin Heidelberg.
- Sugi, Takuma, Yukuo Nishida, and Ikue Mori. 2011. "Regulation of Behavioral Plasticity by Systemic Temperature Signaling in *Caenorhabditis Elegans*." *Nature Neuroscience* 14 (8): 984–92.
- Suh, Greg S. B., Allan M. Wong, Anne C. Hergarden, Jing W. Wang, Anne F. Simon, Seymour Benzer, Richard Axel, and David J. Anderson. 2004. "A Single Population of Olfactory Sensory Neurons Mediates an Innate Avoidance Behaviour in *Drosophila*." *Nature* 431 (7010): 854–59.
- Swierczek, Nicholas A., Andrew C. Giles, Catharine H. Rankin, and Rex A. Kerr. 2011a. "High-Throughput Behavioral Analysis in *C. Elegans*." *Nature Methods* 8 (7): 592–98. 2011b. "High-Throughput Behavioral Analysis in *C. Elegans*." *Nature Methods* 8 (7). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 592–98.
- Tang, Xiangdong, Stuart M. Orchard, and Larry D. Sanford. 2002. "Home Cage Activity and Behavioral Performance in Inbred and Hybrid Mice." *Behavioural Brain Research* 136 (2): 555–69.
- Tecott, Laurence H., and Eric J. Nestler. 2004. "Neurobehavioral Assessment in the Information Age." *Nature Neuroscience* 7 (5): 462–66.
- Tinbergen, N. 1951. "The Study of Instinct." Clarendon Press/Oxford University Press. <http://psycnet.apa.org/psycinfo/2004-16480-000>.
- Uchida, Naoshige, and Zachary F. Mainen. 2003. "Speed and Accuracy of Olfactory Discrimination in the Rat." *Nature Neuroscience* 6 (11): 1224–29.
- Valente, Dan, Ilan Golani, and Partha P. Mitra. 2007. "Analysis of the Trajectory of *Drosophila Melanogaster* in a Circular Open Field Arena." *PloS One* 2 (10): e1083.
- Vanderschuren, Louk J. M. J., Raymond J. M. Niesink, and Jan M. Van Pee. 1997. "The Neurobiology of Social Play Behavior in Rats." *Neuroscience and Biobehavioral Reviews* 21 (3): 309–26.
- Van de Weerd, H. A., R. J. A. Bulthuis, A. F. Bergman, F. Schlingmann, J. Tolboom, P. L. P. Van Loo, R. Remie, V. Baumans, and L. F. M. Van Zutphen. 2001. "Validation of a New System for the Automatic Registration of Behaviour in Mice and Rats." *Behavioural Processes* 53 (1-2): 11–20.
- Vannoni, Elisabetta, Vootele Voikar, Giovanni Colacicco, María Alvarez Sánchez, Hans-Peter Lipp, and David P. Wolfer. 2014. "Spontaneous Behavior in the Social Homecage Discriminates Strains, Lesions and Mutations in Mice." *Journal of Neuroscience Methods* 234 (0): 26–37.
- Vitaterna, M. H., D. P. King, and A. M. Chang. 1994. "Mutagenesis and Mapping of a Mouse Gene, Clock, Essential for Circadian Behavior." *Science (New York, ncbi.nlm.nih.gov)*. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3839659/>.
- Von Frisch, Karl. 1955. *The Dancing Bees: An Account of the Life and Senses of the Honey Bee*. Harcourt, Brace.

- Vorhees, Charles V., and Michael T. Williams. 2006. "Morris Water Maze: Procedures for Assessing Spatial and Related Forms of Learning and Memory." *Nature Protocols* 1 (2): 848–58.
- Wabitsch, M. 2000. "Overweight and Obesity in European Children: Definition and Diagnostic Procedures, Risk Factors and Consequences for Later Health Outcome." *European Journal of Pediatrics* 159 Suppl 1 (September): S8–13.
- Wahlsten, Douglas, Nathan R. Rustay, Pamela Metten, and John C. Crabbe. 2003. "In Search of a Better Mouse Test." *Trends in Neurosciences* 26 (3): 132–36.
- Wallace, R. A., G. P. Sanders, and R. J. Ferl. 1991. *Biology, the Science of Life. 3rd Edn.* Edited by Harper Collins and New York.
- Warby, Simon C., Sabrina L. Wendt, Peter Welinder, Emil G. S. Munk, Oscar Carrillo, Helge B. D. Sorensen, Poul Jennum, Paul E. Peppard, Pietro Perona, and Emmanuel Mignot. 2014. "Sleep-Spindle Detection: Crowdsourcing and Evaluating Performance of Experts, Non-Experts and Automated Methods." *Nature Methods* 11 (4): 385–92.
- Watson, J. D., and F. H. Crick. 1953. "Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid." *Nature* 171 (4356): 737–38.
- Whishaw, I. Q., and L. E. Jarrard. 1996. "Evidence for Extrahippocampal Involvement in Place Learning and Hippocampal Involvement in Path Integration." *Hippocampus* 6 (5): 513–24.
- Wilkinson, L. S., S. S. Killcross, T. Humby, F. S. Hall, M. A. Geyer, and T. W. Robbins. 1994. "Social Isolation in the Rat Produces Developmentally Specific Deficits in Prepulse Inhibition of the Acoustic Startle Response without Disrupting Latent Inhibition." *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 10 (1): 61–72.
- Wiltschko, Alexander B., Matthew J. Johnson, Giuliano Iurilli, Ralph E. Peterson, Jesse M. Katon, Stan L. Pashkovski, Victoria E. Abaira, Ryan P. Adams, and Sandeep Robert Datta. 2015. "Mapping Sub-Second Structure in Mouse Behavior." *Neuron* 88 (6): 1121–35.
- Winter, York, and Andrea T. U. Schaefers. 2011. "A Sorting System with Automated Gates Permits Individual Operant Experiments with Mice from a Social Home Cage." *Journal of Neuroscience Methods* 196 (2): 276–80.
- Wolfer, David P., and Hans-Peter Lipp. 2000. "Dissecting the Behaviour of Transgenic Mice: Is It the Mutation, the Genetic Background, or the Environment?" *Experimental Physiology* 85 (6). Blackwell Science Ltd: 627–34.
- Würbel, H. 2001. "Ideal Homes? Housing Effects on Rodent Brain and Behaviour." *Trends in Neurosciences* 24 (4): 207–11.
- Yamamoto, A., J. J. Lucas, and R. Hen. 2000. "Reversal of Neuropathology and Motor Dysfunction in a Conditional Model of Huntington's Disease." *Cell* 101 (1): 57–66.
- Yemini, Eviatar, Tadas Jucikas, Laura J. Grundy, André E. X. Brown, and William R. Schafer. 2013. "A Database of Caenorhabditis Elegans Behavioral Phenotypes." *Nature Methods* 10 (9): 877–79.
- Zorrilla, Eric P., Koki Inoue, Eva M. Fekete, Antoine Tabarin, Glenn R. Valdez, and George F. Koob. 2005. "Measuring Meals: Structure of Prandial Food and Water Intake of Rats." *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 288 (6): R1450–67.
- Zwaigenbaum, Lonnie, Susan Bryson, Tracey Rogers, Wendy Roberts, Jessica Brian, and Peter Szatmari. 2005. "Behavioral Manifestations of Autism in the First Year of Life." *International Journal of Developmental Neuroscience: The Official Journal of the International Society for Developmental Neuroscience* 23 (2-3): 143–52.

