



Advancing social behavioral neuroscience by integrating ethology and comparative psychology methods through machine learning

Joeri Bordes^a, Lucas Miranda^{b,c}, Bertram Müller-Myhsok^b, Mathias V. Schmidt^{a,*}

^a Research Group Neurobiology of Stress Resilience, Max Planck Institute of Psychiatry, 80804 Munich, Germany

^b Research Group Statistical Genetics, Max Planck Institute of Psychiatry, 80804 Munich, Germany

^c International Max Planck Research School for Translational Psychiatry (IMPRS-TP), 80804 Munich, Germany

ARTICLE INFO

Keywords:

Social behavior
Psychiatric disorders
Ethology
Comparative psychology
Stress

ABSTRACT

Social behavior is naturally occurring in vertebrate species, which holds a strong evolutionary component and is crucial for the normal development and survival of individuals throughout life. Behavioral neuroscience has seen different influential methods for social behavioral phenotyping. The ethological research approach has extensively investigated social behavior in natural habitats, while the comparative psychology approach was developed utilizing standardized and univariate social behavioral tests. The development of advanced and precise tracking tools, together with post-tracking analysis packages, has recently enabled a novel behavioral phenotyping method, that includes the strengths of both approaches. The implementation of such methods will be beneficial for fundamental social behavioral research but will also enable an increased understanding of the influences of many different factors that can influence social behavior, such as stress exposure. Furthermore, future research will increase the number of data modalities, such as sensory, physiological, and neuronal activity data, and will thereby significantly enhance our understanding of the biological basis of social behavior and guide intervention strategies for behavioral abnormalities in psychiatric disorders.

1. Introduction

Understanding behavior in humans and animals is essential, as it forms a central aspect of our existence and interaction with the environment (Heimlich and Ardoin, 2008; Bolhuis et al., 2021). At the same time, behavioral abnormalities are at the core of severe and highly prevalent psychiatric disorders. Specifically, the impairment of social functioning is an important symptom of many different psychiatric disorders, such as post-traumatic stress disorder (PTSD), depression, autism, and schizophrenia (Peleh et al., 2019a; American Psychiatric Association, 2013; Nietlisbach and Maercker, 2009; Dodell-Feder et al., 2015; Katz et al., 2011). Therefore, the accurate measurement of different social behaviors is crucial for a better understanding of psychiatric patients and contributes to solving the complex mechanisms of those disorders. Preclinical rodent models are widely used to investigate the underlying mechanisms of complex social behaviors (Peleh et al., 2019a; Beery and Kaufer, 2015), as well as to evaluate the effectiveness of novel drug interventions (Pöhlmann et al., 2018; Lopez et al., 2022). However, the assessment of social behavior as a holistic construct is challenging as it relies on many different types of social behaviors and is

influenced by different factors, such as stress exposure (Beery and Kaufer, 2015). Therefore, a variety of behavioral tasks have been employed to understand different mechanisms of the social behavioral construct in rodents (Toth and Neumann, 2013; Fan et al., 2019; Zhou et al., 1979; Golden et al., 2011).

1.1. The history of animal behavior

Exposure to external stimuli triggers a distinct pattern of cellular responses in the brain, which ultimately drives the behavioral response. Fundamental concepts of behavior have historically been established using observational research, where animals are left undisturbed and observed in their natural habitat. One of the earliest and most influential animal behavioral research was performed by Charles Darwin in the 1800s (Thierry, 2010), where he observed variations between animals and concluded that species descended from each other (Darwin, 1859). This type of behavioral research led to the development of the field of ethology, which concentrated on the evolutionary basis of animal behavior while using predominantly observational approaches. An overarching principle in ethology is that behavior is described in its

* Correspondence to: Research Group Neurobiology of Stress Resilience, Max Planck Institute of Psychiatry, Kraepelinstr. 2-10, 80804 Munich, Germany.
E-mail address: mschmidt@psych.mpg.de (M.V. Schmidt).

highest complexity by observation in natural or semi-natural environments. This allows the researcher to study behavior in a descriptive way, using a hypothesis-generating approach to discover novel behavioral concepts. Important examples of outstanding work in ethology include the research of Tinbergen and colleagues, who observed that young birds can distinguish the shape of a hawk and a goose and change their behavior according to the perceived shape (Tinbergen, 1939). In addition, the work of Lorenz and colleagues observed an interesting phenomenon in young domestic chicks, where young animals narrow their social preference towards an object during early exposure, which typically is their parent (Bolhuis and Giraldeau, 2005; Sullo way, 1982). The field of ethology and observational research has led to phenomenal research findings, but there are also important limitations in the study design. Observational studies rely on the researcher's ability to assess behavior, which can lead to misinterpretation and consequently differences between researchers in the interpretation of the observed behaviors. In addition, the lack of control over the environment limits the throughput and the research question that can be assessed, which can lead to unreproducible results due to high variability between the experimental conditions.

In response to these limitations, the field of comparative psychology was introduced, in which researchers aimed to study and understand behavior using tightly controlled environmental settings. Here, the focus was on breaking the behavioral construct down into clearly identifiable and quantifiable syllables by reducing the complexity of the behavior and limiting the dimensions that behavior can be expressed in through precise laboratory settings. These laboratory tasks are characterized by a high controllability over the environment and a standardized set of behavioral read-outs and therefore require hypothesis-driven research questions. Several outstanding researchers showed the novelty and strength of this field, including Thorndike, Skinner, and many others. Thorndike demonstrated the construct of trial-and-error learning, as he observed that by increasing the number of trials, animals became more efficient to escape an apparatus and gain their reward (Thorndike, 1898). Subsequently, Skinner developed one of the first and extremely popular behavioral laboratory tasks, the operant conditioning chamber (Skinner, 1948). The task can be used for negative and positive reinforcement learning and is still used in research around the globe to this day forward, underlining its excellence. By this time, a trend to standardize and simplify different behavioral disciplines using laboratory tasks was set in motion (Tolman, 1948; Aggleton, 1985; Barnes, 1979; Morris, 1981), and is currently still ongoing. The excellence of laboratory tasks for behavioral science is undoubtedly great, especially for investigating the influence of external stimuli (e.g., the stress response system) and the role of interventions (e.g., pharmacological) on the behavioral output (Hånell and Marklund, 2014). In addition, the unprecedented options for using different genetic mouse models have enabled the exploration of specific target genes on behavior. Unfortunately, no behavioral tasks are without flaws, and there are specific concerns regarding laboratory tasks that need to be addressed. The laboratory set-up provides an intensive interaction of the researcher with the testing animals, which causes concerns, as inter-individual differences between researchers can influence the outcome of the behavioral performance of the animals (Sorge et al., 2014; Chesler et al., 2002). As an example, a recent study described the influence between male and female experimenters on the largely different behavioral outcomes of drug-treatment efficacy (Georgiou et al., 2022). Another concern is the difficulty of streamlining animal housing and testing. Even though compared to observational research the differences in housing and testing are subtle, they have been found to lead to different behavioral outcomes (Richter et al., 2009), and therefore can give a false sense of reliability. In addition, although group housing of laboratory mice is crucial for their social welfare, it remains a highly unnatural environmental setting and has its limitations. The confined space with several same-sex mice hampers the animals in expressing certain species-typical behaviors, such as the search for food and water, the

vigilance for certain threats, and the motivation for sexual reproduction. In addition, the lack of escaping introduces changes in territorial aggression, as, after an aggressive encounter, the dominant and subordinate animals cannot go their separate paths and will have to stay confronted with each other, which introduces problematic behavior, such as severe aggression (Weber et al., 2017). Finally, the use of inbred mice is a problematic development for investigating naturalistic behaviors. While the genetic models have allowed for important insights into the genome, they have left us with an animal model that behaves rather differently than their conspecifics in the wild, which questions the validity of the animal model and unfortunately has reduced the reproducibility of the behavioral research (Richter, 2017; Wahlsten et al., 2003; Voelkl et al., 2020; Crawley, 1996).

The reductionistic approach in laboratory tasks is a strength for many different behavioral disciplines; however, it can be a pitfall for behavioral constructs that rely on many different behavioral outputs and require more naturalistic environments, which therefore are more complex to assess (Blumstein, 2010). In such cases, laboratory tasks can oversimplify behavior, which can lead to an inadequate or even wrong assessment, that is often lowering the reproducibility rate. A good example of such a complex behavioral construct is social behavior, therefore the next section describes the definition of social behavior, together with the current behavioral tasks that are used to measure social behavior.

1.2. The social behavioral construct

Social behavior is a naturally occurring construct in many different species, including rodents and humans, and holds a strong evolutionary component, as it is critical for the survival of the individual as well as for the species (Wei et al., 2021; Chen and Hong, 2018). These behaviors are innate but are strongly dependent on the environmental circumstances and will adapt the social behavioral output appropriately. For example, adolescent mice will indulge in social play interactions with their conspecifics, but if they are socially isolated during this time, their social behavioral repertoire and the underlying neurobiological mechanisms will be altered during adult age (Musardo et al., 2022). Social behavior is a complex construct and is used as an umbrella term for a plethora of different behaviors, which makes its precise definition challenging. In general, social behavior has been defined as all behaviors that influence or are influenced by other members of the same species. This terminology includes sexual and reproductive activities, all behaviors that bring individuals together, as well as aggressive and dominant behaviors (Chen and Hong, 2018; Whishaw et al., 2006; Mackintosh and Grant, 1963). Although not the focus of this review, it is important to note that this definition excludes social interactions across different species, or objects, which are part of the social behavioral repertoire, even though they might be the minority of the social encounters.

2. Social behavioral tasks

Social behavioral events in rodents typically begin with the detection of a social encounter. The animal can decide to ignore or avoid the social encounter, or to consummate the social behavioral encounter, which relies on one or a combination of different social behaviors, such as parenting, fighting, sexual interaction, following, feeding, playing, and guarding (Wei et al., 2021). These constructs are influenced by other neuronal modalities, including motivation (Bariselli et al., 2018; Solié et al., 2022a), emotion (Sakaguchi et al., 2018; Pisansky et al., 2017; Keum and Shin, 2019; Jabarin et al., 2022), and decision-making (Scheggia et al., 2022), which means that the underlying social behavioral construct is complex to assess, as it relies on many different behavioral read-outs. Currently, the assessment of social behavior relies on using a variety of behavioral tasks measuring different social domains, see Box 1. The majority of the tasks have been developed with the goal of increasing the throughput and the environmental controllability,

Box 1

The different social behavioral domains and their behavioral tasks.

The assessment of social behavior relies on a variety of behavioral tasks that can measure different social domains, including dominance, avoidance, approach memory, transmission, decision-making, and conditioning. Dominance behavior is generally assessed using competition tasks, such as the tube test (Fan et al., 2019; Lindzey et al., 1966) and the warm spot test (Zhou et al., 1979), or via long-term home cage observation (Shemesh et al., 2013; Forkosh et al., 2019). The avoidance and approach behavior can be assessed using tasks that compare the interaction with a non-social stimulus and a social stimulus, using the three-chamber task (Rein et al., 2020) and the social avoidance task (Berton et al., 2006; Lukas et al., 2011), but also with freely-moving multi-animal social interaction tasks (Bordes et al., 2022). In addition, social memory can also be assessed in the three-chamber task and social avoidance task, but then compares a familiar-social stimulus with a novel social stimulus (Gheusi et al., 1994; Winslow, 2003; Engelmann et al., 2011). The transmission of social behavior is a particularly complex domain, which can be subdivided into many different subfields. One example is the social transmission of food preference, in which a naïve mouse follows a demonstrator mouse who has previous knowledge of a food reward (Wrenn, 2004). In the context of stress, a prominent domain is social fear transmission, in which animals receive shocks when interacting with a social conspecific (Toth et al., 2013), or observe other animals getting foot shocks (Jeon and Shin, 2011). The fields of decision-making and conditioning have recently been further adapted to the social behavioral research field and interestingly behavioral tasks have been adapted to specifically measure these domains in a social context, using the two-choice social decision-making task (Scheggia et al., 2022), and the social instrumental task (Solié et al., 2022b).

but therefore rely on a reductionistic approach using single or few behavioral read-outs (univariate behavioral tests), a hallmark of the comparative psychology field. The next section describes the social behavioral domains of social dominance and social avoidance. Although many more variants of social behavior and corresponding tasks exist, these social behavioral domains in particular are of interest, as they are well-established behavioral tasks in the field of stress neuroscience and they showcase the development of the reductionistic approach in social behavioral tasks.

3. Social dominance

An important domain of social behavior is the social hierarchy, which is important for a balanced and structured group of multiple animals from the same species (Dwartz et al., 2022; Qu et al., 2017). Social status impacts the chances of survival and, among other aspects, the individual's reproductive chances, health, and food availability (Tamashiro et al., 2005; Ellis, 1995; Sapolsky, 2005). The social status is conveyed via dominant, aggressive behaviors, and the subsequent submissive, defensive behaviors, which are adaptive behaviors to reduce aggression in established hierarchies. The dominant animal is commonly identified by consistent wins in social conflicts with the subordinate animals (Fan et al., 2019; Dwartz et al., 2022; Wang et al., 2014). Several behavioral tasks have been developed to investigate the hierarchical structure in rodents and the most popular tests include the tube test (Fan et al., 2019; Lindzey et al., 1966) and the warm spot test (Zhou et al., 1979), but also via long-term home cage observations (Shemesh et al., 2013; Forkosh et al., 2019). Other hierarchy tests rely on the competition for food or water, but those tests have decreased in popularity, as they are influenced by the motivation for food and water and can fluctuate based on saturation (Merlot et al., 2004; Cordero and Sandi, 2007; Ujita et al., 2018). The tube test has gained popularity, due to its simplicity and an available standardized protocol by Fan et al., (Fan et al., 2019). The tube test utilizes a narrow transparent tube, in which mice are placed on opposite ends. Upon release, the mice need to go through the tube towards the other side, but can only succeed by forcing the conspecific animal to go backward. The subordinate mouse is identified as the animal that has to go backward, whereas the dominant mouse is the animal that successfully pushes the conspecific away. The tube test relies on the assumption that after several training days, the animals prefer to go to the other side of the tube and use dominant-related mechanisms to establish, which of the animals back out. The animals are tested over several consecutive days, after which a stable and consistent hierarchy can be observed (Fan et al., 2019). However, since the animals are confined within the tube, their agility will influence their winning chances (Fan et al., 2019; Zhou et al., 2018),

which can be different due to many factors, such as age, strain, stress levels, weight, etc. Moreover, the tube test can only test two mice at the same time, which limits the transferability towards a more naturalistic environment, where animals interact and are challenged within group dynamics. A recently developed hierarchical test in mice that overcame this problem is the warm spot test, which works reliably with up to four mice. In this task, the animals are placed on a cold floor with a warm spot in the corner, which is only big enough to warm one animal at a time (Zhou et al., 1979). The test relies on the behavioral construct that mice prefer the warm spot over the cold area and will determine the time spent at the warm spot based on their rank in the hierarchy. The hierarchy state of animals is observed during a 20-minute test, in which the amount of time spent in the warm spot should overlap with the hierarchical position of the animals. Therefore, the dominant animal spends the most time in the warm spot, whereas with every lower ranking, each animal spends less time in the warm spot (Zhou et al., 1979). Most importantly, the tube test and warm spot test show a similar pattern of hierarchical structure (Zhou et al., 1979) and are highly correlating with other markers for dominance, including, among others, agonistic-related behaviors, urine markings, and ultrasonic vocalizations (Zhou et al., 2018; Wang et al., 2011). These tests have contributed significantly to our understanding of hierarchy in mice, are well appreciated for their simplicity and robustness, and have been widely adopted for stress-related research (LeClair et al., 2021; Park et al., 2018; Larrieu et al., 2017; Matthews et al., 2016). However, these tests rely on an instrumental laboratory environment that is particularly unnaturalistic. In the tube test animals are forced within a confined space towards confrontation, and in the warm spot test, the animals are forced into confrontation as there is only one place that is pleasantly warm enough. Therefore, these tests exclude the natural influences of motivation on hierarchy establishment. Several studies have shown individual differences in the motivation to pursue social interaction with conspecifics (Bariselli et al., 2018; Torquet et al., 2018), which includes aggressive confrontation between animals. Thus, these tests reflect the reductionist approach, as they simplify the behavioral read-out.

4. Social avoidance behavior

Clinical research has widely recognized that social withdrawal and avoidance are key symptoms of depression (Ottenbreit et al., 2014; Fernández-Theoduloz et al., 2019) and anxiety disorders (Rinck et al., 2010; Heuer et al., 2007), which is why understanding the underlying mechanisms related to social avoidance is crucial. The domain of social behavior containing social avoidance is an often-used paradigm for investigating the stress response system and in particular, has been well-established to assess stress resiliency versus susceptibility of

individual rodents (Krishnan et al., 2007; Murra et al., 2022; Donahue et al., 2014), but is also used in many other scientific fields to investigate the affectedness of social behavior, for instance in autism (Rein et al., 2021; Rapanelli et al., 2021; Bidinosti et al., 2016; Qin et al., 2018). Therefore, several behavioral tasks have been developed to measure social avoidance behavior (Toth and Neumann, 2013), for example, well-established tasks include the three-chamber task (Rein et al., 2020) and the social avoidance task (Berton et al., 2006; Lukas et al., 2011). These tests have been described in detail in other studies (Rein et al., 2020; Berton et al., 2006; Lukas et al., 2011; Zimprich et al., 2017). Both tests rely on the exploratory behavior of the test mouse towards a novel social stimulus. The social stimulus is an unknown social conspecific, who is constrained in a small grid box or beaker and the interaction time of the test subject is measured in relation to the interaction of an empty box directly before the social interaction test (social avoidance test) or at the same time in a separate chamber of the apparatus (three-chamber test). The gold standard for the calculation of social avoidance behavior relies on the interaction ratio (IR) (Krishnan et al., 2007; Donahue et al., 2014). The IR is calculated by taking the total amount of time spent with the social stimulus (T_S) and dividing it by the total amount of time spent with the empty cage (T_E) resulting in the following formula: $IR = T_S / T_E$ (this score can also be expressed in a percentage which would mean the end result is multiplied by “100”). The expectation is that unaffected social behavior in mice results in an IR higher than “1”, which means that animals spend more time investigating the social stimulus compared to the empty cage. In the same line of thought, animals that have an IR lower than “1” show an avoidance of the social stimulus and therefore have affected social behavior. These tasks allow for a quick and simple analysis of social avoidance behavior and enable high controllability over the environment. In the last decade, they have gained popularity and have been extensively used for observing social avoidance-related behaviors specifically in stress research, where they have shown their value (Murra et al., 2022; Reguilón et al., 2021; Durand-de Cuttoli et al., 2022; Morel et al., 2022; Li et al., 2022). However, these tests also rely on a reductionistic and instrumental environment, as the social stimulus is trapped in a confined space, which increases the risk for cross-over effects by other types of behaviors, such as motivation, emotion, and anxiety. Moreover, the robustness of these tasks has been questioned, due to a lack of standardized experimental set-ups and subsequently high variation in the behavior of individual animals (Pearson et al., 2010). A critical problem that has been overlooked until now is the fact that there is a proportion of nonstressed animals that are also showing social avoidance behavior towards the social stimulus (e.g., 12% (Krishnan et al., 2007), 26% (Golden et al., 2011)). This means that there is a subset of stressed animals that are categorized as susceptible ($IR < 1$), even though they showed avoidance behavior already before the stress exposure. Therefore, the assumption that nonstressed animals prefer the novel social stimulus over non-social environments seems to be highly dependent on the individual. Even though these animals are an extremely interesting group to investigate their social behavioral profile, they cannot be categorized in the same group as animals that initially do show social preference but change upon stress exposure. This can be avoided by screening all animals before the stress exposure so that animals that are already avoiding the social stimulus can be investigated as a separate group or excluded.

Measuring constructs of social behavior using a reductionistic approach has its advantages, as it allows for high controllability over the environment, straightforward comparability across cohorts and conditions, high throughput, and simple analysis of the behavioral read-outs. These reductionistic social behavioral tests have provided important insights, especially into the stress-related mechanisms of social behavior, but there are severe limitations that need to be addressed. These tests often suffer from problems related to the comparative psychology, reductionistic research approach, which means that some of these tasks have limited ecological validity, as they largely restrict free social interaction behavior (such as fighting, following, grooming,

anogenital sniffing, etc.) and only measure one or a few simple behavioral read-outs (e.g. the preference of an animal to be in close proximity of a conspecific). Therefore, these tasks tend to oversimplify the social behavioral repertoire and do not accurately capture the full range of the social behavioral construct. In addition, these tasks have been developed specifically for rodents and have limited comparison to human behavioral tasks, which lowers the translatability to the clinic. Therefore, the development of more naturalistic types of behavioral tasks, including free social interaction between multiple animals, will be crucial for a more accurate and ethologically relevant assessment of social behavior. However, up until recently, no software tools were available to analyze and distinguish multi-animal interactions, but the many different behaviors within a free social interaction task are too complicated, time-intensive, and repetitive to assess manually (Hånell and Marklund, 2014; Goodwin et al., 2020; Sturman et al., 2020). The rapid advances in automatically annotated motion tracking opened the possibility to assess many different social behaviors in complex environments while maintaining high throughput. Therefore, the next section will discuss the current status of motion tracking and machine learning for behavioral classifications, and their implications for social behavioral neuroscience research.

4.1. Automatically annotated motion tracking and machine learning tools advance the behavioral analysis

When looking back at the history of the development of behavioral assessment, there were several influential methods using both observational research approaches in natural environments as well as the reductionistic laboratory research approach. Both methods have their inherent advantages and disadvantages, but by combining the strengths of both methods, it is possible to overcome some of the current problems with (social) behavioral assessment (Fig. 1). This can be achieved by using a semi-naturalistic environment, in which animals can freely interact with each other, but are still limited in their environment and space, to maintain a certain level of control over the environment. Several researchers have contributed to the development of such tasks (Shemesh et al., 2013; de Chaumont et al., 2019), which contain automatic phenotyping systems using top-view camera detection systems, in which up to four mice can live and freely interact for many days without experimenter intervention. Shemesh et al., use different fur color dyes to distinguish the animals, which therefore allows video recordings only with white furred animals (e.g. CD1, BALB/c), as with darker fur (e.g. C57/Bl6) the tracking system cannot distinguish well enough (Shemesh et al., 2013). This is an important drawback as many studies are conducted with C57/Bl6 animals. This problem was solved by combining video recordings with RFID probe identification of the animals via detection antennas under the arena, which therefore does not require different colorings for the animals (de Chaumont et al., 2019; Peleh et al., 2019b). However, the RFID probes are expensive, cannot be reused, and have problems when animals are in close proximity together, making the technique difficult to implement in many different labs. In addition, for both methods, the animals need to be sedated to color dye or implant the RFID probe, which can be seen as a stressor and is therefore not an optimal solution for investigating, e.g. consequences of stress exposure. The tracking and analysis in these systems remain a difficult process and rely on homemade tailored software that is likely too complicated to integrate into many different labs. Also, the tracking of the animals consists of only the center-point of the animals and at best includes the tail base and nose. This limits the information and accuracy that can be obtained about the animal's behavioral profile, especially looking at more complex social behaviors (such as chasing, mounting, and attacks), but also individualistic behaviors (such as grooming, huddling, digging, and sniffing).

In 2018, Mathis et al., launched the open-source software system, DeepLabCut (DLC) (Mathis et al., 2018), which allowed easy access for researchers to track animals with high throughput for any given

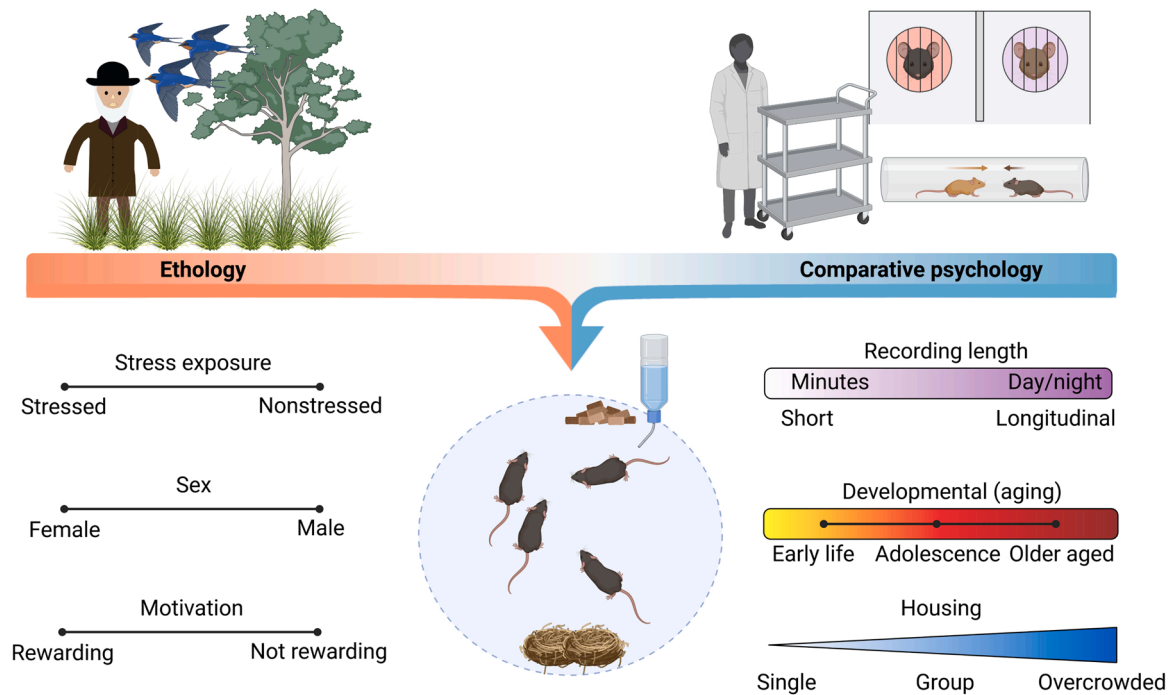


Fig. 1. Next-generation of social behavioral tests. In the top part of the figure the important methods for investigating behavioral neuroscience are illustrated, with ethology on the left and comparative psychology on the right. Next-generation social behavioral tests (illustrated by the arrow in the middle) rely on a semi-naturalistic environment, which includes the strengths of both ethology, in which a naturalistic set-up without the interference of the experimenter is implemented, as well as comparative psychology, whereas certain controllability of the environment remains by limiting the space and influences from the outside world. Different factors, such as stress exposure, sex, motivation, recording length, developmental (aging), and housing conditions, will influence the outcome of the social behavioral assessment.

behavioral video recording. DLC utilizes deep neural networks that enable quick and easy markerless pose estimation and only need little training data to achieve human-level accuracy for object recognition (Mathis et al., 2018). Up until then, the gold standard for tracking animals relied on commercial software systems that could only track the center point of single animal video recordings. The analysis of social interaction data between multiple animals remained to be assessed manually, but as social behavior relies on many different behavioral interactions, this limited the complexity of the behaviors that could be assessed (Sturman et al., 2020; Bordes et al., 2022). DLC enables the tracking of any body part of interest across multiple animals, without the need for additional invasive marking methods (such as RFID tags or animal dye coloring), as the neural networks can also separate identically looking animals (Lauer et al., 2022). Importantly, to gain accurate multi-animal tracking, the DLC models for tracking multiple animals load entire video fragments onto the graphics processing unit (GPU) during training, as opposed to single frames in the single-animal models, which requires substantially more advanced hardware. Since then, other open-source markerless pose estimation software systems have become available, such as DeepPoseKit (Graving et al., 2019), Social Estimates Animal Poses (SLEAP) (Pereira et al., 2019, 2022), and SIPEC (Marks et al., 2022). These different software systems acquire pose estimation in slightly different ways and therefore vary in their accuracy, training speed, and amount of required training data (Mathis and Mathis, 2020). Ultimately, however, they all are open-source tools that allow fast and easy tracking of any body part of interest in single or multiple animal recordings. Their offered features and performance are to date comparable, and choice often relies on community support and downstream analysis compatibility. Importantly, DLC has an extensive user base and has shown a strong dedication to supporting, maintaining, and extending its software ecosystem with new downstream analysis tools (Mathis, 2020; Schneider et al., 2023). These pose-estimation tools have enabled the social behavioral research field to be able to investigate more

naturalistic types of behaviors, as the tracking of many freely moving animals can easily be performed. However, this provides a novel problem, as pose-estimation tools increase the amount of data tremendously. The interpretation of the pose-estimation output, the “X and Y” coordinates over time, needs to be translated into meaningful behaviors. The most basic analyses include tracking the center point of the animals as described before, but in a markerless way, which contributes to automating the quantification of previously used univariate tests (Sturman et al., 2020). However, tracking multiple body parts enables researchers to explore a plethora of different behaviors based on a more complete capture of posture. Different analysis methods can be employed to analyze the pose-estimation data using supervised and unsupervised analysis toolboxes.

The supervised analysis toolbox can be employed using machine learning models or simple rules to identify specific posture-based behaviors of interest. An important package that uses a minimal amount of coding to create supervised behaviors from DLC and other pose-estimation packages, is SimBA (Simple Behavioral Analysis (Nilsson et al., 2020)). SimBA allows practitioners to label behaviors of interest in a graphical user interface to train machine learning models that can learn the rules governing these patterns from data, automating the quantification of arbitrarily complex traits. The provided models are based on extracted static and dynamic features describing animal motion, instead of on the sequences themselves. This makes the models easier to train and does not require a dedicated GPU, but the transferability toward other behavioral setups is rather limited. However, some behaviors do not need machine learning models as they can accurately be deduced via hardcoded rules (Bordes et al., 2022). These can include, but are not limited to, time-in-zone quantifications, certain interactions between individuals (for example, nose to nose and nose to tail), or interactions with objects. Moreover, certain packages, such as SIPEC (Marks et al., 2022) or MARS (Segalin et al., 2021), combine a tracking system with a pipeline for supervised behavioral annotation,

which can show advantages over combining different packages to the same end, since users do not need to worry about software compatibility as everything is contained within the same framework. The current machine learning models that are used in many of the different supervised analysis toolboxes, tend to overfit the nuances present in the dataset they were trained on, which limits the transferability to a new data set. Transferability is an important way to standardize the annotated behavior of interest and allows a minimal amount of work for other labs to implement the tagging of previously annotated behaviors. An important future goal will therefore be to find models that can extend their detection with greater accuracy to completely novel data sets.

Another method to examine behavior is through studying pose-estimation data without prior categorization. This can be performed using unsupervised learning, which is a branch of machine learning that aims to extract insights from the data without using information about the behaviors of interest *a priori*. This method segments animal trajectories across time to recover consistent behavioral patterns, which can coincide with the patterns that were analyzed in a supervised manner. In addition, this method allows for a hypothesis-generating approach, as it can identify behavioral patterns that indicate novel behaviors in the specific behavioral context. Since unsupervised approaches allow to explore the behavioral space without the need for time-intensive labeling, they can be used as an initial screening for behaviors of interest. For example, unsupervised analysis can be used to identify specific behavioral patterns that are showing the biggest difference between a set of defined experimental conditions. As a next step, researchers can then train supervised classifiers to measure the behavior of interest more directly and with less noise, or to initialize classifiers that are further

fine-tuned with human feedback, using an active learning approach (Yttri et al., 2023). Several packages and pipelines have come out that use clever ways to segment behavior in an unsupervised way. The software system B-SOID (Hsu and Yttri, 2021), for example, relies on annotating the motion data with sets of features that can help describe behavior across time without using sequential data directly. Another software system, MoSeq (Wiltchko et al., 2020), takes advantage of the time component of motion using autoregressive hidden Markov models, which can directly capture probabilistic relationships between the input variables. While originally devised for depth sensing camera setups, recent iterations adapt it to pose estimation settings too (Weinreb et al., 2023). Other packages, such as VAME (Luxem et al., 2022) and DeepOF (Bordes et al., 2022) use neural networks that can process the motion sequences directly, which was shown to provide more meaningful clusters of specific behaviors with less noise of other behavioral patterns (Luxem et al., 2022). Furthermore, an advantage of neural network models is that, in parallel to sequence segmentation, they allow researchers to embed motion trajectories into interpretable latent spaces that can be analyzed, for example, differentially across experimental conditions. Moreover, the explainability of the retrieved patterns is key to understanding the underlying behaviors (Goodwin et al., 2022). Being able to interpret how and what the models are doing can help improve transferability across datasets on the one hand, but also categorize unsupervised clusters on the other hand, and this can be done both via visual exploration (taking advantage of video data and going back to the snippets that are assigned to specific patterns) or using machine learning explainability tools, such as Shapley Additive Explanations (SHAP) (Goodwin et al., 2022). A comprehensive overview of

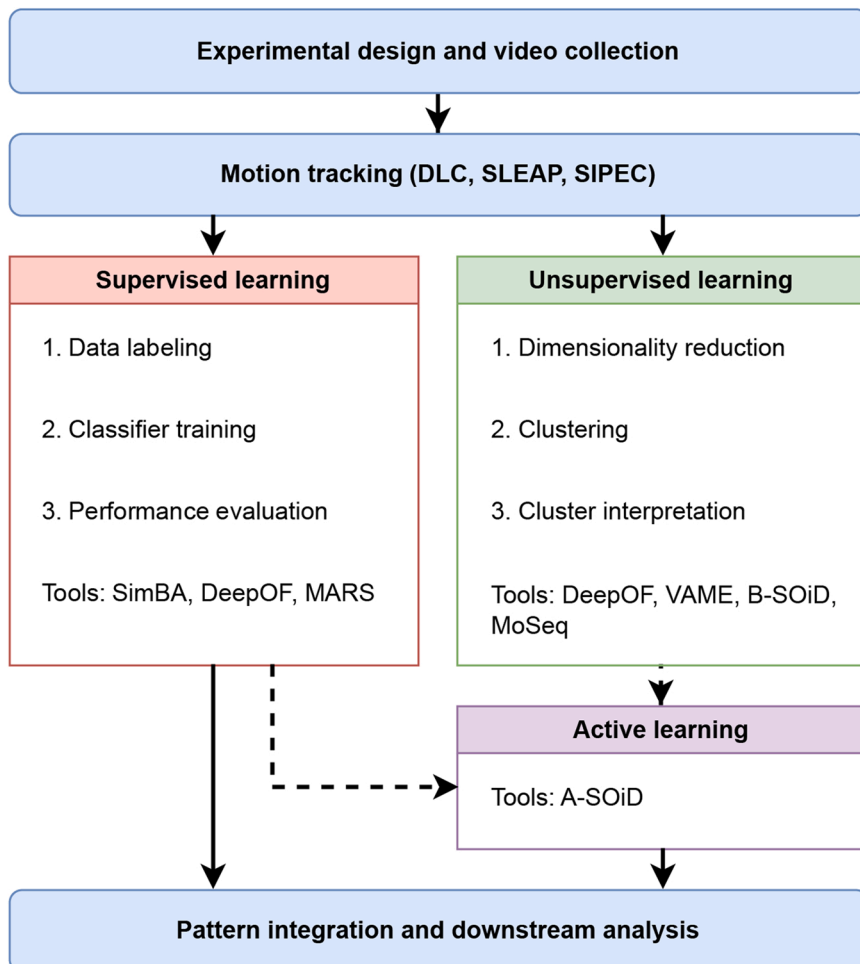


Fig. 2. Automatic behavioral annotation via motion tracking. After the experimental design is defined and videos are collected, key points over time for one or more animals are extracted as time series, using pose-estimation tools, such as DLC, SLEAP, or SIPEC. Predefined behaviors can then be extracted using supervised learning tools such as SimBA or MARS, which typically require data labeling, followed by classifier training and performance evaluation of the extracted behaviors. Other tools, such as DeepOF, provide pre-trained models that bypass these steps. Another way of extracting broader information is unsupervised learning, which does not require labeling, and aims to obtain behavioral syllables (or clusters) after a dimensionality reduction step. Cluster interpretation is a key step, which can be achieved by visual video inspection or using model explainability tools such as SHAP. Tools in this realm include DeepOF, VAME, B-SOid, and MoSeq. Moreover, unsupervised learning results can be used to initialize supervised models with active human feedback, such as seen on the A-SOid framework. Finally, the expression and dynamics of all retrieved patterns can be compared across experimental conditions, for example, to gain insights on behavioral shifts.

this pipeline is presented in Fig. 2.

In contrast to the aforementioned open-source software packages, it is worth mentioning that there are also commercial solutions that have been employed for animal behavioral analysis, such as Smartcube (Lorsch et al., 2021). The promise of delivering an out-of-the-box, easy and high-quality product for a price is valid, however, there are intrinsic disadvantages to using proprietary black-box solutions. Commercial software systems are, for understandable reasons, not subjected to the same rules as scientific software, which means that they are not peer-reviewed and the information about which and how the employed models are working is not openly available. A potential way for bypassing those problems is to visually validate the produced results, which may in turn require more manual labeling. In the end, the fast solution-based development that the open-source community has shown in recent years is hard to compete against. An example of this fast development is the several publicly funded open-source competitions that were released in recent years, with the aim of improving machine-learning algorithms for behavior (Sun et al., 2021). This has the potential to continuously improve the current state of the art and yield a constant stream of open-source tools of excellent quality for behavioral research. Moreover, all data produced by labs while using this type of software package are collected and recycled to further improve previous algorithms, creating a constructive feedback loop from which the whole community can benefit.

Although the field of automatic animal tracking and behavioral annotation has progressed unprecedentedly in the last few years, there are still several issues to overcome (von Ziegler et al., 2021). The increased number of open-source packages to analyze behavioral data is a crucial development that creates novel and better models, but many packages offer similar and overlapping results that are relying on different software systems and are therefore often not compatible with each other. This problem limits users to find the best set of software systems that are needed to answer their research question, especially for researchers from non-computational labs. An important advancement by the DLC development team is the future implementation of pre-trained tracking models that do not require labeling of any data and can export a flexible set of annotated body parts, which means that if competing software packages require a different set of labeled body parts, this would easily be extracted using the DLC Model Zoo Super-Animal software system (Ye et al., 2022). The current fast trajectory of many newly developed software solutions for behavioral analysis is encouraging, but in order for a wide audience to implement these tools, it would be beneficial if the field will converge to a centralized solution that ensures a compatible ecosystem for tracking, annotation, and the embedding of multimodal behavioral data, to which researchers would be able to openly contribute.

The development of more precise tracking tools together with the numerous post-tracking analysis packages have allowed the behavioral research field, and in particular social behavioral research, to rethink the possibilities for the assessment of behavior. The implementation of such tasks will be beneficial for many different neuroscientific behavioral fields including fundamental behavioral research, especially studies focused on the consequences of stress exposure. An increasing number of studies are using these tools to investigate how stress affects social behavior. A selection of those studies is highlighted in the next section to illustrate the possibilities and strengths of such tasks using novel deep-learning tracking tools.

4.2. Investigating social behavior using advanced tracking tools for better stress phenotyping

Recently, several studies have used advanced tracking tools to investigate the social behavioral profile and hierarchy establishment in a semi-naturalistic environment to push our understanding of behavioral phenotyping related to stress exposure (Shemesh and Chen, 2023). Forkosh et al., (Forkosh et al., 2019) used an automated behavioral

monitoring system, called “Social Box” (SB), to investigate hierarchy rankings in mice utilizing a more naturalistic environment. This study uses homemade software to identify pose estimation and subsequently utilizes supervised behavioral analysis tools to calculate a set of social behaviors. The most important classifier for this study is the calculation of aggressive chases that is used, together with some other behaviors, to obtain David’s scores that reflect the hierarchy rankings between mice. The SB system can house up to four mice and can easily be utilized for longitudinal day and night recordings. Interestingly, in a follow-up study, the authors explore how social hierarchy rank can be used as a predictor of chronic stress exposure-outcome in a sex-dependent way, emphasizing the importance of investigating both sexes in the stress-related research field (Karamihalev et al., 2020). The data from the SB system exemplifies the enriched data output using longitudinal observations in a semi-naturalistic environment. This is extremely relevant as it allows for a more natural inference of motivation and decision-making processes crucial for hierarchy rankings. It would be interesting to see how the hierarchy rankings in the SB system compare to the classical approaches, especially in cohorts of animals that show altered motivation or decision-making. Another next step would be to integrate additional behavioral domains via implementing behavioral tasks during the longitudinal overnight recordings, which would be able to abolish the influences of initial arousal, sleep disturbances, and anxiety on the performance of the behavior. An example is the implementation of home cage operant conditioning devices, such as FED3, which allow the investigation of food intake, motivation, and reward processing (Nguyen et al., 2016). In order to boost the implementation of such a system for many different labs, the SB system should implement open-source pose-estimation and classifying software so that the identification of hierarchy in home-cage-like systems would easily be transferable.

A study that implemented a fully open-source software pipeline for stress phenotyping is Bordes et al., (Bordes et al., 2022). The study uses an open field environment with home-cage-like settings (using bedding material inside the maze) to explore chronic social defeat stress (CSDS)-induced social behavioral profiles. CSDS is a well-established stress model for investigating symptoms of depression, which ultimately leads to the avoidance of novel social stimuli in a subset of susceptible mice (Golden et al., 2011; Krishnan et al., 2007). However, the identification of social avoidance behavior is typically measured using the aforementioned social avoidance task or three-chamber task and thus relies on an oversimplified model of the social behavioral repertoire (Pearson et al., 2010; Bordes et al., 2022). In this study, an open-source python package DeepOpenField (DeepOF) was developed to analyze the stress effect using a set of supervised behavioral classifiers related to individualistic- and social behaviors, but also using an unsupervised clustering analysis of the individualistic and social behavioral interaction data. Both supervised and unsupervised pipelines showed a much stronger CSDS-induced difference in the free-social interaction task compared to the classical social avoidance task, indicating a more robust and clearer social behavioral profile using DeepOF. This study identified a subset of different social behavioral classifiers to be altered in stressed animals, which was subsequently used as an indicator for stress affectiveness by calculating a Z-score. Ultimately, the Z-score for stress affectiveness was significantly correlated with markers for stress physiology (end-point body weight and relative adrenal weight), which was not observed using the classical social avoidance IR, indicating that the Z-score analysis of DeepOF-identified behaviors provides a more complete social behavioral profile related to stress exposure than just the social avoidance IR. In addition, the mean-normalization of the data to the control group enables correcting for a potential batch effect. The DeepOF profiling ultimately provides a sensitive tool to assess individual stress resiliency and susceptibility. Interestingly, Ayash et al., (Ayash et al., 2023) observed that the social preference behavior of defeated C57/Bl6 mice in the three-chamber task is different between using a social conspecific from the aggressor strain (CD1) or a novel strain (e.g.,

129/Sv). In addition, they identified that defeated mice who do not avoid CD1 mice (classically labeled as resilient), show impaired learning in a fear conditioning model, and therefore the classification of resiliency is potentially incorrect. This is an interesting perspective and should be incorporated into the free social interaction task, to explore the social preference behavior after CSDS across different species.

Taken together, the increased information of data in video recognition using advanced pose estimation tools in a semi-naturalistic environment has expanded the understanding of the complexity of behavior. Implementations of these tools will lead to an increased understanding of the stress response system, for example, by enabling more ecologically relevant stressors, which are critical for the correct classification of resilient and susceptible animals (Lyons et al., 2023). In addition, these tools can dissect important influences of stress on behavior, such as the type of stressor (physiological versus psychological), stress intensity (chronic versus acute), and the dynamics of behavioral changes following a stressful challenge (Miranda et al., 2023).

4.3. Deep phenotyping benefits the integration of different factors influencing social behavior

Applying advanced tracking tools in semi-naturalistic environments offers the possibility to integrate different factors that shape social behavior, including stress exposure (stressed versus nonstressed), sex (female versus male), motivation (rewarding versus not rewarding), recording length (short versus longitudinal recordings), different developmental stages and aging (early life, adolescence, adulthood, older aged), and housing conditions (single housed, group-housed, or overcrowded), see Fig. 1. For example, the difference between short recordings (minutes to hours) and longitudinal recordings (day and night) is an important methodological consideration. Using shorter tests allows for high-throughput and more easily quantifiable data, and in addition is easily combined with other data modalities, such as neuronal activity data. However, these shorter tests are a snapshot during a certain time and are difficult to fully reflect non-arousal (baseline) behaviors. In contrast, longitudinal recordings are more suitable for investigating baseline behaviors and can be used for a truly free-choice behavioral task that is relying on no interaction with the experimenter. However, adding data modalities, such as neuronal activity is more difficult. Another important aspect is to distinguish between distinct developmental stages, which are known to differentially influence social behavior. First of all, there are specific vulnerable stages during life, in which animals are differentially influenced by environmental circumstances. For example, early life stress exposure influences the social behavioral profile and hierarchy establishment of adult mice in a sex-dependent manner (Bondar et al., 2018; Benner et al., 2014), but also the adolescent stage has been implicated as a critical stage for social behavioral development (Musardo et al., 2022; Endo et al., 2018). Secondly, aging is a strong factor influencing the social behavioral repertoire. Older aged animals have been found to show decreased explorative behaviors and increased grooming (Tran et al., 2021), while also a decreased number of social contacts and social novelty-seeking has been observed (Shoji et al., 2016; Shoji and Miyakawa, 2019). Moreover, the aggressive behavior towards conspecifics changes during aging as well, which is why for chronic social defeat stress only up to 6 months old aggressor mice can be used (Golden et al., 2011). A fundamental question for aging research is how behavior changes over longer periods of time (weeks to months). However, repeated behavioral testing with several classical behavioral tests influences the arousal and anxiety state of the animals, which ultimately leads to different behavioral outcomes between the tests (Voikar et al., 2004; Lad et al., 2010; McIlwain et al., 2001). As a consequence, researchers often use different cohorts of animals and test each cohort at a given age, making within-subject comparisons impossible, thereby increasing the risk of batch effects. A solution to this problem is to observe groups of animals in a semi-natural environment, for which no interference with anxiety,

or differences in the state of arousal can take place, as the animals are not in a specific test environment. Therefore, using advanced tracking tools in a semi-natural environmental set-up will not only facilitate fundamental research of social behavior, but also other research fields, such as stress and aging. The ability to test the same cohort of animals throughout the different stages of life will enable researchers to uncover unbiased behavioral trajectories over time. However, even though motion tracking data provides great insights into behavior, it does not provide the entire picture of the behavior profile. Currently, exciting developments are implementing different data modalities that can be measured in-vivo to provide an even richer understanding of the underlying behavior. The next section will discuss a set of data modalities that are currently being implemented in social behavioral analysis.

4.4. Integrating different behavioral and neuronal data modalities into social behavioral profiling

In recent years there has been an increased focus on the relevance of motion-tracking data for understanding behavior. This has been due to large breakthroughs in pose-estimation tools and their analysis, but also because video data is relatively easy and cheap to acquire. Furthermore, motion analysis can be easily interpreted, since researchers can map back their annotations and results to visual cues, which is significantly more difficult with other relevant types of data. However, behavior cannot be captured completely by motion tracking, as behavior is much more than just the posture of the animals. Therefore, new dimensions need to be included so that an even richer profile of the underlying behavior can be assessed. One option is to increase the data modalities using different sensory information of the animals, such as auditory data (e.g. mouse vocalizations), olfactory data (e.g. animal smell), and visual data (e.g. facial expressions) (Jabarin et al., 2022). Another option is to add physiological data modalities, for example, pupil dilation (Privitera et al., 2020), blood pressure, or heart rate, which can be used as a proxy for perceived stress (Buuse et al., 2001; Lemaire and Mormede, 1995; Farah et al., 2004; Swiercz et al., 2018). Lastly, adding neuronal data modalities will be important for future behavioral phenotyping, for example, by using neuronal activity (Padilla-Coreano et al., 2022), or neurotransmitter release data (Wu et al., 2022), which need to be wireless systems in order to sustain longitudinal recordings without restricting the animals (Fig. 3). The incorporation of different data modalities will increase the amount of data to be analyzed, which provides a new problem on how to tackle data integration. Open-source software is appearing to tackle these specific problems, such as CEBRA (Schneider et al., 2023). CEBRA aims to integrate data modalities to gain a more holistic unsupervised representation of behavior by enabling combined embeddings of motion tracking and neural activity data, which promises to facilitate the identification of the neural processes underlying motion. An interesting study by Padilla-Coreano et al., (Padilla-Coreano et al., 2022) utilized wireless electrophysiology devices to investigate and predict behavioral outcomes based on neuronal activity data. A different activity pattern was observed in the medial prefrontal cortex between social ranks and wins and losses in a reward competition task, which was visible 30 seconds prior to the start of the task. This study shows that the integration of neuronal activity data with behavior adds interesting insights to the understanding of behavior and the underlying mechanisms.

5. Conclusion

The history of the development of behavioral assessment has seen several influential methods that have their inherent advantages and disadvantages. Ethology applies observational research, which can study behavior in a descriptive way, but, among other problems, lacks the standardization of behavioral assessment and control over the environment. Comparative psychology methods reduced the complexity of the behavior by using a highly controllable environment but

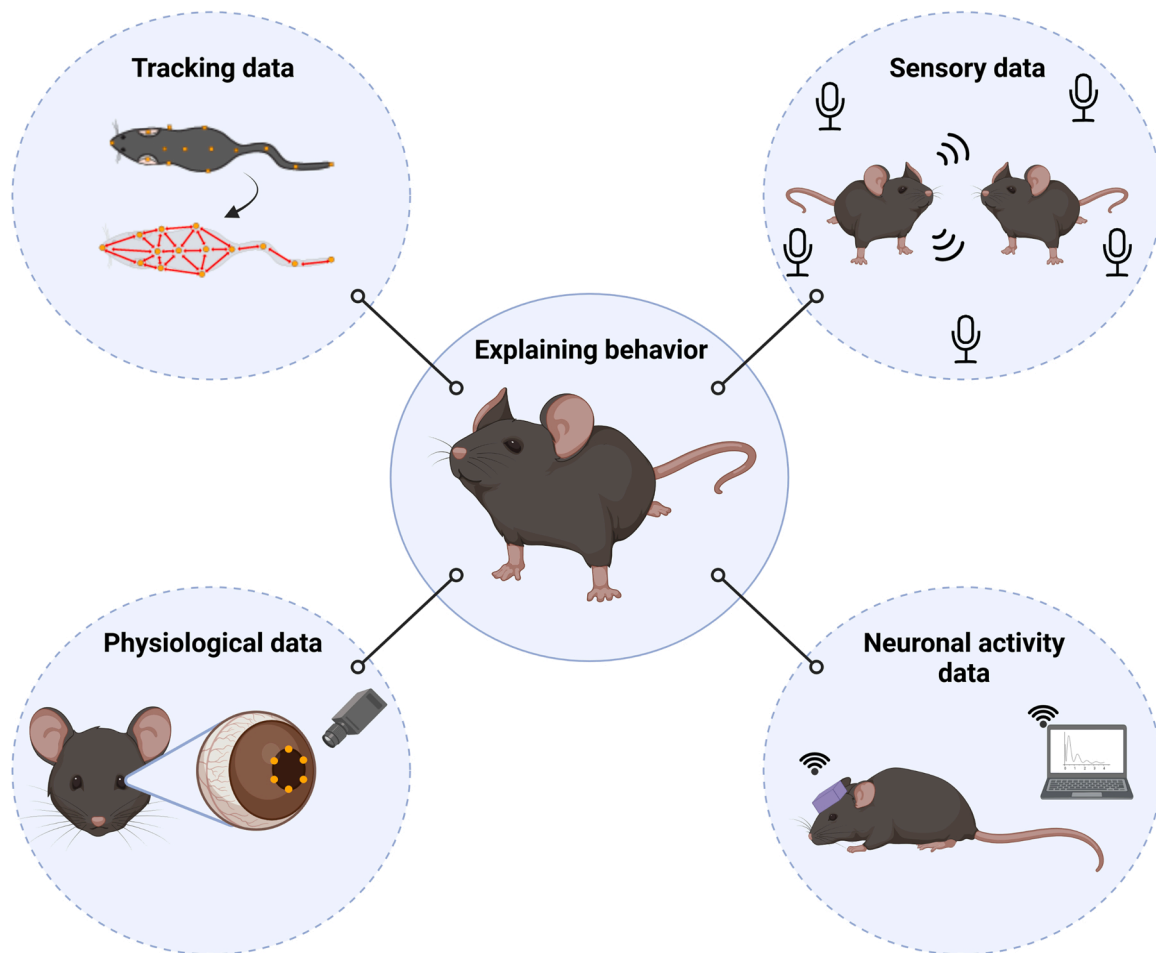


Fig. 3. Integrating data modalities in social behavioral profiling. Different data modalities need to be integrated into in-vivo recordings, in order to increase the understanding of rodent behavior. Diverse data modalities are currently being implemented, which include advanced tracking modalities using pose-estimation, sensory data (e.g. auditory, olfactory, and visual data), physiological data (e.g. pupil dilation, blood pressure or heart rate), and neuronal data modalities (e.g. neuronal activity data or neurotransmitter release).

oversimplified the underlying behavioral construct. The development of advanced and precise tracking tools, together with the numerous post-tracking analysis packages, has facilitated the (social) behavioral research field to assess behavior in a more naturalistic environmental set-up. This allows the investigation of behavior in a more descriptive way, without the interference of test-specific anxiety and arousal of the animals, but maintains a certain level of controllability over the environment. Further, the new behavioral analysis tools can be applied to classic behavioral tests to uncover and quantify novel behavioral traits. The implementation of body posture tracking and machine learning analysis will be beneficial to fundamental social behavioral research, but will also enable an increased understanding of the influences of many different factors that can influence social behavior, such as stress exposure, sex, motivation, recording length, different developmental stages, aging, and housing conditions. However, there are several challenges using such advanced tracking tools and an important step will be to develop an open-source centralized software system to allow a compatible ecosystem for tracking, annotation, and the embedding of multimodal behavioral data. Future research will increase the amount of data modalities, such as sensory, physiological, and neuronal activity data, and will thereby significantly enhance our understanding of the biological basis of social behavior and guide intervention strategies for behavioral abnormalities in psychiatric disorders.

Data Availability

No data was used for the research described in the article.

Acknowledgments

The authors thank London Aman for the language editing of the manuscript. Figures were created with BioRender.com. This work was supported by the Kids2Health grant of the Federal Ministry of Education and Research (01GL1743C) and the SAME-NeuroID grant of the European Research Executive Agency (101079181), and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813533.

Author contributions

The concept of the manuscript was established by JB and MVS. The writing of the manuscript by JB. Feedback and support by MVS, LM, and BMM.

References

- Aggleton, J.P., 1985. One-trial object recognition by rats. *Q. J. Exp. Physiol.* 37 (4), 279–294. <https://doi.org/10.1080/14640748508401171>.
- American Psychiatric Association, 2013. *Diagnostic and statistical manual of mental disorders*. Am. Psychiatr. Assoc. vol. 21.

- Ayash, S., Lingner, T., Ramisch, A., Ryu, S., Kalisch, R., Schmitt, U., Müller, M.B., 2023. Fear circuit-based neurobehavioral signatures mirror resilience to chronic social stress in mouse. *Proc. Nat. Acad. Sci.* 120 (17), e2205576120.
- Bariselli, S., et al., 2018. Role of VTA dopamine neurons and neurexin 3 in sociability traits related to nonfamiliar conspecific interaction. *Nat. Commun.* 9, 3173.
- Barnes, C.A., 1979. Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. *J. Comp. Physiol. Psychol.* 93, 74–104.
- Beery, A.K., Kaufer, D., 2015. Stress, social behavior, and resilience: insights from rodents. *Neurobiol. Stress* 1, 116–127.
- Benner, S., Endo, T., Endo, N., Kakeyama, M., Tohyama, C., 2014. Early deprivation induces competitive subordination in C57BL/6 male mice. *Physiol. Behav.* 137, 42–52.
- Berton, O., et al., 2006. Essential Role of BDNF in the mesolimbic dopamine pathway in social defeat stress (1979). *Science* 311, 864–868.
- Bidinosti, M., et al., 2016. CLK2 inhibition ameliorates autistic features associated with SHANK3 deficiency (1979). *Science* 351, 1199–1203.
- Blumstein, D.T., 2010. Towards an integrative understanding of social behavior: new models and new opportunities. *Front Behav. Neurosci.* <https://doi.org/10.3389/fnbeh.2010.00034>.
- Bolhuis, J.J. & Giraldeau, L.-Alain. The study of animal behavior. (2005).
- Bolhuis, J.J., Giraldeau, L., Hogan, J.A., 2021. The study of animal behavior. In *The Behavior of Animals*, 2nd Edition. Wiley, pp. 1–11. <https://doi.org/10.1002/9781119109556.ch1>.
- Bondar, N.P., Lepeshko, A.A., Reshetnikov, V. v, 2018. Effects of early-life stress on social and anxiety-like behaviors in adult mice: sex-specific effects. *Behav. Neurol.* 1–13.
- Bordes, J., et al., 2022. Automatically annotated motion tracking identifies a distinct social behavioral profile following chronic social defeat stress, 2022.06.23.497350 bioRxiv. <https://doi.org/10.1101/2022.06.23.497350>.
- Buuse, M., Acker, S.A.B.E., Flutterm, M., Kloet, E.R., 2001. Blood pressure, heart rate, and behavioral responses to psychological 'novelty' stress in freely moving rats. *Psychophysiology* 38, 490–499.
- de Chaumont, F., et al., 2019. Real-time analysis of the behaviour of groups of mice via a depth-sensing camera and machine learning. *Nat. Biomed. Eng.* 3, 930–942.
- Chen, P., Hong, W., 2018. Neural circuit mechanisms of social behavior. *Neuron* 98, 16–30.
- Chesler, E.J., Wilson, S.G., Lariviere, W.R., Rodriguez-Zas, S.L., Mogil, J.S., 2002. Influences of laboratory environment on behavior. *Nat. Neurosci.* 5, 1101–1102.
- Cordero, M., Sandi, C., 2007. Stress amplifies memory for social hierarchy. *Front. Neurosci.* 1, 175–184.
- Crawley, J.N., 1996. Unusual behavioral phenotypes of inbred mouse strains. *Trends Neurosci.* 19, 181–182.
- Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life, 1859. John Murray, London.
- Dodell-Feder, D., Tully, L.M., Hooker, C.L., 2015. Social impairment in schizophrenia. *Curr. Opin. Psychiatry* 28, 236–242.
- Donahue, R.J., Muschamp, J.W., Russo, S.J., Nestler, E.J., Carlezon, W.A., 2014. Effects of striatal ΔFosB overexpression and ketamine on social defeat stress-induced anhedonia in mice. *Biol. Psychiatry* 76, 550–558.
- Durand-de Cuttoli, R., et al., 2022. Distinct forms of regret linked to resilience versus susceptibility to stress are regulated by region-specific CREB function in mice. *Sci. Adv.* 8.
- Dwartz, M.F., Curley, J.P., Tye, K.M., Padilla-Coreano, N., 2022. Neural systems that facilitate the representation of social rank. *Philos. Trans. R. Soc. B: Biol. Sci.* 377.
- Ellis, L., 1995. Dominance and reproductive success among nonhuman animals: a cross-species comparison. *Ethol. Socio* 16, 257–333.
- Endo, N., et al., 2018. Multiple animal positioning system shows that socially-reared mice influence the social proximity of isolation-reared cagemates. *Commun. Biol.* 1, 225.
- Engelmann, M., Hädicke, J., Noack, J., 2011. Testing declarative memory in laboratory rats and mice using the nonconditioned social discrimination procedure. *Nat. Protoc.* 6, 1152–1162.
- Fan, Z., et al., 2019. Using the tube test to measure social hierarchy in mice. *Nat. Protoc.* 14, 819–831.
- Farah, V.M.A., Joaquim, L.F., Bernatova, I., Morris, M., 2004. Acute and chronic stress influence blood pressure variability in mice. *Physiol. Behav.* 83, 135–142.
- Fernández-Theoduloz, G., et al., 2019. Social avoidance in depression: a study using a social decision-making task. *J. Abnorm. Psychol.* 128, 234–244.
- Forkosh, O., et al., 2019. Identity domains capture individual differences from across the behavioral repertoire. *Nat. Neurosci.* 22, 2023–2028.
- Georgiou, P., et al., 2022. Experimenters' sex modulates mouse behaviors and neural responses to ketamine via corticotropin releasing factor. *Nat. Neurosci.* 25, 1191–1200.
- Gheusi, G., Bluthé, R.-M., Goodall, G., Dantzer, R., 1994. Social and individual recognition in rodents: methodological aspects and neurobiological bases. *Behav. Process.* 33, 59–87.
- Golden, S.A., Covington, H.E., Berton, O., Russo, S.J., 2011. A standardized protocol for repeated social defeat stress in mice. *Nat. Protoc.* 6, 1183–1191.
- Goodwin, N.L., Nilsson, S.R.O., Golden, S.A., 2020. Rage against the machine: advancing the study of aggression ethology via machine learning. *Psychopharmacol. (Berl.)* 237, 2569–2588.
- Goodwin, N.L., Nilsson, S.R.O., Choong, J.J., Golden, S.A., 2022. Toward the explainability, transparency, and universality of machine learning for behavioral classification in neuroscience. *Curr. Opin. Neurobiol.* 73, 102544.
- Graving, J.M., et al., 2019. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife* 8.
- Hånell, A., Marklund, N., 2014. Structured evaluation of rodent behavioral tests used in drug discovery research. *Front. Behav. Neurosci.* 8.
- Heimlich, J.E., Ardoin, N.M., 2008. Understanding behavior to understand behavior change: a literature review. *Environ. Educ. Res.* 14, 215–237.
- Heuer, K., Rinck, M., Becker, E.S., 2007. Avoidance of emotional facial expressions in social anxiety: the approach-avoidance task. *Behav. Res. Ther.* 45, 2990–3001.
- Hsu, A.I., Yttri, E.A., 2021. B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nat. Commun.* 12, 5188.
- Jabarin, R., Netser, S., Wagner, S., 2022. Beyond the three-chamber test: toward a multimodal and objective assessment of social behavior in rodents. *Mol. Autism* 13, 41.
- Jeon, D., Shin, H., 2011. A mouse model for observational fear learning and the empathetic response. *Curr. Protoc. Neurosci.* 57.
- Karamihalev, S., et al., 2020. Social dominance mediates behavioral adaptation to chronic stress in a sex-specific manner. *Elife* 9.
- Katz, S.J., Conway, C.C., Hammen, C.L., Brennan, P.A., Najman, J.M., 2011. Childhood social withdrawal, interpersonal impairment, and young adult depression: a mediational model. *J. Abnorm. Child Psychol.* 39, 1227–1238.
- Keum, S., Shin, H.-S., 2019. Neural basis of observational fear learning: a potential model of affective empathy. *Neuron* 104, 78–86.
- Krishnan, V., et al., 2007. Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell* 131, 391–404.
- Lad, H.V., et al., 2010. Behavioural battery testing: Evaluation and behavioural outcomes in 8 inbred mouse strains. *Physiol. Behav.* 99, 301–316.
- Larrieu, T., et al., 2017. Hierarchical status predicts behavioral vulnerability and nucleus accumbens metabolic profile following chronic social defeat stress. *e4 Curr. Biol.* 27, 2202–2210. e4.
- Lauer, J., et al., 2022. Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nat. Methods* 19, 496–504.
- LeClair, K.B., et al., 2021. Individual history of winning and hierarchy landscape influence stress susceptibility in mice. *Elife* 10.
- Lemaire, V., Mormede, P., 1995. Telemetered recording of blood pressure and heart rate in different strains of rats during chronic social stress. *Physiol. Behav.* 58, 1181–1188.
- Li, L., et al., 2022. Social trauma engages lateral septum circuitry to occlude social reward. *Nature*. <https://doi.org/10.1038/s41586-022-05484-5>.
- Lindzey, G., Manosevitz, M., Winston, H., 1966. Social dominance in the mouse. *Psychon. Sci.* 5, 451–452.
- Lopez, J.P., et al., 2022. Ketamine exerts its sustained antidepressant effects via cell-type-specific regulation of Kcnq2. *Neuron*. <https://doi.org/10.1016/j.neuron.2022.05.001>.
- Lorsch, Z.S., et al., 2021. Computational analysis of multidimensional behavioral alterations after chronic social defeat stress. *Biol. Psychiatry* 89, 920–928.
- Lukas, M., et al., 2011. The neuropeptide oxytocin facilitates pro-social behavior and prevents social avoidance in rats and mice. *Neuropsychopharmacology* 36, 2159–2168.
- Luxem, K., et al., 2022. Identifying behavioral structure from deep variational embeddings of animal motion. *Commun. Biol.* 5, 1267.
- Lyons, D.M., Ayash, S., Schatzberg, A.F., Müller, M.B., 2023. Ecological validity of social defeat stressors in mouse models of vulnerability and resilience. *Neurosci. Biobehav. Rev.* 145, 105032.
- Mackintosh, J.H., Grant, E.C., 1963. A comparison of the social postures of some common laboratory rodents. *Behaviour* 21, 246–259.
- Marks, M., et al., 2022. Deep-learning-based identification, tracking, pose estimation and behaviour classification of interacting primates and mice in complex environments. *Nat. Mach. Intell.* 4, 331–340.
- Mathis, A., et al., 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* 21, 1281–1289.
- Mathis, M., et al., 2020. DLC2Kinematics: a Post-Deep. Modul. kinematic Anal. <https://doi.org/10.5281/ZENODO.6669074>.
- Mathis, M.W., Mathis, A., 2020. Deep learning tools for the measurement of animal behavior in neuroscience. *Curr. Opin. Neurobiol.* 60, 1–11.
- Matthews, G.A., et al., 2016. Dorsal raphe dopamine neurons represent the experience of social isolation. *Cell* 164, 617–631.
- McIlwain, K.L., Merriweather, M.Y., Yuva-Paylor, L.A., Paylor, R., 2001. The use of behavioral test batteries: effects of training history. *Physiol. Behav.* 73, 705–717.
- Merlot, E., Moze, E., Bartolomucci, A., Dantzer, R., Neveu, P.J., 2004. The rank assessed in a food competition test influences subsequent reactivity to immune and social challenges in mice. *Brain Behav. Immun.* 18, 468–475.
- Miranda, L., Bordes, J., Gasperoni, S., Lopez, J.P., 2023. Increasing resolution in stress neurobiology: from single cells to complex group behaviors. *Stress* 26.
- Morel, C., et al., 2022. Midbrain projection to the basolateral amygdala encodes anxiety-like but not depression-like behaviors. *Nat. Commun.* 13, 1532.
- Morris, R.G.M., 1981. Spatial localization does not require the presence of local cues. *Learn. Motiv.* 12, 239–260.
- Murra, D., et al., 2022. Characterizing the behavioral and neuroendocrine features of susceptibility and resilience to social stress. *Neurobiol. Stress* 17, 100437.
- Musardo, S., Contestabile, A., Knoop, M., Baud, O., Bellone, C., 2022. Oxytocin neurons mediate the effect of social isolation via the VTA circuits. *Elife* 11.
- Nguyen, K.P., O'Neal, T.J., Bolonduro, O.A., White, E., Kravitz, A. v, 2016. Feeding experimentation device (FED): A flexible open-source device for measuring feeding behavior. *J. Neurosci. Methods* 267, 108–114.
- Nietlisbach, G., Maercker, A., 2009. Social cognition and interpersonal impairments in trauma survivors with PTSD. *J. Aggress. Maltreat. Trauma* 18, 382–402.

- Nilsson, S.R.O., et al., 2020. Simple behavioral analysis (SimBA) – an open source toolkit for computer classification of complex social behaviors in experimental animals, 2020.04.19.049452 bioRxiv. <https://doi.org/10.1101/2020.04.19.049452>.
- Ottenbreit, N.D., Dobson, K.S., Quigley, L., 2014. An examination of avoidance in major depression in comparison to social anxiety disorder. *Behav. Res. Ther.* 56, 82–90.
- Padilla-Coreano, N., et al., 2022. Cortical ensembles orchestrate social competition through hypothalamic outputs. *Nature* 603, 667–671.
- Park, M.-J., Seo, B.A., Lee, B., Shin, H.-S., Kang, M.-G., 2018. Stress-induced changes in social dominance are scaled by AMPA-type glutamate receptor phosphorylation in the medial prefrontal cortex. *Sci. Rep.* 8, 15008.
- Pearson, B.L., Defensor, E.B., Blanchard, D.C., Blanchard, R.J., 2010. C57BL/6J mice fail to exhibit preference for social novelty in the three-chamber apparatus. *Behav. Brain Res.* 213, 189–194.
- Peleh, T., Ike, K.G.O., Wams, E.J., Lebois, E.P., Hengerer, B., 2019a. The reverse translation of a quantitative neuropsychiatric framework into preclinical studies: Focus on social interaction and behavior. *Neurosci. Biobehav. Rev.* 97, 96–111.
- Peleh, T., Bai, X., Kas, M.J.H., Hengerer, B., 2019b. RFID-supported video tracking for automated analysis of social behaviour in groups of mice. *J. Neurosci. Methods* 325, 108323.
- Pereira, T.D., et al., 2019. Fast animal pose estimation using deep neural networks. *Nat. Methods* 16, 117–125.
- Pereira, T.D., et al., 2022. SLEAP: a deep learning system for multi-animal pose tracking. *Nat. Methods* 19, 486–495.
- Pisansky, M.T., Hanson, L.R., Gottesman, I.I., Gewirtz, J.C., 2017. Oxytocin enhances observational fear in mice. *Nat. Commun.* 8, 2102.
- Pöhlmann, M.L., et al., 2018. Pharmacological modulation of the psychiatric risk factor FKBP51 alters efficiency of common antidepressant drugs. *Front. Behav. Neurosci.* 12.
- Privitera, M., et al., 2020. A complete pupillometry toolbox for real-time monitoring of locus coeruleus activity in rodents. *Nat. Protoc.* 15, 2301–2320.
- Qin, L., et al., 2018. Social deficits in Shank3-deficient mouse models of autism are rescued by histone deacetylase (HDAC) inhibition. *Nat. Neurosci.* 21, 564–575.
- Qu, C., Ligneul, R., van der Henst, J.-B., Dreher, J.-C., 2017. An integrative interdisciplinary perspective on social dominance hierarchies. *Trends Cogn. Sci.* 21, 893–908.
- Rapanelli, M., et al., 2021. Behavioral, circuitry, and molecular aberrations by region-specific deficiency of the high-risk autism gene *Cul3*. *Mol. Psychiatry* 26, 1491–1504.
- Reguilón, M.D., Ferrer-Pérez, C., Manzanedo, C., Miñarro, J., Rodríguez-Arias, M., 2021. Ethanol intake in male mice exposed to social defeat: Environmental enrichment potentiates resilience. *Neurobiol. Stress* 15, 100413.
- Rein, B., et al., 2021. Reversal of synaptic and behavioral deficits in a 16p11.2 duplication mouse model via restoration of the GABA synapse regulator *Npas4*. *Mol. Psychiatry* 26, 1967–1979.
- Rein, B., Ma, K., Yan, Z., 2020. A standardized social preference protocol for measuring social deficits in mouse models of autism. *Nat. Protoc.* 15, 3464–3477.
- Richter, S.H., 2017. Systematic heterogenization for better reproducibility in animal experimentation. *Lab Anim. (NY)* 46, 343–349.
- Richter, S.H., Garner, J.P., Würbel, H., 2009. Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nat. Methods* 6, 257–261.
- Rinck, M., et al., 2010. Social anxiety predicts avoidance behaviour in virtual encounters. *Cogn. Emot.* 24, 1269–1276.
- Sakaguchi, T., Iwasaki, S., Okada, M., Okamoto, K., Ikegaya, Y., 2018. Ethanol facilitates socially evoked memory recall in mice by recruiting pain-sensitive anterior cingulate cortical neurons. *Nat. Commun.* 9, 3526.
- Sapolsky, R.M., 2005. The influence of social hierarchy on primate health. *Science* 308, 648–652.
- Scheggia, D., et al., 2022. Reciprocal cortico-amygdala connections regulate prosocial and selfish choices in mice. *Nat. Neurosci.* 25, 1505–1518.
- Schneider, S., Lee, J.H., Mathis, M.W., 2023. Learnable latent embeddings for joint behavioural and neural analysis. *Nature* 1–9.
- Segalin, C., et al., 2021. The mouse action recognition system (MARS) software pipeline for automated analysis of social behaviors in mice. *Elife* 10.
- Shemesh, Y., et al., 2013. High-order social interactions in groups of mice. *Elife* 2.
- Shemesh, Y., Chen, A., 2023. A paradigm shift in translational psychiatry through rodent neuroethology. *Mol. Psychiatry*. <https://doi.org/10.1038/s41380-022-01913-z>.
- Shoji, H., Miyakawa, T., 2019. Age-related behavioral changes from young to old age in male mice of a C57BL/6J strain maintained under a genetic stability program. *Neuropsychopharmacol. Rep.* 39, 100–118.
- Shoji, H., Takao, K., Hattori, S., Miyakawa, T., 2016. Age-related changes in behavior in C57BL/6J mice from young adulthood to middle age. *Mol. Brain* 9, 11.
- Skinner, B.F., 1948. 'Superstition' in the pigeon. *J. Exp. Psychol.* 38, 168–172.
- Solić, C., et al., 2022a. Superior Colliculus to VTA pathway controls orienting response and influences social interaction in mice. *Nat. Commun.* 13, 817.
- Solić, C., Girard, B., Righetti, B., Tapparel, M., Bellone, C., 2022b. VTA dopamine neuron activity encodes social interaction and promotes reinforcement learning through social prediction error. *Nat. Neurosci.* 25, 86–97.
- Sorge, R.E., et al., 2014. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat. Methods* 11, 629–632.
- Sturman, O., et al., 2020. Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions. *Neuropsychopharmacology* 45, 1942–1952.
- Sulloway, F.J., 1982. Darwin and his finches: the evolution of a legend. *J. Hist. Biol.* 15, 1–53.
- Sun, J.J., et al., 2021. The Multi-Agent Behavior Dataset. *Mouse Dyadic Soc. Interact.* <https://doi.org/10.48550/arxiv.2104.02710>.
- Swiercz, A.P., Seligowski, A. v., Park, J., Marvar, P.J., 2018. Extinction of fear memory attenuates conditioned cardiovascular fear reactivity. *Front. Behav. Neurosci.* 12.
- Tamashiro, K.L.K., Nguyen, M.M.N., Sakai, R.R., 2005. Social stress: from rodents to primates. *Front. Neuroendocr.* 26, 27–40.
- Thierry, B., 2010. Darwin as a student of behavior. *C. R. Biol.* 333, 188–196.
- Thorndike, E.L., 1898. Animal intelligence: an experimental study of the associative processes in animals. *Psychol. Rev.: Monogr. Suppl.* 2, i–109.
- Tinbergen, N., 1939. Why do birds behave as they do?(II). *Bird-lore* 41, 23–30.
- Tolman, E.C., 1948. Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208.
- Torquet, N., et al., 2018. Social interactions impact on the dopaminergic system and drive individuality. *Nat. Commun.* 9, 3081.
- Toth, I., Neumann, I.D., 2013. Animal models of social avoidance and social fear. *Cell Tissue Res* 354, 107–118.
- Toth, I., Neumann, I.D., Slaterry, D.A., 2013. Social fear conditioning as an animal model of social anxiety disorder. *Curr. Protoc. Neurosci.* 63.
- Tran, T., et al., 2021. Male–female differences in the effects of age on performance measures recorded for 23 hours in mice. *J. Gerontol.: Ser. A* 76, 2141–2146.
- Ujita, W., Kohyama-Koganeya, A., Endo, N., Saito, T., Oyama, H., 2018. Mice lacking a functional NMDA receptor exhibit social subordination in a group-housed environment. *FEBS J.* 285, 188–196.
- Voelkl, B., et al., 2020. Reproducibility of animal research in light of biological variation. *Nat. Rev. Neurosci.* 21, 384–393.
- Voikar, V., Vasar, E., Rauvala, H., 2004. Behavioral alterations induced by repeated testing in C57BL/6J and 129S2/Sv mice: implications for phenotyping screens. *Genes Brain Behav.* 3, 27–38.
- von Ziegler, L., Sturman, O., Bohacek, J., 2021. Big behavior: challenges and opportunities in a new era of deep behavior profiling. *Neuropsychopharmacology* 46, 33–44.
- Wahlsten, D., Metten, P., Crabbe, J.C., 2003. A rating scale for wildness and ease of handling laboratory mice: results for 21 inbred strains tested in two laboratories. *Genes Brain Behav.* 2, 71–79.
- Wang, F., et al., 2011. Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. *Science* 334, 693–697.
- Wang, F., Kessels, H.W., Hu, H., 2014. The mouse that roared: neural mechanisms of social hierarchy. *Trends Neurosci.* 37, 674–682.
- Weber, E.M., Dallaire, J.A., Gaskill, B.N., Pritchett-Corning, K.R., Garner, J.P., 2017. Aggression in group-housed laboratory mice: why can't we solve the problem? *Lab Anim.* 46, 157–161.
- Wei, D., Talwar, V., Lin, D., 2021. Neural circuits of social behaviors: innate yet flexible. *Neuron* 109, 1600–1620.
- Weinreb, C., et al., 2023. Keypoint-MoSeq: parsing behavior by linking point tracking to pose dynamics, 2023.03.16.532307 bioRxiv. <https://doi.org/10.1101/2023.03.16.532307>.
- Whishaw, I.Q., Dall, V.B., Kolb, B., 2006. Analysis of behavior in laboratory rats. In *The Laboratory Rat*. Elsevier, pp. 191–218. <https://doi.org/10.1016/B978-012074903-4/50011-X>.
- Wiltshchko, A.B., et al., 2020. Revealing the structure of pharmacobehavioral space through motion sequencing. *Nat. Neurosci.* 23, 1433–1443.
- Winslow, J.T., 2003. Mouse social recognition and preference. *Curr. Protoc. Neurosci.* 22.
- Wrenn, C.C., 2004. Social transmission of food preference in mice. *Curr. Protoc. Neurosci.* 28.
- Wu, Z., Lin, D., Li, Y., 2022. Pushing the frontiers: tools for monitoring neurotransmitters and neuromodulators. *Nat. Rev. Neurosci.* 23, 257–274.
- Ye, S., Mathis, A. & Mathis, M.W. Panoptic animal pose estimators are zero-shot performers. , 2022 doi:10.48550/arxiv.2203.07436.
- Yttri, E., Hsu, A., Schweihoff, J. & Schwarz, M. A-SOId, an active learning platform for expert-guided, data efficient discovery of behavior. (2023) doi:10.21203/rs.3.rs-2239160/v1.
- Zhou, T., et al., 1979. History of winning remodels thalamo-PFC circuit to reinforce social dominance. *Science* 357 (162–168), 2017.
- Zhou, T., Sandi, C., Hu, H., 2018. Advances in understanding neural mechanisms of social dominance. *Curr. Opin. Neurobiol.* 49, 99–107.
- Zimprich, A., et al., 2017. Assessing sociability, social memory, and pup retrieval in mice. *Curr. Protoc. Mouse Biol.* 7, 287–305.