



Rage Against the Machine: Advancing the study of aggression ethology via machine learning.

Nastacia L. Goodwin^{1,2} · Simon R.O. Nilsson¹ · Sam A. Golden^{1,2,3} 

Received: 7 November 2019 / Accepted: 1 June 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Rationale Aggression, comorbid with neuropsychiatric disorders, exhibits with diverse clinical presentations and places a significant burden on patients, caregivers, and society. This diversity is observed because aggression is a complex behavior that can be ethologically demarcated as either appetitive (rewarding) or reactive (defensive), each with its own behavioral characteristics, functionality, and neural basis that may transition from adaptive to maladaptive depending on genetic and environmental factors. There has been a recent surge in the development of preclinical animal models for studying appetitive aggression-related behaviors and identifying the neural mechanisms guiding their progression and expression. However, adoption of these procedures is often impeded by the arduous task of manually scoring complex social interactions. Manual observations are generally susceptible to observer drift, long analysis times, and poor inter-rater reliability, and are further incompatible with the sampling frequencies required of modern neuroscience methods.

Objectives In this review, we discuss recent advances in the preclinical study of appetitive aggression in mice, paired with our perspective on the potential for machine learning techniques in producing automated, robust scoring of aggressive social behavior. We discuss critical considerations for implementing valid computer classifications within behavioral pharmacological studies.

Key results Open-source automated classification platforms can match or exceed the performance of human observers while removing the confounds of observer drift, bias, and inter-rater reliability. Furthermore, unsupervised approaches can identify previously uncharacterized aggression-related behavioral repertoires in model species.

Discussion and conclusions Advances in open-source computational approaches hold promise for overcoming current manual annotation caveats while also introducing and generalizing computational neuroethology to the greater behavioral neuroscience community. We propose that currently available open-source approaches are sufficient for overcoming the main limitations preventing wide adoption of machine learning within the context of preclinical aggression behavioral research.

Keywords Aggression · Machine learning · Mice · Motivation · Reward

Nastacia L. Goodwin and Simon R.O. Nilsson contributed equally to this work.

✉ Sam A. Golden
sagolden@uw.edu

¹ Department of Biological Structure, University of Washington, Seattle, WA, USA

² Graduate Program in Neuroscience, University of Washington, Seattle, WA, USA

³ Center of Excellence in Neurobiology of Addiction, Pain, and Emotion (NAPE), University of Washington, Seattle, WA, USA

Introduction

There is an increased risk for abnormal or pathological aggression in individuals suffering from neuropsychiatric disorders. A growing literature indicates that certain types of aggression, namely compulsive aggression seeking, may be a distinct externalizing pathology that potentially functions through dysregulation of reward processing in a manner akin to drug addiction (Blair 2016; Chester and DeWall 2016). This is supported by the common finding that, in humans, some neuropsychiatric disorders present comorbid with inappropriate aggression, including depression and substance abuse (Tyrer et al. 2015; Fazel et al. 2015), and recidivism rates of violent offenders closely mimic relapse rates for drug-addicted

individuals (Sinha 2011; Ducrose et al. 2014). We and others have hypothesized that studying the motivational component of aggression, termed appetitive aggression, may provide novel therapeutic approaches to the treatment of maladaptive aggression presenting both within and outside of neuropsychiatric comorbidities (Miczek et al. 2015, 2017; Golden et al. 2019b; Covington et al. 2019). Unfortunately, from a preclinical perspective, there are relatively few established models for mechanistically studying the neurobiological basis of maladaptive appetitive aggression (Hashikawa et al. 2018; Flanigan and Russo 2019; Golden et al. 2019b).

Aggression motivation has long been a focus of preclinical research (Thompson 1963), invariably examined through the lens of ethological analysis, but often stymied due to the tremendous effort and durations required to manually score complex social behavior with sampling frequencies and accuracies that match modern neuroscience techniques. Several recent reviews have highlighted the power of machine learning approaches for creating automated behavioral classifiers to study social behavior, including aggression (Anderson and Perona 2014; Egnor and Branson 2016; Robie et al. 2017; Gris et al. 2017; Brown and de Bivort 2018; Akay and Hess 2019; Mathis and Mathis 2019; Datta et al. 2019). The ability to recognize and categorize common behaviors in model species is an integral component of model reproducibility and extendibility, and these observations can be combined with computational neuroethology to circumvent several inherent issues with manual annotation—most notably including observer drift and bias, long analysis times, and inter-rater reliability (Anderson and Perona 2014; Egnor and Branson 2016; Datta et al. 2019). Additionally, both supervised and unsupervised machine learning algorithms have uncovered previously unknown behavioral repertoires in model organisms and have confirmed foundational assumptions of ethology (Vogelstein et al. 2014; Wiltshcko et al. 2015; Rudolf et al. 2019). Several machine learning–based open-source packages have been developed that can track one or multiple animals during freely moving behavior, while others, including one recently released by our lab (Nilsson et al. 2020), can be used to classify behaviors based on pose estimation tracking (Table 2). These approaches have led to a renaissance in the study of ethology, which is poised to catapult classical behavioral neuroscience into the realm of “big data.”

However, even as these techniques have removed several hurdles impeding high-throughput behavioral analysis, the use of automated behavioral classifiers has proven oddly difficult to adopt and generalize across labs. Implementation of these techniques is often slowed by a lack of computational knowledge, the need for specialized and expensive equipment, and the high computational expense to adequately train new classifiers. Due to the complexity of these programs, even the initial installation can be intimidating or difficult regardless of previous programming experience. In this perspective,

we will briefly review the state of the appetitive aggression literature (also see Golden and Shaham 2018; Golden et al. 2019b), and then within this context, provide a primer on how machine learning approaches (regarding both acquisition/tracking and predictive classifier analysis) may be incorporated into future studies. Definitions of commonly used terms in computational neuroethology are included in Table 1. We propose, predominantly thanks to efforts of numerous labs in developing and advancing machine learning methods for behavioral tracking (He et al. 2017; Mathis et al. 2018; Graving et al. 2019; Pereira et al. 2019), that currently available approaches are sufficient to overcome the main limitations preventing wide adoption of machine learning for scoring complex social behavior within the context of preclinical aggression research.

Appetitive aggression

Winners like to win: revisiting aggression reward

Sixty years of behavioral research have shown that the opportunity for, or experience of, an aggressive encounter with a conspecific can be reinforcing in many species, including select mammals. Early work established the propensity of Siamese fighting fish (*Beta splendens*) to perform operant tasks for the opportunity to attack a static (Thompson 1963; Thompson and Sturm 1965) or animate representation of a conspecific (Craft et al. 2003, 2007; Elcoro et al. 2008). Similar behaviors were also demonstrated in homing pigeons (Cole and Parker 1971) and male fighting cocks (Thompson 1964). Although this pioneering work showed that animals will seek the opportunity for aggression (once thought to be a uniquely human trait), these studies did not explore the relationship between seeking aggression and the actual experience of an aggressive encounter.

Subsequent work allowing physical attacks against conspecifics has highlighted the importance of dominance in solidifying aggression reward seeking. Among many species, those animals that win their first aggressive encounter are more likely to win subsequent bouts, while those that lose are more likely to lose subsequent bouts. Such “winner” and “loser” effects (Ginsburg and Allee 1942; Oyegbile and Marler 2005, 2006; Kudryavtseva et al. 2011), and have been demonstrated in meadow voles (Vlautin and Ferkin 2013), crayfish (Momohara et al. 2016), flour beetles (Okada et al. 2019), *Drosophila* (Trannoy et al. 2015; Kim et al. 2018), lobster cockroaches (Kou et al. 2019), Syrian hamsters (Schwartz et al. 2013), and many species and strains of mice (Oyegbile and Marler 2006). Some mammals do not display the winner effect, however, and species such as the white-footed mouse only do so following experimental manipulations of testosterone levels (Fuxjager et al. 2011). Aggressive encounters may

Table 1 Definitions of common terms in computational neuroethology

Term	Description
Pose	The relationships and location of body part “key-points” (e.g., individual joints/digits in motor control tasks; or the nose, ears, flanks, centroid, and tail base in an open-field task) that are relevant for determining an individual’s activity in a video frame.
Pose estimation	Computer-generated predictions of the individual’s pose in a video frame.
Behavior_classification	A computer-generated categorical prediction describing the presence or absence of a behavior (e.g., attack versus no attack) in a video frame.
Features	Independent variables derived from pose estimation (e.g., animal velocities and the distances between animals) that are used to predict the presence and absence of a behavior in a video frame.
Feature weight/importance	The relevance of a specific feature for accurately classifying the presence and absence of a behavior in a video frame. For example, the distance between the first animal’s nose and the second animal’s tail base has importance or weight for accurately classifying anogenital sniffing.
Ensemble algorithm	The grouping of multiple learning algorithms based on different subsets of features. The individually weak learning algorithms are combined to create a strong or accurate behavior classification (e.g., random forest algorithm).
Hidden states	Unobservable factors determining the probability for ensuing expression of a behavior given the current behavior.
Model interpretability	The extent to which a machine model’s decision processes can be understood by humans (e.g., how does the algorithm discriminate between attack and non-attack events?).
Model explainability	The extent to which a machine learning system allows a qualitative comprehension of the association between the features of a specific observation and its prediction (e.g., why is the video frame classified as containing an attack event?). Note that an explainable model also is an interpretable model.
“Black box” model	A metaphor applied to machine learning algorithms with decision processes that are not readily explainable or interpretable.

be sought prior to exposure, but winning often matters in maintaining the drive to seek aggression and in determining future performance.

To minimize the effects of bout outcome on behavioral measures, many mouse assays for aggression use a variation on the resident-intruder task where a smaller subordinate intruder is introduced into the home cage of a larger, older resident. Early studies using these procedures found that mice will cross electric grids (Lagerspetz 1964) and navigate T-mazes (Tellegen and Horn 1972; Legrand 1978) and runways (Legrand 1970) to attack a subordinate. Furthermore, physical aggression alone is sufficient to condition mice to prefer aggression-paired contexts (Potegal 1979; Taylor 1979). While these early influential studies demonstrate the utility of mouse models in the study of appetitive aggression, subsequent work shifted toward using resident-intruder (Miczek and O’Donnell 1978; Brain et al. 1981) or sensory contact (Kudryavtseva et al. 1991) assays where pairings are repeated and the subsequent interactions are recorded and scored for a variety of behavioral measures. However, such procedures may be biased toward reactive aggression due to their inescapable and involuntary nature (Kudryavtseva et al. 2011, 2014). To overcome these limitations, recent work has focused on behavioral procedures that examine both reactive and appetitive aggression using operant conditioning tasks, and this approach is instrumental for understanding the neurobiological differences underlying different types of aggression.

Individual variability in inbred and outbred lines

A technical confound associated with preclinical aggression research is the relative lack of innately expressed aggression exhibited by commonly used inbred mouse strains relative to outbred strains (Jones and Brain 1987). Unlike outbred strains, inbred mice often require significant experimenter manipulation, either in the form of extended social isolation (Banerjee 1971) or repeated social instigation (Kudryavtseva et al. 2014; Covington et al. 2018), to exhibit significant levels of aggression towards conspecifics. To overcome this confound, preclinical aggression research, and especially studies of appetitive aggression, has focused on the use of outbred mice (Chia et al. 2005) that exhibit a spectrum of innate aggression behavior (Golden et al. 2016). While inbred mice are typically preferred for many research applications, recent meta-analysis shows that inbred strains do not have greater trait stability than outbred mice, and data from outbred mice may be more generalizable across conditions and populations (Tuttle et al. 2018).

We and others have used CFW or CD-1 outbred mice as these strains display several naturally occurring aggression phenotypes ranging from compulsive addictive-like aggression seeking to aggression avoidance (Golden et al. 2016, 2017b). Unfortunately, the use of outbred strains precludes the use of genetically defined Cre recombinase-based methods. To overcome this, we have introduced a hybrid breeding strategy using the F1 hybrid offspring derived from an inbred strain of interest and an outbred CD-1 (Golden et al.

2017a; Aleyasin et al. 2018). Following behavioral phenotyping for baseline reactive and appetitive aggression levels between inbred and hybrid populations, as well as molecular phenotyping for appropriate transgene expression, this approach introduces genetic selectivity to innately aggressive mice populations. We propose that this approach, in combination with more traditional but time-consuming backcrossing of Cre lines onto outbred strains or the impending development of outbred CRISPR transgenic lines, will provide a strong foundation for aggression research moving forward.

Unconditioned versus conditioned aggression

The conditioned place preference (CPP) procedure has historically been used to evaluate the rewarding effects of drugs and alcohol (Beach 1957; Mucha et al. 1982). In this procedure, one distinct context is paired with the conditioned stimulus while another context is paired with the unconditioned stimulus. During a subsequent stimulus-free test, the laboratory animal chooses to spend time within the conditioned or unconditioned context. An increase in preference for the conditioned stimulus paired context is indicative of rewarding effects (Bardo and Bevins 2000). More recently, CPP procedures have been developed to assess the relative reward of affiliative social interactions (Panksepp and Lahvis 2007; Dölen et al. 2013; Goodwin et al. 2018).

Similarly, based on studies in female Syrian hamsters (Meisel and Joppa 1994) and male outbred OF-1 mice (Martínez et al. 1995), we have adapted a CPP procedure in combination with the resident-intruder social defeat procedure (Miczek et al. 1982; Kudryavtseva et al. 1991; Golden et al. 2011) to study aggression reward in CD-1 mice (Golden et al. 2016). Using this method, we first categorized unconditioned reactive aggression in adult CD-1 male mice through repeated daily resident-intruder assays with adolescent submissive C57BL/6J intruder mice. Mice that attacked the intruders during these screening assays (70%) were termed aggressors, while mice that did not attack (30%) were termed non-aggressors. We then evaluated conditioned aggression motivation using the aggression CPP assay. Mice that displayed aggression during the initial screening tests developed aggression CPP, while those that did not attack the intruders demonstrated conditioned place aversion (Golden et al. 2016). In a series of follow-up experiments, we parametrically explored the aggression CPP phenomenon and observed several key findings. First, based on the observation that unconditioned reactive aggression falls along a continuum in CD-1 mice, we characterized individual differences in aggression CPP by testing a third phenotype, termed “variable aggressors,” composed of mice that performed inconsistently when repeatedly exposed to the resident-intruder procedure (Golden et al. 2017a). The variable aggressive mice exhibited significant, although weaker, aggression CPP, suggesting that repeated unconditioned

aggression experiences can transform non-rewarding aggressive encounters into a rewarding experience. Second, aggression CPP is a learned phenomenon that can be acquired even by initially non-aggressive mice. Specifically, we exposed a large cohort of non-aggressors to 10 days of repeated resident-intruder testing and found that 50% transitioned to variable aggressors and exhibited aggression CPP. Lastly, aggression CPP is persistent, lasting several weeks following the final condition session (Golden et al. 2017a).

Notably, the portion of CD-1 mice that fail to show aggression is small and has not been the focus of study due to the difficulty in screening animals and filling group sizes. Advances in automated behavioral tracking, however, may be able to distinguish these phenotypes at an earlier age, alleviating these restrictions, as will be discussed in the second part of this review.

Addiction-like aggression behavior and relapse

Beyond measures of reactive aggression and conditioned aggression reward, several groups have developed operant tasks that measure appetitive aggression. The Miczek group has designed an operant conditioning panel including active and inactive nose poke ports that can be introduced into the home cage of outbred CFW mice. The resident mice are trained to nose poke on fixed ratio and fixed interval schedules of reinforcement to attack intruders (Fish et al. 2002, 2005; Bannai et al. 2007). The development of operant aggression tasks has allowed the pharmacological decoupling of aggression *seeking* versus aggression *consumption* behaviors. The Miczek group reported that the GABA_A positive allosteric modulator allopregnanolone increases operant response rates at lower doses than are required for increased attack behaviors (Fish et al. 2002), but that the effects are inhibited by the rise in corticosterone that are necessary for both operant responding and escalated aggression behaviors (Fish et al. 2005). While alcohol administration increases the motivation to fight, these effects are distinct from fighting performance and were not impacted by the antagonism of NMDA or AMPA receptors (Covington et al. 2018). 5-HT_{1b} receptor agonism was found to decrease attack intensity without changing operant responding (Bannai et al. 2007).

The Kennedy group subsequently replicated and extended these results, finding reliable operant aggression self-administration in mice under progressive ratio, differential reinforcement of low rate behavior, and variable ratio reinforcement schedules (Couppis and Kennedy 2008; May and Kennedy 2009). Local nucleus accumbens dopamine receptor 1 or 2 antagonism in Swiss Webster mice inhibited both operant responding and select attack behaviors (Couppis and Kennedy 2008), and we have found similar results following chemogenetic inhibition

of dopamine receptor (Drd) type 1, but not type 2, in hybrid F1 CD-1 $\times$ Drd1-Cre or Drd2-Cre mice (Golden et al. 2019a, 2019b). Extensions of these operant procedures have also shown that animals rapidly cease aggression self-administration when confronted with a non-submissive intruder (Falkner et al. 2016), further highlighting the necessity of winning in promoting aggression reward.

Pathological aggression in humans mimics cardinal features of drug addiction. Aggressive encounters are often sought despite severe negative consequences, pathological aggression develops only in a minority of individuals (Lacourse et al. 2002; Provencal et al. 2015), and recidivism rates for violent offenders who are incarcerated for repeat violent offenses are similar to the relapse rates of individuals who take addictive drugs (Anthony et al. 1994; Ducrose et al. 2014). Such cardinal features have been reverse-translated from the clinic to create animal models of addiction-like behavior. Deroche-Gamonet (2004) used a combination of operant procedures including fixed and progressive ratio tasks, as well as fixed ratio with cue-contingent shock punishment, within a rodent model of cocaine addiction that has high face validity to the DSM IV criteria (Deroche-Gamonet 2004). In a cohort of rats that initially showed equal levels of cocaine self-administration and sensitization, 17% of animals went on to show addiction-like indicators including difficulty stopping or limiting intake, high motivation for access, and continued use despite negative consequence. These measures also correlated with relapse propensity (Deroche-Gamonet 2004; Piazza and Deroche-Gamonet 2013). Based on the above considerations, we have developed a modified operant chamber to test relapse to aggression-seeking following (i) home cage forced abstinence (Pickens et al. 2011), (ii) voluntary choice-based abstinence (Caprioli et al. 2015), and (iii) punishment-induced abstinence (Krasnova et al. 2014; Marchant et al. 2019).

Using this approach, we have shown that preclinical addiction models can be used to identify the neural mechanisms controlling appetitive aggression and relapse, as well as pathological or compulsive manifestations of aggression (Golden et al. 2017b). About 70% of aggressive mice learn to lever-press for aggressive interactions, and using several gold-standard models derived from the preclinical addiction literature, we observed aggression relapse after forced abstinence, punishment-induced abstinence, or choice-based voluntary abstinence that persists long after the last aggressive act. Through cluster analysis of the aggression-related measures, we also identified a subset of mice that met criteria previously developed to denote compulsive addiction in rodent models (Deroche-Gamonet 2004). Specifically, the cluster analysis identified a subset of compulsive addiction-like aggressive mice (~

19%) that exhibited intense operant-reinforced attack behavior, decreased likelihood to select an alternative palatable food reward over aggression, heightened relapse vulnerability and progressive ratio responding, and resilience to punishment-induced suppression of aggression-reinforced operant responding.

Importantly, this study found that contingent punishment is effective for suppressing aggression-seeking behavior in the majority of aggressive mice, but not in those exhibiting compulsive aggression-seeking behaviors, and ultimately fails to prevent spontaneous recovery of aggression-seeking following extended abstinence in nearly all aggressive mice. These data are especially interesting in light of reports that in rodents, footshock elicits both unconditioned aggression (O'Kelly and Steckle 1939; Azrin et al. 1967) and Pavlovian conditioned aggression to a footshock-paired tone (Vernon and Ulrich 1966). However, work in non-human primates (Ulrich et al. 1969; Azrin 1970) and rats (Baenninger and Grossman 1969; Roberts and Blase 1971) have shown that mechanical pain or shock-induced aggression is suppressed by contingent, but not non-contingent, shock punishment. Further, non-contingent footshock inhibits aggression in dominant but not subordinate mice (Frischknecht et al. 1985). Together, and within the common context of "punishment as a tool to prevent aggression," these data suggest greater nuance and present areas for future study within the context of adaptive and maladaptive compulsive aggressive behavior.

Conclusions

The recent renaissance in aggression research has been assisted by the integration of carefully designed behavioral assays, derived from animal models traditionally used to study compulsive drug use and relapse in the addiction field, and modern neuroscience techniques such as chemogenetics (Coward et al. 1998; Armbruster et al. 2007) and optogenetics (Boyden et al. 2005). These manipulations have begun to help us understand the neural regions driving attack behavior and aggression salience (Miczek et al. 2001; Falkner et al. 2016; Han et al. 2017; Stagkourakis et al. 2018; Golden et al. 2019a), but the timescale of neural manipulations and recordings necessitate higher resolution behavioral scoring than hand scoring is able to consistently and reliably provide. Both aggression-seeking and consumption phases of behavioral assays will benefit from more rapid, reliable, and non-subjective quantification of behavior. Within the field of pharmacology, real-time quantification and behavioral prediction are necessary for closed-loop manipulations that can causally differentiate the behavioral sequence

constituents of reactive and appetitive aggressive behavior. Specific to current operant procedures, such approaches will allow a detailed understanding of the sequela of behavioral events that occur after an operant action is contingently reinforced with an intruder. All of these time-locked events, at previously unfeasible time scales and objectivity, may provide additional information on the motivational state of the aggressor and the neural mechanisms guiding this behavior.

Currently, aggression outcome measures are typically restricted to latency to first attack, or proportion of attack versus non-attack behaviors. Such simplifications leave valuable ethological data unexplored, but these analytical omissions have been necessary due to the complexity of aggression behavior and the length of time and training required to accurately manually score assays. Species typical attack behavior in mice generally includes bites directed toward the back and flanks of intruder mice, and decreased attack behavior upon displays of submission by an intruder. Additionally, shifts toward more damaging bite locations or continued attack despite submission can both be indicative of escalated aggression (Takahashi and Miczek 2014; Newman et al. 2018). Automated behavioral classifiers can simultaneously measure these and other behaviors, of both the resident and intruder animal, allowing for a richer ethological continuum. The classifiers—once validated—can be curated and disseminated through online repositories, which can eliminate observer variability and present significant opportunities for cross-site standards within behavioral analyses. Accessible classifiers that are fast, operational across labs, and interpretable and explainable (Table 1) also have obvious benefits for transparency and scientific rigor.

We and others propose that open-source machine learning techniques will allow for rapid, high-throughput explorations of the incredible nuance of these behaviors without sacrificing accuracy. The integration of these automated behavioral classifiers will overcome the hurdle of hand scoring which has bottlenecked the field since its emergence.

Machine learning

Embracing machine learning

Automated video assessments can exceed human performance (Gris et al. 2017), and behavioral classifiers increase throughput and consistency (Schaefer and Claridge-Chang 2012) in addition to reducing human bias and anthropomorphism in scoring (Robie et al. 2017). The development of specialized machine learning behavioral classifiers have uncovered previously unknown behavioral repertoires in animals including mice, drosophila, bats, and *Caenorhabditis elegans* (Vogelstein et al. 2014; Wiltschko et al. 2015;

Rudolf et al. 2019; Zhang and Yartsev 2019). Markerless pose estimation and behavioral classification algorithms are rapidly improving, and many of the requisites are met for expanding machine learning approaches to classify complex social behavior. A key component to the adoption and success of these efforts is promoting open-source packages that can be easily and widely adopted throughout the behavioral neuroscience community.

There are several requirements for the ready adoption of machine learning approaches:

- (1) **Ease of use.** The field of computational neuroethology is advancing at rapid pace, such that the ability of many behavioral neuroscientists to take advantage of these pipelines is curtailed by their skill behind a command line. Many of the recent additions to the field such as DeepLabCut (Mathis et al. 2018), however, provide graphic user interfaces (GUIs) that allow the user to avoid setting hyperparameters or organizing projects via command line entries. Programs with easy installation and approachable GUIs will open doors for labs without easy access to computationally inclined individuals.
- (2) **Generalizability.** Behavioral assays within and between labs can often be filmed under different lighting conditions, with animals and backgrounds shifting in hues, using variable hardware and acquisition parameters such as frame rate, resolution, color scales, and video format. Many current implementations of machine learning for animal behavior (Table 2) require specialized hardware setups that can be prohibitively expensive or challenging to implement and scale-up for high-throughput behavioral assays, and often do not generalize well from one preparation to another (Anderson and Perona 2014). Machine learning approaches that can use standard, readily accessible recording hardware without the need for specialized builds are vital for generalizability and comprehensive implementation.
- (3) **Cost management.** Current pose estimation and behavioral classifier programs benefit from the use of depth cameras, multiple filming angles, specialized acquisition hardware, and proprietary software. Access to high-end analytical platforms, either in the form of cloud-based or local solutions, is a requirement for generating both tracking and classifier models. These costs propel computational neuroethology out of the reach of many behavioral neuroscience labs. However, the behavioral assays commonly utilized for the study of aggressive behavior are well defined and often recorded with a single camera, making it possible to generate and share

Table 2 Currently available open-source tracking or behavioral classification software

Name	Validated species	Multiple animals?	Main output	Behavioral classifiers validated in original publication	Mechanism	Software website
Citax (Branson et al. 2009)	Drosophila	Yes	Behavior	Walk, stop, sharp turn, crabwalk, backup, touch, chase	Computer vision	http://citax.sourceforge.net/
CADABRA (Dankert et al. 2009)	Drosophila	Yes	Behavior	Lunging, tussling, wing threat/extension, circling, chasing, copulation	Supervised	http://www.vision.caltech.edu/cadabra/
Autotyping (Jhuang et al. 2010)	Mice	No	Behavior	Home-cage behaviors: drink, eat, groom, hang, micromovement, rear, rest, walk	Supervised	http://cbl.mit.edu/software-datasets/mouse/
JAABA (Kabra et al. 2013)	Drosophila, mice	Yes	Behavior	Flies: walk, stop, crabwalk, backup, touch, chase, jump, copulation, wing flick/grooming/extension, righting, center pivot, tail pivot. Mice: follow, walk.	Supervised	https://sourceforge.net/projects/jaaba/files/
MiceProfiler (de Chaumont et al. 2012)	Mice	Yes, if no obstruction	Behavior	Head-head contact, anogenital contact, side by side, approach, leave, follow, chase	Physics engine	http://icy.biomaganalysis.org/plugin/mice-profiler-tracker/
Unnamed* (Burgos-Artiz et al. 2012)	Mice	Yes	Behavior	Approach, walk away, circle, chase, attack, copulation, drink, eat, clean, sniff, rear	Supervised	
Unnamed* (Matsumoto et al. 2013)	Rats	Yes	Behavior	Rearing, head-head and head-hip contact, approach, leave, follow, mount, intromission, ejaculation	Physics engine	See supplementary material
MotionMapper (Berman et al. 2014)	Flies	Yes	Behavior	Grooming: wing, leg, abdomen, head. Wing waggle, running	Unsupervised	https://github.com/gordonberman/MotionMapper
Unnamed* (Hong et al. 2015)	Mice	If distinct	Behavior	Attack, close investigation, mounting	Supervised	
MoSeq* (Wilschko et al. 2015)	Mice	No	Behavior	Dart, micromovement, pause, rear, walk	Unsupervised	MTA and private repository
Unnamed (Unger et al. 2017)	Mice	Yes	Behavior	Direct contact, nose to nose, anogenital sniffing, following, mating, self-grooming	Unsupervised	
DuoMouse (Arakawa et al. 2017)	Mice	Yes	Behavior	Sniffing, following, indifferent	Unsupervised	http://www.mgt-lab.jp/DuoMouse.html
B-SoID (Hsu and Yttri 2019)	Mice	No	Behavior	Pause, rear, groom, sniff, orient left/right, locomotion	Unsupervised	https://github.com/YttriLab/B-SoID
SimBA (Nilsson et al. 2020)	Mice, rats	Yes	Behavior	Attack, pursuit, lateral threat, anogenital sniff, mounting, upright submissive, allogrooming, flee, scramble, boxing, approach, avoidance, drink, clean, eat, rear, walk away, circle	Supervised	https://github.com/simbalab/simba
Stytra (Stih et al. 2019)	Zebrafish larvae	No	Closed-loop tracking and stimulus admin. (real time)		Computer vision	https://github.com/portugueslab/stytra
ToxID (Rodriguez et al. 2017; also see ToxTrac Rodriguez et al. 2018)	Fish, ants, mice	Yes	Individual tracking in groups (real time)		Computer vision	https://sourceforge.net/projects/toxtrac/
MoST (Thanos et al. 2017)	Mice	Yes	Individual tracking in groups		Computer vision	
idTracker.ai (Romero-Ferrero et al. 2019, also see idTracker Pérez-Escudero et al. 2014)	Drosophila, fish, ants, mice	Yes	Individual tracking in groups		Unsupervised	http://idtracker.ai/
LocoMouse* (Machado et al. 2015)	Mice	No	Kinematics		Supervised	https://github.com/careylab/LocoMouse
DeepLabCut (Mathis et al. 2018)	Drosophila, mice, horses, humans, fish	Yes	Pose		Supervised	http://www.mousemotorlab.org/deeplabcut

Table 2 (continued)

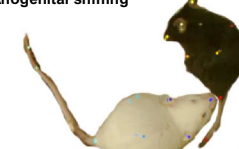
Name	Validated species	Multiple animals?	Main output	Behavioral classifiers validated in original publication	Mechanism	Software website
LEAP (Pereira et al. 2019) Successor: Social LEAP Estimates Animal Pose (SLEAP)	Drosophila, mice	Yes, through SLEAP	Pose		Supervised	https://github.com/talmol/leap https://github.com/murthy/lab/sleap
DeepFly3d* (Günzel et al. 2019)	Drosophila, humans	No	Pose		Unsupervised	https://github.com/NeLy-EPFL/DeepFly3D
DeepPoseKit (Graving et al. 2019)	Drosophila, locusts, zebras	If distinct	Pose (real time)		Supervised	https://github.com/jgraving/DeepPoseKit
Sensory Orientation Software (Gomez-Marín et al. 2012)	Drosophila, mice, fish	No	Posture and trajectory (real time)		Computer vision	https://sourceforge.net/projects/sos-track/

*Next to software name indicates multiple cameras or specialized equipment necessary

“base” tracking and classifier models that preclude the need for individual labs to heavily invest in specialized hardware or commercial software.

- (4) Accuracy. At a minimum, automated classifiers need to meet the accuracy of manual annotation conducted by highly trained observers. Regardless of the adoption of machine learning techniques, the ability to consistently recognize behaviors of interest in model species is integral to behavioral research, and the experimenter aptitude to recognize significant behavioral events determines the success in machine model generation, tuning, and validation. We have found, however, that even observers trained to high rates of inter-rater reliability on strict operational definitions of behaviors (Table 3) may occasionally miss shorter behavioral bouts when scoring long video recordings (e.g., 20k+ video frames). Machine learning approaches provide the advantage of never tiring, and in our experience, classifiers consistently detect shorter behavioral bouts missed by human observers. While ideal algorithms for specific tasks are typically dependent on data structure and the computational knowledge within a lab, we and other labs have found that supervised approaches (Nilsson et al. 2020), such as random forest machine learning algorithms, provide an appropriate balance of ease of use, interpretability, explainability, and accuracy that often exceeds human annotation accuracy.
- (5) Easy expansion of training sets. Ethological datasets provide an interesting challenge for machine learning. Training data provided to algorithms must be representative of the testing data, but behavioral assays are often conducted in discrete cohorts with unique properties. Training sets should contain a balanced mix of control and treatment videos, and diverse and unstable environmental influences (cage change schedules, seasonally noisy HVAC systems, etc.) must be considered as potential confounds if not represented in the original classifier training data. Specific to pharmacological research, when individual animal movement characteristics and social behaviors may be altered by pharmacological interventions, it is not feasible to create representative training sets from the outset. We and others have approached this challenge by creating the ability to easily add additional hand annotated videos (several short videos are typically sufficient) to training sets, which can then rapidly be used to create updated, iterative, behavior classifiers, and animal tracking models.
- (6) Interpretability. Interpretable methods are of utmost importance for reproducibility, transparency, and rigor (Table 1). Methods for creating accurate classification algorithms in behavioral neuroscience should provide transparent processes allowing for

Table 3 Operational definitions for aggression behaviors exhibited between resident CD-1 mice and intruder C57B6/J mice

Classifier	Description	Start frame	Duration of behavior	End frame
Attack 	Clear physical antagonistic interaction initiated by the Resident mouse. Brief pauses can occur (equal or less than 133ms, or 4 frames at 30fps) as long as the Resident mouse remains oriented toward the Intruder.	First frame when the Resident mouse makes physical antagonistic contact with the Intruder. Typically, this is characterized by outstretched Resident forepaw(s) contacting the Intruder, while the Resident has an open mouth to initiate a bite. Can also be characterized by the first frame of a slap or quick bite without the forepaws being outstretched.	Short breaks (equal or less than 133ms, or 4 frames at 30fps) may be present in attack behavior, as long as Resident is still oriented toward Intruder. Attacks can include tussling, biting, boxing, and corralling as part of the attack bout.	First frame when Resident mouse orients away from Intruder. Typically, this is a slight turning of the head to look in a different direction, followed by a relaxation of the body and moving away from the Intruder.
Anogenital sniffing 	Resident mouse is sniffing the anogenital region of the Intruder. Resident must be sniffing at base of tail, not further up on back or on legs.	First frame when the Resident mouse is clearly sniffing anogenital region of Intruder, rather than side, back, or leg.	Uninterrupted sniffing of anogenital region.	First frame when Resident mouse moves head away from anogenital region, either to move away from Intruder or to sniff non-anogenital region.
Lateral threat 	Resident mouse is in close proximity (typically less than one body length away from the Intruder) to face of Intruder with back arched and side displayed toward Intruder. Ears are often pinned with shoulder and side of face nearest to Intruder mouse tilted slightly toward ceiling.	First frame when Resident mouse orients side to Intruder and tilts front of body toward Intruder.	Resident will often circle Intruder or move front half side to side in front of Intruder, feigning attacks prior to actual attack.	First frame when lateral threat posture is dropped. Animal will shift head away to look at other target or will begin an attack. If animal does not attack, the back posture will relax.
Tail rattle 	Resident's tail is curved at two or more points (typically in an "S" shape) and showing rapid back and forth movement.	Defined as rapid back and forth movement of tail. In Resident mice this manifests as curling of tail into "S" shape, where next frames show inversion of shape and rapid whipping back and forth of the tail.	The tail continues to move rapidly back and forth and can occasionally straighten if tail is moving fast enough. Tail side to side movement is continuous.	First frame when the curled side to side movement of the tail stops. The tail typically relaxes and straightens out.
Pursuit 	Resident mouse is following in the Intruder's path as the Intruder moves away from the Resident.	First frame when the Resident is moving toward Intruder as Intruder moves away. Typically, this is characterized by the Resident follows directly after, or when the Intruder walks past the Resident and the Resident markedly changes directions to pursue the Intruder in the Intruder's path.	The Intruder is specifically moving away from the Resident. The Intruder is not sniffing the arena, foraging, etc. Intruder typically moves in a straight line until it reaches the edge of the arena, at which point it will turn to face and watch the Resident. Resident is moving along the same path as the Intruder without sniffing or foraging in the arena.	First frame when either mouse stops moving, or the Resident deviates from path of the Intruder.

clear understanding of how classifications are made and, ideally through GUI control, the capacity to titrate model parameters to operationalized standards. This may include the ability to visualize decision paths and the importances of individual independent variables or features for the classification result, and how the data and features are balanced within the model. These concepts have however—traditionally—not always been favored within computer and data sciences (Rudin 2019) and may introduce additional challenges when creating accessible methods within preclinical social behavior and aggression research.

Aggression behavioral assays provide excellent insight into the current capabilities and limitations of machine learning algorithms due to the numerous technical challenges they present. Behavioral components often include a very rapid succession of bites, lunges, pursuits, anogenital sniffing, tail rattles, lateral threats, boxing, and occasionally head shakes and prosocial behaviors such as grooming.

These sequences can be further reduced to levels of granularity which may or may not convey important information (e.g., face bite versus flank bite). Furthermore, on video, animals frequently occlude each other and significantly change their body shapes as they aggress, creating difficulties for pose estimators. To our knowledge, two groups have published work using automated classifiers for evaluating aggression behaviors. Hong et al. (2015) was able to classify attack behaviors in male mice of different colors using a multi-camera apparatus, and Burgos-Artiz et al. (2012) achieved accuracy rates of ~61% using a multi-camera apparatus with two unmarked mice of the same color. Our lab has recently released Simple Behavioral Analysis (SimBA) (Nilsson et al. 2020), an automated behavioral classification technique for aggressive, submissive, and social behaviors that was built upon pose estimation and the six principles outlined above.

The remainder of this review will discuss the current state of machine learning for social behavior classification in rodents in terms of aggression behavior, and the progress of the field toward meeting the goals discussed above.

Table 4 Recommended potential workflow and hardware specifications for performing supervised machine classification of complex social behaviors in experimental animals using open-source software. For more information, see the GitHub repositories of the Golden lab

Workflow step	Description	Required Hardware	Open-source tools
(i) Record and pre-process videos	Record videos at a quality where behaviors of interest are apparent to trained human observers. Crop/clip video, downsample resolution, and re-sample frame rates, if required.	Camera allowing recordings at ≥ 30 fps and resolutions $\geq 640 \times 480$ (i.e., standard webcam).	FFmpeg OpenCV
(ii) Track animals	Estimate the location of the animals and their body parts in each frame of the videos with convolutional neural network.	CPU supported by graphical processing unit (GPU), through a cloud solution (e.g., Google Colaboratory, AWS SageMaker, Microsoft Azure) or locally (e.g., NVIDIA Rtx2080Ti).	DeepLabCut DeepPoseKit LEAP YOLO mask RCNN
(iii) Supervised decision ensembles	Extract feature sets, annotate behaviors, build/tune algorithms, evaluate classification performance.	Processing time benefits from higher-end CPUs and solid-state drives (SSDs). We recommend a CPU with ≥ 8 cores (e.g., Intel i9-9900K).	scikit-learn imblearn OpenCV Shapely
(iv) Visualize/analyze classification results	Visually evaluate machine classifications. Perform descriptive statistics of behavior sequences, their durations and patterns.	Processing time benefits from higher-end CPUs and solid-state drives (SSDs). We recommend a CPU with ≥ 8 cores (e.g., Intel i9-9900K).	OpenCV FFmpeg matplotlib standard python library

Accurate behavior classifications require arenas and recording conditions that allow detailed animal tracking of relevant body parts through convolutional neural network architectures (workflow step ii). Generally, if the recording conditions allow the behaviors of interest to be reliably observed by human annotators in recorded videos, then the behaviors of interest can also be accurately annotated by machine classification techniques

Supervised versus unsupervised learning

Many machine learning techniques for behavioral analysis first identify the position of animals frame-by-frame in a video via pose estimation or background subtraction. In particular, accurate and rapid pose estimation via recently developed and generalizable convolutional neural network architectures—accessible through packages such as DeepLabCut (Mathis et al. 2018), DeepPoseKit (Graving et al. 2019), and LEAP (Pereira et al. 2019)—provide a platform for generating the pose estimation data needed to create rich feature sets required for machine classification of complex social behavior. For example, the predicted poses may be used to estimate distances between animals, their velocities, angles, and accelerations across rolling windows, and this may correlate with human annotated instances of aggressive behavior and used in further modeling techniques (Table 4). These positional data are then analyzed to cluster statistically similar images that the program either identifies as a predefined behavior (supervised learning) or the cluster is studied by the experimenter who then adds behavioral labels post hoc (unsupervised learning).

Unsupervised learning has successfully classified behaviors in tunicates, drosophila, individual mice, and pairs of mice after cropping the animals into individual videos (Vogelstein et al. 2014; Wiltschko et al. 2015; Klibaite et al. 2017; Rudolf et al. 2019; Dolensek et al. 2020). These

techniques can be tremendously powerful in identifying the inherent structure present in behavior. Because the user does not predefine behaviors of interest, these algorithms have been proposed to be less biased than supervised techniques. Thus far, unsupervised techniques have resulted in important advances in mouse ethology including the identification of facial expressions corresponding to neuronally separable emotion states (Dolensek et al. 2020) and novel sub-second behavioral structures (Wiltschko et al. 2015). The program B-SOiD has also combined unsupervised t-SNE clustering and support vector machines to great success in classifying naturally occurring behaviors in single mice, within a pipeline that is readily accessible to non-specialists (Hsu and Yttri 2019). The power to identify new behavioral repertoires is alluring and has the potential to significantly advance the field of ethology. Unsupervised machine learning tools are notoriously difficult to tune and interpret, however, and we propose that they should not be a first-line option for generalization to labs without significant computational experience.

Training an unsupervised behavioral classifier generally involves feeding many unlabeled video frames into a user-defined algorithm which then identifies and separates images into behavioral clusters based on user-defined mathematical requirements for differences and similarities. The de Bivort lab has conducted one of the only large-scale direct comparisons of

unsupervised learning techniques in classifying animal behavior (the leg movement of flies) and provides an excellent overview of unsupervised learning techniques (Todd et al. 2017). A typical pipeline for clustering often involves data pre-processing, dimensionality reduction, and cluster assignment (Todd et al. 2017; Datta et al. 2019), though see (Wiltchko et al. 2015).

Each of these steps is modular and can be performed using any of several different algorithms. Pre-processing can include transformations such as time frequency analysis, vector normalization, and wavelet transformation, while dimensionality reduction is often accomplished via techniques such as principal component analysis (PCA) or t-distributed stochastic neighbor embedding (t-SNE). The final clustering algorithms, such as k-means clustering or Gaussian mixture modeling, are used to group the data into clusters that researchers can apply behavioral labels towards (Todd et al. 2017). The flexibility of unsupervised learning approaches at each step makes them very powerful in adapting to different datasets, but tuning parameters requires careful consideration.

Importantly, unsupervised learning techniques do not have built-in performance metrics. Tuning them to high perceived accuracy (e.g., all user-defined attacks are also defined by the algorithm as an attack) is in essence training the model to meet particular output benchmarks, i.e., creating a supervised algorithm (Brown and de Bivort 2018). Some classes produced by an unsupervised approach may also be difficult for users to define as recognizable behaviors. Todd et al. (2017) propose evaluating networks via parameters such as minimum dwell time within a behavior class and the reliability of the algorithm when trained multiple times on the same data set (Todd et al. 2017; Brown and de Bivort 2018). These parameters acknowledge the real-world constraints on behavior and the need for replication while imposing less bias onto the model in terms of expected output. We propose that any implementation of algorithmic features during unsupervised training should be accompanied by a description of the experimental objectives driving their necessity.

Supervised learning algorithms require greater user oversight in that users are required to annotate training data that the classification system can use to learn to correctly label images. Several pose estimators are validated for use in *drosophila* (Branson 2014; Berman et al. 2014; Günel et al. 2019), and tools such as DeepLabCut (Mathis et al. 2018), LEAP (Pereira et al. 2019), and DeepPoseKit (Graving et al. 2019) have been validated in mice and other non-human mammals. DeepLabCut and LEAP use variations on the supervised learning technique of deep neural networks to estimate animal pose, while the DeepPoseKit uses a novel Stacked DenseNet algorithm. For in-depth explanations of these estimators, see Mathis et al. (2018), Graving et al. (2019), Pereira et al. (2019), Nath et al. (2019), and Mathis and Mathis (2019). In many of these packages, users label body parts at set coordinates, and these training labels are used to generalize

computed labeling rules to future images. Pose estimators are powerful tools in identifying finer points of animal activity, and are able to identify phenotypes not detected with simply trajectory tracking (Hong et al. 2015). Notably, new packages can include project management and image annotation GUIs, which greatly enhances the user experience and has resulted in more widespread adoption.

Supervised learning techniques can be used for both pose estimation and behavioral classification, but unlike unsupervised algorithms, can only classify predetermined behaviors and cannot identify novel behavioral repertoires. Training accurate supervised machine learning classifiers requires the careful, frame-by-frame annotation of a behavior of interest. Human variability and annotation mistakes, such as erroneous labeling of aggression events as non-aggression events, can generate significant noise that propagates to future machine analyses (Frenay and Verleysen 2014). It is therefore important to have precise behavioral operational definitions that encompass and exclude the behavioral events and non-events, respectively, and we present examples of how such precise operational definitions may look in Table 3. While such precise descriptions are not current standard laboratory practice, and functionally translate to extra time hand-annotating videos, they can promote the introduction of standardized cross-laboratory definitions through shareable classifier repositories that ultimately increase replicability as discussed previously.

Several behavioral classifiers using supervised machine learning techniques have been validated in mice, including JAABA (Kabra et al. 2013), autotyping (Jhuang et al. 2010), and SimBA (Nilsson et al. 2020). Each uses simple, single camera setups with significant differences in underlying machine learning approaches to create predictive classifiers. JAABA classifies walking and following behavior in groups of same colored mice using a boosting ensemble algorithm (Kabra et al. 2013), and autotyping uses a combination of hidden Markov model and support vector machines to classify a variety of home cage behaviors in single-housed mice (Jhuang et al. 2010). SimBA uses random forest classifiers to classify aggressive, defensive, and other social behaviors in mice following pose estimation (Nilsson et al. 2020). Some programs also use multiple cameras, depth cameras, or other specialized setups for behavioral classification. The Anderson Lab has developed a random forest classification technique which can identify attack, mounting, and investigation behaviors in differently colored mice using multiple depth sensing cameras (Hong et al. 2015). Boosting ensemble algorithms have also been used to classify social movement, attack, copulation, exploratory, and drinking and eating behaviors with integrated top and side view filming (Burgos-Artizzu et al. 2012).

Whether implementing supervised or unsupervised machine learning techniques, there is a strong argument for using algorithms that are easily interpretable (Rudin 2019). Black

box models have been helpful in image analysis and uncovering new behavioral patterns, but the inability to readily understand the underlying assumptions of algorithms is a detriment. Black box models—not unlike prevalent behavioral protocols—can appear to be valid and accurate (strong face validity) while measuring factors unrelated to the phenomenon it claims to measure (weak construct validity). Examples of such fallacies, with more trivial consequences, would include convolutional neural networks that “successfully” discriminate husky dogs and wolfs based predominantly on only the presence or absence of snow in the image (Ribeiro et al. 2016), or networks generating “successful” gender classifications from iris texture images through the presence or absence of mascara (Kuehlkamp et al. 2017). More seriously, black box models are notorious for aggregated racial and social-economic biases (Mittelstadt et al. 2016; Rudin 2019; Obermeyer et al. 2019). While clearly a different class of concern than what is anticipated in preclinical behavioral computational neuroethology, these cautionary tales remain relevant and important. Explainable and interpretable models should thus be adopted as best scientific practice (Rudin 2019) to allow for user oversight of the parameters guiding decision processes.

While supervised learning techniques require users to predefine behaviors of interest, they are often much easier to tune and interpret than unsupervised learning techniques, particularly for labs without statisticians or computer scientists on staff. In the spirit of ease of use, generalizability, and accuracy, we propose that supervised learning techniques are a good starting point for the automated examination of aggressive behaviors.

Common classifying algorithms for supervised learning

Within supervised learning, there are many different algorithms from which to choose, and no method is universally superior to the others (Hand 2006). The appropriate selection depends on the structure, noise, and biases within each dataset, as well as the threshold for acceptable training duration and willingness to troubleshoot parameters (Hand 2006; Anderson and Perona 2014; Egnor and Branson 2016; Gris et al. 2017; Akay and Hess 2019). Common techniques for behavioral and image classification include variations of neural networks, support vector machines, gradient boosting machines, random forests, and hidden Markov models. All of these algorithms have potential benefits and may outperform others on specific datasets, but we and others have found that random forest classifiers provide high accuracy classification (Breiman 2001; Liaw and Wiener 2002; Nilsson et al. 2020) with the added benefit of

interpretability and being both easy to tune and robust against overfitting. Here, we describe several of these algorithms and highlight the potential pros and cons they possess regarding the classification of aggression-related behaviors.

Neural networks (Hopfield 1982; LeCun et al. 1989)

Neural networks or convoluted neural networks are typically used for large datasets with a high number (tens of thousands) of features and observations, as well as noise. Features may include the trajectory of individual body parts, Euclidean distances, and body part movement over a small frame of time. Users provide an input layer of data to the neural network and define the desired categories in the output layer. The network then computes additional hidden layers between the input and output layers to correctly classify input data into the appropriate output categories or classifiers. Hidden layers are advantageous as specific features are not user-selected for inclusion within the algorithm, but the cost is an absence of information on what the resulting classifications are ultimately based on. Neural networks can be challenging to interpret and tune, and similar performance can be achieved with other methods (Lietman et al. 1999; Nitze et al. 2012; Liu et al. 2013). Many programs, including the pose estimators described above, successfully use deep neural networks for pose and/or behavioral classifications (Karpathy et al. 2014; Krizhevsky et al. 2017; Mathis et al. 2018; Pereira et al. 2019). Due to the lack of information regarding feature weights, we propose that more transparent algorithms which perform at similar levels are preferable for behavioral classification derived from tracked features.

Random forests (Breiman 2001; Liaw and Wiener 2002)

Random forests are a type of ensemble algorithm. Ensemble algorithms are composed of many independently trained weak models which are combined to make strong predictions. Random forests rely on the bootstrapping of both (i) subsets of data and (ii) predefined features to make a powerful forest of decision trees than can then vote on the classification of a behavior. For example, a decision tree in a random forest starts with a subset of data (e.g., 500 frames of 50,000) then uses a subset of the features (e.g., 10 out of 100 features) to split the data into yes/no classifications. Random forests are useful because each branch is created using the feature which provides the most information, and by combining these data from the forest, it is possible to determine which features were most important for creating the behavioral classification. This technique is computationally rapid as the trees are independent and can be built in parallel, and they are unlikely to overfit data based on noise (Breiman 2001).

One weakness of random forests is their inability to natively support biased datasets. These datasets are common in behavioral videos, in which most frames do not contain the behavior of interest. For example, a 5-min video filmed at 80 fps with 10 total seconds of tail rattle would contain 23,200 frames with no tail rattle and only 800 frames with tail rattle. Constructing a random forest with these data as is would mostly construct trees with no or very few instances of tail rattle frames. In order to train the random forests more robustly, over-sampling and/or under-sampling techniques (Chawla et al. 2002; Batuwita and Palade 2013) can be used to balance the data.

Gradient boosting machines (Freund and Schapire 1997; Friedman et al. 2000; Friedman 2001)

Gradient boosting machines build decision trees iteratively, attempting to fix erroneous classifications after each node split by finding a different feature which better classifies incorrectly classified data. Gradient boosting can be more robust than random forests if tuned correctly but are more prone to overfitting data and incorrectly identifying noise, such as pose misestimation, as legitimate behavioral data. While pose estimation is often highly accurate, there are outliers that may cause overfitting when fed into a gradient boosting machine. With more accurate pose estimation and highly consistent filming conditions across labs, gradient boosting machines may provide superior classification than random forests, but it is difficult to currently meet these conditions and maintain generality of video recording conditions. While the use of gradient boosting machines is a good goal, their widespread adoption may depend on how well “base” models, and their underlying recording acquisition parameters, are standardized across laboratory recording environments.

Support vector machines (Cortes and Vapnik 1995)

Support vector machines are frequently used and work best with small datasets containing few outliers (Hsu and Lin 2002; Xu 2006). Using multiple tuning parameters (kernel, regularization, gamma, and margin), support vector machines identify clusters of similar and dissimilar data in multidimensional space and find the regression line that best separates clusters based on the training data. Understanding linear algebra and principle component analysis, in addition to having appreciable patience with tuning parameters, is helpful when using support vector machines. Like random forests, support vector machines do not work well natively with biased datasets, which can similarly be overcome through re-balancing and over- and under-sampling techniques.

There are few comparative studies of classification algorithms due to lack of generalization to other datasets, but in those that exist, support vector machines do not typically perform as well as random forests or gradient boosting machines

(Caruana and Niculescu-Mizil 2006; Caruana et al. 2008). Due to the larger number of tuning parameters involved in support vector machines and the similar or superior classification performance by random forest classifiers, we suggest that random forests may be a better starting point for creating generalizable and easily adopted classifiers.

Hidden Markov models (Rabiner and Juang 1986)

Hidden Markov models (HMMs) can be integrated with supervised or unsupervised techniques and excel at finding patterns in small sequences of data. These models have recently been extended for use in behavioral classifications due to the similarities from frame to frame of video and the inherent structure of behavior (Carola et al. 2011; Wiltschko et al. 2015; Arakawa et al. 2017). HMMs use hidden weights to predict the probability of a transition from one state to another. For example, if a resident is biting the flank of an intruder in one frame, it is unlikely to be grooming its face in the next frame. As indicated by the name of the algorithm, behavioral classifications depend on a hidden state. Anderson and Perona (2014) raise an interesting point that HMMs may be valuable in objectively measuring behaviors resulting from the “emotion state” of an animal. HMMs have been used in rodent aggression studies to investigate the effects of context (intruder behavior) on the aggression behavior of a home cage resident to parse species typical versus escalated aggression phenotypes (Haccou et al. 1988; Natarajan et al. 2009), and may provide an interesting path forward for understanding contextual aggression. While the hidden states can potentially be found via maximum a posteriori state estimations (Allahverdyan and Galstyan 2009), we propose that more inherently interpretable algorithms are preferable for initial behavioral classification.

Promising directions

We strongly reiterate that there is no single correct pipeline for machine learning classification of animal behavior, but rather that the most successful approaches will likely use multiple pipelines in parallel while taking advantage of their pros and attempting to diminish their cons. The powerful techniques discussed above are summarized in Table 5, and the choice between them often revolves around the applicable knowledge within a lab and which algorithm best classifies data during pilot testing (Table 5). To facilitate generalizability to new labs, however, random forests seem like an excellent option due to their relatively few input parameters, robustness to noise, relative interpretability and explainability, and high performance on a wide array of data. A future goal of automated behavioral analysis is long-term real-time tracking of individuals within groups in a naturalistic setting, and many

Table 5 Overview of advantages and limitations of select open-source options for tracking and behavioral classification

Technique	Open-source alternative(s)	Advantages	Limitations
Segmentation/identification networks for “fingerprinting”	idtracker.ai	• Can track large groups of unmarked individuals • Accessible GUI	• Requires GPU support • No live tracking • Only provides trajectory of animal centroid(s) • Not validated in mouse/rat video with high amounts of aggression • Users must generate their social behavior classifiers from tracking data, which may also require additional layer of pose estimation processing.
RFID/depth camera system	Live Mouse Tracker	• Built-in battery of social behavior classifiers • Group analysis of unmarked animals (≤ 4 mice) • No model training required • Live tracking (30 Hz) • Accessible SQL data format • True longitudinal analysis (days)	• Manufacture cost (~ \$2 k) • Manufacturing time • Dedicated platform (cannot be used to analyze historical datasets) • Arena/RFID implementation require <i>Animal Study Protocol</i> submissions/revision • Validated for mice only
Pose estimation	DeepLabCut LEAP DeepPoseKit	• Accurate tracking of user-defined body parts • Very large user-base • Active code development • Accessible GUI's (LEAP/DLC) • Accepts most recording environments/ and video formats	• CNN training time • Requires GPU access • Group and live tracking features not yet available • Users generate their social behavior classifiers from tracking data
Behavioral classification	SimBA JAABA	• Accessible GUI's • User-defined social behavior classifiers • Can accept any tracking data as input	• Requires third-party tracking (and GPU support) • No live classifications • Social classification restricted by the inability of third-party tracking solutions to track multiple similarly coat-colored animals

The techniques listed above are not mutually exclusive, and the most successful behavioral analyses likely depend on synergistic implementations of disparate computational methods

groups are currently working with these and other techniques to solve parts of this problem.

Distinguishing similarly looking animals has historically been a challenge for automated pose estimators, and ultimately problematic for the computational study of social behavior as many experiments require the use of nearly identical looking animals. Although not a problem when using outbred CD-1 mice as residents and C57 mice as intruders, due to their difference in coat color, this does preclude the generality of machine learning approaches to many rodent social behavioral assays. There are several recent programs that work to track the identity of individual and similarly appearing animals in groups over time, including ToxID, MoST, and idTracker.ai (Thanos et al. 2017; Rodriguez et al. 2017; Romero-Ferrero et al. 2019). Additionally, social tracking extensions of the

program LEAP (sLEAP) and DeepLabCut have recently been released, and provide pose estimation for multiple interacting animals (Mathis et al. 2018; Pereira et al. 2019) by taking advantage of algorithms such as optical flow (Brox et al. 2006), part affinity fields (Cao et al. 2017), and deep learning. The program idTracker uses separate segmentation and identification networks to individually identify animals, and their trajectories and back propagation techniques to identify animals during partial occlusions (Pérez-Escudero et al. 2014).

Beyond image detection, there are several groups working to integrate RFID tracking and video tracking (Weissbrod et al. 2013; de Chaumont et al. 2019; Peleh et al. 2019). Live Mouse Tracker (de Chaumont et al. 2019) uses a combination of RFID tracking and depth imaging, and has been validated with groups of up to 4 mice in a semi-natural environment over time. These

programs are highly influential in allowing for the long-term study of group dynamics in large, enriched home cage environments and currently provide the state-of-art tracking solutions for mice of similar appearance. A commercial version of this program, RFID-Assisted SocialScan, is also available and integrates a large behavioral arena with attached nest boxes (Peleh et al. 2019). Together, these programs can track the identity of individuals while examining social hierarchies over long periods of time without excessive manual oversight or behavioral scoring. However, these approaches require significant investments in highly specialized hardware.

Several groups pioneering pose estimation are also continually working to increase the speed of their algorithms to achieve real-time results (Graving et al. 2019; Mathis and Mathis 2019). Currently, ToxID and Sensory Orientation Software are able to provide real-time tracking of individual mice and their trajectories (Gomez-Marin et al. 2012; Rodriguez et al. 2017), and DeepLabCut has recently been used for real-time pose estimation (Forys et al. 2018). Extending these platforms for in-depth real-time behavioral classification in mice will allow for the evaluation of behavior on the timescale of neural activity and allow fully “closed-loop” recording and manipulation studies. The Stytra package is currently validated in zebrafish larvae, and is able to both analyze pose and activity and rapidly provide stimuli to the experimental animal based on its activity (Stih et al. 2019).

These open-source packages are continuing to improve and emerge, and integration between them is becoming easier. Pose estimators that can rapidly track individual animals despite multiple occlusions is an essential first step for the automation of aggression assays, and our lab has focused heavily on exploring random forest classifiers for the classification of aggression behaviors. Integrating the pose estimators above with high accuracy and easy to use classifiers in an open-source GUI allows for the generalization of these techniques across labs while maintaining the ability to take advantage of newer tracking technologies as they emerge.

For example, our lab has recently developed an open-source graphical workflow (for more information, see the SimBA GitHub repository) for creating supervised decision ensembles from body part tracking data generated through convolutional neural networks in accordance with the considerations and principles discussed above (Nilsson et al. 2020). The package incorporates accessible menus for pre-processing and annotating videos, generating/deploying classification algorithms, with advanced machine learning validation, tuning and visualization tools. The models created through this software—together with extensive documentation and tutorials—are available for the scientific community through online repositories, and we hope that this can spur further data sharing efforts that boosts the accuracy and scope of machine classification techniques in preclinical aggression research.

Conclusion

Aggression is an immutable force that contributes to the suffering and death for millions of people around the world (Sumner et al. 2015). While aggression can be highly rewarding and pursued despite adverse consequences (Chester and DeWall 2016; Gan et al. 2019), as well as sought after lengthy abstinence (Ducrose et al. 2014), aggression is not generally viewed in the medical profession or by the public as a component of neuropsychiatric disease states (Golden et al. 2017a; Golden and Shaham 2018). Indeed, in the recent formulation of Research Domain Criteria (RDoC) of NIMH, the term aggression does not appear under any of the research domains. Not surprisingly, no progress has been made in the treatment of pathological aggression, and little progress has been made using aggression as a diagnostic tool for neuropsychiatric comorbidity (Martin et al. 2013). We believe that drawing greater attention to the neurobiological mechanisms of maladaptive appetitive aggression within the context of neuropsychiatric disease will both act to (i) de-stigmatize aggression comorbidity in a manner similar to current approaches to substance abuse treatments and (ii) support the identification of novel mechanistic therapeutic approaches, which have presently been relegated to neuroleptic dopamine antagonists like haloperidol (Ostinelli et al. 2017) that exert their clinical efficacy through the neuroleptic’s sedative effects (Calver et al. 2015). Therefore, we propose that neurobiological and behavioral tools used to study drug seeking and relapse should be used to study brain mechanisms of appetitive aggression, both preclinically and clinically. Further, expanding beyond reactive aggression to incorporate the full spectrum of ethologically relevant aggressive behaviors is an essential step in increasing the utility and translation of ongoing preclinical research. Machine learning tools enabling high-throughput automated behavioral analysis are key to pursuing these lines of inquiry at a pace and depth necessary for modern neuroscience methods.

Significant progress has been made in developing open-source software packages that are capable of tracking social behavior over long periods of time and across diverse acquisition parameters (Table 5), and we propose that a focus on increasing ease of use, generalizability, cost-consciousness, accuracy, and easy expansion of training sets will allow for their wider adoption. Such automated scoring promises to remove long-standing bottlenecks within aggression research and allow for the high-throughput experiments required for mapping the nuance of aggression phenotypes while simultaneously building a common language of aggression phenotypes through shared machine models across labs.

Acknowledgments We thank Briana Smith, Liana Bloom, Sophia Hwang, and Jia Jie Choong for their thoughtful assistance in discussing the manuscript.

Funding information The research was supported NIDA 4R00DA045662-02 (SAG), NIDA P30 DA048736 (SRON and SAG)), NARSAD Young Investigator Award 27082 (SAG), and NIDA T32 5T32NS099578-04 (NLG).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Akay A, Hess H (2019) Deep learning: current and emerging applications in medicine and technology. *IEEE J Biomed Health Inform* 23:906–920. <https://doi.org/10.1109/JBHI.2019.2894713>
- Aleyasin H, Flanigan ME, Golden SA, Takahashi A, Menard C, Pfau ML, Multer J, Pina J, McCabe KA, Bhatti N, Hodes GE, Heshmati M, Neve RL, Nestler EJ, Heller EA, Russo SJ (2018) Cell-type-specific role of Δ FosB in nucleus accumbens in modulating intermale aggression. *J Neurosci* 38:5913–5924. <https://doi.org/10.1523/JNEUROSCI.0296-18.2018>
- Allahverdyan A, Galstyan A (2009) On maximum a posteriori estimation of hidden Markov processes. arXiv:09061980 [cond-mat, physics: physics, stat]
- Anderson DJ, Perona P (2014) Toward a science of computational ethology. *Neuron* 84:18–31. <https://doi.org/10.1016/j.neuron.2014.09.005>
- Anthony JC, Warner LA, Kessler RC (1994) Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: basic findings from the national comorbidity survey
- Arakawa T, Tanave A, Takahashi A et al (2017) Automated estimation of mouse social behaviors based on a hidden Markov model. In: Westhead DR, Vijayabaskar MS (eds) *Hidden Markov models: methods and protocols*. Springer New York, New York, NY, pp 185–197
- Arbustner BN, Li X, Pausch MH et al (2007) Evolving the lock to fit the key to create a family of G protein-coupled receptors potentially activated by an inert ligand. *Proc Natl Acad Sci U S A* 104:5163–5168. <https://doi.org/10.1073/pnas.0700293104>
- Azrin NH (1970) Punishment of elicited aggression. *J Exp Anal Behav* 14:7–10. <https://doi.org/10.1901/jeab.1970.14-7>
- Azrin NH, Hutchinson RR, Hake DF (1967) Attack, avoidance, and escape reactions to aversive shock. *J Exp Anal Behav* 10:131–148. <https://doi.org/10.1901/jeab.1967.10-131>
- Baenninger R, Grossman JC (1969) Some effects of punishment on pain-elicited aggression. *J Exp Anal Behav* 12:1017–1022. <https://doi.org/10.1901/jeab.1969.12-1017>
- Banerjee U (1971) An inquiry into the genesis of aggression in mice induced by isolation. *Behaviour* 40:86–99
- Bannai M, Fish EW, Faccidomo S, Miczek KA (2007) Anti-aggressive effects of agonists at 5-HT_{1B} receptors in the dorsal raphe nucleus of mice. *Psychopharmacology* 193:295–304. <https://doi.org/10.1007/s00213-007-0780-5>
- Bardo MT, Bevins RA (2000) Conditioned place preference: what does it add to our preclinical understanding of drug reward? *Psychopharmacology* 153:31–43
- Batuwita R, Palade V (2013) Class imbalance learning methods for support vector machines. In: He H, Ma Y (eds) *Imbalanced learning*. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 83–99
- Beach HD (1957) Effect of morphine on the exploratory drive. *Can J Psychol* 11:237–244
- Berman GJ, Choi DM, Bialek W, Shaevitz JW (2014) Mapping the stereotyped behaviour of freely moving fruit flies. *J R Soc Interface* 11: 20140672. <https://doi.org/10.1098/rsif.2014.0672>
- Blair RJR (2016) The neurobiology of impulsive aggression. *J Child Adolesc Psychopharmacol* 26:4–9. <https://doi.org/10.1089/cap.2015.0088>
- Boyden ES, Zhang F, Bamberg E, Nagel G, Deisseroth K (2005) Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci* 8:1263–1268. <https://doi.org/10.1038/nn1525>
- Brain PF, Benton D, Childs G, Parnigiani S (1981) The effect of the type of opponent in tests of murine aggression. *Behav Process* 6:319–327. [https://doi.org/10.1016/0376-6357\(81\)90049-8](https://doi.org/10.1016/0376-6357(81)90049-8)
- Branson K (2014) Distinguishing seemingly indistinguishable animals with computer vision. *Nat Methods* 11:721–722. <https://doi.org/10.1038/nmeth.3004>
- Branson K, Robie AA, Bender J, et al (2009) High-throughput ethomics in large groups of *Drosophila*. *Nat Methods* 6:451–457. <https://doi.org/10.1038/nmeth.1328>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Brown AEX, de Bivort B (2018) Ethology as a physical science. *Nat Phys* 14:653–657. <https://doi.org/10.1038/s41567-018-0093-0>
- Brox T, Rosenhahn B, Cremers D, Seidel H-P (2006) High accuracy optical flow serves 3-D pose tracking: exploiting contour and flow based constraints. In: Leonardis A, Bischof H, Pinz A (eds) *Computer vision – ECCV 2006*. Springer, Berlin, Heidelberg, pp 98–111
- Burgos-Artiz XP, Dollar P, Dayu Lin, et al (2012) Social behavior recognition in continuous video. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Providence, RI, pp 1322–1329
- Calver L, Drinkwater V, Gupta R et al (2015) Droperidol v. haloperidol for sedation of aggressive behaviour in acute mental health: randomised controlled trial. *Br J Psychiatry* 206:223–228. <https://doi.org/10.1192/bjp.bp.114.150227>
- Cao Z, Simon T, Wei S-E, Sheikh Y (2017) Realtime multi-person 2D pose estimation using part affinity fields. arXiv:161108050 [cs]
- Caprioli D, Venniro M, Zeric T, Li X, Adhikary S, Madangopal R, Marchant NJ, Lucantonio F, Schoenbaum G, Bossert JM, Shaham Y (2015) Effect of the novel positive allosteric modulator of metabotropic glutamate receptor 2 AZD8529 on incubation of methamphetamine craving after prolonged voluntary abstinence in a rat model. *Biol Psychiatry* 78:463–473. <https://doi.org/10.1016/j.biopsych.2015.02.018>
- Carola V, Mirabeau O, Gross CT (2011) Hidden Markov model analysis of maternal behavior patterns in inbred and reciprocal hybrid mice. *PLoS One* 6:e14753. <https://doi.org/10.1371/journal.pone.0014753>
- Caruana R, Niculescu-Mizil A (2006) An empirical comparison of supervised learning algorithms. In: *Proceedings of the 23rd international conference on machine learning - ICML '06*. ACM Press, Pittsburgh, Pennsylvania, pp 161–168
- Caruana R, Karampatziakis N, Yessenalina A (2008) An empirical evaluation of supervised learning in high dimensions. In: *Proceedings of the 25th international conference on machine learning - ICML '08*. ACM Press, Helsinki, Finland, pp 96–103
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* 16: 321–357. <https://doi.org/10.1613/jair.953>
- Chester DS, DeWall CN (2016) The pleasure of revenge: retaliatory aggression arises from a neural imbalance toward reward. *Soc Cogn Affect Neurosci* 11:1173–1182. <https://doi.org/10.1093/scan/nsv082>
- Chia R, Achilli F, Festing MFW, Fisher EMC (2005) The origins and uses of mouse outbred stocks. *Nat Genet* 37:1181–1186. <https://doi.org/10.1038/ng1665>

- Cole JM, Parker BK (1971) Schedule-induced aggression: access to an attackable target bird as a positive reinforcer. *Psychon Sci* 22:33–35. <https://doi.org/10.3758/BF03335924>
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20: 273–297. <https://doi.org/10.1007/BF00994018>
- Couppis MH, Kennedy CH (2008) The rewarding effect of aggression is reduced by nucleus accumbens dopamine receptor antagonism in mice. *Psychopharmacology* 197:449–456. <https://doi.org/10.1007/s00213-007-1054-y>
- Covington HE, Newman EL, Tran S et al (2018) The urge to fight: persistent escalation by alcohol and role of NMDA receptors in mice. *Front Behav Neurosci* 12:206. <https://doi.org/10.3389/fnbeh.2018.00206>
- Covington HE, Newman EL, Leonard MZ, Miczek KA (2019) Translational models of adaptive and excessive fighting: an emerging role for neural circuits in pathological aggression. *F1000Res* 8. <https://doi.org/10.12688/f1000research.18883.1>
- Coward P, Wada HG, Falk MS, Chan SDH, Meng F, Akil H, Conklin BR (1998) Controlling signaling with a specifically designed G-coupled receptor. *Proc Natl Acad Sci U S A* 95:352–357
- Craft BB, Velkey AJ, Szalda-Petree A (2003) Instrumental conditioning of choice behavior in male Siamese fighting fish (*Betta splendens*). *Behav Process* 63:171–175
- Craft BB, Szalda-Petree AD, Brinegar JL, Haddad NF (2007) Effect of various discriminative stimuli on choice behavior in male Siamese fighting fish (*Betta splendens*). *Percept Mot Skills* 104:575–580. <https://doi.org/10.2466/pms.104.2.575-580>
- Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P (2009) Automated monitoring and analysis of social behavior in *Drosophila*. *Nat Methods* 6:297–303. <https://doi.org/10.1038/nmeth.1310>
- Datta SR, Anderson DJ, Branson K et al (2019) Computational neuroethology: a call to action. *Neuron* 104:11–24. <https://doi.org/10.1016/j.neuron.2019.09.038>
- de Chaumont F, Ey E, Torquet N, Lagache T, Dallongeville S, Imbert A, Legou T, le Sourd AM, Faure P, Bourgeron T, Olivo-Marin JC (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. <https://doi.org/10.1038/s41551-019-0396-1>
- Deroche-Gamonet V (2004) Evidence for addiction-like behavior in the rat. *Science* 305:1014–1017. <https://doi.org/10.1126/science.1099020>
- Dölen G, Darvishzadeh A, Huang KW, Malenka RC (2013) Social reward requires coordinated activity of nucleus accumbens oxytocin and serotonin. *Nature* 501:179–184. <https://doi.org/10.1038/nature12518>
- Dolensek N, Gehrlach DA, Klein AS, Gogolla N (2020) Facial expressions of emotion states and their neuronal correlates in mice. *Science* 368:89–94. <https://doi.org/10.1126/science.aaz9468>
- Ducrose MR, Alexia D, Cooper, Howard N, Snyder (2014) Recidivism of prisoners released in 30 states in 2005: patterns from 2005 to 2010. U.S. Department of Justice
- Egnor SER, Branson K (2016) Computational analysis of behavior. *Annu Rev Neurosci* 39:217–236. <https://doi.org/10.1146/annurev-neuro-070815-013845>
- Elcoro M, Silva SP, Lattal KA (2008) Visual reinforcement in the female Siamese fighting fish, *Betta splendens*. *J Exp Anal Behav* 90:53–60. <https://doi.org/10.1901/jeab.2008.90.53>
- Falkner AL, Grosenick L, Davidson TJ, Deisseroth K, Lin D (2016) Hypothalamic control of male aggression-seeking behavior. *Nat Neurosci* 19:596–604. <https://doi.org/10.1038/nn.4264>
- Fazel S, Wolf A, Chang Z, Larsson H, Goodwin GM, Lichtenstein P (2015) Depression and violence: a Swedish population study. *Lancet Psychiatry* 2:224–232. [https://doi.org/10.1016/S2215-0366\(14\)00128-X](https://doi.org/10.1016/S2215-0366(14)00128-X)
- Fish EW, De Bold JF, Miczek KA (2002) Aggressive behavior as a reinforcer in mice: activation by allopregnanolone. *Psychopharmacology* 163:459–466. <https://doi.org/10.1007/s00213-002-1211-2>
- Fish EW, DeBold JF, Miczek KA (2005) Escalated aggression as a reward: corticosterone and GABAA receptor positive modulators in mice. *Psychopharmacology* 182:116–127. <https://doi.org/10.1007/s00213-005-0064-x>
- Flanigan ME, Russo SJ (2019) Recent advances in the study of aggression. *Neuropsychopharmacol* 44:241–244. <https://doi.org/10.1038/s41386-018-0226-2>
- Forys B, Xiao D, Gupta P et al (2018) Real-time markerless video tracking of body parts in mice using deep neural networks. *bioRxiv*: 482349. <https://doi.org/10.1101/482349>
- Frenay B, Verleysen M (2014) Classification in the Presence of Label Noise: A Survey. *IEEE Trans Neural Netw Learning Syst* 25:845–869. <https://doi.org/10.1109/TNNLS.2013.2292894>
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55: 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29:1189–1232
- Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Ann Stat* 28:337–407. <https://doi.org/10.1214/aos/1016218223>
- Frischknecht H-R, Siegfried B, Waser PG (1985) Postaggression footshock inhibits aggressive behavior in dominant but not in isolated mice. *Behav Neural Biol* 44:132–138. [https://doi.org/10.1016/S0163-1047\(85\)91286-5](https://doi.org/10.1016/S0163-1047(85)91286-5)
- Fuxjager MJ, Montgomery JL, Marler CA (2011) Species differences in the winner effect disappear in response to post-victory testosterone manipulations. *Proc Biol Sci* 278:3497–3503. <https://doi.org/10.1098/rspb.2011.0301>
- Gan G, Zilverstand A, Parvaz MA, Preston-Campbell RN, d'Oleire Uquillas F, Moeller SJ, Tomasi D, Goldstein RZ, Alia-Klein N (2019) Habenula-prefrontal resting-state connectivity in reactive aggressive men – a pilot study. *Neuropharmacology* 156:107396. <https://doi.org/10.1016/j.neuropharm.2018.10.025>
- Ginsburg B, Allee WC (1942) Some effects of conditioning on social dominance and subordination in inbred strains of mice. *Physiol Zool* 15:485–506
- Golden SA, Shaham Y (2018) Aggression addiction and relapse: a new frontier in psychiatry. *Neuropsychopharmacol* 43:224–225. <https://doi.org/10.1038/npp.2017.173>
- Golden SA, Covington HE, Berton O, Russo SJ (2011) A standardized protocol for repeated social defeat stress in mice. *Nat Protoc* 6:1183–1191. <https://doi.org/10.1038/nprot.2011.361>
- Golden SA, Heshmati M, Flanigan M, Christoffel DJ, Guise K, Pfau ML, Aleyasin H, Menard C, Zhang H, Hodes GE, Bregman D, Khibnik L, Tai J, Rebusi N, Krawitz B, Chaudhury D, Walsh JJ, Han MH, Shapiro ML, Russo SJ (2016) Basal forebrain projections to the lateral habenula modulate aggression reward. *Nature* 534:688–692. <https://doi.org/10.1038/nature18601>
- Golden SA, Aleyasin H, Heins R et al (2017a) Persistent conditioned place preference to aggression experience in adult male sexually-experienced CD-1 mice: persistent aggression conditioned place preference in CD-1 mice. *Genes Brain Behav* 16:44–55. <https://doi.org/10.1111/gbb.12310>
- Golden SA, Heins C, Venniro M, Caprioli D, Zhang M, Epstein DH, Shaham Y (2017b) Compulsive addiction-like aggressive behavior in mice. *Biol Psychiatry* 82:239–248. <https://doi.org/10.1016/j.biopsych.2017.03.004>
- Golden SA, Jin M, Heins C et al (2019a) Nucleus accumbens Drd1-expressing neurons control aggression self-administration and

- aggression seeking in mice. *J Neurosci* 39:2482–2496. <https://doi.org/10.1523/JNEUROSCI.2409-18.2019>
- Golden SA, Jin M, Shaham Y (2019b) Animal models of (or for) aggression reward, addiction, and relapse: behavior and circuits. *J Neurosci* 39:3996–4008. <https://doi.org/10.1523/JNEUROSCI.0151-19.2019>
- Gomez-Marin A, Partoune N, Stephens GJ, Louis M (2012) Automated tracking of animal posture and movement during exploration and sensory orientation behaviors. *PLoS One* 7:e41642. <https://doi.org/10.1371/journal.pone.0041642>
- Goodwin NL, Lopez SA, Lee NS, Beery AK (2018) Comparative role of reward in long-term peer and mate relationships in voles. *Horm Behav* 111:70–77. <https://doi.org/10.1016/j.yhbeh.2018.10.012>
- Graving JM, Chae D, Naik H, et al (2019) DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. 39
- Gris KV, Coutu J-P, Gris D (2017) Supervised and unsupervised learning technology in the study of rodent behavior. *Front Behav Neurosci* 11. <https://doi.org/10.3389/fnbeh.2017.00141>
- Günel S, Rhodin H, Morales D, et al (2019) DeepFly3D: a deep learning-based approach for 3D limb and appendage tracking in tethered, adult *Drosophila*. *bioRxiv* 640375. <https://doi.org/10.1101/640375>
- Haccou P, Kruk MR, Meelis E, van Bavel ET, Wouterse KM, Meelis W (1988) Markov models for social interactions: analysis of electrical stimulation in the hypothalamic aggression area of rats. *Anim Behav* 36:1145–1163. [https://doi.org/10.1016/S0003-3472\(88\)80074-5](https://doi.org/10.1016/S0003-3472(88)80074-5)
- Han W, Tellez LA, Rangel MJ et al (2017) Integrated control of predatory hunting by the central nucleus of the amygdala. *Cell* 168:311–324.e18. <https://doi.org/10.1016/j.cell.2016.12.027>
- Hand DJ (2006) Classifier technology and the illusion of progress. *Stat Sci* 21:1–14. <https://doi.org/10.1214/088342306000000060>
- Hashikawa K, Hashikawa Y, Lischinsky J, Lin D (2018) The neural mechanisms of sexually dimorphic aggressive behaviors. *Trends Genet* 34:755–776. <https://doi.org/10.1016/j.tig.2018.07.001>
- He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. *arXiv:170306870* [cs]
- Hong W, Kennedy A, Burgos-Artizzu XP, Zelikowsky M, Navonne SG, Perona P, Anderson DJ (2015) Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proc Natl Acad Sci U S A* 112:E5351–E5360. <https://doi.org/10.1073/pnas.1515982112>
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *PNAS* 79:2554–2558. <https://doi.org/10.1073/pnas.79.8.2554>
- Hsu C-W, Lin C-J (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw* 13:415–425. <https://doi.org/10.1109/72.991427>
- Hsu AI, Yttri EA (2019) B-SOID: an open source unsupervised algorithm for discovery of spontaneous behaviors. *Neuroscience*
- Jhuang H, Garrote E, Yu X, Khilnani V, Poggio T, Steele AD, Serre T (2010) Automated home-cage behavioural phenotyping of mice. *Nat Commun* 1:1–10. <https://doi.org/10.1038/ncomms1064>
- Jones SE, Brain PF (1987) Performances of inbred and outbred laboratory mice in putative tests of aggression. *Behav Genet* 17:87–96. <https://doi.org/10.1007/BF01066013>
- Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson K (2013) JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat Methods* 10:64–67. <https://doi.org/10.1038/nmeth.2281>
- Karpathy A, Toderici G, Shetty S, et al (2014) Large-scale video classification with convolutional neural networks. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Columbus, OH, USA, pp 1725–1732
- Kim Y-K, Saver M, Simon J, Kent CF, Shao L, Eddison M, Agrawal P, Texada M, Truman JW, Heberlein U (2018) Repetitive aggressive encounters generate a long-lasting internal state in *Drosophila melanogaster* males. *PNAS* 115:1099–1104. <https://doi.org/10.1073/pnas.1716612115>
- Klibaite U, Berman GJ, Cande J, Stern DL, Shaevez JW (2017) An unsupervised method for quantifying the behavior of paired animals. *Phys Biol* 14:015006. <https://doi.org/10.1088/1478-3975/aa5c50>
- Kou R, Hsu C-C, Chen S-C et al (2019) Winner and loser effects in lobster cockroach contests for social dominance. *Horm Behav* 107:49–60. <https://doi.org/10.1016/j.yhbeh.2018.12.002>
- Krasnova IN, Marchant NJ, Ladenheim B, McCoy MT, Panlilio LV, Bossert JM, Shaham Y, Cadet JL (2014) Incubation of methamphetamine and palatable food craving after punishment-induced abstinence. *Neuropsychopharmacology* 39:2008–2016. <https://doi.org/10.1038/npp.2014.50>
- Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. *Commun ACM* 60:84–90. <https://doi.org/10.1145/3065386>
- Kudryavtseva NN, Bakshantovskaya IV, Koryakina LA (1991) Social model of depression in mice of C57BL/6J strain. *Pharmacol Biochem Behav* 38:315–320. [https://doi.org/10.1016/0091-3057\(91\)90284-9](https://doi.org/10.1016/0091-3057(91)90284-9)
- Kudryavtseva NN, Smagin DA, Bondar NP (2011) Modeling fighting deprivation effect in mouse repeated aggression paradigm. *Prog Neuro-Psychopharmacol Biol Psychiatry* 35:1472–1478. <https://doi.org/10.1016/j.pnpbp.2010.10.013>
- Kudryavtseva NN, Smagin DA, Kovalenko IL, Vishnivetskaya GB (2014) Repeated positive fighting experience in male inbred mice. *Nat Protoc* 9:2705–2717. <https://doi.org/10.1038/nprot.2014.156>
- Kuehlkamp A, Becker B, Bowyer K (2017) Gender-from-iris or gender-from-mascara? *arXiv:170201304* [cs]
- Lacourse E, Côté S, Nagin DS et al (2002) A longitudinal–experimental approach to testing theories of antisocial behavior development. *Dev Psychopathol* 14:909–924. <https://doi.org/10.1017/S0954579402004121>
- Lagerspetz K (1964) Studies on the aggressive behaviour of mice. *Annales Academiae Scientiarum Fennicae Series B* 131:1–131
- LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD (1989) Backpropagation applied to handwritten zip code recognition. *Neural Comput* 1:541–551. <https://doi.org/10.1162/neco.1989.1.4.541>
- Legrand R (1970) Successful aggression as the reinforcer for runway behavior of mice. *Psychon Sci* 20:303–305. <https://doi.org/10.3758/BF03329080>
- Legrand R (1978) Reinforcing effect of aggressive behaviors preparatory to fighting in mice. *Bull Psychon Soc* 11:359–362. <https://doi.org/10.3758/BF03336854>
- Liaw A, Wiener M (2002) Classification and regression by random. *Forest* 2:6
- Lietman T, Eng J, Katz J, Quigley HA (1999) Neural networks for visual field analysis: how do they compare with other algorithms? *J Glaucoma* 8:77–80
- Liu M, Wang M, Wang J, Li D (2013) Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar. *Sensors Actuators B Chem* 177:970–980. <https://doi.org/10.1016/j.snb.2012.11.071>
- Marchant NJ, Campbell EJ, Pelloux Y, Bossert JM, Shaham Y (2019) Context-induced relapse after extinction versus punishment: similarities and differences. *Psychopharmacology* 236:439–448. <https://doi.org/10.1007/s00213-018-4929-1>
- Martin LA, Neighbors HW, Griffith DM (2013) The experience of symptoms of depression in men vs women: analysis of the National Comorbidity Survey Replication. *JAMA Psychiatry* 70:1100–1106. <https://doi.org/10.1001/jamapsychiatry.2013.1985>
- Martínez M, Guillén-Salazar F, Salvador A, Simón VM (1995) Successful intermale aggression and conditioned place preference

- in mice. *Physiol Behav* 58:323–328. [https://doi.org/10.1016/0031-9384\(95\)00061-m](https://doi.org/10.1016/0031-9384(95)00061-m)
- Mathis MW, Mathis A (2019) Deep learning tools for the measurement of animal behavior in neuroscience. arXiv:190913868 [cs, q-bio]
- Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M (2018) DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* 21:1281–1289. <https://doi.org/10.1038/s41593-018-0209-y>
- May ME, Kennedy CH (2009) Aggression as positive reinforcement in mice under various ratio- and time-based reinforcement schedules. *J Exp Anal Behav* 91:185–196. <https://doi.org/10.1901/jeab.2009.91-185>
- Meisel RL, Joppa MA (1994) Conditioned place preference in female hamsters following aggressive or sexual encounters. *Physiol Behav* 56:1115–1118. [https://doi.org/10.1016/0031-9384\(94\)90352-2](https://doi.org/10.1016/0031-9384(94)90352-2)
- Miczek KA, O'Donnell JM (1978) Intruder-evoked aggression in isolated and nonisolated mice: effects of psychomotor stimulants and L-dopa. *Psychopharmacology* 57:47–55. <https://doi.org/10.1007/bf00426957>
- Miczek KA, Thompson ML, Shuster L (1982) Opioid-like analgesia in defeated mice. *Science* 215:1520–1522. <https://doi.org/10.1126/science.7199758>
- Miczek KA, Maxson SC, Fish EW, Faccidomo S (2001) Aggressive behavioral phenotypes in mice. *Behav Brain Res* 125:167–181. [https://doi.org/10.1016/S0166-4328\(01\)00298-4](https://doi.org/10.1016/S0166-4328(01)00298-4)
- Miczek KA, Takahashi A, Gobrogge KL, Hwa LS, de Almeida RMM (2015) Escalated aggression in animal models: shedding new light on mesocorticolimbic circuits. *Curr Opin Behav Sci* 3:90–95. <https://doi.org/10.1016/j.cobeha.2015.02.007>
- Miczek KA, DeBold JF, Gobrogge K, et al (2017) The role of neurotransmitters in violence and aggression. In: *The Wiley Handbook of Violence and Aggression*. American Cancer Society, pp 1–13
- Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L (2016) The ethics of algorithms: mapping the debate. *Big Data & Society* 3: 205395171667967. <https://doi.org/10.1177/2053951716679679>
- Momohara Y, Minami H, Kanai A, Nagayama T (2016) Role of cAMP signalling in winner and loser effects in crayfish agonistic encounters. *Eur J Neurosci* 44:1886–1895. <https://doi.org/10.1111/ejn.13259>
- Mucha RF, van der Kooy D, O'Shaughnessy M, Bucenieks P (1982) Drug reinforcement studied by the use of place conditioning in rat. *Brain Res* 243:91–105. [https://doi.org/10.1016/0006-8993\(82\)91123-4](https://doi.org/10.1016/0006-8993(82)91123-4)
- Natarajan D, de Vries H, Saaltink D-J, de Boer SF, Koolhaas JM (2009) Delineation of violence from functional aggression in mice: an ethological approach. *Behav Genet* 39:73–90. <https://doi.org/10.1007/s10519-008-9230-3>
- Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW (2019) Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc* 14:2152–2176. <https://doi.org/10.1038/s41596-019-0176-0>
- Newman EL, Terunuma M, Wang TL, Hewage N, Bicakci MB, Moss SJ, DeBold JF, Miczek KA (2018) A role for prefrontal cortical NMDA receptors in murine alcohol-heightened aggression. *Neuropsychopharmacology* 43:1224–1234. <https://doi.org/10.1038/npp.2017.253>
- Nilsson SRO, Goodwin NL, Choong JJ et al (2020) Simple Behavioral Analysis (SimBA): an open source toolkit for computer classification of complex social behaviors in experimental animals. *bioRxiv* 2020.04.19.049452. <https://doi.org/10.1101/2020.04.19.049452>
- Nitze I, Schulthess U, Asche H (2012) Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. 7
- O'Kelly LI, Steckle LC (1939) A note on long enduring emotional responses in the rat. *The Journal of Psychology* 8:125–131. <https://doi.org/10.1080/00223980.1939.9917655>
- Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366:447–453. <https://doi.org/10.1126/science.aax2342>
- Okada K, Okada Y, Dall SRX, Hosken DJ (2019) Loser-effect duration evolves independently of fighting ability. *Proc Biol Sci* 286: 20190582. <https://doi.org/10.1098/rspb.2019.0582>
- Ostinelli EG, Brooke-Powney MJ, Li X, Adams CE (2017) Haloperidol for psychosis-induced aggression or agitation (rapid tranquillisation). *Cochrane Database Syst Rev* 7:CD009377–CD009377. <https://doi.org/10.1002/14651858.CD009377.pub3>
- Oyegbile TO, Marler CA (2005) Winning fights elevates testosterone levels in California mice and enhances future ability to win fights. *Horm Behav* 48:259–267. <https://doi.org/10.1016/j.yhbeh.2005.04.007>
- Oyegbile TO, Marler CA (2006) Weak winner effect in a less aggressive mammal: correlations with corticosterone but not testosterone. *Physiol Behav* 89:171–179. <https://doi.org/10.1016/j.physbeh.2006.05.044>
- Panksepp JB, Lahvis GP (2007) Social reward among juvenile mice. *Genes Brain Behav* 6:661–671. <https://doi.org/10.1111/j.1601-183X.2006.00295.x>
- Peleh T, Bai X, Kas MJH, Hengerer B (2019) RFID-supported video tracking for automated analysis of social behaviour in groups of mice. *J Neurosci Methods* 325:108323. <https://doi.org/10.1016/j.jneumeth.2019.108323>
- Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SSH, Murthy M, Shaevitz JW (2019) Fast animal pose estimation using deep neural networks. *Nat Methods* 16:117–125. <https://doi.org/10.1038/s41592-018-0234-5>
- Pérez-Escudero A, Vicente-Page J, Hinz RC et al (2014) idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nat Methods* 11:743–748. <https://doi.org/10.1038/nmeth.2994>
- Piazza PV, Deroche-Gamonet V (2013) A multistep general theory of transition to addiction. *Psychopharmacology* 229:387–413. <https://doi.org/10.1007/s00213-013-3224-4>
- Pickens CL, Airavaara M, Theberge F, Fanous S, Hope BT, Shaham Y (2011) Neurobiology of the incubation of drug craving. *Trends Neurosci* 34:411–420. <https://doi.org/10.1016/j.tins.2011.06.001>
- Potegal M (1979) The reinforcing value of several types of aggressive behavior: a review. *Aggress Behav* 5:353–373. [https://doi.org/10.1002/1098-2337\(1979\)5:4<353::AID-AB2480050404>3.0.CO;2-7](https://doi.org/10.1002/1098-2337(1979)5:4<353::AID-AB2480050404>3.0.CO;2-7)
- Provencal N, Booi L, Tremblay RE (2015) The developmental origins of chronic physical aggression: biological pathways triggered by early life adversity. *J Exp Biol* 218:123–133. <https://doi.org/10.1242/jeb.111401>
- Rabiner LR, Juang BH (1986) An introduction to hidden Markov Models. 12
- Ribeiro MT, Singh S, Guestrin C (2016) “Why Should I Trust You?”: explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. ACM Press, San Francisco, California, USA, pp 1135–1144
- Roberts CL, Blase K (1971) Elicitation and punishment of intraspecies aggression by the same stimulus. *J Exp Anal Behav* 15:193–196. <https://doi.org/10.1901/jeab.1971.15-193>
- Robie AA, Seagraves KM, Egnor SER, Branson K (2017) Machine vision methods for analyzing social interactions. *J Exp Biol* 220:25–34. <https://doi.org/10.1242/jeb.142281>
- Rodriguez A, Zhang H, Klaminder J, Brodin T, Andersson M (2017) ToxId: an efficient algorithm to solve occlusions when tracking multiple animals. *Sci Rep* 7:1–8. <https://doi.org/10.1038/s41598-017-15104-2>

- Romero-Ferrero F, Bergomi MG, Hinz RC, Heras FJH, de Polavieja GG (2019) idtracker.ai: tracking all individuals in small or large collectives of unmarked animals. *Nat Methods* 16:179–182. <https://doi.org/10.1038/s41592-018-0295-5>
- Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 1:206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Rudolf J, Dondorp D, Canon L, Tieo S, Chatzigeorgiou M (2019) Automated behavioural analysis reveals the basic behavioural repertoire of the urochordate *Ciona intestinalis*. *Sci Rep* 9:2416. <https://doi.org/10.1038/s41598-019-38791-5>
- Schaefer AT, Claridge-Chang A (2012) The surveillance state of behavioural automation. *Curr Opin Neurobiol* 22:170–176. <https://doi.org/10.1016/j.conb.2011.11.004>
- Schwartz JJ, Ricci LA, Melloni RH (2013) Prior fighting experience increases aggression in Syrian hamsters: implications for a role of dopamine in the winner effect. *Aggress Behav* 39:290–300. <https://doi.org/10.1002/ab.21476>
- Sinha R (2011) New findings on biological factors predicting addiction relapse vulnerability. *Curr Psychiatry Rep* 13:398–405. <https://doi.org/10.1007/s11920-011-0224-0>
- Stagkourakis S, Spigolon G, Williams P, Protzmann J, Fisone G, Broberger C (2018) A neural network for intermale aggression to establish social hierarchy. *Nat Neurosci* 21:834–842. <https://doi.org/10.1038/s41593-018-0153-x>
- Stih V, Petrucco L, Kist A, Portugues R (2019) Stytra: an open-source, integrated system for stimulation, tracking and closed-loop behavioural experiments. 19
- Sumner SA, Mercy JA, Dahlberg LL, Hillis SD, Kleven J, Houry D (2015) Violence in the United States: status, challenges, and opportunities. *JAMA* 314:478–488. <https://doi.org/10.1001/jama.2015.8371>
- Takahashi A, Miczek KA (2014) Neurogenetics of aggressive behavior: studies in rodents. *Curr Top Behav Neurosci* 17:3–44. https://doi.org/10.1007/7854_2013_263
- Taylor GT (1979) Reinforcement and intraspecific aggressive behavior. *Behav Neural Biol* 27:1–24. [https://doi.org/10.1016/S0163-1047\(79\)92720-1](https://doi.org/10.1016/S0163-1047(79)92720-1)
- Tellegen A, Horn JM (1972) Primary aggressive motivation in three inbred strains of mice. *J Comp Physiol Psychol* 78:297–304. <https://doi.org/10.1037/h0032192>
- Thanos PK, Restif C, O'Rourke JR et al (2017) Mouse Social Interaction Test (MoST): a quantitative computer automated analysis of behavior. *J Neural Transm* 124:3–11. <https://doi.org/10.1007/s00702-015-1487-0>
- Thompson TI (1963) Visual reinforcement in Siamese fighting fish. *Science* 141:55–57. <https://doi.org/10.1126/science.141.3575.55>
- Thompson TI (1964) Visual reinforcement in fighting cocks. *J Exp Anal Behav* 7:45–49. <https://doi.org/10.1901/jeab.1964.7.45>
- Thompson T, Sturm T (1965) Visual-reinforcer color, and operant behavior in Siamese fighting fish. *J Exp Anal Behav* 8:341–344. <https://doi.org/10.1901/jeab.1965.8.341>
- Todd JG, Kain JS, de Bivort BL (2017) Systematic exploration of unsupervised methods for mapping behavior. *Phys Biol* 14:015002. <https://doi.org/10.1088/1478-3975/14/1/015002>
- Trannoy S, Chowdhury B, Kravitz EA (2015) Handling alters aggression and “loser” effect formation in *Drosophila melanogaster*. *Learn Mem* 22:64–68. <https://doi.org/10.1101/lm.036418.114>
- Tuttle AH, Philip VM, Chesler EJ, Mogil JS (2018) Comparing phenotypic variation between inbred and outbred mice. *Nat Methods* 15: 994–996. <https://doi.org/10.1038/s41592-018-0224-7>
- Tyrer P, Kendall T, Barnett R, et al (2015) Violence and aggression: short-term management in mental health, health and community settings
- Ulrich R, Wolfe M, Dulane S (1969) Punishment of shock-induced aggression. *J Exp Anal Behav* 12:1009–1015. <https://doi.org/10.1901/jeab.1969.12-1009>
- Vernon W, Ulrich R (1966) Classical conditioning of pain-elicited aggression. *Science* 152:668–669. <https://doi.org/10.1126/science.152.3722.668>
- Vlauthin C, Ferkin M (2013) The outcome of a previous social interaction with a same-sex conspecific affects the behavior of meadow voles, *Microtus pennsylvanicus*. *ETHOLOGY* 119:212–220
- Vogelstein JT, Park Y, Ohyama T, Kerr RA, Truman JW, Priebe CE, Zlatić M (2014) Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science* 344:386–392. <https://doi.org/10.1126/science.1250298>
- Weissbrod A, Shapiro A, Vasserman G et al (2013) Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nat Commun* 4:2018. <https://doi.org/10.1038/ncomms3018>
- Wiltschko AB, Johnson MJ, Iurilli G, Peterson RE, Katon JM, Pashkovski SL, Abaira VE, Adams RP, Datta SR (2015) Mapping sub-second structure in mouse behavior. *Neuron* 88: 1121–1135. <https://doi.org/10.1016/j.neuron.2015.11.031>
- Xu L (2006) Robust support vector machine training via convex outlier ablation. 7
- Zhang W, Yartsev MM (2019) Correlated neural activity across the brains of socially interacting bats. *Cell* 178:413–428.e22. <https://doi.org/10.1016/j.cell.2019.05.023>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.