Research Article: New Research | Cognition and Behavior

High-THC *Cannabis* smoke impairs incidental memory capacity in spontaneous tests of novelty preference for objects and odors in male rats

https://doi.org/10.1523/ENEURO.0115-23.2023

Cite as: eNeuro 2023; 10.1523/ENEURO.0115-23.2023

Received: 8 April 2023
Revised: 19 September 2023
Accepted: 16 October 2023

High-THC *Cannabis* smoke impairs incidental memory capacity in spontaneous tests of novelty preference for objects and odors in male rats

Abbreviated title: High-THC *Cannabis* smoke impairs incidental memory capacity in male rats

Ilne L. Barnard[*1], Timothy J. Onofrychuk[*1], Aaron D. Toderash[4], Vyom N. Patel[4], Aiden E. Glass[1], Jesse C. Adrian[1], Robert. B. Laprairie[2,3], John G. Howland[1,#]

[1] Department of Anatomy, Physiology, and Pharmacology, University of Saskatchewan, Saskatoon, SK, Canada
[2] College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, SK, Canada
[3] Department of Pharmacology, College of Medicine, Dalhousie University, Halifax, NS, Canada
[4] Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada

* These authors contributed equally to this work.

Correspondence to JGH:
Department of Anatomy, Physiology, and Pharmacology
University of Saskatchewan
GD30.7, Health Sciences Building
107 Wiggins Road
Saskatoon, SK S7N 5E5
(e) john.howland@usask.ca

Number of figures: 6
Number of tables: 3
Multimedia: 0

Number of words, Abstract: 248
Number of words, Introduction: 837
Number of words, Discussion: 2519

Acknowledgements and funding sources: Funding for these experiments was obtained from the University of Saskatchewan College of Medicine and the Natural Sciences and Engineering Research Council of Canada (NSERC) to JGH. ILB and TJO were supported by scholarships from NSERC. JCA was supported by the University of Saskatchewan College of Medicine. The authors wish to thank Morgan Schatz for initial pilot research on the behavioral tests reported in this paper, Killian Stacey for contributing to the implementation of automated behavioural analysis methods, and Hassaan Sabir for scoring some of the behavior included in the supplemental information.

Conflict of interest: RBL is a member of the Scientific Advisory Board for Shackleford Pharma Inc.; however this company had no input into this research study. The other authors of this study have no conflicts to declare.

Abstract

Working memory is an executive function that orchestrates the use of limited amounts of information, referred to as working memory capacity, in cognitive functions. *Cannabis* exposure impairs working memory in humans; however, it is unclear if *Cannabis* facilitates or impairs rodent working memory and working memory capacity. The conflicting literature in rodent models may be at least partly due to the use of drug exposure paradigms that do not closely mirror patterns of human *Cannabis* use. Here, we used an incidental memory capacity paradigm where a novelty preference is assessed after a short delay in spontaneous recognition-based tests. Either object or odor-based stimuli were used in test variations with sets of identical (IST) and different (DST) stimuli (3 or 6) for low- and high-memory loads, respectively. Additionally, we developed a human-machine hybrid behavioral quantification approach which supplements stopwatch-based scoring with supervised machine learning-based classification. After validating the spontaneous IST and DST in male rats, 6-item test versions with the hybrid quantification method were used to evaluate the impact of acute exposure to high-THC or high-CBD *Cannabis* smoke on novelty preference. Under control conditions, male rats showed novelty preference in all test variations. We found that high-THC, but not high-CBD, *Cannabis* smoke exposure impaired novelty preference for objects under a high-memory load. Odor-based recognition deficits were seen under both low-, and high-memory loads only following high-THC smoke exposure. Ultimately, these data show that *Cannabis* smoke exposure impacts incidental memory capacity of male rats in a memory load-dependent, and stimuli-specific manner.

Significance Statement

Incidental memory refers to the limited amount of information encoded by chance during behavior. How psychoactive drug exposure affects incidental memory is poorly understood, particularly for *Cannabis* exposure. To address this question, we validated object- and odor-based spontaneous incidental memory tests in male rats using a novel human-machine hybrid scoring method. Using these tests, we show exposure to high-THC, but not high-CBD, *Cannabis* smoke impairs incidental memory under high-memory loads in object-based tests and both high- and low-memory loads in the odor-based tests. Our results highlight cannabinoid-specific effects on incidental memory in male rats using a validated *Cannabis* smoke exposure method, which have broad implications for the impacts of human use of *Cannabis* on cognition.

Introduction

Working memory is an executive function that orchestrates the use of limited amounts of information in cognitive functions like learning and memory (Constantinidis & Klingberg, 2016; D'Esposito et al., 1995; Eriksson et al., 2015; Wilhelm et al., 2013). In humans, Δ^9-tetrahydrocannabinol (THC), the main psychoactive constituent of *Cannabis*, impairs working memory following both acute and chronic *Cannabis* exposure, likely by action at the cannabinoid type 1 receptor (Adam et al., 2020; Bossong et al., 2012; Cousijn et al., 2014; Crane et al., 2013; Curran et al., 2002; D'Souza et al., 2012; Ilan et al., 2004; Ligresti et al., 2016; Owens et al., 2019). The working memory impairments produced by *Cannabis* have been interpreted as resulting from disruptions of the active maintenance, limited capacity, interference control, and flexible updating subconstructs of working memory (Barch & Smith, 2008). In contrast, studies in rodents demonstrate both THC-mediated impairments and improvements in working memory function (Barnard et al., 2022; Blaes et al., 2019; Bruijnzeel et al., 2016; de Melo et al., 2005; Goonawardena et al., 2010; Varvel et al., 2001). These inconsistent findings may be attributable to differences in the behavioral tasks used, cannabinoid dosage, exposure timelines, and routes of administration (Baglot et al., 2021; Hložek et al., 2017; Klausner & Dingell, 1971; Nguyen et al., 2016; Wiley et al., 2021). Importantly, previous rodent studies have not directly assessed the effects of *Cannabis* exposure on working memory capacity. Working memory capacity is essential for higher cognitive operations critical to everyday function and can be impaired in disorders like schizophrenia and Parkinson's disease (Goldman-Rakic, 1999; Piskulic et al., 2007; Gold et al., 2018).

A shortcoming in rodent literature is that traditional rodent working memory capacity tests mimic n-back or recall working memory tests used in humans and require a long training period, learned rules, and considerable experimental involvement (Barnard et al., 2022; Cowan, 2010; Daneman & Carpenter, 1980; Dudchenko, 2004; Dudchenko et al., 2013; Kirchner, 1958; Oomen et al., 2013; Scott et al., 2020; Vorhees & Williams, 2014; Wilhelm et al., 2013). Spontaneous recognition tests circumvent these weaknesses by relying on rodents' innate novelty seeking behavior as shown by preferential interaction with a novel stimulus after a delay (Broadbent et al., 2004; Ennaceur & Aggleton, 1994; Ennaceur & Delacour, 1988; Sannino et al., 2012). These tests measure incidental memory capacity, which is the limited amount of information that is encoded by chance during spontaneous exploration. It is noteoworthy that incidental memory capacity differs from working memory capacity, as information is encoded without the intent for future use. Novelty preference can be used to assess incidental memory capacity in mice under low- and high-memory loads through the Identical and Different Objects Tasks, respectively (Torromino et al., 2022; Olivito et al., 2016, 2019; Sannino et al., 2012). Therefore, the first goal of the present study was to validate these tests in male rats using the Identical Stimuli Test (IST) and Different Stimuli Test (DST) with objects. Our second goal was to develop and validate olfactory versions of these tests to evaluate incidental memory for odors. We chose to perform this initial validation with male rats given the recently reported sex differences in the neural circuitry underlying performance of the tests with objects in mice (Torromino et al., 2022).

For all test variations, novelty preference was inferred by measuring the relative amount of interaction behavior exhibited at novel and previously experienced stimuli after a short delay. Typical approaches to quantifying rodent behavior for spontaneous interaction tests are generally laborious, prone to human subjectivity, and lack objective analysis steps that can be verified and reproduced (Anderson & Perona, 2014). Recent advances in automated behavioral analysis have enabled researchers to obtain a detailed and objective record of a diversity of complex behaviors across species (Cui et al., 2021; Newton et al., 2023; Nilsson et al., 2020; Slivicki et al., 2023; Winters et al., 2022). Here, we automatically quantified interaction events using a supervised machine learning-based analysis approach with DeepLabCut (Mathis et al., 2018) and Simple Behavioral Analysis (SimBA; Nilsson et al., 2020), then upon manual inspection of supervised machine learning predictions, sub-optimal predictions were supplemented by human stopwatch scoring to form a human-machine hybrid scoring method. By automatically predicting interaction events frame-by-frame, several secondary behavioral measures, including approach latency and interaction bout count, were easily calculated and provide a more complete characterization of novelty preference to infer incidental memory capacity. To our knowledge, the present study is the first demonstration of supervised machine learning-based behavioral analysis in the context of a spontaneous interaction-based test.

Using validated spontaneous tests and the hybrid scoring method, our second goal was to assess the effects of *Cannabis* smoke exposure on novelty preference to infer incidental memory capacity. We tested male rats shortly after acute exposure to the smoke of either high-THC or high-CBD-containing *Cannabis* buds using an exposure paradigm validated with rats (Barnard et al., 2022; Roebuck et al., 2022). We found that high-THC, but not high-CBD, smoke impaired performance of male rats in the tests in a stimuli-specific manner.

Materials & Methods

Subjects

Adult (2-4 months of age) male Long-Evans rats (n=92; Charles River Laboratories, Kingston, NY) were pair housed in a vivarium in standard ventilated cages with ad libitum water and food, and a plastic tube for environmental enrichment on a 12-hour light/dark cycle (starting at 0700). For establishment and validation of IST and DST with objects and odors, 52 rats were used; 48 additional rats were used to evaluate the impact of acute *Cannabis* smoke exposure on novelty preference. Rats were tested at the same time of day between the hours of 0730 and 1800. All procedures followed guidelines from the Canadian Council on Animal Care and were approved by the University of Saskatchewan Animal Research Ethics Board.

Apparatus and testing materials

Rats were handled in the testing room (3 mins a day for 3 days) and subsequently habituated to both the testing apparatus (10 min for 2 days) and to the smoke chamber apparatus (20 min for 2 days). Rats were tested in a white corrugated plastic box (60 cm x 60 cm x 60 cm) with the stimuli evenly presented between two opposing walls at three positions (see Fig 1; 9 cm from side of box, 21.5 cm apart from each other). Object stimuli were created from a variety of LEGO™ pieces of different sizes and colors with an average size of 7 cm x 10 cm. LEGO™ was chosen to maintain consistency between different object sets. Odor stimuli were created using 250 mL glass canning jars. The jars were filled with sand for stability, and to provide a resting place for a small plastic vile filled half-way with a powered spice (lemon pepper, dill, sage, onion, anise, cloves, ginger, cumin, cocoa, celery salt, coffee, cinnamon, garlic, or oregano). Holes were drilled in the lids of the jars to allow the rats to smell the spices. All items were affixed to the testing apparatus with Velcro™ at one of six positions to prevent them from being displaced during the test.

Spontaneous incidental memory test protocol

To validate the IST and DST with objects, 24 naïve rats performed both the 3- and 6- object variations (Fig 1). Twenty naïve rats were used to establish the 3- and 6-odor IST and DST. Using a within-subjects design, 48 additional rats performed both the IST and DST with objects and odors 20 min after *Cannabis* smoke exposure (Fig 2A). The order of tests was quasi-counterbalanced, and rats had a 2-day washout period between tests. On the test day, the testing box was prepared with 2 sets of 6 stimuli for the test and paradigm being performed (Figs 2A; 4A,B; 5A,B). The rat was then placed into the testing box for the sample phase, for a duration of 5 min. Following the sample phase, the rat was taken out of the testing box and placed inside a transport cage for 1 min. During the delay, all stimuli were replaced for the test phase. Then, the rat was placed back into the box for the test phase (5 min). The testing box and the stimuli were cleaned with 70% ethanol after each phase.

Cannabis bud preparation and acute smoke exposure protocol

A high-THC (19.51%) and low-CBD (<0.07%) strain, Skywalker (Aphria Inc., Lemington, ON, lot #6216), and a high-CBD (12.98%) and low-THC (0.67%) strain, Treasure Island (Aphria Inc., Lemington, ON, lot #6812), were used for *Cannabis* smoke exposure as previously established (Barnard et al., 2022; Roebuck et al., 2022). All *Cannabis* was stored in light-protected containers at room temperature. On the day of testing, whole *Cannabis* buds were ground in a standard coffee grinder for 5 sec. Then, 300 mg of the ground bud was measured and loaded into a ceramic coil that was part of a 4-chamber inhalation system from La Jolla Alcohol Research, Inc. (San Diego, CA). Rats were then loaded individually into small plastic cages and placed in the airtight Plexiglas chambers. A *Cannabis* combustion session started with a 5-min acclimation period, then a 1-min combustion occurred through three 5 sec ignitions with a 15 sec delay in-between each ignition. The temperature was set to 149°C, with a wattage of 60.1 W on the SV250 mod box. The smoke was then drawn into the clear Plexiglas chambers at a flow rate of 10-12 L/min. Following the 1-min combustion cycle, pumps were turned off for 1 min before they were turned back on for 13-min to gradually evacuate the smoke. Thus, the total exposure time was 15 min following initial ignition of the *Cannabis.* Rats were then moved to the testing apparatus to start the behavioral tests 20 min after the start of the combustion cycle. Boli left by the rats in the small plastic cages that housed them during combustion were then counted by an experimenter.

Behavioral Analysis

For validation of spontaneous incidental memory tests, behavioral videos were collected from an overhead perspective in black and white at a frame rate of 30 frames per second (fps) with a resolution of 720 pixels x 480 pixels (Panasonic WV-BP334 1/3" B&W). Collected videos were manually scored using a conventional stopwatch method, where the duration of interaction at each stimulus was recorded.

To allow for automated behavioral analysis, behavioral videos for the *Cannabis* exposure experiment were recorded from an overhead perspective in full color at a frame rate of 30 fps and a resolution of 1080 pixels x 1080 pixels (Logitech Brio 505, Logitech). To further standardize behavioral videos, we used the "batch preprocessing" module within SimBA to crop videos to only include the apparatus, to ensure standardized resolution and frame rate, and to the trim video length to desired experimental phases. Additionally, we chose to film all videos in a .mp4 format as this format is generally compatible with open-source video analysis software. More details regarding this process, and the subsequent steps in our supervised maching learning pipeline can be found here (**https://github.com/HowlandLab/ILBTJO_NODB_SimBA_2023**).

After filming, DeepLabCut (2.2.3) was utilized to continuously track the spatial location of eight user defined points-of-interest (Fig 2B) (Mathis et al., 2018). Mean tracking confidence for each point-of-interest is shown in Extended Data, Fig 2-1. To train the DeepLabCut model, we randomly extracted 300 frames from 60 representative behavioral videos, with an equal representation of the IST/DST and object/odor stimuli. Next, each frame was manually annotated, where a human annotator placed digital points-of-interest on the rat (Fig 2B). Manually annotated frames were used to train a deep neural network-based model to predict the spatial location of points of interest for each frame across new videos. Nath and colleagues (2019) describe the procedure used in the present experiments for model training and subsequent video analysis using DeepLabCut. A pre-trained ResNet-50 convolutional neural network (CNN) was then trained on 95% of annotated frames for 200,000 iterations, where 5% of frames were reserved for model assessment. After training, we analyzed the CNN learning curve to select an optimal model that performs well on both test and train data. Pose-estimation data was extracted from videos using a model trained for 80,000 iterations, which represents the iteration where test error is minimized, and the training error is saturated. Our model produced a training error of

4.89 and a test error of 4.35 using the default hyperparameters, without a p-cutoff filter applied. Finally, pose-estimation tracking files were filtered using the DeepLabCut native median filter model. It is important to note that annotated training frames for this experiment were added to an existing DLC project (training set = ~1,000 annotated frames). As the CNN was pretrained to predict the spatial position of key points, and all videos were filmed within an identical experimental apparatus, the number of additional required annotated frames to acquire high-fidelity pose-estimation data for the present experiment was likely lower than if the CNN was trained from scratch. The DLC model file used for analysis is freely available on GitHub (**https://github.com/HowlandLab/ILBTJO_NODB_SimBA_2023**), and any additional training data will be freely supplied upon request.

We then trained a supervised machine learning-based behavioral classifier to predict interaction events based on movement features extracted from pose-estimation data (Goodwin et al., 2022). Nilsson and colleagues (2020) describe the detailed procedure used in the present experiments for model training and subsequent video analysis using SimBA. Classifier training was completed using the eight-point classical tracking version of the SimBA pipeline (SimBA-UW-tf-dev = 1.32.2). We trained two classifiers, one for object-based stimuli and one for odor-based stimuli, to predict interaction events across test variation. For each classifier, the training dataset consisted of user-annotated frames from ~30 five-minute videos, where each frame was assigned a binary label of "interaction" or "non-interaction". The object-based and odour-based classifiers were trained on 28,586 and 32,872 frames of target "interaction" behavior, respectively. Prior to manual annotation, trimmed videos and filtered pose-estimation data was imported, then a scale factor was used to normalize variable camera filming heights to a known metric distance (experimental apparatus, dimensions = 60cm x 60cm). Additionally, each stimuli position was assigned a region-of-interest that was centered at each Velcro stimuli attachment point, with a defined radius extending ~2cm beyond the edge of stimuli. In total, 273 features were extracted from tracking data, where 251 features capture spatiotemporal relationships between points-of-interest, and 12 features capture ROI-related movement. We slightly deviated from the standard SimBA feature engineering approach by removing ROI-related features called "zone_cumulative_percent" and "zone_cumulative_time". These features increase the prediction probability of a true class based on animal's preferentially spending time in a defined ROI. While these features may be useful for predicting behaviors that only include in specific regions (e.g., rat dams retrieving pups from a nest), inclusion of these features in our project would bias predictions unequally between the six stimuli positions. For both the object and odor classifiers, the behavioral features most heavily weighted for model predictions include distance to stimuli, nose movements, region-of-interest, and spatial dynamics between points-of-interest (Fig 2C). Feature importance clusters were created by extracting the 40 most important features from SimBA, then splitting features based on the following criteria: 1) features related to the distance to stimuli "distance to stimuli"; 2) features related to nose movements (e.g., Nose_movement_M1_sum_6) were clustered to "nose movements"; 3) features related to a subjects' nose key point being located within a defined ROI surrounding stimuli were clustered to "region-of-interest"; 4) remaining features were clustered to a common "spatial dynamics between points-of-interest". For the object classifier, we defined "interaction" as frames where the rat's nose was within 2 cm of the object, while looking at and/or chewing the stimuli for a duration greater than 50 msec. For the odor classifier, "interaction" was defined as frames where the rat's nose was within 2 cm of the top of the odor jar, while looking at and/or chewing the stimuli for a duration greater than 50 msec. Classifiers were built using the following hyperparameter set: n_estimators = 200, RF_criterion = entropy, RF_max_features = sqrt, RF_min_sample leaf = 2 (Extended Data Fig 2-2,2-3,2-4). Precision, recall, and F1 scores for the classifiers are shown in Fig 2D,E and further described in the Extended Data. To account for instances of sub-optimal supervised machine learning prediction, we created a five-tiered verification rank system, where supervised machine learning-generated predictions on videos with ranks of four or five were replaced by human stopwatch scoring for the final analysis (Fig 3C,D).

<u>Statistical Analysis</u>

For all analyses, the entire 5 min of the sample or test phase was analyzed. Total stimuli exploration times were calculated by taking the sum of the time spent interacting with each stimulus, as measured in sec. A discrimination ratio (DR) was calculated for each test session, which reflects the time spent with the novel stimulus compared to the average time spent with the familiar stimuli. This metric is calculated by the equation DR = (T (novel) – T (avg. familiars) / T (total)), and produces a ratio between -1 and +1, that indicates a familiar and novelty preference, respectively. A DR was also calculated for interaction bout count, while untransformed values were used to assess distance travelled and novel approach latency. Rats were excluded from the final analysis if all stimuli in the box were not visited in the sample phase, if an item was knocked over or moved, or if the video was blurry. From the test establishment experiments, 2 male rats were removed from the 3-object IST, 1 from the 3-odor IST, 1 from the 3-odor DST, and 1 from the 6-odor IST. Due to missing video footage, 8 values are missing from each 3- and 6- object IST and DST sample phase mean ± SEM calculations. From the acute *Cannabis* exposure interaction bout duration data, 6 videos were excluded from the 6-object IST, 2 from the 6-object DST, 1 from the 6-odor IST, and 2 from 6-odor DST. From the bout count data, 7 were excluded from the 6-object IST, 3 from the 6-object DST, and 2 from 6-odor DST.

Data were analyzed using GraphPad Prism 8.0.1 software. To evaluate the DR's generated from interaction times in the test validation and establishment experiment, one-sample t-tests were used against chance (i.e., 0). To evaluate the total exploration times in the test validation and establishment experiment, two-way ANOVAs (followed by Bonferroni's multiple comparisons test) with factors of Phase (sample vs test) and Item Count (3- vs 6-) were used. To evaluate the total exploration times following *Cannabis* smoke exposure, two-way ANOVAs (followed by Bonferroni's multiple comparisons test) with factors of Phase (sample vs test) and Treatment (Air Control vs high-THC [Skywalker] vs high-CBD [Treasure Island]) were used. Following *Cannabis* exposure, to evaluate the DR's and untransformed values measuring interaction time, bout count, distance travelled, and novel approach latency, one-way ANOVAs (followed by Turkey's multiple comparisons test) with a factor of Treatment (Air Control vs high-THC vs high-CBD) were used. Lastly, to evaluate the interaction time DRs (novelty preference) against chance, one-sample t-tests against 0 were used. P values that were < or = to 0.05 were considered significant.

Results

Male rats perform both the IST and DST with objects and odors, using either 3- or 6-stimuli

The 3- and 6-object IST and DST were validated for male rats by adopting protocols similar to those used with mice (Olivito et al., 2016, 2019; Sannino et al., 2012). Male rats spent significantly more time with the novel object in comparison to the familiar objects in the 3-object

IST [$t(14) = -6.29$, $p < 0.001$], and in the 6-object IST [$t(14) = -5.02$, $p < 0.001$] (Fig 1E). Male rats also displayed novelty preference in the 3-object DST [$t(16) = -5.09$, $p < 0.001$], and in the 6-object DST [$t(14) = -3.94$, $p < 0.001$] (Fig 1E). A comparison of the IST and DST DRs showed no differences between the 3-object [$t(30) = 0.98$, $p = 0.36$] or 6-object [$t(28) = 1.40$, $p = 0.17$] variations (Fig 1E). All treatment groups performed better than chance ($t(15) = 7.35$, $p < 0.0001$ (3-object IST); $t(14) = 8.41$, $p < 0.0001$ (6-object IST); $t(15) = 8.52$, $p < 0.0001$ (3-object DST); $t(14) = 7.31$, $p < 0.0001$ (6-object DST) (Fig 1E).

A significant effect of Phase was seen on the total stimuli interaction time in the IST with objects [$F(1, 39) = 9.63$, $p = 0.004$], with no effect of Item Count [$F(1, 39) = 1.62$, $p = 0.21$] or an interaction [$F(1, 39) = 0.11$, $p = 0.74$] present (Table 1). Male rats spent more time exploring stimuli in the sample phase of the object IST than the test phase. There was also a significant effect of Phase on the total stimuli interaction time in the object DST [$F(1, 39) = 13.89$, $p = 0.0006$], with no effect of Item Count [$F(1, 39) = 3.78$, $p = 0.059$] or an interaction [$F(1, 39) = 2.61$, $p = 0.11$] present (Table 1). Inspection of the data revealed that in the object DST, male rats spent more time exploring stimuli in the sample phase than the test phase.

In the tests with odors, male rats also showed novelty preferences in the 3- and 6- odor IST and DST (Fig 1F). Male rats spent significantly more time with the novel odor compared to the familiar odors in the 3-odor IST [$t(7) = -1.87$, $p < 0.05$] and 6-odor IST [$t(10) = -6.59$, $p < 0.001$] (Fig 1F). Novelty preference was also demonstrated in the 3-odor DST [$t(6) = -7.94$, $p < 0.001$], and in the 6-odor DST [$t(11) = -3.92$, $p < 0.01$] (Fig 1F). Lastly, no differences between the IST and DST DR's were found in the 3-odor [$t(13) = -1.44$, $p = 0.17$] or 6-odor [$t(21) = 1.60$, $p = 0.12$] variations (Fig 1F). All treatment groups performed better than chance ($t(7) = 5.04$, $p = 0.0015$ (3-odor IST); $t(11) = 7.36$, $p < 0.0001$ (6-odor IST); $t(7) = 5.40$, $p = 0.0010$ (3-odor DST); $t(11) = 10.61$, $p < 0.0001$ (6-odor DST) (Fig 1F).

In the odor IST, there was no effect of Phase on the total stimuli interaction time [$F(1, 36) = 1.16$, $p = 0.29$], but a main effect of Item Count [$F(1, 36) = 4.55$, $p = 0.040$] and a significant interaction was present [$F(1, 36) = 4.24$, $p = 0.047$] (Table 1). Male rats spent more time exploring odors in the sample phase of the 6-odor IST than in the 3-odor IST ($p = 0.031$). In the odor DST, there was no main effect of Phase [$F(1, 36) = 2.34$, $p = 0.14$], Item Count [$F(1, 36) = 3.79$, $p = 0.06$] or an interaction [$F (1, 36) = 1.49$, $p = 0.23$] present (Table 1).

Combining automated and human stopwatch scoring is a valid behavioral quantification approach

To quantify rat behavior following *Cannabis* smoke exposure using the hybrid scoring method, we created a video set of 288 test phase videos of the 6-stimuli test variations. Sample phase videos were all manually scored, where inclusion criterion was applied as described above, and included test phase videos were analyzed using our automated behavioral quantification pipeline.

To assess the accuracy of model predictions for both pose-estimation and behavioral classification, we utilized software native performance metrics that compare machine-generated predictions to manual annotation. The spatial coordinates of human annotated and machine-predicted points-of-interest differed by a mean Euclidian distance of 4.89 pixels on videos within the model training set and 4.35 pixels on test videos. Pose-estimation quality was further assessed by calculating the average prediction confidence for each point-of-interest by video (Extended Data Fig 2-1). We found that the average prediction confidence ranged between 92.8% and 97.4% by point-of-interest, where no significant differences were observed between object-based and odor-based videos. Behavioral classifier performance was evaluated by a series of confusion matrices (Fig 2D,E) that report the precision, recall, and combined F1 score for each model. In short, both classifiers demonstrate high precision and recall (object F1 = 0.927, odor F1 = 0.897) when assessed by comparing manual annotation to classifier predictions on randomly selected test video frames. However, when classifier performance was assessed by comparing predictions on randomly selected interaction bouts, object classifier performance changed marginally (F1 = 0.93), but odor classifier performance decreased markedly (F1 = 0.63). For both the object and odor classifiers, the behavioral features most heavily weighted for model predictions include distance to stimuli, nose movements, region-of-interest, and spatial dynamics between points-of-interest (Fig 2C). Additional details regarding model training and assessments can be found in the Extended Data.

To verify the reliability of supervised machine learning-generated predictions relative to traditional stopwatch-based and automated region of interest-based scoring, we conducted a three-way correlational analysis on generated interaction DR's (Fig 3A,B). We found that, across stimuli, supervised machine learning-generated predictions were more highly correlated with human stopwatch scoring than region of interest-based scoring; however, supervised machine learning-generated predictions were more highly correlated with human stopwatch scoring for object interaction (r = 0.75) relative to odor interaction (r = 0.53). Additionally, we found that, across stimuli, region of interest-based scoring held a weaker correlation relative to both human stopwatch scoring (object: r = 0.42, odor: r = 0.28) and supervised machine learning-generated (object: r = 0.45, odor: r = 0.42) interaction DR's. To account for instances where supervised machine learning predictions significantly differ from human stopwatch scoring, we created a five-tiered verification rank system, where supervised machine learning-generated predictions on videos with ranks four or five were replaced by human stopwatch scoring for the final analysis (Fig 3C). Upon visual inspection of supervised machine learning-generated predictions, we found that ~80% of object-based videos met inclusion criteria, while only ~60% of odor-based videos met inclusion criteria (Fig 3D). To justify supplementing human stopwatch scoring for sub-optimal supervised machine learning -generated predictions, we conducted a correlational analysis between human stopwatch scoring and supervised machine learning interaction DR's only on videos which met inclusion criteria. We found that human stopwatch scoring and supervised machine learning interaction DR's were moderately-to-highly correlated (Fig 3E: r = 0.83, Fig 3F: r = 0.87) across stimuli type.

High-THC, but not high-CBD, *Cannabis* smoke exposure impairs novelty preference for high- (DST) memory loads with object stimuli

Interaction bout duration DR's were investigated to examine if novelty preference was impacted by treatment within each test variation. No effect of Treatment in the 6-object IST [F(2, 61) = 0.85, p = 0.43] was found (Fig 4C). Using an analysis of the raw effect sizes, there were no notable effect sizes to report (Table 3). A main effect of Treatment was present in the 6-object DST [F(2, 63) = 3.75, p = 0.03], with a significant difference seen between the Air Control and high-THC groups after a Tukey's multiple comparisons test (p = 0.04) (Fig 4C). The difference between the Air Control and high-THC groups represents a moderate effect size [d = -0.66, 95% CI (1.27, -0.035), p =0.03] (Table 3). Most treatment groups performed significantly better than chance (IST-Air Control: t(23) = 3.15, p = 0.004; IST-high-THC: t(19) = 2.24, p = 0.037; IST-high-CBD: t(19) = 4.27, p = 0.0004; DST-Air Control: t(18) = 3.29, p = 0.004; DST-high-CBD: t(24) = 2.14, p = 0.042) except for the high-THC group in the 6-object DST (t(22) = 0.66, p = 0.51) (Fig 4C).

We then investigated novel approach latency values, defined as the interval between rats being placed into the experimental arena and interacting with the novel object. No effect of Treatment on novel approach latency values was observed in either the 6-object IST [$F(2, 70) = 0.77$, $p = 0.46$] or the 6-object DST [$F(2, 67) = 0.076$, $p = 0.93$] (Fig 4D). Next, to examine if male rats visited the novel object at a higher frequency than familiar objects, we evaluated the interaction bout DR's (Fig 4E). Here, we showed a significant main effect of Treatment in the 6-object IST [$F(2, 64) = 8.05$, $p < 0.001$], as the Air Control ($p = 0.001$) and high-THC ($p = 0.01$) groups were different from the high-CBD group. However, we failed to find a main effect of Treatment on bout count DR's in the 6-object DST [$F(2,64) = 0.96$, $p= 0.39$] (Fig 4E). Lastly, the impact of *Cannabis* smoke exposure on locomotion during memory testing was evaluated. We found no main effects of Treatment on distance in either the 6-object IST [$F(2, 70) = 0.58$, $p = 0.56$], or in the 6-object DST [$F(2, 67) = 0.30$, $p = 0.74$] (Fig 4F).

When assessing total stimuli interaction time, a main effect of Treatment [$F(2,129) = 4.07$, $p = 0.019$], and of Phase [$F(1, 129) = 6.45$, $p = 0.012$] was seen in the 6-object IST, with no significant interaction [$F(2, 129) = 0.49$, $p = 0.62$] (Table 2). In the 6-object DST, there was a main effect of Phase on total stimuli interaction time [$F(1, 135) = 7.87$, $p = 0.0058$], with no main effect of Treatment [$F(2, 135) = 1.81$, $p = 0.17$] or an interaction [$F(2, 135) = 0.75$, $p = 0.47$] (Table 2). Following each smoke treatment, the number of boli was counted in the smoke exposure cage (Fig 6). A main effect of Treatment was observed [$F(2, 141) = 172.90$, $p < 0.0001$], with a significant increase in the number of boli recorded following either Skywalker ($p < 0.0001$) or Treasure Island ($p < 0.0001$) smoke exposure after a Tukey's multiple comparisons test. However, there was no difference in the number of boli observed between Skywalker or Treasure Island ($p = 0.40$) smoke exposure groups.

High-THC, but not high-CBD, *Cannabis* smoke exposure impairs novelty preference for high- (DST) and low- (IST) memory loads with odor stimuli

Cannabis smoke exposure impacted the interaction bout duration DRs in the IST and DST. An effect of Treatment in the 6-odor IST [$F(2, 73) = 3.54$, $p = 0.034$] was seen, with a significant difference present between the Air Control and high-THC groups (Tukey's multiple comparisons test, $p = 0.046$) (Fig 5C). A moderate effect size was found between the high-THC and Air Control groups [$d = -0.78$, 95% CI (1.41, -0.19), $p =0.0058$] (Table 3). A main effect of

Treatment for interaction bout duration DRs was also present in the 6-odor DST [F(2, 71) = 4.3, p = 0.017], with a significant difference between the Air Control and high-THC groups (p = 0.024) and between high-THC and high-CBD groups (p = 0.046) after a Tukey's multiple comparisons test (Fig 5C). A moderate effect size was also found between the high-THC and Air Control groups [d = -0.87, 95% CI (1.47, -0.23), p =0.0042] (Table 3). Air Control and high-CBD treatment groups performed significantly better than chance in both tests (IST-Air Control: t(25) = 5.90, p < 0.001; IST-high-CBD: t(22) = 2.47, p = 0.022; DST-Air Control: t(23) = 3.45, p = 0.002; DST-high-CBD: t(27) = 2.25, p = 0.033), whereas the high-THC group did not in either the 6-odor IST (t(26) = 0.47, p = 0.64) or 6-odor DST tests (t(21) = 1.00, p = 0.33) (Fig 5C). There was no effect of Treatment in the 6-odor IST [F(2, 77) = 0.036, p = 0.70], or in the 6-odor DST [F(2, 71) = 0.87, p = 0.42] when investigating novel approach latency (Fig 5D). Interaction bout DR's were also determined to be unaffected by *Cannabis* exposure with no effect of Treatment in the 6-odor IST [F(2, 77) = 1.46, p = 0.24], and the 6-odor DST [F (2, 70) = 2. 19, p= 0.12] (Fig 5E). Treatment also did not impact the distance travelled by male rats in either the 6-odor IST [F(2, 77) = 0.36, p = 0.70], or in the 6-odor DST [F(2, 71) = 0.87, p = 0.42] (Fig 5F).

For exploration times in the 6-odor IST, a main effect of Treatment [F(2,142) = 3.78, p = 0.025], and of Phase [F(1, 142) = 12.90, p = 0.0004] was seen, with no significant interaction [F(2, 142) = 2.27, p = 0.11] (Table 2). Male rats spent more time exploring stimuli in the Air Control sample phase than in the high-THC test phase (p = 0.017). As well, male rats explored stimuli more in the sample phase than in the test phase following high-THC (p = 0.0035), while spending more time exploring stimuli in the test phase following high-THC smoke exposure than following high-CBD smoke exposure (p = 0.009). In the 6-odor DST, there was a main effect of Phase on total stimuli interaction time [F(1, 134) = 10.01, p = 0.0019], with no main effect of Treatment [F(2, 134) = 0.021, p = 0.98] or an interaction [F(2, 134) = 0.85, p = 0.43]. Inspection of the data revealed that male rats spent more time exploring the odors during the test phase of the 6-odor DST, regardless of Treatment (Table 2).

Discussion

In the present study, we showed that male rats display novelty preferences in both the IST and DST with 3 and 6 objects, similar to previous findings using objects in male mice (Olivito et al., 2016, 2019; Sannino et al., 2012). We also demonstrate, for the first time, that male rats exhibit novelty preference with 3 and 6 odor stimuli, as measured in the IST and DST (Fig 1). Overall, male rats spent more time exploring stimuli in the sample phases of the 6 item IST and DST compared to the test phases, with stimuli-specific differences (Table 1). Following high-THC *Cannabis* smoke exposure in the tests with objects, a significant decrease in novelty preference was seen in the 6-object DST, but not in the 6-object IST (Fig 4C). However, for odor-based tests, we observed novelty preference impairments for high- and low-memory loads (Fig 5C). No notable treatment effect on total stimuli exploration time was present in the 6-object IST, but a significant increase in stimuli exploration time was seen in the test phase of the 6-object DST for all treatments (Table 2). In the 6- odor IST, male rats explored stimuli less in the sample phase compared to the test phase following high-THC *Cannabis* smoke exposure, with no notable effects in the 6-odor DST (Table 2). Taken together, these findings suggest that *Cannabis* smoke exposure impacts novelty preference in male rats in a load-dependent and stimuli- specific manner.

Male rats demonstrate novelty preference in both the IST and DST with objects and odors

In the test validation experiment, male rats demonstrated pronounced novelty preference in all test variations (Fig 1). The preferential interaction with novel stimuli compared to familiar stimuli after a brief delay suggests that recognition memory is intact in both object and odor-based tests (Sannino et al., 2012; Shrager et al., 2008; van Vugt et al., 2017). The varying memory loads between the IST and DST also present the opportunity to examine incidental memory capacity (Sannino et al., 2012; Shrager et al., 2008). In this study, 3- and 6-item tests were run to replicate Sannino and others' (2012) results showing that male mice demonstrated novel object discrimination when using up to 6 objects. To enable direct comparisons between object and odor stimuli, sets of 3 odors and 6 odors were chosen as well. Male rats explored the object stimuli a comparable amount between test variations and with varying numbers of stimuli (Table 1). Male rats did, however, spend significantly less time exploring objects in the test phase of the 6-object DST compared to the sample phase (Table 1). As the test phase progressed, male rats would have had increasing familiarization with all items in the test phase, which may explain the decreased total exploration times (Broadbent et al., 2010). Interestingly, there were no notable differences in the total stimuli interaction times between the 3-odor and 6-odor variations, indicating that while the total time male rats spent exploring stimuli was the same, the time spent exploring each individual stimulus in the 6-item variation was about half of that for the 3-item variation (Table 1). In future experiments, it would be interesting to assess novelty preferences and exploration preferences in test with more than 6 stimuli, as has been reported for objects in male mice (Sannino et al., 2012). As well, these tests must be validated for use in female rats. Recent findings show sex differences in delay-dependent incidental memory capacity for objects in mice, which may depend on sub-cortical inhibitory control of the hippocampus (Torromino et al., 2022). These findings in mice raise the possibility that similar sex differences exist in rats, a question that will be investigated in future experiments. Validating the odor-based spontaneous tests in male and female mice would also be worthwhile given their affordability and availability of genetic models.

The IST and DST allow the study of novelty preferences for stimuli arrays of varying size in a spontaneous, simple, and cost-effective manner. The tests do not require rodents to apply learned rules or procedures, eliminating the need for extensive training or researcher involvement. The tests also evoke minimal stress in rodents and do not require typical food-restriction protocols to increase reward-driven performance. Performance on the object tests likely engage a combination of visual and tactile recognition memory, but as the object stimuli were constructed with LEGO™ blocks of similar size, identical smooth textures, and sharp corners, the tests were likely biased to engage visual recognition memory. The object-based test may engage visual, perirhinal, medial prefrontal, parietal, and entorhinal cortices, as well as the hippocampus and thalamus to enable the object-based recognition memory across a delay (Barker et al., 2007; Cazakoff & Howland, 2011; Churchwell & Kesner, 2011; Creighton et al., 2018; Dere et al., 2007; Fernandez & Tendolkar, 2006; Hannesson et al., 2004; Peters et al., 2013; Sugita et al., 2015; Winters et al., 2004). The odor stimuli primarily engage odor-based recognition as identical opaque glass jars were used in the tests. A circuit including piriform, entorhinal, medial prefrontal, and orbitofrontal cortices, along with hippocampus may be involved in the odor-based memory across a delay (Alvarez & Eichenbaum, 2002; Davies et al., 2013; Mouly & Sullivan, 2010; Ramus & Eichenbaum, 2000; Sandini et al., 2020). To examine the brain regions and neural mechanisms underlying working memory capacity in different contexts, a variety of behavioral tasks have been employed. Visuospatial working memory and working memory capacity are examined with the radial-arm maze, Barnes Maze, and operant delayed nonmatching-to-sample and delayed-match-to-sample tasks (Barnard et al., 2022;

Cowan, 2010; Daneman & Carpenter, 1980; Dudchenko, 2004; Dudchenko et al., 2013; Kirchner, 1958; Oomen et al., 2013; Scott et al., 2020; Vorhees & Williams, 2014; Wilhelm et al., 2013). To study odor based working memory capacity, the odor span task and other tests that employ a nonmatch-to-sample-rules have often successfully been used (Dudchenko et al., 2000; Scott et al., 2020). Although these tasks measure working memory capacity, they require food restriction, extensive training, and heavy researcher involvement. Spontaneous recognition tests circumvent these weaknesses, although the cognitive processes involved in incidental memory capacity may differ from those necessary for more goal-directed forms of working memory capacity.

High-THC, but not high-CBD, *Cannabis* smoke exposure impairs novelty preferences for both object and odor stimuli

To evaluate the effects of *Cannabis* smoke exposure on incidental memory over short delays, we used the hybrid scoring approach to assess novelty preference in the IST and DST with objects and odors. The 6-item object and odor tests were selected as they would be expected to engage circuits related to capacity, while still ensuring reliable performance in control groups, as previously established in mice (Sannino et al., 2012; Torromino et al., 2022). Novelty preference was primarily inferred from interaction bout duration, as it was not predicted by interaction bout count or novel approach latency. Following high-THC *Cannabis* smoke exposure in the tests with objects, a significant decrease in novelty preference was seen in the 6-object DST, but not in the 6-object IST (Fig 4C). For odor-based tests, an impairment in novelty preference was observed in both the IST and DST following high-THC *Cannabis* smoke exposure (Fig 5C). In all tests, novelty preference was similar between the Air Control and high-CBD *Cannabis* smoke groups. Additionally, no differences in locomotion were observed among treatment groups. The increased total stimuli exploration time in the sample phases of the object DST compared to the test phases likely indicates familiarity with the items in the test phase that were previously presented during the sample phase (Broadbent et al., 2010). Interestingly, in the 6-odor IST, there was lower stimuli exploration time in the sample phase compared to the test phase following high-THC *Cannabis* smoke exposure (Table 2).

Overall, the deficits in novelty preference following high-THC *Cannabis* smoke exposure in both the object and odor-based tests in male rats are likely attributable to the actions of THC, and not to smoke alone. Interestingly, boli excretion was increased following acute *Cannabis* smoke exposure, but with no differences observed between the high-THC and high-CBD groups (Fig 6). As novelty preference was comparable between the Air Control and high-CBD groups, smoke likely did not provoke stress-induced performance deficits. As behavioral testing was conducted 20 min following the initiation of *Cannabis* smoke exposure, plasma and brain THC concentrations would have been near their peak in the rats (Baglot et al., 2021; Barnard et al., 2022; Hložek et al., 2017; Moore et al., 2022; Ravula et al., 2019). Analysis of plasma from male rats following an identical *Cannabis* smoke exposure paradigm revealed levels of 14.55 ± 1.59 ng/mL with a small amount of CBD (1.98 ± 0.38 ng/mL) 30 min after smoke exposure (Barnard et al., 2022). After high-CBD smoke exposure, negligible amounts of THC were found in plasma, along with 4.47 ± 1.15 ng/mL of CBD (Barnard et al., 2022). Thus, the current smoke exposure protocol increases blood plasma levels of THC to the low end of what is typically observed in humans following *Cannabis* cigarette consumption (Grotenhermen, 2003; Huestis, 2007; Huestis et al., 1992; Newmeyer et al., 2016; Moore et al., 2022; Ramaekers et al., 2009). Although the THC plasma levels in male rats were comparably low, we still observed the impact of *Cannabis* exposure on memory. The different THC-induced novelty preference impairments seen in the male rats between objects and odors may be due to the varying neural circuits underlying stimulus perception and integration (Constantinidis & Klingberg, 2016; Eriksson et al., 2015; Fernandez & Tendolkar, 2006; Galizio, 2016; Mouly & Sullivan, 2010). Under low memory loads (IST), treatment does not impact object novelty preference, consistent with unperturbed WM performance previously observed in a 2-item novel object recognition (NOR) test following chronic exposure to 5.6% THC *Cannabis* cigarettes (Bruijnzeel et al., 2016). The novelty preference deficits observed following high-THC *Cannabis* exposure in the 6-odor IST also might have been affected by the decreased exploration time in the sample phase. Lastly, the similar THC-induced deficits in the DST with objects and odors could be due to sensitivity of the working memory subconstructs evoked under high memory loads to *Cannabis* exposure (Barch & Smith, 2008).

The case for, and caveats of, supervised machine learning-based behavioral analysis at scale

Automated behavioral analysis represents a potential paradigm shift in the way behavioral data are generated and shared (Mathis et al., 2020). In the present study, we demonstrate the case for, and caveats of, using a supervised machine learning-based analysis method for complex behavior at scale. In short, pose-estimation data was used to train two behavioral classifiers to predict interaction events with objects and odors. To assess the reliability of supervised machine learning-generated behavioral predictions, we compared quantified rat-stimulus interaction to human stopwatch and region of interest based scoring. We found that supervised machine learning-generated predictions were more strongly correlated with human stopwatch than region of interest-based scoring; however, we observed that supervised machine learning-generated predictions were more highly correlated with human stopwatch-based scoring for object stimuli than for odor stimuli. As a methodological validation control, we conducted an inter-rater variability analysis to ensure that comparison of human stopwatch and supervised machine learning behavioral scoring is generalizable to manual scorers of varying experience levels (Extended Data Fig 3-1). In short, we found a strong correlation between scorers of all experience levels ($0.85 < r < 0.94$), but a comparatively weaker correlation between experienced and beginner scorers. While a generally strong correlation between all scorers reinforces human stopwatch scoring as a gold-standard, experience-dependent changes in scoring accuracy underscore the value of high-throughput and objective scoring methods, such as the supervised machine learning-based method employed in this study.

Upon visual inspection of supervised machine learning-generated predictions, a near 30% increase in the proportion of excluded supervised machine learning-based odor interaction DR's is striking given that each classifier was trained on the same number of training frames, used identical algorithmic hyperparameters, and no significant treatment differences were observed in the proportion of excluded videos (Extended Data Fig 3-2). We propose that this difference may be explained by divergent operational definitions of interaction in object and odor tests. Rat-object events encompassed interaction along the entire height of the object, while rat-odor interaction was only counted at a narrow space around the lid of the mason jar. As we employed a 2-dimensional (2D) pose-estimation approach, movements along the height of stimuli were not well captured, potentially leading to sub-optimal predictions and grounds for exclusion. While classifiers trained on 2D pose-estimation data show reliability on classifying behaviors restricted to single-plane spatiotemporal movements, recent studies of complex behaviors, such as selfgrooming, generally train classifiers on 3D pose-estimation data to better capture the entirety of a movement and to minimize occlusion (Marshall et al., 2021, 2022; Minkowicz et al., 2023; Newton et al., 2023). Said differently, our assumption is not that the manual scorer and algorithm are using fundamentally different patterns of rat movement to infer behavior, but rather that the human is able to innately infer 3D from a 2D video, which is an important clue for interaction with stimuli that is not well captured in the automated analysis. Finally, software native performance metrics for both behavioral classifiers closely mirror those reported in published studies utilizing supervised machine learning-based analysis; however, manual verification of predictions revealed significant instances of misclassification (Newton et al., 2023; Winters et al., 2022). We contend that supplementing classifier performance metrics with correlational analysis and verification steps are best practices when conducting scaled automated behavioral analysis.

While a full review of best practices in automated behavioral analysis approaches is beyond the scope of this study and has been reviewed in detail by others (Luxem et al., 2022; Mathis et al., 2019), hardware and software optimization is critical for promoting model generalizability. First, to fully capture behaviors of interest, researchers utilizing automated behavioral analysis should be cognisant of the angle, and number, of camera perspectives used during filming (Luxem et al., 2022). Additionally, it is essential to include a diversity of training examples during model training, as a high degree of diversity in a training set will lead to a high degree of generalizability for both pose-estimation (DeepLabCut) and subsequent supervised machine learning-based analysis (SimBA). For example, within the present study, differences in color contrast, filming angle, and resolution likely contributed to a lack of DeepLabCut model generalizability between videos filmed for test validation (Figure 1) and *Cannabis* manipulation (Figure 4, Figure 5). Taken together, supervised machine learning-based analysis is a promising tool for behavioral neuroscience, but this approach still faces some significant limitations, and researchers should adhere to available best practices to maximize the reliability of behavioral measurements.

Conclusion

Using novel spontaneous tests and a hybrid scoring method, the impact of acute exposure to high-THC or high-CBD *Cannabis* smoke on incidental memory was evaluated in male rats.

We show impaired object-based novelty preference after high-THC, but not high-CBD, *Cannabis* smoke exposure under a high-memory load. As well, we show deficits in odor-based novelty preference following high-THC *Cannabis* smoke exposure under both low- and high-memory loads. Ultimately, these data indicate that *Cannabis* smoke exposure impacts novelty preference in a load-dependent, and stimuli- specific manner in male rats.

Figure Captions

Figure 1. The validation and establishment of the IST and DST with objects and odors. A A picture of an example object set-up is shown. Objects are displayed in 6 positions in a white-corrugated plastic box. **B** A picture of an example odor set-up is shown. Odors are displayed in 6 positions in a white-corrugated plastic box. **C** An example of an object stimuli. **D** An example of an odor stimuli. **E** Object interaction was measured using DR's to evaluate novelty preference using 3-objects and 6-objects. Male rats explore the novel object significantly more than the familiar objects in the IST and DST with both 3- and 6- objects. No differences in novelty preference or exploration times are seen between the IST and DST, or between 3-object and 6-object versions. **F** Odor interaction was also measured using DR's to evaluate novelty preference using 3-odors and 6-odors. Male rats explore the novel odor significantly more than the familiar odors in the IST and DST with both 3- and 6- odors. No differences in novelty preference or exploration times are seen between the IST and DST, or between the 3-odor and 6-odor versions. Data is represented as mean ± SEM.

Figure 2. Experimental overview for acute *Cannabis* exposure and behavioral classifier training. A Schematic representation of the experimental design. Male Long-Evans rats (n = 48) were used for this study. Using a repeated measures experimental design, each rat was exposed to high-THC *Cannabis* smoke, low-THC *Cannabis* smoke, and an Air Control condition. Male rats were exposed 20 minutes prior to the start of behavioral testing. Each male rat either underwent the 6-object IST and 6-object DST, or the 6-odor IST and 6-odor DST. The order in which the IST and DST was performed was randomized. Rat behavior was quantified using traditional stopwatch scoring and by automated SML-based behavioral analysis. Sub-optimal SML predictions were replaced by stopwatch scoring, constituting a hybrid scoring approach. **B** Illustration of the point-of-interest configuration used for pose-estimation analysis. We chose the number and position of points in accordance with the SimBA eight-point configuration. SimBA requires a standardized and specific position (and number) of points. Users should decide what SimBA configuration will be used (single animal, multi animal, point number) prior to network training with DeepLabCut. **C** Visualization of the relative feature importance of the four features clusters. In short, the 40 most important features were systematically categorized into distinct clusters, then we summed the feature importance's of individual features within each cluster. The raw features importance log is included under "assessment + logs" for each classifier within our

GitHub repository. **D** Classifier performance metrics for the object (top) and odor (bottom) models. Test frames were randomly extracted from the dataset (20% test, 80% train). **E** Classifier performance metrics for the object (top) and odor (bottom) models. Test bouts were randomly extracted from the dataset (20% test, 80% train). See Extended Data Figs 2-1 to 2-4 for more information regarding the supervised machine learning approach and validation. This figure was created using BioRender.com.

Figure 3. Comparison between human stopwatch and supervised machine-learning generated output. **A** Correlation matrix between methods of quantifying rat-object interaction. This comparison was made between supervised machine-learning (SML), human-stopwatch (HS), and region-of-interest (ROI), generated interaction times. Interaction times by object was quantified using each scoring method, then the correlation between interaction DR's was assessed. **B** Correlation matrix between methods of quantifying rat-odor interaction. Interaction times by odor was quantified using each scoring method, then the correlation between interaction DR's was assessed. **C** Criteria used to rank automated classification. Each video was manually viewed for accurate classification, where a verification rank was assigned based on objective criteria. **D** Frequency of verification rank assignment by type of stimuli. Videos with a verification rank less than three were excluded from final analysis and replaced by human stopwatch scoring. Approximately 80% of object videos and 60% of odor videos met inclusion criteria, respectively. **E** Correlation between human stopwatch and ML-generated DR's on object videos meeting inclusion criteria, indicating a moderate-to-high correlation ($r(109) = .83$, $p < .0001$). **F** Correlation between human stopwatch and ML-generated DR's on odor videos meeting inclusion criteria, indicating a moderate-to-high correlation ($r(77) = .87$, $p < .0001$). See Extended Data Figures 3-1 nad 3-2 for additional information regarding the scoring and the ranking of videos by *Cannabis* treatment.

Figure 4. High-THC *Cannabis* smoke exposure impacts novelty preference under high- (DST) memory loads using object stimuli, with no impact on distance travelled, frequency of item visitation, or approach latencies. **A** An example IST with objects is visualized, showing 6 identical objects in the sample phase, with a novel object introduced after a 1-minute delay in the test phase. **B** A DST with objects variation is shown, with an identical test progression, but instead starts with 6 different objects in the sample phase. **C** Interaction measured as time spent with an object was generated using the human-machine hybrid scoring

approach and visualized using a discrimination ratio for both variations using object stimuli. No difference in treatment groups is seen in the 6-object IST (n = 64). In the 6-object DST (n = 66), a significant decrease in novelty preference is seen in the SW group in contrast to the AC group (p = .04). **D** The mean novel approach latency in the 6-object IST (n = 72) and 6-object DST (n = 69) variations is shown to be consistent between treatment groups. **E** To illustrate the frequency of visitations to the novel object in comparison to the familiar objects, bout counts are visualized using a discrimination ratio. A preference for novel visitations is seen in the 6-object IST (n = 65) AC and SW groups, with no difference in item visitations in the 6-object DST (n = 66). **F** The distance travelled (cm) in the 6-object IST (n = 72) and 6-object DST (n = 69) variations is comparable across treatment groups. Data represents mean ± SEM. *$p < 0.05$. Abbreviations: High-THC *Cannabis* smoke (SW), high-CBD *Cannabis* smoke (TI), Air Control (AC). This figure was created using BioRender.com.

Figure 5. High-THC *Cannabis* smoke exposure impacts novelty preference under high- (DST) and low- (IST) memory loads using odor stimuli, with no impact on distance travelled, frequency of item visitation, or approach latencies. **A** An example IST with odors is visualized, showing 6 identical items in the sample phase, with a novel odor introduced after a 1-minute delay in the test phase. **B** A DST with odors variation is shown, with an identical task progression, but instead starts with 6 different odors in the sample phase. **C** Interaction measured as time spent with an odor was generated using the human-machine hybrid scoring approach and visualized using a discrimination ratio for both variations using odor stimuli. In the 6-odor IST (n = 75), a significant decrease in novelty preference is seen in the AC group in comparison to the SW group (p = .046). Whereas in the 6-odor DST (n = 73), a significant decrease in novelty preference is seen in the SW group from both the AC (p = .023) and TI (p = .046) groups. **D** The mean novel approach latency in the 6-odor IST (n = 79) and 6-odor DST (n = 73) variations is shown to be consistent between treatment groups. **E** To illustrate the frequency of visitations to the novel odor in comparison to the familiar odors, bout counts are visualized using a discrimination ratio. No differences between treatment groups or 6-odor IST (n = 79) and 6-odor DST (n = 73) is seen. **F** Distance travelled (cm) in the 6-odor IST (n = 79) and 6-odor DST (n = 73) variations is comparable across treatment groups. Data represents mean ± SEM. *$p < 0.05$. Abbreviations: High-THC *Cannabis* smoke (SW), high-CBD *Cannabis* smoke (TI), Air Control (AC). This figure was created using BioRender.com.

Figure 6. Boli count following smoke exposure treatment. A significant increase in the number of boli recorded was observed following Cannabis smoke exposure in comparison to the Air Control (AC) condition. However, no difference between Skywalker (SW) or Treasure Island (TI) groups was recorded. **** p < .001.

	OBJECT IST		**OBJECT DST**		**ODOR IST**		**ODOR DST**	
	Sample*	**Test***	**Sample***	**Test***	**Sample**	**Test**	**Sample**	**Test**
3 ITEMS	71.45 ±12.1	47.98 ±6.5	68.43 ±13.4	104.43 ±18.9	31.99$^{\#}$ ±7.3	58.23 ±5.3	35.69 ± 8.4	54.74 ± 5.9
6 ITEMS	63.50 ±5.4	34.30 ±4.1	47.06 ±5.	50.39 ±6.9	38.14$^{\#}$ ±7.6	38.59 ±5.2	33.83 ± 6.3	38.20 ± 3.6

Table 1. Summary of all interaction times for validation of the tests summarized in Fig 1. The mean (± SEM) for the total interaction time seen with stimuli is recorded for each sample and test phase in the IST and DST with objects or odors. * Significant main effect of Phase on object IST and DST ($p<0.05$). $^{\#}$ Significant effect of Item Count on exploration times in the sample phase of the odor IST ($p = 0.047$).

	OBJECT IST		OBJECT DST		ODOR IST		ODOR DST	
	Sample*	**Test***	**Sample#**	**Test#**	**Sample&**	**Test&**	**Sample%**	**Test%**
Air Control	36.21 ±2.9	42.93 ±4.0	35.61 ±3.2	39.23 ±3.4	37.75 ±2.8	47.78 ±5.8	39.16 ±3.1	50.12 ±5.6
high-THC	36.01 ±3.7	46.90 ±4.1	39.65 ±3.5	49.72 ±4.6	34.27 ±3.1	57.94 ±4.8	35.29 ±2.8	55.27 ±6.5
high-CBD	30.09 ±3.0	33.97 ±2.7	33.9 ±3.1	46.96 ±4.2	31.54 ±2	36.93 ±5.5	40.54 ±3.4	48.03 ±6.1

Table 2. Summary of all interaction times for tests with *Cannabis* summarized in Figs 2-5. The mean (± SEM) for the total interaction time seen with stimuli is recorded for the sample and test phases in the different 6- object and 6- odor IST and DST across the Air Control, high-THC, and high-CBD treatment groups. * Significant effect of Treatment ($p = 0.019$) and of Phase ($p = 0.012$) on object IST. # Significant effect of Phase ($p = 0.0058$) on object DST. & Significant effect of Treatment ($p = 0.025$) and Phase ($p = 0.0004$) on odor IST. % Significant effect of Phase ($p = 0.0019$) on odor DST.

	AC-SW Cohen's d	AC-SW P value	AC-TI Cohen's d	AC-TI P value
6-object IST	-0.25 [95.0%CI -0.856, 0.357]	0.409	0.291 [95.0%CI -0.323, 0.872]	0.319
6-object DST	-0.655 [95.0%CI -1.27, -0.035]	0.03*	0.118 [95.0%CI -0.507, 0.716]	0.7
6-odor IST	-0.783 [95.0%CI -1.41, -0.194]	0.0058**	0.0239 [95.0%CI -0.539, 0.637]	0.936
6-odor DST	-0.874 [95.0%CI -1.47, -0.228]	0.0042**	-0.172 [95.0%CI -0.727, 0.413]	0.544

Table 3. Summary of the effect sizes (Cohen's d) and corresponding p-values for Fig 4C and 5C. The unpaired Cohen's d [confidence interval, lower bound; upper bound) for interaction times seen between novel and familiar stimuli is recorded for the test phases in the 6- object and 6- odor IST and DST across the Air Control, high-THC, and high-CBD treatment groups. * P < .05 ** P < .01 ***P < .001.

Extended Data Figure Captions

Fig 2-1. Mean tracking confidence for each point-of-interest, by video. To calculate the mean tracking confidence for each video, the average of the likelihood column associated with each point of interest was calculated.

Fig 2-2. Model hyperparameters used for classifier training. A meta-data csv file is included under "assessment + logs" for each classifier within our GitHub repository.
Previous studies have shown that creating a balanced dataset by using the model hyperparameters of "random under sampling" or "random over sampling" lead to better classifier performance; however, we found that using these features dramatically decreased classifier performance and lead to equal classifier predictions across the data frame. Therefore, we chose to not use these hyperparameters for analysis, and accounted for the unbalanced dataset by setting a relatively low discrimination threshold. For both classifiers, a discrimination threshold of 0.35 and a minimum bout duration of 50ms was used (Extended Data Fig 2-3).

Fig 2-3. Representative plot of classifier predictions across a complete video (9000 frames, 5 min video). We chose a discrimination threshold of 0.35 as it corresponds to the middle segment of obvious probability spikes and excludes the majority of noise below 0.2. We assessed model performance in two ways, both of which are integrated in the SimBA GUI (Extended Data Fig 2-2). First, we generated performance metrics (precision, recall, F1) by randomly splitting the aggregate training set (all human-annotated frames from all videos within the project) into 80% training frames and 20% test frames. Said differently, for a given behavioral video, a fraction of interaction-containing frames was used for model training, then a smaller fraction of frames was used for testing if the model can accurately predict if rat-stimulus interaction occurs in each test frame. As shown below, we found that both the object and odour classifiers generated excellent performance metrics when assessed in this manner. However, a fundamental problem with this assessment method is that for a given interaction bout, there may be both test and training frames, so the model is predicting interaction between two known sub-bouts of interaction (visualized- 1 = known interaction, test = test frame that the model must make a prediction on: 1-1-1-1-1-test-1-1-1-1). Therefore, to assess performance without the confound of intra-bout test frames, we segregated the aggregate training into interaction bouts, then split the segregated training set into 80% training bouts and 20% test bouts. We found that the performance of the object classifier changed marginally with this change, but performance metrics for the odor classifier significantly decreased when assessed in this manner. While we content that assessing classifier performance by-bout is a more conservative and representative method, an important caveat is that classifier performance on a completely model-naïve video is not assessed by either of these methods. This is important to consider because researchers will typically implement this analysis method to automatically quantify behavior for a large dataset, where only a fraction of this dataset is used for training. We did not include a by-video classifier analysis as this is not integrated into SimBA, but we contend that future research and software development should implement this performance assessment method to capture the accuracy of classifier predictions most accurately on model naïve behavioral videos.

Fig 2-4. Precision recall curve visualizing changes in precision, recall, and F1 with classifier training. Raw data is included under "assessment + logs" for each classifier within our GitHub repository. Recall, precision, and by extension the F1 score are calculated from the entries of a confusion matrix. A confusion matrix tells us, given a set of observations belonging to at least 2 different classes and a classifier that attempts to label each, how many and what type of errors were made. The diagonal of the confusion matrix is the correct observations, the off diagonal are the errors. For a binary classifier, we are generally focused on one class over the other, thus the metrics we derive are chosen to represent how we did for the most important class. In our case 'interaction' is the class we care about. In quantifying how our classifier for 'interaction' did, we calculate the recall and precision. Recall is the proportion of all the possible 'interaction' observations that our classifier predicted correctly. That is, the number of True Positives (TP) divided by the total number of 'interaction' observations (note the maximum number of True Positives is all the 'interaction' observations, in which case the recall equals 1, so a classifier that always predicts interaction will have perfect recall). Now there are many other metrics that could be computed, but the next most natural is the precision. Precision is the proportion of predicted 'interaction' observations that were actual 'interactions'. Or mathematically, the number of True Positives divided by the total number of times our classifier predicted 'interaction' (note it's not so easy to get perfect precision). Now we have 2 perfectly good numbers that quantify how our classifier did, the proportion of overall 'interactions' that were recovered (recall) and the proportion of times our classifier predicted 'interaction' and was correct (precision). It's not clear which is more important, so we combined the two as the F1 score as the harmonic mean of recall and precision. Why harmonic mean? We want an average of some kind, and the harmonic mean is the smallest of the 3 Pythagorean means (arithmetic mean, geometric mean, and harmonic mean). So, to have a high F1 score you must have high precision and recall, either one will drag the F1 score down non-linearly.

Fig 3-1. Inter-rater variability analysis between human scorers of varying experience levels. In short, 20 behavioral videos (counterbalanced for IST/DST and objects/odors) were scored for rat-stimulus interaction by three independent scorers of differing experience levels (master, experienced, beginner). We found a strong correlation between scorers of all experience levels, but a comparatively weaker correlation between experienced and beginner scorers.

Fig 3-2. Proportion of excluded videos from verification ranks 4 and 5 as described in Fig 3C,D. The proportion of videos excluded did not differ significantly when grouped by treatment (**A**) or stimuli type (**B**).

References

Adam, K. C. S., Doss, M. K., Pabon, E., Vogel, E. K., & de Wit, H. (2020). Δ9-Tetrahydrocannabinol (THC) impairs visual working memory performance: a randomized crossover trial. *Neuropsychopharmacology*, *45*(11), 1807–1816.

Alvarez, P. & Eichenbaum, H. (2002). Representations of odors in the rat orbitofrontal cortex change during and after learning. *Behavioural Neuroscience*, 116, 421–433.

Anderson, D. J., & Perona, P. (2014). Toward a Science of Computational Ethology. *Neuron*, *84*(1), 18–31.

Baglot, S. L., Hume, C., Petrie, G. N., Aukema, R. J., Lightfoot, S. H. M., Grace, L. M., Zhou, R., Parker, L., Rho, J. M., Borgland, S. L., McLaughlin, R. J., Brechenmacher, L., & Hill, M. N. (2021). Pharmacokinetics and central accumulation of delta-9-tetrahydrocannabinol (THC) and its bioactive metabolites are influenced by route of administration and sex in rats. *Scientific Reports*, *11*(1), 23990.

Barch, D. M. & Smith, E. (2008). The cognitive neuroscience of working memory: relevance to CNTRICS and schizophrenia. *Biological Psychiatry*, 64, 11–17.

Barker, G. R. I., Bird, F., Alexander, V. & Warburton, E. C. (2007). Recognition memory for objects, place, and temporal order: a disconnection analysis of the role of the medial prefrontal cortex and perirhinal cortex. *Journal of Neuroscience*, 27, 2948–2957.

Barnard, I. L., Onofrychuk, T. J., Sandini, T. M., McElroy, D. L., Zagzoog, A., Roebuck, A. J., Austin-Scott, F. V, Laprairie, R. B., & Howland, J. G. (2022). The effects of acute Cannabis smoke or Δ9-THC injections on the trial-unique, nonmatching-to-location and five-choice serial reaction time tasks in male Long-Evans rats. *Neurobiology of Learning and Memory*, *192*, 107624.

Bevins, R. A. & Besheer, J. (2006). Object recognition in rats and mice: a one-trial non-matching-to-sample learning task to study 'recognition memory'. *Nature Protocols*, 1, 1306–1311.

Blaes, S. L., Orsini, C. A., Holik, H. M., Stubbs, T. D., Ferguson, S. N., Heshmati, S. C., Bruner, M. M., Wall, S. C., Febo, M., Bruijnzeel, A. W., Bizon, J. L., & Setlow, B. (2019). Enhancing effects of acute exposure to cannabis smoke on working memory performance. *Neurobiology of Learning and Memory*, *157*, 151–162.

Bossong, M. G., Jansma, J. M., van Hell, H. H., Jager, G., Oudman, E., Saliasi, E., Kahn, R. S., & Ramsey, N. F. (2012). Effects of Δ9-Tetrahydrocannabinol on Human Working Memory Function. *Biological Psychiatry*, *71*(8), 693–699.

Broadbent, N. J., Gaskin, S., Squire, L. R., & Clark, R. E. (2010). Object recognition memory and the rodent hippocampus. *Learning & Memory*, *17*(1), 5–11.

Broadbent, N. J., Squire, L. R., & Clark, R. E. (2004). Spatial memory, recognition memory, and the hippocampus. *Proceedings of the National Academy of Sciences*, *101*(40), 14515–14520.

Bruijnzeel, A. W., Qi, X., Guzhva, L. v, Wall, S., Deng, J. v, Gold, M. S., Febo, M., & Setlow, B. (2016). Behavioral Characterization of the Effects of Cannabis Smoke and Anandamide in Rats. *PLOS ONE*, *11*(4), e0153327-.

Cazakoff, B. N. & Howland, J. G. (2011). AMPA receptor endocytosis in rat perirhinal cortex underlies retrieval of object memory. *Learning & Memory*, 18, 688–692.

Churchwell, J. C. & Kesner, R. P. (2011). Hippocampal-prefrontal dynamics in spatial working memory: interactions and independent parallel processing. *Behavioural Brain Research*, 225, 389–395.

Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, *17*(7), 438–449.

Cousijn, J., Wiers, R. W., Ridderinkhof, K. R., van den Brink, W., Veltman, D. J., & Goudriaan, A. E. (2014). Effect of baseline cannabis use and working-memory network function on changes in cannabis use in heavy cannabis users: A prospective fMRI study. *Human Brain Mapping*, *35*(5), 2470–2482.

Cowan, N. (2010). The Magical Mystery Four. *Current Directions in Psychological Science*, *19*(1), 51–57.

Crane, N. A., Schuster, R. M., Fusar-Poli, P., & Gonzalez, R. (2013). Effects of Cannabis on Neurocognitive Functioning: Recent Advances, Neurodevelopmental Influences, and Sex Differences. *Neuropsychology Review*, *23*(2), 117–137.

Creighton, S. D., Palmer, D., Mitchnick, K. A. & Winters, B. D. Chapter 6 - Exploiting Novelty and Oddity Exploratory Preferences in Rodents to Study Multisensory Object Memory and Perception. in *Handbook of Behavioral Neuroscience* (eds. Ennaceur, A. & de Souza Silva, M. A.) vol. 27 103–123 (Elsevier, 2018).

Cui, Q., Pamukcu, A., Cherian, S., Chang, I. Y. M., Berceau, B. L., Xenias, H. S., Higgs, M. H., Rajamanickam, S., Chen, Y., Du, X., Zhang, Y., McMorrow, H., Abecassis, Z. A., Boca, S. M., Justice, N. J., Wilson, C. J., & Chan, C. S. (2021). Dissociable Roles of Pallidal Neuron Subtypes in Regulating Motor Patterns. *The Journal of Neuroscience*, *41*(18), 4036.

Curran, V. H., Brignell, C., Fletcher, S., Middleton, P., & Henry, J. (2002). Cognitive and subjective dose-response effects of acute oral Δ9-tetrahydrocannabinol (THC) in infrequent cannabis users. *Psychopharmacology*, *164*(1), 61–70.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

de Melo, L. C. S., Cruz, A. P., Valentim, S. J. R., Marinho, A. R., Mendonça, J. B., & Nakamura-Palacios, E. M. (2005). Δ9-THC administered into the medial prefrontal cortex disrupts the spatial working memory. *Psychopharmacology*, *183*(1), 54–64.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, S., & Grossman, M. (1995). The neural basis of the central executive system of working memory. *Nature*, *378*(6554), 279–281.

D'Souza, D. C., Fridberg, D. J., Skosnik, P. D., Williams, A., Roach, B., Singh, N., Carbuto, M., Elander, J., Schnakenberg, A., Pittman, B., Sewell, R. A., Ranganathan, M., & Mathalon, D. (2012). Dose-Related Modulation of Event-Related Potentials to Novel and Target Stimuli by Intravenous Δ9-THC in Humans. *Neuropsychopharmacology*, *37*(7), 1632–1646.

Dere, E., Huston, J. P. & De Souza Silva, M. A. (2007). The pharmacology, neuroanatomy and neurogenetics of one-trial object recognition in rodents. *Neuroscience Biobehavioural Reviews*, 31, 673–704.

Dudchenko, P. A. (2004). An overview of the tasks used to test working memory in rodents. *Neuroscience & Biobehavioral Reviews*, *28*(7), 699–709.

Dudchenko, P. A., Wood, E. R. & Eichenbaum, H. (2000). Neurotoxic hippocampal lesions have no effect on odor span and little effect on odor recognition memory but produce significant impairments on spatial span, recognition, and alternation. *Journal of Neuroscience, 20*, 2964–2977.

Dudchenko, P. A., Talpos, J., Young, J., & Baxter, M. G. (2013). Animal models of working memory: A review of tasks that might be used in screening drug treatments for the memory impairments found in schizophrenia. *Neuroscience & Biobehavioral Reviews*, *37*(9, Part B), 2111–2124.

Ennaceur, A. (2010).One-trial object recognition in rats and mice: methodological and theoretical issues. *Behavioural Brain Research*, 215, 244–254.

Ennaceur, A., & Aggleton, J. P. (1994). Spontaneous recognition of object configurations in rats: effects of fornix lesions. *Experimental Brain Research*, *100*(1), 85–92.

Ennaceur, A., & Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behavioural Brain Research*, *31*(1), 47–59.

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive Architecture of Working Memory. *Neuron*, *88*(1), 33–46.

Fernandez, G., & Tendolkar, I. (2006). The rhinal cortex: 'gatekeeper' of the declarative memory system. *Trends in Cognitive Sciences*, *10*(8), 358–362.

Galizio, M. (2016). Olfactory stimulus control and the behavioral pharmacology of remembering. *Behavior Analysis: Research and Practice*, *16*(4), 169–178.

Gold, J. M. et al. (2019). Working Memory Impairment Across Psychotic disorders. *Schizophrenia Bulletin*, 45, 804–812.

Goldman-Rakic, P. S. (1999). The physiological approach: functional architecture of working memory and disordered cognition in schizophrenia. *Biological Psychiatry*, *46*(5), 650–661.

Goodwin, N. L., Nilsson, S. R. O., Choong, J. J., & Golden, S. A. (2022). Toward the explainability, transparency, and universality of machine learning for behavioral classification in neuroscience. *Current Opinion in Neurobiology*, *73*, 102544.

Goonawardena, A. v., Robinson, L., Hampson, R. E., & Riedel, G. (2010). Cannabinoid and cholinergic systems interact during performance of a short-term memory task in the rat. *Learning and Memory*, *17*(10), 502–511.

Grotenhermen, F. (2003). Pharmacokinetics and Pharmacodynamics of Cannabinoids. *Clinical Pharmacokinetics*, *42*(4), 327–360.

Hložek, T., Uttl, L., Kadeřábek, L., Balíková, M., Lhotková, E., Horsley, R. R., Nováková, P., Šíchová, K., Štefková, K., Tylš, F., Kuchař, M., & Páleníček, T. (2017). Pharmacokinetic and behavioural profile of THC, CBD, and THC+CBD combination after pulmonary, oral, and subcutaneous administration in rats and confirmation of conversion in vivo of CBD to THC. *European Neuropsychopharmacology*, *27*(12), 1223–1237.

Huestis, M. A. (2007). Human Cannabinoid Pharmacokinetics. *Chemistry & Biodiversity*, *4*(8), 1770–1804.

Huestis, M. A., Henningfield, J. E., & Cone, E. J. (1992). Blood Cannabinoids. I. Absorption of THC and Formation of 11-OH-THC and THCCOOH During and After Smoking Marijuana*. *Journal of Analytical Toxicology*, *16*(5), 276–282.

Ilan, A. B., Smith, M. E., & Gevins, A. (2004). Effects of marijuana on neurophysiological signals of working and episodic memory. *Psychopharmacology*, *176*(2), 214–222.

Jiang, L.-X., Huang, G.-D., Wang, H.-L., Zhang, C. & Yu, X. (2022). The protocol for assessing olfactory working memory capacity in mice. *Brain and Behaviour*, 12, e2703.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358.

Klausner, H. A., & Dingell, J. v. (1971). The metabolism and excretion of Δ9-tetrahydrocannabinol in the rat. *Life Sciences*, *10*(1, Part 1), 49–59.

Ligresti, A., De Petrocellis, L., & Di Marzo, V. (2016). From Phytocannabinoids to Cannabinoid Receptors and Endocannabinoids: Pleiotropic Physiological and Pathological Roles Through Complex Pharmacology. *Physiological Reviews*, *96*(4), 1593–1659.

Lueptow, L. M. (2017). Novel Object Recognition Test for the Investigation of Learning and Memory in Mice. *Journal of Visualized Experiments*, 126, 55718.

Luxem, K., Sun, J. J., Bradley, S. P., Krishnan, K., Yttri, E. A., Zimmermann, J., Pereira, T. D., & Laubach, M. (2022). Open-Source Tools for Behavioral Video Analysis: Setup, Methods, and Development. *ELife,* 12, e79305.

Marshall, J. D., Aldarondo, D. E., Dunn, T. W., Wang, W. L., Berman, G. J., & Ölveczky, B. P. (2021). Continuous Whole-Body 3D Kinematic Recordings across the Rodent Behavioral Repertoire. *Neuron*, *109*(3), 420-437.e8.

Marshall, J. D., Li, T., Wu, J. H., & Dunn, T. W. (2022). Leaving flatland: Advances in 3D behavioral measurement. *Current Opinion in Neurobiology*, *73*, 102522.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289.

Mathis, A., Schneider, S., Lauer, J., & Mathis, M. W. (2020). A Primer on Motion Capture with Deep Learning: Principles, Pitfalls, and Perspectives. *Neuron*, *108*(1), 44–65.

Minkowicz, S., Mathews, M. A., Mou, F. H., Yoon, H., Freda, S. N., Cui, E. S., Kennedy, A., & Kozorovitskiy, Y. (2023). Striatal ensemble activity in an innate naturalistic behavior. *BioRxiv*, 2023.02.23.529669.

Moore, C. F., Stiltner, J. W., Davis, C. M., & Weerts, E. M. (2022). Translational models of cannabinoid vapor exposure in laboratory animals. *Behavioural Pharmacology*, *33*(2 & 3), 63–89.

Mouly, A.-M., & Sullivan, R. (2010). Memory and Plasticity in the Olfactory System: From Infancy to Adulthood. *The Neurobiology of Olfaction.* CRC Press/ Taylor & Francis.

Nath, T., Mathis, A., Chen, A.C., Patel, A., Bethge, M., Mathis, M.W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, 14, 2152-2176.

Newmeyer, M. N., Swortwood, M. J., Barnes, A. J., Abulseoud, O. A., Scheidweiler, K. B., & Huestis, M. A. (2016). Free and Glucuronide Whole Blood Cannabinoids' Pharmacokinetics after Controlled Smoked, Vaporized, and Oral Cannabis Administration in Frequent and Occasional Cannabis Users: Identification of Recent Cannabis Intake. *Clinical Chemistry*, 62, 1579–1592.

Newton, K. C., Kacev, D., Nilsson, S. R. O., Saettele, A. L., Golden, S. A., & Sheets, L. (2023). Lateral line ablation by ototoxic compounds results in distinct rheotaxis profiles in larval zebrafish. *Communications Biology*, *6*(1), 84.

Nguyen, J. D., Aarde, S. M., Vandewater, S. A., Grant, Y., Stouffer, D. G., Parsons, L. H., Cole, M., & Taffe, M. A. (2016). Inhaled delivery of Δ9-tetrahydrocannabinol (THC) to rats by e-cigarette vapor technology. *Neuropharmacology*, *109*, 112–120.

Nilsson, S. R. O., Goodwin, N. L., Choong, J. J., Hwang, S., Wright, H. R., Norville, Z. C., Tong, X., Lin, D., Bentzley, B. S., Eshel, N., McLaughlin, R. J., & Golden, S. A. (2020). Simple Behavioral Analysis (SimBA) – an open source toolkit for computer classification of complex social behaviors in experimental animals. *BioRxiv*, 2020.04.19.049452.

Olivito, L., De Risi, M., Russo, F., & De Leonibus, E. (2019). Effects of pharmacological inhibition of dopamine receptors on memory load capacity. *Behavioural Brain Research*, *359*, 197–205.

Olivito, L., Saccone, P., Perri, V., Bachman, J. L., Fragapane, P., Mele, A., Huganir, R. L., & De Leonibus, E. (2016). Phosphorylation of the AMPA receptor GluA1 subunit regulates memory load capacity. *Brain Structure and Function*, *221*(1), 591–603.

Oomen, C. A., Hvoslef-Eide, M., Heath, C. J., Mar, A. C., Horner, A. E., Bussey, T. J., & Saksida, L. M. (2013). The touchscreen operant platform for testing working memory and pattern separation in rats and mice. *Nature Protocols*, *8*(10), 2006–2021.

Owens, M. M., McNally, S., Petker, T., Amlung, M. T., Balodis, I. M., Sweet, L. H., & MacKillop, J. (2019). Urinary tetrahydrocannabinol is associated with poorer working memory performance and alterations in associated brain activity. *Neuropsychopharmacology*, *44*(3), 613–619.

Peters, G. J., David, C. N., Marcus, M. D. & Smith, D. M. (2013). The medial prefrontal cortex is critical for memory retrieval and resolving interference. *Learning & Memory*, 20, 201–209.

Piskulic, D., Olver, J. S., Norman, T. R., & Maruff, P. (2007). Behavioural studies of spatial working memory dysfunction in schizophrenia: A quantitative literature review. *Psychiatry Research*, *150*(2), 111–121.

Ramaekers, J., Kauert, G., Theunissen, E., Toennes, S., & Moeller, M. (2009). Neurocognitive performance during acute THC intoxication in heavy and occasional cannabis users. *Journal of Psychopharmacology*, *23*(3), 266–277.

Ramus, S. J. & Eichenbaum, H. (2000). Neural correlates of olfactory recognition memory in the rat orbitofrontal cortex. *Journal of Neuroscience*, 20, 8199–8208.

Ravula, A., Chandasana, H., Jagnarine, D., Wall, S. C., Setlow, B., Febo, M., Bruijnzeel, A. W., & Derendorf, H. (2019). Pharmacokinetic and Pharmacodynamic Characterization of Tetrahydrocannabinol-Induced Cannabinoid Dependence After Chronic Passive Cannabis Smoke Exposure in Rats. *Cannabis and Cannabinoid Research*, *4*(4), 240–254.

Roebuck, A. J., Greba, Q., Onofrychuk, T. J., McElroy, D. L., Sandini, T. M., Zagzoog, A., Simone, J., Cain, S. M., Snutch, T. P., Laprairie, R. B., & Howland, J. G. (2022). Dissociable changes in spike and wave discharges following exposure to injected cannabinoids and smoked cannabis in Genetic Absence Epilepsy Rats from Strasbourg. *European Journal of Neuroscience*, *55*(4), 1063–1078.

Sandini, T. M. et al. (2020). NMDA Receptors in Visual and Olfactory Sensory Integration in Male Long Evans Rats: A Role for the Orbitofrontal Cortex. *Neuroscience*, 440, 230–238.

Sannino, S., Russo, F., Torromino, G., Pendolino, V., Calabresi, P., & de Leonibus, E. (2012). Role of the dorsal hippocampus in object memory load. *Learning & Memory*, *19*(5), 211–218.

Scott, G. A., Liu, M. C., Tahir, N. B., Zabder, N. K., Song, Y., Greba, Q., & Howland, J. G. (2020). Roles of the medial prefrontal cortex, mediodorsal thalamus, and their combined circuit for performance of the odor span task in rats: analysis of memory capacity and foraging behavior. *Learning & Memory*, *27*(2), 67–77.

Serby, M. J. & Chobor, K. L. *Science of Olfaction.* (Springer Science & Business Media, 2012).

Shrager, Y., Levy, D. A., Hopkins, R. O. & Squire, L. R. (2008). Working memory and the organization of brain systems. *Journal of Neurosceince,* 28, 4818–4822.

Slivicki, R. A., Earnest, T., Chang, Y., Pareta, R., Casey, E., Li, J., Tooley, J., Abiraman, K., Vachez, Y. M., Wolf, D. K., Sackey, J. T., Kumar Pitchai, D., Moore, T., Gereau, R. W., Copits, B. A., Kravitz, A. V., & Creed, M. C. (2023). Oral oxycodone self-administration leads to features of opioid misuse in male and female mice. *Addiction Biology*, *28*(1).

Sugita, M., Yamada, K., Iguchi, N. & Ichitani, Y. (2015). Hippocampal NMDA receptors are involved in rats' spontaneous object recognition only under high memory load condition. *Brain Research*, 1624, 370–379.

Torromino, G. Loffredo, V., Cavezza, D., Sonsini, G., Esposito, F., Crevenna, A. H., Gioffrè, M., De Risi, M., Treves, A., Griguoli, M., De Leonibus, E. (2022).Thalamo-hippocampal pathway regulates incidental memory capacity in mice. *Nature Communications*, 13, 4194.

van Vugt, M. K., Schulze-Bonhage, A., Litt, B., Brandt, A. & Kahana, M. J. (2010). Hippocampal gamma oscillations increase with memory load. *Journal of Neuroscience,* 30, 2694–2699.

Varvel, S., Hamm, R., Martin, B., & Lichtman, A. (2001). Differential effects of Δ9-THC on spatial reference and working memory in mice. *Psychopharmacology*, *157*(2), 142–150.

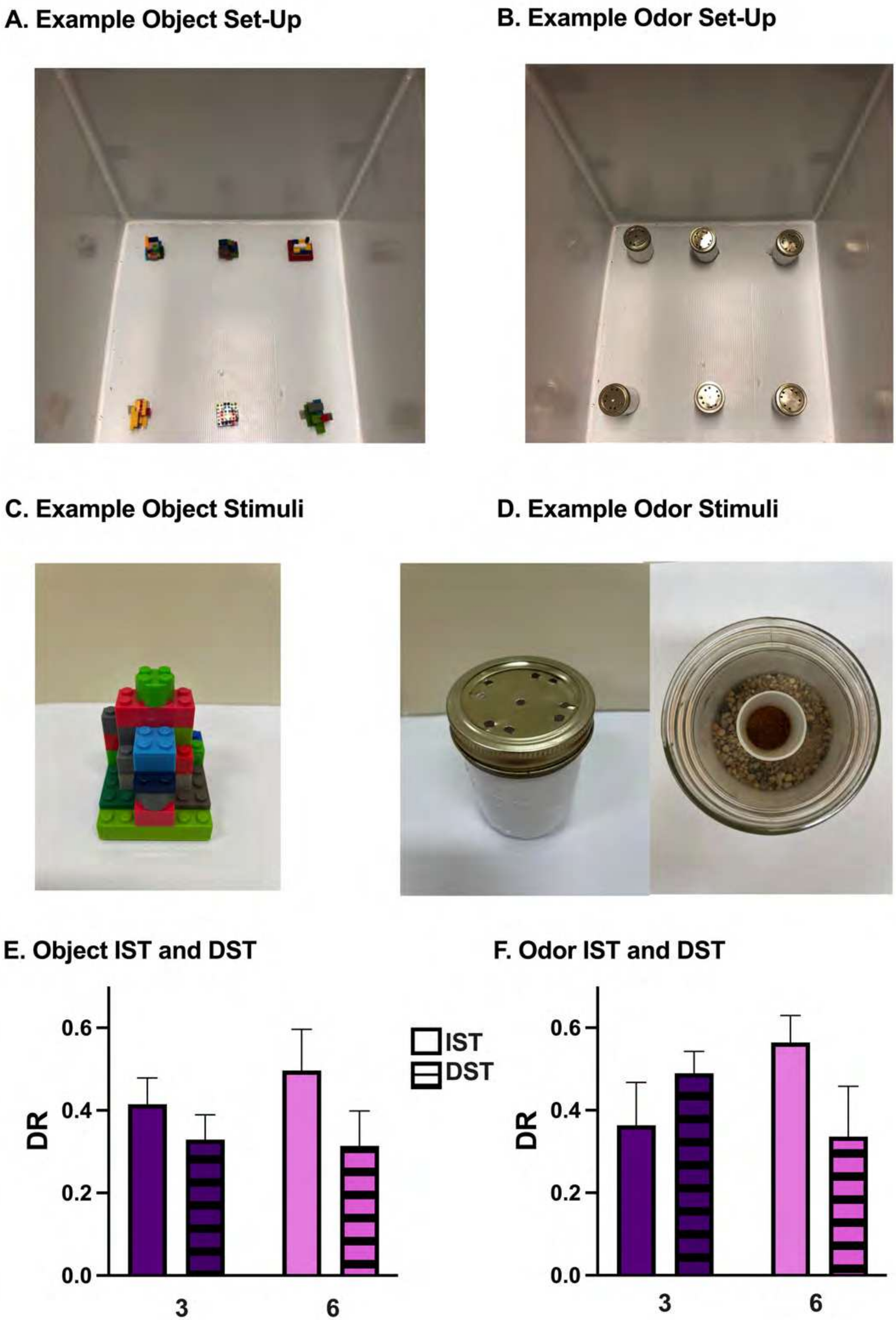
Vorhees, C. V., & Williams, M. T. (2014). Assessing Spatial Learning and Memory in Rodents. *ILAR Journal*, *55*(2), 310–332.

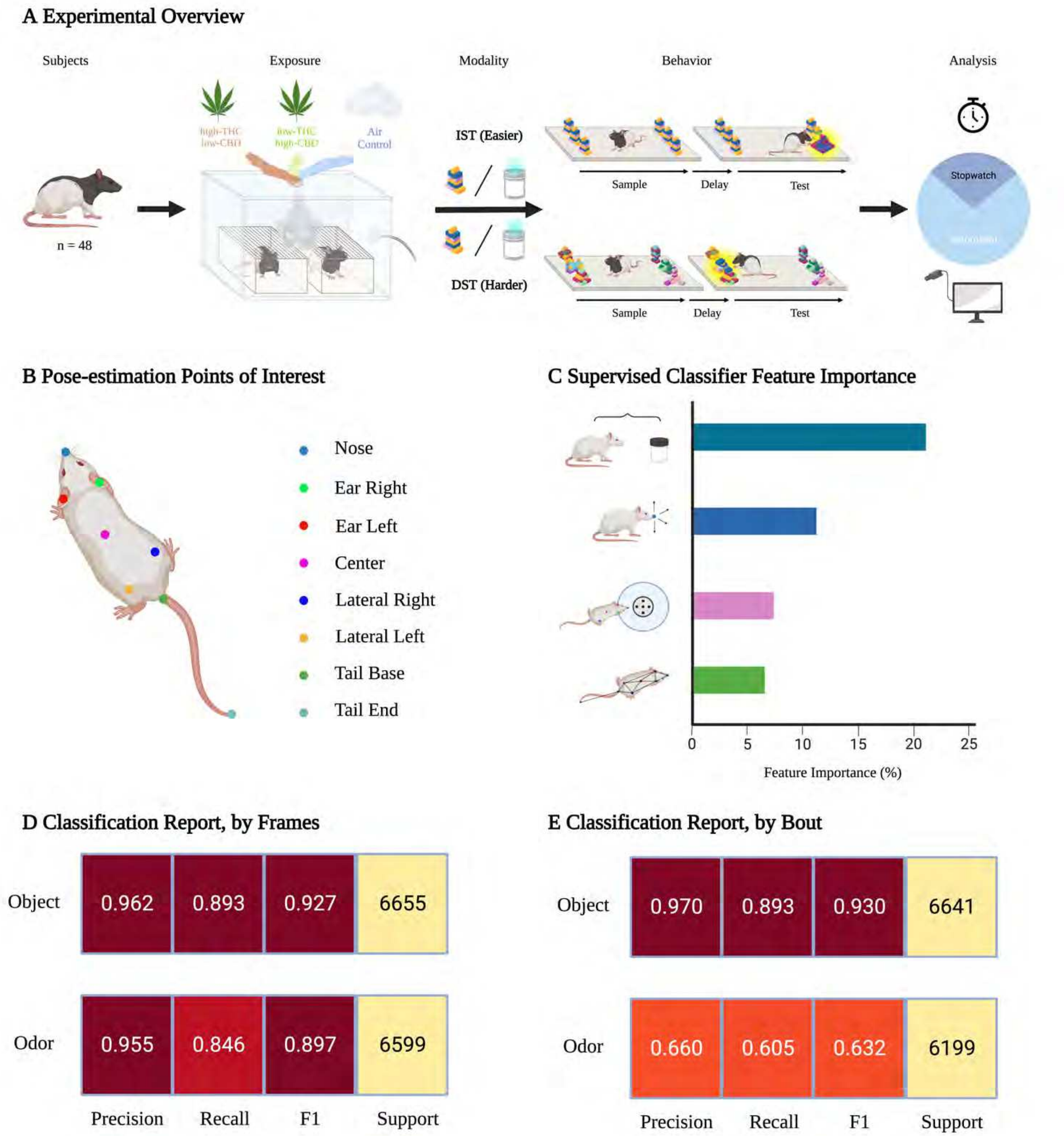
Wiley, J. L., Taylor, S. I., & Marusich, J. A. (2021). Δ9-Tetrahydrocannabinol discrimination: Effects of route of administration in rats. *Drug and Alcohol Dependence*, *225*, 108827.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*.

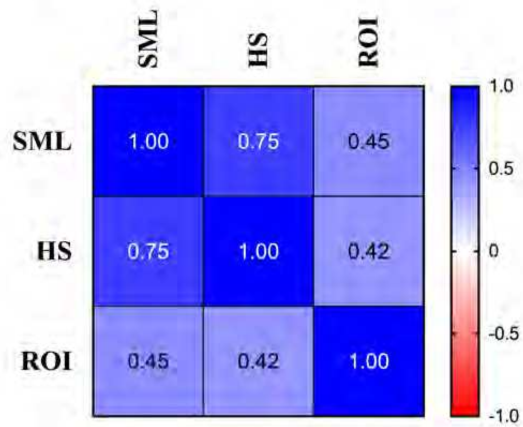
Winters, B. D., Forwood, S. E., Cowell, R. A., Saksida, L. M. & Bussey, T. J. (2004). Double dissociation between the effects of peri-postrhinal cortex and hippocampal lesions on tests of object recognition and spatial memory: heterogeneity of function within the temporal lobe. *Journal of Neurosience*, 24, 5901–5908.

Winters, C., Gorssen, W., Ossorio-Salazar, V. A., Nilsson, S., Golden, S., & D'Hooge, R. (2022). Automated procedure to assess pup retrieval in laboratory mice. *Scientific Reports*, *12*(1), 1663.

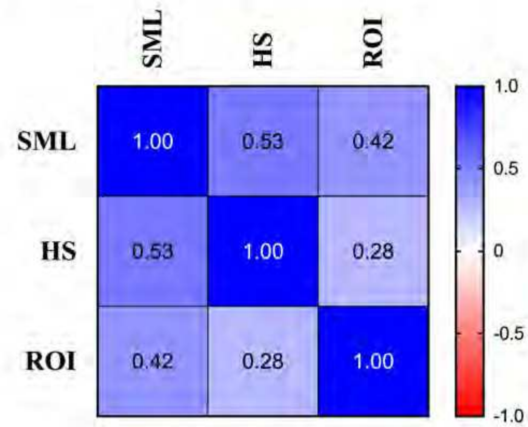

A. Example Object Set-Up
B. Example Odor Set-Up
C. Example Object Stimuli
D. Example Odor Stimuli
E. Object IST and DST
F. Odor IST and DST
DR
0.0
0.2
0.4
0.6
3
6
IST
DST


A Experimental Overview
Subjects
Exposure
Modality
Behavior
Analysis
n = 48
high-THC low-CBD
low-THC high-CBD
Air Control
IST (Easier)
DST (Harder)
Sample
Delay
Test
Stopwatch
B Pose-estimation Points of Interest
Nose
Ear Right
Ear Left
Center
Lateral Right
Lateral Left
Tail Base
Tail End
C Supervised Classifier Feature Importance
0
5
10
15
20
25
Feature Importance (%)
D Classification Report, by Frames
Object
0.962
0.893
0.927
6655
Odor
0.955
0.846
0.897
6599
Precision
Recall
F1
Support
E Classification Report, by Bout
Object
0.970
0.893
0.930
6641
Odor
0.660
0.605
0.632
6199
Precision
Recall
F1
Support

A. Object Correlation by Method



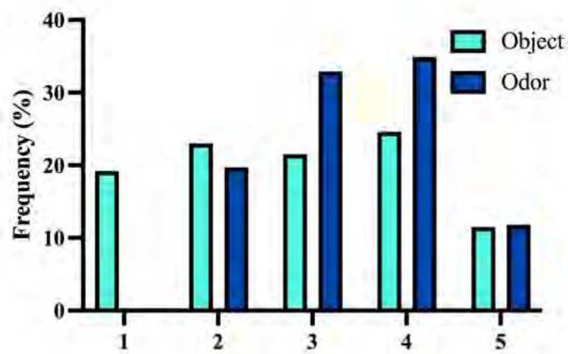
B. Odor Correlation by Method



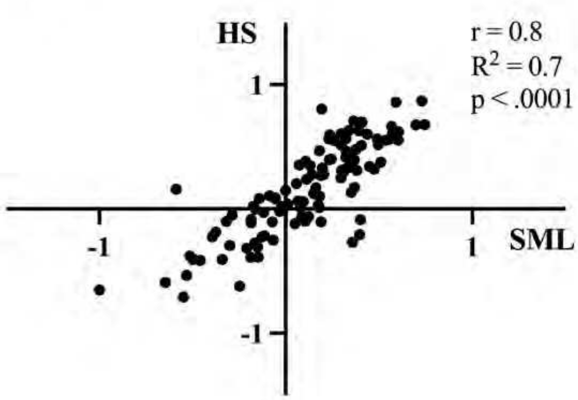
C. Verification Rank Criteria

Verification Rank	Criteria
1	Human level classification
2	Few mistakes < 1 sec impacting all items equally
3	Few mistakes < 1 sec impacting all items unequally
4	Mistakes < 5 secs present across all items
5	Mistakes 5+ secs present across all items

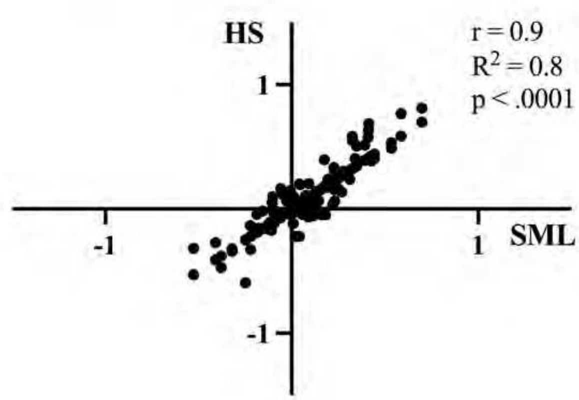
D. Frequency of Verification Ranks

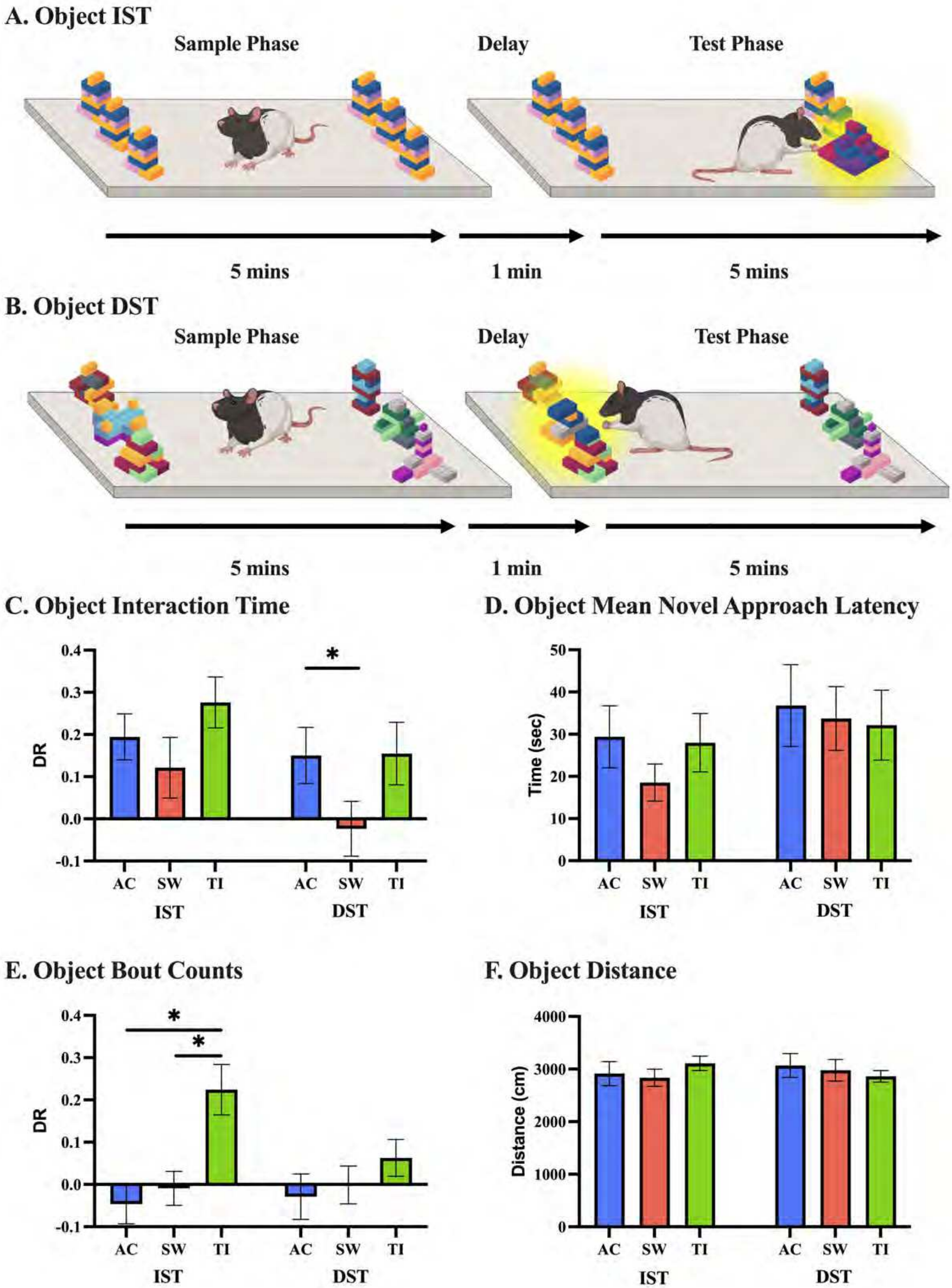


E. Object Correlation

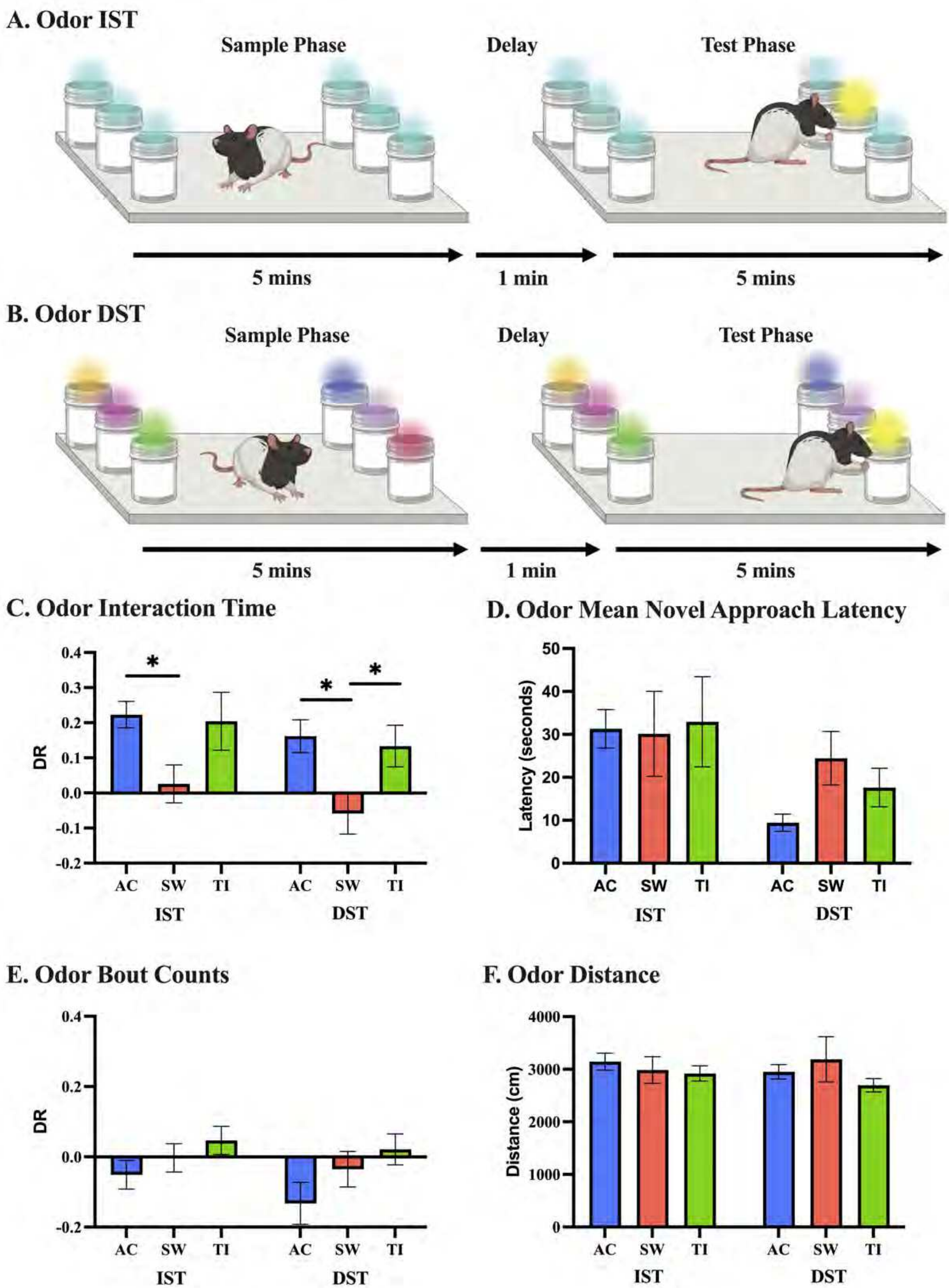


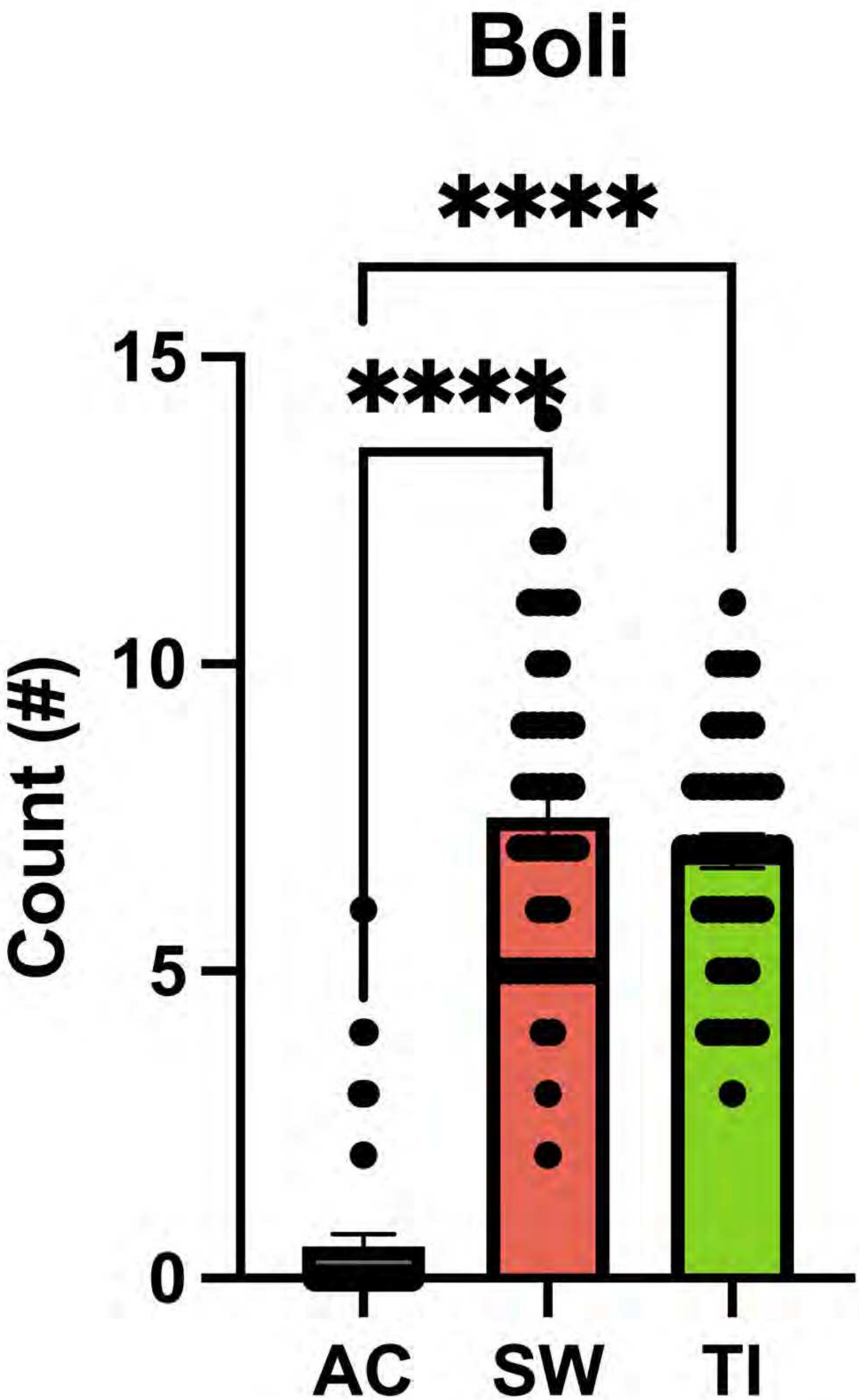
F. Odor Correlation





A. Object IST
Sample Phase
Delay
Test Phase
5 mins
1 min
5 mins
B. Object DST
Sample Phase
Delay
Test Phase
5 mins
1 min
5 mins
C. Object Interaction Time
DR
0.4
0.3
0.2
0.1
0.0
-0.1
*
AC
SW
TI
AC
SW
TI
IST
DST
D. Object Mean Novel Approach Latency
Time (sec)
50
40
30
20
10
0
AC
SW
TI
AC
SW
TI
IST
DST
E. Object Bout Counts
DR
0.4
0.3
0.2
0.1
0.0
-0.1
*
*
AC
SW
TI
AC
SW
TI
IST
DST
F. Object Distance
Distance (cm)
4000
3000
2000
1000
0
AC
SW
TI
AC
SW
TI
IST
DST


A. Odor IST
Sample Phase
Delay
Test Phase
5 mins
1 min
5 mins
B. Odor DST
Sample Phase
Delay
Test Phase
5 mins
1 min
5 mins
C. Odor Interaction Time
DR
AC SW TI AC SW TI
IST DST
D. Odor Mean Novel Approach Latency
Latency (seconds)
AC SW TI AC SW TI
IST DST
E. Odor Bout Counts
DR
AC SW TI AC SW TI
IST DST
F. Odor Distance
Distance (cm)
AC SW TI AC SW TI
IST DST


Boli
****
****
Count (#)
15
10
5
0
AC
SW
TI