

Erratum

Aversive conditioning in honey bees (*Apis mellifera anatolica*): a comparison of drones and workers

Christopher W. Dinges, Arian Avalos, Charles I. Abramson, David Philip Arthur Craig, Zoe M. Austin, Christopher A. Varnon, Fatima Nur Dal, Tugrul Giray and Harrington Wells

10.1242/jeb.098947

There was an error published in *J. Exp. Biol.* **216**, pp. 4124-4134.

The affiliation for Tugrul Giray was incorrect. The correct affiliation is given below:

Department of Biology, University of Puerto Rico, San Juan, PR 00931, Puerto Rico

We apologise to authors and readers for any inconvenience this error may have caused.

RESEARCH ARTICLE

Aversive conditioning in honey bees (*Apis mellifera anatolica*): a comparison of drones and workers

Christopher W. Dinges¹, Arian Avalos², Charles I. Abramson^{1,*}, David Philip Arthur Craig¹, Zoe M. Austin¹, Christopher A. Varnon¹, Fatima Nur Dal³, Tugrul Giray³ and Harrington Wells⁴

¹Department of Psychology, Oklahoma State University, Stillwater, OK 74078, USA, ²Department of Biology, University of Puerto Rico, San Juan, PR 00931, Puerto Rico, ³Beekeeping Research Centre, MKP MYO, Uludag University, Gorukle-Bursa 16059, Turkey and ⁴Department of Biology, University of Tulsa, Tulsa, OK 74104, USA

*Author for correspondence (charles.abramson@okstate.edu)

SUMMARY

Honey bees provide a model system to elucidate the relationship between sociality and complex behaviors within the same species, as females (workers) are highly social and males (drones) are more solitary. We report on aversive learning studies in drone and worker honey bees (*Apis mellifera anatolica*) in escape, punishment and discriminative punishment situations. In all three experiments, a newly developed electric shock avoidance assay was used. The comparisons of expected and observed responses were performed with conventional statistical methods and a systematic randomization modeling approach called object oriented modeling. The escape experiment consisted of two measurements recorded in a master-yoked paradigm: frequency of response and latency to respond following administration of shock. Master individuals could terminate an unavoidable shock triggered by a decrementing 30s timer by crossing the shuttlebox centerline following shock activation. Across all groups, there was large individual response variation. When assessing group response frequency and latency, master subjects performed better than yoked subjects for both workers and drones. In the punishment experiment, individuals were shocked upon entering the shock portion of a bilaterally wired shuttlebox. The shock portion was spatially static and unsignalled. Only workers effectively avoided the shock. The discriminative punishment experiment repeated the punishment experiment but included a counterbalanced blue and yellow background signal and the side of shock was manipulated. Drones correctly responded less than workers when shock was paired with blue. However, when shock was paired with yellow there was no observable difference between drones and workers.

Key words: honey bees, drones, workers, aversive conditioning.

Received 22 April 2013; Accepted 22 July 2013

INTRODUCTION

The present study uses a newly developed protocol to examine aversive learning in honey bee drones and workers. Honey bees provide an excellent model system to elucidate the relationship between brain architecture and social complex behaviors within the same species: workers (females) are highly social and males are more solitary. Patrilineal effects have been correlated with learning ability, aggression and other variables, making the comparative examination of these behaviors significant for further evolutionary and molecular analysis (Bhagavan et al., 1994; Ferguson et al., 2001; Guzman-Novoa et al., 2005).

Previous evidence offers a positive correlation between socio-behavioral roles and changes in brain architecture in worker bees (Fahrbach et al., 1995; Grozinger et al., 2007; Fahrbach et al., 1998; Withers et al., 1993; Farris et al., 2001). Neural architecture has been shown to be influenced by age, task and experience. Principally socially mediated plasticity in the mushroom body (MB) has been observed in the age-induced transition from inside tasks, such as nursing the offspring, to more complex external foraging tasks (Fahrbach et al., 1998; Smith et al., 2010; Farris et al., 2001; Fahrbach et al., 2003). Foraging demands new types of learning, such as identifying and memorizing the colony's location, spatial

orientation using environmental cues for navigation and social communication such as the dance language to recruit foragers (Seeley, 1995).

In males, similar changes in the MB have been documented as they begin orientation and mating flights (Fahrbach et al., 1997). The described expansion also coincides with greater expression of dopamine receptors in the mushroom body and an increase in juvenile hormone titers that mirrors those described in workers (Giray and Robinson, 1996; Humphries et al., 2003). Dopamine levels in honey bees have been shown to influence associative (punishment) learning (Vergoz et al., 2007; Agarwal et al., 2011). Juvenile hormone in both males and females has been shown to work as a regulator of behavioral development (Robinson et al., 1989; Fahrbach et al., 1995; Giray and Robinson, 1996; Giray et al., 1999; Giray et al., 2005). These results suggest that the transition to more complex tasks that require learning coincides with broad expansions of neural tissue in the MB.

Similarities and differences in neuroanatomy prompt further study, for an analysis of learning in honey bees would be incomplete if restricted to only the worker honey bees [queens (Aquino et al., 2004); for drones, see below]. Few reported studies examine learning in drones; those available used classical conditioning of

the proboscis extension reflex (PER) [first described by Frings (Frings, 1944) and refined over the years (Kuwabara, 1957; Takeda, 1961; Vareschi, 1971; Bitterman et al., 1983; Abramson and Boyd, 2001) to investigate the heritability of conditioning in workers (Bhagavan et al., 1994; Benatar et al., 1995; Chandra et al., 2001; Ferguson et al., 2001)]. These studies used harnessed individuals motivated for food rewards to measure response. Here, we used a newly developed electric shock avoidance assay (Agarwal et al., 2011) to study escape and punishment; in addition, both contingencies have been previously demonstrated using formic acid in a shuttlebox assay (Abramson, 1986). In escape conditioning, an aversive event is presented and is terminated by a response. In addition, the results of escape conditioning are similar to those found with food rewards (see Mackintosh, 1974). Here, an aversive event is presented when the response occurs (Kaczer and Maldonado, 2009), and is ideal for researchers looking for a protocol that produces the opposite effect of escape training.

The study of aversive conditioning is practical and ecologically relevant. Under field conditions, honey bees face numerous challenges related to the escape and avoidance of aversive stimuli, including those related to predators, repellents and pesticides. Free-flying honey bees stopped flying to a target after being punished with a single taste of an essential-oil-based pesticide (Abramson et al., 2006). Negative feedback signaling during forager dances has been observed to reduce visitations to feeders on which peril or competition has been experienced, thus demonstrating not only individual perception of dangers but also a medium for social dissemination (Nieh, 2010). However, the use of aversive stimuli in studies of honey bee learning and memory is extremely rare, and no investigations describe learning in drones. Of 62 honey bee learning citations, none describe experiments using aversive stimuli (Wells, 1973); and in a review of the advantages that honey bees offer students of cognition, none of the 105 citations discuss aversive conditioning (Srinivasan, 2010).

The study of aversive conditioning can provide new paradigms to further advance our understanding of neurobiological and genomic learning mechanisms. Much of what is known about these mechanisms comes from work related to the Pavlovian conditioning of the PER. Issues with PER conditioning, and recent work on the role of biogenic amines in reward and punishment pathways, has stimulated the search for new conditioning paradigms in the area of aversive conditioning (Kaczer and Maldonado, 2009; Abramson et al., 2011; Agarwal et al., 2011; Vergoz et al., 2007; Giray et al., 2007).

MATERIALS AND METHODS

Subjects

Subjects were honey bee foragers (female) and drones (males) (*Apis mellifera anatolica* Maa 1953) located in the apiary of the Beekeeping Development and Research Center of the Uludağ University in Bursa, Turkey. Foragers were randomly collected from a feeder containing clove-scented 1 mol l^{-1} sucrose solution ($5 \mu\text{l l}^{-1}$ clove oil). Subjects were experimentally naive and allowed to feed to satiation before capture. Drones were collected from the hive in the mornings and stored in drone crates fitted with queen-/drone-excluding mesh to allow worker entrance and escape but exclude drones from leaving the cell. These drones were stored in their home hive, where they were fed by workers until data collection was initiated. Drones were tested for maturity upon completion of each session by inducing eversion and ejaculation, and those with mucus and sperm on the aedeagus were considered 'reproductively mature' (Giray and Robinson, 1996). Only mature drones were used for data analysis. Each subject was used for one session and then terminated.

Apparatus

An automated apparatus was used for the aversive conditioning assessments (Fig. 1A). The shuttlebox ($14 \times 1.5 \times 0.8 \text{ cm}$) was modeled from the apparatus employed by Agarwal et al. (Agarwal et al., 2011) and was constructed from black plastic and outfitted with a clear plastic roof and two infrared beam ports positioned 1 cm from the center line on both sides of the shuttlebox (Fig. 1B). A stainless steel 29-pin shock grid ($14 \times 20 \text{ cm}$ useable area) served as the floor of the shuttlebox. This grid was wired bilaterally (14 pins on each side) with a neutral center pin to allow for discrete shock application to one or both sides of the shock grid as needed for individual experimental and trial constraints. Shock was administered *via* a 1.2 A variable voltage Universal AC Adapter (model number: DX-AC1200, Dynex, Lincoln, UK). When set at 9 V DC, actual measurements on the shock grid were 8.71 V at 1.0 A. A clear plastic sheath was placed below the grid to allow easy cleaning and to prevent bees from attempting to escape through the pins of the shock grid. The shuttlebox lid was coated with a thin layer of petroleum jelly to prevent bees from walking upside-down to ensure the bees remained in contact with the shock grid for the length of the trial.

The shuttlebox was designed to house one subject and contained two side-looking infrared photodiode–phototransistor pairs (512-QEE113, 512-QSE113, Fairchild Semiconductor, San Jose, CA, USA) mounted in the infrared beam ports positioned parallel to the centerline of the shuttlebox. The orientation of one photodiode–phototransistor pair was reversed with respect to the other pair in each shuttlebox so that light from the photodiode in one pair would not inadvertently buffer triggering of the other pair's phototransistor. A mutually exclusive activation circuit was constructed such that responses would only be detected if the subject interrupted the infrared beam farthest from them to ensure the subject must cross the center line to trigger a response. A PIC18F2580-I/SO microcontroller and relay were used to construct this circuit. A plastic top-justified hurdle at the center line ensured that subjects would break the infrared beam. Two shuttleboxes were run in tandem on the same shock grid and placed upon a 17 inch Dell 1704FPt flat panel monitor (Round Rock, TX, USA) set at default color settings. This monitor was used as the source of visual stimuli when required. To synchronize other aspects of the apparatus with the visual stimuli presented on the computer monitor, a photoresistor-based light-activated relay (Fk401, Backatronics, Meriden, CT, USA) was positioned on the computer monitor. When stimuli were presented on the monitor, the light-activated relay would activate and send a signal to the other equipment.

An experiment controller, developed by Palya and Walter (Palya and Walter, 1993), was interfaced with the previously described equipment. The experiment controller administered shock in accordance with the experimental design, and was interfaced with the infrared beam circuit to detect responses and with the light-activated relay to synchronize the experiment with the video display. The experiment controller ran custom programs written in ECBASIC (Jacksonville State University, Jacksonville, AL, USA), to define input properties, and organized data and administered stimuli to the specifications of each experiment. After each experimental session, the experiment controller saved labeled time-stamped data files to a computer.

Pre-trial preparation

Experimentally naive honey bees were collected for each trial. Prior to each trial, the plastic sheath and shock grid were cleaned and

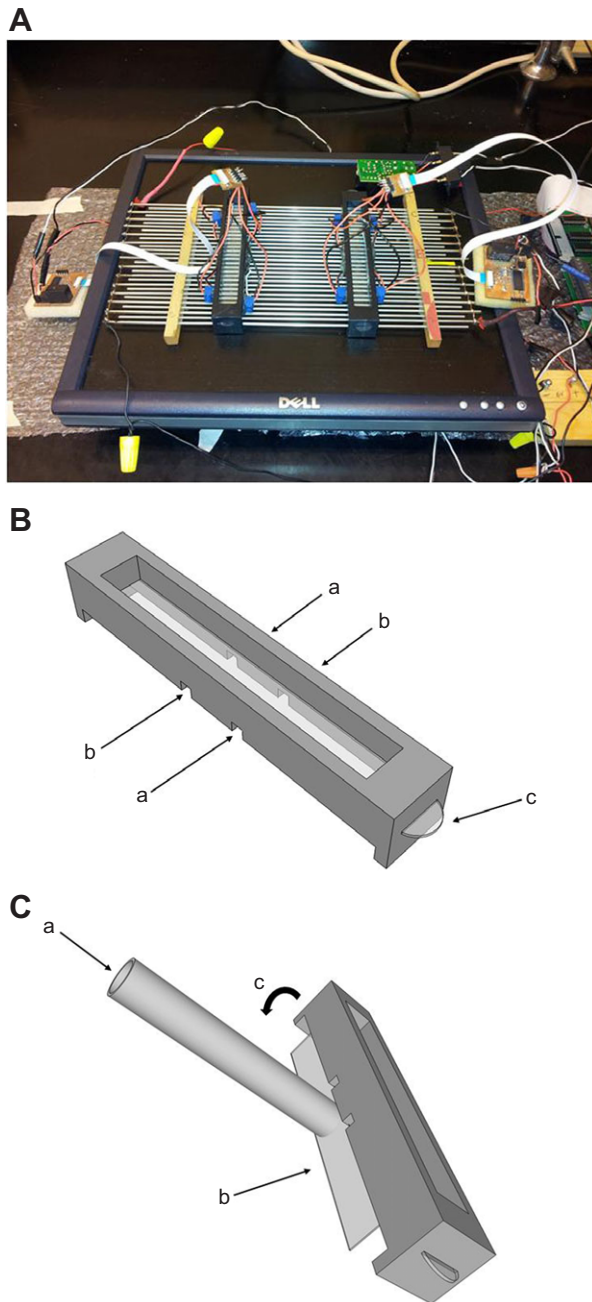


Fig. 1. (A) Shock shuttlebox assay apparatus and computer monitor. The two shuttleboxes, run in tandem, were positioned ~8 cm apart on the shock grid and placed atop the computer monitor. The infrared beams were positioned adjacent to the center line (yellow tape) and mounted *via* a detachable connector port (blue boxes on the sides of each shuttlebox). The mutually exclusive activation relays (copper printed circuit boards) are attached *via* a ribbon cable and input hub. At the top (green printed circuit board) is the light-activated relay fixed to the monitor to communicate the monitor status to the interface (not shown). (B) Computer rendering of the shuttlebox apparatus, showing the photodiode ports (a) and the phototransistor ports (b) placed 1 cm from the center line on both sides. A clear plastic sheath (c) was used to limit the shuttlebox area to the ideal dimensions (14×1.5×0.8 cm). (C) Deposit procedure. After coaxing the bee into the tube of the device and briefly plugging the device to prevent escape, the translocation device was positioned under the shuttlebox as shown. A puff of air to the entrance end (a) was used to push the bee into the shuttlebox. The shield (b) provided some measure of restraint to keep the bee from immediately escaping. The shuttlebox was then quickly lowered into position to trap the bee (c).

placed upon the powered-off computer monitor. Each shuttlebox was counterbalanced for each session (i.e. a master shuttlebox for one trial would be the yoked shuttlebox for next session). To avoid the deleterious effects of anesthetizing foragers and to allow for between-sex comparisons of non-anesthetized drones, an original translocation device was constructed for transporting and depositing forager bees into the apparatus.

The translocation device consisted of a hollow clear plastic cylinder, 9 cm in length, with a 1.2 cm diameter internal tube and a clear plastic strip 12 cm long and 1.5 cm wide to serve as a shielding extension (Fig. 1C). This device takes advantage of the phototropic nature of honey bees by coaxing them into the tube from their holding cell using an LED flashlight. The tube was briefly plugged and positioned beneath the shuttlebox whereby a puff of air was administered opposite the depositing end to push the bee out of the tube and into the shuttlebox. The depositing end of the tube was affixed with a perpendicular plastic strip to prevent the bee from escaping before the shuttlebox could be lowered into position (Fig. 1C).

Ambient light was minimized to reduce interference with the apparatus detection circuitry and to reduce unintended phototactic responses. Both foragers and drones were run throughout the day from 09:00 to 16:00 h to limit the effects of circadian-influenced behavioral variations. Following random selection in groups of three bees from the feeder, two of the three bees were chosen based on their similarity in activity levels. Bees that did not move for more than 50% of the trial time were considered non-responsive and were discarded. Each experiment was run to completion before beginning the next.

Experiment 1: Escape

A sample size of 40 honey bees (20 drones and 20 foragers) was used for the escape experiment. Half of each group was designated as master and half as yoked. Each subject was placed in a shuttlebox atop a stainless steel shock grid positioned on a powered-off computer monitor. Each subject was exposed to a 10 min trial period in which an unavoidable shock was administered by a decrementing 30 s timer. Upon reaching zero, the timer would trigger shock (DC 8.71 V, 1.0 A) on both sides of the shuttlebox, for both master and yoked subject, then restart the timer. Master subjects could terminate this indefinite shock for both the master and yoked subject by crossing to the other side of the shuttlebox. At the onset of shock, the master subject was required to break the infrared beam farthest from them to deactivate the shock. The yoked subject had no control over the shock.

Experiment 2: Punishment

A sample size of 40 honey bees (20 drones and 20 foragers) was chosen for this place preference experiment. Each subject was placed in a shuttlebox atop a stainless steel shock grid positioned on a powered-off computer monitor to maintain environmental consistency with other experiments.

Following a 2 min habituation period, each subject was exposed to a 10 min trial period where shock (DC 8.71 V, 1.0 A) was continuously administered to one half of the shock grid; shock areas were counterbalanced. Spatial positioning of the shock was static and its spatial orientation was the only discriminative cue. When placed in the shuttlebox, a bee would repeatedly shuttle from end to end. A decrease in this baseline shuttling behavior would allow the bee to avoid punishment. The subject's shuttling behavior was recorded *via* interruption of the infrared beams. Each 10 min trial was partitioned into 60 s bins for data analysis.

Experiment 3: Punishment with discrimination

A sample size of 80 honey bees [40 drones (half yellow, half blue) and 40 foragers (half yellow, half blue)] was chosen for this place preference experiment. Each subject was placed in a shuttlebox atop a stainless steel shock grid positioned on a powered-off computer monitor. Each subject was exposed to a 10 min training period, followed by a 2 min testing period. Following a 2 min habituation period, the computer monitor was turned on and displayed two colors. One color was paired with the shock side (DC 8.71 V, 1.0 A), the other color with the no-shock side of the shuttlebox. To remain consistent with Agarwal et al. (Agarwal et al., 2011), the selected colors were Microsoft Paint default swatches: blue (R: 0, G: 0, B: 255, Hue: 160, Sat: 240, Lum: 120) and yellow (R: 255, G: 255, B: 0, Hue: 40, Sat: 240, Lum: 120). These were counterbalanced and displayed on a 17 inch Dell 1704FPt flat panel monitor set at factory default color settings. Each 10 min trial was partitioned into 60 s bins for data analysis. To account for spatial place preference, these colors and paired shock positioning would be transposed once per 60 s bin.

Following the 10 min training period, the computer monitor and shock were turned off for a 20 s loading period. During this time, no responses were recorded. At the end of the loading period, the computer monitor was turned back on to display the same colors as during the training period; however, no shock was administered. Response recording was continued at this point in an extinction phase. During these two bins of data collection, one spatial color switch occurred after 65 s following the initial presentation.

Data analysis

We used observation oriented modelling (OOM) (Grice, 2011; Grice et al., 2012) to analyze our data. OOM is a data analysis technique that allows comparisons of observed results with expected patterns of outcomes for each bee and group, and the evaluation of these differences with an accuracy index and a randomization test. OOM assesses the individual subject observations and does not rely on traditional summaries of data such as measures of central tendency. By using these methods, we were able to eschew the assumptions of null hypothesis significance testing (e.g. homogeneity, normality) as well as avoid construing learning as an abstract population parameter to be estimated from our data. We have successfully used this approach in recent experiments (e.g. Craig et al., 2012).

Within OOM, we used an ordinal analysis that produces a percent correct classification (PCC) value and a chance value (a probability statistic). The PCC value is computed by comparing an *a priori* ordinal prediction with the observed data, and is the ratio of the observed data that matches the expected pattern compared with the number of comparisons that were made. A chance value (*c*-value) ranging from zero to one displays how many randomized versions of the observed data yielded higher PCCs compared with the observed data. A *c*-value of 0.01 indicates a 99% chance that the PCC value is not due to chance based on a range of values obtained from randomized versions of the data. In a two-order comparison, a *c*-value could be considered as conceptually similar to a binomial probability.

However, as *c*-values are calculated from randomizations of the observed data points, each PCC value's likelihood of being due to chance is assessed on an adaptable distribution that is based on observed data rather than a hypothetical distribution (e.g. the standard normal curve). Two procedures of ordinal pattern analysis were conducted to compare between and within groups to thoroughly analyze each dependent variable in each experiment.

To assess learning within subjects and trials, each respective variable was binned, and these bins were compared with every other bin for an individual bee (e.g. bin 1 *versus* bin 2, bin 1 *versus* bin 3...bin 1 *versus* bin 20, etc.); consequently, the number of observations that fit the ordinal pattern can range from zero to combination ${}_nC_2$ (n choose 2). For example, chunking responses into 20 bins would result in 190 bin comparisons. A PCC value of the comparisons matching the expected patterns is computed for each bee, and a randomization *c*-value is obtained by comparing the observed data with 100 randomizations of the observed data.

To assess learning between groups, each respective variable was binned for all subjects, and combinations of each group's individual's bins were compared against another group's individual's bins. We avoided simple mean comparisons *via* this method, for every response bin of every individual in a first group was compared against every response bin of every individual in a second group. Instead of relying on *t*-tests, which simply indicate whether differences in means were observed, we were able to predict the ordinal direction of the pairwise comparison.

RESULTS

Experiment 1

Two dependent variables were assessed to investigate how quickly subjects would escape a shock occurring every 30 s. Because of dimorphism in base response rates, a direct comparison between workers and drones could not be made; thus, a relative comparison was performed. Two possible strategies of shock termination were analyzed: continuous responding (response frequency) and reactive responding (latency to respond). Response frequency and first response latencies from 20 trials were compared between each individual master subject and its yoked counterpart using the between-group pairwise method to assess responding differences between the pair. This assessment was performed under the prediction that master animals would have both higher rates of responding and lower latency to respond following the onset of shock. Tables 1 and 2 display the individual and group master-yoke comparisons for response frequency and master first response latency, respectively. Response rates and latencies to respond varied greatly across individuals. Visual representations of response rates and latency comparisons of master-yoked pairs of aggregated worker and drone groups are displayed in Figs 2 and 3, respectively.

In the ordinal analysis, workers were predicted to have shorter latencies and higher response rates than drones. Workers' response frequency ranged from PCC values of 1% (*c*-value: 1.00) to 98% (*c*-value: 0.01). Workers' latency to respond ranged from PCC values of 14% (*c*-value: 1.00) to 97%, (*c*-value: 0.01). Sixty percent of the worker master subjects had higher response rates than yoked subjects; 70% of the worker master subjects had shorter latencies to respond than yoked subjects. Individual drone response frequency was varied: PCC values ranged from 0% (*c*-value: 1.00) to 83% (*c*-value: 0.01); master response latency PCC ranged from 13% (*c*-value: 1.00) to 80% (*c*-value: 0.01). Sixty percent of the drone master subjects had higher response rates than yoked subjects, and 60% of the drone master subjects had shorter latencies to respond than yoked subjects. When assessing group response frequency and latency, worker master subjects performed better than yoked worker subjects (PCC response: 52%, *c*-value: 0.01; PCC latency: 63%, *c*-value: 0.01), as did drone master subjects (PCC response: 43%, *c*-value: 0.01; PCC latency: 49%, *c*-value: 0.01).

We compared worker and drone master subjects to assess sex differences under the prediction that workers would have higher response rates and lower latencies to respond. Consistent with our

Table 1. Escape: comparison in response rate between individual yoked and master subjects under the prediction that master subjects would have higher response rates than yoked subjects

	PCC (%)	c-value
Workers		
Group	52	0.01
B1	66	0.01
B2	68	0.01
B3	29	1
B4	1	1
B5	26	1
B6	98	0.01
B7	90	0.01
B8	88	0.01
B9	66	0.01
B10	57	0.01
B11	22	1
B12	12	1
B13	80	0.01
B14	50	0.01
B15	12	1
B16	94	0.01
B17	18	1
B18	22	1
B19	48	0.01
B20	95	0.01
Drones		
Group	43	0.01
B1	0	1
B2	65	0.01
B3	62	0.01
B4	57	0.01
B5	47	0.01
B6	18	1
B7	23	1
B8	43	0.05
B9	36	0.81
B10	20	1
B11	35	0.43
B12	79	0.01
B13	17	1
B14	65	0.01
B15	26	0.03
B16	65	0.01
B17	64	0.01
B18	33	0.01
B19	11	1
B20	83	0.01

PCC, percent correct classification.

prediction, workers had both higher response rates (PCC: 76%, *c*-value: 0.01) and lower latencies to respond (PCC: 73%, *c*-value: 0.01) compared with drones. We then compared worker and drone yoked subjects to assess sex differences under the prediction that workers would have higher response rates and lower latencies to respond. Consistent with our prediction, workers had both higher response rates (PCC: 62%, *c*-value: 0.01) and lower latencies to respond (PCC: 60%, *c*-value: 0.01) compared with drones.

To facilitate a comparison in data analysis strategies, we performed a repeated-measures ANOVA with an *a priori* alpha level of 0.05 to assess differences in latency scores. Mauchly's test of sphericity was significant ($W=0.007$, $\chi^2_{189}=349.759$, $P<0.000$). A within-subject trial Greenhouse–Geisser correction ($\epsilon=0.678$) was not significant ($F_{12,890,979,644}=1.60$, $P<0.080$, $\eta^2=0.021$), and neither was a Huynh–Feldt correction ($\epsilon=0.858$, $F_{16,308,1239,416}=1.60$, $P<0.060$). Additionally, the subject sex and bin interaction was not significant

Table 2. Escape: comparison in latency between individual yoked and master subjects under the prediction that master subjects would have shorter latencies to respond than yoked subjects

	PCC (%)	c-value
Workers		
Group	63	0.01
B1	67	0.01
B2	66	0.01
B3	62	0.01
B4	14	1
B5	49	0.55
B6	97	0.01
B7	89	0.01
B8	85	0.01
B9	72	0.01
B10	70	0.01
B11	53	0.13
B12	60	0.01
B13	80	0.01
B14	66	0.01
B15	43	0.01
B16	94	0.01
B17	37	1
B18	47	0.9
B19	53	0.22
B20	84	0.01
Drones		
Group	49	0.01
B1	13	1
B2	79	0.01
B3	71	0.01
B4	60	0.01
B5	52	0.01
B6	26	1
B7	22	1
B8	57	0.01
B9	40	0.92
B10	51	0.47
B11	55	0.06
B12	80	0.01
B13	22	1
B14	63	0.01
B15	28	0.03
B16	63	0.01
B17	78	0.01
B18	29	0.01
B19	20	1
B20	79	0.01

with a Greenhouse–Geisser correction ($F=1.191$, $P<0.281$, $\eta^2=0.015$), nor was the interaction significant with a Huynh–Feldt correction ($P<0.267$). Moreover, the subject's master or yoked status and bin interaction was not significant with a Greenhouse–Geisser correction ($F=1.372$, $P<0.167$, $\eta^2=0.018$), nor was the interaction significant with a Huynh–Feldt correction, $P<0.146$. However, a significant between-sex difference was observed ($F_{1,76}=25.724$, $P=0.000$, $\eta^2=0.253$); drones ($M=17.005$, 95% CI=14.971, 19.040) had higher latencies than workers ($M=9.678$, 95% CI=7.643, 11.712). Moreover, a significant difference was observed between a subject's master or yoked status, ($F_{1,76}=5.378$, $P=0.023$, $\eta^2=0.066$), even though master subjects' confidence intervals ($M=11.666$, 95% CI=9.632, 13.701) overlapped with those of yoked subjects ($M=15.017$, 95% CI=12.982, 17.051). This inconsistency may be related to a subject's master or yoked status existing as a truly dependent between-subjects comparison, for a repeated-measures ANOVA can only assume

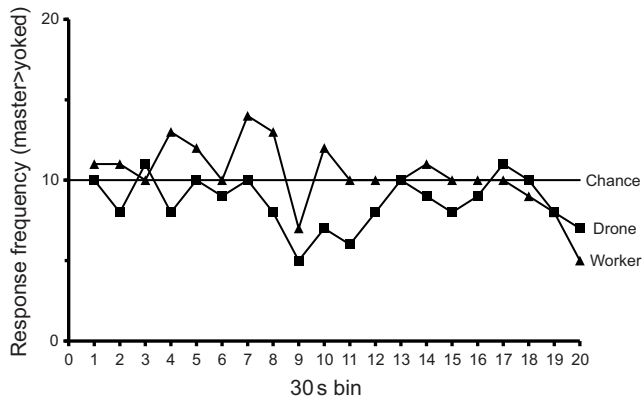


Fig. 2. Experiment 1: Escape. Frequency of master responses per bin that are greater than yoked responses per bin. Shown are the total occurrences in which a master subject responded more times per bin than its paired yoked subject. The ordinates indicate how many of the total 20 master–yoked pairs had master subjects responding more than their paired yoked subject compared with a chance occurrence of 50%. Filled triangles represent workers and filled squares represent drones.

independent between-subject comparisons; our master and yoke subjects are dependent. A significant between-subject interaction between sex and master or yoked status was not observed ($F=1.476$, $P<0.228$, $\eta^2=0.019$).

Additionally, we performed a repeated-measures ANOVA with an *a priori* alpha level of 0.05 to assess differences in the number of escape responses. Mauchly's test of sphericity was significant ($W=0.004$, $\chi^2_{189}=390.116$, $P<0.000$). A within-subject trial Greenhouse–Geisser correction ($\epsilon=0.551$) was significant ($F_{10.477,796.257}=1.839$, $P<0.047$, $\eta^2=0.024$), as was a Huynh–Feldt correction ($\epsilon=0.672$, $F_{12.761,969.868}=1.839$, $P<0.035$). However, the subject sex and bin interaction was not significant with a Greenhouse–Geisser correction ($F=0.822$, $P<0.613$, $\eta^2=0.011$), nor was the interaction significant with a Huynh–Feldt correction ($P<0.635$). Moreover, the subject's master or yoke status and bin interaction was not significant with a Greenhouse–Geisser correction ($F=1.437$, $P<0.155$, $\eta^2=0.019$), nor was the interaction significant with a Huynh–Feldt correction ($P<0.138$). However, a significant between-sex difference was observed ($F_{1,76}=49.157$, $P=0.000$, $\eta^2=0.393$): workers ($M=3.104$, 95% CI=2.734, 3.474) made more responses than drones ($M=1.263$, 95% CI=0.893, 1.632). A significant difference was not observed between a subject's master or yoked status ($F_{1,76}=3.827$, $P=0.054$, $\eta^2=0.048$): master subjects ($M=2.440$, 95% CI=2.070, 2.810) did not make more responses than yoke subjects ($M=1.926$, 95% CI=1.556, 2.296). A significant between-subject interaction between sex and master or yoke status was not observed ($F=1.615$, $P<0.208$, $\eta^2=0.021$).

To demonstrate the complications of continuity assumptions in null hypothesis significance testing, we also performed a series of Friedman's tests to assess within-subject changes in escape behavior across trials. Friedman's test assumes the sample is a single group; hence, we performed four assessments. Master worker subjects ($\chi^2_{19}=30.057$, $P<0.051$) did not significantly change response performance across trials, nor did yoked worker subjects ($\chi^2_{19}=22.578$, $P<0.256$), master drone subjects ($\chi^2_{19}=19.423$, $P<0.430$) or yoked drone subjects ($\chi^2_{19}=26.627$, $P<0.114$).

Experiment 2

We measured avoidance learning by creating ten 60s bins to develop a percentage of the subject's correct responding (i.e. time spent not

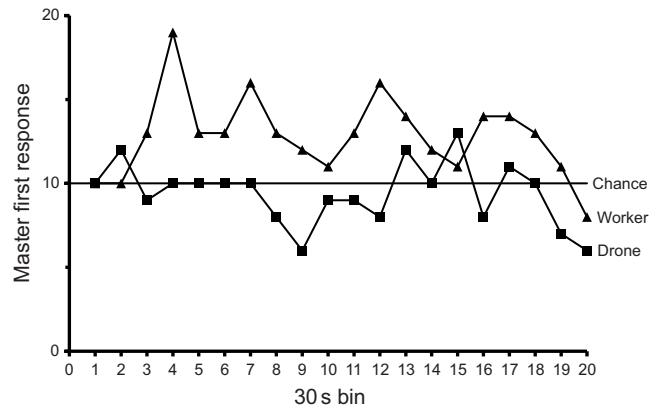


Fig. 3. Experiment 1: Escape. Master first response totals. Shown are the total occurrences in which a master subject of each master–yoked pair was first to make a response following the onset of the unavoidable shock. Each shock occurred every 30 s, indicating the start of each bin, and continued until terminated by the master subject for a total of 20 trials (10 min). Each bin was exactly 30 s long. The ordinates indicate the number of master first responses of the total sample group at each bin compared with a chance occurrence of 50%. Filled triangles represent workers and filled squares represent drones.

being shocked) across each 60 s bin. To assess within-subject improvement across the 10 min trial, we predicted a monotonic increase in correct response length percentages across the 10 bins while separating for each subject. Group PCC values were calculated by pooling each subject within a group to allow assessment of group differences of within-subject improvement across the trials without relying on measures of central tendency. Individual and group ordinal PCC values and corresponding *c*-values are displayed in Table 3; subjects did not tend to monotonically improve avoidance performance across bins and no systematic differences in monotonic improvement between workers and drones were observed. A visual representation of an aggregated worker–drone comparison of correct responding is portrayed in Fig. 4.

To assess between-sex differences, we predicted that workers would have higher PCC responding values compared with drones, and observed that workers successfully avoided shock for longer periods of time than drones (PCC: 64%; *c*-value: 0.01). Additionally, we assessed subject performance by comparing each individual's percentage of correct responding to a chance performance of 50% correct responding, or avoiding shock for 30 s of the 60 s bin. We predicted that workers and drones would both perform better than 50% correct responding. Consistent with our between-sex analysis, drones (PCC: 52%, *c*-value: 0.13) performed at a lower responding PCC value compared with workers (PCC: 64%, *c*-value: 0.01).

To facilitate a comparison in data analysis strategies, we performed a repeated-measures ANOVA with an *a priori* alpha level of 0.05 to assess within-trial differences in avoidance learning. Mauchly's test of sphericity was significant ($W=0.113$, $\chi^2_{44}=76.015$, $P<0.002$). A within-subject trial Greenhouse–Geisser correction ($\epsilon=0.692$) was not significant ($F_{6.224,236.522}=1.997$, $P<0.064$, $\eta^2=0.050$), but a Huynh–Feldt correction ($\epsilon=0.864$) was significant ($F_{7.772,236.522}=1.997$, $P<0.048$). These differences in corrections complicate interpretations of this within-subject trial assessment; however, overlapping confidence intervals surrounding mean performances of workers' and drones' first and last trials dissuades a rejection of the null hypothesis. Additionally, the subject sex and trial interaction was not significant with a Greenhouse–Geisser correction ($F=0.764$, $P<0.604$, $\eta^2=0.020$), nor was the interaction

Table 3. Punishment: within-subject improvement across bins of time spent on the safe side of the shock grid under the prediction that subjects would show a monotonic increase in time spent on the safe side

	PCC (%)	c-value
Workers		
Group	49	0.07
B1	62	0.1
B2	42	0.09
B3	69	0.06
B4	42	0.73
B5	27	0.56
B6	42	0.73
B7	62	0.11
B8	40	0.53
B9	64	0.13
B10	44	0.31
B11	53	0.44
B12	56	0.14
B13	24	0.96
B14	49	0.56
B15	49	0.56
B16	47	0.63
B17	47	0.48
B18	40	0.74
B19	24	0.98
B20	84	0
Drones		
Group	51	0.04
B1	51	0.46
B2	40	0.75
B3	49	0.58
B4	27	0.94
B5	38	0.84
B6	36	0.86
B7	29	0.92
B8	47	0.62
B9	60	0.12
B10	62	0.15
B11	58	0.25
B12	51	0.48
B13	80	0
B14	58	0.28
B15	22	0.44
B16	73	0.03
B17	64	0.06
B18	56	0.16
B19	56	0.28
B20	56	0.32

significant with a Huynh–Feldt correction ($P < 0.631$). However, significant between-sex differences were observed ($F_{1,38} = 4.393$, $P < 0.043$, $\eta^2 = 0.104$): workers ($M = 37.710$, 95% CI = 31.765, 43.655) had higher percentages of correct responding than drones ($M = 29.005$, 95% CI = 23.060, 34.950), even though overlapping confidence intervals around the means were observed.

Experiment 3

Using the same methods in Experiment 2, we assessed the percentage of a subject’s correct responding to determine whether honey bees could discriminate between colors. To assess within-subject improvement across the trial, we predicted a monotonic increase in correct response length percentages across the 10 bins while separating for each subject. Additionally, group PCC values were calculated by pooling each subject within a group to allow assessment of group differences of within-subject improvement

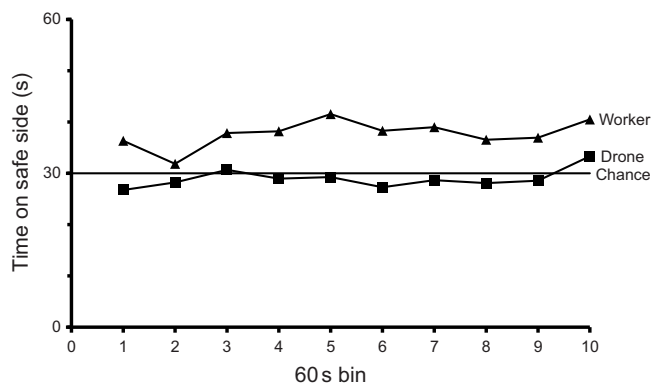


Fig. 4. Experiment 2: Punishment. Average total time on safe side per bin. Shown are the group averages for time spent correctly responding (time spent on the no-shock portion of grid: safe portion) per bin compared with a chance correct responding rate of 50%. A 10 min trial was partitioned into ten 60 s bins for data analysis. Filled triangles represent workers and filled squares represent drones.

across the trials without relying on measures of central tendency. Drones’ individual and group ordinal PCC values and corresponding c -values are displayed in Table 4 while workers’ individual and group ordinal PCC values and corresponding c -values are displayed in Table 5; subjects did not tend to monotonically improve avoidance performance across bins, and no systematic group differences in monotonic improvement between colors, or in workers and drones were observed. A visual representation of an aggregated worker–drone comparison of correct responding for both yellow and blue is portrayed in Fig. 5.

Table 6 displays our between-group assessments for both training and extinction conditions. Subjects that did not respond during the extinction trial were not included in the group extinction analysis. Before assessing sex differences, we investigated if any color discrimination bias was observable in workers and drones. We first separated which color was the punished S+ for each sex before assessing general between-sex differences under the prediction that workers would have higher percentages of correct responding than drones. Based on these between-group ordinal comparisons reported in Table 4, workers with blue as the S+ performed better than each of the three other groups.

Additionally, we assessed subject performance by comparing each individual’s responding PCC value to a chance performance of 50% correct responding, or avoiding shock for 30s of the 60 s bin. We predicted that workers and drones would both perform better than 50% correct responding. Consistent with our between-sex analysis, drones (PCC: 53%, c -value: 0.20) performed at a lower percentage of correct responding compared with workers (PCC: 78%, c -value: 0.00) when shock was paired with blue. However, when shock was paired with yellow there was no observable difference between drones (PCC: 46%, c -value: 0.92) and workers (PCC: 45%, c -value: 0.95).

To facilitate a comparison in data analysis strategies, we performed a repeated-measures ANOVA with an *a priori* alpha level of 0.05 to assess within-bin differences and between sex and color differences. Mauchly’s test of sphericity was significant ($W = 0.327$, $\chi^2_{44} = 81.351$, $P < 0.001$). A within-subject bin Greenhouse–Geisser correction ($\epsilon = 0.804$) was not significant ($F_{7,235,549.896} = 0.850$, $P < 0.549$, $\eta^2 = 0.011$), and a Huynh–Feldt correction ($\epsilon = 0.932$) was not significant ($F_{8,389,637.533} = 0.850$, $P < 0.563$). Additionally, the subject sex and bin interaction was not significant with a

Table 4. Punishment with discrimination: within-trial improvement across bins of time spent on the safe side for drones under the prediction that subjects would show a monotonic increase in time spent on the safe side

	PCC (%)	c-value
Yellow		
Group	55	0.01
B1	53	0.45
B2	51	0.52
B3	60	0.24
B4	69	0.09
B5	73	0.04
B6	33	0.92
B7	38	0.84
B8	58	0.3
B9	62	0.19
B10	51	0.52
B11	60	0.29
B12	53	0.41
B13	51	0.46
B14	53	0.44
B15	47	0.64
B16	60	0.27
B17	44	0.71
B18	76	0.01
B19	51	0.48
B20	62	0.22
Blue		
Group	51	0.34
B1	31	0.96
B2	44	0.7
B3	67	0.22
B4	33	0.93
B5	49	0.58
B6	38	0.8
B7	67	0.1
B8	62	0.21
B9	56	0.41
B10	69	0.08
B11	33	0.95
B12	44	0.74
B13	49	0.59
B14	49	0.41
B15	60	0.22
B16	67	0.14
B17	51	0.49
B18	49	0.5
B19	78	0.01
B20	27	0.97

Table 5. Punishment with discrimination: within-trial improvement across bins of time spent on the safe side for workers under the prediction that subjects would show a monotonic increase in time spent on the safe side

	PCC (%)	c-value
Yellow		
Group	51	0.38
B1	47	0.6
B2	40	0.84
B3	56	0.31
B4	80	0.01
B5	51	0.48
B6	60	0.21
B7	47	0.65
B8	47	0.64
B9	33	0.9
B10	51	0.49
B11	62	0.24
B12	38	0.88
B13	58	0.37
B14	42	0.76
B15	60	0.25
B16	47	0.58
B17	53	0.42
B18	51	0.55
B19	69	0.1
B20	29	0.96
Blue		
Group	51	0.44
B1	40	0.8
B2	58	0.28
B3	47	0.64
B4	36	0.91
B5	58	0.31
B6	51	0.55
B7	40	0.81
B8	64	0.15
B9	22	1
B10	64	0.16
B11	71	0.03
B12	42	0.77
B13	51	0.47
B14	44	0.73
B15	47	0.58
B16	62	0.12
B17	56	0.28
B18	27	0.98
B19	78	0.01
B20	53	0.45

Greenhouse–Geisser correction ($F=0.975$, $P<0.450$, $\eta^2=0.013$), nor with a Huynh–Feldt correction ($P<0.457$). The bin and color interaction was not significant with a Greenhouse–Geisser correction ($F=1.318$, $P<0.238$, $\eta^2=0.017$), nor was the interaction significant with a Huynh–Feldt correction ($P<0.229$). The bin, subject and color interaction was also not significant with a Greenhouse–Geisser correction ($F=1.245$, $P<0.275$, $\eta^2=0.016$), nor was the interaction significant with a Huynh–Feldt correction ($P<0.268$). However, significant between-sex differences were observed ($F_{1,76}=27.986$, $P<0.000$, $\eta^2=0.269$): drones ($M=29.145$, 95% CI=27.755, 30.534) had lower percentages of correct responding than workers ($M=34.363$, 95% CI=32.974, 35.752). Moreover, significant between-color differences were observed ($F=47.667$, $P<0.000$, $\eta^2=0.385$): the percentages of correct responding on yellow ($M=28.348$, 95% CI=26.959, 29.738) were lower than percentages of correct responding on blue ($M=35.159$, 95% CI=33.770, 36.548).

A significant interaction between sex and color was observed ($F=14.930$, $P<0.000$, $\eta^2=0.164$).

Additionally, we performed a repeated-measures ANOVA with an *a priori* alpha level of 0.05 to assess within-bin differences and between-sex and -color differences during extinction trials. Nine subjects did not respond during extinction; hence their data were not included in the extinction analysis. A significant within-subject bin effect was not observed ($F_{1,67}=0.318$, $P<0.575$, $\eta^2=0.005$). Additionally, the subject sex and bin interaction was not significant ($F=0.838$, $P<0.450$, $\eta^2=0.001$). The bin and color interaction was not significant ($F=0.245$, $P<0.622$, $\eta^2=0.004$). The bin, subject and color interaction was also not significant ($F=0.158$, $P<0.692$, $\eta^2=0.002$). However, significant between-sex differences were observed ($F_{1,67}=16.559$, $P<0.000$, $\eta^2=0.198$): drones ($M=23.785$, 95% CI=21.238, 26.332) had lower percentages of correct responding than workers ($M=31.551$, 95% CI=28.718, 34.384).

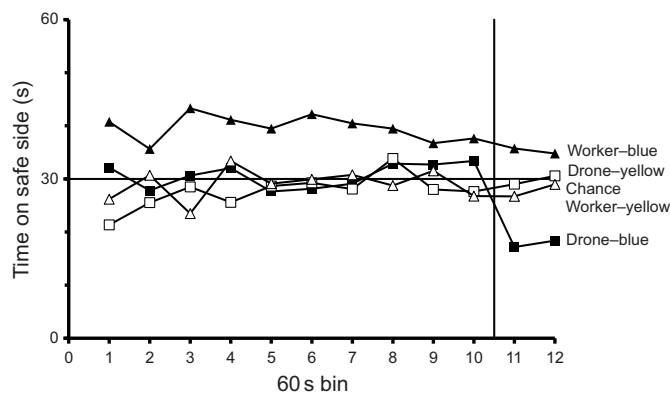


Fig. 5. Experiment 3: Punishment with color discrimination. Average total time on safe side per bin. Shown are the group averages for time spent correctly responding (time spent on the no-shock portion of grid: safe portion) per bin compared with a chance correct responding rate of 50%. A 10 min trial was partitioned into ten 60 s bins for data analysis. Filled triangles represent workers in the treatment group whose shock was paired with blue, and filled squares represent drones in the same treatment group. Open triangles represent workers in the treatment group whose shock was paired with yellow, and open squares represent drones in the same treatment group. The vertical line at the 10th bin indicates the end of the training period and the beginning of the extinction period, where no shock was administered.

Significant between-color differences were not observed ($F=1.453$, $P<0.232$, $\eta^2=0.021$): correct responding on yellow ($M=28.818$, 95% CI=26.018, 31.618) did not differ from correct responding on blue ($M=26.518$, 95% CI=23.935, 29.101). However, a significant interaction between sex and color was observed ($F=25.927$, $P<0.000$, $\eta^2=0.279$).

Data analysis comparison

In addition to utilizing OOM, we also analyzed our data *via* null hypothesis significance testing (NHST) to demonstrate four main differences in data analysis methods and results between OOM and NHST. First, assumptions of continuity are not made in OOM; discriminating between ordinal data or ratio data and determining whether a parametric or non-parametric assessment should be performed is not paramount, for both forms of data are assessed in the same manner in OOM and the ordinal analysis. To exemplify the difficulties of continuity concerns in NHST, we performed a series of Friedman’s tests and a repeated-measures ANOVA and argue that responding can be conceived as ordinal data or ratio data, for a response count per trial can be argued to be ordinal and ratio data. Each trial in Experiment 1 lasted for 30s; dividing each trial’s ordinal response frequency by a constant 30s to create a trial response rate produces truly continuous data. Even though NHST conclusions are not affected by constant multiplicative scale changes at a trial level, the same data can undergo two radically different assessment procedures and could result in different rejection

decisions of the null hypothesis. Indeed, while using the same response data points from Experiment 1, a repeated-measures ANOVA reported significant within-subject differences in response rate while a Friedman’s test failed to report significant within-subject differences in response numbers. Second, NHST assumptions about sphericity or homogeneity are eschewed in OOM because means, variances and sums of squares are not compared. As such, concerns about corrections and the observed inconsistent results from different corrections are avoided in OOM (e.g. in Experiment 2, a Greenhouse–Geisser correction was not significant but a Huynh–Feldt correction was significant). Third, OOM tests do not utilize critical alpha levels; hence, there are no concerns about alpha-level adjustments following numerous tests. In focusing on the individual observations of the collected data, generalizations to population parameters are not made; rather, uniqueness of the specific observations of the data are assessed. Fourth, testing between-group dependency or between-group independency does not involve inherently different methods. In Experiment 1, latency data were assessed within subjects (trial) in a repeated-measures design, and between subjects independently (sex) and dependently (master status). The assumptions and requirements to properly run a between-subjects dependent *t*-test while controlling for within-subject trials in a single assessment are not met in NHST.

Several clear advantages of OOM are identifiable when compared with comparisons of measures of central tendency. Concerns of unrepresentative aggregates because of outlier effects or dichotomous trends in individual performances are irrelevant in OOM. Adjusting critical alpha levels after performing multiple tests is unneeded in OOM. Complications with missing trial data do not result in all of the subject’s data remaining unassessed in OOM. Abstract, often impossible population parameters are not compared as if they are concrete individual observations in OOM. Instead of providing a probability value of a dataset’s extremity based on pre-determined alpha levels, OOM provides a chance value of the observed dataset’s uniqueness compared with a series of randomizations of the dataset. Finally, the PCC value indicates the percentage of data points in a group that are larger or smaller than an alternative group; we believe the information in a PCC value offers an easily comprehensible summary of the dataset compared with the required hodgepodge of tests required to thoroughly assess a between-subject repeated-measures design.

DISCUSSION

The neural plasticity and complex social interactions within a colony make honey bees an advantageous model. Female workers constantly interact and transition from tasks inside the nest when they are young. In contrast, a drone’s life history primarily constitutes finding and memorizing specific congregation sites in which mating will occur with no involvement in other colony-associated tasks. This system provides an ideal opportunity for comparative examination of learning. Further analyses of learning across the more solitary drones and the more social worker bees

Table 6. Punishment with discrimination: between-groups ordinal comparison of time spent on safe side during training and extinction

Ordinal prediction	PCC (%)		c-value	
	Training	Extinction	Training	Extinction
Drone–yellow < drone–blue	57	21	0.01	1.00
Worker–yellow < worker–blue	74	67	0.01	0.01
Drone–yellow < worker–yellow	53	45	0.01	1.00
Drone–blue < worker–blue	71	85	0.01	0.01
Drone < worker	61	69	0.01	0.01

sets the stage for later phylogenetic studies of learning in solitary and social bees (Fischman et al., 2011; Woodard et al., 2011). Here, we present such an analysis using two forms of learning: escape and punishment.

The escape experiment consisted of two measurements recorded in a master–yoked paradigm: frequency to respond and latency to respond following administration of shock. For both workers and drones and for both measurements, there was large individual response variation. Worker individual master–yoked comparisons of PCC values ranged from 1 to 98% for response frequency and 14 to 97% for latency to respond. Drone individual master–yoked comparisons of PCC values ranged from 0 to 83% for response frequency and 13 to 80% for latency to respond. Despite the range of variation between individuals for workers and drones, master subjects as a whole for both sexes performed better under both measurements. When comparing workers with drones, workers performed better under both measurements.

Escape behavior is a basic and ecologically significant strategy for avoiding predation. As such, our results match expectations that both drones and workers are capable of escape learning. Foraging workers experience increased mortality rates as they age. Dukas (Dukas, 2008) argues both age-dependent and independent mortality rates are due primarily to predation and posits that learning contributes to some decrease in the observed age-specific mortality rate. Additionally, male honey bees experience similar increases in mortality rates once flight is initiated, which is, at least in part, attributed to predation (Rueppell et al., 2005); one should expect little to no difference in their ability to learn under this paradigm, as our results demonstrate. Disparity in the response of workers and drones was observed and does not seem to be constrained to learning. In one set of studies, aggressive response of workers and drones were analyzed, and results showed that worker reaction is more threshold dependent compared with drone reaction when under exposure to the same aversive stimuli (A.A. and T.G., unpublished). Together, these observations might indicate a difference in threshold for reaction to negative stimuli, which prompts further examination.

In the punishment experiment, there was no indication of monotonic increase in correct response patterns across bins for either drones or workers. A between-sex comparison revealed that workers performed better than drones and effectively avoided shock for longer periods of time. In addition to the improvement measure and the between-sex comparison, a flat ‘to chance’ comparison was performed comparing correct responses per bin with the 50% chance response value. Workers performed better than chance and outperformed drones in this measure. Drones did not perform significantly better than chance.

A lack of monotonic increase in response pattern across bins for both drones and workers was also observed in the punishment with discrimination experiment. For this discrimination experiment, the color of the discriminative stimulus played a pivotal role on the avoidance potential of the bee. As with the punishment experiment, correct response values per bin were compared with a chance correct response value of 50%. When blue was paired with shock, subjects performed better than subjects whose shock was paired with yellow. Between-sex comparisons revealed workers performed better than drones when shock was paired with blue; however, there was no observable difference between workers and drones when the shock was paired with yellow. Both workers and drones performed worse than chance when shock was paired with yellow and performed better than chance when shock was paired with blue. Both sexes displayed a preference for spending time on the yellow side; this

bias inhibits a clear conclusion on learning ability and demonstrates the importance of counterbalancing discriminative stimuli.

Negative associations have been shown to be an important component with regards to resource patch visitation and signaling by honey bee workers (McNally and Westbrook, 2006; Abramson et al., 2006; Nieh, 2010). Similarities can be drawn between the importance of resource patches for workers and mating congregation sites for drones. As such, we expected associative (punishment) learning ability to be similar, but our results did not support this expectation. An explanation for workers outperforming drones in this type of learning might stem from our methods. The described complications might also account for the differences in response demonstrated in Experiment 1: escape.

A point to consider in interpreting our drone data is the suitability of the shuttlebox. In contrast to the smooth performance of the worker bees, the movement of the drones often appeared ‘clumsy’. To keep the experimental design as similar as possible between the worker and drones, we opted to use the same dimensions. Our pilot work showed both the workers and drones could turn around in the apparatus; hence, we attempted to address inherent locomotion differences between drones and workers. While they could easily turn around in the shuttlebox, they would often run to the end of the compartment before turning in the opposite direction. We believe this behavior may have been caused, in part, by their momentum and is reminiscent of what the early comparative psychologist Schneirla called ‘centrifugal swing’ in his discussion of the errors made by ants in a complex maze (Abramson, 1997).

Our experiment represents the first time that a drone has been tested in a shuttlebox; however, the development of a suitable apparatus to test the learning of drones in non-appetitive situations is still required [for a review of apparatus used for the study of invertebrate learning, see Abramson (Abramson, 1994)]. The popular PER in honey bee research has been shown to have methodological inconsistencies across laboratories (Abramson et al., 2011; Matsumoto et al., 2012; Frost et al., 2012); these inconsistencies have led to the renewed interest in developing aversive conditioning shuttlebox situations for honey bees (Agarwal et al., 2011).

In the punishment with discrimination experiment, color significantly affected response behavior. Foraging workers experience and use color as a significant environmental cue. Past experience by the examined individuals to specific color associations might explain the observed bias. Constancy towards an experienced color has been described (Hill et al., 1997). Furthermore, subspecies of bees have been shown to vary in their experience-dependent preference and constancy towards flower/resource coloration (Cakmak et al., 2010). Previous analysis of punishment learning with discrimination on a gentle Africanized hybrid (gAHB) found in Puerto Rico did not find differences in color preferences (Agarwal et al., 2011). We therefore conclude that in future experiments that use color as a discriminatory cue, care must be taken in the selection process.

ACKNOWLEDGEMENTS

We thank Ian Finley and Corey Vyhldal of Oklahoma State University for their assistance with circuit design and construction.

AUTHOR CONTRIBUTIONS

C.W.D., C.I.A., T.G. and H.W. conceived and designed the experiments. C.W.D., F.N.D. and C.I.A. performed the experiments. C.W.D., A.A., D.P.A.C., Z.M.A. and C.A.V. analyzed the data. C.W.D., C.A.V. and F.N.D. contributed reagents, materials and/or analysis tools. C.W.D., A.A., C.I.A., D.P.A.C., Z.M.A., C.A.V., T.G. and H.W. wrote the paper.

COMPETING INTERESTS

No competing interests declared.

FUNDING

This research was supported in part by National Science Foundation grant DBI 0552717.

REFERENCES

- Abramson, C. I. (1986). Aversive conditioning in honey bees (*Apis mellifera*). *J. Comp. Psychol.* **100**, 108-116.
- Abramson, C. I. (1994). *A Primer of Invertebrate Learning: The Behavioral Perspective*. Washington, DC: American Psychological Association.
- Abramson, C. I. (1997). Where have I heard it all before: some neglected issues of invertebrate learning. In *Comparative Psychology of Invertebrates: The Field and Laboratory Study of Insect Behavior* (ed. G. Greenberg and E. Tobach), pp. 55-78. New York, NY: Garland Publishing.
- Abramson, C. I. and Boyd, B. J. (2001). An automated apparatus for conditioning proboscis extension in honey bees (*Apis mellifera* L.). *J. Entomol. Sci.* **36**, 78-92.
- Abramson, C. I., Singleton, J. B., Wilson, M. K., Wanderley, P. A., Ramalho, F. S. and Michaluk, L. M. (2006). The effect of an organic pesticide on mortality and learning in Africanized honey bees (*Apis mellifera* L.) in Brasil. *Am. J. Environ. Sci.* **2**, 37-44.
- Abramson, C. I., Sokolowski, M. B. C. and Wells, H. (2011). Issues in the study of proboscis conditioning. In *Social Insects: Structure, Function, and Behavior* (ed. F. Columbus), pp. 25-49. Hauppaug, NY: Nova Science Publishers.
- Agarwal, M., Giannoni Guzmán, M., Morales-Matos, C., Del Valle Díaz, R. A., Abramson, C. I. and Giray, T. (2011). Dopamine and octopamine influence avoidance learning of honey bees in a place preference assay. *PLoS ONE* **6**, e25371.
- Aquino, I. S., Abramson, C. I., Soares, A. E., Fernandes, A. C. and Benbassat, D. (2004). Classical conditioning of proboscis extension in harnessed Africanized honey bee queens (*Apis mellifera* L.). *Psychol. Rep.* **94**, 1221-1231.
- Benatar, S. T., Cobey, S. and Smith, B. H. (1995). Selection on a haploid genotype for discrimination learning performance: correlation between drone honey bees (*Apis mellifera*) and their worker progeny (Hymenoptera: Apidae). *J. Insect Behav.* **8**, 637-652.
- Bhagavan, S., Benatar, S., Cobey, S. and Smith, B. H. (1994). Effect of genotype but not age or caste on olfactory learning performance in the honey bee, *Apis mellifera*. *Anim. Behav.* **48**, 1357-1369.
- Bitterman, M. E., Menzel, R., Fietz, A. and Schäfer, S. (1983). Classical conditioning of proboscis extension in honeybees (*Apis mellifera*). *J. Comp. Psychol.* **97**, 107-119.
- Cakmak, I., Song, D. S., Mixson, T. A., Serrano, E., Clement, M. L., Savitski, A., Johnson, G. A., Giray, T., Abramson, C. I., Barthell, J. F. et al. (2010). Foraging response of Turkish honey bee subspecies to flower color choices and reward consistency. *J. Insect Behav.* **23**, 100-116.
- Chandra, S. B. C., Hunt, G. J., Cobey, S. and Smith, B. H. (2001). Quantitative trait loci associated with reversal learning and latent inhibition in honeybees (*Apis mellifera*). *Behav. Genet.* **31**, 275-285.
- Craig, D. P., Grice, J. W., Varnon, C. A., Gibson, B., Sokolowski, M. B. C. and Abramson, C. I. (2012). Social reinforcement delays in free-flying honey bees (*Apis mellifera* L.). *PLoS ONE* **7**, e46729.
- Dukas, R. (2008). Mortality rates of honey bees in the wild. *Insectes Soc.* **55**, 252-255.
- Fahrbach, S. E., Giray, T. and Robinson, G. E. (1995). Volume changes in the mushroom bodies of adult honey bee queens. *Neurobiol. Learn. Mem.* **63**, 181-191.
- Fahrbach, S. E., Giray, T., Farris, S. M. and Robinson, G. E. (1997). Expansion of the neuropil of the mushroom bodies in male honey bees is coincident with initiation of flight. *Neurosci. Lett.* **236**, 135-138.
- Fahrbach, S. E., Moore, D., Capaldi, E. A., Farris, S. M. and Robinson, G. E. (1998). Experience-expectant plasticity in the mushroom bodies of the honeybee. *Learn. Mem.* **5**, 115-123.
- Fahrbach, S. E., Farris, S. M., Sullivan, J. P. and Robinson, G. E. (2003). Limits on volume changes in the mushroom bodies of the honey bee brain. *J. Neurobiol.* **57**, 141-151.
- Farris, S. M., Robinson, G. E. and Fahrbach, S. E. (2001). Experience- and age-related outgrowth of intrinsic neurons in the mushroom bodies of the adult worker honeybee. *J. Neurosci.* **21**, 6395-6404.
- Ferguson, H. J., Cobey, S. and Smith, B. H. (2001). Sensitivity to a change in reward is heritable in the honeybee, *Apis mellifera*. *Anim. Behav.* **61**, 527-534.
- Fischman, B. J., Woodard, S. H. and Robinson, G. E. (2011). Molecular evolutionary analyses of insect societies. *Proc. Natl. Acad. Sci. USA* **108** Suppl. 2, 10847-10854.
- Frings, H. (1944). The loci of olfactory end-organs in the honeybee. *J. Exp. Zool.* **97**, 123-134.
- Frost, E. H., Shutler, D. and Hillier, N. K. (2012). The proboscis extension reflex to evaluate learning and memory in honeybees (*Apis mellifera*): some caveats. *Naturwissenschaften* **99**, 677-686.
- Giray, T. and Robinson, G. E. (1996). Endocrine-mediated behavioral development in male honey bees and the evolution of division of labor. *Proc. Natl. Acad. Sci. USA* **93**, 11718-11722.
- Giray, T., Huang, Z. Y., Guzmán-Novoa, E. and Robinson, G. E. (1999). Physiological correlates of genetic variation for rate of behavioral development in the honeybee, *Apis mellifera*. *Behav. Ecol. Sociobiol.* **47**, 17-28.
- Giray, T., Giovanetti, M. and West-Eberhard, M. J. (2005). Juvenile hormone, reproduction, and worker behavior in the neotropical social wasp *Polistes canadensis*. *Proc. Natl. Acad. Sci. USA* **102**, 3330-3335.
- Giray, T., Galindo-Cardona, A. and Oskay, D. (2007). Octopamine influences honey bee foraging preference. *J. Insect Physiol.* **53**, 691-698.
- Grice, J. W. (2011). *Observation Oriented Modeling: Analysis of Cause in the Behavioral Sciences*. San Diego, CA: Academic Press.
- Grice, J. W., Barrett, P. T., Schlingens, L. A. and Abramson, C. I. (2012). Toward a brighter future for psychology as an observation oriented science. *Behav. Sci.* **2**, 1-22.
- Grozinger, C. M., Fan, Y., Hoover, S. E. R. and Winston, M. L. (2007). Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Mol. Ecol.* **16**, 4837-4848.
- Guzman-Novoa, E., Hunt, G. J., Page, R. E., Jr, Uribe-Rubio, J. L., Prieto-Merlos, D. and Becerra-Guzman, F. (2005). Paternal effects on the defensive behavior of honeybees. *J. Hered.* **96**, 376-380.
- Hill, P. S. M., Wells, P. H. and Wells, H. (1997). Spontaneous flower constancy and learning in honey bees as a function of colour. *Anim. Behav.* **54**, 615-627.
- Humphries, M. A., Müller, U., Fondrk, M. K. and Page, R. E., Jr (2003). PKA and PKC content in the honey bee central brain differs in genotypic strains with distinct foraging behavior. *J. Comp. Physiol. A* **189**, 555-562.
- Kaczor, L. and Maldonado, H. (2009). Contrasting role of octopamine in appetitive and aversive learning in the crab *Chasmagnathus*. *PLoS ONE* **4**, e6223.
- Kuwabara, M. (1957). Bildung des bedingten reflexes von pavlovian typus bei der honigbiene, *Apis mellifera* [Establishment of Pavlovian conditioned reflexes in honeybees]. *J. Sci. Hokkaido Univ. Zool.* **13**, 458-464.
- Mackintosh, N. J. (1974). *The Psychology of Animal Learning*. New York, NY: Academic Press.
- Matsumoto, Y., Menzel, R., Sandoz, J. C. and Giurfa, M. (2012). Revisiting olfactory classical conditioning of the proboscis extension response in honey bees: a step toward standardized procedures. *J. Neurosci. Methods* **211**, 159-167.
- McNally, G. P. and Westbrook, R. F. (2006). Predicting danger: the nature, consequences, and neural mechanisms of predictive fear learning. *Learn. Mem.* **13**, 245-253.
- Nieh, J. C. (2010). A negative feedback signal that is triggered by peril curbs honey bee recruitment. *Curr. Biol.* **20**, 310-315.
- Palya, W. L. and Walter, D. E. (1993). A powerful, inexpensive experiment controller or IBM PC interface and experiment control language. *Behav. Res. Methods* **25**, 127-136.
- Robinson, G. E., Page, R. E., Jr, Strambi, C. and Strambi, A. (1989). Hormonal and genetic control of behavioral integration in honey bee colonies. *Science* **246**, 109-112.
- Rueppell, O., Fondrk, M. K. and Page, R. E., Jr (2005). Biodemographic analysis of male honey bee mortality. *Aging Cell* **4**, 13-19.
- Seeley, T. D. (1995). *The Wisdom of The Hive*. Cambridge, MA: Harvard University Press.
- Smith, A. R., Seid, M. A., Jiménez, L. C. and Wcislo, W. T. (2010). Socially induced brain development in a facultatively eusocial sweat bee *Megalopta genalis* (Halictidae). *Proc. Biol. Sci.* **277**, 2157-2163.
- Srinivasan, M. V. (2010). Honey bees as a model for vision, perception, and cognition. *Annu. Rev. Entomol.* **55**, 267-284.
- Takeda, K. (1961). Classical conditioned response in the honeybee. *J. Insect Physiol.* **6**, 168-179.
- Vareschi, E. (1971). Duftunterscheidung bei den honigbiene: Einzellenz-ableitungen und verhaltensreaktionen [Odor discrimination by the honeybee: single cell recording and behavior reaction]. *Z. Vgl. Physiol.* **75**, 143-173.
- Vergoz, V., Roussel, E., Sandoz, J. C. and Giurfa, M. (2007). Aversive learning in honeybees revealed by the olfactory conditioning of the sting extension reflex. *PLoS ONE* **2**, e288.
- Wells, P. H. (1973). Honey bees. In *Invertebrate Learning: Arthropods and Gastropod Mollusks*, Vol. 2 (ed. W. C. Corning, J. A. Dyal and A. O. D. Willows), pp. 173-185. New York, NY: Plenum Press.
- Withers, G. S., Fahrbach, S. E. and Robinson, G. E. (1993). Selective neuroanatomical plasticity and division of labour in the honeybee. *Nature* **364**, 238-240.
- Woodard, S. H., Fischman, B. J., Venkat, A., Hudson, M. E., Varala, K., Cameron, S. A., Clark, A. G. and Robinson, G. E. (2011). Genes involved in convergent evolution of eusociality in bees. *Proc. Natl. Acad. Sci. USA* **108**, 7472-7477.