BIOLOGY LETTERS

royalsocietypublishing.org/rsbl

CC BY

Research

Cite this article: Baciadonna L, Nityananda V. 2026 Temporal consistency of judgement biases in bumblebees. *Biol. Lett.* **22**: 20260011.
https://doi.org/10.1098/rsbl.2026.0011

Received: 7 January 2026
Accepted: 9 April 2026

Subject Category:
Animal behaviour

Subject Areas:
behaviour

Keywords:
active choice, bees, individual differences, invertebrates, optimism, pessimism, repeatability

Author for correspondence:
Luigi Baciadonna
e-mail: luigi.baciadonna@newcastle.ac.uk

Supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.8512245.

THE ROYAL SOCIETY PUBLISHING

Temporal consistency of judgement biases in bumblebees

Luigi Baciadonna[1,3] and Vivek Nityananda[2]

[1]School of Arts and Cultures, and [2]Biosciences Institute, Newcastle University, Newcastle Upon Tyne, UK
[3]Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy

VN, 0000-0002-2878-2425

Judgement bias tasks are increasingly used to assess affective states in animals, yet the extent to which they might reflect transient states or stable traits remains unclear. Here, we tested bumblebees (*Bombus terrestris*) in an active choice task across three repeated sessions to assess individual consistency in the absence of any manipulation. Bees were trained to associate each of two colours with either a high or a low reward, presented in separate chambers. During testing, they were presented with ambiguous colours. Bees were more likely to choose the high-reward chamber and to choose more quickly in response to colours closer to the positive colour. The latency to choose the cues showed significant and moderate repeatability across sessions, suggesting a stable, trait-like underlying component. In contrast, the repeatability of the chamber choices was negligible, indicating that such responses might be largely state-dependent and influenced by situational factors. These findings suggest that judgement biases, particularly as assessed through an active choice task, reflect states affected by external factors. Active choice tasks may help disentangle stable behavioural traits from transient affective states in invertebrates.

1. Introduction

Reliable measures for assessing animal emotions are essential, especially when they can inform and guide welfare-related decisions or contribute to translational research. A well-known paradigm to assess these states is the judgement bias test, where emotional states are inferred based on the evaluation of an ambiguous stimulus [1,2]. In humans, studies employing this paradigm have demonstrated that people with depression or anxiety disorders tend to interpret ambiguous situations negatively [3,4]. Over recentt decades, this paradigm has been adapted to non-human animal species, including invertebrates [5,6].

Since the seminal study that tested this approach in rats [7], several studies have applied the same approach across multiple species. Research on invertebrates, specifically honeybees, bumblebees and fruit flies, demonstrates that they are malleable to the experimental induction of positive or negative emotional states, as demonstrated by cognitive biases in response to ambiguity [6]. A recent meta-analysis of 71 studies across 22 species supports the overall validity of experimental manipulations to probe affective states in animals [5]. However, the effect sizes are relatively small and appear to be affected by several methodological and biological moderators, including task type, training cue reinforcement and the sex of the animals.

Task choice, for example, can impact findings and their interpretation, as suggested by differences between studies using a go/no-go or active choice paradigm [5,8,9]. Go/no-go paradigms require responding to an ambiguous stimulus after learning to respond to a positive one and suppressing the response to a negative one. Active choice paradigms rely on active responses

© 2026 The Author(s). Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

Downloaded from royalsocietypublishing.org/rsbl/article-pdf/doi/10.1098/rsbl.2026.0011/6145623/rsbl.2026.0011.pdf
by Newcastle University user
on 19 November 2026

for each stimulus. Since the animal must make a clear choice, active choice paradigms mitigate against the influence of confounding factors, specifically motivational changes, thus increasing the validity and interpretability of the results [5].

While judgement bias tasks are widely used to assess animal affect, the responses may reflect either transient states or stable traits [5,9,10]. Contextual factors, including housing, stressors or mood-altering drugs can modulate the bias, indicating situational changes typical of an emotional change [5,11]. At the same time, individual animals can display consistent predispositions to interpret ambiguous stimuli optimistically or pessimistically, suggesting a trait component [11]. Similar patterns have been reported in pigs, companion animals and wild species [12–18]. Baseline differences across breeds, test designs or motivational contexts further complicate interpretation [19–22]. Assessing repeatability can help separate these aspects. While state and trait effects are not mutually exclusive, quantifying repeatability provides a baseline estimate of behavioural variability and provides an estimate of how much variation is due to consistent individual differences. This baseline is crucial for meaningful interpretation. Even when treatment effects are detected, their magnitude and reliability may depend on the level of underlying variability. Despite this, repeatability of judgement biases has been investigated only in a few species, showing that individual consistency can range from stable to highly variable [23–26]. Even studies without manipulations are rare; two go/no-go studies found strong learning effects across sessions [24,27]. Thus, the extent to which judgement biases reflect transient affective states rather than stable traits remains unclear.

We used an active choice judgement bias test to measure the effect of repeated testing of judgement bias in bumblebees (*Bombus terrestris*). We hypothesized that, in the absence of any manipulation, responses to ambiguous stimuli would remain stable across repeated tests, indicating a trait component of judgement bias in bumblebees.

2. Material and methods

(a) Animals and experimental set-up

The study was conducted on female worker bumblebees (*Bombus terrestris*). To control for age, only individuals tagged within 24 h of eclosion were included. When not tested, bees had ad libitum access to a 20% (w/w) sucrose solution and approximately 3 g of pollen per day. During experiments, bees were presented with solid-colour visual stimuli on a monitor in the flight arena (figure 1*a*), using the colours detailed in Procenko *et al.* [28].

(b) Pretraining

Before conditioning, bees were familiarized with the reward chambers and allowed to perform independent foraging bouts. Only bees that reliably located the reward without assistance proceeded to training. See electronic supplementary material for additional details.

(c) Training

The conditioning procedure was identical to the one described previously [28]. Two colours (blue and green) were each associated with a different sucrose concentration (low: 30% w/w; high: 50% w/w) in one of two reward chambers (figure 1*a*). On each trial, one of two colours was presented. One colour indicated that a reward (either high or low value) was available in the chamber on one side of the apparatus, while the opposite chamber contained distilled water. For the other colour, the reward and the water location were reversed. Colour, reward value and side were counterbalanced across bees (electronic supplementary material, table S1). Training continued until individuals reached ≥80% correct choices in the last 20 trials. A total of 24 out of 25 bees passed the initial conditioning, with one bee excluded due to strong side biases. Only full entries, when the bee's body fully entered the chamber, were recorded. All trials were video recorded and choice latencies were calculated as the mean of the last two training trials for each respective colour. See electronic supplementary material for additional details.

(d) Judgement bias testing

The test phase consisted of five trials, each presenting a different colour cue on the monitor. Test colours included the two conditioned colours (green and blue) and three ambiguous colours of intermediate value: near-blue (RGB: 0, 140, 150), medium (RGB: 0, 170, 120) and near-green (RGB: 0, 200, 100). Depending on their proximity to the colour indicating a high- or low-reward, ambiguous cues were classified as near-high, medium or near-low (figure 1*a*). The order of colour presentation was pseudorandomized across bees, with the first test colour always being one of the three ambiguous colours. During testing, all cues were unrewarded, with both chambers containing 0.2 ml of distilled water. After the bee made its first choice, it was gently captured and returned to the tunnel connecting the nest box and the arena. Between presentations of test cues, bees received refresher trials consisting of two presentations of each conditioned colour with the corresponding reward. If a bee made an incorrect choice, an additional set of four trials was administered. This procedure was repeated up to two more times (for a total of three sets of four trials), and if errors persisted, refresher training continued until the bee met the learning criterion (≥80% correct choices in the last 20 trials). All trials were video recorded. Bees' first choices, defined as fully entering a reward chamber, and the choice latency, measured as the time from leaving the tunnel to entering the reward chamber, were recorded.

Downloaded from royalsocietypublishing.org/rsbl/article-pdf/doi/10.1098/rsbl.2026.0011/6145623/rsbl.2026.0011.pdf
by Newcastle University user
on 19 November 2026

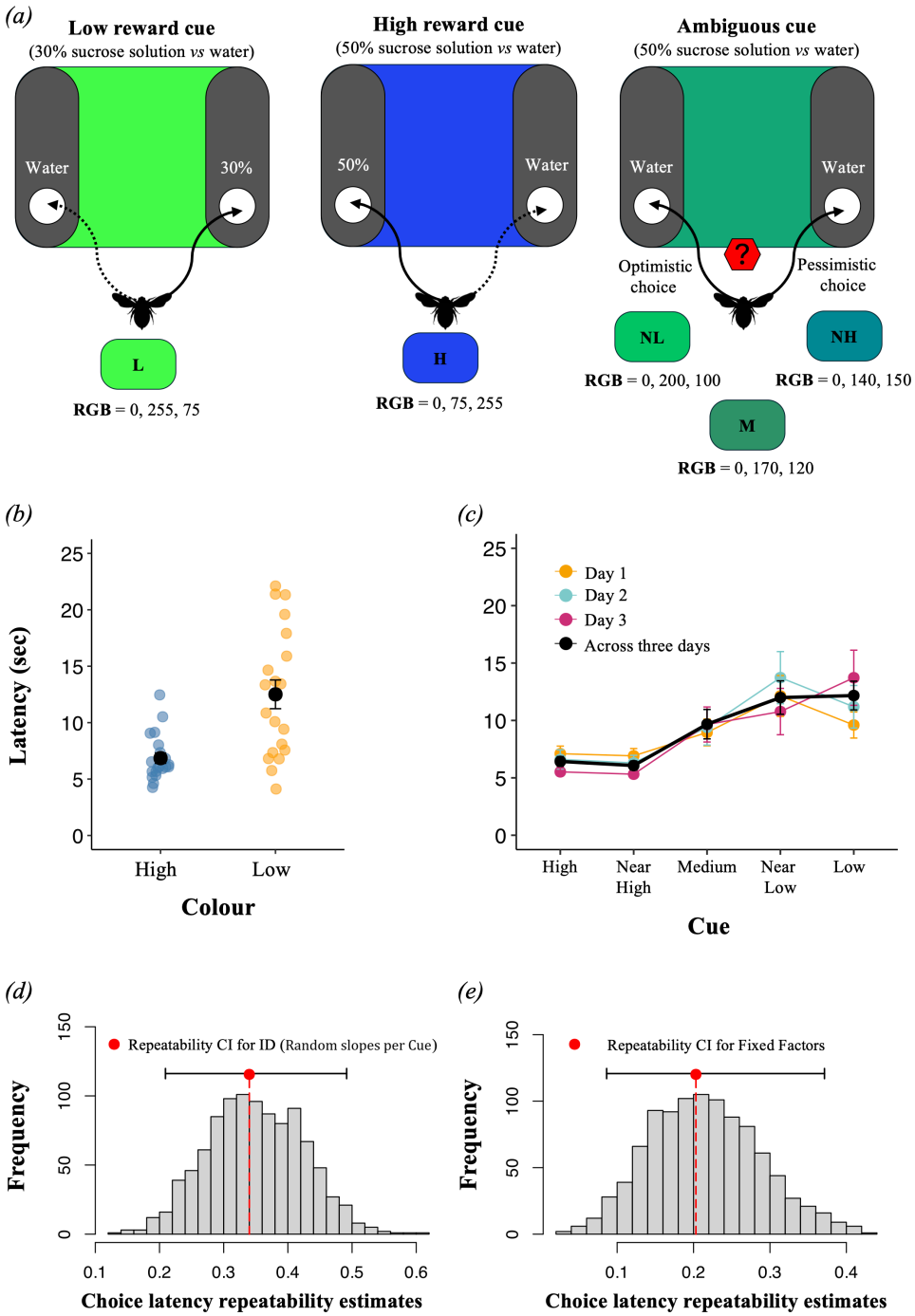

Figure 1. Experimental paradigm and choice latency analyses. (*a*) An example scenario during bee training, with green indicating a low reward (a 30% sucrose solution) in the right chamber, and blue indicating a high reward (a 50% sucrose solution) in the left chamber. In both cases, the other chamber would contain only distilled water. The association between colour, reward and location was counterbalanced across trials. The test phase consisted of five trials in which different colours were presented on the screen in a pseudorandom order (cue values: L, low; NL, near low; M, medium; NH, near high; H, high). (*b*) Latency to approach high and low rewards at the end of training. (*c*) Response latency to each of the five cues during tests. Each colour represents a different day and the black line is the average across all days. The error bars represent the standard error of the mean. (*d*,*e*) Histograms of repeatability estimates (*r*) of choice latency for (*d*) bee identity and (*e*) fixed factors. The histograms represent the distribution of *R* values generated from 1000 bootstrap iterations. The *R* value represents the proportion of total variance explained by individual identity (ID) in (*d*) and by the fixed effect included in the model in (*e*). In both plots, the dot and dashed line represent the point estimate. For bee identity (*d*), the estimate ($R = 0.34$) indicates that approximately 34% of the variation in speed was due to consistent differences between individual bees. Similarly, the estimate for fixed effects ($R = 0.20$) indicates that approximately 20% of the variance in choice latency was attributable to the fixed factor included in the model.

(e) Judgement bias testing across consecutive days

The judgement bias test was carried out on the following two days as described in §2d. Each session began with four refresher trials (two presentations of each conditioned colour with its reward). If a bee made an incorrect choice, the steps described in §2d were followed until the bee reached the learning criteria. After the refresher trials, judgement bias tests were carried out as in §2d. From the initial pool of 24 bees, four did not complete all tests because they died. Specifically, one bee completed only the first test and three bees completed two out of the three planned tests.

Downloaded from royalsocietypublishing.org/rsbl/article-pdf/doi/10.1098/rsbl.2026.0011/6145623/rsbl.2026.0011.pdf
by Newcastle University user
on 19 November 2026

(f) Statistical analyses

Our hypothesis and statistical analyses were preregistered at aspredicted.org (https://aspredicted.org/zp8qu4.pdf). The data were plotted and analysed using RStudio v. 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). Subsequent statistical models were fitted by maximum-likelihood estimation. Models were compared using the 'model.sel' function in the MuMIn package [23] and the model with the lowest Akaike information criterion (AIC) score was selected as the best model. We used the package DHARMa [29] for residual testing of all models. The *p*-value of each factor was derived using the 'drop1' function [30] and *post hoc* analyses were performed using 'emmeans' function [12].

Latency data in all analyses were log-transformed and latencies greater than 1.5 times the interquartile range were excluded (18 of 120 datapoints for training, 46 of 300 for tests). Choice latency for the last two training bouts was analysed using a linear mixed-effects model (lmer function, lme4 package), including only bees that completed all three tests [13]. The model included 'colour' (categorical: high or low), 'training day' (categorical: 1, 2 and 3) and their interaction, 'age', 'high colour' (categorical: blue or green), 'high side' (categorical: right or left, position of the high reward) and 'colony' (categorical: 1, 2 and 3) as fixed factors. Bee identity ('ID') was included as a random intercept variable.

The model analysing choice latency for the judgement bias test included 'cue' (categorical: high, near high, medium, near low and low), 'testing day' (categorical: 1, 2 and 3) and their interaction, 'age', 'high colour side' (categorical: blue/right, blue/left, green/right and green/left) and 'colony' (categorical: 1, 2 and 3) as fixed factors. Bee identity ('ID') was included as a random intercept variable.

To assess the consistency of individual differences in latency to choose across days and cues, we estimated repeatability using the rptR package [14]. The model included the same fixed factor with individual identity ('ID') specified as a random effect with random slopes for 'cue'. Repeatability was calculated as the proportion of total variance in choice latency explained by differences between individuals. Confidence intervals were obtained using 1000 parametric bootstrap iterations. The variance explained by fixed effects was also estimated to quantify their contribution relative to individual consistency.

Choices of reward chambers previously associated with high- and low-value cues were coded as 1 or 0, respectively. The choices of high-reward chambers were classified as optimistic, and choices of low-reward chambers as pessimistic. We analysed the probability of an optimistic choice with a generalized linear mixed-effect model (glmer function, lme4 package) with binomial distribution and a logit link function. The fixed factors were 'cue' (categorical: high, near high, medium, near low and low), 'testing day' (categorical: 1, 2 and 3) and their interaction, 'age', order (categorical: 1, 2, 3, 4, 5 representing the sequence of colour cue presentations) and 'colony' (categorical: 1, 2 and 3). Bee identity ('ID') was included as a random intercept variable. Repeatability of individual choices was estimated using the rptR package [14]. Due to lack of convergence using the model described, repeatability was assessed using a model which included only the random intercept for bee identity and the binary response variable. Repeatability was calculated as the proportion of total variance explained by differences between individuals, and confidence intervals were obtained via 1000 bootstrap iterations. This analysis provided an estimate of how consistent bees were in their optimistic choices across trials.

3. Results

(a) Training

During training, bumblebees achieved the learning criterion and continued to the judgement bias test. The best model (AIC = 134.6) for choice latency data included the main effects of colour cue (GLMM, $\chi^2_{(1)} = 37.20$, $p = 3.386 \times 10^{-8}$). Bees were significantly faster to choose a high-reward colour compared with a low-reward colour in the last training choice (estimate ± s.e. = –0.50 ± 0.08, $t = -6.05$, $p < 0.0001$; figure 1*b*). The difference in latencies indicated that the bees could differentiate between both the colour cues and the rewards. Including the full dataset, which contains latencies exceeding 1.5 times the interquartile range, produced similar results (electronic supplementary material, table S2).

(b) Judgement bias tests

The best model (AIC = 346.1) for the test choice latency data included the main effect of cue. Cue significantly affected bee latency to choose a chamber (GLMM, $\chi^2_{(4)} = 22.89$, $p = 5.497 \times 10^{-16}$; figure 1*c*), with bees showing progressively longer latencies for colours closer to that associated with lower reward (high versus medium: estimate ± s.e. = –0.33 ± 0.08, $t = -4.00$, $p = 0.0008$; high versus near low: –0.55 ± 0.08, $t = -6.54$, $p < 0.0001$; high versus low: –0.56 ± 0.08, $t = -6.35$, $p < 0.0001$; see electronic supplementary material for additional details). Including the full dataset, which includes latencies exceeding 1.5 times the interquartile range, showed similar results (electronic supplementary material, table S3). Repeatability analysis indicated that individual bees were moderately and significantly consistent in their latency to choose ($R = 0.34 \pm 0.07$, 95% CI = 0.20 – 0.49, $p = 1.23 \times 10^{-7}$; figure 1*d*), reflecting stable individual differences (see electronic supplementary material for results including the full dataset). Fixed effects explained a comparable proportion of variance ($R = 0.20 \pm 0.07$, 95% CI = 0.08 – 0.37; figure 1*e*). Bees' latencies to choose across testing days for each cue are plotted in figure 2*a*.

The best model (AIC = 214.7) for reward chamber choice included the main effect of cue. Cue significantly influenced the likelihood of bees making an optimistic choice (GLMM, $\chi^2_{(4)} = 206.72$, $p = 2.2 \times 10^{-16}$; figure 2*b*). Bees were progressively less likely to respond optimistically to cues further from the high cue. They made more optimistic responses to the near high cue than the near low (estimate ± s.e. = 4.30 ± 0.59, $z = 7.20$, $p < 0.0001$) and low cues (4.61 ± 0.62, $z = 7.35$, $p < 0.0001$; see electronic

Downloaded from royalsocietypublishing.org/rsbl/article-pdf/doi/10.1098/rsbl.2026.0011/6145623/rsbl.2026.0011.pdf
by Newcastle University user
on 19 November 2026

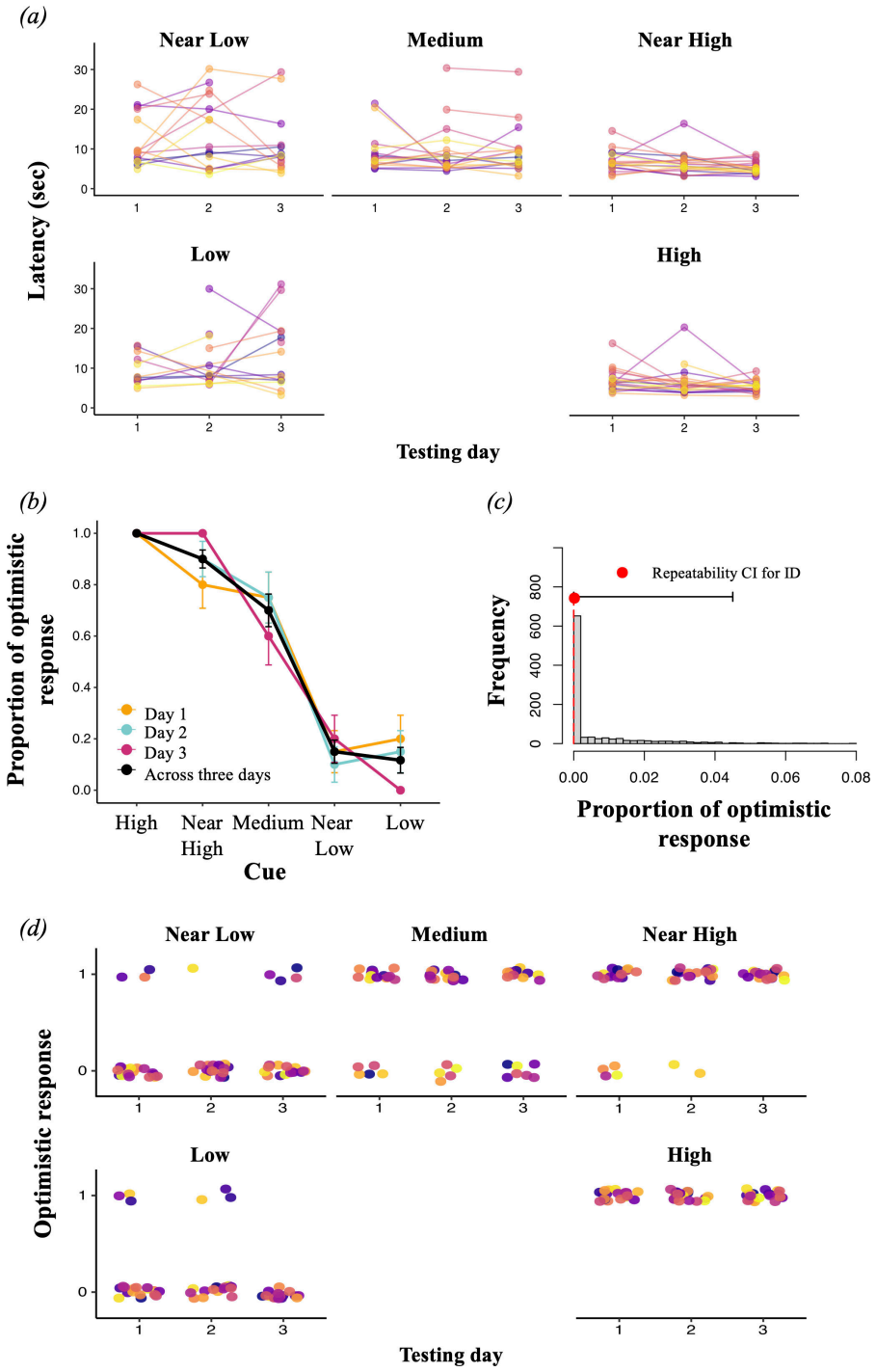

Figure 2. Bee choice analyses. (*a*) Individual bee choice latencies in tests across testing days for each cue. Each colour and line depicts a different individual. (*b*) The proportion of bees that made an optimistic choice in response to each of the five cues. Other details as in figure 1*d*. (*c*) Histogram of repeatability estimates of the optimistic response. (*d*) Individual bee optimistic responses across testing days for each cue.

supplementary material for additional details). Bees always responded optimistically to the high cue. Repeatability analysis across all cues indicated no detectable variance attributable to individual identity ($R = 0$, s.e. = 0.013, 95% CI = 0–0.044, $p = 1$; figure 2*c*). We therefore excluded the conditioned cues, leaving only the ambiguous ones (near high, medium, low). However, the repeatability remained negligible ($R = 0$, s.e. = 0.022, 95% CI = 0–0.078, $p = 0.5$). These results indicate that bees responded similarly to each cue, leaving little inter-individual variation for the model to detect (figure 2*d*). To further support this, we calculated the coefficient of variation (CV) for responses to each cue, which showed that bees' responses were highly uniform for the conditioned extreme cues (high: mean ± s.d.; CV = 1 ± 0, CV = 0; low: 0.10 ± 0.31, CV = 2.94;) and moderately variable for ambiguous cues (near high: 0.89 ± 0.31, CV = 0.34; medium: 0.70 ± 0.46, CV = 0.65; near low: 0.14 ± 0.35, CV = 2.50).

4. Discussion

We tested the hypothesis that a judgement bias reflects a stable individual trait rather than a transient affective state and that responses to cues would therefore remain consistent across multiple tests. Our repeatability analyses of behaviour in judgement

Downloaded from royalsocietypublishing.org/rsbl/article-pdf/doi/10.1098/rsbl.2026.0011/6145623/rsbl.2026.0011.pdf
by Newcastle University user
on 19 November 2026

bias tests revealed moderate consistency in individual latency to choose but negligible repeatability in the probability of an optimistic choice in response to ambiguous stimuli.

An important aspect in judgement bias paradigms, especially in go/no-go tasks, is that subjects often learn that ambiguous cues are not rewarded, which can progressively reduce their ambiguity and lead to increased latency and number of no responses across repeated sessions [9,15–17]. Secondary or partial reinforcement has been suggested to mitigate the loss of ambiguity [18–22]. However, we argue that the use of active choice tests is a convenient and effective method to avoid learning effects and other confounding factors. We found no interaction between cue and testing time for either latency or optimistic choice, indicating that bees' responses to learnt and ambiguous cues remained stable over time. This suggests that repeated exposure did not lead bees to learn that the ambiguous cues were unrewarded, maintaining their uncertainty across days. The persistence of graded responses further supports the idea that the active choice task, along with the refresher trials performed during the test, were effective in preserving cue ambiguity. These results demonstrate the potential benefit of using an active choice task for assessing judgement bias over repeated tests, as it reduces the confounding effects of learning and motivation often reported in go/no-go tasks [28].

In our study, bees showed moderate repeatability in their latency to choose across cues (R = 0.34), with some individuals being consistently faster or slower across repetitions. Such repeatability aligns with the level of consistency found in literature, where typically 37% of individual variation is attributed to stable individual differences [24]. This suggests that choice latency may underpin a stable trait supporting the covariation hypothesis between cognition and personality [25–27,31]. Differences in personality traits may underpin emotional states, and this relationship can be bidirectional. Individuals with a specific personality type can be more prone to experience positive or negative affective states, leading to corresponding cognitive biases. At the same time, repeated exposure to negative or positive experiences may change these traits. The observed stability in choice latency suggests that while decision speed may be driven by stable metabolic and physiological constraints, optimistic choice (i.e. judgement remains flexible [32,33]). This distinction is ecologically important as it allows bees to maintain consistent activity levels while modulating their evaluation of ambiguous cues in response to fluctuating rewards. To further validate the judgement bias test, it would be valuable to include additional behavioural measures that are hypothesized to covary with emotional states.

Repeatability of optimistic choices for ambiguous cues was negligible, suggesting that the likelihood of making an optimistic choice fluctuated across the three repetitions, whereas responses to the trained cues, based on the coefficient of variation, were highly consistent. In contrast to latency, optimistic responses appear to be largely state-dependent and influenced by situational factors rather than stable individual traits. The absence of consistent individual biases indicates that bees were unbiased in the absence of any experimental manipulation. Therefore, when a manipulation elicits a detectable shift in active choices in a judgement bias test, such a change is more likely to reflect state-dependent variation rather than a stable individual difference, supporting an interpretation of them as a shift in affective states. For instance, we found that the variability in the likelihood of making an optimistic choice across three testing days was lower than the variation observed when bumble bees were shaken or trapped in a comparable experimental set-up [11]. This suggests that variability in the absence of manipulation is substantially lower than when the animals' affective state is experimentally altered.

In conclusion, our findings demonstrate that repeated testing can be performed without confounding effects of learning and support the use of active choice tasks to assess transient and stable emotional states in invertebrates.

Ethics. Studies on Hymenoptera are not regulated under the Animals (Scientific Procedures) Act 1986 (ASPA). See electronic supplementary material for additional details.

Data accessibility. All data and R code used for analysis are provided as supplementary materials [34].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. L.B.: conceptualization, data curation, formal analysis, methodology, visualization, writing—original draft, writing—review and editing; V.N.: conceptualization, funding acquisition, methodology, project administration, resources, supervision, validation, writing—original draft, writing—review and editing.

Both authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. L.B. and this research are supported by a Leverhulme Trust Research Project Grant (RPG-2021-358).

References

1. Blanchette I, Richards A. 2010 The influence of affect on higher level cognition: a review of research on interpretation, judgement, decision making and reasoning. *Cogn. Emot*. **24**, 561–595. (doi:10.1080/02699930903132496)
2. Paul ES, Harding EJ, Mendl M. 2005 Measuring emotional processes in animals: the utility of a cognitive approach. *Neurosci. Biobehav. Rev*. **29**, 469–491. (doi:10.1016/j.neubiorev.2005.01.002)
3. Beevers CG, Mullarkey MC, Dainer-Best J, Stewart RA, Labrada J, Allen JJB, McGeary JE, Shumake J. 2019 Association between negative cognitive bias and depression: a symptom-level approach. *J. Abnorm. Psychol*. **128**, 212–227. (doi:10.1037/abn0000405)
4. Vos LMW, Nieto I, Amanvermez Y, Smeets T, Everaert J. 2025 Do cognitive biases prospectively predict anxiety and depression? a multi-level meta-analysis of longitudinal studies. *Clin. Psychol. Rev*. **116**, 102552. (doi:10.1016/j.cpr.2025.102552)
5. Lagisz M, Zidar J, Nakagawa S, Neville V, Sorato E, Paul ES, Bateson M, Mendl M, Løvlie H. 2020 Optimism, pessimism and judgement bias in animals: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev*. **118**, 3–17. (doi:10.1016/j.neubiorev.2020.07.012)
6. Perry CJ, Baciadonna L. 2017 Studying emotion in invertebrates: what has been done, what can be measured and what they can provide. *J. Exp. Biol*. **220**, 3856–3868. (doi:10.1242/jeb.151308)

Downloaded from royalsocietypublishing.org/rsbl/article-pdf/doi/10.1098/rsbl.2026.0011/6145623/rsbl.2026.0011.pdf
by Newcastle University user
on 19 November 2026

7. Harding EJ, Paul ES, Mendl M. 2004 Cognitive bias and affective state. *Nature* **427**, 312–312. (doi:10.1038/427312a)
8. Baciadonna L, McElligott A. 2015 The use of judgement bias to assess welfare in farm livestock. *Anim. Welf*. **24**, 81–91. (doi:10.7120/09627286.24.1.081)
9. Roelofs S, Boleij H, Nordquist RE, van der Staay FJ. 2016 Making decisions under ambiguity: judgment bias tasks for assessing emotional state in animals. *Front. Behav. Neurosci*. **10**, 119. (doi:10.3389/fnbeh.2016.00119)
10. Faustino AI, Oliveira GA, Oliveira RF. 2015 Linking appraisal to behavioral flexibility in animals: implications for stress research. *Front. Behav. Neurosci*. **9**, 104. (doi:10.3389/fnbeh.2015.00104)
11. Neville V, Nakagawa S, Zidar J, Paul ES, Lagisz M, Bateson M, Løvlie H, Mendl M. 2020 Pharmacological manipulations of judgement bias: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* **108**, 269–286. (doi:10.1101/612382)
12. Lenth RV. 2016 Least-squares means: the R package lsmeans. *J. Stat. Soft.* **69**, 1–33.
13. Bates D, Mächler M, Bolker B, Walker S. 2015 Fitting linear mixed-effects models using lme4. *J. Stat. Softw*. **67**, 01. (doi:10.18637/jss.v067.i01)
14. Stoffel MA, Nakagawa S, Schielzeth H. 2017 rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol*. **8**, 1639–1644. (doi:10.1111/2041-210x.12797)
15. Murphy E, Nordquist RE, van der Staay FJ. 2013 Responses of conventional pigs and Göttingen miniature pigs in an active choice judgement bias task. *Appl. Anim. Behav. Sci*. **148**, 64–76. (doi:10.1016/j.applanim.2013.07.011)
16. Wilson C, Hall N, Aviles-Rosa EO, Campbell K, Arnott G, Reeve C. 2023 The effect of repeated testing on judgement bias in domestic dogs (*Canis familiaris*). *Anim. Cogn*. **26**, 477–489. (doi:10.1007/s10071-022-01689-3)
17. Doyle RE, Vidal S, Hinch GN, Fisher AD, Boissy A, Lee C. 2010 The effect of repeated testing on judgement biases in sheep. *Behav. Process*. **83**, 349–352. (doi:10.1016/j.beproc.2010.01.019)
18. Neave HW, Daros RR, Costa JHC, von Keyserlingk MAG, Weary DM. 2013 Pain and pessimism: dairy calves exhibit negative judgement bias following hot-iron disbudding. *PLoS One* **8**, e80556. (doi:10.1371/journal.pone.0080556)
19. Keen HA, Nelson OL, Robbins CT, Evans M, Shepherdson DJ, Newberry RC. 2014 Validation of a novel cognitive bias task based on difference in quantity of reinforcement for assessing environmental enrichment. *Anim. Cogn*. **17**, 529–541. (doi:10.1007/s10071-013-0684-1)
20. Bethell E, Holmes A, MacLarnon A, Semple S. 2012 Cognitive bias in a non-human primate: husbandry procedures influence cognitive indicators of psychological well-being in captive rhesus macaques. *Anim. Welf*. **21**, 185–195. (doi:10.7120/09627286.21.2.185)
21. Bateson M, Emmerson M, Ergün G, Monaghan P, Nettle D. 2015 Opposite effects of early-life competition and developmental telomere attrition on cognitive biases in juvenile European starlings. *PLoS One* **10**, e0132602. (doi:10.1371/journal.pone.0132602)
22. Bateson M, Matheson S. 2007 Performance on a categorisation task suggests that removal of environmental enrichment induces 'pessimism' in captive European starlings (*Sturnus vulgaris*). *Anim. Welf.* **16**, 33–36. (doi:10.1017/S0962728600031705)
23. Barton K. 2009 *Mu-min: multi-model inference*. R-Forge. See http://R-Forge.R-project.org/projects/mumin/.
24. Bell AM, Hankison SJ, Laskowski KL. 2009 The repeatability of behaviour: a meta-analysis. *Anim. Behav*. **77**, 771–783. (doi:10.1016/j.anbehav.2008.12.022)
25. Lucon-Xiccato T, Carere C, Baracchi D. 2024 Intraspecific variation in invertebrate cognition: a review. *Behav. Ecol. Sociobiol.* **78**. (doi:10.1007/s00265-023-03413-8)
26. Dougherty LR, Guillette LM. 2018 Linking personality and cognition: a meta-analysis. *Phil. Trans. R. Soc. B* **373**, 20170282. (doi:10.1098/rstb.2017.0282)
27. Sih A, Del Giudice M. 2012 Linking behavioural syndromes and cognition: a behavioural ecology perspective. *Phil. Trans. R. Soc. B* **367**, 2762–2772. (doi:10.1098/rstb.2012.0216)
28. Procenko O, Read J, Nityananda V. 2024 Physically stressed bees expect less reward in an active choice judgement bias test. *Proc. R. Soc. B* **291**, 20240512. (doi:10.1101/2023.10.06.561175)
29. Harting F. 2020 DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. *R package version 0.3.3*. See https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html.
30. Barr DJ, Levy R, Scheepers C, Tily HJ. 2013 Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang*. **68** 255–278. (doi:10.1016/j.jml.2012.11.001)
31. Carere C, Locurto C. 2011 Interaction between animal personality and animal cognition. *Curr. Zool*. **57**, 491–498. (doi:10.1093/czoolo/57.4.491)
32. Burns JG. 2005 Impulsive bees forage better: the advantage of quick, sometimes inaccurate foraging decisions. *Anim. Behav*. **70**, e1–e5. (doi:10.1016/j.anbehav.2005.06.002)
33. Vasilieva NA. 2022 Pace-of-life syndrome (POLS): evolution of the concept. *Biol. Bull. Russ. Acad. Sci.* **49**, 750–762. (doi:10.1134/S1062359022070238)
34. Luigi B, Nityananda V. 2026 Supplementary material from: Temporal consistency of judgement biases in bumblebees. Figshare. (doi:10.6084/m9.figshare.c.8512245)

Downloaded from royalsocietypublishing.org/rsbl/article-pdf/doi/10.1098/rsbl.2026.0011/6145623/rsbl.2026.0011.pdf
by Newcastle University user
on 19 November 2026